

Fay Huang
Reen-Cheng Wang (Eds.)



30

Arts and Technology

First International Conference, ArtsIT 2009
Yi-Lan, Taiwan, September 2009
Revised Selected Papers



 Springer

The Springer logo features a stylized chess knight icon to the left of the word 'Springer' in a serif font.

Lecture Notes of the Institute
for Computer Sciences, Social-Informatics
and Telecommunications Engineering

30

Editorial Board

Ozgur Akan

Middle East Technical University, Ankara, Turkey

Paolo Bellavista

University of Bologna, Italy

Jiannong Cao

Hong Kong Polytechnic University, Hong Kong

Falko Dressler

University of Erlangen, Germany

Domenico Ferrari

Università Cattolica Piacenza, Italy

Mario Gerla

UCLA, USA

Hisashi Kobayashi

Princeton University, USA

Sergio Palazzo

University of Catania, Italy

Sartaj Sahni

University of Florida, USA

Xuemin (Sherman) Shen

University of Waterloo, Canada

Mircea Stan

University of Virginia, USA

Jia Xiaohua

City University of Hong Kong, Hong Kong

Albert Zomaya

University of Sydney, Australia

Geoffrey Coulson

Lancaster University, UK

Fay Huang Reen-Cheng Wang (Eds.)

Arts and Technology

First International Conference ArtsIT 2009

Yi-Lan, Taiwan, September 24-25, 2009

Revised Selected Papers



Springer

Volume Editors

Fay Huang
National Ilan University
Institute of Computer Science
and Information Engineering
No. 1, Sec. 1, Shen-Lung Rd.
Yi-Lan, 26047, Taiwan, R.O.C
E-mail: fay@niu.edu.tw

Reen-Cheng Wang
National Dong Hwa University
Department of Computer Science
and Information Engineering
No. 1, Sec.2, Da Hsueh Rd.
Shoufeng Hualien 97401, Taiwan, R.O.C.
E-mail: rcwang@mail.ndhu.edu.tw

Library of Congress Control Number: 2009942770

CR Subject Classification (1998): I.3, I.4.8, I.6, I.3.7, C.3, J.5, K.8.1

ISSN 1867-8211
ISBN-10 3-642-11576-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-11576-9 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© ICST Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12833696 06/3180 5 4 3 2 1 0

Preface

We welcome you to the First International Conference on Arts and Technology (ArtsIT 2009), hosted by CSIE of the National Ilan University and co-organized by the National Science Council, ICST, College of EECS at National Ilan University, Software Simulation Society in Taiwan, ISAC, TCA, NCHC, CREATE-NET, and Institute for Information Industry. ArtsIT2009 was held in Yilan, Taiwan, during September 24–25, 2009.

The conference comprised the following themes:

- **New Media Technologies** (Evolutionary systems that create arts or display art works, such as tracking sensors, wearable computers, mixed reality, etc.)
- **Software Art** (Image processing or computer graphics techniques that create arts, including algorithmic art, mathematic art, advanced modeling and rendering, etc.)
- **Animation Techniques** (2D or 3D computer animations, AI-based animations, etc.)
- **Multimedia** (Integration of different media, such as virtual reality systems, audio, performing arts, etc.)
- **Interactive Methods** (Vision-based tracking and recognition, interactive art, etc.)

The conference program started with an opening ceremony, followed by three keynote speeches and four technical sessions distributed over a period of two days. Two poster sessions, one hour each, were scheduled before the afternoon oral sessions. An Interactive Arts Exhibition was held in conjunction with ArtsIT 2009. Twelve well-known digital arts teams from Taiwan exhibited 15 artworks in this event, including 10 interactive installation arts, 4 video arts, and 1 digital print.

The conference received around 50 submissions from 15 different countries. Each of these submissions was evaluated by a minimum of two reviewers. In the end, there were 18 papers accepted for oral presentation and 14 for poster presentation plus 1 invited paper. The overall paper acceptance rate was about 63%.

We are deeply grateful to all the Organizing Committee members for their great efforts in making this a successful event. We would like to acknowledge the hosts of the conference, led by the Local Chair Meng-Hsueh Chiang, and thank them for their hospitality and hard work, especially Prof. Chiang for his thoughtful coordination. We were very fortunate to have Hao-Chiang Lin as our Exhibition Chair, and extremely grateful for his enthusiasm. The Interactive Art Exhibition really added great value to the conference. We would like to express our gratitude to the 35 Technical Program Committee members and the 8 additional reviewers for sharing with us their knowledge and experiences. Furthermore, we would like to acknowledge the generosity of our sponsors: Intel, CISCO, Stark Technology Inc., BDTC Technology Co. Ltd., AceNet Technology, Chief Telecom Inc., DMAX Technology Ltd., Weblink International Inc., and Tatung Co.,

and thank them for their kind support. Finally, we would like to express our most sincere thanks to the three keynote speakers, Takeo Igarashi, Reinhard Klette, and David G. Stork, for their inspirational speeches.

We hope you find these conference proceedings enjoyable and enlightening.

Han-Chieh Chao
Fay Huang
Reen-Cheng Wang

Organization

Steering Committee Chair

Imrich Chlamtac Create-Net

General Co-chairs

Han-Chieh Chao National Ilan University, Taiwan
Athanasios Vasilakos University of Western Macedonia, Greece

Program Co-chairs

Su-Chu Hsu Taipei National University of Arts, Taiwan
Fay Huang National Ilan University, Taiwan
Tai-Fang Pan National Taiwan University of Arts, Taiwan
Reen-Cheng Wang National Dong Hwa University, Taiwan

Conference Coordinator

Maria Morozova ICST

Exhibition Chair

Hao Chiang (Koong) Lin National University of Tainan, Taiwan

Finance Chair

Po Zone Wu National Ilan University, Taiwan

Local Chair

Meng Hsueh Chiang National Ilan University, Taiwan

Publicity Chair

Yun-Sheng Yen Fo Guang University, Taiwan

Technical Program Committee

Joost Batenburg	University of Antwerp, Belgium
Tony Brooks	Aalborg University Esbjerg, Denmark
Pai-Ling Chang	Shih Hsin University, Taiwan
Chin-Chen Chang	National United University, Taiwan
Chun-Fa Chang	National Taiwan Normal University, Taiwan
Chia-Yen Chen	National University of Kaohsiung, Taiwan
Chu-Yin Chen	University Paris 8, France
Lieu-Hen Chen	National Chi Nan University, Taiwan
Yung-Yu Chuang	National Taiwan University, Taiwan
Stuart Eleanor Gates	Australian National University, Australia
Magy Seif El-Nasr	Simon Fraser University, Canada
Paul Fishwick	University of Florida, USA
Chiou-Shann Fuh	National Taiwan University, Taiwan
Oliver Grau	Danube University, Austria
Jun Hu	Eindhoven University of Technology, The Netherlands
Shyh-Kang Jeng	National Taiwan University, Taiwan
Ruyi Jiang	Shanghai Jiao Tong University, China
Bin-Shyan Jong	Chung Yuan Christian University, Taiwan
Tong-Yee Lee	National Cheng Kung University, Taiwan
Wen-Pin Hope Lee	National Taiwan Normal University, Taiwan
Chia-Hsiang Lee	National Taipei University of Technology, Taiwan
Rung-Huei Liang	National Taiwan University of Science and Technology, Taiwan
I-Chen Lin	National Chiao Tung University, Taiwan
Wen-Chieh Lin	National Chiao Tung University, Taiwan
Karen Lo	National Chiao Tung University, Taiwan
Charalampos Z. Patrikakis	National Technical University of Athens, Greece
Sara Owsley Sood	Pomona College, USA
Wen-Kai Tai	National Dong Hwa University, Taiwan
Ming-Chang Tien	National Dong Hwa University, Taiwan
Shyh-Kuang Ueng	National Taiwan Ocean University, Taiwan
Tobi Vaudrey	The University of Auckland, New Zealand
Ching-Sheng Wang	Aletheia University, Taiwan
Dai-Yi Wang	Providence University, Taiwan
Neal Naixue Xiong	Georgia State University, USA
Chii-Zen Yu	Toko University, Taiwan

Table of Contents

Keynote Speeches

From Digital Imaging to Computer Image Analysis of Fine Art	1
<i>David G. Stork</i>	
Panoramic and 3D Computer Vision	9
<i>Akihiko Torii and Reinhard Klette</i>	

Full Papers

Meaningful Engagement: Computer-Based Interactive Media Art in Public Space	17
<i>Jiun-Jhy Her and Jim Hamlyn</i>	
Interactive WSN-Bar	25
<i>Jiun-Shian Lin, Su-Chu Hsu, and Ying-Chung Chen</i>	
Immersive Painting	33
<i>Stefan Soutschek, Florian Hoenig, Andreas Maier, Stefan Steidl, Michael Stuermer, Hellmut Erzigkeit, Joachim Hornegger, and Johannes Kornhuber</i>	
Sonic Onyx: Case Study of an Interactive Artwork	40
<i>Salah Uddin Ahmed, Letizia Jaccheri, and Samir M'kadmi</i>	
An Interactive Concert Program Based on Infrared Watermark and Audio Synthesis	48
<i>Hsi-Chun Wang, Wen-Pin Hope Lee, and Feng-Ju Liang</i>	
Butterfly Effect Fractal	56
<i>Yin-Wei Chang and Fay Huang</i>	
Pattern Formation in Networks Inspired by Biological Development	64
<i>Kei Ohnishi, Kaori Yoshida, and Mario Köppen</i>	
A Watercolor NPR System with Web-Mining 3D Color Charts	72
<i>Lieu-Hen Chen, Yi-Hsin Ho, Ting-Yu Liu, and Wen-Chieh Hsieh</i>	
Gas and Shadow Swing	80
<i>Chi-Hung Tsai, Mei-Yi Lai, Che-Wei Liu, Shiang-Yin Huang, Che-Yu Lin, and Jeng-Sheng Yeh</i>	
Artist-Oriented Real-Time Skin Deformation Using Dynamic Patterns	88
<i>Masaki Oshita and Kenji Suzuki</i>	

Interaction Analysis in Performing Arts: A Case Study in Multimodal Choreography	97
<i>Maria Christou and Annie Luciani</i>	
BWAIN: An Artistic Web Interpretation	105
<i>Linda Huber, Michael Bißmann, Stefan Gottschalk, Alexander Keck, Andreas Schönefeldt, Marko Seidenglanz, Norbert Sroke, and Alexander Radke</i>	
Treasure Transformers: Novel Interpretative Installations for the National Palace Museum	112
<i>Chun-Ko Hsieh, I-Ling Liu, Quo-Ping Lin, Li-Wen Chan, Chuan-Heng Hsiao, and Yi-Ping Hung</i>	
JacksonBot – Design, Simulation and Optimal Control of an Action Painting Robot	120
<i>Michael Raschke, Katja Mombaur, and Alexander Schubert</i>	
RF Sounding: A System for Generating Sounds from Spectral Analysis	128
<i>Fabio Graziosi, Claudia Rinaldi, and Francesco Tarquini</i>	
Depicting Time Evolving Flow with Illustrative Visualization Techniques	136
<i>Wei-Hsien Hsu, Jianqiang Mei, Carlos D. Correa, and Kwan-Liu Ma</i>	
Augmenting a Ballet Dance Show Using the Dancer’s Emotion: Conducting Joint Research in Dance and Computer Science	148
<i>Alexis Clay, Elric Delord, Nadine Couture, and Gaël Domenger</i>	
Automatic Skin Color Beautification	157
<i>Chih-Wei Chen, Da-Yuan Huang, and Chiou-Shann Fuh</i>	
Tracking Small Artists	165
<i>James C. Russell, Reinhard Klette, and Chia-Yen Chen</i>	
RealSurf – A Tool for the Interactive Visualization of Mathematical Models	173
<i>Christian Stussak and Peter Schenzel</i>	
Creating Wheel-Thrown Potteries in Digital Space	181
<i>Gautam Kumar, Naveen Kumar Sharma, and Partha Bhowmick</i>	
Low-Level Image Processing for Lane Detection and Tracking	190
<i>Ruyi Jiang, Mutsuhiro Terauchi, Reinhard Klette, Shigang Wang, and Tobi Vaudrey</i>	
Lane Detection on the iPhone	198
<i>Feixiang Ren, Jinsheng Huang, Mutsuhiro Terauchi, Ruyi Jiang, and Reinhard Klette</i>	

Traditional Culture into Interactive Arts: The Cases of Lion Dance in Temple Lecture	206
<i>Wen-Hui Lee, Chih-Tung Chen, Ming-Yu He, and Tao-i Hsu</i>	
Improving Optical Flow Using Residual and Sobel Edge Images	215
<i>Tobi Vaudrey, Andreas Wedel, Chia-Yen Chen, and Reinhard Klette</i>	
3 Case Studies: A Hybrid Educational Strategy for ART/SCI Collaborations	223
<i>Elif Ayiter, Selim Balcisoy, and Murat Germen</i>	
A Multimedia, Augmented Reality Interactive System for the Application of a Guided School Tour	231
<i>Ko-Chun Lin, Sheng-Wen Huang, Sheng-Kai Chu, Ming-Wei Su, Chia-Yen Chen, and Chi-Fa Chen</i>	
Reconstruction of Cultural Artifact Using Structured Lighting with Densified Stereo Correspondence	239
<i>H.-J. Chien, C.-Y. Chen, and C.-F. Chen</i>	
Motion Generation for Glove Puppet Show with Procedural Animation	247
<i>Chih-Chung Lin, Gee-Chin Hou, and Tsai-Yen Li</i>	
Can't See the Forest: Using an Evolutionary Algorithm to Produce an Animated Artwork	255
<i>Karen Trist, Vic Ciesielski, and Perry Barile</i>	
Automatic Generation of Caricatures with Multiple Expressions Using Transformative Approach	263
<i>Wen-Hung Liao and Chien-An Lai</i>	
The Autonomous Duck: Exploring the Possibilities of a Markov Chain Model in Animation	272
<i>Javier Villegas</i>	
MixPlore: A Cocktail-Based Media Performance Using Tangible User Interfaces	279
<i>Zune Lee, Sungkyun Chang, and Chang Young Lim</i>	
Author Index	291

From Digital Imaging to Computer Image Analysis of Fine Art

David G. Stork^{1,2}

¹ Ricoh Innovations, 2882 Sand Hill Road Suite 115, Menlo Park, CA 94025 USA
artanalyst@gmail.com

www.diatrope.com/stork/FAQs.html

² Department of Statistics, Stanford University, Stanford CA 94305 USA

Abstract. An expanding range of techniques from computer vision, pattern recognition, image analysis, and computer graphics are being applied to problems in the history of art. The success of these efforts is enabled by the growing corpus of high-resolution multi-spectral digital images of art (primarily paintings and drawings), sophisticated computer vision methods, and most importantly the engagement of some art scholars who bring questions that may be addressed through computer methods. This paper outlines some general problem areas and opportunities in this new inter-disciplinary research program.

Keywords: Computer vision, image analysis, fine art, art history, perspective analysis, color analysis, authentication.

1 Introduction

Ever since the discovery of x-rays, imaging has been used in the service of art scholarship, generally to reveal invisible, partially obscured, or illegible images. Particularly important in this development is infra-red reflectography, which uses long-wavelength radiation to penetrate layers of paint to reveal underdrawings and *pentimenti* (brush strokes or forms that have been changed or painted over) in paintings. Likewise multi-spectral imaging can reveal properties of paint and media that otherwise elude the unaided human eye. Such techniques have been invaluable for understanding artists' aesthetic decision and working methods and for dating and authenticating works. In these and nearly all other cases, the revealed image is interpreted by art scholars. A great success in this imaging and image processing tradition is the work decoding an important lost text by Archimedes in the *Archimedes palimpsest*. [\[8\]\[16\]](#)

The large number of high-resolution multi-spectral images of art, and the growing body of rigorous techniques from computer vision and computer graphics set the stage for a new era in this interdisciplinary field: true computer analysis, where computer algorithms detect patterns or properties in an image that are too subtle, diffuse or of a type difficult for humans to detect or interpret.

Computer vision—the discipline seeking to make computers “see”—has addressed a wide range of problems in robotics, forensic image analysis, automated

inspection, biometrics, security, remote sensing, automated lipreading, optical character recognition, document analysis, and so on. Its many successes include algorithms for object recognition, shape recognition, shape-from-shading, color analysis, lighting analysis, perspective analysis, and many others. Of course, the application of these methods in technical areas depends greatly upon the problem and this is especially true in their application to problems in the history of art. Scholars must be deeply familiar with the strengths and limitations of algorithms and the assumptions underlying their use, but also must bring deep knowledge of artist and the work in question, its cultural context, documentary evidence, and much more. These new computer methods do not *replace* other methods, such as the trained eye of an art historian or connoisseur, but rather extend and refine them.

This brief article seeks to provide an introduction for computer scientists to some of the more important classes of problems from art history that have benefitted—and are likely to continue to benefit—from the application of computer image analysis methods.

1.1 Authentication and Attribution

The technical distinction between *authentication* and *attribution* is that the task of authentication is to determine whether a particular, candidate artist created a given work, whereas the task of attribution is to determine who among a candidate *set* of artists likely created the work. (In some attribution studies, this set of artists might be large.) A subclass of this general problem domain is to determine the number of “hands” in a work, that is, the number of artists contributing to a given work, as was quite common in Renaissance workshops. For example, Andrea del Verrocchio collaborated with his young apprentice Leonardo on *Baptism of Christ* (c. 1472), where Verrocchio painted Christ and St. John and Leonardo painted the angels. Classical art scholarship approaches to such problems rely heavily on careful visual study and informed opinions of experts.

Computer vision methods may contribute to such efforts through the automatic recognition of subtle differences in artists’ styles (cf., Sect. 1.2). Of course which aspect of style is relevant depends upon the art work and artists in question. Statistical classifiers based on wavelet decompositions of brush strokes have been applied to identifying the number of hands in Perugino’s *Holy family*, [17] and a closely related method has been applied to the authentication of paintings by Vincent van Gogh. [11]

A related approach relies on sophisticated computational methods is the recent work on image-based authentication of Jackson Pollock’s drip paintings. (If one determines that a work is not by Pollock, the task of attribution is rarely of interest.) Richard Taylor and his colleagues were intrigued by the prospect that Pollock’s drip paintings might exhibit fractal or scale-space properties—visual features not previously considered by the art community and virtually impossible to infer “by eye.” This research group introduced a box-counting method for estimating such properties and reported that genuine Pollock paintings revealed a characteristic scale-space signature while fake paintings generally did not. [32]

This method has been criticized on a number of grounds, [13] but recent theoretical and empirical results show that such criticisms may not be warranted. [26,10] In particular, classifiers trained with scale-space features *as well as* other features, may indeed perform better than chance, and thus complement traditional methods for the authentication of such drip paintings.

1.2 Stylometry

Stylometry is the discipline of quantifying artistic style, and has been applied to a range of works and creators in music, literature, dance, and other modes of artistic expression. The particular aspects of style that are relevant depend on the class of art works and the problem at hand, and might involve color, brush stroke, composition and so forth, or even combinations of these properties. Stylometry nearly always exploits pattern recognition and machine learning techniques to infer subtle properties of style. [7] Stylometry underlies much of the work in authentication, but there are other applications as well. For example, one might consider just the color schemes or *palette* of an artist such as Vincent van Gogh as an aspect of his style and track quantitatively the variations in his palette through his career. Indeed, there are three main stages in his career—Netherlands, south of France, and Paris—each with a rather distinct palette. In this way, color stylometry can reveal objectively the effects in van Gogh's paintings of influences such as his response to colorful Japanese Ukiyo-e woodblock prints in Paris around 1886.

1.3 Dating of Works

Most of the technical methods for dating a work, such as a painting, are based on chemical and physical analyses of pigments, supports (canvas, wood panel, paper, parchment, ...) and knowledge of contemporary working methods, for instance preparation of supports, marking tools, and so on. Art historians also consider compositional style, the types of objects or events depicted, the influences of other artists or works whose dates are known, iconography and of course their expert visual analysis of style, brush strokes, colors, techniques, and so on.

A particularly intriguing automated image analysis approach to this problem applies to printed documents, such as maps printed using an intaglio process. [9] Each metal intaglio master plate is cleaned before each major print run, and such cleaning abrades the surface and thus affects the width of printed lines. Statistical models of such degradation, learned from optical scans of printed lines and textures from works of known dates, can then be applied to works of unknown date. Similar problems and opportunities arise in the dating of Japanese woodblock prints, where degradation of marks and lines, including breaks, indicate the wear on block masters and hence provide an estimate of age.

1.4 Predicting the Effects of Conservation Treatment

Cleaning a painting (eg., removing darkened varnish) is a very slow and expensive task; it is not unusual for the full cleaning of a large painting to take several

years and cost \$1M US or more. Curators, conservators and art scholars face the difficult task of deciding which of their works most deserve such cleaning and often their most important consideration is how the painting will appear after such cleaning—the appearance of colors, details that are revealed, and so on. Digital image processing methods can help predict how a given painting will appear after such cleaning. The methods involve modeling the optical physics of varnish as well as empirical statistical estimation from physical tests and small samples. [2]

1.5 Rejuvenating Faded Works

Some pigments are *fugitive*, that is, they fade over time; thus, some old master paintings we see today are faded versions of what the artist intended. Much as described in Sect. [1.4], one can develop computer models of such fading—based on experimental colorimetric evidence and chemical analysis. One then starts with the painting as it appears today and runs the fading model “backward” in time, to the date of the painting’s creation. In this way the algorithms “rejuvenate” the colors, as for example recent research rejuvenating the colors in Georges Seurat’s *Un dimanche après-midi à l’Île de la Grande Jatte* (1884–86). [1]

1.6 Inferring Artists’ Working Methods

An important set of problems in art scholarship is to infer artists’ studio conditions and working methods. The most direct class of such problems is understanding an artist’s marking tools and implements, for instance the nature of paint brushes, pencils, pastel sticks, conte crayons, or other markers. Statistical classifiers have been trained on features extracted direct from the image of lines and marks drawn with different markers, such as the shape of the ends of such lines. [15]

A somewhat less direct approach is used for testing whether an artist used a mechanical aid, such as a *Reductionszirkel* or *reducing compass* or even traced optically projected images. Such image analysis relies on the Chamfer metric or distance transform to measure the fidelity of copies. [22,6] A more sophisticated approach to testing for artists’ possible use of projected images is computer ray tracing. Scholars build models the putative projector and then explore matters such as the depth of field, blurriness, and overall brightness that would arise in the artist’s studio. [20]

Perspective is one of the richest sources of information in realist paintings. Sophisticated perspectival tests, image metrology and the estimation of perspective transformations have been used to test for artists’ use of projected images. [23,22,5] Moreover, algorithms for analyzing lighting in images have been used to infer the position of illuminants in tableaus, and thus test claims about lighting and possible use of optics. [30]

Johnson and his colleagues applied a number of lighting analysis techniques to Jan Vermeer’s *Girl with a pearl earring* (c. 1665) in order to infer the primary direction to the illuminant. The excellent agreement among the directions estimated from different sources of visual information (cast shadows, lightness along occluding contours, highlight reflections from the eye, ...) strongly imply



Fig. 1. Computer graphics rendering of Caravaggio's *The calling of St. Matthew* (1599–1600). [25][18] Such detailed computer graphics renderings of tableaus of paintings allow art scholars to visualize the work in new ways and explore hypothetical or “what if” scenarios to infer artists’ working methods. For instance, this model can reveal what would have been seen by each of the figures (eg., who sees Christ, who does not). Likewise, one can explore Caravaggio’s working methods, for instance the nature of the source of light in this tableau, specifically whether it is local artificial light or instead distance solar illumination. Computer experts and art scholars together thus simulate plausible studio conditions, always aware that artists are not “photographers” and that scholars must be careful to consider both the objective physical image in the studio and the artist’s perception and expressive ends. Specifically, this computer graphics model includes detailed models of the reflectance properties of the rear wall, expressed as a bidirectional reflectance distribution function (BRDF) that can be adjusted to best approximate the pattern of lightness found in the painting.

that Vermeer worked from a live model, rather than from his imagination. [12] One can infer the position of the illuminant based on the pattern of light on the rear wall in Caravaggio’s *The calling of St. Matthew* (Fig. 1) to determine whether it was distant solar illumination or nearby artificial illumination, a result that has important implications for the question of whether Caravaggio employed optical projections. [14][24]

Some of the most sophisticated algorithms ever applied to the analysis of images in art infer the complex pattern of illumination based on measured lightness along the outer or occluding boundaries of objects. Stork and Johnson used a method based on spherical harmonics to reveal that the lighting on different figures in Garth Herrick’s paintings differed significantly, thus showing that Herrick

worked under two different lighting conditions. In essence, Herrick “composited” one figure into his painting—a fact difficult to discern by eye. [31]

1.7 Rendering New Views for Visualization

Computer methods can provide new views into art works. Criminisi and his colleagues dewarped the images in spherical mirrors depicted in Renaissance paintings such as Jan van Eyck’s *Arnolfini portrait* (1434) and Robert Campin’s *Heinrich von Werl and St. John the Baptist* (1438), revealing new views into the respective tableaux. [4] Criminisi also developed uncalibrated methods to reconstruct three-dimensional models from two-dimensional artworks, including Masaccio’s *Trinità* (1427–28) and Piero della Francesca’s *Flagellazione di Cristo* (1455–60). [3] Art scholars can now rotate, zoom in and “fly” through these virtual spaces, study the consistency of the perspective and artists’ deviations from geometric accuracy. Smith, Stork and Zhang used multi-view reconstruction to form a three-dimensional model of the tableau in Scott Fraser’s *Three way vanitas* based on the images depicted in multiple plane mirrors within the tableau. [21]

Computer graphics models of studios allow scholars to study artists’ working methods (cf. Sect. [L6]) and gain insights for new interpretations. For example, Stork and Furuichi built a computer graphics model of Diego Velázquez’s *Las meninas* (1656) to determine what was the source of the images reflected in the plane mirror on the rear wall. [29] In this way they confirmed the claim that the source was not the king and queen in position of the viewer, but instead an otherwise hidden portrait on the large canvas depicted within the painting itself.

2 Conclusion

This interdisciplinary research program, while fairly new, has had several successes and a growing number of adherents, among both computer scientists and art scholars. [1] There have been several conferences in major scientific and technical organizations [27,28] and recently interest from traditional art organizations as well. [25,19] Computer scientists interested in contributing to this field should work as closely as possible with art scholars to ensure that their work is grounded in art historical knowledge and addresses questions of interest to the art community.

Acknowledgements

The author would like to thank the Getty Research Institute, Santa Monica, CA, for granting reader’s privileges to its superb reference library.

References

1. Berns, R.S.: Rejuvenating Seurat’s palette using color and imaging science: A simulation. In: Herbert, R.L. (ed.) *Seurat and the making of La Grande Jatte*, pp. 214–227. Art Institute of Chicago, Chicago (2004)

¹ www.diatrope.com/stork/FAQs.html

2. Chahine, H., Cupitt, J., Saunders, D., Martinez, K.: Investigation and modelling of color change in paintings during conservation treatment. In: *Imaging the past: Electronic imaging and computer graphics in museums and archaeology*. Occasional papers of the British Museum, vol. 114, pp. 23–33 (1996)
3. Criminisi, A.: *Accurate visual metrology from single and multiple uncalibrated images*. ACM Distinguished Dissertation Series. Springer, London (2001)
4. Criminisi, A., Kemp, M., Kang, S.-B.: Reflections of reality in Jan van Eyck and Robert Campin. *Historical methods* 3(37), 109–121 (2004)
5. Criminisi, A., Stork, D.G.: Did the great masters use optical projections while painting? Perspective comparison of paintings and photographs of Renaissance chandeliers. In: Kittler, J., Petrou, M., Nixon, M.S. (eds.) *Proceedings of the 17th International Conference on Pattern Recognition*, vol. IV, pp. 645–648 (2004)
6. Duarte, M., Stork, D.G.: Image contour fidelity analysis of mechanically aided enlargements of Jan van Eyck’s Portrait of Cardinal Niccolò Albergati. *Leonardo* 42 (in press, 2010)
7. Duda, R.O., Hart, P.E., Stork, D.G. (eds.): *Pattern classification*, 2nd edn. John Wiley and Sons, New York (2001)
8. Easton Jr., R.L., Knox, K.T., Christens-Barry, W.A.: Multispectral imaging of the Archimedes palimpsest. In: *Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop, AIPR 2003*, El Segundo, CA, pp. 111–116. IEEE, Los Alamitos (2003)
9. Blair Hedges, S.: Image analysis of Renaissance copperplate prints. In: Stork, D.G., Coddington, J. (eds.) *Computer image analysis in the study of art*, Bellingham, WA, vol. 6810, pp. 681009–1–20. IS&T/SPIE (2008)
10. Irfan, M., Stork, D.G.: Multiple visual features for the computer authentication of Jackson Pollock’s drip paintings: Beyond box-counting and fractals. In: Niel, K.S., Fofi, D. (eds.) *SPIE Electronic Imaging: Machine vision applications II*, vol. 7251, pp. 72510Q1–11. SPIE/IS&T, Bellingham (2009)
11. Johnson, C.R., Hendriks, E., Berezhnoy, I.J., Brevdo, E., Hughes, S.M., Daubechies, I., Li, J., Postma, E., Wang, J.Z.: Image processing for artist identification. *IEEE Signal Processing magazine* 25(4), 37–48 (2008)
12. Johnson, M.K., Stork, D.G., Biswas, S., Furuichi, Y.: Inferring illumination direction estimated from disparate sources in paintings: An investigation into Jan Vermeer’s Girl with a pearl earring. In: Stork, D.G., Coddington, J. (eds.) *Computer image analysis in the study of art*, vol. 6810, pp. 68100I–1–12. SPIE/IS&T, Bellingham (2008)
13. Jones-Smith, K., Mathur, H., Krauss, L.M.: Drip paintings and fractal analysis. *Physical Review E* 79(046111), 046111–12 (2009)
14. Kale, D., Stork, D.G.: Estimating the position of illuminants in paintings under weak model assumptions: An application to the works of two Baroque masters. In: Rogowitz, B.E., Pappas, T.N. (eds.) *Electronic Imaging: Human vision and electronic imaging XIV*, vol. 7240, pp. 72401M1–12. SPIE/IS&T, Bellingham (2009)
15. Kammerer, P., Lettner, M., Zolda, E., Sablatnig, R.: Identification of drawing tools by classification of textural and boundary features of strokes. *Pattern Recognition Letters* 28(6), 710–718 (2007)
16. Knox, K.T.: Enhancement of overwritten text in the Archimedes Palimpsest. In: Stork, D.G., Coddington, J. (eds.) *Computer image analysis in the study of art*, vol. 6810, pp. 681004–1–11. IS&T/SPIE, Bellingham (2008)
17. Lyu, S., Rockmore, D., Farid, H.: A digital technique for art authentication. *Proceedings of the National Academy of Sciences* 101(49), 17006–17010 (2004)

18. Nagy, G., Stork, D.G.: Inferring Caravaggio's studio lighting and praxis in 'The calling of St. Matthew' by computer graphics modeling. In: Stork, D.G., Coddington, J., Bentkowska-Kafel, A. (eds.) *Computer vision and image analysis of art*. SPIE/IS&T, Bellingham (2010)
19. Noble, P., Stork, D.G., Meador, S.: Computer image analyses of brick patterns in paintings by Jan van der Heyden (1637–1712). American Institute for Conservation, Paintings specialty group, 20 (2009) (poster abstract)
20. Robinson, M.D., Stork, D.G.: Aberration analysis of the putative projector for Lorenzo Lotto's Husband and wife: Image analysis through computer ray-tracing. In: Stork, D.G., Coddington, J. (eds.) *Computer image analysis in the study of art*, vol. 6810, pp. 68100H–1–11. SPIE/IS&T, Bellingham (2008)
21. Smith, B., Stork, D.G., Zhang, L.: Three-dimensional reconstruction from multiple reflected views within a realist painting: An application to Scott Fraser's Three way vanitas. In: Angelo Beraldin, J., Cheok, G.S., McCarthy, M., Neuschaefer-Rube, U. (eds.) *Electronic imaging: 3D imaging metrology*, vol. 7239, pp. 72390U1–10. SPIE/IS&T, Bellingham (2009)
22. David, G.: Stork. Optics and realism in Renaissance art. *Scientific American* 291(6), 76–84 (2004)
23. Stork, D.G.: Optics and the old masters revisited. *Optics and Photonics News* 15(3), 30–37 (2004)
24. Stork, D.G.: Locating illumination sources from lighting on planar surfaces in paintings: An application to Georges de la Tour and Caravaggio. In: *Optical Society of American Annual Meeting*, Rochester, NY. Optical Society of America (2008)
25. Stork, D.G.: New insights into Caravaggio's studio methods: Revelations from computer vision and computer graphics modeling. In: *Renaissance Society of American Annual Meeting*, Los Angeles, CA, p. 102 (2009) (abstract)
26. Stork, D.G.: Comment on Drip paintings and fractal analysis: Scale-space features in multi-feature classifiers for drip painting authentication. *Physical Review E* (submitted, 2010)
27. Stork, D.G., Coddington, J. (eds.): *Computer image analysis in the study of art*, vol. 6810. SPIE/IS&T, Bellingham (2008)
28. Stork, D.G., Coddington, J., Bentkowska-Kafel, A. (eds.): *Computer vision and image analysis of art*. SPIE/IS&T, Bellingham (forthcoming 2010)
29. Stork, D.G., Furuichi, Y.: Computer graphics synthesis for inferring artist studio practice: An application to Diego Velázquez's Las meninas. In: McDowall, I.E., Dolinsky, M. (eds.) *Electronic imaging: The engineering reality of virtual reality*, vol. 7238, pp. 7238061–7238069. SPIE/IS&T, Bellingham (2009)
30. Stork, D.G., Johnson, M.K.: Estimating the location of illuminants in realist master paintings: Computer image analysis addresses a debate in art history of the Baroque. In: *Proceedings of the 18th International Conference on Pattern Recognition*, Hong Kong, vol. I, pp. 255–258. IEEE Press, Los Alamitos (2006)
31. Stork, D.G., Johnson, M.K.: Lighting analysis of diffusely illuminated tabeaus in realist paintings: An application to detecting 'compositing' in the portraits of Garth Herrick. In: Delp III, E.J., Dittmann, J., Memon, N.D., Wong, P.W. (eds.) *Electronic Imaging: Media forensics and security XI*, vol. 7254, pp. 72540L1–8. SPIE/IS&T, Bellingham (2009)
32. Taylor, R.P., Micolich, A.P., Jonas, D.: Fractal analysis of Pollock's drip paintings. *Nature* 399, 422 (1999)

Panoramic and 3D Computer Vision

Akihiko Torii¹ and Reinhard Klette²

¹ Czech Technical University, Prague, Czech Republic

² The University of Auckland, Auckland, New Zealand

Abstract. Image sensor technology provides in recent years the tools for wide-angle and high-resolution 3D recording, analysis and modeling of static or dynamic scenes, ranging from small objects such as artifacts in a museum to large-scale 3D models of a castle or 3D city maps, also allowing real time 3D data acquisition from a moving platform, such as in vision-based driver assistance. The paper documents a particular subfield of these developments by providing mappings of omnidirectional images (catadioptric or dioptric images) into panoramic images on a cylinder. The mappings are specified by using the geometries of omnidirectional cameras. Image examples illustrate potentials for projects in arts, science or technology.

Keywords: Panoramic imaging, omnidirectional viewing.



Cusco mural by Juan Bravo (1992, length: about 17 m)

1 Introduction

Panoramic photography is needed for recording a painting as shown above. We briefly review the geometry of catadioptric and dioptric cameras, which are relatively new tools for panoramic photography. Omnidirectional camera systems [2] have been developed to observe a 360-degree field of view; see Fig. 1 for an example. Popular omnidirectional imaging systems are catadioptric or dioptric camera systems [1]. A catadioptric camera system is constructed as a combination of a quadric mirror and a conventional camera [8,9,10]. The dioptric camera system is constructed with a specially designed refractor, which controls the angles of rays passing through the lens [9], as the optical lens of the camera. Cylindric panoramas are, for example, generated by rotating line cameras [7]; we consider

¹ *Catadioptric*: pertaining to, or involving both the reflection and the refraction of light; *dioptric*: relating to the refraction of light



Fig. 1. Omnidirectional cameras. Left: a digital camera with a hyperboloidal shaped mirror. Right: a fish-eye camera. [12]

single-projection center cylindric panoramas in the following (i.e., where the off-axis distance equals zero). Omnidirectional images are either of the catadioptric or dioptric kind.

Omnidirectional imaging systems are used for controlling mobile robots [1] or in vision-based driver assistance [5]. Therefore, image analysis is also required for omnidirectional images. Basically, analysis may happen directly in those catadioptric or dioptric images, or after a transform of those into traditional formats (preferably cylindric panoramas).

Besides robotics, another significant use of catadioptric or dioptric cameras is the generation of panoramic views from captured images. Panoramic images are essential for virtual reality applications, such as Google Earth [6]. Furthermore, panoramic images play an important role in the creation of artistic photographs [12]. In this paper, we identify mappings of catadioptric and dioptric images into images on a cylinder (i.e., cylindric images). See Fig. 2 for three examples.



Fig. 2. Upper row: original fish-eye images (180-degree fields of view, showing Prague castle and a group of people). Lower row: resulting panoramic images.

Omnidirectional images are required to be pre-processed before applying computer vision or image recognition tools (e.g., detectors and trackers of pedestrians, cars, and other objects [34]), because available tools are typically established and trained on perspective images. The aim is often to generate perspective cutouts from omnidirectional images, and to adapt common tools without special modifications. However, such cutouts completely lose the benefit of the large field of view (e.g., surveying a large area, without a need to switch between multiple images). Consequently, a mapping of the captured 360-degree field of view into a cylindric panorama is often the best solution, because single-center cylindric images possess perspective-like appearances while suppressing circular distortion due to catadioptric or dioptric projection.

The structure of this short paper is as follows. In Section 2, we first introduce central hyperbolic and parabolic camera models for characterizing common catadioptric cameras, and then we derive mappings from hyperbolic and parabolic images into cylindric images. In Section 3, we describe a fish-eye camera model as an example of dioptric cameras, and derive a mapping of fish-eye images into cylindric images. Finally, we show examples of cylindric images obtained from recorded hyperbolic or fish-eye images.

2 Catadioptric into Cylindric Images

Hyperbolic or parabolic mirrors are two commonly used in catadioptric cameras, and further mirrors are under consideration (e.g., to ensure a single projection center and uniform image resolution at the same time).

2.1 Hyperbolic into Cylindric Images

A central hyperbolic camera consists of a hyperboloidal shaped mirror and a conventional digital camera [8,10]. Such a hyperbolic camera has a unique center of projection at the focus of the hyperboloidal mirror by locating the projection center of the digital camera at the other focus of the hyperboloid, as illustrated on the left in Fig. 3. In this configuration, light rays which pass through the focus of the hyperboloid are reflected on the mirror, and the reflected rays pass through the other focus of the hyperboloid.

For deriving the transformation equation, we consider at first the case as illustrated on the left in Fig. 3. Let $\mathbf{C} = (0, 0, -2e)$ be the center of the pinhole camera. Letting the focus \mathbf{F} of the hyperboloid C^2 be the origin of the world coordinate system, the hyperboloid is expressed in the quadric form:

$$(x, y, z, 1) \begin{pmatrix} \frac{1}{a^2} & 0 & 0 & 0 \\ 0 & \frac{1}{a^2} & 0 & 0 \\ 0 & 0 & -\frac{1}{b^2} & -\frac{e}{b^2} \\ 0 & 0 & -\frac{e}{b^2} & -\frac{e^2}{b^2} + 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = 0$$

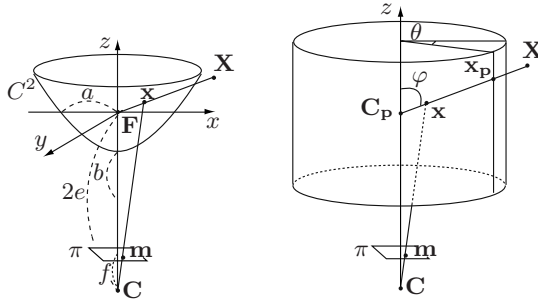


Fig. 3. Mapping from hyperbolic to cylindric images. Left: hyperbolic camera system. Right: hyperbolic and cylindric camera systems combined.

where $e = \sqrt{a^2 + b^2}$. The projection of a point $\mathbf{X} = (X, Y, Z)^\top$ in 3D space into a point $\mathbf{x} = (x, y, z)^\top$ on the hyperboloid C^2 is then expressed as $\mathbf{x} = \chi \mathbf{X}$ where

$$\chi = \frac{a^2}{b|\mathbf{X}| - eZ}$$

The projection of point \mathbf{x} into a corresponding point $\mathbf{m} = (u, v)^\top$ on the image plane π is expressed as

$$\begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} = \frac{1}{z + 2e} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}$$

Accordingly, the total mapping, from \mathbf{X} to \mathbf{m} , is formulated as

$$u = \frac{fa^2X}{(a^2 - 2e^2)Z + 2be|\mathbf{X}|} \quad \text{and} \quad v = \frac{fa^2Y}{(a^2 - 2e^2)Z + 2be|\mathbf{X}|} \quad (1)$$

The hyperbolic to cylindric image transform is illustrated on the right in Fig. 3. Let \mathbf{C}_p be the center of the cylindric projection. By setting $\mathbf{C}_p = \mathbf{F}$, a point \mathbf{x}_p on the cylindric image and a point \mathbf{x} on the hyperboloid lie on a line, connecting point \mathbf{X} in 3D space with the focus \mathbf{F} of the hyperboloid. The cylindric coordinate system expresses a point $\mathbf{x}_p = (x_p, y_p, z_p)$ on the cylindric surface as

$$x_p = r \cos \theta, \quad y_p = r \sin \theta, \quad z_p = r \tan \varphi \quad (2)$$

where $0 \leq \theta < 2\pi$ and $-\pi/2 \leq \varphi < \pi/2$. Hereafter, without loss of generality, we set $r = 1$. The mapping between the hyperbolic image $I(u, v)$ and the cylindric image $I_P(\theta, \varphi)$ can then be formulated as

$$u = \frac{fa^2 \cos \theta}{(a^2 \mp 2e^2) \tan \varphi \pm 2be \sqrt{1 + \tan^2 \varphi}}$$

$$v = \frac{fa^2 \sin \theta}{(a^2 \mp 2e^2) \tan \varphi \pm 2be \sqrt{1 + \tan^2 \varphi}}$$

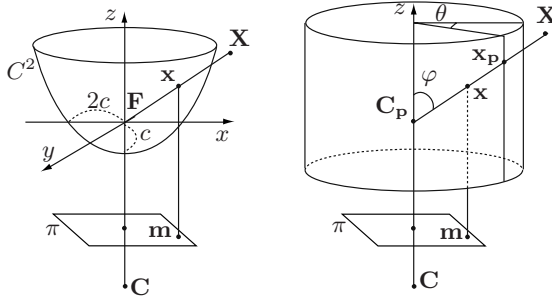


Fig. 4. Mapping from parabolic to cylindric images. Left: parabolic camera system. Right: parabolic and cylindric camera systems combined.

We thus derived a one-to-one correspondence between points on a hyperbolic image and points on a cylindric image. This allows to transform these images using common interpolation techniques, such as bilinear, cubic convolution, or B-spline interpolation. Note that substituting Equ. (2) into Equ. (1) derives a mapping with central cylindric projection. Cylindric images can also be derived for different projections (e.g., setting $z_p = r\varphi$ achieves equi-rectangular projection).

2.2 Parabolic into Cylindric Images

We consider at first the case illustrated on the left in Fig. 4. Let $\mathbf{C} = (0, 0, -\infty)$ be the center of the orthographic camera. Letting the focus \mathbf{F} of the paraboloid C^2 be the origin of the world coordinate system, the paraboloid C^2 is expressed in the quadric form:

$$(x, y, z, 1) \begin{pmatrix} \frac{1}{4c} & 0 & 0 & 0 \\ 0 & \frac{1}{4c} & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = 0$$

The projection of a point $\mathbf{X} = (X, Y, Z)^\top$ in 3D space into a point $\mathbf{x} = (x, y, z)^\top$ on the paraboloid is then expressed as $\mathbf{x} = \chi\mathbf{X}$ where

$$\chi = \frac{2c}{|\mathbf{X}| - \mathbf{Z}}$$

The projection of point \mathbf{x} into a point $\mathbf{m} = (u, v)^\top$ in the image plane π is expressed as

$$\begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}$$

Accordingly, the mapping from \mathbf{X} into \mathbf{m} is formulated as

$$u = \frac{2cX}{|\mathbf{X}| - \mathbf{Z}}, \quad v = \frac{2cY}{|\mathbf{X}| - \mathbf{Z}}$$

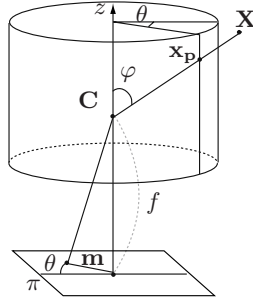


Fig. 5. Transform of a fish-eye into a cylindric image

Furthermore, as show on the right in Fig. 4, by setting $C_p = F$, a point x_p on the panoramic image and a point x on the paraboloid lie on a line, connecting point X in 3D space with the focal point F of the paraboloid. Let $x_p = (x_p, y_p, z_p)$ be a point on the cylindric surface as in Equ. (2). A mapping between the parabolic image $I(u, v)$ and the cylindric image $I_P(\theta, \varphi)$ is given as follows:

$$u_h = \frac{2c \cos \theta}{\sqrt{1 + \tan^2 \varphi} - \tan \varphi} \quad \text{and} \quad v_h = \frac{2c \sin \theta}{\sqrt{1 + \tan^2 \varphi} - \tan \varphi}$$

3 Fish-Eye into Cylindric Images

A fish-eye camera generates an image on the basis of a stereographic, equi-solid angle, orthogonal, and equi-distant projection. This projection is illustrated in Fig. 5. Let $m = (u, v)^T$ and $X = (X, Y, Z)^T$ be a point on an image acquired by fish-eye camera and a point in a 3D space, respectively. Depending on the projection model, the mapping of X into m can be stated as follows:

$$\begin{aligned} \text{stereographic} : u &= 2fa \tan(\cos^{-1}(c)/2), \quad v = 2fb \tan(\cos^{-1}(c)/2) \\ \text{equi-solid angle} : u &= 2fa \sin(\cos^{-1}(c)/2), \quad v = 2fb \sin(\cos^{-1}(c)/2) \\ \text{orthogonal} : u &= fa \sin(\cos^{-1}(c)), \quad v = fb \sin(\cos^{-1}(c)) \\ \text{equi-distant} : u &= fa \cos^{-1}(c), \quad v = fb \cos^{-1}(c) \end{aligned}$$

where

$$a = \frac{X}{\sqrt{X^2 + Y^2}}, \quad b = \frac{Y}{\sqrt{X^2 + Y^2}}, \quad c = Z/|X|$$

Next, let $x_p = (x_p, y_p, z_p)$ be a point on the cylindric surface as in Equ. (2). A mapping between the fish-eye camera image $I(u, v)$ and the cylindric image $I_P(\theta, \varphi)$ is then as follows (where $\gamma = \pi/2 - \varphi$):

$$\begin{aligned}
 \text{stereographic} : u &= 2f \tan(\gamma/2) \cos \theta, & v &= 2f \tan(\gamma/2) \sin \theta \\
 \text{equi-solid angle} : u &= 2f \sin(\gamma/2) \cos \theta, & v &= 2f \sin(\gamma/2) \sin \theta \\
 \text{orthogonal} : u &= f \sin(\gamma) \cos \theta, & v &= f \sin(\gamma) \sin \theta \\
 \text{equi-distant} : u &= f(\gamma) \cos \theta, & v &= f(\gamma) \sin \theta
 \end{aligned}$$

4 Experiments

We show some examples of cylindric images transformed either from hyperbolic or fish-eye images into this format. The top left and the middle left in Fig. 6 show two hyperbolic images with 360-degree fields of view, acquired with a hyperbolic camera as shown on the left of Fig. 1. Using the mapping derived in Section 2.1, the original images are instantly transformed into cylindric images as shown on the top right and middle right of Fig. 6. The bottom row in Fig. 6 shows the Inca shield of Cusco (part of Bravo’s mural as shown above) and its cylindric image, assuming that the shield was created by imagining a ‘hyperbolic camera’.

Furthermore, the top row in Fig. 2 shows three original images recorded with a hemispherical field of view, acquired with a fish-eye camera as shown on the right of Fig. 1. The resulting panoramic images are shown in the bottom row of Fig. 2.

Figure 7 just illustrates a stereo pair of fish-eye images, recorded on a roof-rack of HAKA1, a research vehicle in the driver assistance project *.enpeda..* at Auckland University. Stereo images require geometric rectification before applying a

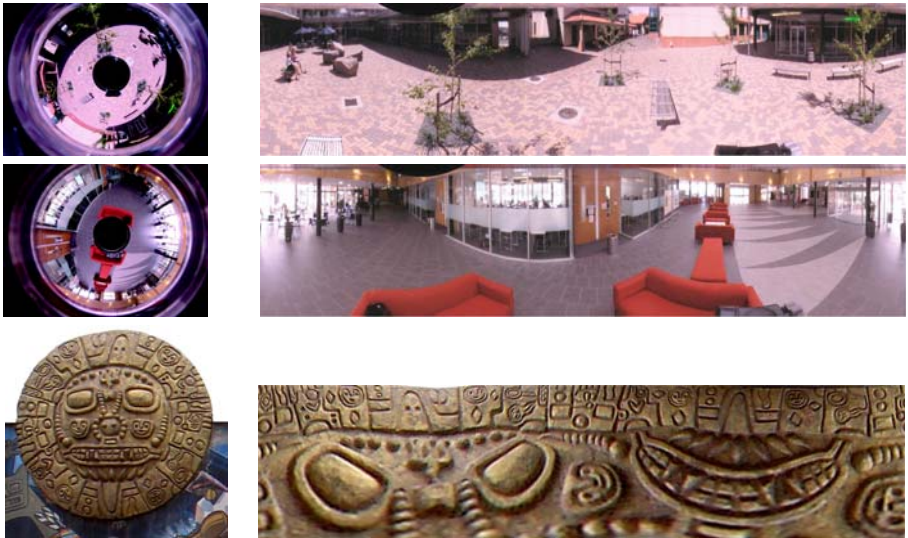


Fig. 6. Hyperbolic into panoramic images (top and middle row: Tamaki campus, The University of Auckland; bottom row: Inca shield of Cusco). Left: original images; right: resulting panoramic images.



Fig. 7. Stereo pair of fish-eye images, for ensuring a wide field of view in vision based driver assistance

stereo matching algorithm [5]. Prior to rectification, these images are mapped into cylindric format as described above.

5 Concluding Remarks

This brief paper accompanies a keynote talk at this conference by providing transformation formulas of catadioptric and dioptric images into cylindric images, thus detailing one particular subject in the large field of panoramic or 3D computer vision. The provided ideal geometric transforms still need to be “refined” by further analysis of potential impacts, such as optical distortions, to ensure resultant panoramas of ‘perfect quality’.

References

1. Daniilidis, K., Papanikolopoulos, N. (eds.): Panoramic Robots. Special issue of IEEE Robotics Automation Magazine 11 (2004)
2. Daniilidis, K., Klette, R. (eds.): Imaging Beyond the Pinhole Camera. Springer, New York (2007)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR, vol. 1, pp. 886–893 (2005)
4. Ess, A., Leibe, B., Schindler, K., Van Gool, L.: A mobile vision system for robust multi-person tracking. In: Proc. CVPR, CD (2008)
5. Gehrig, S., Rabe, C., Krueger, L.: 6D vision goes fisheye for intersection assistance. In: Proc. Canadian Conf. Computer Robot Vision, pp. 34–41 (2008)
6. Google. Google earth (2004), <http://earth.google.com/>
7. Huang, F., Klette, R., Scheibe, K.: Panoramic Imaging. Wiley, Chichester (2008)
8. Nayar, S.K.: Catadioptric omnidirectional cameras. In: Proc. CVPR, pp. 482–488 (1997)
9. Ray, S.F.: Applied Photographic Optics : Lenses and Optical Systems for Photography, Film, Video, Electronic and Digital Imaging, 3rd edn. Focal Press, Oxford (2002)
10. Svoboda, T., Pajdla, T., Hlaváč, V.: Epipolar geometry for panoramic cameras. In: Proc. Czech Technical University Workshop 1998, Prague, pp. 177–178 (1998)
11. Wei, S.K., Huang, F., Klette, R.: Color anaglyphs for panorama visualizations. Technical Report, CITR-TR-19, The University of Auckland (1998)
12. Wikipedia. Panoramic photography (2009), http://en.wikipedia.org/wiki/Panoramic_photography

Meaningful Engagement: Computer-Based Interactive Media Art in Public Space

Jiun-Jhy Her and Jim Hamlyn

Gray's School of Art, The Robert Gordon University, UK
{j.her, j.hamlyn}@rgu.ac.uk

Abstract. Interactive technologies, including electronic devices are increasingly being utilized as a medium for artistic expression and have been placed in freely accessible public environments with mixed results. When audiences encounter computer-based interactive media arts in a public space they are drawn by various interactivities, to play and experiment with them. However, whether the audience is able to gain a meaningful experience through those physical interactivities has remained an issue of both theoretical and practical debate. This paper will focus on these aspects, most specifically through the study of interactive art in freely accessible public space. The author proposes four new conceptual/analytical tools for examining the subject. It is anticipated that this paper will provide possible alternative strategies for both artists and art researchers in this field with a purpose to enhance intellectual engagement with their audiences, so as to succeed in leading interactors to obtain meaningful experience and rewards.

Keywords: Interactive Art, Media Art, Public Art, Meaningful Engagement.

1 Introduction – The Research Background

‘Interactive art is said to be ‘created’ by the people engaged in the active experience of it’ [1].

The research has drawn upon both the allure and the awareness of issues of the subject. The allure of interactivity and the dynamics of computer-based interactive media arts, which often actively grasp the attention of audiences with its diverse presentations, in contrast to the conventional ways of viewing art, this active participation is often encouraged as an artistic input so as to obtain meaningful experiences and reveal the artistic intent [2]. Since the 1990’s there has been a growing study of aesthetic and emotional experiences in the area of computer-based interactive media art research (Edmonds and Graham et al) [3], [4]. Artists and art researchers have been investigating the interface of these media arts in an attempt to reveal forms which engage audiences. Various strategies for engaging audiences have been developed. ‘Playfulness’ might be considered one of the most viable tactics that is often employed as the main ingredient as for an initial engagement as well as being a catalyst to arouse subsequent more meaningful experiences (Polaine and Moggridge et al) [5], [6]. In addition to this, others have been proposed by the author, which include: Dominance Transfer, Mind Orientedness and Accessible Challenge, all of which have been examined to some extent by previous research, the pilot and field studies which will be discussed further in the following sections.

2 The Pilot and Field Studies

In order to establish the scope for further research and tests for the viability of the research methodology, an experimental interactive installation was made and a pilot study conducted. The work “Event Horizon” is a screen based interactive installation, which is equipped with an infrared-sensor as means of detecting the audiences’ presence. The image patterns of the work changes randomly and dynamically and is triggered by the audiences’ movement. The pilot study was conducted in The Robert Gordon University, in a public hallway of the Scott Sutherland School; it is one of the major thoroughfares that leads to both the main exit and lecture rooms in the building. The criteria for choosing the pilot site were: 1) It must be a mundane and non exhibition space 2) Is accessible for everyone 3) It is a major route for everyday use.

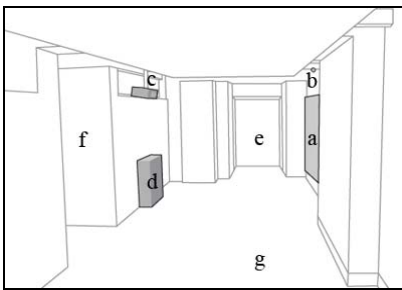


Fig. 1. The pilot site

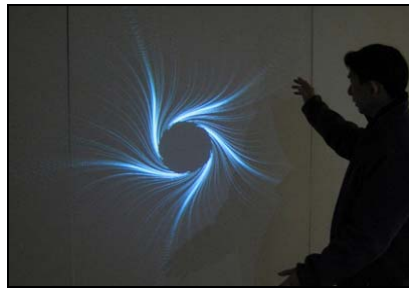


Fig. 2. Event Horizon

Fig 1. Label Descriptions

a. Screen	b. Infrared-sensor	c. Projector	d. Computer and speakers
e. To the atrium	f. The elevator	g. To the lecture theaters	

The first field study was carried out in Taiwan, Kaoshiung County, Fongshan West (Mass Rapid Transit) station. A phoenix shaped computer-controlled interactive installation was made from articulated stainless steel pipes, which were hung beneath the ceiling inside the station near the exit one. The form of the installation symbolizes the legend of Fongshan City (Fong Sang in Chinese means Phoenix Mountain). The streamlined phoenix shaped installation resembles Chinese calligraphy, and is not only made for reflecting the cultural value but also with the purpose to elicit affection from the passengers toward their hometown. Several stainless maracas are attached to the end of the pipes. The sound of maracas is triggered when the passengers pass underneath it.

Ethnographical approaches were employed in both studies that included non-participant observation and semi-structured interviews with the audiences and passengers concerning their physical reaction and sensory responses with the art installations. Thirty research questionnaires were issued at the conclusion of the verbal interviews for the pilot study of which twenty were retrieved at the end. In consideration of any ethical issues and with regard to the efficiency of the interview process, a digital voice recorder was utilized as a major research data collecting tool. This methodology was applied to research gathered from passengers at the Fongshan West MRT station, of which ultimately fifteen passengers were interviewed.



Fig. 3-4. The field study site (Fongshan West MRT station)

2.1 Initial Findings

The sounds and image patterns of the works were constantly changing as people passed the installation, and as such it did not require any active physical intervention to trigger an initial interaction; the installation instantly gained people's attention in the hallway. The same response to the installations acoustics' was also identified during the field study at the Fongshan West MRT station. As the interactors realized that they were the stimulus to trigger the interaction, they instantly became involved, and enjoyed this 'creative authorship' [7]. Within the retrieved questionnaires from the pilot study, ninety-five percent reported that the interactive effect stimulated their curiosity and thought the shifting sounds and image patterns, which were not repeated, kept their attention and made them want to explore and try to understand how it worked.

While the majority of interviewees reported initially being mainly attracted by the sound of the installation, they were then subsequently curious to ascertain the meaning of the artwork. Ninety percent of interviewees indicated that they were trying to discern how the installation worked and actively interacted with the installation. They waved hands, shook feet, moved back and forth, and some even danced in front of the installation with the goal of trying to manipulate the changing the image patterns. During the pilot study, an interviewee (A1) was asked about his thoughts when he first saw the installation, he wrote: "Aroused and helped brighten what was a grey boring day, and even helped for a moment to think about different happy thoughts and possibility it would be song as I could dance to"[sic]. This is in contrast with the first field research site where none of the physical reactions mentioned above were exhibited by passengers.

When the interviewees were asked what they felt when learning that their movement triggers the changing of the patterns, nineteen people gave positive responses ("excited" and "interested") and some even responded that the interaction had encouraged discussion between their friends. Though the passengers from the Fongshan West MRT station were not seen to be as active as those from the pilot study, they all gave positive responses. Some of them discussed the audible interaction with their friends too. In spite of the two different forms of presentation between "Event Horizon" and "The legend of the Phoenix", the finding from both research studies were very similar in a number of respects: for instance, the non repeated sounds succeeded in engaging the audiences initially, and together with changing patterns this evoked interactors' curiosity so as to lead to further explorations.

Despite this success, there are two issues which must be mentioned here: 1) Both art installations were installed in fairly quiet surroundings, which means the audiences were able to hear the sounds produced, however, these sounds might not be audible at other busier stations or bustling spaces e.g., the Taipei Main MRT station. 2) Both art installations were new to the environments and in general, people tend to be drawn to new. The interviewees from the pilot study were asked about their thoughts when they first saw the installation, they replied: (A2) “It was a bit fresh to see a piece of art at an unexpected place like that, (Corridor of the Scott Sutherland School)”. (B3) “What is this? It was quite strange”. (B5) “What is this and why is here and I stop, and read [sic]”. However, how long this sense of novelty can be sustained, and whether the interactors will be able to obtain a meaningful interactive experience, remains to be seen. These questions will be discussed further in the following sections, along with other audience’s characteristics of engagement.

3 Characteristics of Engagement

3.1 Playfulness

In both research subjects, the interactivities acted as the bait which lured the audience to further play with the art installations. ‘Play is a core of human value; even a core of mammalian value, we used an analogy that lion cubs learn to hunt and fight by play together’ [6]. By playing, the audiences are urged to participate, to stand closer, and to become involved and even to touch. At this point, they become active players; they question, explore and test the possibility of the art installations. This brings about a child-like state of joy. ‘In game-type interactive works, the term player is common’ [8]. The encouragement of participation is not the sole purpose of the installation; a key function is also to bring about a sense of empathy and to simulate the interactors imagination.

‘Playfulness’ has been spotted in many computer-based interactive media arts that successfully engaged their audiences in various public contexts, e.g., United Visual Artists’ Volume at the V&A (2006) [9] and Lozano Hemmer’s Body Movie (2001-2008) [10]. Once the audiences’ role becomes that of an active player, they could further become ‘an independent causal agent’ [11], who might unwittingly take over the stimulus while interacting with the art installations. Their responses could motivate the curiosity of spectators nearby and further influence their behavior. As soon as the interactivity reaches this level the effect is no longer constrained between art installations and the individual but provokes further interaction among people. This is an ideal psychological state, which indicates one has reached a state of ‘Immersion, The sensation of being surrounded by a completely other reality, as different as water is from air, which takes over all our attention, our all perception apparatus’ [7].

However, not only the result of the state is unpredictable, but it is also delicate. Eisenberg points out ‘The fragility of the jamming is evident in that it can never be routinized, habitual, linked to a specific set of antecedents, or necessarily self-sustaining once begun. On the other hand, it is possible to court these experiences and to cultivate the attitudes and expectations that make jamming most likely to occur’ [12]. Based on this argument ‘Playfulness’ may not be the sole functional characteristic and other strategies and processes may be required in order to invoke meaningful experiences.

3.2 Dominance Transfer

This is the power of ‘transformation’ [7] that is bestowed on interactors. ‘Digital narratives add another powerful element to this potential by offering us the opportunity to enact stories rather than to merely witness them’ [7], giving the sense of control to the audiences may be another crucial element in arousing interactivity. Works that have this characteristic are often deliberately designed to function as a charming and hospitable ‘host’ [13] who invites the guests to the game. In the meantime the space for the guests is given [14] where hearty play is allowed and, in fact, is strongly encouraged.

One interviewee (A5) from the pilot study, who was asked, did the interactive effect of the installation stimulate your curiosity, he reported: “I was fascinated at the way the patterns change. I had to stop severally [sic] trying to figure out what it’s all about”. Here the audiences do not feel any demand, on contrary they act enthusiastically and actively engage with the art installation. This echos Murray’s ‘Agency, it is the satisfying power to take meaningful action and see the results of our decisions and choices’ [7].

In contrast to interactive installations, Dominance Transfer is not usually sensed in static art forms. ‘Traditionally the interaction of the viewer with the work of art has been via looking and respectfully appreciating’ [15] under this, the audience is often positioned as inferior to the work and always physically passive [15]. Whereas, the exclusive intellectual and physical hegemony of the artist does not exist in computer-based interactive media art. The host relinquishes dominance to its viewers and gives them conceptual space and a sense of control, whilst ‘creative authorship’ is shared. Once the audiences perceive interactivities are activated by their physical movements, they will go further and try to manipulate them, at this stage the artist’s dominance has been overcome and creativity is shared with the audience.

3.3 Mind-Orientedness

Traditionally aesthetics in general encourage viewers to think about author’s intention [16], which usually through visual admiration and contemplation, the action of which is deemed as ‘Reading’ [15]. To truly comprehend traditional arts requires a high degree of fluency in the artistic techniques and history concerning the work. This is not to say that art cannot be appreciated by the untrained eye, but merely that to appreciate it fully, knowledge is required. In relation to computer-based interactive media arts, the barrier to appreciation is lowered, thanks in part to the themes discussed in the paper, and as such may be accessible more widely than static art forms.

The sound of maracas “The legend of the Phoenix” is activated when the audience enters the reception/exhibition area that instantly attracted the attention of the audience. This installation succeeded in attaining the first goal in this regard, secondly its form reflects a strong cultural value and represents the story of the land, which closely ties to passengers who originate from this area. This generates resonance with the locality. Some interviewees from the field study reported that they can tell the form of the installation resembled a phoenix and that they were interested in finding out the meaning it represents.

Another example of “Metamorphosis” [17] is a holographic artwork installed at C.K.S Memorial Hall MRT Station, Taipei, Taiwan. There are separate 3-D images, e.g. paper airplanes, birth certificates, textbooks, graduation photos, personal identification, wedding

certificates, and near the end an image of a dove flying away. The images are installed from the bottom to the top alongside the escalator leading to the exit. Passengers are engaged as soon as they step onto the escalator. The static holographic images become a slow animation while following the moving escalator. The images represent different stages of life that connect to the people who were born and raised in the countryside and in the city. The sentiment between the country and the passengers is invoked and reflects to the viewers following the movement with their eyesight. These are Mind Oriented characteristics. 'Not relegated solely to self-reflexive aesthetic concerns, artworks increasingly reflected cultural values, responded to political issues, and directly engaged their audiences in critical dialogue of the day' [18].

3.4 Accessible Challenges

An issue that was mentioned at the very beginning of this article: how long can fresh feelings be sustained? In many cases the majority of the audience showed curiosity at the beginning. Curiosity as a widely accepted aspect of human nature can 'act as facilitators of the process of making sense of the objects' [19]. However, this effect usually does not endure. In order to allow the audience to develop their 'Optimal experience' [11], a viable strategy may be to increase the challenge with the intention of prolonging the time of engagement. Challenges and Skills are two indices used for measuring aesthetic/emotional experience. The 'Flow research model' [11] has been adopted by researchers in attempts to decipher the codes leading to engagement in various contexts (Costello, 2005, Forlizzi 2004, Stuart 2005 et al). One of the most remarkable findings is that people give positive reports when challenges and skills are balanced, and when both indices reach a high level, people reported entering the 'Flow', whereas reports showed boredom when challenges and skills lay on a low level [11].

Other indices such as 'use-time' are often used as a reference to indicate the level of engagement (Graham, 1997, Candy, 2006, Brigid, 2005 Ann, 2007 et al). Over time the audience might react to the 'emancipatory effect' [15], which allows the development of thought, imagination or might even produce interpretation. Thus to raise a challenge might be one of the crucial components to extend 'use-time'. However, high challenges in artistic appreciation may become counterproductive. It is understandable that people feel intrigued when they are in charge and able to cope with challenges, whereas they lose interest when things get boring or when there is no foreseeable hope in solving the challenges [6]. Hence an accessible challenge for art themes may be a way to provoke the interest of the audience, tempt them to explore, and lead them to obtaining meaningful experiences. This is where "Intelligent" interactivity may become increasingly important since, in theory, it can respond to different levels of skills and present an appropriate challenge for different interactors.

4 Interactivity in Free Accessible Public Spaces

When art is encountered in freely accessible public space the level of complexity and difficulty is increased, not only concerning the aspects of security, maintenance etc., but also concerning the audience which consists of people who are normally unaccustomed to being intellectually engaged in such a context. Variables related to the

audience might also be more difficult to account for. 'As Harriet Senie asserts, the public is often an "involuntary audience" for public art' [18]. Also as Birchfield argued 'Public art in this scope is housed outside of traditional art settings and is intended to engage a public audience that might not otherwise seek art experiences'[20]. However, by equipping the concepts of functionality and site specifics (Miles, M. 1997, Kaye, N. 2000, Kwon, M. 2004 et al) upon art installations, the tactics have long been considered the antidotes to revitalize the environmental surroundings. Nowadays 'an active interactivity' is also deemed a functional alternative as technology is becoming a mature part of artistic creative professions. Many interactive installations are installed in free access public places and have successfully provoked audiences' reactions and responses e.g. Jaume Plensa's "The Crown Fountain" (2004), Lozano-Hemmer's "Under Scan" series (2005-2008) and Hsiao's "The Legend of the Phoenix" (2009). Some of them have even been placed as permanent installations with specific artistic intentions that relate to various public contexts. Hence the combination of functionality, site specifics and interactivity (Meaningful Engaging Characteristics) could prove a feasible strategy to revitalize and engage wider public audiences in non art public space.

5 Conclusion and Future Works

There is no doubt that many previous research efforts about audience response have laid a strong foundation in the field. The engaging characteristics such as Playfulness, Dominance Transfer, Mind-Orientedness and the Accessible Challenge have been identified as functional strategies in the pilot study, the field studies as well as in the review of this context. However, those characteristics can not be treated as an overall solution to the issues but solely as tactics that may engage audiences. As more and more computer-based interactive media art installations are seen in open public space, the general public is starting to recognize and become familiar with this new genre. At this point, engaging audiences is not the only issue, a more important issue is how a meaningful experience can be obtained by interactors. Muller and Edmonds suggested that 'we must begin to question how interactivity as a medium produces meaning' [21]. In order to obtain in-depth research data about how meaningful experiences are triggered by interaction with interactive media art installations in open public space further studies are required. These will remain focused on the computer-based interactive media art installations that are placed in public spaces, specifically non-art spaces. Research into existing media artworks and audience responses in Taipei and Kaohsiung (Mass Rapid Transit) stations has been carried out and will be continued in other similar public settings. It is intended that the outcome will provide useful alternative methods for researchers and artists in the field, so as to deepen our understanding of the capacity of interactive art to realize meaningful engagement in the wider public context.

References

1. Candy, L., Amitani, S., Bilda, Z.: Practice-led strategies for interactive art research 2(4), 209-223 (2006)
2. Huhtamo, E.: Trouble at the Interface, or the Identity Crisis of Interactive Art. MAHArchive (2004)

3. Creativity and Cognition Studios, <http://www.creativityandcognition.com/>
4. Graham, B.: A Study of Audience Relationships with Interactive Computer Based Visual Artworks 32(4), 326–328 (1999)
5. Polaine, A.: The flow principle in interactivity. In: Proceedings of the second Australasian conference on Interactive entertainment, pp. 151–158. Creativity & Cognition Studios Press, Sydney (2005)
6. Moggridge, B.: Designing Interactions, pp. 321–325. The MIT Press, Cambridge (2007)
7. Murray, J.H.: Hamlet on the holodeck, pp. 97–182. The MIT Press, Cambridge (1997)
8. Ann, J.M., Peta, M., Margot, B.: The lens of ludic engagement: evaluating participation in interactive art installations. In: Proceedings of the 15th international conference on Multimedia. ACM, Augsburg (2007)
9. UnitedVisualArtists, <http://www.uva.co.uk/archives/49>
10. Lozano-Hemmer, R.: <http://www.lozano-hemmer.com/>
11. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper & Row (1990)
12. Eisenberg, E.: Strategic Ambiguities: Essays on Communication, Organization, and Identity, pp. 75–97. Sage Publications Inc., Thousand Oaks (2007)
13. Graham, C.: A Study of Audience Relationships with Interactive Computer-based Visual Artworks in Gallery Settings, Through Observation, Art Practice, and Curation, University of Sunderland (1997)
14. Petersen, M., Iversen, O., Krogh, P., Ludvigsen, M.: Aesthetic interaction: a pragmatist's aesthetics of interactive systems. ACM, New York (2004)
15. Graham Coulter-Smith, E.C.-S.: Art games: Interactivity and the embodied gaze. *Techno-etic Arts: a Journal of Speculative Research* 4(3), 169–182 (2006)
16. Manovich, L.: Post-media Aesthetics, http://netart.incubadora.fapesp.br/portal/referencias/Post_media_aesthetics1.pdf
17. Lin, S.-M.: Metamorphosis, <http://english.taipei.gov.tw/dorts/index.jsp?catid=2058>
18. Knight, C.: Public Art: Theory, Practice and Populism, pp. 21–28. Blackwell, Malden (2008)
19. Ciolfi, L., Bannon, L.: Designing Interactive Museum Exhibits: Enhancing visitor curiosity through augmented artifacts, <http://richie.idc.ul.ie/luigina/PapersPDFs/ECCE.pdf>
20. Birchfield, D., Phillips, K., Kidané, A., Lorig, D.: Interactive public sound art: a case study. In: Proceedings of the 2006 conference on New interfaces for musical expression. IRCAM Centre Pompidou, Paris (2006)
21. Muller, L., Edmonds, E.: Living laboratories for interactive art, vol. 2, pp. 195–207. CoDesign (2006)

Interactive WSN-Bar

Jiun-Shian Lin, Su-Chu Hsu, and Ying-Chung Chen

Graduate School of Art and Technology
Taipei National University of the Arts

No.1, Hsueh-Yuan Rd., Beitou, Taipei 112, Taiwan (R.O.C.)
{jiunshianlin, suchu.hsu, wadanabe}@gmail.com

Abstract. Based on the concept of ambient intelligence, we utilized wireless sensor network (WSN) and vision-based tracking technologies to create an interactive WSN-Bar. WSN-Bar is an interactive and innovative creation which has two modules: Garden of Light and Vivacious Bushes. It refers the variety of natural environmental factors and focuses on the relationship between human and nature. WSN-Bar can also detect the changes of brightness, temperature, CO₂ density outdoors and the movement of people inside the building. Besides, WSN-Bar is an interactive installation art which creates the opportunity to reduce the estranged gape among the participants.

Keywords: Bar, wireless sensor network, ambient intelligence, interactive art, vision-based tracking technology.

1 Introduction

In recent years, the concept of Ambient Intelligence (AmI) has been gradually granted more attention in fields of information design and digital art design. AmI emphasizes ubiquitous computing, invisible technology, and natural interaction. Mark Weiser was the first person who mentioned the basic concept of AmI in 1991 [1]. The term of Ambient Intelligence was first proposed by Philips in 1998; then the meaning of this term was clearer explained by ISTAG (Information Society Technology Advisory Group) in 1999 [2, 3]. The concept is to blend computing devices into a daily living environment; therefore, it can detect human behaviors and body conditions, and then automatically provide users with the matching feedbacks. In 1999, MIT cooperated with Philips to create Oxygen project [4]. In the same year, AIR&D was established to integrate industrial resources for achieving AmI idea [5]. Artists also pay much attention to the concept of AmI. Marshall McLuhan, Guy Debord, and Jean Baudrillard successively queried the impact of media technology. These scholars claimed that the spirit of the Arts should not be puzzled by technology [6], which matches the main idea of AmI. Briefly speaking, the main concept of AmI is to make technology humble and invisible existence in our living space.

Based on the concept of AmI, we developed an interactive environment-sensing bar, “WSN-bar”, which integrates wireless sensor network (WSN), environmental sensing, human locating, multi-touch, and vision-based tracking technologies. We collected the information from sensors and visualized them into art animation. Participants can enjoy

WSN-Bar by art appreciation and be aware of the environmental variations through the animation in the bar.

Because of the maturity of WSN technology, WSN-Bar can be developed by applying smaller, smarter, and cheaper sensing and computing devices. It makes WSN-Bar easily to achieve the characteristics of AmI. We mainly used WSN wireless transmitting and locating system to achieve intelligence, ubiquity, and context awareness and we utilized the bar-like appearance and small WSN routers to attain invisibility; we applied the intuitional interaction system to implement natural interaction.

Moreover, we developed some vision-based tracking technology to make WSN-Bar with multi-touch interface. Previously iBar [7], MERL's DiamondTouch table [8], and Microsoft Surface [9] have developed multi-touch table. But our WSN-Bar is the first interactive bar applying WSN technology and can reflect the change of the environment and increase communications between human and natural environment.

WSN-Bar has two modules: "Garden of Light" and "Vivacious Bushes". In "Garden of Light" module: WSN-Bar can receive the environmental data from brightness, temperature, and CO₂ density sensors which were built outdoors by WSN short-distance wireless transmitting technology. Then WSN-Bar transfers the collected information into the interface of WSN-Bar to make interactive flower images on the bar. In "Vivacious Bushes" module: WSN-Bar can detect people's movement by WSN locating technology. It also symbolizes the information by bushes and butterflies. Butterflies' flying among different bushes signifies people's moving among different rooms in the building.

According to what has been mentioned upon, WSN-Bar is not only a practical, meaningful and innovative bar, but also an appreciative and interactive art work.

2 Concepts of Interactive WSN-Bar

In this section, we will describe the concepts of how we made these two modules, "Garden of Light" and "Vivacious Bushes". In WSN-Bar, participants can switch them using a cup as a marker and can communicate with natural environment easily by operating its multi-touch interface. We are glad that WSN-Bar also creates opportunities to reduce the estranged gap among the participants when people interact with it.

WSN-Bar video: http://techart.tnua.edu.tw/eTaiwan/contents/elif_wsn-bar.html

2.1 Module 1 – Garden of Light

The content of the multi-touch interface is an interactive garden. Flowers in the garden can move, grow, and change their colors to show the variations of outdoor brightness, temperature, and CO₂ density. Through WSN technology, the environmental sensors can detect and transmit the data from outdoors to the bar. By the artistic flowers, the participants can easily realize the real-time outdoor environment condition (Table 1, Fig. 2). The change of flowers' colors looks like using different colorful light to bright the garden, so this module is called "Garden of Light".

WSN-Bar contains vision-based tracking technology [10] that we have developed, so participants can interact with it by multi-touch interface. As soon as participants put their hands or objects on the top of WSN-bar, the flowers will track the objects



Fig. 1. The plastic design and two modules

and congregate slowly to their positions, which makes WSN-Bar interactive and entertaining.

Table 1. Relationships between environmental data and image content

	Temperature	brightness	CO ₂ density
High	Flowers get warmer color	Flowers bloom	Plants become single color
Low	Flowers get colder color	Flowers wither	Plants become full color

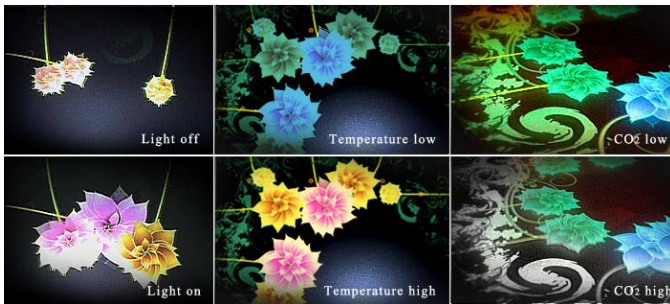


Fig. 2. Relationships between environmental data and image content



Fig. 3. The flowers congregate slowly to the position of participant's hand

2.2 Module 2 – Vivacious Bushes

ZigBee is the core WSN technology we used. WSN-Bar can detect the amount and positions of people automatically through ZigBee routers set in each room and ZigBee tags worn by each person.

The content of the multi-touch interface is bushes and butterflies. Bushes and butterflies symbolize the rooms and the people in the building. Each bush corresponds to each room; each butterfly corresponds to each person. When people move from one room to another, butterflies will fly from one bush to another correspondingly. In addition, people’s identity numbers are shown in butterflies’ flying paths at the same time (Fig. 4).

Just like the “Garden of Light” module, when participants put their hands or objects on the top of WSN-bar, bushes will trace the objects and congregate slowly to the object positions (Fig. 5). The bushes move according to user’s behaviors, so we call this module “Vivacious Bushes”.



Fig. 4. People’s movements are according to butterflies’ movements



Fig. 5. The bushes trace participants’ hand positions

3 Interactive Technology

3.1 Vision-Based Tracking System

Vision-based tracking technology is the core of WSN-Bar’s multi-touch system. There are infrared projectors and an infrared camera installed inside the bar (Fig. 6). WSN-Bar detects the positions of participant’s hands and objects on it by receiving

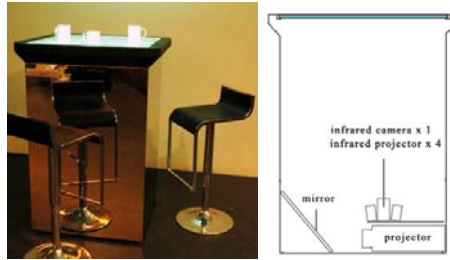


Fig. 6. The layout inside WSN-Bar

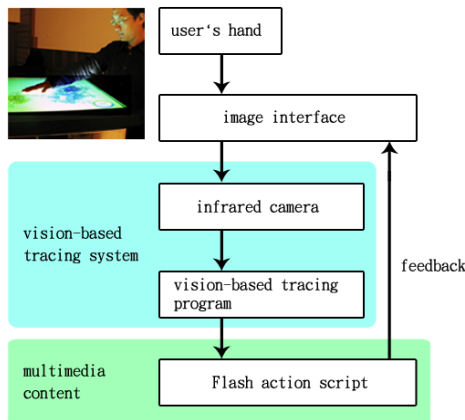


Fig. 7. The procedure of the vision-based tracing system

the reflection of the infrared. After receiving the position information, Flash program will change the media content according to the data. And then participants can see and appreciate the result of the interaction (Fig. 7).

3.2 Transmission and Positioning of WSN

ZigBee technology used in WSN-Bar was developed by Networks and Multimedia Institute of III (Institute for Information Industry) [11]. It follows the standards of the ZigBee Alliance and IEEE 802.15.4. It has the abilities of power saving, high stability and two-way transmission. It has been widely applied in health-care, access, and environmental sensing system. WSN-Bar primarily uses the environmental sensing, data transmitting, and locating functions of ZigBee technology.

The data from the brightness, temperature, CO₂ density sensors and ZigBee tags are sent to ZigBee routers, which transmit the data from one to another, and finally send the information to the ZigBee coordinator. After receiving the data from routers, the coordinator transmits these data to the ZigBee server. Finally, Flash program controls the changes of multi-touch interface according to the data passed by XML socket from ZigBee server (Fig. 8).

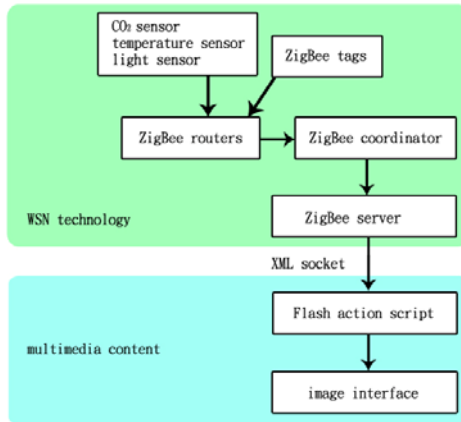


Fig. 8. The transmission procedure of WSN-Bar

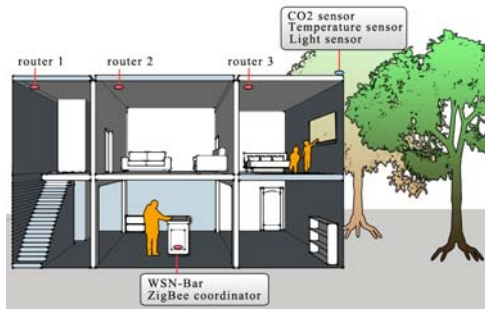


Fig. 9. The positions of WSN-Bar, WSN routers, and sensors

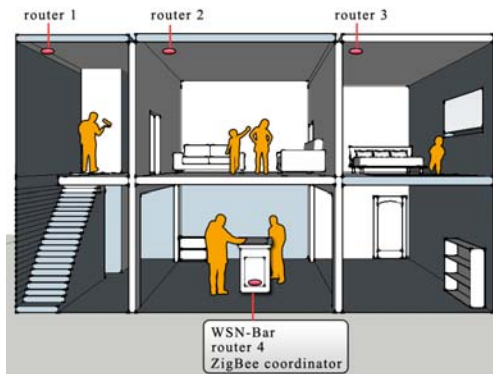


Fig. 10. The positions of WSN-Bar, WSN routers, and sensors

“Garden of Light” is primarily developed using the environment sensing technology and short-distance transmitting function of WSN technology. No matter where the bar is (Fig. 9), ZigBee routers can pass the information from outdoors to WSN-Bar. The application of WSN technology makes the concept of AmI easier to implement.

“Vivacious Bushes” can know people’s positions by comparing different signal degrees from ZigBee tags. We installed ZigBee routers about every 5~7 meters (Fig. 10) to detect ZigBee tags’ signals. Through comparing and calculating these signal degrees by RSSI-based and Area-based algorithm, WSN-Bar gets the positions of people. Therefore, participants can know other people’s positions by tracing the butterflies’ locations in WSN-Bar.

4 Conclusion

By integrating environmental sensing technology, multi-touch technology, and art design, WSN-Bar has finished up some goals: (1) Through ZigBee sensors, the bar can detect the brightness, temperature, and CO₂ density outdoors. (2) It can locate people’s movement inside the building. (3) Participants can interact with the bar under entertainment atmosphere. (4) It is an art installation embedded with the environmental information visualization.

The emergence of AmI proves that people are full of dreams and exceptions for living in the future. Undoubtedly, our WSN-Bar achieves the idea which is to make our environment with ambient intelligence. As WSN technology can be blended into the home environment naturally, people can enjoy the convenience offered from the high-tech technology. In addition, our research and design is concerned that humans have become estranged to each other in digital age. WSN-Bar creates an opportunity to reduce the gaps among the participants.

Acknowledgments. We gratefully acknowledge support for this work from the Department of Industrial Technology, Ministry of Economic Affairs (ROC) under the “Digital Life Sensing and Recognition Application Technologies” Project of the Institute for Information Industry (III) and grant 96-EC-17-A-02-S1-049 to the Center for Art and Technology, TNUA. Special thanks to the Network & Multimedia Institute of III which provided us with substantial technical support and samples of its ZigBee technology. Thanks to Chung-Fang Tsao, Chia-Wen Chen, Yu-Hsiung Huang and Yi-Wei Chia for their help with this artwork.

References

1. Weiser, M.: The Computer for the Twenty-First Century. *Scientific American*, pp. 94–100 (1991)
2. Information Society Technology Advisory Group, <http://www.cordis.lu/ist/istag.htm>
3. Weyrich, C.: Orientations for Workprogramme 2000 and beyond, ISTAG (2000)
4. MIT Oxygen project, <http://oxygen.csail.mit.edu/>
5. Ambient Intelligence R&D Consortium, <http://www.air-d.org>

6. Hsu, S.-C.: Interaction and Creation in DIGIart@eTaiwan, pp. 73–78. Ylib Publisher, Taipei (2007)
7. Mindstorm iBar, <http://mindstorm.com/solutions?category=ibar>
8. MERL DiamondTouch table,
<http://www.merl.com/projects/DiamondTouch/>
9. Microsoft Surface, <http://www.microsoft.com/surface/>
10. eBar, FBI Lab.,
http://techart.tnua.edu.tw/eTaiwan/contents/elif_bar.html
11. The project of implement sensing and identifying technology in digital living space, III,
http://www.iii.org.tw/2_2_2_1_1_2.asp?sqno=TDB200812018&yy=2008&pv=

Immersive Painting

Stefan Soutschek¹, Florian Hoenig², Andreas Maier², Stefan Steidl²,
Michael Stuermer², Hellmut Erzigkeit¹, Joachim Hornegger²,
and Johannes Kornhuber¹

¹ Department of Psychiatry and Psychotherapy, University of Erlangen-Nuremberg,
Erlangen, Germany

² Chair of Pattern Recognition, Department of Computer Science,
University of Erlangen-Nuremberg, Erlangen, Germany

Abstract. This paper presents a human machine interface, which helps elderly people learn how to become aware of their physical state and how to influence it. One of the biggest requirements for such a system is to provide an intuitive interface which does not overexert an elderly person, while also being easily accepted. Here, the connection of art and computer science offers the ideal outlet for such an interface. In our work, we show an user interface that is pleasant, expressive and does not look like the traditional computer interactions. We use classical biosignals, such as skin temperature or skin conductance, to get the necessary information of the physical state of the user. This information is presented to the user as individual artwork, which is created from the measured biosignals and the position of a cursor. Of course, the traditional scientific graph output is also made available. Another aspect of the system is that its design, allows its smooth integration into a normal home environment, or art studio. All necessary components are off-the-shelf, commercially available products to reduce costs and to allow a quick setup time.

Keywords: Computer Science, Art, Ambient Assisted Living, Elderly People, Human Machine Interface, Biosignals.

1 Introduction

Within the next decades, the continuous improvement of medical care, the increasing standard of living in the world will result in an increase of the anticipated average age [1]. In Germany, for example, the population of people older than 80 years is the fastest growing segment of population. This segment is expected to grow from currently four million people to approximately ten million people in 2050 [2]. This implies a huge challenge to society, government and industry but also for each individual person. It is necessary to address the different kinds of limitations due to for example illnesses, injuries or natural deterioration, that occur or arise with aging. Sustainable solutions need to be created that support elderly people in their everyday activities and to provide assistance

in case of emergency. Interdisciplinary research is required for the development of solutions that help cope with the upcoming changes in our society.

Especially for systems developed for the focus group of elderly people it is important that such developments are attractive, easy to use and address individual needs. The majority of elderly persons the advantages of a new acquisitions need to be clear. Price, training period and personal benefit are only a few of the important factors, which influence their purchase decision.

In our work, we investigate a new approach, where we combine art, psychology and computer science to develop interfaces for elderly people. We utilize artwork as access to modern communication technology and additionally we want to use it as a means for promoting the collection of clinical relevant data. The aim is to direct the attention, especially of elderly people, to clinically relevant topics, to impart knowledge and to offer screening and diagnostic instruments. In this paper, we show how the combination of art and computer science can help elderly people gain an intuitive and playful access to modern technologies, without the need of complicated instruction manuals. We show that one can develop systems that make it possible to learn how to interpret and manipulate ones physical state while having fun.

2 State of the Art

An example setup that combines artwork and computer science, to arouse interest and to gain the attention of a user was developed by Shugrina et al. [3]. Here the properties of an image, e.g. coloring or style, are changed according to the classified emotion of the user. Features are continuously extracted from a face, which is recorded by a camera. This example shows that the utilization of adaptive art, independent of the actual media, has big advantages. Art as intuitive and attractive interface for the visualization of the acquired data significantly increases usability and joy. Our approach, in comparison, is based on the ideas of classical biofeedback applications, such as described in [4]. In principle, such applications, much like our system, aim at teaching the user how to interpret and manipulate his body signs. There is however a significant difference between traditional biofeedback and our setup. While classical biofeedback approaches perform a predefined action that is influenced by the change of the biosignals, e.g. show a movie of a flower (see Fig. 1) which starts to bloom [5], our system involves the user's creativity, to create unique artwork everytime it is used. The basic idea, to combine artwork with the interpretation of biosignals has also been introduced in the field of biofeedback, e.g. by [9]. But once again there, the user is just able to modify an existing artwork and not to create his own, individual artwork. One advantage of our approach is that it gives a larger freedom of expression to the user and thus keeps the fun factor elevated. One can use it for entertainment or artistic purposes instead of only a means of necessary biosignal data collection. Furthermore, while previous systems could not be integrated in a home environment, our session uses a regular TV screen as a display unit, which is available in nearly every home nowadays.



Fig. 1. The state of the flower depends on biosignals (images from [5])

3 Feature Extraction from Biosignals

As input for the painting device, we need to know, whether a user is currently relaxed or tensed up. To obtain this information, we first of all measure the skin temperature (ST) and the skin conductivity (SC) with the nexus-10 from Mind Media [5]. These signals were chosen from the available signals of the nexus-10, because they are easy to record, compared to other signals, e.g. the electrocardiogram (ECG), but still robust enough to allow for an analysis of the physical state of the user. For the ECG, we would need to place additional electrodes on the body, whereas the recording of ST and SC can be realized by small sensors, shown in Fig. 2, which are attached to the hand. The measured raw values are sent to the computer via Bluetooth, which avoids unnecessary cables and allows the user a higher range of motions.

The big advantage of choosing sensors that are easy to use is, that our system can be used without a complicated preparation, which in turn would lead to a loss of acceptance. There are already systems under development, where the biosignals can be collected via a glove [6], a vest [7] or a T-shirt [8] in a comfortable way. Such devices can also be integrated in our system in the future so as to obtain additional signals, like the ECG, or the respiratory frequency in a comfortable way. Our system is flexible enough to allow the possibility to process additional signals besides skin temperature and skin conductivity. As



(a) Nexus-10

(b) Temperature sensor

(c) Skin conductance sensor

Fig. 2. Sensors for acquiring skin temperature and skin conductivity (images from [5])

soon as these devices become commercially available and can be comfortably and intuitively used by elderly persons, it will be investigated, if these devices can be integrated in future versions of our painting device.

To ensure that the measured signals have the consistent influence on the painting process, whether the user tends to relax or tends to tense-up, the nexus-10 is calibrated and the received raw values are normalized. The normalized signal values $s(n)$ of each sensor n are multiplied by a sign factor $f(n)$, which guaranties that an increase of the values implies a tendency for tension and a decrease of the values indicates relaxation. Afterward they are summed up to a total sum S , which is the sum over all corrected signals received from the number of available sensors N .

$$S = \sum_{n=1}^N f(n) \cdot s(n) \quad (1)$$

The sum S in Eq. 1 is calculated for each time frame t . In the next step the sum of the current frame is compared to the average A sum of the last five frames. If the current sum is larger than this average, this indicates that the user has become more tense. If the sum of the current frame is smaller than the average, then this is an indication that the user is more relaxed. With this information one part of the necessary information that is needed as input for the painting device is available. The workflow of the device is shown in figure 3.

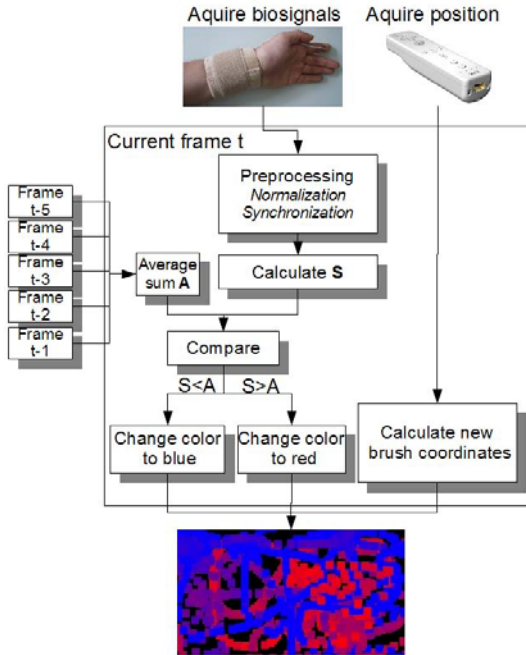


Fig. 3. Workflow of the painting device

4 The Painting Device

To finally setup the Painter, the position of a paint brush and its color need to be defined. To realize the paint brush, we currently use a Wii Mote™ [10], which is used as pointing device. The user just needs to point to the direction of the desired spot with this device to define the position of the paint brush. For people, who are used to a computer, it is also possible to use a standard mouse device to define the current position of the brush.

The color of the paint brush is defined according to the sensed psychological state as described in Section 3. The information of the state of the user, relaxed or tensed up, is used to shift the current color between two colors for each frame. To avoid the case where a single wrong measurement in one frame could result in a large change of the color, the color only changes gradually from frame to frame. In the images shown in Fig. 4, the colors are set to vary between blue and red. As long as the user relaxes, the color of the paint brush continuously shifts from red to blue over time. If the user is tensed, the color changes from blue to red. The user is always able to interrupt the painting, just by pushing a button on the Wii Mote or on the computer mouse. This enables the user to precisely position the brush before the point is colored. When the user is pointing to a position, which is already colored, the mean of the old and the current color is set. If the painting is finished, the user can stop the painting, the image is saved and one can then either start with the next one or quit the software.

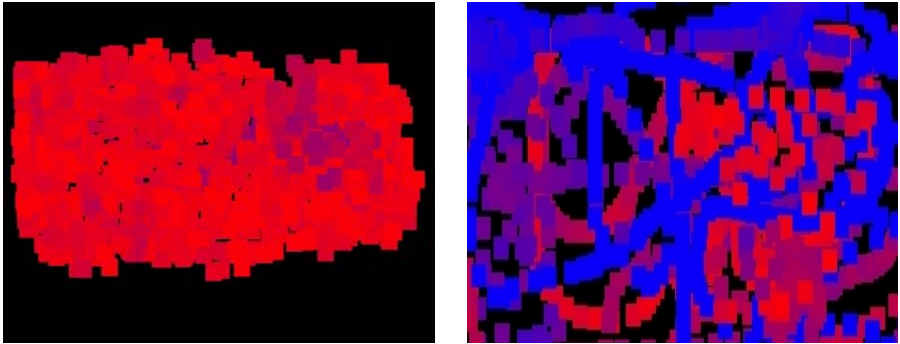


Fig. 4. Individual artwork created with the painter

5 Conclusion and Future Work

With our approach we show one possible way to combine artwork and computer science to build attractive and intuitive applications that motivate usage and direct the interest to clinical relevant data at the same time. People, who use this application, playfully learn, how to interpret and manipulate their psychological state. Additionally, we are able to integrate our system into a home environment. This is important, as future work will focus on two main applications.

We want to use the different biosignals from the available sensors to enhance the existing painter. These different signals could be utilized to control the input to the artwork. For example, one signal could be used to choose the color, another signal could be used to choose the type of the paint brush. Additionally, the time the brush is at the same spot in the image could determine the thickness of the brush. We also plan to have artists use our prototype and guide us in the next version of our immersive painting method.

Furthermore, we want to provide a continuous training, which can be utilized as a screening instrument. It will be investigated, if our system can be used as platform to build a screening instrument to support the early diagnostic of dementia. Therefore, a set of pictures of known persons or events of the past will be shown to an elderly person on the TV screen. As reference, also pictures with neural content will be in this selection. For each picture, known or neutral, an artwork from our painting device is created, but in contrast to the described painting device, the artwork will be created in the background. After each session these artworks can be used to compare the session with previous ones. Although, the spatial resolution of the artwork and the delay in the change of the biosignals will probably not be sufficient to give a detailed prediction, which detail in the picture had the most influence, we expect the user to show different reactions, between known and unknown pictures. With the additional information, which picture shows known persons or events to the user, we will examine, if a system can be set up, which automatically detects, whether a picture is known or unknown to the elderly person with the help of the created artwork. The aim of this enhancement is to gain continuous information if the memory of an elderly person acts according to the age.

Acknowledgement

This research project is supported by the BFS¹ Bavarian Research Association “Zukunftsorientierte Produkte und Dienstleistungen für die demographischen Herausforderungen - FitForAge”².

References

1. Commission of the European Communities, Green Paper - Confronting demographic change: a new solidarity between the generations, COM(2005), 94 final (2005)
2. Eisenmenger, M., et al.: Bevölkerung Deutschlands bis 2050. 11. koordinierte Bevölkerungsvorausberechnung. Statistisches Bundesamt, Wiesbaden (2006)
3. Shugrina, M., Betke, M., Collomosse, J.: Empathic painting: interactive stylization through observed emotional state. In: NPAR 2006: Proceedings of the 4th international symposium on Non-photorealistic animation and rendering, Annecy, France, pp. 87–96 (2006)

¹ Bayerische Forschungsstiftung.

² www.fit4age.org

4. Horowitz, S.: Biofeedback Applications: A Survey of Clinical Research. *Alternative and Complementary Therapies* 12(6), 275–281 (2006)
5. <http://www.mindmedia.nl/english/index.php> (06.07.2009)
6. Peter, C., Ebert, E., Beikirch, H.: A Wearable Multi-sensor System for Mobile Acquisition of Emotion-Related Physiological Data. In: *Affective Computing and Intelligent Interaction*, pp. 691–698. Springer, Heidelberg (2005)
7. Paradiso, R., Loriga, G., Taccini, N.: A wearable health care system based on knitted integrated sensors. *IEEE Transactions on Information Technology in Biomedicine* 9(3), 337–344 (2005)
8. Pola, T., Vanhala, J.: Textile electrodes in ECG measurement. In: *ISSNIP 2007 Third International Conference of Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, Australia (2007)
9. Tirel, T., Hahne, S., Garancs, J., Muller, N.: BIOS - Bidirectional Input/Output System. In: *VISION - image and perception*, Budapest, Hungary (2002)
10. <http://www.nintendo.com> (06.07.2009)

Sonic Onyx: Case Study of an Interactive Artwork

Salah Uddin Ahmed¹, Letizia Jaccheri¹, and Samir M'kadmi²

¹ Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim 7491, Norway

² Artist, Oslo, Norway

{salah, letizia}@idi.ntnu.no, s-mkadmi@online.no

Abstract. Software supported art projects are increasing in numbers in recent years as artists are exploring how computing can be used to create new forms of live art. Interactive sound installation is one kind of art in this genre. In this article we present the development process and functional description of Sonic Onyx, an interactive sound installation. The objective is to show, through the life cycle of Sonic Onyx, how a software dependent interactive artwork involves its users and raises issues related to its interaction and functionalities.

Keywords: Interactive artwork, software dependent artwork, case study, artist technologist collaboration.

1 Introduction

The relation between art and computer dates back to the 60s. The intersection of art and software interests, includes and attracts people with diverse background to come together and work in common projects [1]. The use of computer in art has increased with time covering computer graphics, computer composed and played music, computer-animated films, computer texts, and among other computer generated material such as sculpture. The development of software based sound technology and the interaction with art can relate back to a pre-digital era, which goes from the Italian futurist, Luigi Russolo (1913) to John Cage, Pierre Schaeffer, Pierre Boulez, Olivier Messiaen, Karlheinz Stockhausen, Iannis Xenakis, etc. [2]. MUSICOMP (Music Simulator Interpreter for Compositional Procedures), one of the first computer systems for automated composition was written in the late 1950s by Hiller and Robert Baker [3]. Popular software for creating interactive sound installations in recent years includes Max/MSP and Pure Data. Often these kind of interactive artworks rely heavily on software applications for realizing artistic expressions and interaction. Like Oates [4] who looks at computer artworks as Information Systems and proposes to extend Information System research agenda to include computer art, we suggest to extend software engineering research to include software dependent art projects. Our experience from projects involving both artists and software engineers such as *Flyndre* [5], and *Open Wall* [6] shows that software engineering can play important role in such interdisciplinary projects [7]. The struggle of software engineers to make technology more accessible to artists is recognized by researchers [8]. Besides, researchers have addressed collaboration between artists and technologists as an important

issue in interdisciplinary projects which is often challenging due to the differences among the involved parties [9]. In this article we present the case study of Sonic Onyx by describing its technical and functional features and showing how it raises different issues related to its development, interaction and user experience.

The rest of the paper is organized in following way; section 1.1 describes our research background and research method. Section 2 gives technical and functional description of Sonic Onyx, where 2.1 describes the hardware and 2.2 describes the software components of the artwork and 2.3 describes how it works. Section 3 addresses the development context. In section 4 we evaluate the project and identify the issues raised by the project during and after development. This is in a way our lessons learned from the project. Section 5 concludes the article with indication of possible future extension of Sonic Onyx.

1.1 Research Background and Method

The work presented in this article is part of SArt research project at the intersection of art and technology [10]. The objective of SArt is to assess, develop, and propose model and process for developing art using technology and software. Sonic Onyx was chosen as a case study for SArt because of its involvement of both artist and technologists and its placing as a public art in a school with a purpose to use and experiment the sound installation with the pupils of the school. Besides, Media Lab at Norwegian University of Science and Technology (NTNU) was also involved in the process of development of the artwork. The objective is to find answers to the following research question set by SArt, *“What are issues and challenges that software engineering has to tackle to implement software dependent art projects?”*

The software of Sonic Onyx was created by a group of computer science students. The development lasted six months and researchers from SArt took part in the process as observers. They took part in the group meetings and received copies of all documentations produced during development. The case study was done following research strategy defined by Oates [11]. The type of the case study can be defined as explanatory and longitudinal with duration of six months. The data presented in this article about the project is collected from i) interviews, ii) project report, iii) project documentations and iv) participation in the group meetings. Interviews were conducted on the developers, the sponsor and the artist who is also an author of this article. Feedbacks from users were collected from school teachers. Documents consist of pre-project report, weekly reports, meeting minutes and the final project report. The developers of the project created a status report every two weeks. They also created meeting minutes on each meeting. Besides, project activities were closely observed by taking part in the project meetings.

2 Sonic Onyx – An Interactive Art Installation

Sonic Onyx is an interactive sound sculpture placed in front of a secondary school in Trondheim. The sculpture is about four meters high and the diameter of the “space” is about seven meters. There are fourteen static metal arms and a globe like sphere in the middle. The sculpture has seven loudspeakers located in seven arms and a subwoofer

located in ground at the center making it possible to provide 3D sound effects within the space created by the sculpture. People can communicate with the sculpture by sending texts, image and sound files from their Bluetooth enabled handheld devices such as mobiles or PDAs. These media files are processed and converted into unrecognizable sound that is played by the sculpture. The artwork was commissioned by the Municipality of Trondheim based on the following propositions made by the artist: i) the artwork should be interactive ii) use latest technology and iii) allow learning and pedagogy.



Fig. 1. Sonic Onyx during night and daytime

2.1 The Hardware

Sonic Onyx consists of four main parts: i) Central server ii) Bluetooth server iii) Sound system and iv) Light system. Central server is a desktop computer with high configuration hardware. It's main task is to store art music, run the main application and support remote administration. The interactive part of the sculpture is based on Bluetooth technology that allows spectators to communicate with it by i) sending instructions to handheld devices in proximity of the sculpture and ii) receiving files from the handheld devices.

Table 1. Hardware components of Sonic Onyx

Part	Configuration/Tools
Server	PC with Dual Core 64 bit CPU, OS: Linux Debian
Bluetooth Systems	Bluegiga WRAP access server with 3 Bluetooth radios
Sound Systems	Sound Card - M-audio Delta 1010 Speakers - CAP-15 (7 piece) Subwoofer - TIC GS50 omni Power Amplifier - IMG STA 1508 Microphone mixer - Moncor MMX-24
Light System	DMX 512 Lanbox-LCX system

A Bluegiga WRAP access server [12] makes the Bluetooth system. It has an Ethernet interface which makes it possible to place it inside the sphere of the sculpture while the main server resides in a building close by. The sound system was carefully chosen since the installation is meant to be operational for many years in outdoor setting. The

speakers need to have robust design, good sound quality and outdoor specification. The light system consists of a LanBox DMX controller which controls the light located in the sphere of the sculpture [13]. The changing of color of the light is programmable

2.2 The Software

The main program which is called “Spamalot” runs in background all the time and calls other subprograms when necessary. The application has two major tasks which are: to receive files from the Bluegiga Bluetooth server and to process the files in order to modify and/or generate sound. The processing of the received file is done by Pure Data (called Pd here after) patches. Pd is an open source graphical programming language used for interactive computer music and multimedia works [14]. Modular, reusable units of code written in Pd are called ‘Patches’. The transfer script is a shell script used to rename, timestamp and copy the files received from the Bluetooth server. The script is executed every time a file is received.

Table 2. Software components of Sonic Onyx

Name	Description	Tools Used
Spamalot	265 lines of code	Python
Processing Part	Pd patches – 20 patches for Sound, 13 for Image and 4 for text	Puredata, pd-zexy, pd-audio, gem, decoders, speech synthesizers
Transfer Script	60 lines of code	Linux Shell
Web site	5 pages and a guestbook	PHP, MySQL,

Besides the main application for running the system, a website was considered as an important part of the project which will provide a) remote administration of art music, b) provide sample sound files online and c) a guest book. Thus it is an extension of the artwork that would allow flexibility in administration of the sculpture sound system. The website that was primarily built using PHP and MySQL and it supported uploading music files and guest book but no remote administration.

2.3 How Does It Work

The working process of the artwork can be described sequentially as follow: First, a user sends a file from Bluetooth enabled device. The Bluetooth server inside the globe receives the file and sends it to the central server located in the school building. Obexsender [15] software that comes with the Bluegiga access server is used to transfer the file.

An NFS-share is mounted on the main server as part of the access server file system; in that way all incoming files are copied to the nfs-shared drive of the server. Once the user file is in the server, it is converted into an audio file by the application ‘Spamalot’. Appropriate file conversion libraries are called and many encoders and decoders are used in order to support a wide range of file formats. In case of text file, it is converted into wav file. But in case of image it is converted into a jpeg which is used to manipulate a preselected audio file by reverb and pitch shifting depending on

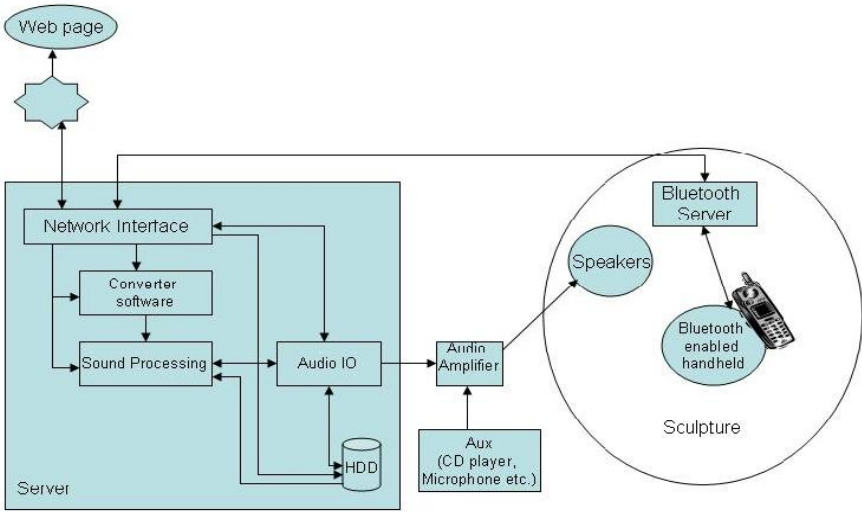


Fig. 2. The architecture of Sonic Onyx sound system

the color values of the picture. In case of sound file, it is processed and modified using random Pd patch. The modified file is then played by the sculpture using the amplifiers and loudspeakers.

3 Development Context of the Project

The project was developed by a group of five computer science students from Trondheim and Sør-Trøndelag University College (HiST) as part of their bachelor’s project. The group had some prior knowledge on programming, electronics, web design and project work but no experience of working in art projects. Besides the developers, there was one project supervisor who is a teacher at HiST, the artist, who is the main client of the project and a technical consultant from Media Lab, NTNU.

The development process that the group have followed can be described as an agile or rapid development method more specifically [16]. They chose Python and used DrPython IDE [17] to take the advantage of Python’s fast testing, and debugging opportunities. They followed a two week iteration process with meetings every second week and created a report documenting work done in last iteration as well as the plan for next iteration. Communication between artist and the developers were pretty good. Artist’s familiarity with technology and previous experience with similar project made the communication smooth. But since the artist was not living in the same town, he was invited to the meetings only when the developers had something significant to consult him. Communication among developers was easy since they were all students of the same class and many of them knew each other before. As the developers had little experience and technical skill needed for this kind of artwork, a technically experienced person was needed as a consultant in the development team. Experts from Midgard Media Lab [18] took part in group meetings and helped the

group regarding domain related knowledge and expertise. Technologists also took part with the artist in the conceptualization phase to define the concept of the artwork. The functionalities of the system were defined, modified and changed through explorations.

4 Evaluation of the Project

The project is special in a way that students did not have previous experience of working artist and on the other hand artist did not work with student developers. But all in all, the project ran smoothly. The management of the project was good; everyone was serious about the meetings and project deliveries and collaboration between all members was very good. The group was very optimistic during the development period. The project was a success in a way that students covered most of the requirements except a few due to lack of time or load of complexity. For example, sending a welcome instruction by the Bluetooth receiver was not completed. Besides, the web site [19] of the project was not completed. The website made by the group was not used by the artist mainly because: i) he did not like the layout and the design of the web site and ii) the site was incomplete.

Unlike traditional sculpture or painting, the development of interactive art does not end with the completion of the artwork; rather it continues enhancement and changes with users' responses and interaction with the sculpture. Besides, software dependent artwork needs timely maintenance and upgrading. Interactivity and experimenting with artwork increases the importance of maintenance and upgrading.

Sonic Onyx was built with the objective that pupils from the school will use it and experiment different possibilities around the artwork. It was not meant to be an end product, rather a developing one which will evolve with response to the users. Often the sponsors of the artwork do not consider this difference and as such they overlook the after delivery maintenance and upgrading of the system. This might be because of their lack of awareness of maintenance and upgrading issues related to the interactive installations or their unfamiliarity with technology dependent artworks. So the artist has to take care of this factor and make the sponsor understand the necessity of maintenance and upgrading. In case of Sonic Onyx, the artwork came across some after-ward maintenance problems such as hardware and configuration related problems. Since there was no one responsible for maintaining, it was hard for the school to use it continuously without interrupt. Placing an artwork in the public space specially in outdoor setting without any supervision bring other safety issues, such as the sub-woofer placed in the ground was destroyed by the pupils from the school even though it had protective cover.

Locating the artwork in a public space with a neighborhood raises some issues due to its interruption of the environment such as how loud should it play, which time it will play, when the light should be turned off during day time etc. These issues are based on users' feedbacks which actually evoke enhancements and modification of the artwork and its software.

Involving students in the project as part of a coursework has certain issues such as limitation of time. Besides developing the system, students need to get a good grade in the course. In sonic Onyx, when a part of project was lagging behind due to manufacturing delay of the arms, students could not wait till the end for deploying the sculpture and the software system on the site. This was a bit surprising for the artist; on the other hand students did not have any contract in the project, so they were not

bound to wait until everything was ready. The artist was expecting a proper delivery of the system including in site deployment and testing.

Backing up of the whole system at regular interval is important especially for a system like Sonic Onyx which is supposed to be upgraded. For a complex interactive system to carry on experimenting and adding, updating or modifying functionalities it is necessary that the artist can roll back to a safe state after any kind of problems during the experimentation. The system can also go down or crash any time and back up is important to restart the system. Otherwise it can be a complete mess with the application.

Sonic Onyx is targeted to involve school pupils with music activities. Besides playing files received from user's handheld devices, it provides the possibility to add auxiliary instruments such as microphone, mixer etc. Students can interact with Sonic Onyx directly through its MIDI interfaces to control both light and sound allowing them live performances. In this way it allows students of the school to create and store music in the server, play directly and or mix different instruments with the help of the sculpture. Thus the artwork not only enhances the school yard with its aesthetics value but also involves the students with activities for learning and practicing music.

When in rest, the sculpture plays songs stored in the hard drive of the server. The idea is to promote contemporary artists by playing their music in stand by mode and bringing them closer to the audience. The web site of the project is planned to give the artists the possibility to register, create account, and upload music files which will be later played by the sculpture. In this way, besides the students it has the possibility of engaging contemporary artists as well.

5 Conclusion

In this article we have presented the technical and functional descriptions of an interactive artwork. We have presented here how artwork is very similar to a computer application and raises many issues related to maintenance, upgrading, interaction, and development process. Technology dependent artwork placed in a public place might have similar issues that deserve consideration. We have also presented through Sonic Onyx how interactive art has the possibilities of engaging people in an artful way. We have shown how the artwork can involve students in learning and playing with technology as well as attract contemporary artists to use the artwork to play their music. Future work of Sonic Onyx includes live streaming of the playing of sculpture through the project web site. Thus it will allow contemporary artists to upload, store and administer music files through their web account and reach to the audience through artwork playing their music. It would be also a nice extension to have the possibility of user sending request via SMS or Bluetooth from the site of the sculpture to play certain music. Based on user request ranking of music and playlists can be made which will help engage all participant more enthusiastically.

References

1. Ahmed, S.U., Jaccheri, L., Trifonova, A., Sindre, G.: Conceptual framework for the intersection of software and art. In: Braman, J., Vincenti, G., Trajkovski, G. (eds.) *Handbook of Research on Computational Arts and Creative Informatics*. Information Science Reference, p. 26 (2009)

2. A Brief History of Algorithmic Composition, <http://ccrma.stanford.edu/~blackrse/algorithm.html>
3. Techniques for Algorithmic Composition of Music, <http://alum.hampshire.edu/~adaF92/algocomp/algocomp95.html>
4. Oates, B.J.: New frontiers for information systems research: computer art as an information system. *European Journal of Information Systems* 15, 617–626 (2006)
5. Trifonova, A., Jaccheri, L.: Software Engineering Issues in Interactive Installation Art. In: *Designing Interactive Systems Conference (DIS 2008)*, Cape Town, South Africa (2007)
6. Ahmed, S.U.: Achieving pervasive awareness through artwork. In: *3rd International Conference on Digital Interactive Media in Entertainment and Arts*, pp. 488–491. ACM, Athens (2008)
7. Trifonova, A., Ahmed, S.U., Jaccheri, L.: SArt: Towards Innovation at the intersection of Software engineering and art. In: *16th International Conference on Information Systems Development*. Springer, Galway (2007)
8. Machin, C.H.C.: Digital artworks: bridging the technology gap. In: *Proceedings of the 20th Eurographics UK Conference, 2002*, pp. 16–23. IEEE Computer Society, Los Alamitos (2002)
9. Meyer, J., Staples, L., Minneman, S., Naimark, M., Glassner, A.: Artists and technologists working together (panel). In: *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 67–69. ACM Press, San Francisco (1998)
10. SArt Project Web Page, <http://prosjekt.idi.ntnu.no/sart/>
11. Oates, B.J.: *Researching Information Systems and Computing*. SAGE Publications Ltd., Thousand Oaks (2006)
12. Bluegiga Access Servers, http://www.bluegiga.com/more_229x
13. LanBox DMX Controller, <http://www.lanbox.com/>
14. Pure Data, <http://puredata.info/>
15. Access Server: User's and Developer's Guide, <http://bluegiga.com/as/current/doc/html/c2034.html>
16. Asphaug, T., Bielsa, C., Grytå, E.M., Thorsrud, E., Tollefsen, T.: Sculpture with 3D Interactive Sound. Department of Electrical and Computer Engineering, Faculty of Technology, vol. Bachelor HiST, p. 110 (2007)
17. DrPython, <http://drpython.sourceforge.net/>
18. Midgard Media Lab., <http://www.ntnu.no/midgard>
19. Sonic Onyx, <http://www.soniconyx.org>

An Interactive Concert Program Based on Infrared Watermark and Audio Synthesis

Hsi-Chun Wang¹, Wen-Pin Hope Lee², and Feng-Ju Liang¹

¹ Department of Graphic Arts and Communications, College of Technology

² Department of Music, College of Music,

National Taiwan Normal University, Taipei, Taiwan, 10610, R.O.C.

hsiwang@ntnu.edu.tw, {chimolee, aluileung}@gmail.com

Abstract. The objective of this research is to propose a video/audio system which allows the user to listen the typical music notes in the concert program under infrared detection. The system synthesizes audio with different pitches and tempi in accordance with the encoded data in a 2-D barcode embedded in the infrared watermark. The digital halftoning technique has been used to fabricate the infrared watermark composed of halftone dots by both amplitude modulation (AM) and frequency modulation (FM). The results show that this interactive system successfully recognizes the barcode and synthesizes audio under infrared detection of a concert program which is also valid for human observation of the contents. This interactive video/audio system has greatly expanded the capability of the printout paper to audio display and also has many potential value-added applications.

Keywords: Data hiding, 2D barcode, digital image processing, halftoning.

1 Introduction

For more than a thousand years, paper has been esteemed as the most popular interface [1] between human and information, like texts, graphics and music notes. As the advance of digital technology, a major part of information carrier has been shifted to internet and disk storage. However, most people still enjoy reading printed books/newspapers/magazine in a traditional way. In this research, the printed concert program is designed for both human vision of the printed texts/graphics contents under visible light and computer vision of embedded musical scores under infrared detection. Then, the concert program is not just a 2D printed medium but also an interactive interface to play some typical music notes in the concert. Some techniques involved in this printed concert program are reviewed in the following subsections.

1.1 Digital Halftoning Technique

Halftoning is a traditional printing process. Due to the limited tone which the output device can provide, a con-tone image has to be halftoned or binarized before it can be actually output. Since human eyes have the ability to integrate the neighboring halftone

dots, the output binary image is perceived like a continuous tone image. With the fast innovation of computer technology, digital halftoning technique [2] is adopted by nowadays printing industry and desktop printers. In general, methods of digital halftoning fall into one of two categories: amplitude modulation (AM) and frequency modulation (FM), shown in Fig. 1.

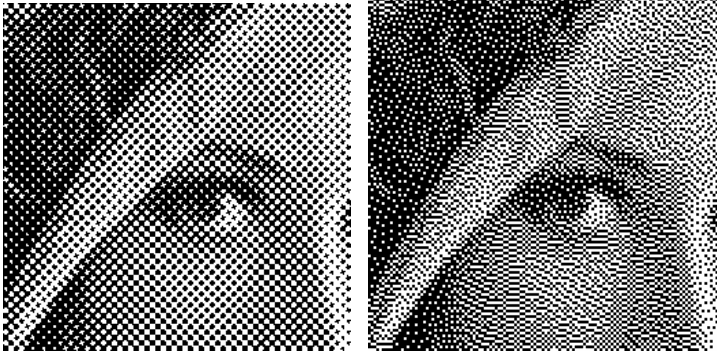


Fig. 1. Typical halftone images. Left: Amplitude modulation (AM), and Right: Frequency modulation (FM).

The amplitude modulation is also known as ordered dithering, and the smallest individually printed halftone dots form a regular pattern in Fig. 1(a). The original gray-scale image is divided into many 8×8 or other sized blocks to compare with the entries in the threshold matrix and to complete the binarization. On the other hand, the frequency modulation screening by error diffusion (shown in Fig. 1(b)) takes a totally different approach to generate a binary image for reproducing a continuous tone image. Although the computational complexity is higher for error diffusion, frequency modulation can provide better spatial resolution of the details.

1.2 Digital Watermarking in Printed Medium

With rapid development of modern computer technology, multimedia and network, researches on information hiding in image, audio and other media have attracted lots of researchers [3,4], including at MIT Media Lab. Due to limited depth for hiding information, data hiding in bi-level image is considered much more challenging and less addressed in the related research fields. However, data hiding in bi-level image is indeed on demand because printing and some display systems are in bi-level. Most of all, data hiding especially has great potential in security document printing, and may significantly help to strike counterfeiting and illegally duplicating. Works have reported several document protection schemes to modify the digital halftoning processes and to obtain deformed or shifted halftone dots to encode the hidden data [5,6]. Wang et al. [7] also blend the AM/FM halftone dots to an anti-copy security printed image and invisible marker design for augmented reality (AR). By careful manipulation of the formation of halftone dots, various applications to obtain secured

document can be derived. It is the key technology to link the computer digital world and printed physical document.

1.3 The Characteristics of Inks under Infrared

Carbon black, which consists of pure element of carbon, is known as a common black pigment. Modern ink-jet printer uses it as the major presentation of black (K) ink. Carbon appears dark because it absorbs almost all wavelength of light from ultraviolet to infrared of the spectrum. In contrast, inks of cyan (C), magenta (M) and yellow (Y) absorb much less infrared light. The transmission percentage for each ink at different wavelength is shown in Fig. 2.

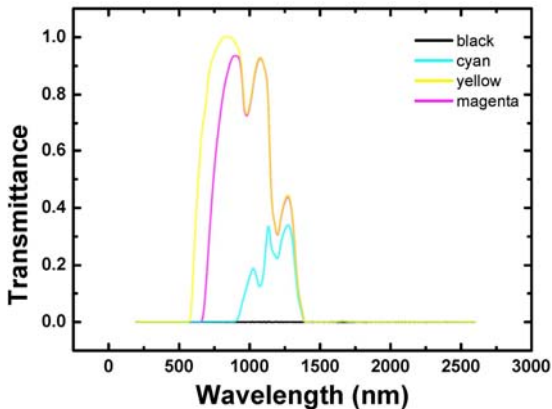


Fig. 2. The transmission percentage for each ink

There are some specially-manufactured fluorescent/infrared inks which can display distinct difference under visible light and infrared light [8]. With carefully manipulating the CMYK halftone dot distribution, watermark can be hidden in a printed image without involving special material. By using of the optical characteristic of carbon black, ImageSwitch was proposed by researchers in Japan [9]. It's not only using the digital watermark but also using the irregular shaped cell to allocate the CMYK halftone dots. The hidden pattern or watermark can be observed under the infrared illumination.

2 Methods

This research combines audio/image processing, watermark embedding, image capture and content design. The multi-disciplinary feature especially fits quite well with the recent development of the integration of technology and art.

The equipment used in this study is listed in Table 1. No special material, such as fluorescent ink, is involved. Desktop PC and printer can serve as the core tools to obtain the infrared watermark.

Table 1. Equipment used in this study.

Devices	Desktop PC
	HP DeskJet 1280 inkjet printer
	Infrared webcam
	PerkinElmer Lambda 900 spectrometer
	X-Loupe Portable Microscope Camera from Lumos Technology
Tools	Matlab 2007a (Programming Language Software)
	Adobe Illustrator (Graphics Software)
	Adobe Photoshop (Image Software)

2.1 Music Data Encoding

The music scores (shown in Fig. 3) are originally composed by Professor Wen-Pin Lee, the coauthor of this paper. The first 8 notes are selected to encode a 2-D barcode to be hidden in the concert program. Eight bits are used to encode one music note. Six of the eight bits are for pitch, and the other 2 bits are for the duration (rhythm), respectively. There are 64 bits in total encoded. Finally, the 64 ones and zeros is randomized and reshaped as an 8x8 square matrix which is shown in Fig. 4. Four black brackets are added to the 2-D barcode to provide the ability to register the input image. In the following procedures, this 2-D barcode is processed as a 12x12 binary square matrix. Though only the central part of the binary image sized 8x8 is the embedded music information.



Fig. 3. The original notes, scored by Prof. Wen-Pin Lee, are to be embedded in the watermark of concert program



Fig. 4. The 2-D barcode contains the music notes

2.2 Design of the Infrared Watermark

Since different multimedia elements, like texts, image, graphics, video and audio, are integrated in the digital era, the concert program in this research is also carefully

designed. It is essential that the concert program be both aesthetically pleasing and practically designed for its end application [10]. A painting is chosen as the cover image of infrared watermark. In order to make the hidden watermark invisible, some procedures have to be adopted. The composition of AM (S_2) halftone image and FM (S_1) halftone image is carried out by using a binary mask, M . The composed watermark image, S_A , can be obtained according to Equation (1).

$$S_A(i,j) = S_1(i,j) \cdot M(i,j) + S_2(i,j) \cdot \sim M(i,j) \quad (1)$$

The close-up microscopic image of the composed AM halftone dots (ink of black, K) and FM halftone dots (inks of C, M, Y) is shown in Fig. 5. At a normal viewing distance (30cm), density of black ink and density of the combination of CMY inks will be balanced. Due to the different dot gain of AM and FM halftone dots, a calibration chart is designed to get the optimum output parameters. That is, the watermark will achieve the best hidden effect for human perception.

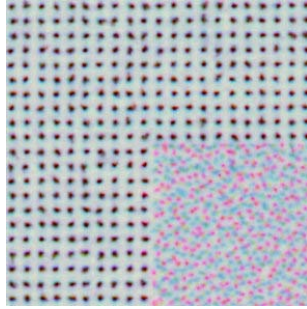


Fig. 5. The micro-structure of the AM halftone dots (ink of black, K) and FM halftone dots (inks of C, M, Y)

2.3 Audio Synthesis and System Installation

The synthetic audio signals are based on the fundamental frequency. The sinusoidal waveform is generated according to Equation (2). [11]

$$y(t) = A \sin 2\pi f t \quad (2)$$

where A is the signal amplitude, f is the musical note's frequency (in Hz) and t is time (in second). In discrete format, t can be represented as $m\Delta t$, where m is an integer (or number of samples) and Δt is the sampling period which is represented in Equation (3).

$$\Delta t = 1 / f_s \quad (3)$$

where f_s is the sampling frequency (in Hz). In general, sound is composed of sine waves and cosine waves, and its waveform is very complex. During the synthesis of audio signal in this study, the sinusoidal wave based on the fundamental frequency generates the fork's timbre. In this study, the synthesized waveforms at sampling rate of 11025 Hz are used.

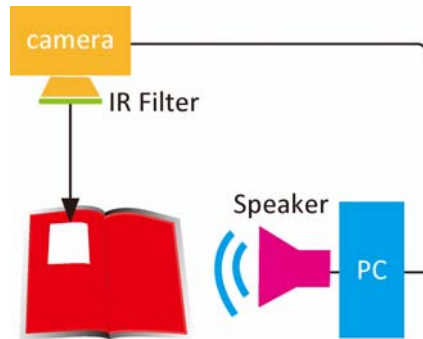


Fig. 6. The schematic diagram of experimental setup

The schematic diagram of the interactive system is illustrated in Fig. 6. It can demonstrate that an image is captured by web camera, and the data in the 2D barcode is recognized as music scores. Then the audio is synthesized and the system can play the encoded audio signals. It should be noted that, the IR filter in Fig. 6 is to block up the visible light and to let the invisible IR light pass through the filter. The CCD/CMOS has the IR detection ability in near IR range.

3 Results and Discussion

The designed concert program in this research displays very different scenes by human observation and infrared detection (shown in Fig. 7). The watermark hidden in the cover image of a painting is visible under IR detection. However, the signal of watermark is not strong enough and is subjected to various noise attacks. In Fig. 8(a), the portion of the 2D barcode has further been enhanced by adaptive histogram

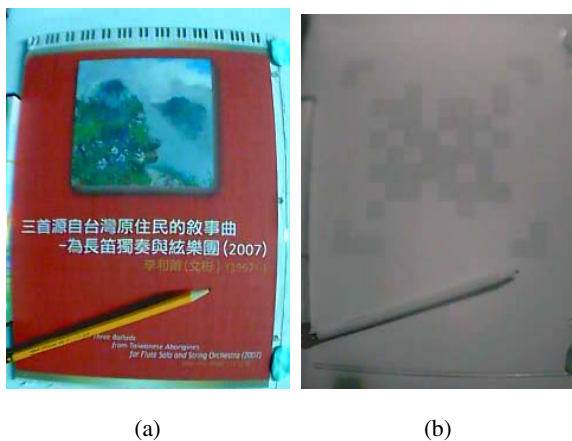


Fig. 7. The designed concert program shows distinct features by human observation and computer vision. (a) under visible light. (b) under infrared detection.

equalization to eliminate the non-uniform lighting. By digital image processing techniques, including geometrical transformation and image registration (the red points in Fig. 8(b)), the music data in the 2-D barcode is detected and recognized. The audio waveforms are synthesized according to decoded data in 2D barcode. The first 4 notes with frequency of 392Hz, 294Hz, 494Hz and 392Hz at sampling frequency of 11025Hz are illustrated in Fig. 9. The x-axis is the number of the samples and the y-axis is the signal amplitude. It shows that these waveforms have very different fundamental frequencies.

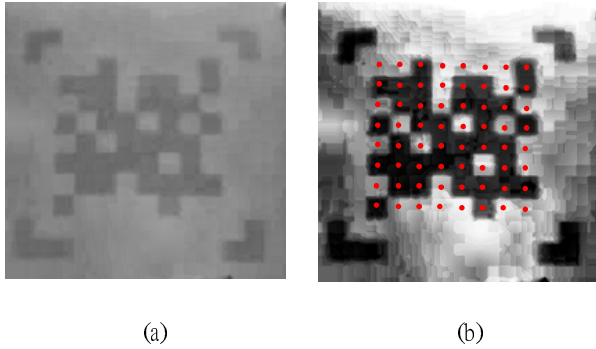


Fig. 8. (a) The close-up image of the infrared watermark. (b) The image analysis of the notes embedded in the 2D barcode.

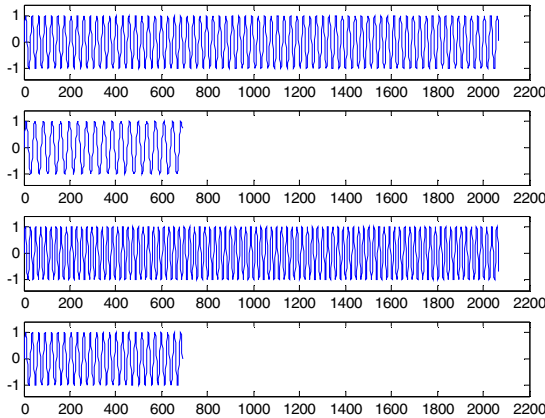


Fig. 9. The synthesized waveforms for the first 4 notes with frequency of 392Hz, 294Hz, 494Hz and 392Hz at sampling frequency of 11025Hz

4 Conclusions

The results of this research show that the proposed concert program is successfully implemented for both human vision of the printed texts/graphics contents under visible

light and computer vision of embedded musical scores under infrared detection. This interactive video/audio system has greatly expanded the capability of the printout paper to audio display and also has many potential applications to interactive art, e-learning, game development and anti-counterfeiting technology. In the future works, the system will be fine-tuned to synthesize higher quality audio sound by MIDI format or by newly developed IEEE 1599 music encoding standard for interactive application [12]. The pattern recognition technique will be used in this system to automatically extract the control points and let the user to use the mobile device such as camera phone to perform the interactive audio/video system.

Acknowledgments. The help to measure the YMCK inks' transmission spectrum in Fig. 2 from Professor Liu, Hsiang-Lin in the Department of Physics at National Taiwan Normal University (NTNU) is specially appreciated. Also thanks to the valuable assistances from Miss Liu, Wen-Hsin and Mr. Lu, Chang-Chen of Half Media Lab in the Department of Graphic Arts and Communications at NTNU.

References

1. Chiang, P.J., et al.: Printer and scanner forensics. *IEEE Signal Processing Magazine* 26(2), 72–83 (2009)
2. Ulichney, R.: *Digital Halftoning*. MIT Press, Cambridge (1987)
3. Bender, W., et al.: Techniques for data hiding. *IBM System Journal* 35(3&4), 313–336 (1996)
4. Gruhl, D., Bender, W.: Information Hiding to Foil the Casual Counterfeiter. In: Aucsmith, D. (ed.) *IH 1998. LNCS*, vol. 1525, pp. 1–15. Springer, Heidelberg (1998)
5. Hecht, D.L.: Printed embedded data graphical user interfaces. *IEEE Computer* 34(3), 47–55 (2001)
6. Sharma, G., Wang, S.: Show-through watermarking of duplex printed documents. In: *SPIE proceedings, Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, pp. 670–684 (2004)
7. Wang, H.C., et al.: Design of halftone-based AR markers under infrared detection. In: *International conference on computer science and software engineering*, vol. 6, pp. 97–100 (2008)
8. Auslander, J., Cordery, R.: Black fluorescent ink and applications. In: *SPIE proceedings, Optical Security and Counterfeit Deterrence Techniques VI*, vol. 6075, pp. 60750D.1–60750D.12 (2006)
9. Nagashima, H., Saito, K.: New Security System for ID Certificates in IT Society. In: *SPIE proceedings, Conferences on Optical Security and Counterfeit Deterrence Techniques V*, vol. 5310, pp. 142–150 (2004)
10. van Renesse, R.L.: *Optical Document Security*, 3rd edn. Artech House, London (2005)
11. Dodge, C., Jerse, T.A.: *Computer Music: Synthesis, Composition, and Performance*, 2nd edn. Schirmer, New York (1997)
12. Baggi, D., Haus, G.: IEEE 1599: Music Encoding and Interaction. *IEEE Computer* 42(3), 84–87 (2009)

Butterfly Effect Fractal

Yin-Wei Chang and Fay Huang

Institute of Computer Science and Information Engineering
National Ilan University, Taiwan
fay@niu.edu.tw

Abstract. In this paper, a new concept of integration of fractal and the butterfly effect is proposed and implemented. A new fractal program was designed and developed to perform such integration. Among many existing fractal and chaos software programs, none of them allow us to achieve the resulting patterns demonstrated in this paper. Moreover, it is the first time that a fractal program provides functional concepts of overlapping results in 3D space and sequential transformations, which allow us to generate a wider variety of patterns. Therefore, potentially an artist can use this program to create 2D digital artworks.

Keywords: Butterfly effect, Lorenz attractor, fractal and chaos, mathematical art.

1 Introduction

In recent years there has been an increasing interest in the fields of digital art and mathematical art while computing power increases rapidly. Computers offer a wide range of creative possibilities for digital artists to explore new ideas and generate or deliver their artworks. Combining fractal and chaos theory is a well-known approach which enables to create amazing artistic patterns [3,7]. In this paper, we elaborate particularly on the potential of *The Butterfly Effect* in chaos theory for creating artistic patterns and animations.

There are software programs that support various fractal formulas and a wide variety of color filters for creating fractal art images [4]. Eye-catching fractal images are easily accessible through web-based fractal art galleries. Fractal techniques have become a powerful tool not only for digital art design but also for architectural design [1,6]. Moreover, fractal formulas have been applied to the creation of simulated natural landscapes and plants in computer graphics [8].

Among all the studies of fractal and chaos theories, to the best of our knowledge, no one has reported so far a mixed model of *The Butterfly Effect* fractal. The integration of The Butterfly Effect formulas into fractal art is a newly proposed concept, which provides a novel variety of possibilities in terms of forms and shapes in fractal art creation. Figure 1 illustrates the potential of our developed program, in which a typical fractal and a typical butterfly effect pattern are shown on the left and middle, respectively, as well as a resulting artistic pattern generated by our program on the right.

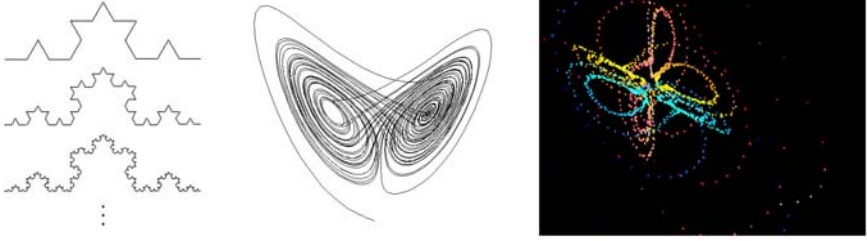


Fig. 1. Left: a typical fractal structure. Middle: a typical Lorenz attractor. Right: a 2D artwork generated by our program.

2 The Butterfly Effect

The Butterfly Effect is not only a fascinating mathematical theory, it also happens to be the name of a well-known movie series. Perhaps more people know about the stories of those movies than about the history and essence of this theory. We provide some useful mathematical background and a brief history of The Butterfly Effect (partially recalled from [2]) in this section.

2.1 Lorenz Equations and Lorenz Attractor

Edward Lorenz was a Mathematician and Meteorologist at the Massachusetts Institute of Technology who was interested in weather prediction. He constructed a mathematical model of the weather, defined by a set of twelve differential equations that represented changes in temperature, pressure, wind velocity, and so forth. On a particular winter day in 1961, Lorenz wanted to re-examine a sequence of data coming from this weather model. He discovered that his model exhibits the phenomenon known as "sensitive dependence on initial conditions". This is sometimes referred to as the *Butterfly Effect* (i.e., a butterfly flapping its wings in South America may affect the weather in Central Park, London). The data obtained by two different runs diverged dramatically due to rounding-off errors; results differed in more than three decimal places. This led to the question, why does a set of completely deterministic equations exhibit this behavior? It is due to the nature of the equations themselves which were nonlinear equations. Nonlinear systems are central to chaos theory and often exhibit fantastically complex and chaotic behavior.

Lorenz first reported his weather model and the discovery in *Deterministic Nonperiodic Flow* [5], which was a journal paper published in 1963 and had great influence on many subsequent related studies. His work is important even today because he had an insight into the essence of chaos, and his work settled in today's chaos theory. Later, Lorenz decided to look for complex behavior in an even simpler set of equations, and was led to the phenomenon of rolling fluid convection. He attempted to simplify a few fluid dynamic equations (called the

Navier-Stokes equations) and ended up with a set of three nonlinear equations known as the *Lorenz equations*:

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= x(\rho - z) - y \\ \frac{dz}{dt} &= xy - \beta z\end{aligned}$$

where σ is the *Prandtl number* representing the ratio of the fluid viscosity to its thermal conductivity, ρ is the *Rayleigh number* representing the difference in temperature between top and bottom of the system, and β is the ratio of width to height of the box used to hold the system. All $\sigma, \rho, \beta > 0$, but usually $\sigma = 10$, $\beta = 8/3$ and ρ is varied. The system exhibits chaotic behavior for $\rho = 28$ but displays knotted periodic orbits for other values of ρ .

Although the equations look quite simple, they are nonlinear depending on the products of the variables xz and xy , and an analytic solution is impossible in general. We employed the iterative method that allows solving the system of nonlinear equations numerically [9]. To compute a numerical approximation for the solution, we employ the following iterative equations:

$$\begin{aligned}x_{n+1} &= x_n + 10(y_n - x_n)h \\ y_{n+1} &= y_n + (-x_n z_n + 28x_n - y_n)h \\ z_{n+1} &= z_n + (x_n y_n - \frac{8}{3}z_n)h\end{aligned}\tag{1}$$

We obtained the 3D plot as shown in Figure 1(middle), which is known as Lorenz attractor, generated by an iterative process as in Equations (1) with initial values $x_0 = 0.0001$, $y_0 = 0.0001$, and $z_0 = 0.000001$. This illustrates the typical butterfly-like pattern when projecting data points onto the xz -plane.

2.2 Integration of Fractal and Lorenz Equations

We launched the concept of fractals for Equations (1). The Lorenz attractor becomes a basic element of a fractal structure. As a result of a two-leveled fractal structure, the initial base Lorenz attractor will grow into smaller version of Lorenz attractors along its path as shown in Figure 2. The equations used to generate such fractal Butterfly Effect pattern are as follows:

$$\begin{aligned}A_{n+1} &= A_n + 10(B_n - A_n) \\ B_{n+1} &= B_n + (-A_n C_n + 28A_n - B_n) \\ C_{n+1} &= C_n + (A_n B_n - \frac{8}{3}C_n) \\ a_{m+1} &= a_m + 10(b_m - a_m) \\ b_{m+1} &= b_m + (-a_m c_m + 28a_m - b_m)\end{aligned}\tag{2}$$

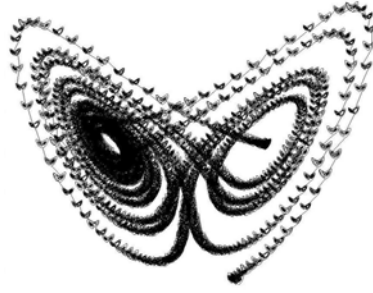


Fig. 2. The fractal *Butterfly Effect*

$$\begin{aligned}
 c_{m+1} &= c_m + (a_m b_m - \frac{8}{3} c_m) \\
 x_{nm} &= 20A_n + a_m \\
 y_{nm} &= 20B_n + b_m \\
 z_{nm} &= 20C_n + c_m
 \end{aligned}$$

We show only the equations for a two-leveled fractal structure here. Equations for n -leveled fractal structures, with $n > 3$, would involve huge amounts of computations. Since we are interested to achieve real-time calculation, and accelerating computation is not our major objective in this work, our program offers at most three-leveled fractal.

3 Program Design

In this section, we explain the design issues and functions of control buttons provided in our fractal program. Target users of this program are people who do not have mathematical background on chaos theory but are interested in creating fractal art. Thus, there is no formula appearing in the user interface. An artistic pattern can be created purely by pressing buttons, selecting data-points, shape or color, and changing the view point by simple mouse drag.

Functionality provided by our program can be classified into five categories, namely initialization, transformation, viewing, point style, and coloring filters. We explain the purpose of each of these categories in the following:

Initialization. There are three initial values defining the x , y , and z coordinates of the Butterfly Effect's starting position in 3D space. A different starting position leads to different butterfly pattern, which is the core meaning of "Butterfly Effect". Moreover, there are two initial values indicating the complexity of the butterfly pattern, namely *Big loop* and *Small loop*. The Big (Small) loop corresponds to the first (second) level of a fractal structure. The larger the value the more complex the butterfly pattern and thus more computation time.

Transformation. There are three menus provided in this category. The first menu is called *Old Shape*, in which a user may decide whether to keep the current

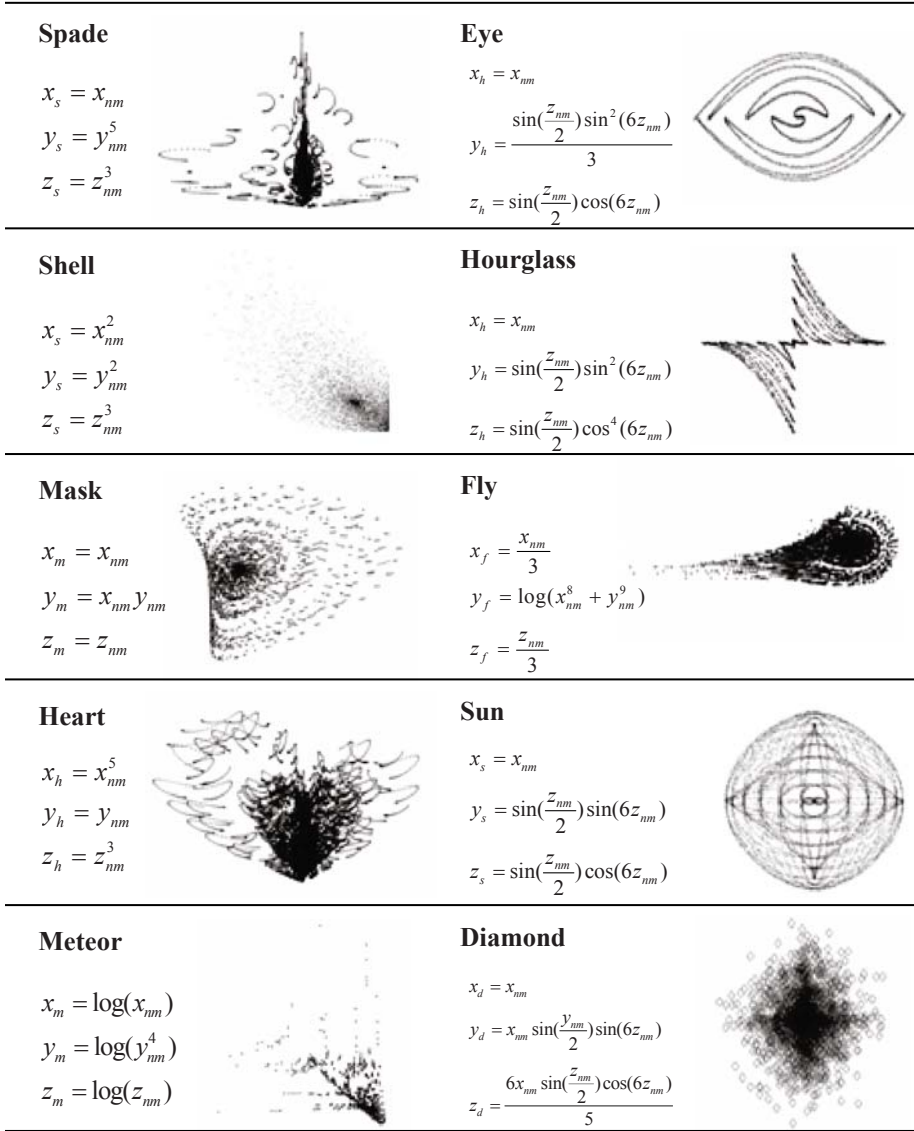


Fig. 3. Predefined transformations and their equations

data or to transform them. If we choose *Keep*, then the current pattern(s) will be kept in the displaying window. In this case, the next generated pattern will overlap the previous one(s) in the window. Figure 5 (middle) is an example of overlapping butterfly patterns generated using different colors and by different looping values. If we choose *Release*, then the last generated pattern will be transformed according to the next transformation command. The performed

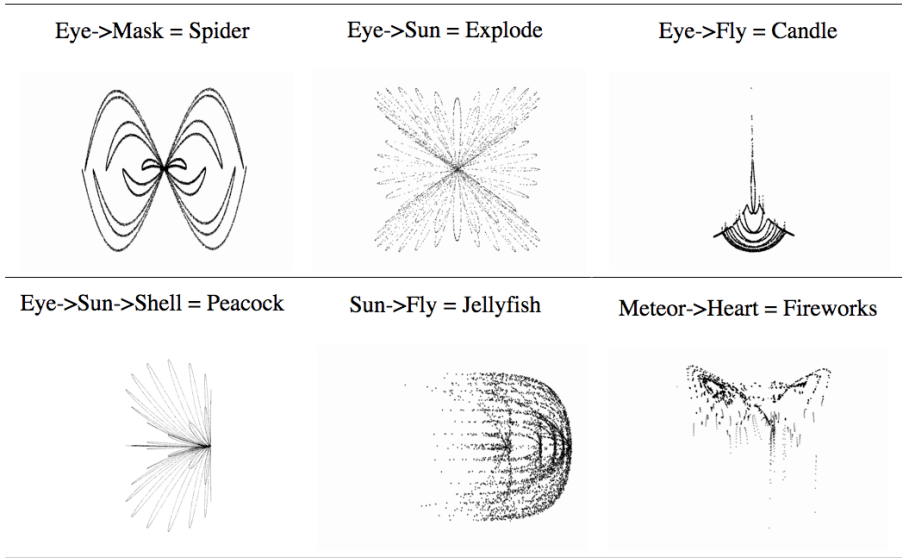


Fig. 4. Patterns generated by different combinations of predefined transformations

transformations are displayed in a text region. The second menu is called *Evolve*, in which there are many predefined transformations. The key to explore a new pattern is to try different combinations of transformations and in different order. The third menu is called *Mirror*, which performs mirror transformation to the current data.

Viewing. This changes user's viewing position and direction. The buttons and menus provided here are quite trivial for computer graphics users. For instance, you can choose the predefined view by menu selection, rotate the data by mouse drag, or continuously rotate by pressing a button. The program is able to perform real-time rendering which changing viewing position and angle, and hence the sequence of rendered images can be recorded and converted to an animation format. This function enables users to design their own "viewing paths" within the 3D data, and offers the potential of creating artistic (abstract) animations.

Point Style. This changes the style (shape, size, color) of the data point.

Coloring Filter. Sometimes only changing the point style is not sufficient to create impressive results. Here, we provide some coloring methods in the *Color Map* menu to create different coloring effects. This category is called *Sphere Controller* because those coloring filters are only applicable to facets, and thus it is necessary to represent the data by some 3D geometric shapes before applying coloring filters. In this first version of the program, we only implemented spheres and six coloring methods. Other coloring filters (or methods) can be easily included in the program.

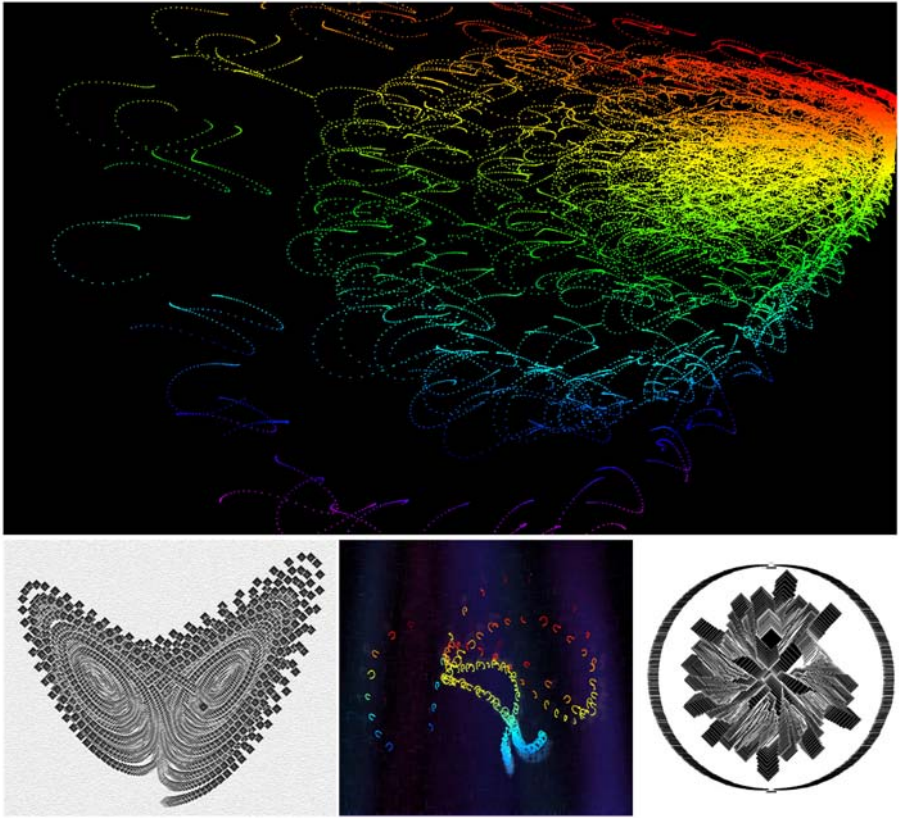


Fig. 5. Generated artworks utilizing the fractal butterfly effect

3.1 Results

Typical patterns and corresponding equations of ten predefined transformations are illustrated in Figure 3. Those are meant to show the structure of the pattern, and are thus displayed in black only. Applying different transformations in different sequential order will result in very different patterns. Some example combinations of transformations are illustrated in Figure 4. Figure 5 illustrates variety of 2D digital artworks generated by our program. In particular, the structures of the images on the bottom row of Figure 5 are generated by our program, with subsequent minor graphical fine-tuning in Adobe's Photoshop.

4 Conclusions

A new fractal art program was developed. The friendly user interface enables users who have no mathematical background to create artistic fractal patterns.

The major contribution is the proposal and implementation of a fractal Butterfly Effect concept; we believe that this is the first program that allows us to generate such an integration. Other contributions include the specification of equations of the predefined transformations and concepts for applying overlapping results and sequential transformations in the fractal program. There are various existing programs for generating various fractals or Lorenz attractors, but none of them provides overlapping and sequential transformation functions.

References

1. Bovill, C.: *Fractal Geometry in Architecture and Design*. Birkhäuser, Boston (1996)
2. Bradley, L.: *Chaos & Fractals* (2009),
<http://www.pha.jhu.edu/~ldb/seminar/index.html>
3. Gleick, J.: *Chaos: Making a New Science*. Penguin Books (1987)
4. Lee, P.N.: *Fractal Links* (2009),
http://home.att.net/~Paul.N.Lee/Fractal_Software.html
5. Lorenz, E.N.: Deterministic Nonperiodic Flow. *J. Atmospheric Science* 20, 130–141 (1963)
6. Ostwald, M.J.: Fractal Architecture: Late Twentieth Century Connections Between Architecture and Fractal Geometry. *Nexus Network J.* 3(1), 73–84 (2001)
7. Parker, B.: *Chaos in the Cosmos: The Stunning Complexity of the Universe*. Plenum Press, New York (1996)
8. Peitgen, H., Saupe, D. (eds.): *The Science of Fractal Images*. Springer, New York (1988)
9. Tucker, W.A.: Rigorous ODE Solver and Smale's 14th Problem. *Found. Comp. Math.* 2, 53–117 (2002)

Pattern Formation in Networks Inspired by Biological Development

Kei Ohnishi, Kaori Yoshida, and Mario Köppen

Kyushu Institute of Technology,
680-4 Kawazu, Iizuka, Fukuoka 820-8054 Japan
ohnishi@cse.kyutech.ac.jp, kaori@ai.kyutech.ac.jp,
mario@ndrc.kyutech.ac.jp

Abstract. We present a system for pattern formation in networks that is inspired by biological development. The system hierarchically forms a specific pattern in a rough-to-detail manner as seen in biological development. We apply the system to generating gray-scaled images. In this application, a network topology is first generated from a given original image in which pixels correspond to network nodes and the nodes are linked to one another in a certain way, and then the system is run in the generated network. The pattern formed by the system within the network is a distribution of brightness of pixels which are equal to nodes. The system is shown to create a variety of images which are influenced by the structure in the original image in a rough-to-detail manner. This feature would be useful for designing patterns under some structural constraints.

Keywords: Biological development, pattern formation, feedback system, image generation, network topology.

1 Introduction

Recently, many proposed engineering methods have been inspired by biological behavior and structures, such as evolutionary computation [1], neural networks, artificial immune systems [2], ant colony systems [3], and so on. The reason is that we can expect to obtain useful hints on creating new methods from the study of biological behavior and structures.

Some of these methods are equivalent to the approaches used in artificial life (A-Life). These approaches create macroscopic structures or functions by only using local interaction between their elements. In these approaches, the macroscopic structures or functions are often quantitatively undefined but qualitatively defined. Consequently, we need to adjust the local interaction rules for our purpose. However, the A-Life approaches have several advantages. One advantage is that these approaches have the possibility to realize structures or functions that are barely realized by a centralized control system.

Simple systems using the A-Life approaches are feedback systems such as cellular automata [7] and L-systems [4][6]. These systems can create complex structures from a simple initial state. Complex structures are created by repeatedly applying the rules of rewriting symbols corresponding to their structures. It

is simple but difficult task for us to estimate the final structure from the initial state and rules because the rules are applied locally. Their effective use may be to discover or to create interesting structures by adjusting the rules and the initial state by trial and error. Interesting structures that we could have never imagined may be discovered by chance.

Following this idea, we previously proposed a new feedback system inspired by biological development, which we referred to as the “BiDevelopment System” [5]. The BiDevelopment System is supposed to form patterns in Euclidean spaces. In the present paper, we present a new algorithm by extending an example algorithm of the BiDevelopment System. The algorithm is able to form patterns in networks. In addition, we apply the algorithm to generating gray-scaled images.

The advantage of the BiDevelopment System is the embedded rough-design structure and various possible generated structures under the given rough-design structure. Conventional feedback systems, such as the L-System or cellular automaton, apply only one rule to each element at each feedback loop. It is difficult to design an entire final structure from the beginning and embed it into the system. Since each module of the proposed system determines final detail from the basic structure, it is easier to embed the rough-design structure before the system runs.

The present paper is organized as follows. Section 2 describes the BiDevelopment System. In Section 3, we present the algorithm of the BiDevelopment system for pattern formation in networks and also apply the algorithm to generating gray-scale images. Section 4 describes our conclusion and future work.

2 Pattern Formation in Euclidean Spaces

The new algorithm, which will be presented in Section 3, is realized based on an example algorithm of the recently proposed BiDevelopment System. So, we recall the BiDevelopment System in this section. The hints for the BiDevelopment System, which is a mechanism of biological development, are described in [5]. A biological development process forms an adult body from a mother cell based only on design information of the mother cell in a rough-to-detail manner. In this process, proteins hierarchically diffuse among all or some cells and gives them positional information, telling the cells what organs they will eventually become.

2.1 The Conceptual Framework

The BiDevelopment System outputs values on a space, which is its input. Values of the system parameters are provided in a character code. It has an initialization module and three major modules. The module numbers like (0), (1) etc below correspond to numbers in Figure 1. The initialization module (0) defines the area where system outputs exist. The module (1) generates the global positional information that determines what each element processes. The module (2) lets each element autonomously behave based on its positional information. Local mutual interaction among elements may take place. The module (3) exchanges

the output of the module (2) into the output range of elements. The final output value of each element is determined by the feedback module (4b).

The module that generates data related with structure is the module (3). The values that the feedback process at the module (3) generates after certain number of feedback loops determine the final structure of the system output. The modules (1) and (2) sequentially generate the structure of the final output pattern from the BiDevelopment System.

The modules (1) and (3) include feedback modules (4a) and (4b), respectively. The feedback module (4a) generates the positional information of the next state from that of the previous state. The feedback module (4b) gradually narrows the range of output values of each feedback process.

2.2 Example Algorithm

An example algorithm under the framework of the BiDevelopment generates values of y for given coordinates of x . Figure 1 shows the flow of modules in the algorithm.

The task here is to draw a figure in a (x, y) space by generating values of y for coordinates of x in a fixed closed area, $x \in [x_1, x_n]$ ($x_1 < x_n$). Then, the modules of the algorithms are as follows. The modules are (0), (1), (2), (3), (4a), and (4b) in Figure 1. A character code presented in Figure 1 encodes values of the algorithm parameters.

(0) initialization

The module (0) in Figure 1 inputs the range of x space where y values are generated to draw a figure. The initial data inputted here are (x_1, x_n) .

(1) generating positional information

The module (1) in Figure 1 gives positional information of each x coordinate using diffusion functions such as Equation (1) (see Figure 2(a) (1).)

$$M(x, A_k, \mu_k, \sigma_k) = M_k(x) = A \exp \frac{-(x-\mu_k)^2}{2\sigma_k^2} \quad (1)$$

Let us call the $M_k(x)$ in Equation (1) a diffusion function from the analogy to diffusing protein among cells in biological development. The iteration number of feedback is represented by k . Any function is available for the diffusing functions, and we adopt a Gaussian function for M_k in this paper. The positional information is determined by the $M(x, A_k, \mu_k, \sigma_k)$ whose parameters, A_k, μ_k , and σ_k are read from the character code. Although the algorithm presented here uses only one diffusion function as in Equation (1), other algorithms may use multiple diffusion functions. That is, $\sum_i M(x, A_{ki}, \mu_{ki}, \sigma_{ki})$ is used.

(2) gene expression

The module (2) in Figure 1 divides x into several areas according to the s_i read from the character code and values of $M_k(x)$ (see Figure 2(a) (2)). Labels are

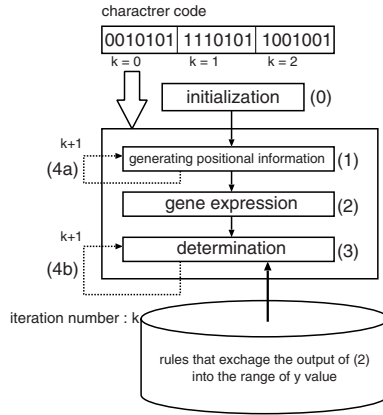


Fig. 1. Process flow of the algorithm

given to the divided areas of x . We compared the given labels to the genetic information and named this module as a *gene expression*.

(3) determination

The module (3) in Figure 1 converts the group labels outputted from the previous module (2) together with previous y range into the range of values of y (see Figure 2(a) (3)). The conversion rules that are different in each processing time, k , must be prepared.

(4) feedback

These modules (4a) and (4b) in Figure 1 are feedback modules at the modules (1) and (3), respectively.

(4a) feedback for the module (1)

The module (4a) calculates μ_k for diffusion function $M_k(x)$ from $M_{k-1}(x)$. It first reads m from the character code, substitutes the m for Equation (2), and determines μ_k , where μ_0 is given in the character code.

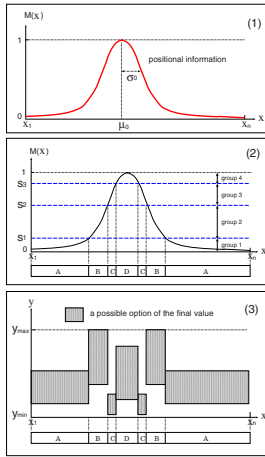
$$\mu_k = x, \text{ if } M_{k-1}(x) < m < M_{k-1}(x + 1), \text{ or} \tag{2}$$

$$\text{if } M_{k-1}(x + 1) < m < M_{k-1}(x).$$

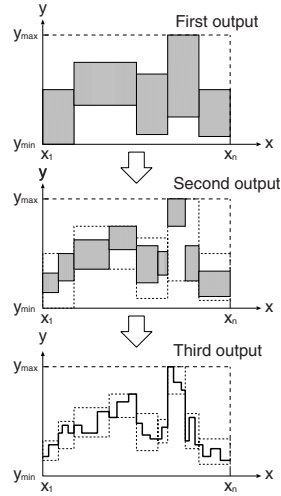
When multiple μ_{ki} are obtained for the given m , $\sum_i M(x, A_k, \mu_{ki}, \sigma_k)$ is used in the module (1). Other necessary parameters for Equation (1), A_k and σ_k , are read from the character code.

(4b) feedback for the module (3)

According to the feedback module (4b), the range of y is gradually reduced (see Figure 2(b)). The rules that determine the new ranges of y from the past ranges y and group labels must be previously prepared in a rule-base at the module (3).



(a) Example outputs of the first processing, $k = 0$



(b) Example outputs of the module (3)

Fig. 2. Example outputs

2.3 Generating Gray-Scaled Images

In this section, we demonstrate how the BiDevelopment System creates gray-scaled images. The algorithm used here is similar to the algorithm in Section 2.2 except for a 2-D target space, (x_1, x_2) , and the initial diffusion function combining 50 Gaussian functions. The algorithm used here obtains a final output just after the third feedback. Images are displayed by converting the generated y values to the brightness of pixels. An example of generated images is shown in Figure 3. In Figure 3, outputs at each feedback are shown. Although outputs



(a) The output at $k = 1$ (b) The output at $k = 2$ (c) The final output at $k = 3$

Fig. 3. Example gray-scaled images generated by the BiDevelopment System

except the final output represent ranges of z values on the x - y plane, we create images from those outputs by using the average of the ranges of z values.

3 Pattern Formation in Networks

In the image generation shown in Section 2.3, we first considered correspondence between discrete coordinates in a two dimensional Euclidean space (a x - y plane) and pixels in a gray-scaled image, and then, we generated gray-scaled images by using z values that the algorithm generated in the x - y plane as brightness of the pixels in the images. Therefore, the generated images basically, as shown in Figure 3, include shape of cross section between multiple Gaussian functions and a plane parallel to the x - y plane.

One of the ways to create more various images is to change a field or a space in which the algorithm forms patterns. So, in this section, we extend the example algorithm of the BiDevelopment System presented in Section 2.3 to the one which is able to form patterns in networks. To form patterns in a network with the extended algorithm, the algorithm has to prepare a diffusion function for a network. A diffusion function used herein is represented as Equation (3). The diffusion function is generated at some node.

$$M(h_s, A_k, \mu_k, \sigma_k) = M_k(h_s) = A \exp^{-\frac{(h_s - \mu_k)^2}{2\sigma_k^2}}, \quad (3)$$

where h_s is the minimal number of hops from a node of focus to the node, at which the diffusion function is generated. The diffusion function is basically the same as Equation (1) shown in Section 2.2, but x in Equation 1 is changed to h_s in Equation (3).

In addition, we apply the extended algorithm to generating gray-scaled images. In this application, a network is first generated from a given original image, in which pixels correspond to network nodes and the nodes are linked to one another in a certain way (see Figure 4), and then the algorithm is run in the generated network. The formed pattern in the network is a distribution of brightness of pixels which are equal to nodes.

How to generate a network from a given image is as follows. Each node, which has a corresponding pixel, makes L links to other nodes. Nodes to which each node makes links are selected from among all of the nodes with probability inversely proportional to difference between brightness of each node and other node. That is, nodes with similar brightness are likely to link to one another. The brightness of the pixels of the give image are used only to generate a network, which becomes a field, to which the algorithm is applied. In the image generation here, we set L to be 6.

The algorithm here uses the initial diffusion function combining 50 Gaussian functions, and obtains a final output just after the third feedback. Figure 5 shows three examples of the generated images and the original image used for generating the network. In Figure 5, outputs at each feedback are shown. Although outputs except the final output represent ranges of values (brightness) on the

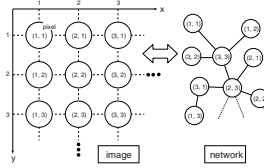
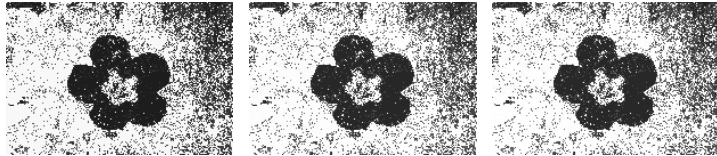
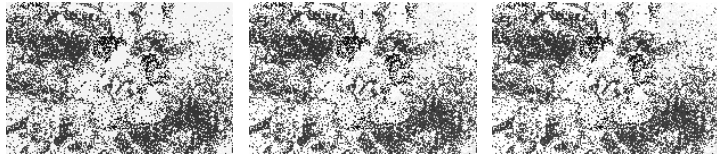


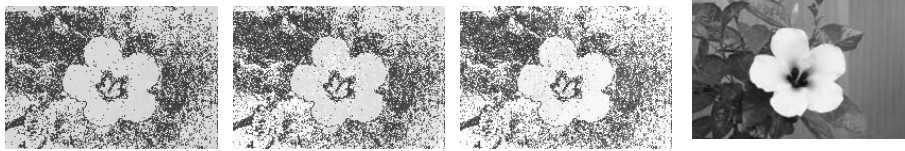
Fig. 4. Correspondence between nodes in a network and pixels in a gray-scaled image



(a) First example. (b) First example. (c) First example. The output at $k = 1$. The output at $k = 2$. The final output at $k = 3$.



(d) Second example. (e) Second example. (f) Second example. The output at $k = 1$. The output at $k = 2$. The final output at $k = 3$.



(g) Third example. (h) Third example. (i) Third example. (j) Original image. The output at $k = 1$. The output at $k = 2$. The final output at $k = 3$.

Fig. 5. Three examples of gray-scaled images generated by the extended algorithm with three different character codes

network, we create images from those outputs by using the average of the ranges of values.

Figure 5 suggests that the algorithm presented in this section can create various images. In the application here, we generated the network using the distribution of brightness of pixels in the original image, so that the images generated by the algorithm seem to be influenced by the distribution of brightness.

4 Concluding Remarks

In this paper we realized the new algorithm under the framework of the recently proposed BiDevelopment System. The realized algorithm is meant to form patterns in networks. In addition, the algorithm was applied to generating gray-scaled images. In this application, a network was first generated based on the given image and then the algorithm was run in the network. The results of the image generation suggested that the algorithm can create various images which are influenced by the distribution of brightness of an original image, which is used for generating a network.

As shown in this paper, it would be possible to represent fundamental desired structures or structural constraints as a form of network topologies. So, if we combine the use of network topologies, the strategy of rough-to-detail pattern formation of the BiDevelopment System, and an optimization technique which adjusts the parameter of the BiDevelopment System based on human evaluation, the BiDevelopment System might be able to be an emergent design tool that is handleable for humans. In the future work, we will tackle with this problem.

Acknowledgment

This work was supported by the Japan Society for the Promotion of Science through a Grant-in-Aid for Scientific Research (S)(18100001).

References

1. Bäck, T.: Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford University Press, Oxford (1996)
2. Dasgupta, D. (ed.): Artificial immune systems and their applications. Springer, Heidelberg (1999)
3. Dorigo, M., Maniezzo, V., Coloni, A.: The ant system: optimization by a colony of cooperating agents. *IEEE Trans. on Systems, Man, and Cybernetics-Part B* 26(2), 29–41 (1996)
4. Lindenmayer, A.: Mathematical models for cellular interaction in development, PartI and PartII. *Journal of Theoretical Biology* 18, 280–315 (1968)
5. Ohnishi, K., Takagi, H.: Feed-Back Model Inspired by Biological Development to Hierarchically Design Complex Structure. In: *IEEE International Conference on System, Man, and Cybernetics (SMC 2000)*, Nashville, TN, USA, October 8-11, pp. 3699–3704 (2000)
6. Prusinkiewicz, P., Lindenmayer, A.: The algorithmic beauty of plants. Springer, New York (1990)
7. Wolfram, S.: Cellular automata and complexity: collected papers. Addison-Wesley, Reading (1994)

A Watercolor NPR System with Web-Mining 3D Color Charts

Lieu-Hen Chen, Yi-Hsin Ho, Ting-Yu Liu, and Wen-Chieh Hsieh

Department of Computer Science & Information Engineering, National Chi Nan University
No.1, University Rd, Puli, Nantou County, Taiwan (R.O.C.)
lhchen@csie.ncnu.edu.tw,
{s96321523, s96321543, s95321025}@ncnu.edu.tw

Abstract. In this paper, we propose a watercolor image synthesizing system which integrates the user-personalized color charts based on web-mining technologies with the 3D Watercolor NPR system. Through our system, users can personalize their own color palette by using keywords such as the name of the artist or by choosing color sets on an emotional map. The related images are searched from web by adopting web mining technology, and the appropriate colors are extracted to construct the color chart by analyzing these images. Then, the color chart is rendered in a 3D visualization system which allows users to view and manage the distribution of colors interactively. Then, users can use these colors on our watercolor NPR system with a sketch-based GUI which allows users to manipulate watercolor attributes of object intuitively and directly.

Keywords: Computer Graphic (CG), Non-Photorealistic Rendering (NPR), Watercolor, Color Charts, Web Mining, Sketch-based Graphic User Interface.

1 Introduction

Choosing colors is an important issue not only for professional artists and designers, but also for amateur and beginner groups. By carefully arranging the tints and shades of colors, the artist allows viewers to feel what the artist is feeling with strong impression.

Color chart is the basis of color designing. Using color charts can avoid the visual error which is caused by misunderstanding and intermingling colors. Furthermore, the appropriate colors can be determined by their color value and color hue, and picked up from the color charts. Unfortunately, the conventional color charts are designed and represented in 2-dimensional way. It is very difficult for users to survey the whole color model and choose colors in an intuitive way [1].

In this paper, we propose a watercolor image synthesizing system which integrates the user-personalized color charts based on web-mining technologies with the 3D Watercolor NPR system.

Non-Photorealistic Rendering (NPR) has been an important research topic in the field of Computer Graphic. Instead of highlighting the realism of CG synthesized images, NPR focus on the representation of the stroke and painter's style. For the research of NPR watercolor, simulating water flow and dyes diffusion on paper is an

important task. In our system, the Lattice Boltzmann Method (LBM) is adopted to simulate these watercolor effects. In addition, the Line Integral Convolution (LIC) method is adopted to automatically synthesize the computer-generated images using the fluid dynamics. Finally, we develop a sketch-based GUI for simulating hand drawing.

In the next section, a brief description of 3D color charts (user-customized color charts) and the structure of color system is explained. Our NPR watercolor system is discussed in section 3. The experiment results of our system are shown in section 4. Finally, we conclude this paper in section 5.

2 3D Color Charts

In this section, we introduce the basis of color combination and the way how we customize the color charts.

2.1 The Basis of Color Combination

We selected HSL system as the color system for this project, because it describes perceptual color relationships more accurately than RGB and reflects the intuitive notion of "saturation" and "lightness" better. [2][3]

HSL describes colors as points in a cylinder whose central axis ranges from black at the bottom to white at the top with neutral colors between them, where angle around the axis corresponds to "hue", distance from the axis corresponds to "saturation", and distance along the axis corresponds to "lightness" (Fig. 1 HSL Color system).

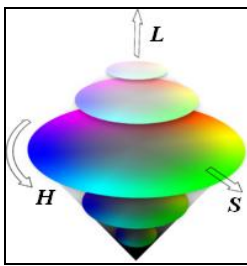


Fig. 1. HSL color system

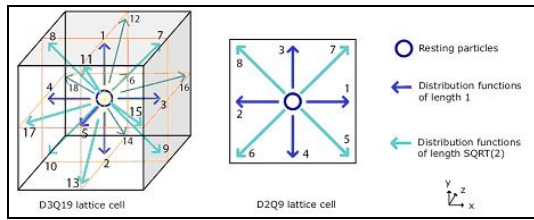


Fig. 2. LBM model, D3Q18 and D2Q9

Hue is one of the main properties of a color described with names such as "red", "yellow".

Value, or lightness, varies vertically along the color solid, from black (value 0) at the bottom, to white (value 10) at the top. Neutral grays lie along the vertical axis between black and white.

Chroma, or saturation, measured radially from the center of each slice, represents the "purity" of a color, with lower chroma being less pure (more washed out, as in pastels). Note that there is no intrinsic upper limit to chroma.

2.2 Customizing the Color Charts

Color psychology is a science of chromatology and psychology [4] [5] [6] [7]. Different colors give people different feeling by the relation between color and mood. It is practical to provide an intuitive mapping function between colors and user feeling. Therefore, we adopted the web-mining technique and 160 intuition tables to assay the possible meaning of colors to users.

In our system, we construct the user-customized color charts by performing the following procedures as shown in the lower part of figure 3. First, users can define their current psychological situation directly by using a 3D emotion map, or indirectly by choosing keywords such as the name of artist or styles of painting from tables. Then, the IPhoto -Web Pictures Searcher of Keronsoft is adopted to search related images from web automatically. The appropriate colors are extracted to construct the color chart by analyzing the histograms of these images. Finally, the color chart is rendered in a 3D visualization system which allows users to view and manage the distribution of colors interactively.

3 System Architecture of the 3D Watercolor NPR System with a Sketch-Based GUI

In this section, we briefly introduce our system architecture and the basic technique of watercolor NPR images synthesis. The flowchart of the whole system is shown in the figure 3.

3.1 Compose 3D Scene and Synthesize Strokes

Our system is designed to use the X file format of 3D Studio Max for inputting data of 3D models. Because the advantage of object class of X file, the system utilizes the object class to divide and rebuild the data structure. Then, the modified z-buffer is used as the virtual canvas to store and access the information of 3D objects in the scene. Finally, the web-mining 3D chart is adopted to synthesize watercolor painting image.

Module for NPR images synthesis is summarized as following [8]:

(A) 3D model loader (Pre Processing part):

Read in the X file and take apart the object class, rebuilt the data structure, and construct the mesh data.

(B) Modified Z-Buffer (Rasterizing & Depth test – Per Frame part):

1. Initialize the modified Z-buffer.
2. Raster the 3D-model float data, and fill into Z-buffer.
3. Differentiate the line segment of object and decide the line those have to be rendered.
4. Test the depth buffer to remove the hidden lines.

(C) Stroke Calculator and Stroke Drawer:

1. Choose the styles of stroke and calculate the trends of stroke.
2. Calculate the weight of object frame.

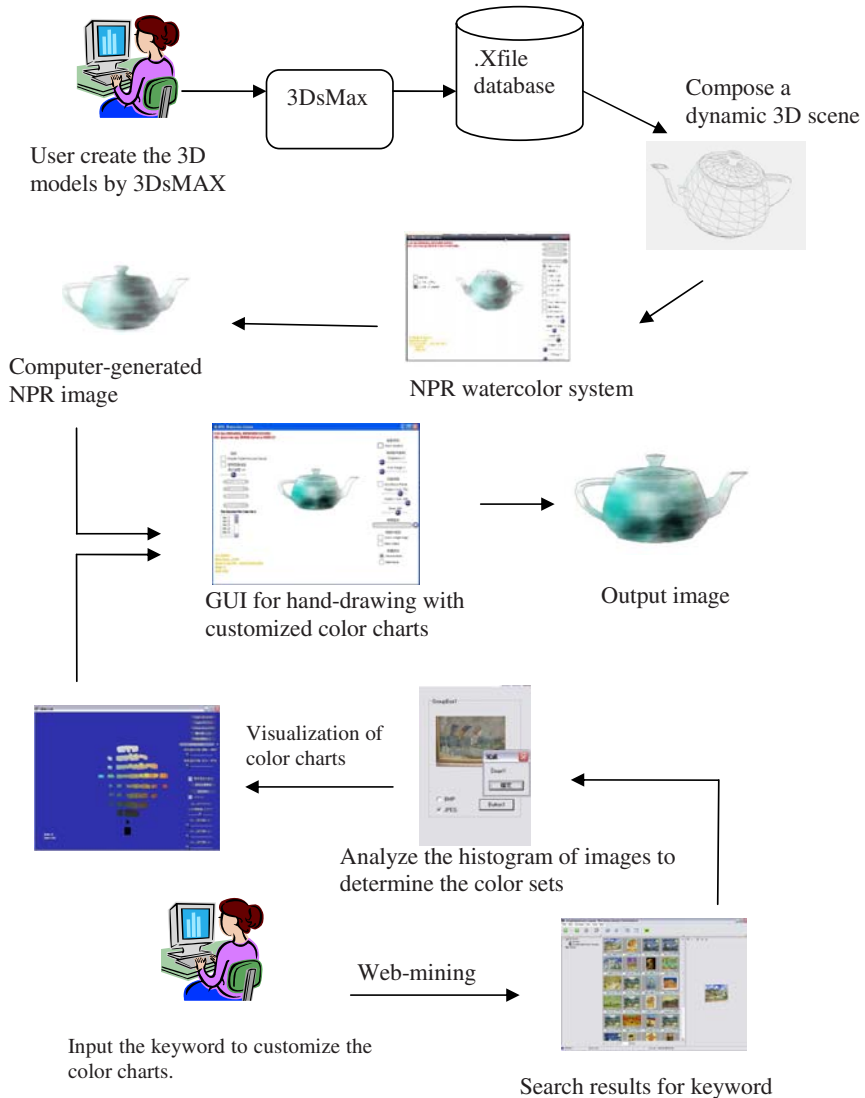


Fig. 3. System architecture

3. Choose the recommendation color from Web-minin3D color charts.
4. Draw the painting automatically by system or manually by Graphic table.

We assumed that the stroke is constructing by a series of points what link alone a track. The different styles of stroke can be easily presented. In our system, we use Line Integral Convolution (LIC) to produce strokes. A simple LIC is combined by a color/noise image and a vector field. LIC is based on vector field to trace every pixel on image. Pixels that can make a fixed length are recorded as a group. Color mixing and filling are performed for each group to compose an image.

3.2 Watercolor Simulation

There are many watercolor painting techniques. We simulated the "wet in wet" technique in our system. This technique creates blurring boundaries. Because watercolor is constituted by pigment and water, it can be considered as fluid. We use lattice Boltzmann Method [9] to simulate the watercolor effects. Adopting LBM along the paths of strokes creates computer-generated NPR images automatically. Finally, our system support users to interactively hand-draw images by using our sketch-based GUI.

3.3 Lattice Boltzmann Method (LBM)

LBM models, an incompressible fluid by particles, are allowed to move only along the lattice velocity vectors. It also can be described as a kind of cellular automaton. All cells are based on the state of the surrounding cells, and all cells are updated each time by simple rules.

D3Q18 and D2Q9 are the two standard lattice models. (Fig.2 LBM Model) In our system, we use D2Q9, here 2 represents the number of dimension and 9 represents the number of different velocity vectors.

At every time, LBM repeats two steps: stream step and collide step. First, the stream step calculates the particle movement on lattice. Then the collide step uses the distribution from the stream step to calculate velocity. The velocity is used to calculate the local equilibrium distribution function. Finally, it gets the new distribution function from adding the two functions by weight. This weighting value is decided by the density of fluid.

3.4 Automatic and Hand-Drawing Phases

The image synthesizing stage can be divided into two phases, including the automatic rendering phase and the hand drawing phase.

In the automatic rendering phase, our system allows users choosing the color palette from 3D color charts and the stroke type first. Then, the system combines the user-customized color palette with the stroke colors. Finally, the system calculates the stroke positions and synthesizes the watercolor painting effects along the paths of strokes automatically. [10]

In the hand drawing phase, a sketch-based GUI with graphics table is developed to provide artists and designers a more intuitive way of input. Users can use the picked-up color from their customized color palette to add different color strokes on the virtual canvas. The process of watercolor mixing and filling simulates the real process of hand-drawing on paper. Using this sketch-based GUI, users can modify and add visual effects to object and scene interactively.

4 Experimental Results

Our experiment results are shown in the figures 4 to 9. In addition, a demonstration video clip is available at <http://www.youtube.com/watch?v=snbLKL676wI>.

Figure 4 to 6 illustrate our visualization system of 3D color charts. The yellow text in the upper-left corner of figure 5 which introduces the information about the author is obtained through web-mining techniques. The pink text in the upper-right corner of

figure 6 which describes the emotional meaning of current color set is referred from a color-emotional database containing 160 intuition tables.

Figure 7 illustrates the wire-frames of the 3D scene. Figure 9 illustrates the computer-generated watercolor NPR image. Figure 10 illustrates the final image result which is edited and enhanced by user through the hand-drawing GUI.

We also designed a questionnaire with four questions to verify our proposal. This test is performed to 40 visitors after they have used our system (for more than 10 minutes, at least). The four questions are list below.

Q1. In comparison with conventional 2D color charts, the 3D visualized color charts provide an easier way to select and survey colors.

Q2. It is helpful for users to understand the coloring preference of an artist by using our system.

Q3. It is convenient for users to arrange colors to reflect their mood by using the 160 sense color forms.

Q4. Compared with figure 7, users can feel the emotional changes, i.e. sadness, peace, and joy, respectively from the images shown in figure 8.

According to the statistics of this survey, we give the graphs of comparison between our 3D visualized color charts with the conventional 2D charts in table 1.

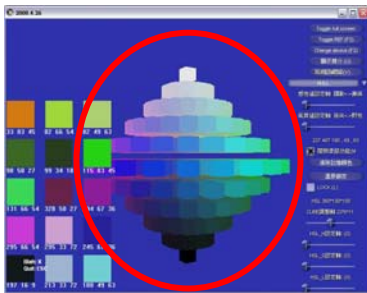


Fig. 4. 3D Color Charts and color information



Fig. 5. Artists' colour combination

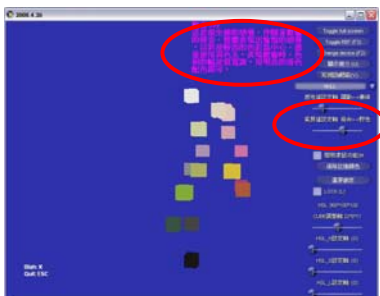


Fig. 6. Emotional map

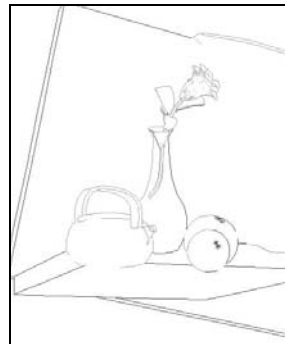


Fig. 7. Wire frame of NPR system

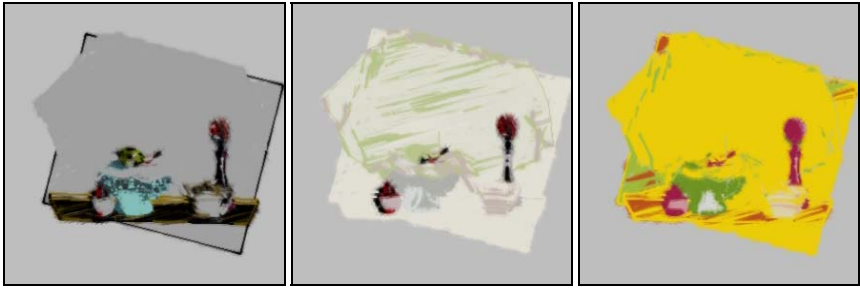


Fig. 8. Images with (a) sad, (b) peaceful, (c) joyful feeling

Table 1. Statistic Results of Comparison

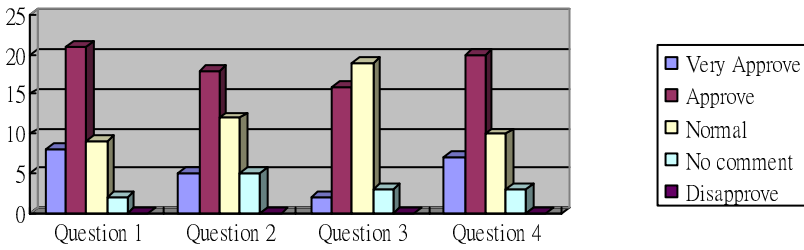


Fig. 9. Computer-Generated NPR Image



Fig. 10. The NPR Watercolor Image Result

5 Conclusion and Future Work

In this paper, we integrate the user-customized color charts with sketch-based GUI of 3D Watercolor NPR. The system allows users to build their own color model by keywords such as the name of artists. The web mining technology offers the capability for search images what needed, and the appropriate colors are extracted to construct the color chart by analyzing these images. Finally, the color chart is rendered in a 3D

visualization system which allows users to view and manage the distribution of colors interactively. Then, users can use those colors on Watercolor NPR System which allows user to manipulate watercolor attributes of object intuitively and manually.

More research efforts are needed to provide users better designing environment and more convenient tools to create their art works. For example, we are still surveying new web-mining techniques for better ways to extract the color charts more appropriately. Also, the GUI of our system still needs to be improved. Finally, the way for users to customize color charts should be with more flexibility and multi-modality.

References

1. Meier, B.J., Spalter, A.M., Karelitz, D.B.: Interactive Color Palette Tools. Department of Computer Science Brown University
2. Chu, J.-Y.: Chromatics: color scheme & match colors. Art, Designing and Technology (2001) ISBN:957202731X
3. Li, S.-Q.: Interpretation of Chromatics. Artist Publishing, Taipei (1996)
4. Nagumo, H.: Color Combination Charts. Long Sea International Book Co., Ltd., Taipei (2003)
5. Nagumo, H.: Color Combination Samples. Long Sea International Book Co., Ltd., Taipei (2003)
6. Cheng, K.-Y., Lin, P.-S.: "Color Plan". Yi Fong Tang Publisher, Taipei (1987)
7. Yang, Y.-P.: 'Color Combination'. Go Top Information Inc., Taipei (2003)
8. Chen, L.-H., Lin, H.-H.: Shape-Oriented Brush Stroke Synthesis in Non-Photorealistic Rendering. In: CGW 2006 (2006)
9. Yu, D., Mei, R., Luo, L.-S., Shyy, W.: Viscous flow computations with the method of lattice Boltzmann equation. Progress in Aerospace Sciences 39, 329–367 (2003)
10. Chen, L.-H., Lee, Y.-J., Ho, Y.-H.: Simulating NPR Watercolor Painting based on the Lattice Boltzmann Method. In: CGW 2008 (2008)

Gas and Shadow Swing

Chi-Hung Tsai, Mei-Yi Lai, Che-Wei Liu, Shiang-Yin Huang, Che-Yu Lin,
and Jeng-Sheng Yeh*

Computer and Communication Engineering Department
Ming Chuan University
{windydodo, cravendo, drummerwei, shirleysky922, comouse}@gmail.com,
jsyeh@mail.mcu.edu.tw

Abstract. In our digital art, we design a folding fan as an interactive magic device. You can use it to play with gas around the world of illusions. Although gas could not be seen in our real world, we still want to interact with it in our illusions by the element of bubble shadows. Opening and swinging the folding fan can blow the bubble shadows away; closing and swinging it can break bubbles. If the magic fan touches the shadow of gas, the bubble shadows will explode and release colorful particles to surround you. Those actions are controlled and located by our circuits with Arduino board.

Keywords: Interactive digital art, folding fan, shadow, Flash, Arduino.

1 Introduction

The wind blows the ground gently. The air flows into the world slowly. The shadow passes through by people quietly. Suddenly, we remember the childhood memories that we played with shadows. So we choose the shadows to be our key idea: we focus on the interaction between people and shadows. Furthermore, we would like to add one more element into our digital work. In summers, what we usually think about is the folding fan. When the fan blows the wind in our work, it can make shadow to swing just like gas can flow. We design a folding fan to be our interactive controller. It lets us feel the magic by influencing shadows. With the help of dancing shadows, we can also feel the existence of gas flow.

2 Creative Ideas

We major in Computer and Communication Engineering. When we step into the field of Technology Art, we find computer and art can combine to become another work. So we are trying to develop it. Gas and shadow are always in our life everywhere. They appear sometimes and disappear at other times. Although we can see the shadow, we can't sight the gas. Even they have thousand or million, people excite to see them.

* Corresponding author.

According to our literature review, we find interactive works which is seldom to use folding fan. Therefore, we try to use folding fan to design an interactive work. We imagine the fan that has magic, then user can use fan to swing and breeze. After that you can see the gas elements. Most of the proportion of gas in aerosphere are making by hydrogen, oxygen, nitrogen, etc. We think the gas is just like shadow, and it follows and fills around us. We can simply aware the shadow, but we can't feel those gas elements existed. Therefore, we let shadow to display the appearance of gas elements. It can explain those gas accompany with us. They are transparent, but actually they touch us all the time. Those gas elements have life; also they have cycle, and they afraid of touching with people. That is the secret of nature. When we enter to this space, folding fan can help you to see those gas elements easily; it can be not only seen but also touched. Be careful, they are very weak.

3 Literature Review

“Sharing the meal” created by Kuo [1] is a funny creation about the shadow. The shadow will change their position or shape when you turn the plane. For example, some animals go along with the rim of the table and run very fast when you turn the plane with fast speed. The creation can be better relationship for people who get the meal on the same dining table (Fig. 1(a)).

In SIGGRAPH 2002, “Molecular Bubbles” [2] is an interactive art by Simpson. The screen will be affected by the body motion. For example, bubbles drifting on the screen and you can cut and circumscribe it by your shadow (Fig. 1(b)).

Then, the other creation is “Shadow Monsters” by Worthington [3]. It will make strange noise and shape when you put your hand or body in the area. Some shape adhered on your shadow. It makes you look like a monster. Your shadow will be difference kinds of monsters, when you shake your body. It looks like those monsters are fighting with the other, so you can play it with someone together at this space (Fig. 2(a)).

There is another interactive art created by Adam Frank. The creation's name is “Shadow” [4]. This is a special interactive creation with people. You would see



(a) Sharing the meal.



(b) Molecular Bubbles.

Fig. 1. Sharing the meal [1] and Molecular Bubbles [2]

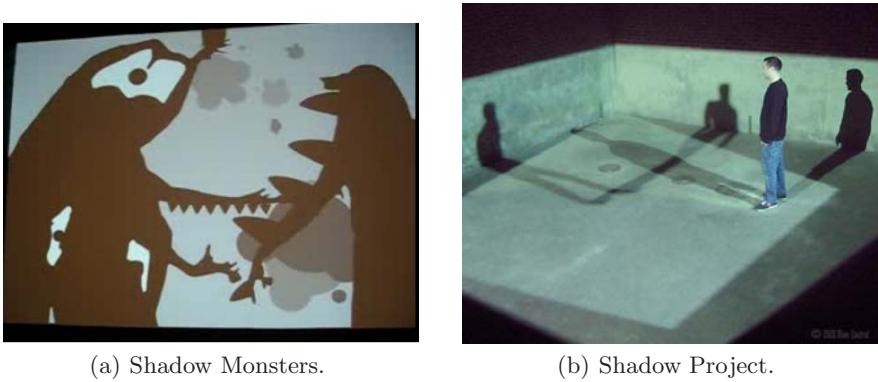


Fig. 2. Shadow Monsters [3] and Shadow Project [4]

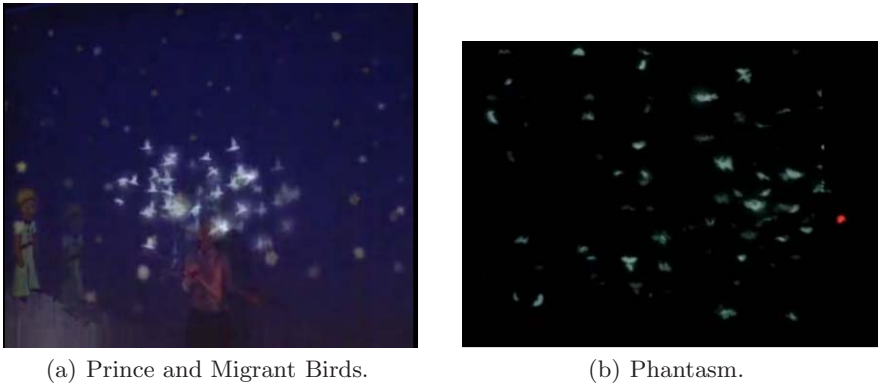


Fig. 3. Prince and MigrantBirds [5] and Phantasm [6]

a shadow beside you and scare of you. It likes the shadow says “Don’t approach me!” But, the shadow will be near to you when you intend to go away (Fig. 2(b)).

In SIGGRAPH 2006, “Phantasm” is one of creations by Takahiro Matsuo. The interactive art makes a lot of people to go back childhood. They don’t scare the insects and feel that the creation was very amazing. Takahiro Matsuo have another creation is “Prince and Migrant Birds” It’s depended on a story about prince. People like holding a magic ball and feel beautiful for the whole area with stars (Fig. 3(a) and Fig. 3(b)).

4 Research Methods

We use Flash [7] to paint the interacting contents and use Flash Action Script to receive the information of the folding fan, including position, shaking frequency, acceleration, and the opening status. Image processing technologies are used to

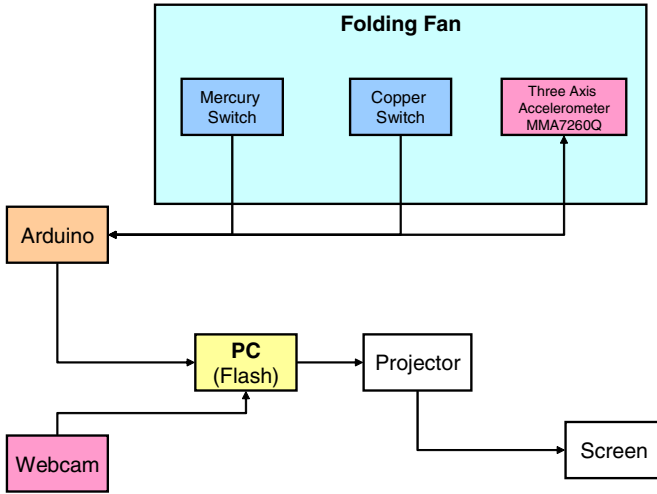


Fig. 4. Hardware setting diagram of our work

decide the region of user's shadow. The shadow regions are important for the interaction by testing whether the user is touching the shadow bubbles or not. The hardware setting diagram is displayed in Fig. 4.

We design a folding fan which is an interactive device in the user interface. It can be also an interactive media between user and screen. When the fan swings,



Fig. 5. Swing the folding fan can stir the object on the screen. If the fan moves faster with higher frequency, the objects will move faster.

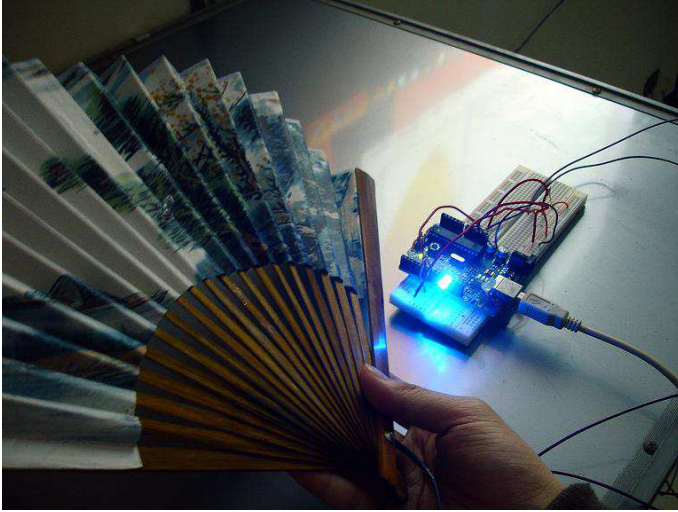


Fig. 6. Our sensor on the folding fan with Arduino



Fig. 7. A closer look of our sensor device on the folding fan

we capture the fan's statuses and play a Flash animation corresponding to the statuses (Fig. 5). Our concept is to simulate an invisible fan virtually in computer to show the amazing effects of fan. Therefore we attach a sensing device to the fan (Fig. 6). The device is built with three sensors: a mercury switch, a three-axis low g acceleration sensor, and a touching copper switch.



Fig. 8. We attach the sensor to the folding fan



Fig. 9. A newer and smaller version of sensor is attached to the folding fan

The mercury switch is used to calculate the swing frequency of fan. When the user swings the fan back and forth one time, the mercury switch will shake once and generate one bit signal. In order to receive the signal from the sensor, we use an Arduino board to receive the signal, to calculate the frequency, and to send the frequency to computer. Those signals will be transferred into PC via USB Port, and our interaction system, written in Flash, can receive those signals using Serial Proxy Server [8]. We can calculate the shaking rate per second then we can obtain the swing frequency from fan. When the user swings the fan and speeds it up, the colorful bubbles the screen will move faster and faster. It will make the simulation of virtual fan more realistically. Other interaction events can also use this information.

A three-axis accelerometer on the fan is used to measure the direction of movement, shaking speed and force, and to help roughly locating the position of fan. When the user open the fan to swing, the objects on screen will swing out slowly from the position of the fan. The speed of movement is based on the fan's shaking frequency. More and more virtual objects will be generated

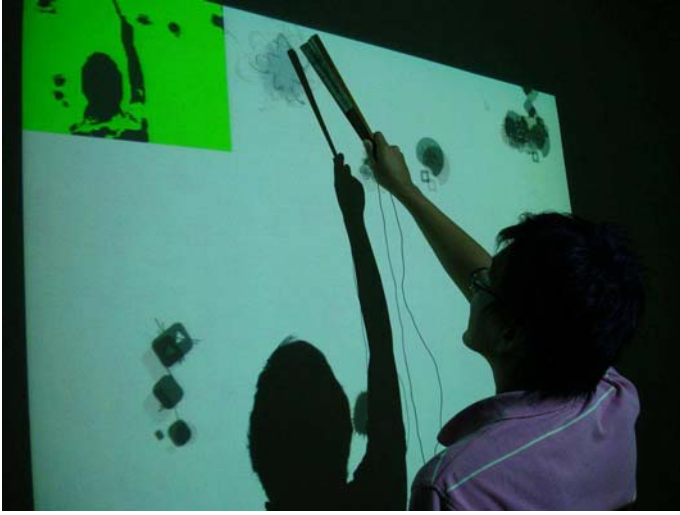


Fig. 10. The user can use the folding fan touch the shadow objects on screen. The shadow object can interact with the player.

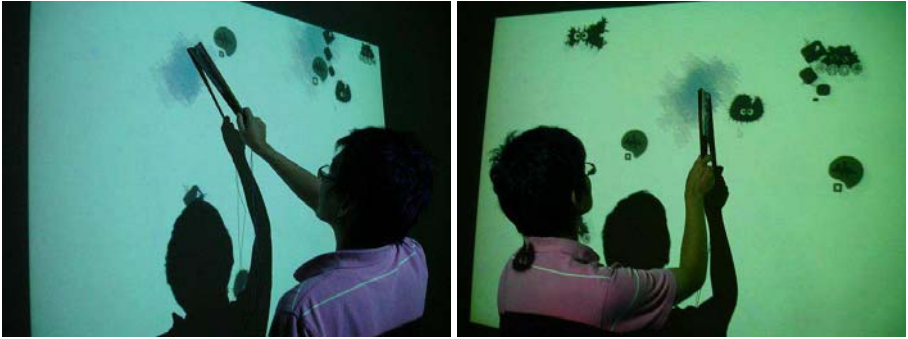


Fig. 11. When the user touch the shadow object, the object will break and release some color smoke spreading the smell of the gas

when swinging the fan and gradually there are full of the objects in the space. When closing the fan, the objects on screen will become touchable with magic exploding. At the moment of exploding, the colorful particles will be released to surround with you.

On the other hand, a webcam is also used in our work. We use the “Bitmap-Data” object in Flash Action Script to create array. We can use this array as a video cube to find where the user’s shadow is. We hope the user can not only swing the fan, but also use the fan to touch the gas shadows. When the user

touches the shadow objects on projector screen, the shadow object can interact with the player (Fig. 10). The shadow will be boomed by user, and make difference color on the screen (Fig. 11).

When we were working in progress, we faced two important problems. The first one is to compare two pictures, one is captured by camera and the other is the projecting screen. Because of projection geometry, there are some perspective projection effects. So we used the projection matrix to solve it. A homographic matrix with eight unknowns is solved and used here. Therefore, an initial calibration is necessary. This method can help the image processing to find the regions of human's shadow accurately.

The second problem is caused by the interactive contents on screen. When we draw any magic effect on screen, it will overlap with the shadow of the user and the fan. The overlapping will cause the image processing a little confusing. This problem can be solved by some image morphology operations and a difference operation. We implement and use them in Flash Action Script 3.0.

5 Conclusion

The proposed system provides a new direction of interaction. Image processing for video input from webcam and folding fan user interface are combined together. Some operations of the folding fan including opening, closing, swinging and touching, are designed to interact with the virtual world. When the users use the fan to touch the magic bubbles, bubbles will explode and release some colorful particles spread the virtual smell of the gas. Users can not only swing the amazing fan to generate different flying objects but also play the colorful bubbles or particles by themselves. The thousands of particles will interact and surround you.

References

1. Kuo, Z.K.: Sharing the meal (2006), <http://www.youtube.com/watch?v=GHeBod2Atfg>
2. Simpson, Z.B.: Molecular Bubbles. SIGGRAPH (2002), <http://www.mine-control.com/bubbles.html>
3. Worthington, P.: Shadow Monsters (2006), http://www.youtube.com/watch?v=g0T0Qo_7te4
4. Frank, A.: Shadow (2009), http://www.youtube.com/watch?v=oVAywSF_eVM
5. Matsuo, T.: Prince and Migrant Birds (2006), <http://www.youtube.com/watch?v=U14X5bEzsDU>
6. Matsuo, T.: Phantasm. SIGGRAPH (2006), <http://www.monoscape.jp>
7. Flash official reference document, <http://help.adobe.com/>
8. Busti, S.: Serial Proxy Server (1999), <http://www.lspace.nildram.co.uk/freeware.html>

Artist-Oriented Real-Time Skin Deformation Using Dynamic Patterns

Masaki Oshita and Kenji Suzuki

Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan
oshita@ces.kyutech.ac.jp, suzuki@cg.ces.kyutech.ac.jp

Abstract. In this paper, we propose an easy-to-use real-time method to simulate realistic deformation of human skin. We utilize the fact that various skin deformations such as wrinkles, bulging and protruding bones can be categorized into various deformation patterns. A deformation pattern for a local skin region is represented using a dynamic height map. Users of our system specify a region of the body model in the texture space to which each deformation pattern should be applied. Then, during animation, the skin deformation of the body model is realized by changing the height patterns based on the joint angles and applying bump mapping or displacement mapping to the body model. The pattern deformation process is efficiently executed on a GPU.

Keywords: Skin deformation, deformation pattern, height map, GPU.

1 Introduction

Character animation is used in many applications such as movies and computer games. However, creating realistic human skin is still difficult to achieve in computer animations. The rendering technique for human skin has recently been greatly improved [1]. However, simulating skin deformations remains a challenge, because the skin shape deforms in a complex way. Although it is possible to simulate human skin deformations using an anatomy-based physical simulation [14][15], in order to apply such methods to a human model, anatomical information such as bones, muscles, tendons, and tissues needs to be supplied in addition to the conventional human figure model. Specifying such information is both tedious for the designer and requires in-depth anatomical knowledge. Moreover, simulating the deformation requires a great deal of computational time, and as such cannot be used in interactive applications such as computer games and web graphics. Currently the SSD (Skeletal Subspace Deformation) model [9] is used extensively in computer games. The SSD simply deforms the geometry of a human model based on the posture of the skeleton and ensures smoothness of the deformed geometries. However, skin-like deformations are not simulated and it lacks reality.

In this paper, we propose an easy-to-use real-time method to simulate realistic deformation of human skin. We utilize the fact that various skin deformations such as wrinkles, bulges and protruding bones can be categorized into various deformation

patterns. A deformation pattern for a local skin region is represented using a dynamic height map. Users of our system specify a region on a body model in the texture space to which each deformation pattern should be applied. Then, during the animation, the skin deformation is realized by changing the height patterns based on the joint angles and applying bump mapping or displacement mapping to the body model.

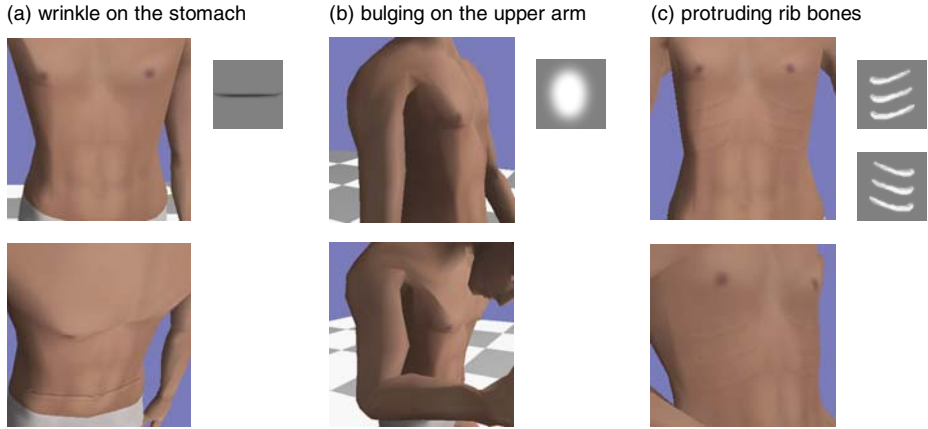


Fig. 1. Examples of skin deformations using dynamic height patterns. As the character moves, the height patterns are dynamically changed according to the angle of the corresponding joint.

We propose two deformation models for the dynamic patterns (Fig. 1). In the first model, the height patterns are scaled in response to the corresponding joint angle. This model is used to represent wrinkles and bulges (Fig. 1(a) and (b)). In the second model, the height patterns are biased in response to the corresponding joint angle. This model is used to represent protruding bones (Fig. 1(c)). The pattern deformation process is efficiently executed on a GPU.

Human bodies are considered to have common deformation patterns, because they share the same musculoskeletal structure, even if the sex, age, height, or weight thereof differs. Once we have created good patterns, these can be applied to various human models. With the pre-created patterns, users of our system merely have to place the required pattern on the texture space and specify appropriate deformation parameters. Although a user needs to have basic knowledge of skin deformation such as knowing to which region of the human body each pattern can be applied, knowledge of the anatomy is not necessary. Our method is thus easy to use and is expected to be useful.

The dynamic height map has previously been used in research studies and applications [6]. The main contribution of this research is the categorization of skin deformations into three types (wrinkles, bulging and protruding bones) and the realization thereof using two different deformation models. Moreover, by patternizing the deformations, our method allows users to apply common patterns to a new human model and to make these look believable more easily and more quickly.

The rest of this paper is organized as follows. In Section 2, we review related works. In Section 3, we analyze human skin deformations and describe how our

method realizes the deformations. Section 4 details the implementation of our method. Experimental results and a discussion are presented in Section 5. Finally, Section 6 concludes the paper.

2 Related Work

Methods for skin deformation can be categorized as either model-based or data-based methods.

Anatomy based physical simulation of skin deformation has been widely researched [14][15][19]. This model-based approach realizes very human-like skin deformation by modeling bones, tendons, muscles, and tissues as a real human body and by simulating the interactions of these with one another. In practice, however, such anatomy-based simulation is not widely employed since specifying anatomical information is tedious for creators and requires in-depth anatomical knowledge. Moreover, the simulation is generally computationally expensive, although it can be computed more efficiently by using pre-computed muscle deformations [19].

There are generic methods for simulating deformable objects that employ either mass-spring models [3] or finite element methods [17][2]. These simulations can be driven by a skeleton [2]. Although bulging deformations can be simulated in this way, wrinkles and protruding bones cannot.

In addition to the methods specific to skin deformations, wrinkle generation or deformation methods for cloth simulation can also be applied to skin deformation. These methods generate wrinkles on a coarse cloth mesh based on the compression of the mesh computed using physics-based simulation or geometric deformation. Oshita et al. [12] proposed a geometric method to add wrinkles to the edges of a cloth mesh based on the length variation of the edges. Larboulette et al. [8] described the generation of wrinkles on a cloth or skin mesh along the position and orientation specified by the designer. The curves of the wrinkles are controlled by the compression of the mesh. Li et al. [10] presented a method to generate wrinkles on a face based on the deformation of the keynodes and wrinkle lines specified by the user. Loviscach [11] developed a method to generate wrinkles over a cloth mesh using sinusoidal curves based on the wrinkle magnitude and direction computed for each vertex from the compression of the cloth mesh. These methods generate wrinkles based on the compression of the geometry. However, since human skin is more elastic than cloth, wrinkles cannot simply be computed from compression. Moreover, these methods cannot simulate bulging or protruding bones.

Various data-based methods have also been developed. Using sample data acquired from real human or physics simulations, realistic skin deformation can be represented without using any anatomical model or time consuming physics simulations. Sloan et al. [16] applied a shape blending method to scanned body shapes. Kry et al. [7] proposed the EigenSkin method to compute shape blending in real-time on a GPU using linear approximation. This method was applied to hand skin using sample shapes computed from a physics based simulation. However, in order to employ these blending methods, high-quality sample shape data must be prepared.

Hapap et al. [6] proposed a method for generating cloth wrinkles using height patterns. The height patterns, drawn by a designer, are scaled based on the compression

of the coarse cloth geometry. Cutler et al. [4] developed a wrinkle generation system for cloth animation, that uses curve-based wrinkle patterns created by a designer for each sample pose. These sample patterns are then blended according to a stress map, computed from the coarse cloth geometry, to generate the final wrinkles. Our method uses the height patterns created by a designer as input in a similar way to [6] and [4]. However, we deform the height patterns according to the joint angles, and not the compression of the geometry, because this is more suited to human skins. Moreover, our method simulates not only wrinkles, but also bulging muscles and protruding bones by changing the scales and offsets of the height patterns.

3 Skin Deformation

Based on the anatomy and observation of the human body [14][13], we categorize skin deformations into three types: wrinkles, bulging, and protruding bones (Fig. 1). Wrinkles are observed on the stomach, waist, wrist, neck, elbow, knee, etc. and are caused when the skin is compressed by bending or twisting an adjacent joint (Fig. 1(a)). Bulging occurs on the thigh, upper arm, etc. when muscles contract or tissue is compressed by bending an adjacent joint (Fig. 1(b)). Protruding bones are observed on the clavicles, ribs, scapulae, etc. and occur when bones are pressed against the skin by extending joints and the shape of the bones is seen through the skin (Fig. 1(c)). All these deformations are caused by moving a corresponding joint.

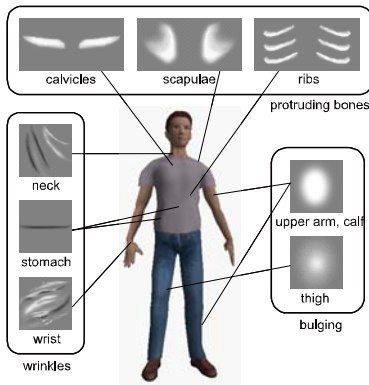


Fig. 2. Examples of the three types of deformation with height patterns for each. The patterns are dynamically altered according to the angle of the corresponding joint.

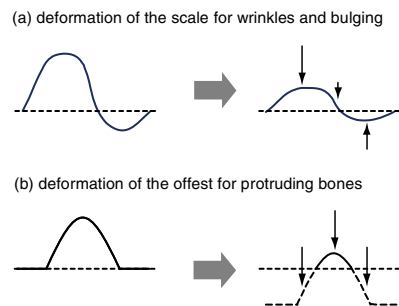


Fig. 3. Two deformation models for height patterns. (a) For wrinkles and bulging, the scale of the height pattern is controlled. (b) For protruding bones, the offset of the height pattern is controlled.

Our method simulates these deformations using dynamic height patterns. The height pattern for each deformation is created by a designer (Fig. 2). By changing the scaling or offset of each height pattern according to the angle of the corresponding joint, a height map is dynamically generated. For wrinkles or bulging (Fig. 3(a)), the

scaling of the height pattern is changed to vary the height and depth of the wrinkles or bulging. In the case of protruding bones (Fig. 3(b)), the offset of the height pattern is changed to vary the area revealing bones, without changing the shape of the bones.

We use a height image for each pattern. Each pixel has a signed height value between -1.0 and 1.0. Each pattern is dynamically altered based on the angle around a rotational axis of the corresponding joint. For example, wrinkles on the stomach are caused by bending at the waist (that is, by moving the waist joint around the front-back rotational axis). Rotations around a difference axis of a particular joint cause different deformations. For example, twisting around the waist joint causes wrinkles on the sides of the waist. Since such deformations from different axes are considered to be independent, we can assign different patterns to each axis.

To date we have prepared 8 basic patterns as shown in Fig. 2. As explained in Section 1, these patterns are common to every human body, and can therefore, be applied to any human body by placing a pattern in the texture space and setting the appropriate deformation parameters. Users may also add their own custom patterns.

4 Implementation

An overview of our method is shown in Fig. 4. A conventional SSD (Skeletal Subspace Deformation) model [9] and pattern information are given to the system. The pattern information contains the positions, scales and orientations in the texture coordinates for each height pattern. It also contains deformation parameters. The pattern information is specified by the user via an easy-to-use interface.

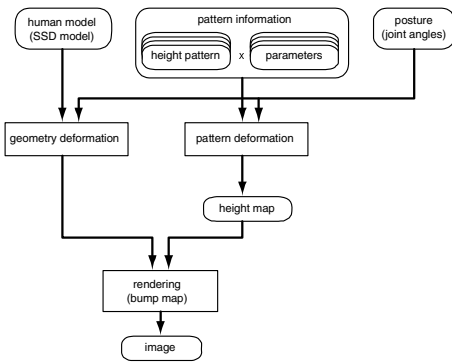


Fig. 4. System overview

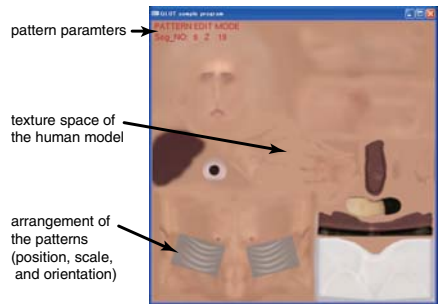


Fig. 5. User interface

Rendering of a human mode is done in two stages. In the first stage, the height map is generated by rendering each pattern on the specified region on the height map while changing the scale and offset of the patterns according to the joint angles of the character given by the animation system. This process is computed on a GPU using a render-to-texture technique and programmable shaders. In the second stage, the SSD model is rendered with the height map using bump mapping and displacement mapping. The

deformation of the SSD model is also computed efficiently on a GPU using vertex shaders [5]. Since all processes are computed on a GPU, our method executes very quickly. In addition, our system can easily be integrated with conventional rendering frameworks of existing animation systems or computer games.

For rendering the SSD model using a height map, displacement mapping is expected to generate better images than bump mapping. However, it is still a challenge to apply displacement mapping in real-time on the current graphics hardware and the shader model. Therefore, in our current implementation, we use both bump mapping and geometry deformation with two height maps. The height map for wrinkles and protruding bones is used for bump mapping, while the height map for bulging is used to move the vertices of the geometry of the human model in their normal direction. Since bulging affects the contour of the model, geometry deformation is considered necessary, whereas wrinkles and protruding bones are subtle and thus bump mapping is sufficient to represent these.

4.1 User Interface

Users of our system specify how each deformation is applied to a human figure model. We have developed a simple interface as shown in Fig. 5. Users can choose a pattern via the keyboard and change the position, scale, and orientation of the pattern in the texture space with the mouse. They can also specify the necessary deformation parameters, as explained in Section 4.2, using the keyboard. The result of the deformation can be reviewed by switching the view from the texture space to a 3D space showing the rendered model with animation. The user repeats this process for each deformation pattern. Typically, it takes between ten to twenty minutes to specify all patterns on a new human model.

Users can add their own custom patterns to the system. These patterns are created using existing digital painting software such as Adobe Photoshop or Corel Painter.

4.2 Deformation Model

As explained in Section 4, each pattern is rendered on the specified region on the height map. The output height of each pixel is computed based on the height pattern using the following equations. Recall that we use two different models for the different types of skin deformations.

For wrinkle and bulging patterns, as shown in Fig. 3(a), the height of the corresponding pixel of the pattern is scaled based on the angle of the specified joint,

$$f = \frac{\theta - \theta_{min}}{\theta_{max} - \theta_{min}} + 1, L_{out} = \begin{cases} 0 & \text{if } \theta \leq \theta_{min} \\ L \cdot h & \text{if } \theta \geq \theta_{max} \\ L \cdot h \cdot \log(f) & \text{otherwise} \end{cases}, \quad (1)$$

where L is the height value from the corresponding pixel of the pattern and θ is the joint angle. θ_{max} is the joint angle at which the height reaches its maximum value (the value of the ordinal pattern), while θ_{min} is the joint angle at which the height is zero. Since wrinkles and bulging are non-linear, we use a logarithmic function. h is the

parameter that controls the overall scale of the pattern and is specified by the user. This parameter is used to make the patterns more or less noticeable for individual figure models and body parts (the default value is 1.0). The deformation parameters $\theta_{max}, \theta_{min}, h$ are changed by the user for each pattern.

For protruding bones, as shown in Fig. 3(b), the height of the corresponding pixel of the pattern is biased according to the angle of the specified joint.

$$g = \frac{\theta - \theta_{max}}{\theta_{min} - \theta_{max}} + 1, L_{out} = \begin{cases} 0 & \text{if } \theta \leq \theta_{min} \\ L \cdot h & \text{if } \theta \geq \theta_{max} \\ \max[h(L - \log(g)), 0] & \text{otherwise} \end{cases} \quad (2)$$

where $\max[a, b]$ returns the higher value of a or b . As for bones, the shape of the bones does not change even if the joints are moved; what changes is how much of the bones is noticeable. Therefore, in our model, the offset of the height is biased according to the angle of the specified joint.

5 Experiments and Discussion

We have tested our method with a human model and the deformation patterns shown in Fig. 2. The human model consists of 4340 triangles and a texture image. Its skeleton has 50 bone segments. We have developed a program for the experiments using Visual Studio 2005 (C++), OpenGL, and Cg [5] for GPU programming. The program has been tested on a standard PC (Pentium 4 3.2 GHz, 1 GB Memory, GeForce 6800GT) with Windows XP Pro.

The time taken to render one human model with 15 patterns is 11 milliseconds with a frame rate of 150 fps when all processes are executed on a GPU using render-to-texture and programmable shaders. However, the speed drops to 19 milliseconds and 48 fps when the height map generation is executed on a CPU. These experiments show that our method achieves real-time performance by utilizing a GPU.

Examples of generated animations are shown in Fig. 1 and 6. Additional generated animations are available on the authors' website (<http://www.cg.ces.kyutech.ac.jp>). Included in the animations is a comparison with static patterns (standard bump map), showing that our method generates more plausible animations by dynamically changing the height patterns. Although we currently use a simple shading model, the skin shading can be improved by using a more advanced model [1]. This is beyond the scope of this paper, since the aim of our method is to improve the deformation of skin and as such shading is not dealt with at present.

As explained in Section 4.2, we use a simply approximated deformation model that does not reflect the physics. Although we have tried more complex models, the differences were not noticeable because skin deformations are very subtle. On the contrary, the differences between our model and the linear deformation model were indeed noticeable; as were the differences between the scaling deformation for wrinkles and bulging and the offset deformation for bones. We believe that the complexity and quality of our model suits our purpose admirably.

(a) wrinkles on the stomach and protruding scapula bones



(b) bulging on the upper arm

**Fig. 6.** Images from generated animations

A weakness of our model is that at present it cannot handle deformation from external forces (e.g., from contact with other characters or objects) or gravity. In order to handle such deformation, physics of some kind needs to be introduced. We believe that simplified physics such as that used in [2] can easily be integrated into our method since our method relies only on height maps and does not change the geometry which can be deformed by using a physics-based method.

In the latest generation of computer games, characters have been given their own height maps to improve the quality of the rendering with a relatively small number of polygons. Although the human model in our experiments does not have such a height map, our method can be applied to character models with height maps by adding the original height map and the dynamic height patterns on the fly.

6 Conclusion

In this paper, we proposed an artist-oriented pattern-based real-time skin deformation technique. Experiments show that our method works in real-time and produces plausible animations. Users of our system are able to make a character model look more human-like merely by specifying how predefined patterns are applied to the model. We believe that our method is very practical. Our future work includes improving the deformation model and integrating it with a simplified physics simulation. Improving the default deformation patterns is also seen as future work.

References

1. Beeson, C., BJORKE, K.: Skin in the “Dawn” demo, GPU Gems, pp. 45–72. Addison-Wesley, Reading (2004)
2. Capell, S., Green, S., Curless, B., Duchamp, T., Popovic, Z.: Interactive Skeleton-Driven Dynamic Deformations. ACM Transactions on Graphics (Proc. of SIGGRAPH 2002) 21(3), 586–593 (2002)

3. Chadwick, J.E., Haumann, D.R., Parent, R.E.: Layered construction for deformable animated characters. *Proc. of SIGGRAPH 1989* 23(3) (1989)
4. Cutler, L.D., Gershbein, R., Wang, X.C., Curtis, C., Maigret, E., Prasso, L.: An art-directed wrinkle system for CG character clothing. In: *Proc. of ACM SIGGRAPH Symposium on Computer Animation*, pp. 117–126 (2005)
5. Fernando, R., Kilgard, M.J.: *The Cg tutorial: the definitive guide to programmable real-time graphics*. Addison-Wesley, Reading (2003)
6. Hadap, S., Bangerter, E., Volino, P., Magnenat-Thalmann, N.: Animating wrinkles on clothes. In: *Proc. of the conference on Visualization 1999*, pp. 175–182 (1999)
7. Kry, P.G., James, D.L., Pai, D.K.: EigenSkin: real time large deformation character skinning in hardware. In: *Proc. of ACM SIGGRAPH Symposium on Computer Animation 2002*, pp. 153–160 (2002)
8. Larboulette, C., Cani, M.-P.: Real-time dynamic wrinkles. In: *Proc. of Computer Graphics International 2004*, pp. 522–525 (2004)
9. Lewis, J.P., Corder, M., Fong, N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In: *Proc. of SIGGRAPH 2000*, pp. 165–172 (2000)
10. Li, M., Yin, B., Kong, D., Luo, X.: Modeling expressive wrinkles of face for animation. In: *Proc. of Fourth International Conference on Image and Graphics 2007*, pp. 874–879 (2007)
11. Loviscach, J.: Wrinkling coarse meshes on the GPU. *Computer Graphics Forum (Proc. of EUROGRAPHICS 2006)* 20(3), 467–476 (2006)
12. Oshita, M., Makinouchi, A.: Real-time cloth simulation with sparse particles and curved faces. In: *Proc. of Computer Animation 2001*, pp. 220–227 (2001)
13. Sakuragi, A., Takeda, M.: *Anatomy for CG Creators* (in Japanese). Bone Digital Inc. (2002)
14. Scheepers, F., Parent, R.E., Carlson, W.E., May, S.F.: Anatomy based modeling of the human musculature. In: *Proc. of SIGGRAPH 1997*, pp. 163–172 (1997)
15. Simmons, M., Wilhelms, J., Gelder, A.V.: Model-based reconstruction for creature animation. In: *Proc. of ACM SIGGRAPH Symposium on Computer Animation 2002*, pp. 139–146 (2002)
16. Sloan, P.-P.J., Rose III, C.F., Cohen, M.F.: Shape by example. In: *Proc. of Symposium on Interactive 3D Graphics 2001*, pp. 135–143 (2001)
17. Terzopoulos, D., Platt, J., Barr, A., Fleischer, K.: Elastically deformable models. In: *Proc. of SIGGRAPH 1987*, pp. 205–214 (1987)
18. Wilhelms, J., Gelder, A.V.: Anatomically based modeling. In: *Proc. of SIGGRAPH 1997*, pp. 173–180 (1997)
19. Yang, X., Chang, J., Zhang, J.J.: Animating the human muscle structure. *Computing in Science & Engineering* 9(5), 39–45 (2007)

Interaction Analysis in Performing Arts: A Case Study in Multimodal Choreography

Maria Christou and Annie Luciani

Laboratoire ICA
Grenoble INP
46 av. Félix Viallet, 38 000 Grenoble, France
{maria.christou, annie.luciani}@imag.fr
www.acroe.imag.fr

Abstract. The growing overture towards interacting virtual worlds and the variety of uses, have brought great changes in the performing arts, that worth a profound analysis in order to understand the emerging issues. We examine the performance conception for its embodiment capacity with a methodology based on interaction analysis. Finally, we propose a new situation of multimodal choreography that respects the aforementioned analysis, and we evaluate the results on a simulation exercise.

Keywords: Embodiment, Digitally enhanced performance, Interaction methodology, Dance.

1 Introduction

Even though technologies have since very early on been correlated to performing arts, it is in the digital era that great changes were introduced in the way we treat and perceive the body, space, time and the interactivity of a performance [1]. Contemporary artistic creations use more and more frequently information and communication technology as composite of an art installation or spectacle, and sometimes even for appointing virtual interactive worlds. Under this context, the interactive environment - the interface of interaction - in an art performance becomes a powerful dramaturgic element. This ever-growing use of modern technology in art creation and variety of different uses, make a great field for analysis research from which interesting results can be subtracted concerning interactivity.

To examine the properties and the relations between the participants of an interactive performance, we propose a complete methodology based on interactive theories that promote embodiment.

The proposed methodology is applied to a case study of a Dancer – Public – Environment interaction. Embodiment is inherent to dance for a dancer; it is also this notion that makes the choreographer and the public able to understand the communicated feelings through the motion and the interaction. It is then important to design the interaction having interactive scenarios in mind.

In the triangle of participating actors of a dance performance, dancer- choreographer-public all have different ways of seeing and perceiving the motion in the digital environment. Most computer programs produce movements based on descriptive models and not generating ones. However, choreographer needs to express the movement in terms of a cause and not by describing the result [2]. Thus, it is important to adopt dynamic models to generate movements. We argue that a physically based particle system, such as MIMESIS [3], can make the procedure of designing a digital performance, and also the simulation results, to be more plausible to the dancer and more intuitive as a concept of both dancer and choreographer.

Our study case has been developed and evaluated during our collaboration with the choreographer Wolf Ka, creator of an interactive piece designed for dancer and spectator [4].

2 The Typology of Actors and of Their Interactions

The procedure for defining the interaction system in a spectacle installation is to find the basic roles and assign them properties, then to put them in relation and examine their interactions. Although the beginning of this deliberation was an installation for a dancer and a spectator, it is essential to place this problem within a larger framework in order to better understand the nature of interaction and to continue seeking possibilities that are not yet employed.

After having tested different cases for modelling the roles in an interactive dance performance, we came to the conclusion of five types of factors.

We can see below a diagram overview of the factors we are going to use in this paper.

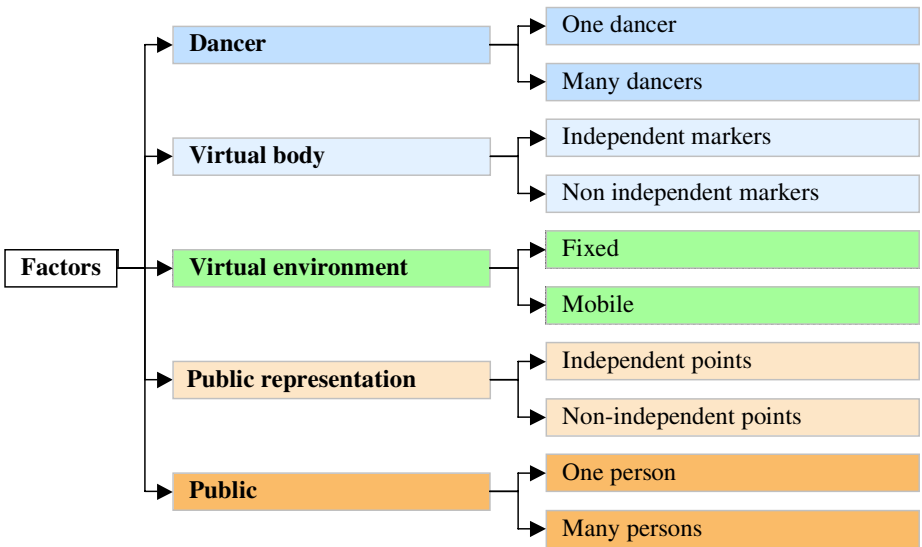


Fig. 1. The typology of the interacting factors

The first factor is the dancer or dancers. It is important to note that the properties of the dancer's role are different if there are several dancers on the scene; so we define two sub-types correspondingly.

A second factor will be what is called the "virtual body". The virtual body is the representation of the dancer in the virtual world. Motion sensors can be used to capture the dancer's movement. We also have two subcategories here, depending on if the sensors appertain to different persons or to the same one.

In terms of the virtual world we come to the third factor which is the animation environment, that is to say the virtual space of the interaction. We can distinguish two cases of this environment, either it is a fixed milieu or it allows movement.

The public in an interactive performance has an active role too. As the dancer, the audience too has a virtual equal. Its representation in the virtual space can be made by one or multiple points. Again we set two subtypes, according to the dependence of the points, whether they belong to the same person of the public or to different ones.

Finally the last factor is the public, members of the audience, who are able to use the digital interaction system.

3 The Interactive Analysis

We examine all the possible types of interaction between the different factors in order to verify how the adopted representations could be suitable in supporting the embodiment criteria of the perceptibility of the artistic intention. We try to explore how to enforce the sensation of "being with" at the time of the realisation and the memorability of the performance, even before considering aesthetic appraisal. We examine through this prism some preliminary cases of the dance performance "*Moving by Numbers*" by Wolf Ka.

The analysis process has two stages - firstly, it is necessary to define the behavioural possibilities of the factors. Secondly, we define the type of interaction between them. The combination of these two can give great graphic variations. We can summarise this result into a metaphor in order to render the interaction more familiar to the human experience.

Our analysis showed that there were some challenging points in the conception of the interactions in a preliminary version of "*Moving by Numbers*". The first point that challenges the comprehension for the dancer is the integration of multiple spatial scales. In fact, the dancer can move freely into space whereas the virtual environment on which he/she acts remains fix. This leads to ruptures of the relation between the dancer and the virtual environment. The second point is the change of the virtual environment by the spectator. Indeed these kind of changes break the movements effected and perceived under the natural assumption of the context permanence, if we don't give any prior indication that allows us to anticipated them.

4 A New Situation

While getting in more depth of these study cases, we felt the need to explore the interaction levels for manipulating the models of the virtual world. Claude Cadoz has summarised the essential levels of gestural control [5] to the following three, which we explain for our case:

- Instrumental level: where the relation between the human and the model is direct and results in the excitation of the structure (triggering the movement).
- Parametric level: where the gesture does not result in an animation but it modifies the properties of the model. This modification aims at modulating the type of movement (i.e. how the model behaves).
- Structural level: where the gesture changes completely the structure of the model.

We argue throughout our experimental experience in the context of the Enaction Research, that the intervention to the Instrumental Level is the easiest to perceive and to embody. Changes on parametric and structural level, during the performance and without full attention can be harder to understand.

Under the light of these observations, we have proposed a new artistic situation. We suggest another kind of interaction between the user and the environment. In our case, the spectator does not alter the parameters or the structure; he is rather an equal co-player on this Instrumental Level. In this new situation, all the factors will be integrated in the same installation in an interaction with an artistic value.

The interaction between the dancer and the spectator, who now becomes a user, is made through and with the virtual environment model. It is a case of a triptych conception of dancer-environment-user where the interactivity respects the instrumental metaphor.



Fig. 2. The interaction scheme for the new situation

Our interaction system involves all the factors seen in the typology. More precisely we have:

1. A dancer equipped with a motion capture point,
2. his virtual body which in this case is represented by a mass that is interaction with
3. the environment which is a formation of masses in a flock of birds or a fish school, and with which also interacts
4. the representation of the user in the virtual space by a point
5. and the user itself who moves his pointer in the environment trough a manipulator.

We could employ a metaphor for this system, where the dancer guides the flight of birds and the spectator comes to disturb this flock. The virtual environment is constituted of a physical model of a flock and the interactions between it and the other actors: the dancer's and user's virtual representations.

5 The Modelling of the Virtual Elements and Their Interactions

Methodology and modelling should not be regarded as two independent fields of development. The physical modeling approach suggests exactly the opposite. For the modeling work we have used the software MIMESIS whose approach to motion simulation it is based to Newtonian physics principle of Action-Interaction.

The interaction system we suggest has three virtual agents: the dancer's and the user's virtual representations, and the flock formation. The three are present in our virtual environment and in our model.

We chose to simulate the relation between the masses of the flock in a way to approach the behaviour of boids [6] for representing birds' formations. The inertial masses of the flock are connected in cohesion forces on agglomeration (Fig. 4, left). The cohesion function that we used is like the Van der Waals intermolecular cohesion forces. It is composed of three zones according to the distances between the interacting bodies. For long distances, it makes the birds fly free. For medium distances the interacting bodies are attracted, as for very short distances a repulsive force tends to separate them. Then at the medium and short distances, the combination between attractive and repulsive forces creates a cohesion zone keeping the birds together to form an ensemble.

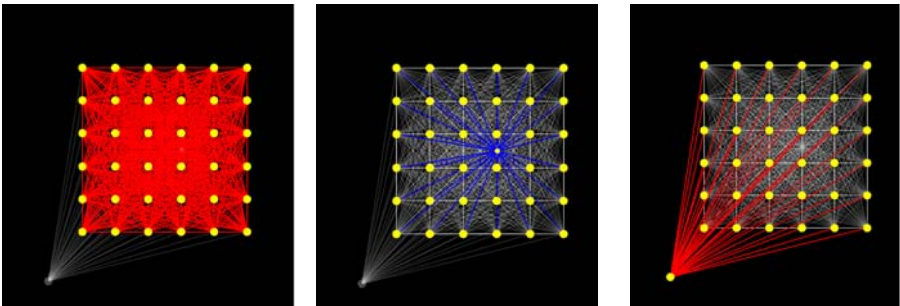


Fig. 3. The physical model and the interactions as seen in MIMESIS for: The Flock (on the left) here highlighted the cohesion interactions between the agglomerate masses - The Dancer (on the middle) spring-type interaction between the dancer and the agglomerate - The Spectator (on the right) repulsion interaction between the spectator and the agglomerate

Then each of these masses of the flock of birds is connected to two other masses that represent respectively the dancer and the user. With the virtual dancer there is an damping spring relation (Fig. 4, in the middle). It makes the “birds” flying around it. And with the virtual spectator, the force is repulsive, making it to repel the flock, like a predator (Fig. 4, on the right).

6 The Simulation Exercise

Our proposed model needs two simultaneous and real-time inputs, one from the dancer and one from the spectator who is a member of the audience. For the audience

input we suggest a joystick (or mouse) input. The dancer’s motion capture introduced aspects we took into account. We decided not to change the scale factor for the dancer, so the correspondence between his movements and the animated ensemble would be in the closest association possible. Concerning the motion sensor and taking into consideration the type of the movement we imagine to acquire, we concluded to place one sensor on the head of the dancer. The artistic approach proposed does not consist a final suggestion for performance. It is a scenic proposition answering to our interaction design demands and using the initial concept of interaction proposed by Wolf Ka.

For the new situation, we transferred the CORDIS-ANIMA algorithms [7] used in the MIMESIS software to a platform able to host the motion capture and to treat the computation demands of the simulation and projection. There, we implemented six examples with different parameter settings, and two different visualisations. A very important part in this methodology is assigned to parametrizing the model. Proper parameters are able to promote the embodiment. This game of parameters, the phenomenology, plays a great role in the final experience as we remarked during the response of the dancer to different settings.

Table 1. Parameters of the relations between the masses of the model as used in MIMESIS

Agglomerate (3-level cohesion)	Dancer (Damped spring)		Spectator (Repulsion)		
1rst level elasticity	1,00e-05	Elasticity	1,00e-06	1rst elasticity	level 0,001
1rst level viscosity	1,00e-04	Viscosity	1,00e-04	1rst viscosity	level 1,00e-05
1rst level threshold	1	Threshold	2,5	1rst threshold	level 2,5
2nd level elasticity	1,00e-10			2nd elasticity	level 1,00e-06
2nd level viscosity	1,00e-06			2nd viscosity	level 1,00e-05
2nd level threshold	1,2			2nd threshold	level 6

In the simulation exercise that we have conducted with the dancer who was implied at the work of origin, we tested the different settings (variations to the elasticity, viscosity and distance level of the interactions), the different visualizations (coatings), and we examined the users’ response.

We noticed that the dancer integrated with great facility the interaction with the virtual world, the immersion was intuitive and we didn’t need to give any instructions. There was a high degree of engagement from the part of the dancer and the spectator, and the given animations and dancing movements were of great interest. It was also interesting to see how the different coatings, one where the masses were spheres and a second where they had trails, changed the dancing behavior. In the first case the dancer tried to round up the flock around her, and in the second she was dancing so to promote movements of the flock that crossed her body.

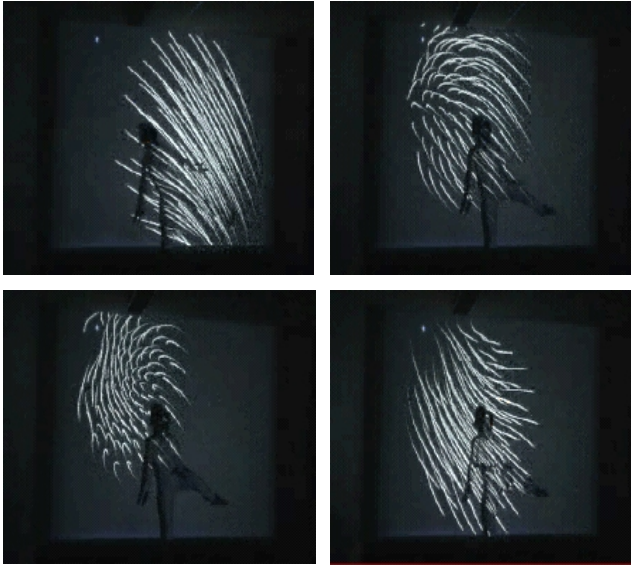


Fig. 4. Snapshots of the Flock simulation exercise. The flock of birds exhibits a cohesive behavior and it follows the motion of the dancer.

7 Conclusion and Perspectives

There is a question that is always present when designing an interface for a computer-mediated interaction: how do we offer to users a meaningful experience and create a believable virtual world? The method that we developed and followed for this case-study is based on the Embodiment theory, which tries to describe possible methods and practices to overcome such questions.

The proposed simulation exercise does not mean to be an integral spectacle; it is suggested as an example for implementing all parts of the produced theory and methodology. It can be though easily modified to employ new ideas, as it is very flexible.

The physical models are also by nature very modular and flexible to modifications. All the physical models are combinable and compatible. That allows us to integrate a lot of models within the same scenography.

Finally this work proposes well-founded and specific schemes of interaction for choreographers interested in designing an interactive performance, and it might give them the need to explore the interactions with physical models.

As a theoretical perspective, the underlying fascinating question of embodiment in a complex interactive environment may have here an experimental support. One fruitful track of research is, in our opinion, the analysis of what we called “the interaction metaphors” considering the whole system.

A very interesting subject of research that sprung from this work, is the question of the duality of the body (physical – virtual) and with that the possibility to play with the frontier of what it is interior and what exterior. The passage from “object of the environment” to “object near the hand” to “object in hand” is a transforming process

for both the instrument and the man [8]. Man transforms the object to instrument; instrument transforms the man into an instrumentalist. This last phase of incorporating the object is exactly the beginning of the embodiment process. Embodiment as well as disembodiment (the inverse process) are basic processes of artistic creation and constitute a rich field of research by virtue of interactive technologies.

References

- [1] Dixon, S.: *Digital Performance: A History of New Media in Theatre, Dance, Performance Art, and Installation*. The MIT Press, Cambridge (2007)
- [2] Hsieh, C., Luciani, A.: Generating dance verbs and assisting computer choreography. In: *Proceedings of the 13th Annual ACM international Conference on Multimedia* (2005)
- [3] Evrard, M., Luciani, A., Castagné, N.: MIMESIS: Interactive Interface for Mass-Interaction Modeling. In: *Proceedings of CASA 2006* (2006)
- [4] Ka, W.: Man in lelspace.mov / Motion Analysis in 3D Space. In: *ACMM 2005, Singapore* (2005)
- [5] Cadoz, C.: *Gesture and Music in Trends in Gestural Control of Music*. IRCAM Editions (2000)
- [6] Reynolds, C.W.: Flocks, Herds, and Schools: A Distributed Behavioral Model, in *Computer Graphics*. In: *SIGGRAPH 1987 Conference Proceedings* (1987)
- [7] Cadoz, C., Luciani, A., Florens, J.-L.: *CORDIS-ANIMA: a Modeling and Simulation System for Sound and Image Synthesis - The General Formalism* (1993)
- [8] Luciani, A., Magnusson, C., Castagné, N.: Final report on technological specifications for Enactive Interfaces. In: *Deliverable D.IA2.6., November 3. NOE Enactive Interfaces* (2006)

BWAIN: An Artistic Web Interpretation

Linda Huber, Michael Bißmann, Stefan Gottschalk, Alexander Keck,
Andreas Schönefeldt, Marko Seidenglanz, Norbert Sroke, and Alexander Radke

Art_Works

Technical University of Dresden, Dresden, Germany

art_works@web.de

Abstract. BWAIN is a creative web browser which uniquely combines graphics and sound using the HTML structure of any selected web page. The structure behind the web page which normally remains unseen, is revealed in a process of visualization. The HTML Code acts as a genome for building an image which looks and sounds quite distinct. Several web pages may be linked with each other and thus reflects the character of the web. The user may achieve a wide range of styles in the visualization and actively navigate through the scene by scaling and translating the image. The opportunity to scale and transform the image becomes a creative process resulting in varied compositions.

Keywords: Browser, Generative Art, Sound, Interactive Media.

1 Introduction

As both artists and computer scientists, we determined to conceive of a project which bridged both art and technology. How could we unite our visual abilities with our technical experiences in a unique work? Brainstorming began and we looked at significant projects, both local and international, so as to assemble inspirations which we could build upon. Computing, aesthetics and sound fascinated us and we pursued some ideas. We decided to build a program based on accidental occurrences. Other standardized graphic programs are limited in design quality, so that became yet another goal. Digital techniques are often only imitating real things rather than developing new approaches. This thought led us to contemplate and experiment.

The basic thinking behind this project had to do with the connections and parallels in pictures and sound. Both areas hold common ground: colour space, harmonies, contrasts and spatial compositions exist in art as well as music. Could a visual piece be woven together with a musical piece? What would happen if visual principles were applied to musical concepts? Is it possible to represent mathematical or physical models acoustically? Could one apply sound to a spatial representation which is two-dimensional or three-dimensional? And could one possibly navigate through this space?

Analysis of the paintings by the American artists, Elizabeth Murray and Al Held provided us with ideas about colour and shape assemblage. Lauri Groehn [\[4\]](#)

of Finland found a means to digitally translate paintings into sound. An earlier software program called, Create and Learn [1] was developed to work artistically in a digital manner. Here the user may learn basic principles in colour and design and be able to create multiple works which may be saved.

Several discussions about music led us to question digital music. We found two sources to enhance our idea search. One was the work of a group of computer science students [5] led by Claudio Midolo, who designed a musical instrument which combined graphic drawing with sounds and melodies. When a person holds this instrument, moves it and makes gestures, the result is a graphical translation of these movements on a digital screen. Here drawings in line, shape and colour can be realized and questioned. The second influence was the work of Yu Nishibori [6] who developed an electronic instrument with a graphic panel for controlling the sound. Still another project by Marcel Salathé [7] influenced our work in the use of the HTML structure to produce an image. The result was simplified geometrical configurations in colour. Our interest was to build upon this in creating more complex imagery.

The concept behind BWAIN was that one could access the endless source of structured data from the World Wide Web and use it as a basis for an acoustic and visual representation (Fig. 1). The result is an alternative browser which over the Internet, uses structure and layout information provided by the HTML and CSS files. This allows a process of representation which grows and develops. Through HTML links, there are navigation possibilities between the web pages. This gives the user the chance to work in an artistic manner, in an abstract search through the World Wide Web.

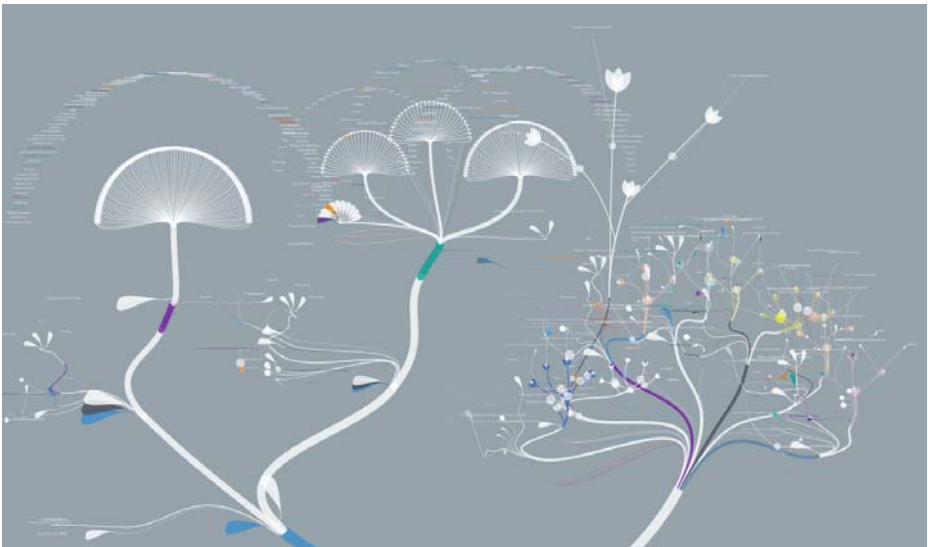


Fig. 1. The hidden structure of the web page www.ard.de is revealed in this artistic realization in the program BWAIN

2 Program Development

A previous project called Cellarium [2], served as a basis for discussion. Cellarium is a microbic simulation using mathematics and biology in a visual orientation. Microbes and their paths, represented through various functions in a colour composition use generated rules to create ever-changing results. The interaction between microbes in eating, pairing and hunting were used in conjunction with melodies, chords and harmonies which resulted in abstract compositions of a particular character. This idea led us to consider how we might build dynamic structures which were visual and acoustic in nature. Using static information masses to represent and influence them, we struck upon the notion of *creative browsing*. We began by reading a web page and analyzing its structure in terms of basic elements. The original structure of the web page would remain unchanged, however tags and attributes were used to define visual and acoustic elements, resulting in a completely new representation. The user sees an abstract structure growing and simultaneously, hears an acoustic composition. Through the selection of links, the structure may divide further and develop.

The results produced by BWAIN in terms of growth and aesthetics, is similar to the idea behind generative art. Professor of Art at New York University, Philip Galanter [3] defined this as practice by which an artist uses a system, and sets it in motion so that a relevant creation may evolve. Presently, there is a variety of generative art programs which may be built and navigated using computer programs. From random colour shapes to complex simulations, observers may react and view this as art. Two artists, Marius Watz [8] and Martin Wattenberg [9] take music, analyze it and generate a graphical output.

This project BWAIN is connected as well to generative art, while the results allow the viewer to enjoy basic formal principles, as one would when looking at an artwork. The fact that the visualization is derived from the Internet, promotes the idea that a system was used to develop the program. Every web page has in its basic HTML text, a tree with specific features, individual functions and changing complexities. This information is applied from web pages and then transformed, structured and represented using the web browser. In daily life and use, the content of a web page is important, but most users do not realize what lies behind the scenes. Its background structure is the supporting frame for all textual and graphical components, making the facade readable and comprehensible. BWAIN interprets this structure and substantiates the invisible nodes of a web page, making this a visual reality. If one departs from this expected and ordinary perception, the structure of a web page runs parallel to the interpretation of a musical piece in which rhythm and melody develop. Simultaneously, visual and acoustic elements swing together in cadence. After selecting a favorite web page, a creative process unfolds: a structure develops, grows and produces results. This process may be intuitively manipulated and directed by the user, whereby influences from the page and user choice may cause homogeneous results in both picture and melody.

After creative browsing was conceived, we sought to develop a strategy. The concept of *Director* was designated as the center point of our project. The

responsibility of the Director was to control the visual and acoustic results through time limitations in repeated rhythms. Next, a data structure was necessary so that the web page could be transformed.

3 Sound

Along with visualization, sound plays an important part in the presentation of the web page. The visual construction is dependent upon the speed of the music and provides orientation in terms of tact. Each web page possesses an individual flow in melody line with a particular length which is repeated endlessly, but with subtle variations. The various HTML nodes summoned from the particular web pages are given an additional tone which during the interaction of the Program, can be integrated in the existing melody line. The *SoundView* consists of various components with different functions for the acoustic realization.

The *Metronome* provides the tact for the music as the visualization process builds. It is given a speed in terms of beats per minute (BPM) which determines the periodical intervals in which it triggers actions for the visualization. The *Metronome* is instrumental in defining the *SoundView*. It generates a melody track which is sent to the synthesizer and may be played as a loop. Composed of a soundbank, the synthesizer is the actual producer of sound. All information about the music is transmitted by the synthesizer in the form of MIDI (Musical Instrument Digital Interface) data which are transformed into acoustically perceptible signals. At the moment, the synthesizer is realized by a standard Java soundbank, which uses audio samples. The soundbank may be replaced at any time by another type of sound production. At present, the generated melodies only depend on the form of MIDI data and therefore, may be linked to any desired synthesizer. The *Composer* is responsible for the generation of the actual music. It contains algorithms which generate melodies for various instrumental elements (bass, chords, rhythm). Every web page has a pattern of melody which consists of many tracks, in which various levels of instruments are represented. The various tracks are generated with different algorithms and then brought together in a MIDI sequence. This generated sequence is in turn submitted to the *Metronome* which plays them back. The *WebMusic* component is the connecting link between the HTML code of a particular web page and the generated music. In this part, the web page is parsed for meaningful input values, suitable for the algorithms of the *Composer* element. Here musical values, such as key, length of the melody and number of tracks, etc. are determined for an individual acoustical performance. Thereby, attributes such as the length of the URL, the number and the existence of particular tag types, etc. are taken into consideration. It should be mentioned that the *WebMusic* component does not generate the music completely out of the HTML code. One would need too many parsed values for each web page to sound identical and therefore, the melodic patterns are varied at random, but maintain the same character in tempo and key.

4 Visualization

If one should analyze BWAIN from its functional side, it is a web browser. This means that documents in the HTML format are parsed, read, analyzed and the components separated. These basic elements or tags are instantiated as objects and sequenced according to importance. Every object is represented by a tag on the HTML page. Each object possesses an attribute which is defined by the tags. Attributes possess characteristics such as colour, typeface, height and width, etc. How may we transmit the abstract information to the user? For one, the acoustic signals which are generated from the *SoundView* and two, in the form of visual clues provided by the *VisualViews*. In order to achieve synchronicity between the visual and audio components, a beat was introduced. Through a metronome this beat functions as the initiator for the growth process.

VisualViews are the visual representation of a corresponding HTML tag. Attributes from the HTML tag define the size and colour of the views. Different tag types result in variations of *VisualViews*. For example, the view of a body-tag looks different from the view of an image-tag. An important constructive element of those *VisualViews* is the Bézier curve. The thickness of a curve depends on the quantity of the load it must carry. One may compare this to a tree branch which must support the weight of its descending parts. Generally, one can distinguish between three groups of *VisualViews*: the organic, the geometric and the typographic. Organic views have a plant-like appearance (Fig. 2). The Bézier curve serves as a stem with leaves, fruits or blossoms emanating. Geometric views consist of basic forms such as circles, rectangles and lines. The typographic view is applied to text nodes in the HTML tree. It simply

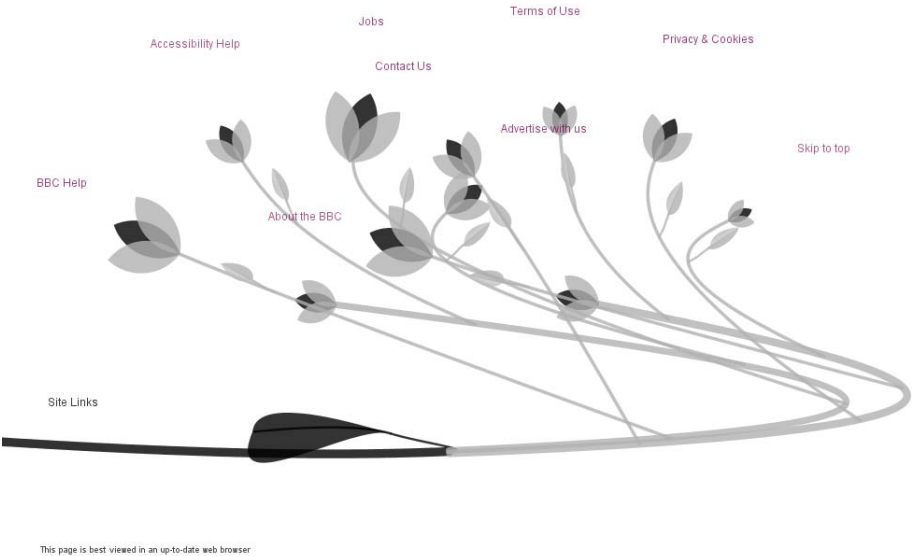


Fig. 2. Detail of the visualization of www.bbc.co.uk showing organic forms and text

shows the text of the web page. The way these views are mapped to the HTML tags is partially pre-defined in the program, in part due to the choice by the user. Using a slider bar in the *Control Panel* one may select an organic or geometric appearance with all possible variations between them.

5 User Interface

From the beginning, it was clear that we did not want a standard desktop application, but rather, one with aesthetical demands with an unconventional user interface. In addition, we had no interest in competing with functional applications. Standard *GUI* elements and the *Menu* were reduced for the sake of simplicity. Please refer to figure 3. The user has a *Control Panel* from which to begin, but then has access to a *Debug Panel* for further possibilities.



Fig. 3. Main Window with the Control Panel at the top and Debug Panel at the right

The user has two options to navigate through the scene. By dragging the mouse, you may *move* the field of vision. A *zoom* feature is possible by rolling the wheel of the mouse. The position of the mouse cursor may center on an object and zoom in and out, depending upon the desired movement. Yet another possibility occurs through *picking*, or selecting a particular node from which a sound emanates. *Rolling over* an object with the cursor causes a change to occur. For example, when the cursor moves across a word, rollover produces an unfolding of the full text. Other functions were programmed into BWAIN: *screenshot* and *soundshot* offer one the chance to save the visual and acoustic results. Both the *Debug Panel* and the *Control Panel* may be blended out, so that one may view a design in its entirety. The *Debug Panel* offers an array of other possibilities for the user to experiment. Here one has the opportunity to change the look of the image through

selection of programmed *Views (visibility)* which generate different reflections of a website. One may alter the grouping or direction of the image through its tree construction, or *branching mode*. These adjustments affect the way the elements are composed on the surface. Apart from the default tree setting, there is a random mode, which creates a very abstract look. When the river mode is chosen, the elements align themselves in a global direction, much like the flow of water. The orthogonal mode aligns all elements in horizontal and vertical directions. This may result in a composition that resembles technical imagery such as roadmaps or circuitry. The node type mode assigns individual branching modes depending upon the type of element used.

6 Conclusion

BWAIN, in its current form is transformable, and has the potential to expand in several directions. For example, one could show the entire free navigational three-dimensional environment. This form of representation would further develop the tree-like character of the web page. It would also be conceivable that in combination with quicker drawing components, continual animation could be implemented such as a perpetual wind or simple movements represented through shifting and turning. BWAIN serves to creatively inspire. A local artist reflected on the work by saying, that the typically unseen structure of the web page becomes the focus and therefore, the information content is unimportant. The function of the web page has changed and thus, becomes a work of art.

References

1. Huber, L., Christoph, N., Meissner, R., Radke, A., Weiss, C.: On Learning Colour – A Software Program. In: Proceedings for International Conference on Colour Harmony, Budapest (2007)
2. Huber, L., Bach, B., Foerster, M., Gottschalk, S., Schliebenow, A., Seidler, K.: Cellarium: A Software Application for Creating and Learning Colour Through Basic Biological Processes. In: AIC 2009, Sydney, Australia (2009)
3. Petersen, T., Ploug, K.: Generative art is as old as art. An interview with Philip Galanter (2004), <http://www.artificial.dk/articles/galanter.htm> (last visit: January 29, 2009)
4. Groehn, L.: Sound of Paintings (2007), <http://www.synesthesia.fi/> (last visit: January 29, 2009)
5. Midolo, C., Castellanos, E., Sinigaglia, N., Mari, P.: Musical Instruments, <http://de.posi.to/wiiwiiwiiwii/> (last visit: January 29, 2009)
6. Nishibori, Y.: TENORI-ON, http://www.youtube.com/watch?v=pU_rG0w0bsA (last visit: January 29, 2009)
7. Salathé, M.: Webpages as Graphs, <http://www.aharef.info/static/htmlgraph/> (last visit: January 29, 2009)
8. Watz, M.: <http://www.unlekker.net> (last visit: January 29, 2009)
9. Wattenberg, M.: The Shape of a Song, <http://www.turbulence.org/Works/song> (last visit: January 29, 2009)

Treasure Transformers: Novel Interpretative Installations for the National Palace Museum

Chun-Ko Hsieh^{1,2}, I-Ling Liu¹, Quo-Ping Lin¹, Li-Wen Chan², Chuan-Heng Hsiao²,
and Yi-Ping Hung²

¹ Department of Education Promotion, National Palace Museum
{kenneth, iliu, jameslin}@npm.gov.tw

² Graduate Institute of Networking and Multimedia,
Department of Computer Science and Information Engineering, National Taiwan University
{allenjam, hsiao.chuanheng}@gmail.com, Hung@csie.ntu.edu.tw

Abstract. Museums have missions to increase accessibility and share cultural assets to the public. The National Palace Museum intends to be a pioneer of utilizing novel interpretative installations to reach more diverse and potential audiences, and Human-Computer Interaction (HCI) technology has been selected as the new interpretative approach. The pilot project in partnership with the National Taiwan University has successfully completed four interactive installations. To consider the different nature of collections, the four systems designed against different interpretation strategies are uPoster, i-m-Top, Magic Crystal Ball and Virtual Panel. To assess the feasibility of the project, the interactive installations were exhibited at the Taipei World Trade Center in 2008. The purpose of this paper is to present the development of the “Treasure Transformers” exhibition, design principles, and effectiveness of installations from the evaluation. It is our ambition that the contributions will propose innovative media approaches in museum settings.

Keywords: Museum, interactive exhibition, HCI, context-aware, tabletop, virtual panel.

1 Introduction

The National Palace Museum holds a great quantity of Chinese artifacts and artworks as part of its world-leading museum service. With around 655,713 objects in collections, many artworks are masterpieces of Chinese emperors’ treasures. To share the invaluable cultural assets to the public, the museum has continually sought to make its collections accessible to greater and more diverse audiences. Nevertheless, museums in the 21st century have been confronted with many challenges from sifting societies. With the exception of modernist practices, museums need to be more aware of the power of learning, and engage audiences with more sophisticated and complex approaches [1]. In order to present stories or learning resources in galleries, interpretation is required to transfer material cultures into philosophical and historical narratives. In the last decade, new media have been widely implemented in galleries as means of interpretation. Besides, the *American Association of Museums* has given the

following definition “Interpretive Interactive Installations are made up of multiple kiosks or a full gallery installation and are interactive and educational.”[2] Human-Computer Interaction (HCI) has made interpretation strategies more versatile. As HCI technology is interactive computing systems designed against human behaviors, it can practically provoke conversations between museum objects and visitors. To foresee the new trend, the pilot project “Treasure Transformers”, in partnership with the National Taiwan University, was initiated to develop four interpretative installations exhibited in the IT Expo at Taipei World Trade Center in 2008. Four HCI systems are utilized to accommodate the different nature of the collections, including artifacts, paintings and documents. The context-aware system is employed to develop *uPoster: Ladies, people are coming*; the *i-m-Top* tabletop system is for *Livening up paintings on the Tabletop*; the virtual display system is for *Magic Crystal Ball*; and the virtual panel is for *Animals of Fantasy*.

2 Related Works

The following definition has been offered: "Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them." [3] In a museum context, HCI would be applied to reshape a knowledge place as an interactively intelligent environment. The project-related studies are discussed as below.

Context-Aware. Context-Aware was first introduced in 1994, when Schilit and Theimer stated “context-aware computing is the ability of a mobile user’s applications to discover and react to changes in the environment they are situated in.”[4] Similarly, Dey and Abowd stated that context-aware is very important in interactive applications, particularly when the users’ context changes rapidly in the environment [5].

Interactive Tabletop System. During the last decade, interactive tabletop systems have been gradually implemented in galleries. According to Geller’s survey, there are many examples to prove that tabletop systems are successful applications in museum settings [6]. The main reason is that the natural interface of tabletop systems can tackle the different levels of computer literacy of diverse visitors. Hence, the project employed the tabletop system *i-m-Top* developed by Hu and Hung. The system features multi-touch and multi-resolution display, which accommodates the multi-resolution characteristics of human vision [7].

Virtual Display System. Displaying three-dimensional images of museum artefacts requires a volumetric display technique, such as i-ball system [8]. i-ball is a transparent ball, which can display produced images through a special optical system. Based on the display module of i-ball2 [9], the Magic Crystal Ball, a gesture-based interaction system, was developed by Chan and Hung. Users can see virtual objects appearing within the transparent sphere, and manipulate them with barehanded interactions [10].

Virtual Panel. The virtual panel is a privacy-protected virtual screen, created by a special optical mechanism. As the virtual panel is intangible and floats in the air, apparently all interactions must be taken in the air. However, without tactile feedback,

users might not perceive whether interactions have happened or not. To solve the problem, a water ripple is used as a metaphor to personify the virtual interaction [11]. In this project, the intangible interaction in the air is applied to the design as a game scenario.

3 Design Principles

To consider the feasibility of interactive multimedia in museum settings, the team has defined four criteria as design principles: novelty, accessibility, reliability and learning potential, which are discussed below.

Novelty. As modern societies have shifted dynamically with innovations of new technology, the traditional interpretative devices may not fulfil the diverse museum visitors any more. For example, research conducted in SFMOMA (2006) indicated that visitors under 40 years old prefer to download a gallery tour to their own mobile devices rather than to rent an audio tour device [12]. In other words, the development of new technology inevitably influences visiting patterns in museums. That is to say, museums have to consider the novelty of new interpretation devices, which not only retain existing museum audiences, but also attract potential new ones.

Accessibility. Museum visitors generally with diverse backgrounds in age, language or education might have different computer literacy to interact with multimedia installations. Take the Churchill Museum for example, the evaluation undertaken in 2005 indicated that visitors required more guidance on how to operate the interactive initiatives, although less than a half the visitors actually used guidance in the gallery [13]. Therefore, it is important to bear in mind how to make multimedia interactives accessible to all visitors. In this respect, the concept of “Natural Interaction” [14] is utilized to increase the accessibility of the initiatives.

Reliability. IEEE defines reliability as “the ability of a system or component to perform its required functions under stated conditions for a specified period of time” [15]. From this point of view, the quality of system services can be evaluated to indicate the reliability. For example, theme parks like Disneyland must provide reliable entertaining facilities to sustain visitors’ repeat visits in the future. Similarly, to provide high-quality museum services, the new media installations must achieve a high degree of reliability before implementation.

Learning potential. In museum education provision, multimedia installations can be utilized to accommodate different learning styles and levels of learning. So far, the effectiveness of interactives is constantly being evaluated and proven in galleries. For example, the study “Interactive and Visitor Learning” [16] conducted in Australia (2001) illustrated that interactives contributed both short-term and long-term impacts to visitors, whether in a museum or a science centre. Later, the study undertaken in the Churchill Museum (2005) indicated that multimedia and interactive featured exhibitions encourage visitors to learn in different ways [13].

4 System Implementation

4.1 uPoster: Ladies, People Are Coming

Hardware configuration. The uPoster employs the concept of context-aware. Two major parts of the hardware are the multimedia display module and the web camera sensing module. The web camera is used to detect the location of visitors; then, the multimedia workstation is triggered to response the different location of visitors with video content displayed on a 42-inch screen.

Content scenario. The T'ang Dynasty painting “A Palace Concert” shows ten ladies from the inner court, sitting around a large table, sharing tea, or playing music. It seems a merry palace concert. Therefore, the scenario is designed to make people aware of the content of the painting. While there is on one around, the ladies will be animated. On the contrary, ladies will be static as in the normal painting when people walk into the detecting area. (Fig.1) The scenario is set out to cause the psychological effect of cognitive inconsistency, and then provokes people’ imagination towards the court life of the T’ang ladies.



Fig. 1. Scenario of the uPoster. (a) Ladies are animated while there is on one around. (b) Ladies are static while people walk into the detecting area of web camera. (c) Ladies are playing again.

4.2 *i-m-Top*: Livening Up Paintings on the Tabletop

Hardware configuration. The architecture of the *i-m-Top* system is shown in Figure2a. The major components of the installation are multi-resolution display module and multi-touch sensing module [7, 17].

Content scenario. For the reason that paintings cannot be displayed in galleries under long periods of exposure to lighting, the installation has offered the best solution to this concern. The user interface is designed as a handscroll of Chinese paintings with a magnifier in the centre and the paintings arranged inside (Fig.2b). The area of the magnifier is a spotlight reflected by the foveal projector, which provides higher resolution than the peripheral area. Paintings can be selected by the use of hand gestures, zooming in/out or moving to the magnifier for careful appreciation. Moreover, colours following the fingertips make users more aware of multi-touch on the tabletop (Fig.2c).

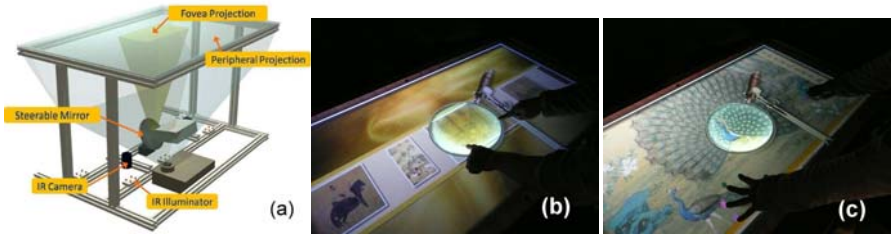


Fig. 2. (a) The architecture of the *i-m-Top*. (b) The handsroll-shaped menu of the paintings. (c) Multi-resolution display and multi-touch of colors following fingertips.

4.3 Interactive Virtual Display: Magic Crystal Ball

Hardware configuration. The architecture of the Magic Crystal Ball is shown in Figure 3a. There are two major modules of the construction: one is the display module to which the optical system of *i-ball2* is applied; the other is the detection module composed of an infrared camera and three pressure sensors [10].

Content scenario. It is said that crystal balls contain incredible power to make prophecies, tell fortunes or see into the past. With this fantastic nature, the stories of five artifacts from the National Palace Museum are told in the Magic Crystal Ball. The installation is designed as a small booth with a mysterious atmosphere created by special lighting (Fig. 3b). While an artefact is selected from the display panel, a virtual 3D image appears inside the crystal ball. Every finger motion on the crystal ball will be computed and translated into changes in the viewing direction. As a result, it enables users to feel as if they are acting with magic power (Fig. 3c).

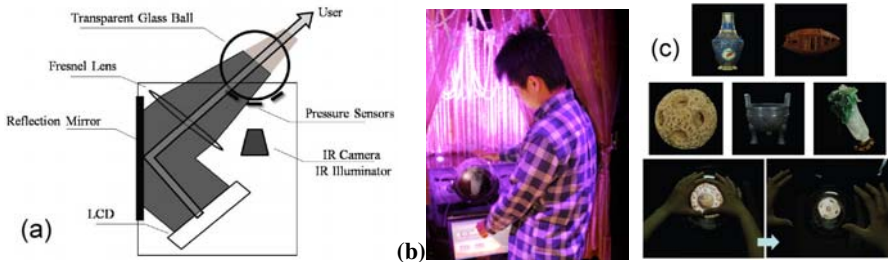


Fig. 3. Magic Crystal Ball. (a) The architecture of the Magic Crystal Ball system. (b) The layout of the installation (c) Five artifacts and a user manipulates the 3D virtual image.

4.4 Virtual Panel: Animals of Fantasy

Hardware configuration. First, in the display module, a real image forms in the air. Then, in the finger-touch detection module, an IR line illuminator is set out to create an IR plane aligned with the virtual panel. The architecture is shown in Fig. 4a [11].

Content scenario. Inspired by the 2009 Chinese New Year of the Ox, the scenario is designed as a game *Animals of Fantasy*. The ox is selected from an ancient Chinese painting as the main character, accompanying other animals. It is a guessing game. At the beginning, visitors must concentrate on the location of the ox. Then, the tree cups will cover over the three animals, randomly moving around in the surface. Once the cups stop, visitors must point out where the ox is in the air. Finally, a mark will be rewarded to visitors as feedback (Fig. 4b).

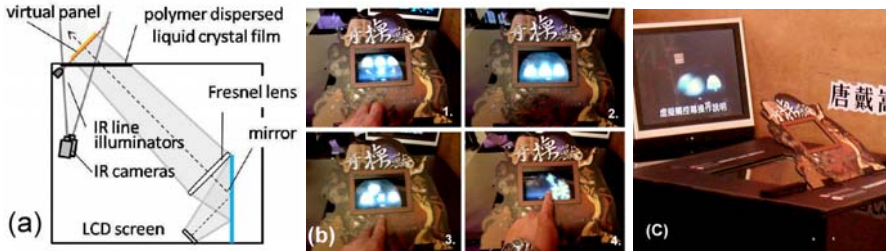


Fig. 4. (a) The architecture of the Virtual Panel (b) User interactions of guessing where the ox is. (c) The side of the installation and an instructional film standing in front of the frame.

5 Evaluation

The evaluation objectives are set out to explore the effectiveness of each installation against design principles. Mini-survey, semi-structured interview and observation were employed to generate quantitative and qualitative data. The executive findings are summarized in Table 1. If the accessibility is described with absolute statistics, the other three indicators are described with relative statistics, which can be summed up to 100%. The reason for that is that it is easy to compare the merits and shortcomings between each installation. Then, according to the performance ranking in each indicator, the overall effectiveness of installations is ranked from 1 to 4 (1= the best).

Table 1. Executive findings of the evaluation

Indicators	Ladies, people are coming	Livening up Paintings on the Tabletop	Magic Crystal Ball (MaC Ball)	Animals of Fantasy
Novelty Ranking	12.4% 4	36.0% 1	34.8% 2	16.9% 3
Accessibility Ranking	Mean=4.05 3	Mean=4.39 1	Mean=4.13 2	Mean=3.97 4
Reliability Ranking	13.0% 3	47.8% 1	29.3% 2	9.8% 4
Learning Potential Ranking	17.4% 3	30.4% 2	49.3% 1	2.9% 4
Overall Effectiveness Ranking	3	1	2	4

Novelty. Visitors were asked which installation surprised them most in the exhibition context, both around 35% of respondents indicated that *the Tabletop* and *Magic Crystal Ball*. One respondent commented that it is a very fresh thought to integrate ancient artifacts with high technology, which creates a new sensational experience.

Accessibility. Visitors were asked to rate the interpretative devices on a score from 1-5 (5 standing for the highest score) according to their experience when operating each installation. The tabletop had the highest mean score of 4.39. Visitors commented that the tabletop enabled the users to feel authentic interactions with real objects without a mouse or keyboard. On the contrary, *Animals of Fantasy* reached a rather low mean score. In observation, the majority of visitors required guidance from instructors. The likely reason for this is that the user interface is lacks of an appropriate metaphor of human thinking, resulting in many users not knowing how to operate it.

Reliability. To indicate the reliability, the quality of services was evaluated from end users' perspectives and stability of computer systems. It is significant that 47.8% of respondents indicated the tabletop system providing the highest quality of service in operation. Most of them commented that high-resolution displays gave them very clear images of paintings rather than unclear images in dim galleries. On the other hand, according to maintenance record of the systems, the *Magic Crystal Ball* was the most stable system that it never crashed during the exhibition period.

Learning Potential. Visitors were asked which object in the installations is the most memorable. The main reason is that visitors will recall the information, which had entered their Short-Term Memory while operating the installations. Significantly, nearly 50% of respondents answered that the objects shown in the *MaC Ball* were the most memorable. The likely reason for this is that artefact descriptions allow visitors to acquire more concrete knowledge. Compared with the tabletop, its learning potential is second to the *MaC Ball* at around 30%. There is sufficient evidence to prove that learners preferred different learning styles associated with different interpretative devices.

6 Conclusion

The paper presents how the team developed interpretive installations to accommodate different collections and learning styles via innovative HCI technologies. The pilot project has provided invaluable experience to museum professionals. First, the most effective application of the tabletop would be implemented as interpretative devices in galleries, information kiosks or educational programs in public area. Second, the *Magic Crystal Ball* can contain more high-quality images and develop more functions, e.g. magnify. Third, the *uPoster* is simple, economical and easy to be adjusted into any scenario or theme setting. Finally, as the virtual panel is too innovative to users; it requires more user studies in the future. Also, the project led to the conclusion that interpretative installations of HCI indeed facilitated meaningful interactions between visitors and museum collections. In the future, we would like to explore more interpretative multimedia based on what has been achieved and the barriers encountered. The ultimate goal is to create a positive and joyful museum learning environment, inspire creativity of the society, and play a leading role in the museums in Taiwan.

Acknowledgements

We are indebted to the Image and Vision Laboratory at National Taiwan University, coordinator Mr. Yi-Jin Gou and the team from the Department of Education Promotion, National Palace Museum, EeCom Multimedia Co. Ltd, EeRise Co. Ltd., and Digital Media Concept Co. Ltd.

References

1. Hooper-Greenhill, E.: *Museum and Education: Purpose, Pedagogy, Performance*. Routledge, Oxon (2007)
2. Muse Awards, <http://www.mediaandtechnology.org/muse/criteria.html>
3. ACM SIGCHI Curricula for Human-Computer Interaction, http://sigchi.org/cdg/cdg2.html#2_1 (accessed March 17, 2009)
4. Schilit, B., Theimer, M.: Disseminating Active Map Information to Mobile Hosts. *IEEE Network* 8(5), 22–32 (1994)
5. Dey, A.K., Abowed, G.D.: Towards a better understanding of context and context-awareness. In: *The Workshop on The What, Who, Where, When, and How of Context-Awareness*, as part of the 2000 Conference on Human Factors in Computing Systems (CHI 2000), The Hague, The Netherlands (2000)
6. Geller, T.: Interactive Tabletop Exhibits in Museums and Galleries. *IEEE Computer Graphics and Applications* 26(5), 6–11 (2006)
7. Hu, T.T., Chia, Y.W., Chan, L.W., Hung, Y.P., Hsu, J.: i-m-Top: An Interactive Multi-Resolution Tabletop System Accommodating to Multi-Resolution Human Vision. In: *IEEE International Workshop on Tabletops and Interactive Surfaces*, Netherlands (2008)
8. Ikeda, H., Naemura, T., Harashima, H., Ishikawa, J.: i-ball: Interactive Information Display like a Crystal Ball. In: *Conference Abstract and Applications of SIGGRAPH 2001*, p. 122 (2001)
9. Ikeda, H., Harashima, H., Ishikawa, J.: I-ball2: An interaction Platform with a crystal-ball-like display for multiple users. In: *International Conference on Artificial Reality and Teleexistence* (2003)
10. Chan, L.W., Chuang, Y.F., Yu, M.C., Chao, Y.L., Lee, M.S., Hung, Y.P., Hsu, J.: Gesture-based Interaction for a Magic Crystal Ball. In: *14th International Conference on ACM Virtual Reality Software and Technology (VRST)*, California (November 2007)
11. Chan, L.W., Hu, T.T., Hung, Y.P., Hsu, J.: On Top of Tabletop: a Virtual Touch Panel Display. In: *IEEE International Workshop on Tabletops and Interactive Surfaces*, Amsterdam, Netherlands (2008)
12. Samis, P.: New Technologies as Part of a Comprehensive Interpretive Plan. In: Din, H., Hecht, P. (eds.) *The Digital Museum: A Think Guide*. American Association of Museums, Washington (2007)
13. McIntyre, M.H.: I am easily satisfied with the very best: Summative evaluation of the Churchill Museum London. Morris Hargreaves McIntyre Ltd., London (2005)
14. Natural Interaction, <http://www.naturalinteraction.org/> (accessed March 23, 2009)
15. Reliability-Wikipedia, <http://en.wikipedia.org/wiki/Reliability> (accessed March 17, 2009)
16. Dierking, L.D., Falk, J.H., et al.: Interactive and Visitor Learning. *Curator* 47(2), 171–198 (2004)
17. Chan, L.W., Chuang, Y.F., Chia, Y.W., Hung, Y.P., Hsu, J.: A new method for multi-finger detection using a regular diffuser. In: *HCI* (3), pp. 573–582 (2007)

JacksonBot – Design, Simulation and Optimal Control of an Action Painting Robot

Michael Raschke¹, Katja Mombaur^{1,2}, and Alexander Schubert¹

¹ Interdisciplinary Center for Scientific Computing, University of Heidelberg
Im Neuenheimer Feld 368, 69120 Heidelberg, Germany

michael.raschke@gmx.de, kmombaur@uni-hd.de,
alexander.schubert@iwr.uni-heidelberg.de

² LAAS-CNRS, Université de Toulouse
7 av du Colonel Roche, 31077 Toulouse, France

Abstract. We present the robotics platform JacksonBot which is capable to produce paintings inspired by the Action Painting style of Jackson Pollock. A dynamically moving robot arm splashes color from a container at the end effector on the canvas. The paintings produced by this platform rely on a combination of the algorithmic generation of robot arm motions with random effects of the splashing color. The robot can be considered as a complex and powerful tool to generate art works programmed by a user. Desired end effector motions can be prescribed either by mathematical functions, by point sequences or by data glove motions. We have evaluated the effect of different shapes of input motions on the resulting painting. In order to compute the robot joint trajectories necessary to move along a desired end effector path, we use an optimal control based approach to solve the inverse kinematics problem.

Keywords: Action painting, robotic tool to generate art works, algorithmic motion generation, dynamic motion.

1 Introduction

In this paper we present JacksonBot - an algorithmic robotics platform established to explore the interdisciplinary research field between Graphical Arts, Robotics, Dynamical Motions and Mathematics. The painting style of the robot is inspired by the Action Paintings of the famous American artist Jackson Pollock (1912 - 1956) who dripped and splashed color on the canvas.

The purpose of this platform is not to generate a completely autonomous art robot but rather to set up a complex tool that can be used and easily programmed by a human artist to generate action paintings. With respect to classical computer generated art, the introduction of the robotic system as a tool between computer and art work represents an interesting addition due to its ability to move in a dynamical fashion.

There is a strong relationship between arts and motions. While the motion itself is the main focus in performing arts such as dancing, it also plays an important supporting role in fine arts, since the resulting artwork heavily depends

on the artists motion during its creation. This is particularly true for Action Paintings.

The robot presented in this paper relies on a combination of different techniques which are connected in an modular software and hardware environment. The platform consists of a robotic manipulator arm which is installed upside down in its working environment and carries a color container. This is completed by a color supply system, different input devices and a computation and visualization environment. The generation of robot arm motions is based on algorithmic techniques. Different possibilities to enter desired end effector motions are provided, such as specifying the motion by mathematical functions in the Cartesian space, or by a sequence of points. Alternatively motions can be freely prescribed by the user via a data glove. The computation of robot joint trajectories necessary to move along any of these desired end effector paths follows an optimal control based approach to solve the inverse kinematics problem, using efficient numerical techniques. These precise algorithmic components of the approach are complemented by more arbitrary and stochastic effects such as speed-related tracking errors, elastic oscillations of the robot arm and of the color container, and unpredictable behavior of the splashing color. These random effects make the paintings of our robotics platform more human-like than classical robotics art.

We have used the robot platform to investigate the effect of different shapes and types of input motions on the resulting painting. This also allows us to explore the relationship between mathematics and art and to find mathematical criteria or input functions that lead to a visually pleasing output in terms of the resulting painting. This can serve as a basis to implement more autonomous art robots.

Several authors have worked in the interdisciplinary field linking robots, art and motions. One of the first to combine robot motions and art was the Swiss artist Jean Tinguely [1], the founder of the so-called kinematic art, who became particularly famous for his kinematic fountain installations in Basel and Paris. The Strandbeest art works of the Dutch artist Theo Jansen [2] are immense mechanisms that move efficiently along beaches, only powered by wind energy, in a very similar way as passive-dynamic walking robots. During the last decade there has been a variety of robots producing some kind of visual arts. Several robots such as HOAP at EPFL [3] and [4] in Karlsruhe are capable to autonomously draw realistic portraits of people using on their vision systems. There are other robots that generate paintings with independently chosen styles and colors [5]. The annual exhibition "Artbots Robot Talent Show" [6] features many different art robots, e.g. the pneumatic robot gossamer-1 which splashes color powered by sound signals, and has some resemblance to the robot presented in this paper. There also is some relationship to the computer-controlled rolling balls or moving pens that create the Algorithmic Art of Hébert [7].

The remainder of the paper is organized as follows. In section 2, we briefly present the robotics platform JacksonBot. Section 3 describes the formulation of the inverse kinematics problem to generate robot arm motions and its solution

via numerical optimal control. In section 4 we present three different paintings generated by JacksonBot. In section 5 we finally give some conclusions about past and future work.

2 The Robotics Platform

In this section, we briefly describe the hardware and software environment of the art robot platform JacksonBot which can be considered as a complex and powerful tool to generate art works guided by a user. It is a first test platform that has been designed with a very small budget in order to validate the concept. But since modularity and flexibility were important guidelines during the design process it will be easily possible for us in future work to replace e.g. the simple manipulator arm by a more sophisticated mobile robot with feedback, or to add other ways of desired motion input.

As shown in figure 1, the current platform uses a six DOF (degree of freedom) robot manipulator arm which is installed upside down on the ceiling of its protected working environment. The canvas is positioned on the floor below the robot. Every joint of the robot is driven by a servo motor and can be independently controlled by prescribing its angular history. Potentiometers were attached to the four main axes which allows to record executed trajectories and analyze tracking errors. An interface module connects the servo motors and the potentiometers to the PC via micro controllers. A small color container is attached to the end effector of the robot with the possibility to swing freely. A small pump guarantees a continuous refill of the color in the container through a flexible polymer hose.

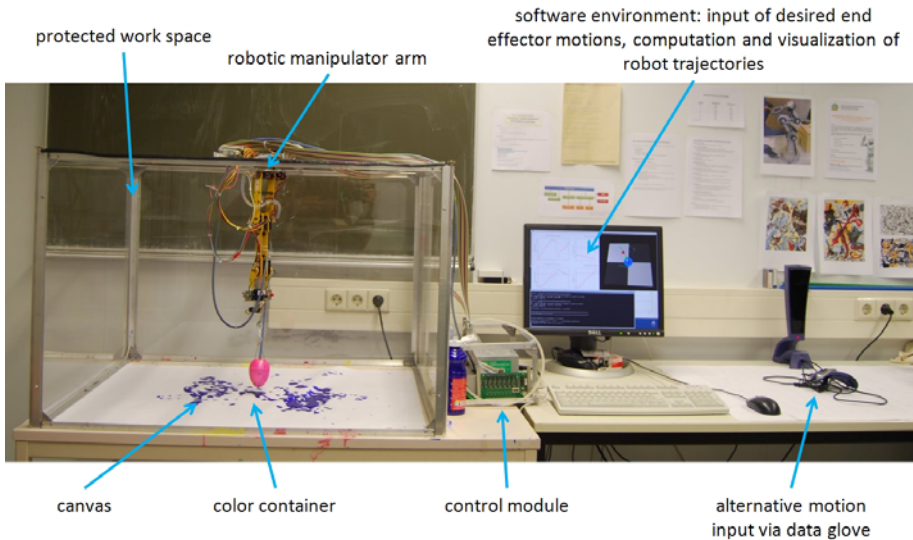


Fig. 1. Hardware and software environment of the JacksonBot art robot platform

Desired end effector motions can be prescribed either by mathematical functions, by point sequences or by data glove motions. A library of mathematical plot functions supports the user during the generation of desired trajectories. From simple shapes like lines, circles and other geometrical standard forms to very general curves, all types of trajectories can be programmed and added to the library. It is also possible to request the end effector to run through a defined sequence of points at given times. In addition, a P5 Data Glove allows a direct transfer of human hand movements to the robot arm which makes it possible to compare the mathematical descriptions of trajectories with free human-like motions.

Once a desired end effector motion is specified, it has to be translated to trajectories of the individual robot joints by solving an inverse kinematics problem. This issue is handled automatically on the PC via efficient numerical techniques, and we will describe this part in more detail in the next section. Once the resulting joint motions are determined, they can be prescribed to the servos via the micro controllers. In an OpenGL visualization (see figure 3) all desired end effector motions and the resulting computed motions of the robot arm can be displayed on the screen prior to execution on the real robot.

3 Optimal Control Based Solution of Inverse Kinematics to Generate Robot Motions

The inverse kinematics problem for a robot manipulator arm consists in determining the combination of joint angles that results in a prescribed end effector position¹. Depending on the DOF of the robot arm and the requested situation this inverse kinematics problem may have a unique solution, multiple solutions or no solution. The problem of moving the end effector along a prescribed trajectory could be solved by studying the inverse kinematics independently for a sequence of positions along this trajectory.

However, we choose a different approach based on the solution of an optimal control problem which handles the full trajectory at once instead of pointwise. This has the following two advantages:

- In the case of redundant solutions for individual time points it avoids jumps between different families of solutions from one time point to the next, since appropriate measures guarantee that the overall joint trajectories are as smooth as possible and that the change in joint angle per time is as small as possible.
- In the case of no existing solution of the inverse kinematics problem (since the requested positions are outside the work space) it is still possible to compute the best possible solution within the workspace since the cost function requires a best possible approximation, and not a zero distance.

¹ Note that in some cases it also may be interesting to additionally prescribe end effector orientations. In the case of the JacksonBot platform with its freely swinging color container it however only makes sense to prescribe end effector position histories and no orientations (which might be different e.g. for a brush or a pencil).

We do not need to explicitly formulate the inverse kinematics equations, but can use a model of the forward kinematics of the robot end effector positions

$$z_E(t) = f(\phi(t), p), \quad z_E \in \mathbb{R}. \tag{1}$$

as a function of its joint angles ϕ and the robot parameters p such as segment lengths. The function $f(\cdot)$ essentially describes the sequential rotations and translations induced along the kinematic chain up to the end effector.

We consider the angular velocities as input or control functions u_i of the model which are linked to the joint angles ϕ_i - the state variables of the robot arm - by the simple differential equations

$$\dot{\phi}_i = u_i \quad , i = 1, \dots, 6. \tag{2}$$

As optimization cost function, we use a minimization of deviation from the desired reference end effector positions, evaluated at m discrete points. We add a regularizing term to punish sudden changes of the joint angles, i.e. large values of the angular velocities. This results in the cost function

$$\min_{u, \phi} \Psi = \sum_i^m (z_E(\phi(t_i), p) - z_{E,ref}(t_i))^2 + \gamma \int_0^T u^T u dt. \tag{3}$$

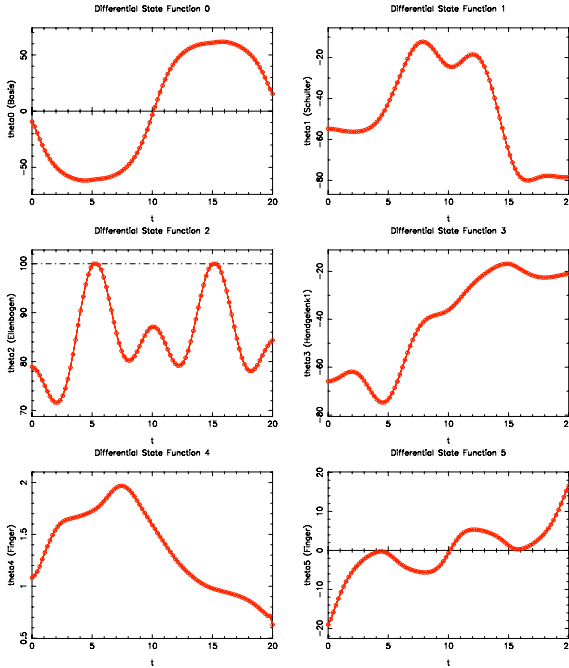


Fig. 2. Joint angle histories computed to follow the figure eight base function

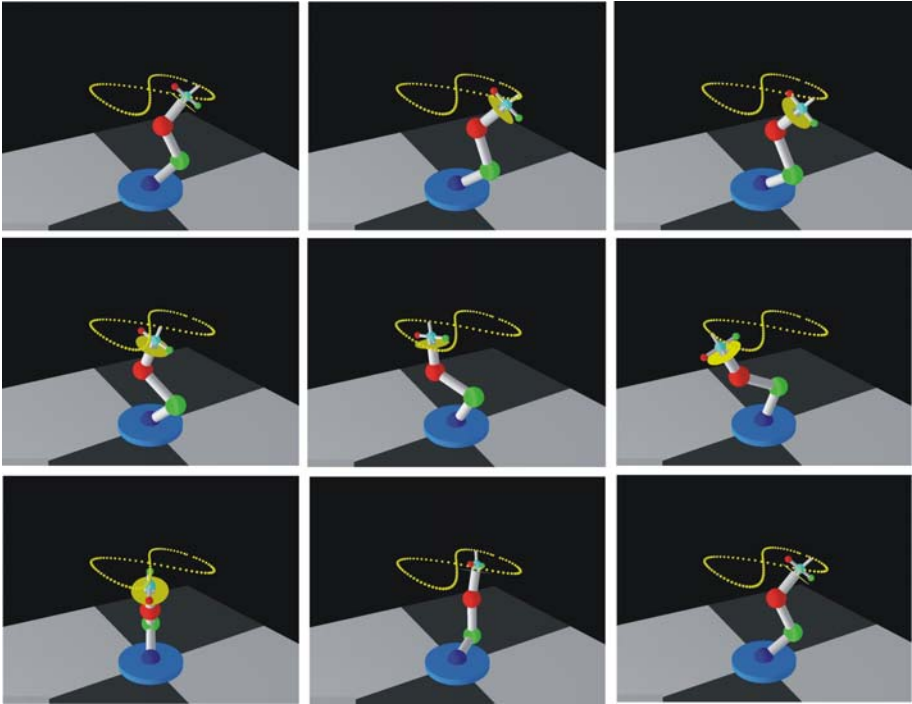


Fig. 3. Visualization of the resulting robot arm motion corresponding to figure 2

We solve this problem using the optimization techniques implemented in the optimal control code MUSCOD developed at IWR in Heidelberg [89]. The MUSCOD core algorithm is based on a direct method for the solution of the optimal control problem, i.e. a discretization of control variables, and a multiple shooting state parameterization. The technique which allows a fast and efficient solution of the described problems, is integrated to be called autonomously within the JacksonBot software environment.

In order to illustrate this approach, we briefly show results of an example elementary robot motion along a desired trajectory. In this case, the task is to move along a figure eight with a given length and width. This is one of the elementary plot functions contained in the mathematical library. Figure 2 shows the optimization results for all individual joint motions. In figure 3 the OpenGL visualization of the resulting robot and end effector motion is shown.

4 Example Paintings Generated by JacksonBot

In this section, we show three examples of paintings generated by the robot. Paintings are usually produced by a sequence of different motions using different colors and different strategies for defining motions.

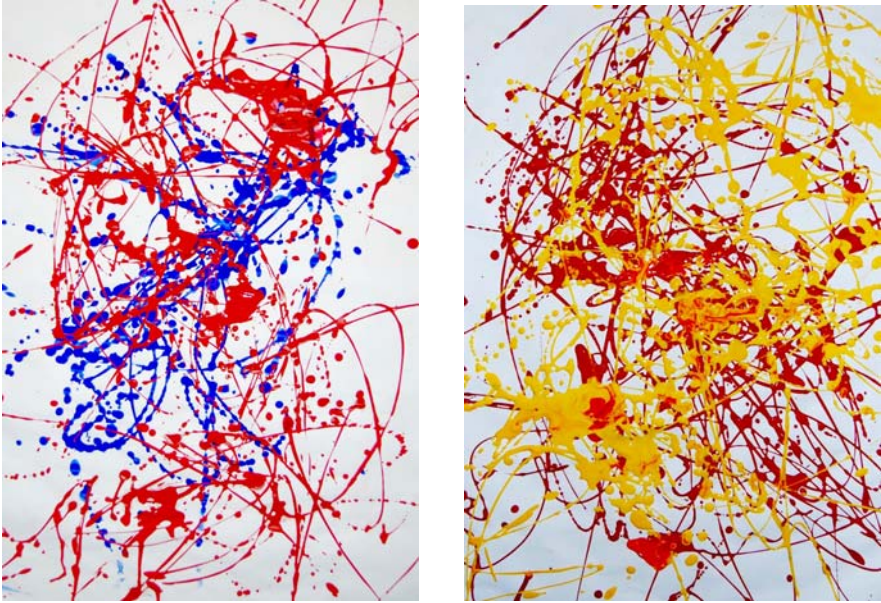


Fig. 4. Two paintings generated by dynamic motions passing through series of points

Both paintings in figure 4 were generated by specifying sequences of points to be passed at given time points. They represent a combination of very dynamic fast movements and slower ones. The dynamic movements produce more continuous lines, the slower ones leave behind more pronounced drops of color. Since the mechanics of the current robotics arm do not allow to equally access all areas, the canvas was manually rotated once during the creation of the painting.

In figure 5 the red circle was generated by a circle plot function. In the middle of the picture a blue eight can be identified. The galaxy-like shape on the left

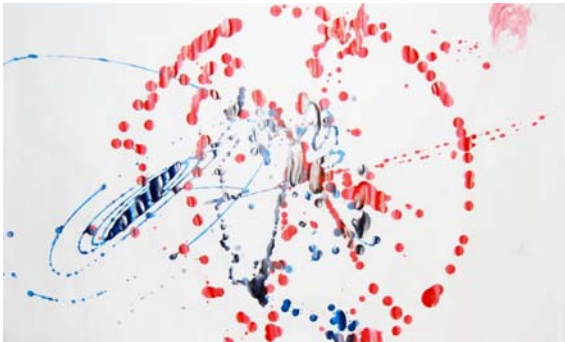


Fig. 5. Painting produced based on elementary mathematical plot functions and visible random effects

side was caused by a rapid movement of the end effector to this position over the canvas with a sudden stop. Because of the sudden stop the color container started to swing in a chaotic fashion. This is a good example for a combination of a determined, algorithmic with a chaotic and not precomputed, but random movement.

5 Conclusion

The presented robot platform JacksonBot is capable to automatically generate action paintings requiring only a few input functions or parameters. In its current state it can therefore be considered as a complex tool for the generation of art under the direction of a human user. It is possible to extend this approach and to continue from this *automatic* generation of art towards a truly *autonomous* generation of art by a robot alone without any interaction of a human. In order to do this it will be necessary to better understand the process of art generation as well as art evaluation by humans and to incorporate this knowledge in a model to be used by the robot. We have studied several different input modes and functions describing the motions but will continue this study further including evaluations of the resulting paintings by test persons. Based on these findings the robot would then be capable to autonomously select the most appropriate motions. In addition we plan to exchange the robot arm by a more sophisticated robot hardware such as a small humanoid robot that is capable to execute faster motions in a larger workspace and that also is equipped with visual sensing allowing an autonomous visual evaluation of the painting during the creation.

References

1. Wikipedia: Jean Tinguely (2009), http://en.wikipedia.org/wiki/Jean_Tinguely
2. Theo Jansen: Strandbeest (2009), <http://www.strandbeest.com>
3. Calinon, S., Epiney, J., Billard, A.: A Humanoid Robot Drawing Human Portraits. In: IEEE-RAS International Conference on Humanoid Robots (2005)
4. Gommel, M., Haitz, M., Zappe, J.: Autoportrait- Portaitzeichnungen aus der Hand eines Roboters (2002), <http://www.robotlab.de>
5. Van Arman, P.: Zanelle, the painting art robot, www.zanelle.com
6. ArtBots: The Robot Talent Show, <http://www.artbots.org>
7. Hébert, J.-P.: Mac-controlled Art (2009), <http://www.apple.com/science/profiles/hebert/>
8. Bock, H.G., Plitt, K.J.: A multiple shooting algorithm for direct solution of optimal control problems. In: Proceedings 9th IFAC World Congress Budapest, pp. 243–247. Pergamon Press, Oxford (1984)
9. Leineweber, D.B., Bauer, I., Bock, H.G., Schlöder, J.P.: An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization - part I: theoretical aspects. *Comput. Chem. Engng.* 27, 157–166 (2003)

RF Sounding: A System for Generating Sounds from Spectral Analysis

Fabio Graziosi, Claudia Rinaldi, and Francesco Tarquini

Center of Excellence DEWS,

Faculty of Engineering, University of L'Aquila, 67040 Monteluco di Roio, L'Aquila, IT
{fabio.graziosi,claudia.rinaldi,francesco.tarquini}@univaq.it

Abstract. In this paper we present *RF Sounding*, an open space installation which comprises both artistic and technological innovations. The aim of this project is to provide the user entering a specifically defined area, with awareness of radio frequency signals characterizing the cellular networks band. Indeed, radio signals are shifted, with proper elaboration, to the audible band and the result is spread all over the specific area through a certain number of loudspeakers. The system produces different reactions depending on the communication phase (e.g. initial handshake procedure, reception or initiation of a call, etc.). Moreover, the sound produced after translation of signals to the audible band, is assumed to be spatialized as a function of user movement; localization is indeed achieved through a wireless sensor network that is installed in the defined area.

Keywords: Radio-frequency propagation, audio signal processing, wireless sensor network, localization, sound spatialization.

1 Introduction and Motivations

The opportunity to integrate technology and arts represents one of the most stimulating and creative research areas. The contamination between scientific knowledge and artistic components is too often limited to casuality, mainly stemming from the convergence on the same researcher of technical and artistic interests. In this scenario we are mainly interested on the integration of wireless communications and audio wave propagation, being both characterized by the same propagation medium, even if with substantially different propagation modes. More in particular, in this paper we want to highlight the limitations of our senses which are capable of feeling audio waves (sounds) but are not capable of directly feeling radio frequency (RF) waves. We propose to overcome this limitation by translating RF waves to the audible band. This translation can be exploited as an interesting opportunity for modern music composers which will be called to introduce a proper signal processing on revealed RF signals. Some features of RF sensed signals will be kept unchanged in order to give to the listener the clear feeling of the amount (power) of RF signals present in the monitored area. The IT technology has evolved in furnishing the most advanced services in terms of communication [1,2,3], monitoring [4,5,6], and security [7].

Our project aims in particular to investigate the possibility to exploit wireless sensor networks (WSNs) in artistic applications, with particular efforts on electronic music. In order to properly design the RF sounding the main topics to be investigated are: WSNs with main attention on localization, spatialization and sound elaboration, cellular networks communication.

Localization in the context of WSNs is a well known topic in the IT research sphere, [4,5,11,12,13,14], but the problem of localizing a passive target through small sensors is still a very open research area and it appear of fundamental importance for our project since it is thought for end users that can use the installation with their cellular phones only.

On the other hand a lot of possibilities have been proposed since the birth of the computer music for sound spatialization and elaboration, [15,16,17,18] and it appears that most of them are suitable to be realized and improved in our proposed prototype.

The aim of this project is twofold. Indeed, from one side we want to increase end users knowledge of the strength of the power emitted by their cellular phones with respect to the electromagnetic fields produced in the environment, on the other hand we want to provide for an artistic and interactive installation that can also be remotely joined through a web interface.

The paper is organized as follows: the system architecture together with its general components is presented in section 2, section 3 is concerned with the specific functioning of RF Sounding, while the localization aspect is detailed in section 4; techniques for sound processing and spatializations are presented in section 5; finally conclusions and future works are drawn in section 7.

2 System Architecture

The general project for RF Sounding, fig. 1, assumes a scenario comprising a circular area with a 10m diameter. The area is accessible through at least 2 entrances equipped with gating sensors. Within the area an hexagon will be defined on whose vertex six speakers will be placed. Along with the speakers there will be six or more wireless sensor nodes that will feed a positioning algorithm allowing to evaluate the user position and movement in the equipped area. The positioning algorithm will be based on [23] where the body of a person in the monitored area is reviewed as on obstacle to the radio transmissions among sensors. There will be one node sending a low power radio signal and all other nodes will evaluate the strength of the received signal. A signal processing will follow to determine the user position. These sensors will allow the installation to interact with user's movements by changing lights and sounds conditions. The system interaction with the user will be set according to suitable psychoacoustics evaluations. In the center of the hexagon, at a level of 2.5-3m from the ground, a receiving antenna is placed in order to gather all signals in the band of interest and to send them to a spectrum analyzer. The analyzer will be linked to an elaboration unit, equipped with an audio processing board, that will implement sound's elaboration and spatialization algorithms. This unit will also handle the

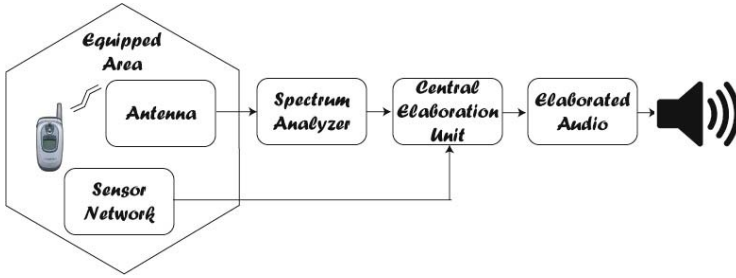


Fig. 1. RF Sounding: general functioning

processing of the localization data obtained from sensors in order to produce a suitable sound spatialization. In order to obtain a real interaction with the user and the system, we plan to implement a web interface consistent with WEB 2.0 technology. This web interface will allow the user to take control of his/her experience in the equipped area, for example using an handled device on site. The remote interface makes possible to start and stop the experience and also to ask the system for a specific response, such as sounds and lights variations, on the basis of the received signal and user position. Any time and anywhere, through the web interface, it will be possible to hear and visualize the signal received by the spectrum analyzer and the system status and to take control of the system.

The architecture of the system is shown in figures 2, 3. The RF Sounding project is characterized by a double aim. Firstly we want users to become aware of the spectral occupancy of phenomena arising while using the services of a certain cellular network and then we want to exploit an aesthetic elaboration of these spectral phenomena in order to allow the user to listen to a likable final sounding result.

The starting point of this project is given by the design and validation of a prototype exploiting signals coming from the GSM public network, 1, 2. We

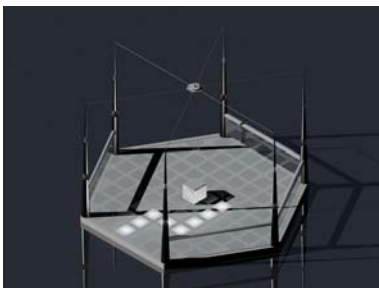


Fig. 2. Axonometric view of the equipped area

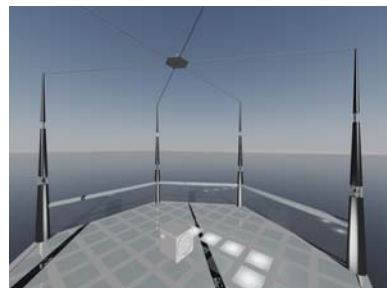


Fig. 3. Subwoofer placed at the center of the hexagon

choose the GSM standard since it is still one of the most diffused all around the world and this represents an advantage in terms of usability of the installation. GSM uses a time-division/frequency-division multiple access technique (TDMA/FDMA). This means that many radio channels are shared by many users. In particular each channel is characterized by a 200Hz bandwidth and is responsible for the transmission of data belonging to 8 users. It has to be noticed that different frequencies are used in different countries, indeed 900 and 1800 MHz are the frequency bands used in Europe, while United States use 850 and 1900 MHz frequency bands. Obviously, we are not going to consider each GSM version, at least in this first phase of RF sounding project. We are indeed mainly interested in exploiting the multiple access technique: the FDMA part involves the division by frequency of the (maximum) 25 MHz bandwidth into 124 carrier frequencies spaced 200 kHz apart. One or more carrier frequencies are assigned to each base station. Each of these carrier frequencies is then divided in time slots, using a TDMA scheme. Eight time slots are grouped into a TDMA frame, which forms the basic unit for the definition of logical channels.

On the basis of this brief information it is easy to believe that a user getting in the equipped area may experiment different reactions of the installation. Actually, the mobile terminal makes operations even without the intervention of its owner. For instance this happens when the mobile terminal (MT) is switched on and it starts scanning all radio channels in the service band in order to find out the beacon frequencies. The term beacon frequencies indicates the carrier frequencies continuously emitted by each cell at a constant power. It is worth noting that most control channels are allocated over the beacon frequencies. As a result of the scanning procedure, the terminal creates a list where frequencies are ordered basing on strength and constancy of the received power and it establishes to connect to the cell characterized by the strongest power.

Once the carrier frequency is known, the local oscillator (LO) of the mobile terminal is tuned to this frequency through a phase locked loop operating on a periodically emitted Frequency Burst (FB) [8] carried by the Frequency Correction Channel (FCCH) of the Broadcast Control Channel (BCCH) of the GSM signaling frame structure. The information traveling through the FCCH is transmitted at the maximum power since it has to be received by all terminals. This procedure takes place even in absence of mobile terminals and this aspect can be exploited in the installation, after a proper elaboration, as a persistent basic element. This basilar aspect represents the starting point of the real performance.

Other interesting procedures to be exploited in the installation performance are the reception of a call, the initiation of a call, the switch-off of the terminal and so on.

Moreover, a wireless sensor network is placed around the equipped area in order to follow the movements of the user. The sound will then be "moved" in the area as a function of user movements. The overall result will thus be a function of both user required services or MT procedures and user movements.

3 Localization

The user localization within the system area is carried out in a passive way: the user is not equipped with any active device collaborating with the localization system or aware of its location. We approach to the localization problem by the technique known as LBE (learning by example) [11]. Basically we customize and apply the multistatic radar concept, [4], [5] to the wireless sensor networks. Our localization algorithm is made up by two main phases: the first one (calibration), made offline, basically is made up by some measurement of the RSSI without any passive target, while the second one consists in the evaluation of the perturbations in the RSSI values, introduced by the target in the equipped area. In the user localization we introduce spatial operators. Spatial operators are used to capture all the relevant geometric properties of objects embedded in the physical space and the relations between them as well as to perform spatial analysis. More in particular we use topological operators [21]: intersection test and point in polygon test, to perform a qualitative analysis of user position within the area. This is made possible by the partition of the base, or floor, of the equipped area into zones identified by reinforced concrete and glass tiles. Thus, once we get the supposed user position by RSSI measurements, we test the *intersection* of this position with floor tiles identifying the one with non-empty intersection and we can light this tile up. Due to the fact that the presumed user position is a broad boundary region we are forced to use spatial operator defined for Broad Boundaries objects, [22].

4 Spatialization and Sound Processing

Sound spatialization is essentially related to the movement of sound through space. Techniques for sound spatialization are different if the target display is by means of headphones or loudspeakers, [23]. In this context we are mainly concerned with spatialization through loudspeakers in the installation area, while spatialization through headphones is achieved in the web remote application.

The most popular and easy way to spatialize sounds using loudspeakers is amplitude panning. This approach can be expressed in matrix form for an arbitrary number of loudspeakers located at any azimuth though nearly equidistant from the listener. Such formulation is called Vector Base Amplitude Panning (VBAP) [15] and is based on a vector representation of positions in a cartesian plane having its center in the position of the listener. In figure 4 the case for two loudspeakers is represented while the generalization to more than two loudspeakers and to three dimensions can be simply achieved.

A different approach to spatialization using loudspeakers can be taken by controlling the relative time delay between the loudspeaker feeds. The metaphor underlying the Moore model, [17], is that of the Room within a Room, where the inner room has holes in the walls, corresponding to the positions of loudspeakers, and the outer room is the virtual room where sound events have to take place, [5].

The simplest form of spatialization is obtained by drawing direct sound rays from the virtual sound source to the holes of the inner room. The loudspeakers

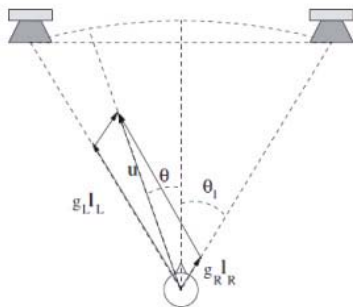


Fig. 4. Stereo panning

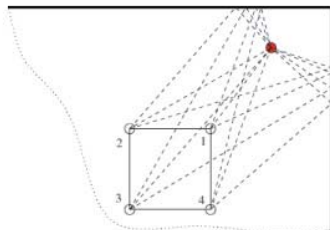


Fig. 5. Moore's room in a room model

will be fed by signals delayed by an amount proportional to the length of these paths, and attenuated according to relationship of inverse proportionality valid for propagation of spherical waves. The two controlled parameters are thus the gain of each loudspeakers and their relative delays.

Actually, while concerning with spatialization more than two parameters can be concurrently varied, [16] and we will control the following parameters: total power of signals emitted, number of loudspeakers emitting sound at the same time for each time interval, envelope characterizing the cross fade from one channel to another or between more than one channel per time, remaining power offset on loudspeakers not directly involved in sound movement.

For what concerns spatialization algorithms it has to be noticed that the proposed configuration allows almost an infinite number of variations. A first characterization could be done as a function of the user's speed; indeed a speed threshold will be established. If the user's speed will be below this threshold the loudspeakers-subwoofer systems will provide for a moving sound that will be characterized by a slow speed and by only circular spatialization, that is, the sound will move only between adjacent loudspeakers. Other mechanisms are instead assumed for the case of a fast user's movement:

- diagonal movement between single loudspeakers,
- sound motion between pairs of loudspeakers,
- varying power offset for loudspeaker not directly involved in the primary spatialization.

All these mechanisms can be properly automatized depending on user position as well as varied in a random fashion in order for the sounding experience to not be predictable.

In terms of sound processing a first distinction has to be done between sound synthesis and sound elaboration techniques. For what concerns sound synthesis techniques, we will base our elaborated sounds on shifts of RF signals to the audio frequency range but we will also process signals through the following type of synthesis: subtractive, additive, granular, wavetable, sample-based,

frequency modulation, phase distortion, [18]. On the other hand, for what concerns sound elaboration techniques a first fundamental variation will be given to sound amplitude as a function of antenna's received power. Subsequently a certain number of effects could be applied, depending on e.g. channel carrier frequency, the specific telecommunication operator furnishing a service for a certain user, the number of users inside the equipped area and so on. Some of these techniques are reported in [18], [24].

5 Conclusions and Future Works

In this paper an innovative project integrating technologies and experimental music has been proposed. The project focuses on the creation of an interactive installation with a double aim, to increase users awareness of the spectral occupancy in the cellular networks bands and to provide for a spectral phenomena aesthetic elaboration in order to produce a sounding experience. The very next step of our project will be the system test together with electronic music composers.

References

1. ETSI EN 302 403 V8.0.1 (2000-10), Digital cellular telecommunications system (Phase 2+); GSM Cordless Telephony System (CTS), Phase 1; Service description; Stage 1 (GSM 02.56 version 8.0.1 Release 1999)
2. ETSI EN 302 404 V8.0.1 (2000-09), Digital cellular telecommunications system (Phase 2+); GSM Cordless Telephony System (CTS), Phase 1; Lower Layers of the CTS Radio Interface; Stage 2 (GSM 03.52 version 8.0.1 Release 1999)
3. Universal Mobile Telecommunications System (UMTS); Network Functional Model for IMT2000 networks [ITU-T Recommendation Q.1711] (2000)
4. Bistatic: Multistatic Radar H. D. Griffiths University College London Dept. Electronic and Electrical Engineering Torrington Place, London WC1E 7JE, UK
5. Chernyak, V.S.: Fundamentals of Multisite Radar Systems: Multistatic Radars and Multistatic Radar Systems. CRC Press, Boca Raton (1998)
6. Alesii, R., Gargano, G., Graziosi, F., Pomante, L., Rinaldi, C.: WSN-based Audio Surveillance Systems. In: Proc. of WSEAS European Computing Conference 2007 (ECC 2007), Athens, Greece (September 2007)
7. Feller, S.D., Cull, E., Kowalski, D., Farlow, K., Burchett, J., Adelman, J., Lin, C., Brady, D.J.: Tracking and Imaging Humans on Heterogeneous Infrared Sensor Array for Law Enforcement Applications. Department of Electrical and Computer Engineering, Duke University, Durham, NC
8. Acquisition of frequency synchronization for GSM and its evolution systems. In: 2000 IEEE International Conference on Jha, U.S. Personal Wireless Communications, pp. 558–562 (2000)
9. Pahlavan, K., Li, X., Makela, J.P.: Indoor geolocation science and technology. IEEE Communication Magazine (2002)
10. Catovic, A., Sahinoglu, Z.: The Cramer-Rao bounds of hybrid TOA/RSS location estimation schemes. IEEE Communication Letters (2004)

11. Viani, F., Lizzi, L., Rocca, P., Benedetti, M., Donelli, M., Massa, A.: Object tracking through RSSI measurements in wireless sensor networks
12. Tennina, S., Di Renzo, M., Graziosi, F., Santucci, F.: Distributed and Cooperative Localization Algorithms for WSNs in GPS-less Environments. The Italian Institute of Navigation (I.I.N.), No. 188, pp. 70–76 (June/July 2008)
13. Tennina, S., et al.: Locating zigbee nodes using the TI's CC2431 location engine: a testbed platform and new solutions for positioning estimation of WSNs in dynamic indoor environments. In: ACM Int. Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (September 2008)
14. Tennina, S., Di Renzo, M., Santucci, F., Graziosi, F.: On The Positioning Estimation of WSNs in Dynamic Indoor Environments. In: NEWCOM++ Emerging Topic Workshop, Brixen, July 1 (2008)
15. Pulkki, V.: Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* 45(6), 456–466 (1997)
16. Stockhausen, K.: Musik im Raum, 1958, trad. it. Musica nello spazio, in *La Rassegna Musicale* 32(4) (1961)
17. Moore, F.R.: A General Model for Spatial Processing of Sounds. *Computer Music J.* 7(3), 6–15 (1982)
18. Dodge, C., Jerse, T.A.: *Computer Music: Synthesis, Composition, and Performance*. Schirmer (July 1997)
19. Clementini, E., Di Felice, P.: Spatial Operators. *ACM SIGMOD Record* 29, 31–38 (2000)
20. Clementini, E., Billen, R.: Modeling and computing ternary projective relations between regions. *IEEE Transactions on Knowledge and Data Engineering* 18, 799–814 (2006)
21. Clementini, E.: Dimension-Extended Topological Relationships. In: Liu, L., Özsu, M.T. (eds.) *Encyclopedia of Database Systems*, vol. à Parâitre. Springer, Berlin (2009)
22. Clementini, E.: Objects with Broad Boundaries. In: Shekhar, S., Xiong, H. (eds.) *Encyclopedia of Geographical Information Science*, pp. 793–799. Springer, Heidelberg (2008)
23. Rocchesso, D.: *Introduction to Sound Processing*, ebook, University of Verona, Departement of Computer Science (2003), <http://www.scienze.univr.it/?rocchess>
24. Proakis, J.G.: *Communication Systems Engineering*. Prentice Hall International Editions (1994)
25. Gardner, W.G.: *Transaural 3D audio*, MIT Media Laboratori Perceptual Computing Section Technical Report No. 342, July 20 (1995)

Depicting Time Evolving Flow with Illustrative Visualization Techniques

Wei-Hsien Hsu¹, Jianqiang Mei^{1,2}, Carlos D. Correa¹, and Kwan-Liu Ma¹

¹ University of California, Davis
One Shields Avenue, Davis CA 95616, USA
whhsu@ucdavis.edu

² Tianjin University, Tianjin
92, Weijin Road, Nankai District, Tianjin 300072, P.R. China

Abstract. Visualization has become an indispensable tool for scientists to extract knowledge from large amounts of data and convey that knowledge to others. Visualization may be exploratory or illustrative. Exploratory visualization generally provides multiple views of the data at different levels of abstraction and should be highly interactive, whereas illustrative visualization is often made offline at high quality with sufficient knowledge about the data and features of interest. Techniques used by professional illustrators may be borrowed to enhance the clarity and aesthetics of the visualization. This paper presents a set of visualization techniques for presenting the evolution of 3D flow. While the spatial features of the data is rendered in 3D space, the temporal behaviors of the flow are depicted using image-based methods. We demonstrate visualization results generated using three data sets obtained from simulations.

Keywords: Volume visualization, time-varying data visualization, image processing, evolution drawing, non-photorealistic rendering.

1 Introduction

Visualization transforms numbers into vivid pictures that capture the essence of the data generated in the study of a complex problem in physical sciences, social science, engineering, biology and medicine. Visualization if properly made can help a scientist to see what she expects to see, possibly see previously unknown phenomena, and communicate with others her findings more effectively. Advances in visualization over the years have been made possible largely due to the incorporation of concepts in art and visual perception into visualization design, rapid development of graphics hardware technology, and coupling of advanced display technology and interaction techniques into an overall visualization solution. A visualization system should support its users to make visualizations according to the purpose of visualization, be it for exploration or illustration.

This paper presents a set of techniques for crafting scientific visualization to be illustrative, expressive and aesthetic. We show how to achieve our goals by combining advanced rendering methods, image processing techniques, together with styles, compositions, and layouts commonly found in art. Skilled artists are

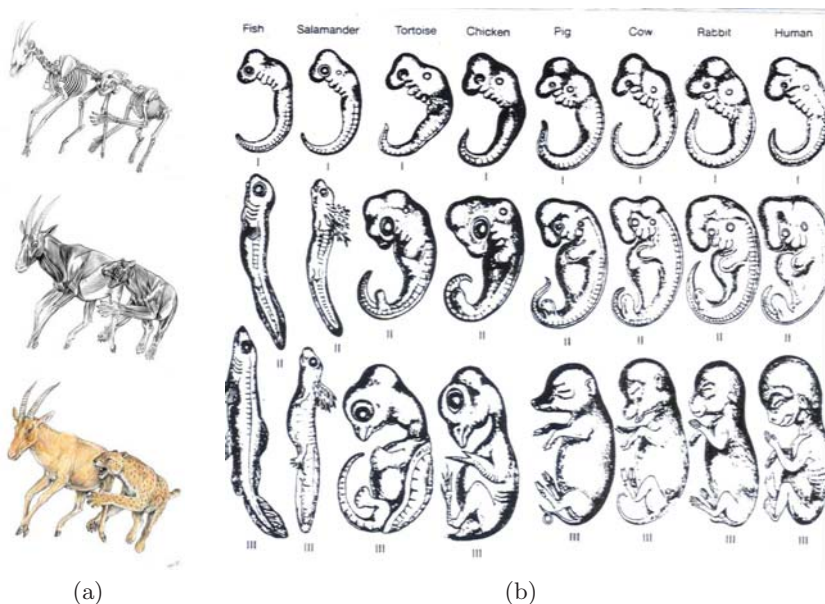


Fig. 1. Examples of professionally made illustrations (a) Depiction of *Metailurus* major based of the proper body proportions. Preying on the Miocene antelope, *Tragoportax*, by Zoologists and artist Velizar Simeonovski. (b) A comparison of vertebrate embryos (from *What Evolution Is* by E. Mayr [11]).

able to apply pigments using a brush or pen in a manner that creates meaningful abstractions of reality. On a piece of paper, they can unambiguously convey three-dimensional shape, express spatial relationship, and give a sense of movement. Figure 1(a) uses three pictures of the same pose to depict the anatomy of an extinct genus of false saber-toothed cats called *Metailurus*, illustrated by Zoologists and artist Velizar Simeonovski. Figure 1(b) shows a sequence of vertebrate embryo images for the comparison of different species at different time periods [11]. Like an artist, we can create visualization emphasizing features that might not otherwise be visible, while de-emphasizing those features of less importance. With proper artistic composition and arrangement, the resulting visualization is often more expressive and compelling. In this paper, we demonstrate techniques operating in either 3D volume data space or 2D image space to create desirable visualization of time-varying 3D flow fields.

2 Related Work

Time-varying volume data visualization leads to many unique challenges from data encoding to feature extraction and volume rendering. Ma has given a survey of early work on time-varying volume data visualization [10].

Researchers have been paying attention to the approaches to visually represent the time evolving phenomena in the physical space. Attempts have been made to

visualize time evolving data without animation. Inspired by the idea of chronophotography, Woodring and Shen [19] integrated a time-varying 3D dataset into a single 3D volume, called a chronovolume, using different integration functions. In [20], Woodring et al. considered time-varying dataset as a four-dimensional data field and applied high dimensional projection techniques to create an image hyperplane. Lu and Shen [9] introduced a method by first reducing time steps and only visualizing the essential information in a storyboard layout. Balabanian et al. [1] used temporal style transfer functions to visualize temporal characteristics into a single image.

Depicting time evolving data by focusing on a small number of features in the data is becoming increasingly popular especially when the data size is large. The feature-based approach requires multiple steps, including feature extraction, feature tracking, and the visual representation. Wang and Silver proposed several volume tracking methods for regular grid [15] and unstructured grid [14] by utilizing the spatial overlap among corresponding isosurfaces in consecutive time steps. Muelder and Ma [12] introduced a prediction-correction method to make the best guess of the feature region in the subsequent time step to achieve a better tracking result. Instead of explicitly tracking features, Lee and Shen [7] used Time Activity Curves to describe time-varying features and visualized the vector field of the motion tendency of the feature. To draw time-varying features, Post et al. [13] used icons as a symbolic representation of essential characteristics of features to visualize time-dependent data. Joshi and Rheingans [6] used the techniques inspired by conventional illustration such as speedlines and flow ribbons to depict the motion of the feature.

Non-photorealistic rendering (NPR) techniques as used in [6] have been shown great promise in conveying information. In contrast to the traditional photorealistic rendering, illustrative rendering aims to mimic various styles of human paintings and artistic illustrations to highlight hidden structures where a specific meaning has to be conveyed to viewers. Considerable work has been done in making stylish artistic images from both 2D images [5,18] and 3D objects [4,8]. NPR techniques have also been applied to volume visualization. Stompel and Ma [16] introduced a set of feature enhancement techniques based on NPR to generate more perceptually effective visualization of multivariate and time-varying volume data. Svakhine et al. [17] used a combination of illustrative and photorealistic rendering techniques to visualize flows and volumes.

3 Object-Based Approach

Our approach to 3D time-varying data visualization can be performed on either object data space, i.e. the 3D domain of the data, or in image-space, i.e., using the 2D images generated by 3D volume rendering. In this section, we describe our technique for illustrating time-evolving flow in 3D data space. To obtain a good 3D volume rendering of time-varying data, we follow a series of steps. First, we find a suitable transfer function that maps the data values to color and opacity. To follow the evolution of a feature, we employ a feature extraction and tracking method [12]. Finally, we render the extracted features using illustration-inspired techniques.

3.1 Transfer Function Specification

Transfer functions are used to map data values to rendering characteristics, such as opacity and color. Before assigning data values to color and opacity, we must first normalize the entire collection of time steps, since the data range for each time step might be different. This normalization is necessary to ensure that the regions highlighted with a transfer function are consistent across the different time steps. We then use a 1D transfer function based on the normalized scalar data value. Transfer functions let users select regions of interest and highlight intervals in the data domain, such as regions of high vorticity.

3.2 Feature Extraction and Tracking

Feature Extraction and Tracking(FET) lets us extract meaningful features and follow their evolution over time. The FET technique we used in this paper predicts each feature's region and then adjusts the surface of this region to extract the correct feature. It utilizes information from previous time steps to predict a feature's region in the current time step, then correcting the predicted region to extract the feature's actual boundaries via region growing and shrinking. Users can select an individual feature, and the corresponding features over time would be extracted and tracked. If no special features are selected, the system automatically extracts and tracks all the features.

3.3 Spatio-temporal Fade-in and Contour Rendering

Once we have extracted different features, we render them in a way to convey their evolution over time. To this purpose, we use fade-in effects. The oldest time step is depicted using only gray levels, while the newest time step is rendered with color. To convey a smooth transition, we use linear blending between these two extremes for intermediate time steps. The idea of this transition is to convey a natural direction of time, often found in art forms. Older instants of time of a moving object appear blurred whereas the current instant is crisp and clear. Since we want to preserve the rich detail of each time step, we choose to use less saturated colors to achieve the same effect. In addition, we compute contours based on the 3D depth of features. Contours are useful to simulate line drawing, useful when we want to preserve the overall shape of features with a minimal impact in the image. Combined with contour rendering, the motion of interesting features and their evolution can be conveyed and easily obtained by users. Furthermore, interactive tools to modify the color and opacity of each time-varying feature let users create any desired fading effect to emphasize earlier or more recent stages of evolving flow.

3.4 Results

Our technique has been implemented and tested on several data sets. We show here an argon bubble data set, which is $640 \times 256 \times 256$, and the turbulent vortex data set(used by D. Silver [15]), which is $128 \times 128 \times 128$ voxels. Both data sets

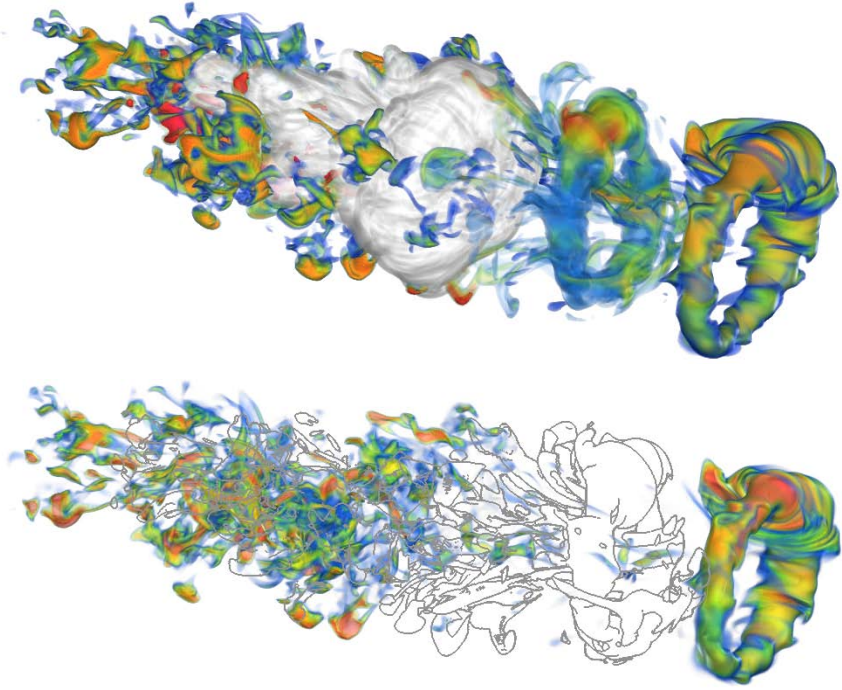


Fig. 2. Top: The separation process of the ring structure and its left-to-right motion is presented. There are three time steps in the image, the first time step is rendered in gray-color, while the last time step is full volume rendering, and the middle one is transparent color. Bottom: The contour line drawing depicts the exact position of the smoke ring in earlier time step more clearly.

are time-varying scalar volumes. For the bubble data set, the ring structure is what scientists are most interested in; while for the turbulent vortex data set, the features of interest we focus on are regions of high vorticity. All tests were run on a Dual-Core of a Mac Pro with a $2 \times 2.66\text{GHz}$ Intel Xeon processor and 6GB RAM.

Figure 2 is an example created with the argon bubble data set. This data set shows the result of the simulation of a shockwave interacting with a bubble of argon gas surrounded by air creating a swirling torus-shaped "smoke ring" along with smaller turbulence structures. By looking at how the bubble deforms, scientists can understand how the air shockwave interacts with the argon bubble and generates vorticity, in an effort to model hyperbolic conservation laws. As we can see in Figure 2(top), the separation process of the ring structure and the left-to-right motion of the interested region is presented. As time progresses, gray-scale color becomes more and more transparent, meanwhile, the chromatic color becomes more apparent. Finally, a full-colored volume rendering is shown to depict the last time step of the data. Because the different time steps overlap

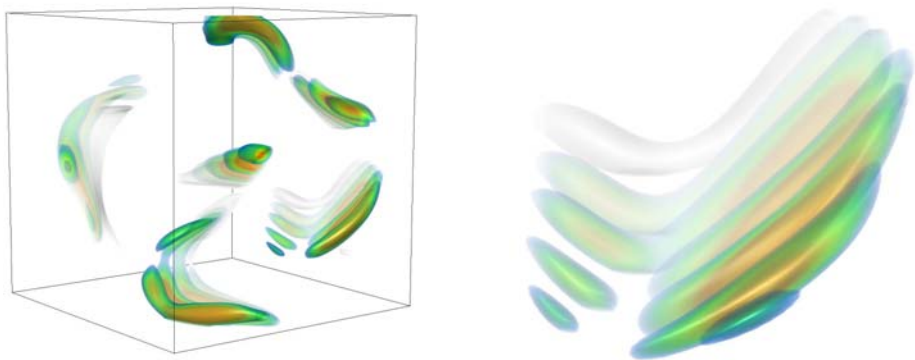


Fig. 3. (a) Evolution of five features in the original volume from vortex data set. Time-varying behavior of the volume could be obtained. (b) From birth to death, one single feature from vortex data set. As time progresses, more transfer function colors have been used to render the feature, and the evolution through its lifetime could be captured at all.

in 3D space, obtaining a clear view of the smoke ring may be difficult. Figure 2(bottom) uses a line drawing style for one of the time steps to show the exact position of the smoke ring in relation to the other stages.

Figure 3(a) and figure 3(b) are examples created with a Computational Fluid Dynamics(CFD) simulation with turbulent vortex structures. The vortex structures are line-type features with strongly curved, and tube-like shape. The relationship between different features at different time steps could be easily delivered through Figure 3(a), which contains six features with five time steps, from frame 1 to frame 17, 4 frames interval. One special feature, from birth to death, has been shown in Figure 3(b), the spatio-temporal changing of the feature through its life can be obtained from the result, which could also convey a visual impact to the users.

4 Image-Based Approach

Some visual effects which are rather difficult to implement in 3D volume rendering can be easily done in image space. An obvious example is the layout of the temporal depiction. It is not always true the time-varying volume data have spatial translation over time. Sometimes its temporal characteristic is the shape or the density. In such a case, rendering several time steps into a single volume becomes too crowded to effectively show the evolution of the data. Instead, we use a combination of image processing techniques to create a time evolving flow depiction entirely in image-space.

The overall process of our image-based approach is shown in Figure 4. The input is two sets of sequences of the volume rendering snapshots and the contour

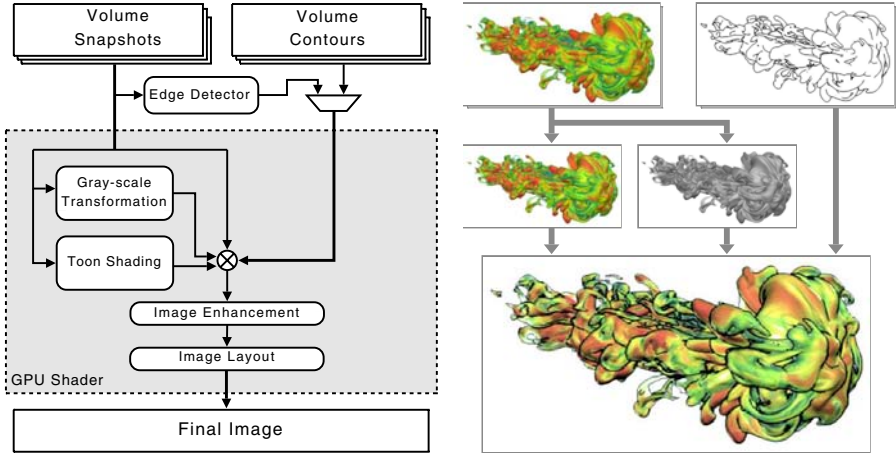


Fig. 4. The data flow of the image-based approach. The snapshots and the contours generated from volume rendering are sent to the system. Additional edge maps, gray-scale luminances and toon colors are computed to compose a compact depiction.

images extracted in volume space. Each snapshot is first sent to an edge detector. We used these images and the edge maps to obtain line-drawing renderings. Both the original frames and the edge maps are then sent to a GPU shader, which applies illustrative rendering such as toon-shading, contour enhancement and sharpening. Finally, the final image is composed according to a specified 2D layout.

4.1 Edge Detection

Edge detection attempts to localize significant variations of the gray level image in the hopes of finding meaningful object boundaries and contours [21]. However, edge detection is not a trivial task due to the similarity of the image content, the contrast of the image, and noise. In this paper, we use the Canny edge detector [2], for its robustness and quality. The output of this edge detector is an image that highlights the most important edges in a flow that can be used to enhance the original images obtained from volume rendering. Since the original images are obtained from volume rendered images, there is not much noise when compared to natural images, which makes this approach appropriate for simulating a line-drawing rendering style.

4.2 Toon-Shading

Toon-shading is a rendering technique which aims to mimic the style of a comic book or cartoon and makes the computer graphics appear to be hand-drawn [4]. To achieve the same effect from a 2D image, we simply quantize the color of images at a number of levels as : $C^* = \text{Round}[C \times t]/t$, where C^* is the resulting

toon color, C is the original pixel color in the range of $[0, 1]$ (for each red, green and blue components), and t is the quantization level.

4.3 Color Blending and Enhancement

In addition to shading, we perform color blending and enhancement to highlight different stages using rendering properties. In our work, the final color of a pixel is given by:

$$C = (t_1C_0 + t_2T + t_3G)(1 - t_4E) \quad (1)$$

where C_0 is the original color in the input image, T is the tone color computed in the previous step, G is the gray intensity of the color, and E is the edge intensity as provided by the previous step or as an input. The gray level intensity is computed as the luminance $G = 0.3R + 0.59G + 0.11B$, where R, G, B are the red, green, blue components of a pixel color, respectively. t_1, t_2 and $t_3 \in [0, 1]$ are weighting parameters that add up to one and are used to smoothly blend from gray-scale images, through tone shading, to full color images, depending on the position of the snapshots in the final sequence. In our examples, earlier time steps are assigned to gray-scale, while the final time step is rendered at full color. t_4 is a sequential factor that determines the saturation of the edge map in a snapshot. As t_4 becomes closer to 1, the contours are more apparent.

There are several methods for image enhancement [3]. Here we simply use gray-scale manipulation which maps the original gray/color intensity to a new value according to a specific curve. Gray-scale manipulation can effectively increase the image contrast. In our system, the manipulation is done in the graphics hardware on the fly, which also enables interactive adjustment.

4.4 Interactive Composition

Once all the snapshots are processed, the final step is to arrange each snapshot to the right place according to the specific layout. In order to accommodate to different types of images, our system provides interactive image composing. Users can add or remove a snapshot in real-time and the system will automatically update the sequential index of each snapshot. The layout can also be altered by moving, rotating, or scaling up and down the snapshots.

4.5 Results

Figure 5 shows the visualization of the simulation conducted by researchers at the National Center for Atmospheric Research (NCAR) that models 3D thermal starting plumes descending through a fully-compressible adiabatically-stratified fluid. In order to better estimate the penetration of the plume into the solar interior, scientists mostly focus on the stability of the plume as it evolves. The vertical layout of the plume in different time periods in Figure 5 provides a better view of how it changes over time. The line drawing and the luminance highlight in early and intermediate snapshots show the structural evolution more clearly.



Fig. 5. Plume Data Set. The vertical layout shows the plume in different time periods and provides a better view for how it changes over time.

And at the final step, we use the full-colored snapshot from the volume rendering to highlight the intensity of the plume vorticity.

As for the argon bubble data set, Figure 6(top) shows a similar layout to the plume illustration. However, we also apply the masking technique to the snapshots which de-emphasizes the turbulence structure and only show the feature of interest, the smoke ring. In doing so, we are able to put the ring structure compactly as shown in Figure 6(bottom). Likewise, the contour drawing and the fading gray-scale coloring depict the evolving structure more clearly.

5 Discussion

The traditional approach to visualizing time varying data relies on animations, which may be lengthy and only convey one instant at a time. As an alternative, we have produced illustrations that capture flow motion in a single image using a combination of both 3D object- and 2D image-based approaches. Object-based approaches have the advantage of showing all steps in the evolving flow in their original 3D position and orientation. This is useful for correlating the different stages in time and the space they occupy. Image-based approaches, on the other hand, decouple the space dimension and lets us experiment with different layouts to depict the different stages of an evolving flow. Although the actual 3D

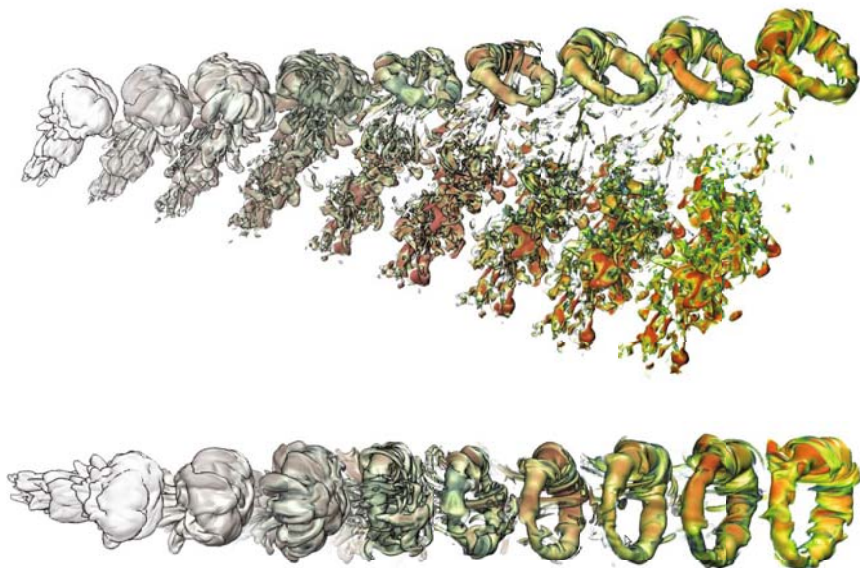


Fig. 6. Argon Bubble Data Set. The top image is a sequential arrangement of the snapshots showing the entire evolving process. By applying a color mask, we can achieve a more compact image showing on the features of interest (the smoke ring), as the bottom image shown.

information is not encoded, a different layout lets us see different stages without the inherent overlapping that may occur otherwise. Compare, for example, the illustrations generated in Figure 6 (image-based) and Figure 2 (object-based). Although the latter clearly shows the actual advance of the argon bubble, it is not possible to show many of these stages in the same image. The former, on the other hand, lets us place many stages in a single image. In general, different illustration needs demand different strategies. For example, we have shown that the use of transparency and contour rendering is important for de-emphasizing certain stages so that others are clearly depicted. In other cases, transparency is used to emphasize entire regions, such as the smoke ring in the argon bubble illustration, Figure 6 (bottom). With the use of transparency, we have traded completeness for compactness, and produce an effective illustration.

6 Conclusions and Future Work

We have presented a series of techniques for obtaining effective illustrations of time evolving flow from volume data. At the core of our approach is the premise that flow and motion can be represented effectively in a single 2D image. To convey the idea of flow, we apply different rendering styles to different stages in a flow, such as contours and toon shading. Inspired by hand-crafted illustrations, we have also experimented with different layouts to convey the different stages of flow clearly

without the ambiguity that may be raised with naive image overlapping. Obtaining a good illustration that communicates the essence of time evolving flow remains a craft. For this reason, a combination of both 3D object- and 2D image-based techniques is necessary. Coupled with interactive tools to allow the user to explore the different rendering and layout parameters, these techniques highlight important features of the data, such as contours and shading, that help understand the evolution of a complex shape over time. There are a number of directions for future work. We will explore other metaphors for depicting flow motion extensively used in art forms, such as speed lines, arrows, and motion blur. Directionality is also important when conveying motion. We will experiment with other layouts that preserve the directionality of flow.

Acknowledgement. This work is sponsored in part by the U.S. National Science Foundation's ITR and PetaApps programs and the U.S. Department of Energy's SciDAC program. The argon bubble data set was provided by the Center for Computational Sciences and Engineering at the Lawrence Berkeley National Laboratory. The plume data set was provided by the National Center for Atmospheric Research. Chris Muelder of the VIDI group at UC Davis provided the feature extraction and tracking code.

References

1. Balabanian, J.-P., Viola, I., Möller, T., Gröller, E.: Temporal styles for time-varying volume data. In: Gumhold, S., Kosecka, J., Staadt, O. (eds.) *Proceedings of 3DPVT 2008 - the Fourth International Symposium on 3D Data Processing, Visualization and Transmission*, June 2008, pp. 81–89 (2008)
2. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8(6), 679–698 (1986)
3. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 3rd edn. Prentice-Hall, Inc., Upper Saddle River (2006)
4. Gooch, B., Sloan, P.-P.J., Gooch, A., Shirley, P., Riesenfeld, R.: Interactive technical illustration. In: *I3D 1999: Proceedings of the 1999 symposium on Interactive 3D graphics*, pp. 31–38 (1999)
5. Healey, C.G., Tateosian, L., Enns, J.T., Remple, M.: Healey, Laura Tateosian, James T. Enns, and Mark Remple. Perceptually based brush strokes for nonphotorealistic visualization. *ACM Trans. Graph.* 23(1), 64–96 (2004)
6. Joshi, A., Rheingans, P.: Illustration-inspired techniques for visualizing time-varying data, pp. 679–686 (October 2005)
7. Lee, T.-Y., Shen, H.-W.: Visualizing time-varying features with tac-based distance fields, April 2009, pp. 1–8 (2009)
8. Lee, Y., Markosian, L., Lee, S., Hughes, J.F.: Line drawings via abstracted shading. In: *ACM SIGGRAPH 2007*, p. 18 (2007)
9. Lu, A., Shen, H.-W.: Interactive storyboard for overall time-varying data visualization, March 2008, pp. 143–150 (2008)
10. Ma, K.-L.: Visualizing time-varying volume data. *Computing in Science and Engg.* 5(2), 34–42 (2003)
11. Mayr, E.: *What evolution is*. Basic Books, New York (2001)

12. Muelder, C., Ma, K.-L.: Interactive feature extraction and tracking by utilizing region coherency. In: Proceedings of IEEE Pacific Visualization 2009 Symposium (April 2009)
13. Post, F.H., Post, F.J., Van Walsum, T., Silver, D.: Iconic techniques for feature visualization. In: VIS 1995: Proceedings of the 6th conference on Visualization 1995, pp. 288–295 (1995)
14. Silver, D., Wang, X.: Tracking scalar features in unstructured data sets, pp. 79–86 (October 1998)
15. Silver, D., Wang, X.: Volume tracking. In: VIS 1996: Proceedings of the 7th conference on Visualization 1996, pp. 157–164 (1996)
16. Stoppel, A., Lum, E., Ma, K.-L.: Feature-enhanced visualization of multidimensional, multivariate volume data using non-photorealistic rendering techniques. In: Proceedings of Pacific Graphics 2002, pp. 1–8. IEEE, Los Alamitos (2002)
17. Svakhine, N.A., Jang, Y., Ebert, D., Gaither, K.: Illustration and photography inspired visualization of flows and volumes, pp. 687–694 (October 2005)
18. Wen, F., Luan, Q., Liang, L., Xu, Y.-Q., Shum, H.-Y.: Color sketch generation. In: NPAR 2006: Proceedings of the 4th international symposium on Non-photorealistic animation and rendering, pp. 47–54 (2006)
19. Woodring, J., Shen, H.-W.: Chronovolumes: a direct rendering technique for visualizing time-varying data. In: VG 2003: Proceedings of the 2003 Eurographics/IEEE TVCG Workshop on Volume graphics, pp. 27–34 (2003)
20. Woodring, J., Wang, C., Shen, H.-W.: High dimensional direct rendering of time-varying volumetric data. In: VIS 2003: Proceedings of the 14th IEEE Visualization 2003 (VIS 2003), pp. 417–424 (2003)
21. Ziou, D., Tabbone, S.: Edge Detection Techniques-An Overview. *Pattern Recognition & Image Analysis* 8, 537–559 (1998)

Augmenting a Ballet Dance Show Using the Dancer's Emotion: Conducting Joint Research in Dance and Computer Science

Alexis Clay^{1,2}, Elric Delord¹, Nadine Couture^{1,2}, and Gaël Domenger³

¹ ESTIA, Technopole Izarbel 64210 Bidart, France

² LaBRI, Université de Bordeaux 1, CNRS, 33405 Talence, France

³ Malandain Ballet Biarritz, 64200 Biarritz, France

{a.clay,e.delord,n.couture}@estia.fr, g.domenger@balletbiarritz.com

Abstract. We describe the joint research that we conduct in gesture-based emotion recognition and virtual augmentation of a stage, bridging together the fields of computer science and dance. After establishing a common ground for dialogue, we could conduct a research process that equally benefits both fields. As computer scientists, dance is a perfect application case. Dancer's artistic creativity orient our research choices. As dancers, computer science provides new tools for creativity, and more importantly a new point of view that forces us to reconsider dance from its fundamentals. In this paper we hence describe our scientific work and its implications on dance. We provide an overview of our system to augment a ballet stage, taking a dancer's emotion into account. To illustrate our work in both fields, we describe three events that mixed dance, emotion recognition and augmented reality.

Keywords: Dance, Augmented Reality, Affective Computing, Stage Augmentation, Emotion Recognition, Arts and Science Research.

1 Introduction and Related Works

Joint research between a technical domain and an artistic domain seems a difficult task at first glance. How to conduct research in Science and Arts at the same time? Augmented Reality (AR) seems to be able to give a way to answer. A growing interest risen particularly from the artistic community. Within the scope of our research in AR and human-machine interaction, we developed a close relationship between a research laboratory and a ballet dance company. Our interests lie in gesture-based emotion recognition and the augmentation of a ballet stage during a dance show.

Mixed Reality [19] is an interaction paradigm born from the will to merge computers processing abilities and our physical environment, drawing a computer abilities out from its case. The goal is to eliminate the limit between the computer and the physical world, in order to allow interweaving information

from the real world and information from the virtual world. On the continuum of Mixed Reality, from real world to virtual world, the AR paradigm appears. AR consists in augmenting the real world with virtual elements such as images. For example, scanner information of a patient, collected before surgery, can be directly projected onto the patient during the surgery [1], virtual information is then added to the physical space.

We seek to explore the potential of AR in the context of a ballet dance show to better convey the choreographer message and suggest innovative artistic situations. Several augmented shows were conducted since about fifteen years ago. The evolution of technologies and systems in the field of AR allowed performance artists to use them as tools for their performances. First, The Plane [16] unified dance, theater and computer media in a duo between a dancer and his own image. With more interactive features, Hand-Drawn Spaces [17] presented a 3D choreography of hand-drawn graphics, where the real dancer's movements were captured and applied to virtual characters. Such interaction coupled with real time computing were achieved in "The Jew of Malta" [18] where virtual buildings architecture cuts and virtual dancer costumes were generated, in real time, depending on the music and the opera singer's position on the stage. However, as far as we know, using emotion recognition to enhance spectator experience by the way of AR in a ballet dance show is a new challenge. This is the aim of our thought and of our prototype described in this article.

Computer-based emotion recognition is a growing field of research that emerged roughly fifteen years ago with R.W. Picard's book "Affective Computing" [8]. The process of emotion recognition is usually divided into three steps: capturing data from the user, extracting emotionally-relevant cues from this data, and interpreting those cues in order to infer an emotion. Main issues in this area cover the identification and validation of emotionally-relevant cues according to different affective channels (e.g. facial expression, voice subtleties), and interpreting those cues according to a certain model of emotion [12]. In our work on emotion recognition, we rely on Scherer's definition of affect [9] as a generic term and emotions as a category of affect.

In the following, we describe how we conducted a research jointly between the computer science field and the dance field, taking from each domain to progress in the other. We describe the computer systems used for recognizing emotions and augmenting a ballet dance show, as well as how those technologies were used for research in dance. We then describe three events that we jointly conducted. One is dominantly scientific, another one is dominantly artistic, and the last one is a balanced mix of both.

2 Gesture-Based Emotion Recognition

There is a large literature about bodily expression of emotions. Darwin [11] listed several body movements linked to different emotions. In the field of psychology, de Meijer [2] identified and validated, through human evaluations of actor performances, affect-expressive movement cues, such as trunk curvature, position of

hands or velocity of a movement. Later, Wallbott [3] conducted a similar study with different sets of emotions and body movements. The analysis of evaluation data enabled both of them to compute the weights of each movement cue in the expression of a set of particular emotions. Coulson [4] extended this research by working on affective cues in static body postures. In the domain of dance, Laban's Theory of Effort is a seminal work focused on expressive movements. As a choreographer, his work was drawn from and applied to dance. Laban divides human movement into four dimensions: body, effort, shape and space [13]. These dimensions focus on describing the expressiveness of a gesture by identifying how a particular gesture is performed, as opposed to what gesture is performed. In the field of computer science, the Infomus Lab in Genoa based their research on Laban's theory to identify and validate formally-described cues for computer-based recognition. Their studies cover several artistic contexts such as dance [5] and piano performances [6].

This research frame pushes the dancer to question himself on what is "interpretation". Sharing emotions with an audience and a computer are clearly two different things, but it implies the same process of research. This process that goes through his body and implicates his mind follows a long tradition of research and theories that dance carries through history to its today practice. The body being the focus point of dancers, the whole process questions the relationship that the dancer developed during his career between his mind and his body. It forces him to come back to basics and fundamentals. What is space? What is time? What are the others? Here are the questions that he has to ask to himself and to the computer. Computer-based emotion recognition forces us to establish a distinction between emotions and their expression. This nuance implies a deeper analysis of movement qualities and on the ways we have to translate them into a computer language. This translation needs to be regularly upgraded each time scientists and dancers manage to move forward together. The relationship between scientists and dancers remain, in that research, at the centre of every progress made by the computer concerning recognition of emotions through observation of movement. It is somehow imperative for the scientist to understand the modifications of qualities produced by the dancer's dance in order to refine the recognition parameters given to the computer. The dancer in his side needs to explain and dissect his practice to reach a better understanding by the computer. This whole process generates an atmosphere that is clearly particular and unusual for the practice of dance. We have reached a point where we can observe the apparition of emotions that needs to take their place in the global experiment that scientist and dancer are going through, together.

3 Augmenting a Ballet Dance Show: Use Case

We developed a framework to augment a ballet dance show built on several applications, which aims at providing with a generic and reusable solution for augmenting a performing art show.

3.1 eMotion Application

eMotion is a computer-based gestural emotion recognition system that relies on three successive steps: acquiring data, extracting gestural and postural emotional cues, and interpreting them as an emotion. eMotion relies on gestural cues drawn from de Meijer [2] for inferring an emotion at each frame. Those cues are verbally described and have to be interpreted to be implemented in a computer system. De Meijer's work provides, along with those cues, their respective weight for interpreting an emotion. We hence interpret an emotion as the one corresponding with the maximum sum of weighted characteristics values, using de Meijer's weights values.

The eMotion software relies on motion capture which can send over the network the coordinates of 23 segments of the dancer's body in an open XML format. From the flow of coordinates provided by the Moven application which is described in [3,2], the eMotion software computes trunk and arm movement, vertical and sagittal directions, and velocity. The interpretation is then performed by choosing the maximum weighted sum of each cue over each of the six basic emotions: joy, sadness, anger, disgust, fear and surprise. The eMotion application delivers an emotion label at each frame and is able to send it over the network through an UDP connection.

3.2 Moven Studio Application

The Moven suit is a motion capture suit that uses 16 inertial motion trackers composed of 3D gyroscopes, 3D accelerometers and magnetometers. The body model is a biomechanical model composed of 23 segments. Sensors allow for absolute orientation of each segment; translations are computed from each segment's orientation. Moven Studio is a commercial solution software provided by XSens with the Moven motion capture suit. Its role is to acquire and process data in real time from the suit. It also offers an interface for post-processing of recorded movements. Network features allow Moven Studio to send motion capture data over the network.

3.3 ShadoZ Application

ShadoZ is the contraction of the word "Shadow" and the Z-axis of a three-dimensional space. Its purpose is to use both movement information from Moven studio and the emotion expressed by the dancer from the eMotion application to create a shadow. The shadow's color and size depends on the dancer's emotion and is projected on stage. The ShadoZ application is composed of a core architecture supplemented by plugins for augmentations. Both the core and plugins can be multi-threaded. This architecture relies on design patterns for better evolutivity. The ShadoZ application is implemented using C++ language in conjunction with Trolltech's Qt library and OpenGL for rendering 3D scenes.

The system is distributed over three computers, communicating through UDP connections (see figure [1]). The first computer hosts the Moven Studio software

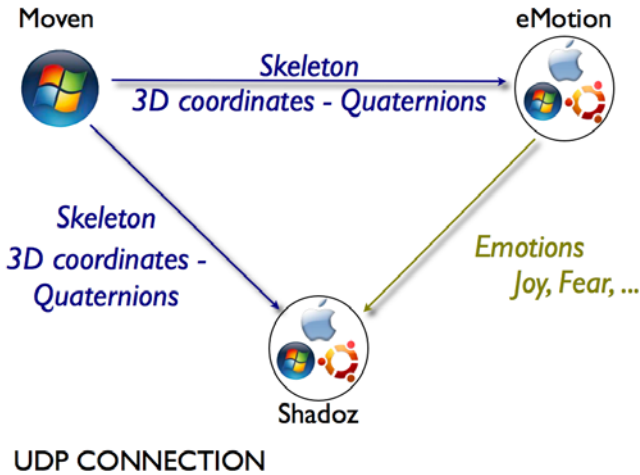


Fig. 1. The heterogeneous system distributed with Moven Studio, eMotion and ShadoZ on three operating systems: Windows Vista, Mac OSX Leopard and Linux Ubuntu

and send dancer's body coordinates over the network. The eMotion software is on a second computer and takes the dancer's movement data to infer an emotion at each frame. This emotion is sent continuously over the network to ShadoZ. Finally the ShadoZ application uses the coordinates from the Moven suit to create a virtual shadow that mimics the dancer's movement. Dancer's emotions are mapped to the virtual shadow, which changes size and color accordingly. Mapping between emotions, color and size ratio were drawn according to Birren [10] and Valdez [7] studies.

3.4 Augmented Technologies, a Support for Research in Dance

The different propositions made by AR technologies allow dance to step forward in its principles and to share the staging process with other art forms that have not participated until today to its development. The different steps of research that have been made to understand the particularities of the relationships between humans and machines have generated an aesthetic of its own that we find interesting to include into the staging work. To correspond with that aesthetic we observed that the collaboration between science and the particular art form of dance has created a form of thinking that could also be included in the staging process and that would give access to the audience to the questions generated by our research. We would like as well to build a story line that strips the modalities of that research on emotion and give meaning to its presentation as a show. The materials of texts, graphics, sounds, lights, that represent this research need something to link them to each other to be able to appear on stage and create an artistic logic reachable and clear for the audience. The tools proposed by AR technologies offer the opportunity to bring forward research on dance and help

dancers finding solutions on the path of sharing with an audience the emotions that are being exchanged between a machine and a human.

4 Experience and Events: Bases of Joint Research

4.1 Affective Dance Collection

In order to design augmenting modalities and test the recognition of emotion by the eMotion module, we collected motion-captured movements of a dancer. Dance sequences were performed by a single professional ballet dancer (see figure 2a). Recordings were performed using the Moven capture suit, a firewire video camera and a camcorder. Expressive dance improvisations were collected in two sessions. In the first one, the dancer was told a single emotion label from our chosen set. Each affective state had to be performed three times. The 24 resulting sequences were randomly performed. The second session mixes the Ekman's six basic emotions [20]. Seven pairs were studied and randomly performed. This recording session allowed us to obtain 31 affective dance sequences as a collection of materials. Seven human evaluators were asked to choose, for each sequence, the emotions they thought the dancer expressed among the set of eight emotions. These sequences and their evaluation provided us with testing material for our eMotion application. These recollection sessions were clearly scientifically-oriented, though their setting proved interesting from an artistic point of view, as the dancer explored how to express a particular emotion through an improvised dance.

4.2 Ethiopiques: An Improvised Augmented Show Combining Dance, Music, Text Reading and a Virtual World

This event took place in Bayonne, South-West of France, in March 2009 (see video on [21]). It took the form of an improvised show in one of the artists' flat and was open to the public, providing a comfy, but a bit strange atmosphere. A saz (a kurd lute) player and an accordion player improvised a musical atmosphere over which poetry was read. A professional dancer wore the Moven motion capture suit and improvised a dance over the music and text. Then, a 3D skeleton of the dancer was projected onto the wall. At the same time, a draftsman triggered flash animations in real time that superposed with the virtual scene. At the beginning of the show, the dancer was in a separate room. The audience could only see its virtual *alter ego* projected on the wall, and the superposed flash animations (as in figure 2b). In the end of the show, the dancer moved and danced within the audience, which at this moment could see the real dancer interacting with them. The virtual dancer, the real dancer and his shadow on the wall formed a trio of dancers that bridged together the virtual and the physical world, as the three of them interacted with themselves, the audience and with the flash elements that were superposed to the 3D scene.



(a) Affective dance collection: the Dancer wearing the Moven motion capture suit



(b) Scene of the Ethiopiques show with the dancer, his virtual representation, and virtual elements surimposed to the scene

Fig. 2. Jointly performed events

4.3 An Open Danced Talk

The danced conference form is still rarely used but has been applied to domain such as paleoanthropology [14] and chaos theory [15]. In this danced conference, we mixed scientific communication and improvized danced to withdraw from the classical forms of scientific presentation and artistic representation. The dancer wore the Moven motion capture suit and improvized over a text, accompanied by music and lights. The form of a danced talk became naturally when trying to bridge the domains of computer science and dance. This form allows the dance audience to integrate a research problematic and processes, and allows the scientific audience to withdraw from a purely technical approach and better grasp the interest of research on emotion recognition. In the frame of our research, the danced talk explicited a constant and rewarding interaction between dance and research and allowed an equally rewarding interaction between dancers, researchers and the audience.

5 Conclusion

For computer scientists, such a collaboration is an occasion to explore the domain of dance and to study body language and movements in order to allow a computer recognizing the emotion expressed by a dancer. For dancers and choreographers, it is an occasion to go back to the source of movement and to re-question the fundamental themes of dance, time, space, and the others, while being able to see its animating concepts shape themselves as a virtual representation casts in the real world. The first step of the collaborative research was establishing a common ground for dialogue. For scientists, it allowed understanding some of

the world of dance and the significance of some concepts such as time, space, and the other. For dancers, it opened the doors on the reality of scientific research and to better understand what it could and could not bring to dance. We hence experienced conducting research jointly, between the computer science domain and the dance domain. Such a collaboration brings forward many advantages. For scientists, dance and dancers can be used as an application case and an experimental tool. Artists creativity makes them formulating new needs that drives research forward. For dancers, science presents itself as a world to explore throughout their arts. Its constant questioning and attempts to model reality implies revisiting the fundamentals of dance. Finally, developed technologies provides artistic tools for visiting new virtual worlds.

Acknowledgments. This joint research was conducted within the frame of the CARE project (Cultural experience - Augmented Reality and Emotion), funded by the french National Agency for Research (ANR) [22].

References

1. Grimson, W.E.L., Ettinger, G.J., White, S.J., Lozano-Perez, T., Wells, W.M., Kikinis, R.: An Automatic Registration Method for Frameless Stereotaxy, Image Guided Surgery and Enhanced Reality Visualisation. *IEEE Trans. on Medical Imaging* 15(2), 129–140 (1996)
2. De Meijer, M.: The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior* 13(4), 247–268 (1989)
3. Wallbott, H.G.: Bodily expression of emotion. *European Journal of Social Psychology* 28, 879–896 (1998)
4. Coulson, M.: Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior* 28, 117–139 (2004)
5. Camurri, A., Lagerlof, I., Volpe, G.: Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies* 59(1-2), 213–225 (2003)
6. Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., Scherer, K.R.: Automated analysis of body movement in emotionally expressive piano performances. *Journal Article, Music Perception* 26(2), 103–120 (2008)
7. Valdez, P., Mehrabain, A.: Effects of Color on Emotions. *Journal of Experimental Psychology* 123(4), 394–409 (1994)
8. Picard, R.: *Affective Computing*, *Comput. Entertain.* MIT Press, Cambridge (1997)
9. Scherer, K., WP3 Members: Preliminary Plans for Exemplars: Theory, Version 1.0, In HUMAINE Deliverable D3c, 28/05/2004 (2004), <http://emotion-research.net/deliverables/D3c.pdf>
10. Birren, F.: *Color psychology and color therapy*. University Books Inc., New Hyde Park (1961)
11. Darwin, C.: *The expression of the emotions in man and animals*. The Portable Darwin (1872)

12. Clay, A., Couture, N., Nigay, L.: Towards Emotion Recognition in Interactive Systems: Application to a Ballet Dance Show. In: Proceedings of the World Conference on Innovative Virtual Reality (WinVR 2009) - World Conference on Innovative Virtual Reality (WinVR 2009), Chalon-sur-Saone, France (2007)
13. Hodgson, J.: *Mastering Movement: The Life and Work of Rudolf Laban*. Paperback edition (2001)
14. http://www.cite-sciences.fr/francais/ala_cite/evenemen/danse_evolution06/
15. http://www.koakidi.com/rubrique.php3?id_rubrique=97
16. Troika Ranch, <http://www.troikaranch.org/>
17. Kaiser, P.: Hand-drawn spaces. In: SIGGRAPH 1998: ACM SIGGRAPH 98 Electronic art and animation catalog, Orlando, Florida, United States, p. 134. ACM, New York (1998)
18. The Jew of Malta, <http://www.joachimsauter.com/en/projects/vro.html>
19. Milgram, P., Kishino, F.: A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D(12), 1321–1329 (1994)
20. Ekman, P., Friesen, W.V.: *Unmasking the face. A guide to recognizing emotions from facial clues*. Prentice-Hall Inc., Englewood Cliffs (1975)
21. The Ethiopiques show (2009), <http://www.youtube.com/watch?v=XGXzXmFwr68>
22. The CARE Project, <http://www.careproject.fr>

Automatic Skin Color Beautification

Chih-Wei Chen, Da-Yuan Huang, and Chiou-Shann Fuh

Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan
{d95013, r97022, fuh}@csie.ntu.edu.tw

Abstract. In this paper, we propose an automatic skin beautification framework based on color-temperature-insensitive skin-color detection. To polish selected skin region, we apply bilateral filter to smooth the facial flaw. Last, we use Poisson image cloning to integrate the beautified parts into the original input. Experimental results show that the proposed method can be applied in varied light source environment. In addition, this method can naturally beautify the portrait skin.

Keywords: Skin beautification, color-temperature-insensitive, bilateral filter, Poisson image cloning.

1 Introduction

Recently, digital still camera becomes more widespread. To provide more capability for users and satisfy their different requirements, many techniques have been integrated into digital still camera.

Skin beautification is the technique to smooth the skin of the photographed portrait. The result of skin beautification is preferable to the original one. Many consumer cameras have embedded the function of skin beautification. Besides, there are many approaches proposed to deal with skin beautification. However, most of these techniques rely on accurate skin color detection.

In this paper, we propose an original framework based on color-temperature-insensitive skin-color detection. Then, we use the characteristic of bilateral filter, which can preserve the significant image edge and can smooth the low gradient parts in given image, to polish the selected skin regions. Finally, to satisfy human perception, we apply Poisson seamless cloning algorithm to integrate the polished part into original image.

2 Backgrounds

2.1 Skin Color Detection

There are many skin color detection algorithms proposed to extract the skin parts from given image [1] [2]. Most of these approaches use statistical method to find out

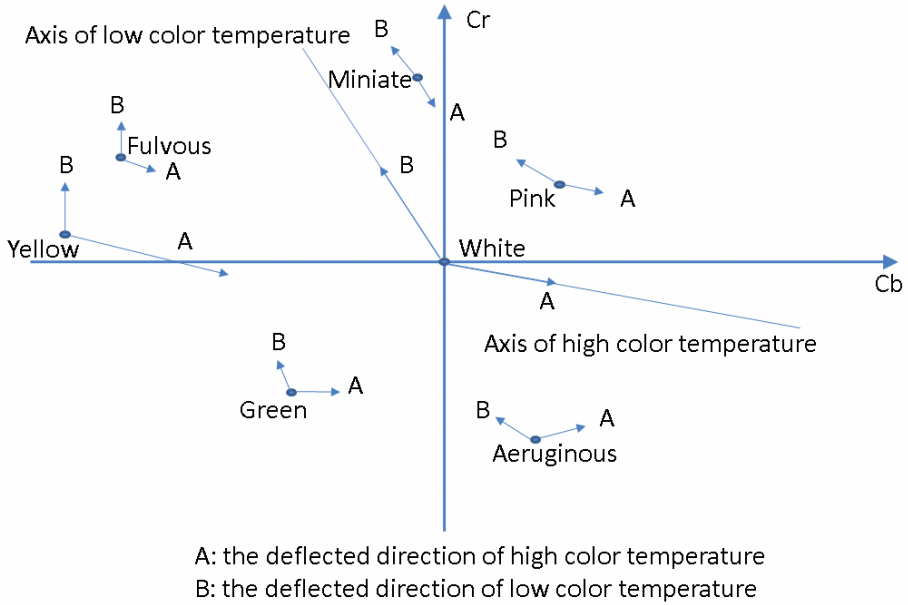


Fig. 1. The experimental data of color patches under different light sources [3]

the boundary of human skin color in specific color space. However, skin color is sensitive to color temperature. The relation between color and color temperature is shown in Fig. 1.

To lower the effect of color temperature, the better for us to use face detection algorithm as preprocessing step. Then, we can determine the characteristic of skin color from the abovementioned face detection.

2.2 Face Detection

To extract human face from given image, many algorithms have been proposed. Viola et al. proposed an efficient face detection approach [4]. There are three main principles in their algorithm, including the integral image, Ada boost, and cascade classifier. Furthermore, Lienhart et al. proposed Haar-like features for object detection [5]. In this paper, we apply Lienhart’s frontal face classifier [6] and Hameed’s eye classifier [7] to extract human face and human eyes from given image respectively. The objective of human eye detection is to subtract the effect of non-skin tone element from detected human face. The result of human face detection and human eye detection is shown in Fig. 2(b). The white color in Fig. 2(c) is the region of interest, which is a rough skin-color data.

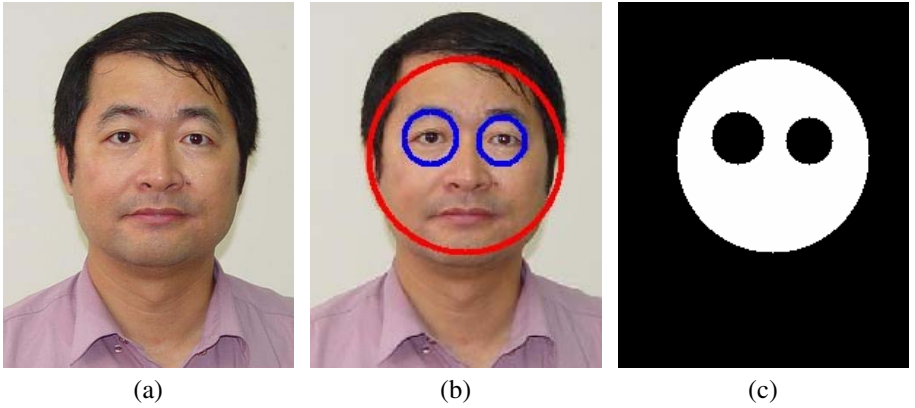


Fig. 2. (a) Input image. (b) Red circle and blue circle mean face region and eyes region respectively. (c) White region represents interested skin color.

2.3 Bilateral Filter

The bilateral filter is a nonlinear technique that can smooth the input image while preserving its main edges [8] [9]. With a Gaussian function $g_{\sigma}(x) = \exp(-x^2 / \sigma^2)$ the bilateral filter of image I at pixel p is defined by:

$$BF(I)_p = \frac{1}{W} \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(\|I_p - I_q\|) I_q, \quad (1)$$

$$w = \sum_{q \in I} g_{\sigma_s}(\|p - q\|) g_{\sigma_r}(\|I_p - I_q\|),$$

where w is a normalized factor; g_{σ_s} and g_{σ_r} are typical Gaussian kernel functions; and p and q indicate spatial locations of pixels.

2.4 Seamless Cloning

The objective of Poisson seamless cloning is to find out a function, which makes the gradients of target image closest to the gradients of source image. We implement the discrete Poisson solver [10] to obtain finished result of skin beautification. In this paper, the guidance field vector v is

$$v = \nabla g, \quad (2)$$

where ∇g means the gradient of source image.

3 Our Proposed Method

The proposed framework is shown in Fig. 3.

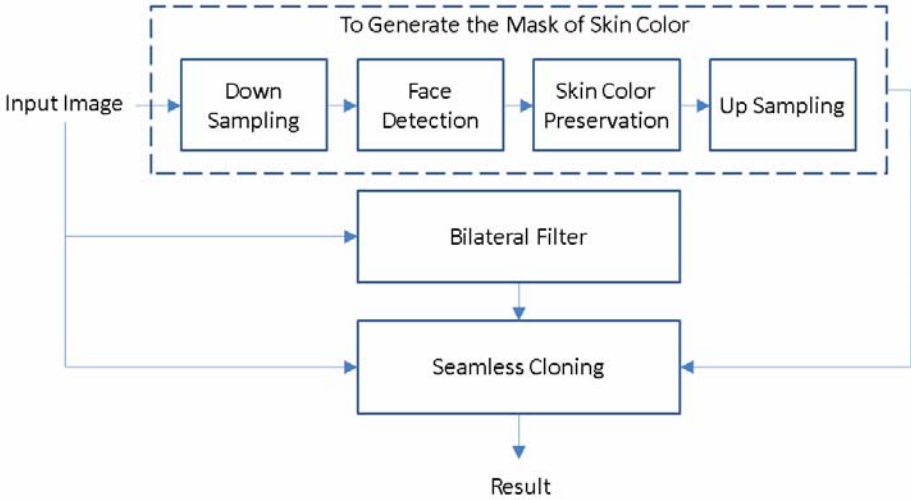


Fig. 3. System flowchart

We use down-sampling approach to lower the effect of facial flaw (e.g. scar, zit, spot, and so on). Besides, to downsize the input image can improve the efficiency of face detection. After face detection, the interested skin parts have been preserved, the sampling result is shown in Fig. 2(c). To achieve color-temperature-insensitive skin-color preservation, we translate the interested skin parts into YCbCr color space. Then, we calculate each mean value and standard deviation of Y, Cb, and Cr respectively. The abovementioned procedure is shown in Fig. 4. To prevent the effect of extreme value, we skip 10% of the head and tail in Y, Cb, and Cr histogram.

The skin-color mask of image I at pixel p is defined by

$$M(I_p) = \begin{cases} 1, & \text{if } \mu_Y - 2\sigma_Y \leq Y(I_p) \leq \mu_Y + 2\sigma_Y \\ & \text{and } \mu_{Cb} - 2\sigma_{Cb} \leq Cb(I_p) \leq \mu_{Cb} + 2\sigma_{Cb} \\ & \text{and } \mu_{Cr} - 2\sigma_{Cr} \leq Cr(I_p) \leq \mu_{Cr} + 2\sigma_{Cr} \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where $Y(I_p)$, $Cb(I_p)$, and $Cr(I_p)$ mean the YCbCr value of image I at pixel p respectively; μ_Y , μ_{Cb} , and μ_{Cr} represent the mean value of interested region in YCbCr color space respectively; σ_Y , σ_{Cb} , and σ_{Cr} mean the standard deviation value of interested region in YCbCr color space respectively.

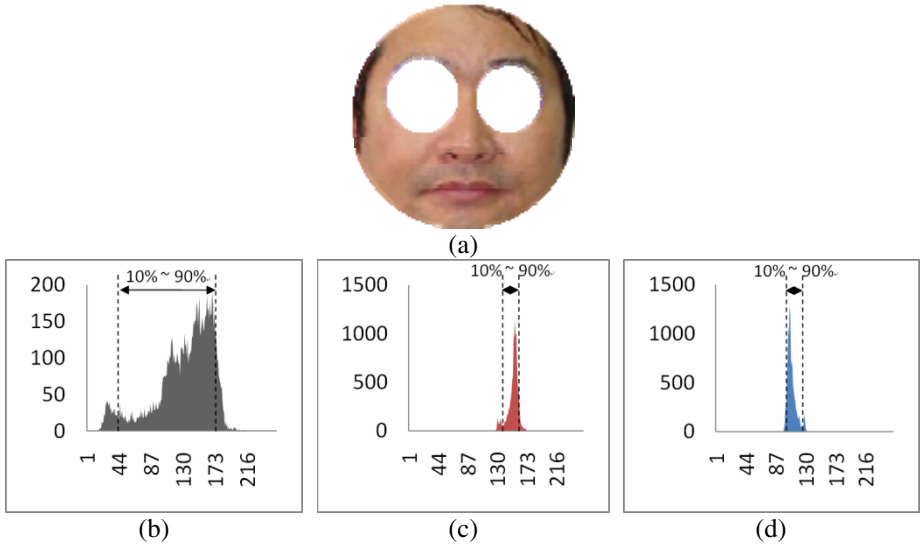


Fig. 4. (a) Region of interest. (b) The Y-channel of (a). (c) The Cr-channel of (a). (d) The Cb-channel of (a).

The abovementioned mask is generated in downsized form due to the down-sampled input image. To accomplish the skin-color mask generation, we apply bilinear interpolation to adjust the size of the downsized form to equal to the size of input image.





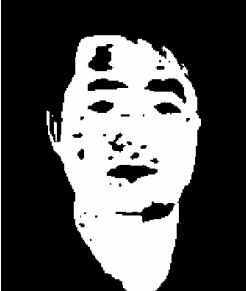


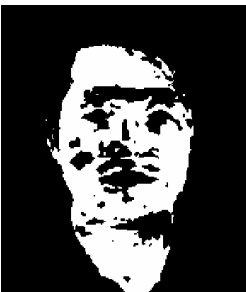

Concurrently, we apply bilateral filter to smooth the original input image. The result of bilateral filter is called base image. The result of bilateral filter is similar to the output of low-pass filter. Moreover, bilateral filter can preserve the image edge simultaneously.

Finally, we integrate the input image, mask of skin-color, and the base image obtained from bilateral filter algorithm into result image. In this paper, we apply Poisson image cloning as image integration algorithm. The advantage of Poisson image cloning is to integrate the given image into harmonious image, even if there are few mis-detections or false-alarms in the skin-color mask.

4 Experimental Result

To prevent the influence of color temperature, we use flexible method to determine the skin-color boundary. The performance of our skin-color detection is stable in various color temperatures. Table 1 shows the results of the proposed method. The size of input image is 349×419 , and the σ_s and the σ_r in bilateral filter is set to 60 and 0.1 respectively.

Table 1. The skin-color mask and beautification result of our proposed method in various color temperatures

		Input Image	Mask of Skin Color	Result
Color Temperature	Low			
	Normal			
	High			

According to human perception, we can generate a touching portrait by smoothing the blemish (e.g. facial wrinkle, sweat pore, rash, and so on) of human face. Besides, our approach can be applied to lower the image noise in portrait photography. Table 2 shows the result of skin beautification in normal color temperature.

Table 2. The results of skin beautification in normal color temperature

Input Image	Mask of Skin Color	Result
		
		
		
		

5 Conclusion

To judge the performance of skin beautification is quite subjective. In this paper, we propose an automatic approach, which is color-temperature-insensitive, to extract skin-color from given image. Then, we lower the blemish in interested part. There still are many challenging topics, including automatic skin tone adjustment in various light sources, automatic make up, automatic skin beautification in video stream, and so on.

Acknowledgements. This research was supported by the National Science Council of Taiwan, under Grants NSC 94-2213-E-002-032 and NSC 95-2221-E-002-276-MY3, by EeRise Corporation, EeVision Corporation, Primax Electronics, Altek, Alpha Imaging, Utechzone, TRI, Genesys Logic, Lite-on, and Syntek Semiconductor.

References

1. Tomaz, F., Candeias, T., Shahbazkia, H.: Fast and accurate skin segmentation in color images. In: Proceedings of First Canadian Conference on Computer and Robot Vision, 2004, pp. 180–187 (2004)
2. Kukharev, G., Nowosielski, A.: Visitor Identification - Elaborating Real Time Face Recognition System. Short Communication Papers Proceedings of WSCG 2004, pp. 157–164 (2004)
3. Yung-Cheng, L., Wen-Hsin, C., Ye-Quang, C.: Automatic white balance for digital still camera. IEEE Transactions on Consumer Electronics 41, 460–466 (1995)
4. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. Int. J. Comput. Vision 57, 137–154 (2004)
5. Lienhart, R., Maydt, J.: An extended set of Haar-like features for rapid object detection. In: Proceedings of 2002 International Conference on Image Processing, vol. 901, pp. I-900–I-903 (2002)
6. Lienhart, R.: Stump-based 20x20 gentle adaboost frontal face detector, <http://www.lienhart.de/>
7. Hameed, S.: Stump-based 20x20 frontal eye detector, <http://umich.edu/~shameem>
8. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision, 1998, pp. 839–846 (1998)
9. Paris, S., Kornprobst, P., Tumblin, J., Durand, F.: A gentle introduction to bilateral filtering and its applications. In: ACM SIGGRAPH 2008 classes. ACM, Los Angeles (2008)
10. P'erez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Trans. Graph. 22, 313–318 (2003)

Tracking Small Artists

James C. Russell¹, Reinhard Klette², and Chia-Yen Chen³

¹ University of California - Berkeley, Berkeley, USA

² University of Auckland, Auckland, New Zealand

³ National University of Kaohsiung, Taiwan

Abstract. Tracks of small animals are important in environmental surveillance, where pattern recognition algorithms allow species identification of the individuals creating tracks. These individuals can also be seen as artists, presented in their natural environments with a canvas upon which they can make prints. We present tracks of small mammals and reptiles which have been collected for identification purposes, and re-interpret them from an esthetic point of view. We re-classify these tracks not by their geometric qualities as pattern recognition algorithms would, but through interpreting the ‘artist’, their brush strokes and intensity. We describe the algorithms used to enhance and present the work of the ‘artists’.

Keywords: Rats, reptiles, footprints, image processing, pseudo coloring.

1 Introduction

Tracks of animals have been used since humans were first hunter-gatherers as a means of identifying and stalking species of interest, historically for food [5], but more recently in application for environmental surveillance [7]. Animal tracks are a useful means of confirming species presence, and hence community composition at sites. Tracking methods are particularly useful as they are non-invasive, using inked tracking cards in tracking tunnels, thus providing very little disturbance to an animal’s normal behavior (cf. methods such as trapping). After an individual walks through a tunnel and across the inked pad of the card, it leaves some footprints on the white part of that card (see Fig. 1, left). Those cards are later collected, and the “canvas” parts are visually scanned for species identification (see Fig. 1, right). Recent applications of pattern recognition methods have also shown that a geometric approach to such tracks (namely, of various species of rats) can allow very accurate (about 80%) confirmation of the species responsible for leaving the print [9,11]. However, this is not the subject of this paper.

We have collected tracks of many small animal species such as rats, lizards and insects over the course of our work. When presenting this research at various occasions, these tracks also attracted interest from an esthetic point of view, with comments such as ‘similar to an abstract painting’, ‘looks like an alien visited Earth’, or ‘having that printed on a T-shirt would certainly sell well’. Building on those comments, we present and comment here on a selection of image-processed footprints of some species, with a brief characterization of those species (introducing the ‘artist’) and of the selected processing techniques. The

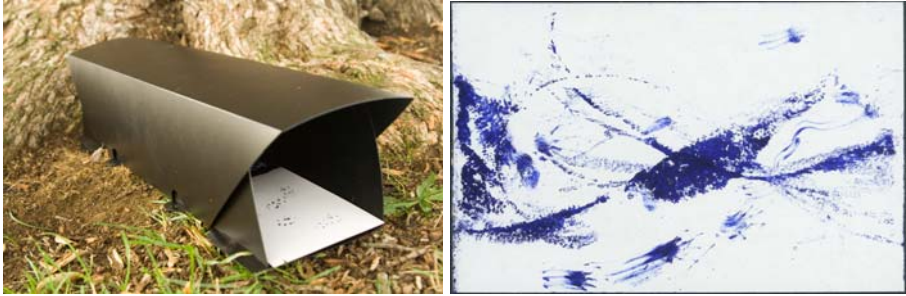


Fig. 1. Left: Tracking tunnel system - the artist's canvas. Right: Original print of a robust skink *Oligosoma alani*.

shown processed images can also be seen online at [\[3\]](#) (in color, and reasonable resolution).

We treat the cards by performing different algorithms upon them, much as one would do for photo processing. Our treatments can be broadly classified by pseudo-colorings (i.e., mapping of the given gray-scale into a set of selected colors), posterization (i.e., reduction of gray levels), removal of artifacts (i.e., elimination of small dark regions by assigning a brighter gray-value or color), or some kind of contrast enhancement or blurring; however, in any case we aimed at being careful to stay close to the artist's strokes or patterns, and the processed result may just provide another interpretation of the given original (see Fig. [2](#)). The paper is organized as follows. We present some of the artists with samples of their work in Section 2, and then discuss the overall collection of works in Section 3.

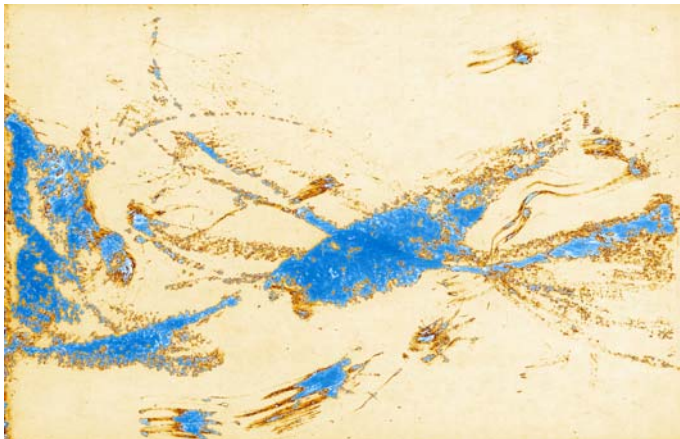


Fig. 2. Processed print of a robust skink *Oligosoma alani* using contrast enhancement and subsequent pseudo-coloring; for the original scan see Fig. [1](#) right

2 The Artists

From our large gallery of available expressions, we only present here a few New Zealand-based artists in this short paper; introduced rats and native New Zealand skinks and geckos.

2.1 Rats

Rats are by far the most wellknown of the artists we consider here. Three species of rats (*R. rattus*, *R. norvegicus* and *R. exulans*) have been introduced around the world [1]. The tracks of rats have previously been used in medical studies as a measure of damage to sciatic nerves, via a functional index [6]. More recently, rat tracks have been used to identify new arrivals on rat-free islands [8]. The genus *Rattus* is quite morphologically conserved, so differences between species can be cryptic to identify, although recent work has found subtle geometric differences [9]. The process of a rat walking across an inked card can introduce significant variability into the tracking ('artistic') process however. Biological differences among the rats such as age, size and sex can all play a role, while the quality and amount of ink and canvas will also affect the work. Together these factors interact, with the speed and agility of the rat upon the ink, leading to a myriad of patterns, such as foot-sliding and tail-dragging.

We present here (see Fig. 3) the work of a large Wistar-strain male laboratory rat (*R. norvegicus*). The footprints themselves remain discrete, founding the background of the work, while body-drag of the individual draws unifying links among the footprints, indicating more precisely the exact course the artist took across the canvas. The final touches include lighter footprints of the animal as it re-explores the canvas for a second time. The movement from left to right across the canvas provides a logical flow for the work, facilitated by our ability

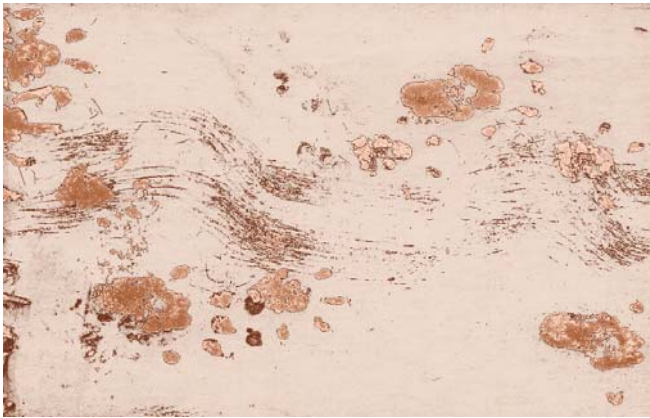


Fig. 3. The work of a large Wistar-strain male laboratory rat *R. norvegicus*. Pseudo-colored in earthy tones.

to readily identify the geometric orientation of the footprints. We have further processed this print by applying an earthy pseudo-coloring to the original scan.

2.2 Skinks

In contrast to our introduced rats, lizard populations display remarkable diversification in morphology among species as adaptations to their environments [10]. The New Zealand skink fauna, by example, is incredibly diverse, having radiated from a mono-phyletic lineage into a large number of species and variations in morphological characters [2]. We follow the revised nomenclature of [2].

The robust skink (*Oligosoma alani*) has been a victim of introduced rats in New Zealand. Once widespread, it is currently restricted to small islands off the north-eastern coast of New Zealand. Its large size is evident from its work where its body has dragged across the canvas, and the usually clear differentiation between footprints and body is diminished in this example (see Fig. 4). When viewed in a particular way an emblematic print of an entire gecko body can almost be made out on the canvas, as if the artist has left an impression of himself - perhaps a record lest they be forgotten as their range diminishes. Minimal processing has been applied to this print so as not to detract from the uniqueness of its character in comparison to other works. The artist has been allowed to ‘speak for himself’.

Like the robust skink, the moko skink (*Oligosoma moco*) is also most commonly found on islands off north-eastern New Zealand. The characteristics of its distinctly long and slender toes and tail are clearly reflected in this print (see Fig. 5). We have gray-scale-inverted this print and processed it for minor simplification and coloring. This abstraction of the details removes the artist from its



Fig. 4. Robust skink *Oligosoma alani*. Small artifacts have been removed from original scan, followed by pseudo-coloring.



Fig. 5. Moko skink *Oligosoma moco*. Inverted gray scale. Posterization and pseudo-coloring.

morphological reality and almost lifting it into a third dimension. Such prints can occur naturally however in softer substrates such as mud or soil, where water flow softens the print.

In comparison to its northern island coastal relatives, the Otago skink (*Oligosoma otagense*) is found in the southern inland areas of New Zealand; it is intimately linked with schist rock formation habitats. In contrast to the previous two artists, the pattern of this one is less decisive (see Fig. 6). The artist appears to be exploring the canvas, and to reflect this different approach we chose strongly contrasting pseudo-coloring.

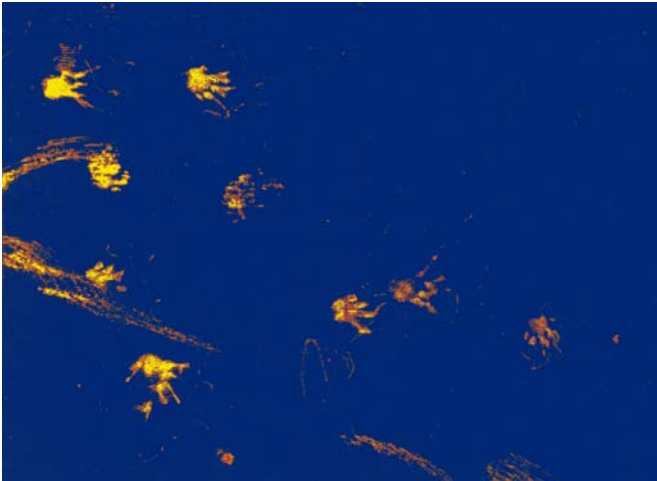


Fig. 6. Print of an Otago skink *Oligosoma otagense*, after applying pseudo-coloring only

2.3 Geckos

Similarly to their close relatives, the skinks, geckos in New Zealand have displayed a remarkable diversification. Subtle differences in their morphology are already clear when comparing prints between the skinks and geckos. Geckos have much fatter digits, allowing them great adaptability in their more arboreal environments. Geckos prefer to work by night, and we have chosen to emphasise this by using darker tones, if not outright black, in the prints.

The forest gecko (*Hoplodactylus granulatus*) is found throughout most of New Zealand, although is very secretive and often not found unless actively searched for. In this case the artist has made a direct path across the canvas, perhaps reflecting a determined curiosity of the canvas but only passing over once (see Fig. 7). We have only ‘illuminated’ the print against a dark background, highlighting once again the boldness of this individual in its nocturnal habitat - not afraid to stand out.

The common gecko (*Hoplodactylus maculatus*) is found throughout New Zealand, and is the most widespread and abundant of geckos. As with other New Zealand geckos the mother gives birth to twins once a year but provides no parental care. As if reflecting this wandering off from the offspring, we can see the fading prints in this work (see Fig. 8). Small artifacts have not been removed, and they create a uniform noise pattern on the yellow background. This noise could be much like life for common geckos among so many other individuals, where time to raise offspring is more valuably spent elsewhere such as protecting your own survival.

Duvacel’s gecko (*Hoplodactylus duvaucelii*) is restricted to a very small number of undisturbed islands in the north-east of New Zealand, but has been raised in captivity and reintroduced to many other island locations following pest eradication [12]. As if reacting to this intense human management, we can see that the

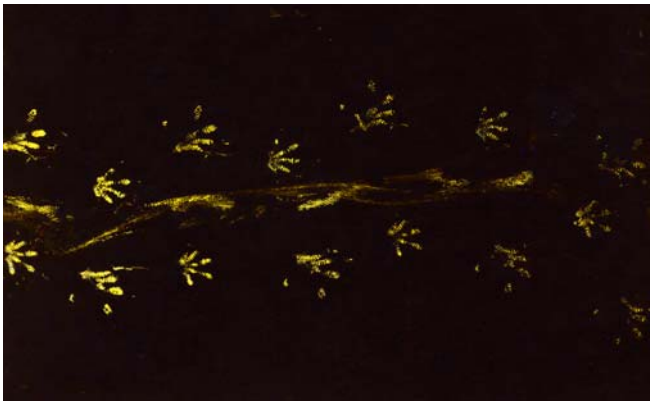


Fig. 7. Forest gecko *Hoplodactylus granulatus*. An enlightening color scale has been applied after gray-scale inversion.

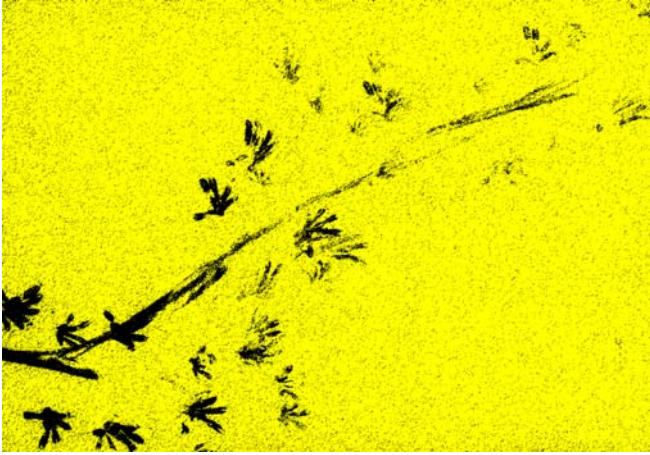


Fig. 8. Common gecko *Hoplodactylus maculatus*. Coloring of background of original print, and contrast enhancement.



Fig. 9. Duvacel's gecko *Hoplodactylus duvaucelii*. Gray-scale inversion only.

artist has almost stomped his way around the print (see Fig. 9). Only one or two truly clear impressions have been left, the remaining work being destroyed almost immediately after it was made. Minimal change (gray-scale inversion only) is required to present this work.

3 Discussion

The works we have presented here share much in common with [4], where small animals are naturally encouraged to create works. In contrast to that school of work however, we consider it as being important for the presented work that artists were able to produce these prints within their natural habitat. Without

such control over the process of artistic impression however, the work here must necessarily draw upon post-processing digital techniques, much as digital photography does today when conditions for the original work cannot be controlled much as in a laboratory setting.

The nature of this post-processing also overlaps significantly with our other work in pattern recognition algorithms [9]. Within an educational setting these overlaps may not be so mutually exclusive, where children could both select tracks for identification, while simultaneously ‘coloring them in’. A school class could, for example, move fluidly from learning about conservation by collecting prints from a field trip in the wild, to learning about the mathematics of algorithms for identifying prints in the class, and then finally create artistic souvenirs to take home for themselves or their parents.

Acknowledgments. Lizard tracking cards for our work have generously been provided by Subair Siyam, and rat footprints were processed by Guannan Yuan. The authors acknowledge comments by Garry Tee.

References

1. Atkinson, I.A.E.: The spread of commensal species of *Rattus* to oceanic islands and their effects on island avifaunas. In: Conservation of Island Birds, ICBP Technical Publication No. 3, Cambridge, pp. 35–81 (1985)
2. Chapple, D.G., Ritchie, P.A., Daugherty, C.H.: Origin, diversification, and systematics of the New Zealand skink fauna (Reptilia: Scincidae). In: Molecular Phylogenetics and Evolution (in press, 2009)
3. Klette, R.: Scanned and processed footprints of small mammals, reptiles and insects, <http://www.mi.auckland.ac.nz/ScanT>
4. Kutcher, S.R.: Bug art, <http://www.bugartbysteven.com/>
5. Liebenberg, L.: Persistence hunting by modern hunter-gatherers. *Current Anthropology* 47(6), 1017–1025 (2006)
6. de Medinaceli, L., Freed, W.J., Wyatt, R.J.: An index of the functional condition of rat sciatic nerve based on measurements made from walking tracks. *Experimental Neurology* 77(3), 634–643 (1982)
7. Ratz, H.: Identification of footprints of some small mammals. *Mammalia* 61(3), 431–441 (1997)
8. Russell, J.C., Towns, D.R., Clout, M.N.: Review of rat invasion biology: implications for island biosecurity. *Science for Conservation* 286, Department of Conservation, Wellington (2008)
9. Russell, J.C., Hasler, N., Klette, R., Rosenhahn, B.: Automatic track recognition of footprints for identifying cryptic species. *Ecology* (in press, 2009)
10. Warheit, K.I., Forman, J.D., Losos, J.B., Miles, D.B.: Morphological diversification and adaptive radiation: a comparison of two diverse lizard clades. *Evolution* 53(4), 1226–1234 (1999)
11. Yuan, G., Russell, J.C., Klette, R., Rosenhahn, B., Stones-Havas, S.: Understanding tracks of different species of rats. In: *Proc. Int. Conf. Image Vision Computing, New Zealand*, pp. 493–499 (2005)
12. Howald, G., Donlan, C.J., Galván, J.P., Russell, J.C., Parkes, J., Samaniego, A., Wang, Y., Veitch, C.R., Genovesi, P., Pascal, M., Saunders, A., Tershy, B.: Invasive rodent eradication on islands. *Conservation Biology* 21(5), 1258–1268 (2007)

RealSurf – A Tool for the Interactive Visualization of Mathematical Models

Christian Stussak and Peter Schenzel

Martin Luther University Halle-Wittenberg, Institute of Computer Science,
Von-Seckendorff-Platz 1, D – 06120 Halle (Saale), Germany
{stussak,schenzel}@informatik.uni-halle.de

Abstract. For applications in fine art, architecture and engineering it is often important to visualize and to explore complex mathematical models. In former times there were static models of them collected in museums respectively in mathematical institutes. In order to check their properties for esthetical reasons it could be helpful to explore them interactively in 3D in real time. For the class of implicitly given algebraic surfaces we developed the tool REALSURF. Here we give an introduction to the program and some hints for the design of interesting surfaces.

Keywords: Implicit surface, ray tracing, models for fine art.

1 Introduction

Most of the mathematical institutes of "traditional" german universities hold a collection of mathematical models, see e.g. the Steiner surface in figure [1](#).

In the book [\[2\]](#) (edited by Gerd Fischer) the authors described several of such models, often made by gypsum, and gave mathematical explanations of them. Unfortunately this mathematical pearl is out of print today.

Because of their symmetries and their interesting esthetical behaviour this kind of objects are collected in museums. Moreover, there is a large interest in fine art for this extraordinary and sometimes beautiful mathematical constructions. See also the book by Glaeser [\[3\]](#) for relations of mathematical objects in fine art, architecture and engineering.

In recent times there is a strong effort in order to visualize mathematical models on a computer by the aid of methods from computer graphics, in particular (implicit) algebraic surfaces. One basic algorithm for implicitly given algebraic surfaces is the so-called ray tracing, see section [3](#). One of the programs for this kind of visualization is SURF, a ray tracing program for algebraic surfaces, developed by a group around Stephan Endraß, see [\[1\]](#).

Movies for the exploration of certain surfaces were built by a group of Herwig Hauser based on the free renderer Persistence of Vision Raytracer, see Herwig Hauser's homepage [\[5\]](#) for these examples. During 2008, the year of Mathematics in Germany, the Oberwolfach Research Institute of Mathematics provided



Fig. 1. The Steiner surface made of gypsum

an exhibition IMAGINARY combining art and algebraic geometry. The catalogue of the exhibition and many interesting programs for the visualization of mathematical problems are available at the homepage of the exhibition [4]. For the interactive visualization of algebraic surfaces there is the program SURFER, based on S. Endraß's program SURF, which creates high-resolution images of the surfaces in a background process. The renderings appear after a certain rendering time depending on the complexity of the surface.

The aim of this report is to introduce the program REALSURF [6] for an interactive visualization of algebraic surfaces in real time. The program allows the construction of mathematically and esthetically interesting surfaces, for instance for the use in fine art, and their interactive exploration. It is based on a recent technique of programming on the graphics processing unit (GPU) with shader languages. It works well for computers with recent NVIDIA graphics cards.

2 An Example

In the following consider the affine twisted cubic C curve in \mathbb{A}^3 . That is the curve given parametrically by $x = t, y = t^2, z = t^3$. The parametric equation of the tangent surface is

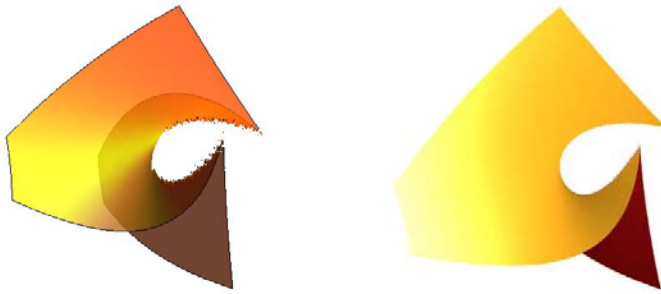
$$x = t + u, \quad y = t^2 + 2tu, \quad z = t^3 + 3t^2u. \quad (1)$$

An easy elimination shows that the tangent surface in its implicit form is given by

$$F = 3x^2y^2 - 4x^3z - 4y^3 + 6xyz - z^2 = 0. \quad (2)$$

The curve C is a singular curve on its tangent variety. This could be a challenge for a presentation in architecture. Figure 2a shows an image of the implicit form of the tangent variety, which was created by MATHEMATICA.

One might see the difficulty for drawing the singular locus, the original curve, correctly. In fact, this requires high numerical stability for the computation of zeros of polynomial equations.



(a) MATHEMATICA 6 rendering

(b) REALSURF rendering

Fig. 2. The tangent surface $3x^2y^2 - 4x^3z - 4y^3 + 6xyz - z^2 = 0$

3 Real Time and Singularities

A popular technique for the visualization of surfaces is based on polygonal meshes. It works well for algebraic surfaces given in parameter form, e.g. BÉZIER surfaces. This method does not work correctly for singularities. It could be possible that a singularity is not exhibited in a polygonal mesh, and hence does not occur for the visualization.

In Computer Graphics ray tracing is an appropriate technique in order to visualize scenes with complex details like singularities. A popular ray tracer is POV-RAY. Ray tracing techniques are also implemented in SURF. Typically ray tracing is a time consuming process, which stays in contrast to the requirement of real time visualization with interactions. An overview of the ray tracing process is given in the next subsection.

3.1 Ray Tracing

In our model of ray tracing there is a fixed world coordinate system and a scene with mathematically described objects. For each pixel of the screen a ray is sent from the eye into the scene and the nearest one among the hit points of the ray with all scene objects is computed. Afterwards the illumination at the object with the nearest hit point is calculated, which can introduce new rays in order to simulate complex visual phenomena like reflection, refraction or shadowing. To obtain fine details this iteration may require minutes respectively hours of computing time. The particular case, where only primary rays are considered, is also known as ray casting.

Most recently there is a new hardware development to allow computations and programming on the graphics processing unit (GPU) using specially designed programming languages, typically called shading languages due to their computer graphics background. Examples of shading languages are Cg, GLSL and HLSL. Prior to the execution of the program the graphics driver translates the shader code into machine instructions for the graphics card. Because of the

multiprocessor concepts of recent graphics cards this procedure ensures an essential increase in computing speed as required for instance for computer games and complex visualizations.

4 RealSurf

REALSURF is a program for the interactive visualization of (implicit) algebraic surfaces. It is based on the hardware development as mentioned above and uses the OPENGL Shading Language (GLSL). Figure 3 shows a screen shot of the program for the exploration of Barth's surface of degree 10.

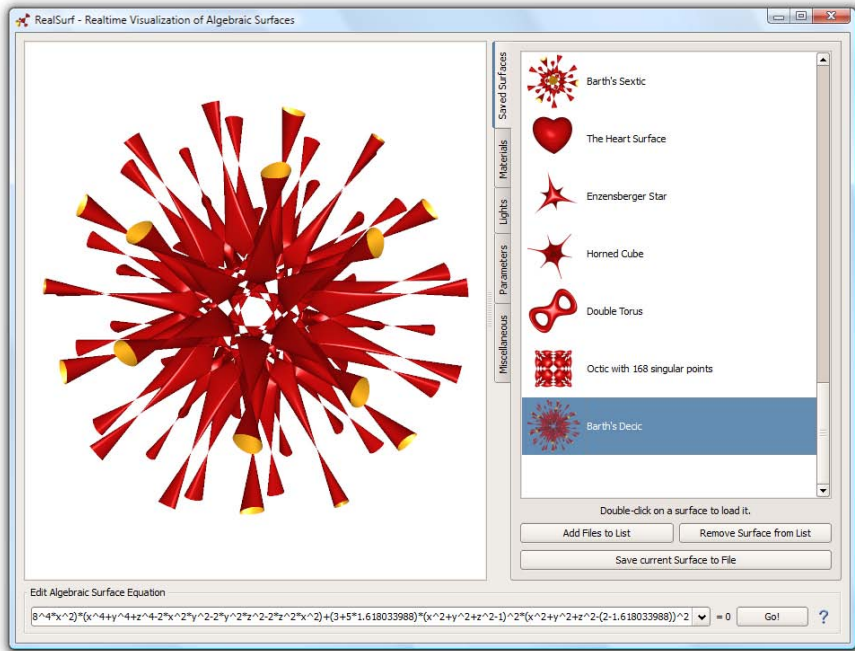


Fig. 3. The main window of RealSurf, including a gallery of predefined surfaces

Besides the shader programming techniques we had to consider several numerical aspects in order to calculate the zeros of algebraic equations in the ray-surface intersection step. Due to the numerical instabilities in the neighborhood of singularities and the low precision of calculations on graphics hardware, we first separate the roots of the polynomial equation and afterwards refine the isolating intervals to a sufficient precision for the visualization. For the details see [6].

Features of RealSurf

REALSURF allows to scale, rotate and translate the surface in real time by mouse actions. A collection of classical surfaces is included, own surface equations can be entered – saving and loading surfaces is supported as well.

Constants in the equation can be replaced by parameters. This procedure permits e.g. interactive deformations of surfaces. As an example we investigate the cubic surface

$$F(x, y, z) = x^2 + y^2 + z^2 - \alpha xyz - 1 = 0 \quad (3)$$

with the parameter α , see Figure 4. The program allows an interactive visualization of the surface by a continuous change of the parameter. In particular for $\alpha = 2$ it yields the well known Cayley cubic.

Furthermore several scene parameters like lighting, surface materials and objects to clip the surface against are adjustable in REALSURF. The rendered images can be saved as PNG files for use in publications and presentations.

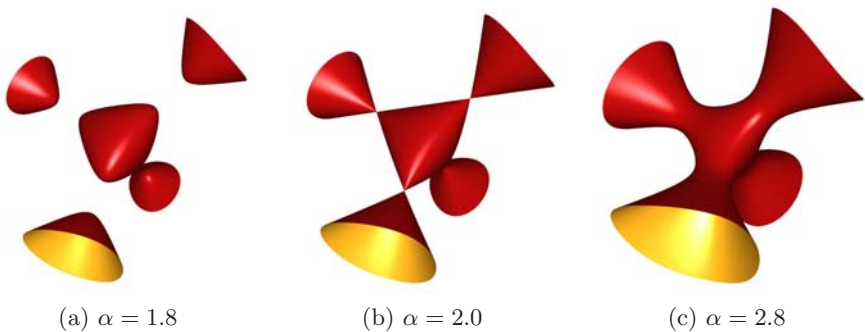


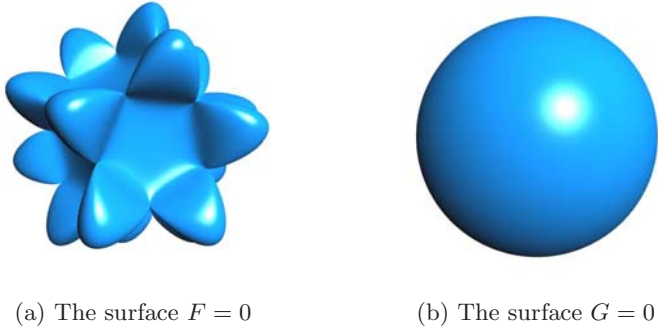
Fig. 4. The cubic for different values of α

5 Some Recipes for Creating Interesting Figures

From an artists point of view the mathematical properties of an algebraic surface may not be very interesting. Nevertheless one can start with some higher order surfaces like the Barth sextic or with variants of them like the surface in figure 5a. One possibility is to remove some terms from the complex equation, add new terms or modify coefficients and parameters.

Visually more appealing surfaces can often be constructed by playing with the level sets of combinations of two or more surfaces. For example one can multiply a complex surface $F = 0$ and a simple surface $G = 0$ like in figure 5 and add a small offset to generate the level set surface $F \cdot G + \gamma = 0$, which gives a slightly disturbed union of both surfaces, thus "melting" both surfaces together. An example is shown in figure 6.

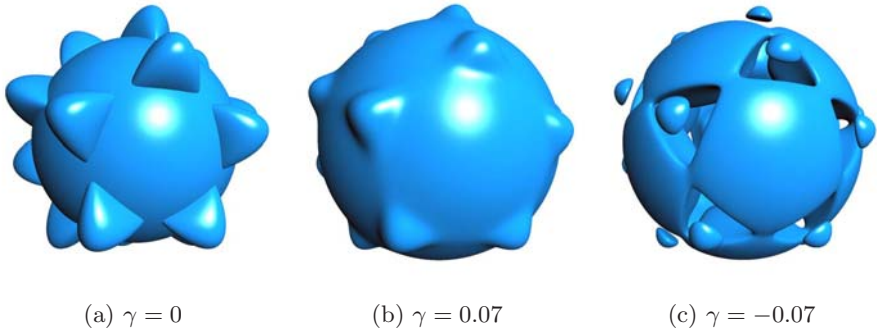
Another form of combining two surfaces is to calculate their intersection curve, which is given by $F^2 + G^2 = 0$. An example is shown in figure 7a. The infinitely thin intersection curve is thickened by the offset γ , resulting in the equation $F^2 + G^2 - \gamma = 0$, which often gives interesting surfaces like in figure 7.



(a) The surface $F = 0$

(b) The surface $G = 0$

Fig. 5. Surfaces used in the following examples: (a) a variation of the Barth sextic with 30 cusps ($F = 4(\alpha^2 x^2 - y^2)(\alpha^2 y^2 - z^2)(\alpha^2 z^2 - x^2) - (1 + 2\alpha)(x^2 + y^2 + z^2 - 1)^3 = 0$, $\alpha = \frac{1}{2}(1 + \sqrt{5})$), (b) a sphere ($G = x^2 + y^2 + z^2 - \beta^2 = 0$)

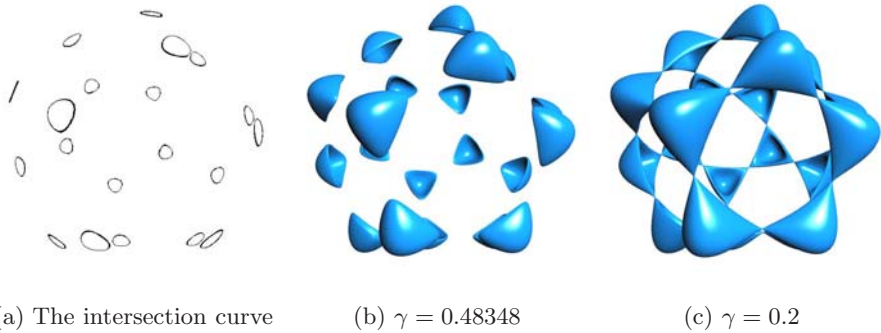


(a) $\gamma = 0$

(b) $\gamma = 0.07$

(c) $\gamma = -0.07$

Fig. 6. Level sets of the surface $F \cdot G + \gamma = 0$ for $\alpha = \frac{1}{2}(1 + \sqrt{5})$ and $\beta = \sqrt{1.3}$



(a) The intersection curve

(b) $\gamma = 0.48348$

(c) $\gamma = 0.2$

Fig. 7. Starting with the intersection curve $F^2 + G^2 = 0$ of two surfaces and considering its level sets $F^2 + G^2 - \gamma = 0$ often leads to very nice surfaces. The images above were created with $\alpha = \frac{1}{2}(1 + \sqrt{5})$ and $\beta = \sqrt{2}$.



(a) Cassini surface



(b) Enzensberger star



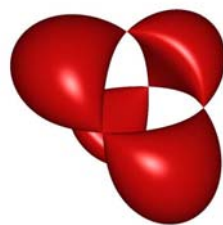
(c) double torus



(d) Stagnaro quintic



(e) Kummer quartic



(f) tetrahedral symmetry



(g) heart surface



(h) Boy surface



(i) Chmutov octic

Fig. 8. For further inspiration we present some images of well known algebraic surfaces. Equations of the above surfaces can be found in [7], for example.

6 Concluding Remarks

The program REALSURF allows the interactive exploration of implicit given algebraic surfaces with singularities in real time. It is based on programming with shading languages and currently renders most surfaces up to degree 13 correctly and in real time, if the equation is not too complex.

It is available for Microsoft Windows and requires recent NVIDIA graphics hardware (GeForce 6 series and up). Upon request it is freely available via [6].

The program allows the investigation of mathematical models in an interactive way. In addition one might be able to create new esthetical objects for further use in fine art, architecture a.o. The advantage over static models is the ability to modify a surface by parameters and to obtain a 3D imagination of the model.

References

- [1] Endraß, S.: Surf (2001), <http://surf.sourceforge.net>
- [2] Fischer, G. (ed.): Mathematische Modelle. Vieweg (1986)
- [3] Glaeser, G.: Geometrie und ihre Anwendungen in Kunst, Natur und Technik, 2nd edn. Elsevier, Spektrum Akademischer Verlag (2007)
- [4] Greuel, G.M.: Imaginary – Mit den Augen der Mathematik (2008), <http://www.imaginary2008.de>
- [5] Hauser, H.: Animation von Flächen (2005), <http://homepage.univie.ac.at/herwig.hauser/>
- [6] Stussak, C.: Echtzeit-Raytracing algebraischer Flächen auf der GPU. Diploma thesis, Martin Luther University Halle-Wittenberg (2007), <http://realsurf.informatik.uni-halle.de>
- [7] Weisstein, E.W.: Algebraic surface (2009), <http://mathworld.wolfram.com/AlgebraicSurface.html>

Creating Wheel-Thrown Potteries in Digital Space

Gautam Kumar, Naveen Kumar Sharma, and Partha Bhowmick

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur, India
gautamkumar.iit@gmail.com, nkrsharma@gmail.com, bhowmick@gmail.com
<http://www.facweb.iitkgp.ernet.in/~pb>



A set of digital potteries generated by our algorithm

Abstract. This paper introduces a novel technique to create digital potteries using certain simple-yet-efficient techniques of digital geometry. Given a *digital generatrix*, the proposed wheel-throwing procedure works with a few primitive integer computations only, wherein lies its strength and novelty. The created digital surface is *digitally connected and irreducible* when the generatrix is an irreducible digital curve segment, which ensures its successful rendition with a realistic finish, whatsoever may be the zoom factor. Thick-walled potteries can also be created successfully and efficiently using non-monotone digital generatrices to have the final product ultimately resembling a real-life pottery.

Keywords: Digital geometry, digital surface of revolution, generatrix, geometry and art, potteries.

1 Introduction

Creating wheel-thrown potteries is a millennium-old artistry, which, with the passing time, has gained an enduring popularity with today's state-of-the-art

ceramic technology [3,14]. Hence, with the proliferation of digitization techniques in our computerized society, digital creation and artistic visualization of potteries is a call of the day. The existing graphic tools are mostly based on complex trigonometric procedures involving computations in the real space, which have to be tuned properly to suit the discrete nature of the 3D digital space [6,8]. For, the method of *circularly sweeping* a shape/polygon/polyline/generating curve about the axis of revolution is done in discrete steps, which requires a discrete approximation. And choosing k equispaced angles about the y -axis requires the following transformation matrices for a real-geometric realization:

$$M_i = \begin{pmatrix} \cos \theta_i & 0 & \sin \theta_i & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \theta_i & 0 & \cos \theta_i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{where, } \theta_i = 2\pi i/k, i = 0, 1, \dots, k-1 \text{ [8].}$$

Clearly, in order to generate circles of varying radii describing the surface of revolution, the number of steps, k , becomes a concerning issue.

Existing methods for modeling potteries mostly start with an initial cylindrical ‘clay’ piece and use deformation techniques, as described in [9,10]. The deformation is based either on shifting the individual points horizontally or on using devices that accept feedback by means of touch-sensation at multiple points [13]. As a result, the user has a direct control over the geometry of the shape to be constructed, which, however, poses difficulties to represent the surface mathematically or to store it efficiently. Some of these practices may be seen in [6,7,12].

Features of Our Method: The proposed algorithm of *digital wheel-throwing* works purely in the digital domain and banks on a few primitive integer computations only, wherein lies its strength and novelty. Given a digital generatrix as an irreducible digital curve segment, the digital surface produced by it is both connected and irreducible in digital-geometric sense. This, in turn, ensures its successful rendition with a realistic finish that involves conventional processing of quad decomposition, texture mapping, illumination, etc. Another distinguishing feature is that the whole surface can be represented by and stored as a sequence of few control points and the axis of rotation. For, storing these control points in sequence is sufficient to reconstruct the digital generatrix, if it is defined as a sequence of cubic B-splines. The method is robust and efficient, and guarantees easy implementation will all relevant graphic features. The rendered visualization is found to be absolutely free of any bugs or degeneracies, whatsoever may be the zoom factor. Producing a monotone or a non-monotone digital surface of revolution is feasible without destroying its digital connectivity and irreducibility by respective input of a monotone or a non-monotone digital generatrix. To create exquisite digital products resembling on-the-shelf potteries in toto, a double-layered generatrix can be used, which is its ultimate benefit.

2 Proposed Method

A digital curve segment $\mathcal{G} := \{p_i : i = 1, 2, \dots, n\}$ is a finite sequence of digital points (i.e., points with integer coordinates) [11], in which two points $p(x, y) \in \mathcal{G}$

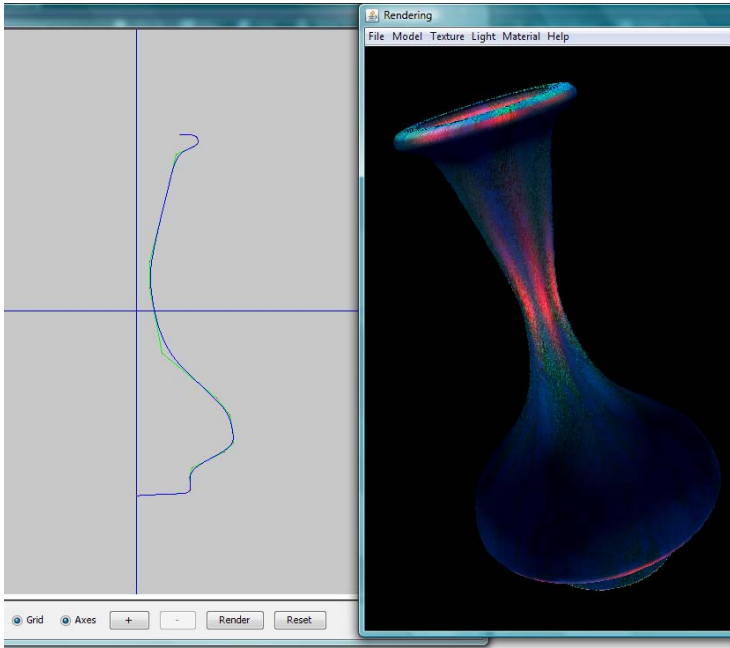


Fig. 1. A snapshot of our algorithm. Left: digital generatrix. Right: a flowerpot.

and $p'(x', y') \in \mathcal{G}$ are 8-neighbors of each other, if and only if $\max(|x - x'|, |y - y'|) = 1$. To ensure that \mathcal{G} is simple, irreducible, and open-ended, each point $p_i (i = 2, 3, \dots, n - 1)$ should have exactly two neighbors, and each of p_1 and p_n should have one, from \mathcal{G} . Thus, the chain code of a point $p_i \in \mathcal{G}$ w.r.t. its previous point $p_{i-1} \in \mathcal{G}$ is given by $c_i \in \{0, 1, 2, \dots, 7\}$ [5]. In order to generate a digital surface of revolution, \mathcal{S} , we consider a digital curve segment \mathcal{G} as the digital generatrix, as shown in Fig. 1. The input digital generatrix may be taken either as a sequence of chain codes or as a sequence of control points. Given $m (\geq 4)$ control points as input, the digital generatrix is considered as the digital irreducible curve segment that approximates the sequence of $m - 3$ uniform non-rational cubic B-spline segments interpolating the sequence of these control points [4, 8]. The reasons for B-spline interpolation of the generatrix is its unique characteristic of having both parametric and geometric continuities, which ensure the smoothness and the optimal exactness of the fitted curve against the given set of control points. In addition, a local change (insertion/deletion/repositioning) of control point(s) has only a local effect on the shape of the digital generatrix (and a local effect on the generated surface, thereof). This, in fact, simulates the local effect of a “potter’s hand” rolling and maneuvering the “clay” on his rotating wheel (Fig. 2).

The sets $N_4(p) := \{(x', y') : |x - x'| + |y - y'| = 1\}$ and $N_8(p) := \{(x', y') : \max(|x - x'|, |y - y'|) = 1\}$ corresponding to a point $p(x, y) \in \mathbb{Z}^2$ define the respective 4-neighborhood (4N) and 8-neighborhood (8N) of p . Similarly, if

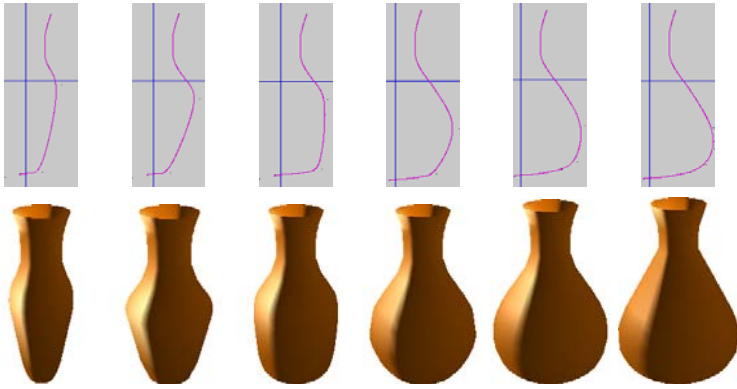


Fig. 2. Simulating the local effect of a “potter’s hand” by inserting control points

$p(x, y, z) \in \mathbb{Z}^3$, then $N_6(p) := \{(x', y', z') : |x - x'| + |y - y'| + |z - z'| = 1\}$ and $N_{26}(p) := \{(x', y', z') : \max(|x - x'|, |y - y'|, |z - z'|) = 1\}$ define the respective 6N and 26N of p . Each point in $N_k(p)$ is said to be a k -neighbor of p ; $k \in \{4, 8\}$ if $p \in \mathbb{Z}^2$ and $k \in \{6, 26\}$ if $p \in \mathbb{Z}^3$. We consider $k = 8$ and $\bar{k} = 4$ in \mathbb{Z}^2 , and $k = 26$ and $\bar{k} = 6$ in \mathbb{Z}^3 . Thus, in a (finite) digital set $S \subset \mathbb{Z}^2(\mathbb{Z}^3)$, two points $p \in S$ and $q \in S$ are k -connected if and only if there exists a sequence $\langle p := p_0, p_1, \dots, p_n := q \rangle \subseteq S$ such that $p_i \in N_k(p_{i-1})$ for $1 \leq i \leq n$. The digital set S is said to be a *connected digital set* if and only if every pair of points in S is k -connected. If a connected digital set S be such that $\bar{S} := \mathbb{Z}^2 \setminus S(\mathbb{Z}^3 \setminus S)$ contains at least two points, $\bar{p} \in \bar{S}$ and $\bar{q} \in \bar{S}$, which are not \bar{k} -connected, then S is said possess a *hole*. A hole H is a finite and maximal subset of \bar{S} such that every pair of its points are \bar{k} -connected and none of its points is \bar{k} -connected with any point from $\bar{S} \setminus H$. If $S \subset \mathbb{Z}^2(\mathbb{Z}^3)$ be a connected digital set containing exactly one hole, namely H , then S forms a *closed digital curve (surface)*, of possibly arbitrary thickness. A closed digital curve (surface) S is *irreducible* (i.e., of unit thickness) if and only if exclusion of *any* point p from S (and its inclusion to \bar{S}) gives rise to a \bar{k} -connected path from each point of H to each point of $\bar{S} \setminus H$.

If L be a real line that partitions an irreducible and closed digital curve S into two or more (k -connected) components such that all points in each component are on the same side of L or lying on L , and no two components share a common point, then each such component of S becomes an *open digital curve* (irreducible segment). Partitioning (iteratively) of an open digital curve segment, in turn, gives rise to more and more smaller open segments. Clearly, if the closed digital curve S be such that any horizontal (vertical) line L_y (L_x) always partitions S into at most two components, then S is a y -monotone (x -monotone) digital curve. Similarly, partitioning an irreducible and closed digital surface S by a plane Π produces two or more components such that all points in each component are on the same side of Π or lying on Π , and no two components share a common point. Each component obtained by partitioning S with Π is an *open digital surface* (irreducible). If the closed digital surface S be such that any plane Π_y

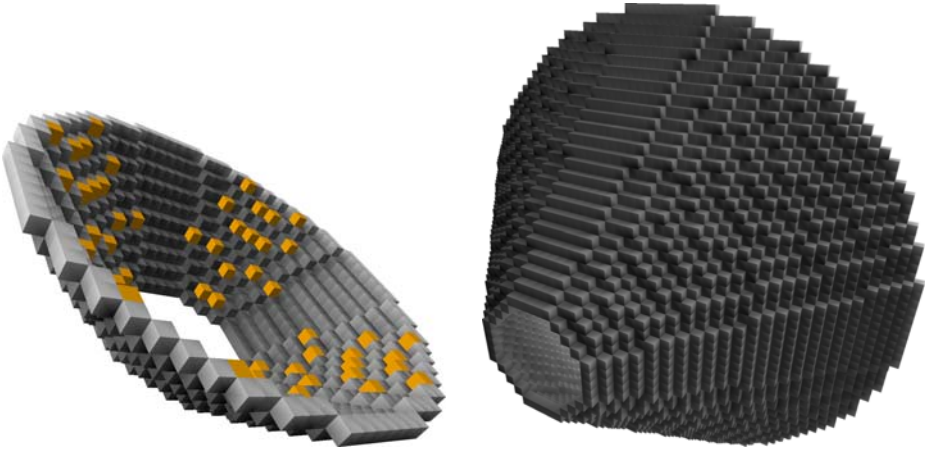


Fig. 3. Shown in yellow are the missing voxels (left), which are detected and included to successfully create the digitally connected and irreducible surface (right)

(Π_x, Π_z) orthogonal to y -axis (x -, z -axis) always partitions S into at most two components, then S is a y -monotone (x -, z -monotone) digital surface. It may be mentioned here that, if the generatrix be an open, irreducible, and y -monotone digital curve segment, C , then the digital surface of revolution, S , produced by our algorithm is also a y -monotone digital surface.

2.1 Algorithm for Digital Wheel-Throwing

We generate a wheel-thrown piece by revolving the digital generatrix $\mathcal{G} := \{p_i : i = 1, 2, \dots, n\}$ about an axis of revolution given by $\alpha : \langle z = -c, x = a \rangle$, where a and c are two positive integers. In order to achieve this, for each digital point $p_i \in \mathcal{G}$, we construct the digital circle $\mathcal{C}^{\mathbb{Z}}_i$ by revolving p_i around the specified axis of revolution, α . Depending on whether the radius and the center of a digital circle are real or integer values, several definitions of digital circles may be seen in the literature [2]. If the radius be $r \in \mathbb{Z}^+$ and the center be $c = o(0, 0)$, then the first octant of the corresponding digital circle is given by $\mathcal{C}^{\mathbb{Z}}_1(o, r) = \{(i, j) \in \mathbb{Z}^2 : 0 \leq i \leq j \leq r \wedge |j - \sqrt{r^2 - i^2}| < \frac{1}{2}\}$, and so the entire digital circle is $\mathcal{C}^{\mathbb{Z}}(o, r) = \{(i, j) : \{|i|, |j|\} \in \mathcal{C}^{\mathbb{Z}}_1(o, r)\}$. With $p(i_p, j_p) \in \mathbb{Z}^2$ as center, $\mathcal{C}^{\mathbb{Z}}(p, r) = \{(i + i_p, j + j_p) : (i, j) \in \mathcal{C}^{\mathbb{Z}}(o, r)\}$.

As \mathcal{G} is irreducible in nature, two consecutive points p_i and p_{i+1} have their distances, namely r_i and r_{i+1} , measured from α , differing by at most unity. If their distances from α are same, then it is easy to observe that the surface $\mathcal{C}^{\mathbb{Z}}_i \cup \mathcal{C}^{\mathbb{Z}}_{i+1}$ is digitally connected and irreducible. The problem arises when p_i and p_{i+1} have their respective distances from α differing by unity. Then there may arise some missing voxels trapped between $\mathcal{C}^{\mathbb{Z}}_i$ and $\mathcal{C}^{\mathbb{Z}}_{i+1}$, which results in digital disconnectedness in the surface $\mathcal{C}^{\mathbb{Z}}_i \cup \mathcal{C}^{\mathbb{Z}}_{i+1}$, as shown in Fig. 3. Detection

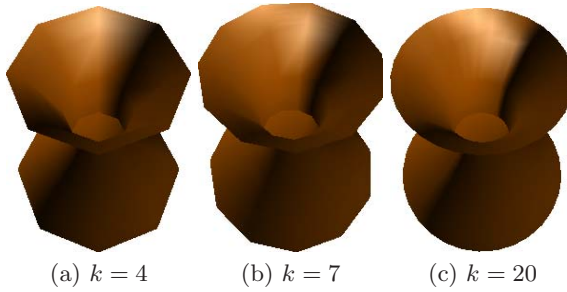


Fig. 4. Approximating generating digital circles by regular $2k$ -gons

of these missing voxels is performed to achieve a digitally connected surface, namely $\mathcal{S}_G := \mathcal{C}^{\mathbb{Z}}_1 \cup \mathcal{C}^{\mathbb{Z}}_2 \cup \dots \cup \mathcal{C}^{\mathbb{Z}}_n$, as follows.

Case 1 ($r_{i+1} > r_i$): While generating the digital circle $\mathcal{C}^{\mathbb{Z}}_{i+1}$ parallel to the zx -plane corresponding to the point $p_{i+1} \in \mathcal{G}$, there is either an east (E) transition or south-east (SE) transition from the current point $q(x, y, z)$ in Octant 1 ($z \leq x \leq r_{i+1}$) [4]. If we take the respective projections, $\mathcal{C}^{\mathbb{Z}'}_i$ and $\mathcal{C}^{\mathbb{Z}'}_{i+1}$, of $\mathcal{C}^{\mathbb{Z}}_i$ and $\mathcal{C}^{\mathbb{Z}}_{i+1}$ on the zx -plane, then $\mathcal{C}^{\mathbb{Z}'}_i$ and $\mathcal{C}^{\mathbb{Z}'}_{i+1}$ become concentric with their radii differing by unity. As $r_{i+1} > r_i$, the cumulative run-length $\sum_{j=0}^k \lambda_{i+1}^{(j)}$ of $\mathcal{C}^{\mathbb{Z}'}_{i+1}$ in Octant 1 is either same as the corresponding cumulative run-length $\sum_{j=0}^k \lambda_i^{(j)}$ of $\mathcal{C}^{\mathbb{Z}'}_i$ or greater by unity [2]. Hence, a “missing voxel” between $\mathcal{C}^{\mathbb{Z}}_i$ and $\mathcal{C}^{\mathbb{Z}}_{i+1}$ is formed only if there is a transition towards SE (a change in run, thereof) from a point/pixel in $\mathcal{C}^{\mathbb{Z}'}_{i+1}$ giving rise to a “missing pixel” between $\mathcal{C}^{\mathbb{Z}'}_i$ and $\mathcal{C}^{\mathbb{Z}'}_{i+1}$. We detect such missing pixels by determining whether or not there is a “miss” during each SE transition for $\mathcal{C}^{\mathbb{Z}'}_{i+1}$. More precisely, if the point next to the current point $q'(x, z) \in \mathcal{C}^{\mathbb{Z}'}_{i+1}$ in Octant 1 is $(x - 1, z + 1)$ and the point $(x - 1, z)$ does not belong to $\mathcal{C}^{\mathbb{Z}'}_i$, then we include the point $(x - 1, y, z)$ in $\mathcal{C}^{\mathbb{Z}}_{i+1}$ between $(x, y, z) \in \mathcal{C}^{\mathbb{Z}}_{i+1}$ and $(x - 1, y, z + 1) \in \mathcal{C}^{\mathbb{Z}}_{i+1}$.

Case 2 ($r_{i+1} < r_i$): If the point next to the current point $q'(x, z) \in \mathcal{C}^{\mathbb{Z}'}_{i+1}$ in Octant 1 is $(x - 1, z + 1)$ and the point $(x, z + 1)$ does not belong to $\mathcal{C}^{\mathbb{Z}'}_i$, then we include the point $(x, y, z + 1)$ in $\mathcal{C}^{\mathbb{Z}}_{i+1}$ between $(x, y, z) \in \mathcal{C}^{\mathbb{Z}}_{i+1}$ and $(x - 1, y, z + 1) \in \mathcal{C}^{\mathbb{Z}}_{i+1}$.

2.2 Creating Textured/Painted Potteries

As explained in Sec. 2.1, the surface \mathcal{S}_G is generated from the digital generatrix \mathcal{G} as an ordered set of voxels. For each digital circle $\mathcal{C}^{\mathbb{Z}}_i$ corresponding to each voxel $p_i \in \mathcal{G}$, the voxels inclusive of the missing voxels are generated in a definite order, starting from Octant 1 and ending at Octant 8. All the circles, namely $\mathcal{C}^{\mathbb{Z}}_1, \mathcal{C}^{\mathbb{Z}}_2, \dots, \mathcal{C}^{\mathbb{Z}}_n$, in turn, are also generated in order. The above ordering to represent a wheel-thrown piece helps to map textures in a straightforward way. We map adjacent and equal-area rectangular parts from the texture image to

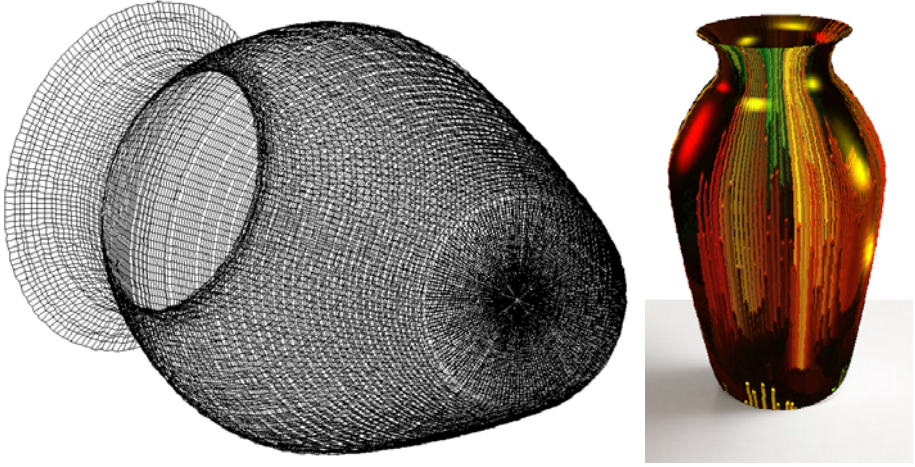


Fig. 5. Quad decomposition of a wheel-thrown digital vase for texturing

each of the quads. To do this, each digital circle is approximated as a regular polygon. Clearly, there lies a trade-off between the number of vertices of the approximate polygon and the rendering speed, since a coarse approximation gives a polyhedral effect (Fig. 4). The nature of approximation error has been studied in a recent work [1], which shows that if we have polygons with $k \geq 20$ or so, then for $r \leq 100$, we have 2% or less error. For higher radii, of course, we should increase k accordingly to have the desired error.

In order to map the adjacent square portions of a texture image to the corresponding adjacent quads representing the digital surface of a pottery, the number of sides, χ , in the polygonal approximation is kept same for all the digital circles corresponding to the digital points in \mathcal{G} . To do this, the distances of all the points in \mathcal{G} from the axis of revolution, α , are computed, from which their maximum, r_{\max} , is obtained. We use r_{\max} to determine the value of χ . Since the distance between two adjacent digital points in \mathcal{G} is either 1 or $\sqrt{2}$, we determine the value of χ by approximating the circle of radius r_{\max} with a regular polygon having side-length d . Thus, $\chi = \left\lfloor \frac{\pi}{\sin^{-1}\left(\frac{d}{2r_{\max}}\right)} + \frac{1}{2} \right\rfloor$.

Now, for each point on \mathcal{G} , we approximate the circle of revolution with a regular polygon with χ sides. For every pair of adjacent points in \mathcal{G} , we join the corresponding vertices of the approximate polygon corresponding to the circle of revolution. In this way, we obtain a simple quad-decomposition of the surface of revolution. For a faster rendering, we can approximate the digital generatrix/B-spline by selecting alternate points (or 1 in every $d > 1$). (In our experiments, we have taken $d = 4$.) In this way, the digital continuity of the original image is easily maintained on the texture map, with satisfactory results. It may be noted that, only for such an artistic finish of the digital pottery, floating-point computations have been resorted to. Fig. 5 demonstrates the quad

decomposition of a wheel-thrown vase and the resultant texture mapping. To create realistic potteries (one shown in the front page) having thick walls, we need to supply (connected and irreducible) digital generatrices having two parallel segments — one for the outer wall and another for the inner wall including the rim top.

3 Experiments and Results

We have developed the software in Java3D™ API, version 1.5.2. Snapshots of a few potteries are given in Fig. 5 and in the front page, and some statistical figures and CPU times are presented in Table 1. The total CPU time required to create a digital pottery is for constructing the digital surface \mathcal{S}_G from a given digital generatrix, \mathcal{G} , and for quad-decomposition followed by texture mapping coupled with necessary rendering. The number of voxels constituting \mathcal{S}_G not only increases with the length of \mathcal{G} and the distance of \mathcal{G} from the axis of revolution, α , but it also depends on the shape of \mathcal{G} . For, if the each pair of consecutive points comprising \mathcal{G} has the radius difference unity, then the number of missing voxels increases, which, in turn, increases the total number of voxels defining \mathcal{S}_G .

Table 1. Some statistical figures and CPU times for some digital potteries

#Control Points	#Points in original generatrix	#Points in approximate generatrix	#Sides of polygon	#Quads	Time (milliseconds.)
4	103	26	163	4075	475
7	227	57	185	10360	805
10	508	127	87	10962	792
15	599	150	113	16837	1009
20	592	148	186	27342	1624

4 Conclusion and Future Possibilities

We have shown how wheel-throwing can be done using certain efficient techniques of digital geometry for creating various digital potteries, which satisfy both mathematical and aesthetic criteria. Future prospects include generating irregular/bumpy surfaces and sub-surfaces of revolution, and in defining various morphological/set-theoretic operations on digital surfaces of revolution so as to generate more interesting potteries out of the “digital wheel”.

References

1. Bhowmick, P., Bhattacharya, B.B.: Real polygonal covers of digital discs — Some theories and experiments. *Fundamenta Informaticae* 90, 487–505 (2009)
2. Bhowmick, P., Bhattacharya, B.B.: Number-theoretic interpretation and construction of a digital circle. *Discrete Applied Mathematics* 156(12), 2381–2399 (2008)
3. Bryant, V.: Web tutorials for potteries (2004), <http://www.victor.bryant.hemscott.net>

4. Foley, J.D., et al.: *Computer Graphics—Principles & Practice*. Addison-Wesley, Reading (1993)
5. Freeman, H.: On the encoding of arbitrary geometric configurations. *IRE Trans. Electronic Computers* EC-10, 260–268 (1961)
6. Galyean, T.A., Hughes, J.F.: Sculpting: An interactive volumetric modeling technique. *Computer Graphics (Proc. ACM Siggraph)* 25(4), 267–274 (1991)
7. Han, G., et al.: Virtual Pottery Modeling with Force Feedback Using Cylindrical Element Method. In: *Proc. Intl. Conf. Next-Gen., Computing (ICON-C)*, pp. 125–129 (2007)
8. Hill, F.S., Kelley, S.M.: *Computer Graphics Using OpenGL*. Prentice Hall, Englewood Cliffs (2007)
9. Kameyama, K.: Virtual clay modeling system. In: *Proc. ACM Symp. Virtual Reality Software and Technology (VRST)*, pp. 197–200 (1997)
10. Korida, K., et al.: An Interactive 3D Interface for a Virtual Ceramic Art Work Environment. In: *Proc. Intl. Conf. Virtual Systems & Multimedia (VSMM)*, p. 227 (1997)
11. Klette, R., Rosenfeld, A.: *Digital Geometry: Geometric Methods for Digital Picture Analysis*. Morgan Kaufmann, San Francisco (2004)
12. Lee, J., et al.: Haptic Pottery Modeling Using Circular Sector Element Method. In: Ferre, M. (ed.) *EuroHaptics 2008*. LNCS, vol. 5024, pp. 668–674. Springer, Heidelberg (2008)
13. Ueda, E., et al.: Virtual Clay Modeling System Using Multi-Viewpoint Images. In: *Proc. 5 Intl. Conf. 3-D Digital Imaging & Modeling (3DIM)*, pp. 134–141 (2005)
14. Woody, E.S.: *Pottery on the Wheel*. Allworth Press, New York (2008); Original publisher: Farrar, Straus and Giroux (1975)

Low-Level Image Processing for Lane Detection and Tracking

Ruyi Jiang¹, Mutsuhiro Terauchi²,
Reinhard Klette³, Shigang Wang¹, and Tobi Vaudrey³

¹ Shanghai Jiao Tong University, Shanghai, China
{jiangrui, wangshigang}@sjtu.edu.cn

² Hiroshima International University, Japan
mucha@he.hirokoku-u.ac.jp

³ The University of Auckland, Auckland, New Zealand
{r.klette, t.vaudrey}@auckland.ac.nz

Abstract. Lane detection and tracking is a significant component of vision-based driver assistance systems (DAS). Low-level image processing is the first step in such a component. This paper suggests three useful techniques for low-level image processing in lane detection situations: bird's-eye view mapping, a specialized edge detection method, and the distance transform. The first two techniques have been widely used in DAS, while the distance transform is a method newly exploited in DAS, that can provide useful information in lane detection situations. This paper recalls two methods to generate a bird's-eye image from the original input image, it also compares edge detectors. A modified version of the Euclidean distance transform called *real orientation distance transform* (RODT) is proposed. Finally, the paper discusses experiments on lane detection and tracking using these technologies.

Keywords: Lane detection and tracking, DAS, bird's-eye view, distance transform.

1 Introduction

Lane detection plays a significant role in driver assistance systems (DAS), as it can help estimate the geometry of the road ahead, as well as the lateral position of the ego-vehicle on the road [7]. Lane detection is used in intelligent cruise control systems, for lane departure warning, road modelling, and so on. Typically, lane detection and tracking are used for localizing lane boundaries in given road images. In general, a procedure for lane detection includes low-level image processing as pre-processing, lane boundary detection, and some post-processing.

The first step in lane detection is low-level image processing (LLIP), which deals with the input image [from the camera(s) installed on the ego-vehicle] and generates useful information for the detection part.

For edge-based lane detection, three relative stages are to be considered in LLIP. First, a transformation of the input image as lane detection has been conducted (to a large extent) directly on the input image. However, a proper transformation (e.g., a homographic transform) of the input image, prior to further analysis, will facilitate and improve the accuracy of lane detection. Secondly, an edge detection method is

performed. All-purpose edge operators (e.g., Sobel, Canny) are also applied for lane detection, some specially designed operators will fit better to the properties of lane marks, which are to be detected. Third, some other operations on the edge map will be discussed below.

Complete LLIP will not always include these three stages according to various lane detection situations and applied methods. This paper will discuss three useful technologies that can be applied as LLIP in lane detection and tracking. Bird’s-eye view mapping has been widely used in lane detection [115] and robot navigation [4], as it can provide a top-down view through a homographic transform. A specialized edge detector, as introduced in [1], is recalled and compared with other gradient-based edge operators. Furthermore, this paper fully exploits the advantages of distance transforms for lane detection as a powerful way to represent relevant information.

This work is organized as follows: Section 2 discusses bird’s-eye view mapping, and provides two methods for such a transformation. Section 3 describes an edge detection method. Section 4 recalls distance transform and discusses its useful properties for lane detection. Experimental results of lane detection and tracking are presented in Section 5 using LLIP technologies as introduced in this paper.

2 Bird’s-Eye View Mapping

Generally, detecting lanes directly on the input image will waste no time over transforming the input image into some other space. However, as introduced in [1], at least two disadvantages will come with apparent perspective effects: non-constant lane mark width, and different distances (see Fig. 1). A removal of such perspective effects will greatly facilitate lane detection. The main function of a bird’s-eye view mapping is that a remapped image represents a (scaled) view from the top towards an assumed planar ground manifold. From this remapped view, perspective effects are approximately removed (i.e., approximate with respect to the assumed planarity). Two methods are briefly reviewed here to generate a bird’s-eye image from the input image.

2.1 Warp Perspective Mapping

As in [5], a four-points correspondence can be used for the mapping from the input image into the bird’s-eye image. The mapping is achieved by selecting four points P_W on the ground plane when calibrating the ego-vehicle’s camera(s), and by using the

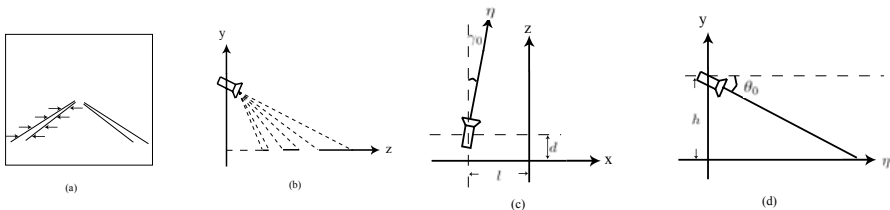


Fig. 1. Perspective effects in the original input image: (a) Non-constant lane mark width, (b) Variation of distances to different pixels. The model of IPM: (c) The xz plane in space \mathbb{W} , (d) The $y\eta$ plane.

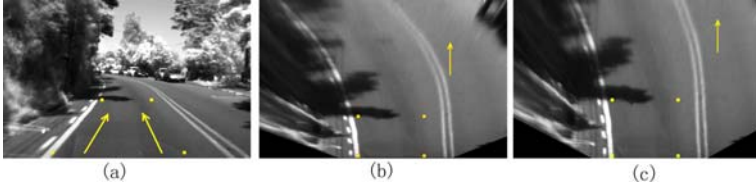


Fig. 2. (a) Input Image. (b) and (c) are bird’s-eye images using a WPM but based on different distance definitions. Four-point correspondence (points shown in yellow) is established by calibration; the driving direction is indicated by the arrow.

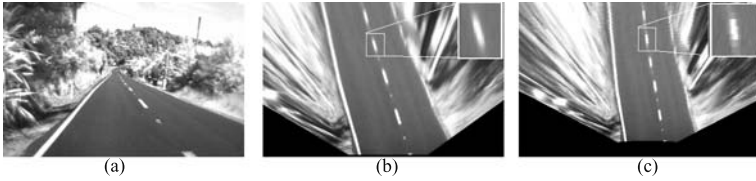


Fig. 3. Bird’s-eye view using warp perspective mapping and inverse perspective mapping. (a) The original input image. (b) Bird’s-eye view using WPM. (c) Bird’s-eye view using IPM. In general, those views generated by WPM appear to be more blurred, and IPM is more pixelated.

planar ground manifold assumption. P_W in the input image will be viewed as P_I . The mapping from P_I to P_W can be achieved by an affine matrix A as $P_W = A \cdot P_I$. The calculation of A using the above equation can be easily achieved. Then, the whole input image can be mapped, pixel-by-pixel, using A . One main benefit of this warp perspective mapping (WPM) is that the used distance scale can be adjusted by selecting different sets of four corresponding points (i.e., by scaling the “length” of the rectangle). This proved to be useful for detecting discontinuous lane markers as well as for further forward looking situations. Another benefit is that no intrinsic or extrinsic parameters of camera(s) are needed.

2.2 Inverse Perspective Mapping

Inverse perspective mapping (IPM) as introduced in [6] is another way to generate a bird’s-eye view from the input image [12]. The knowledge of the camera parameters (extrinsic and intrinsic) is required for the application of the IPM transform:

Viewpoint: camera position in world coordinate system $C = (l, h, d)$.

Viewing direction: optical axis is defined by two angles:

γ_0 : the angle formed by the projection of the optical axis on the xz plane [as shown in Fig. 1(c)],

θ_0 : the angle formed by the optical axis and axis η [as shown in Fig. 1(d)].

Aperture: camera angular aperture is $2\alpha_u$ in row direction and $2\alpha_v$ in column direction. While [note that (u_0, v_0) is the camera’s focal center, and F is focal length]: $\alpha_u = \arctan\{u_0/F\}$, $\alpha_v = \arctan\{v_0/F\}$.

Resolution: camera resolution is $n \times m$.

Mathematically, IPM can be modeled as a projection from a 3D Euclidean space \mathbb{W} , containing elements $(x, y, z) \in \mathbb{R}^3$, onto a planar (2D) subspace of \mathbb{R}^3 , denoted by \mathbb{I} , with elements $(u, v) \in \mathbb{R}^2$. The mapping from \mathbb{I} to \mathbb{W} is as follows,

$$\begin{aligned} x(u, v) &= h \cdot \cot\left\{(\theta_0 - \alpha_u) + u \frac{2\alpha_u}{m-1}\right\} \cdot \cos\left\{(\gamma_0 - \alpha_v) + v \frac{2\alpha_v}{n-1} + l\right\} \\ y(u, v) &= 0 \\ z(u, v) &= h \cdot \cot\left\{(\theta_0 - \alpha_u) + u \frac{2\alpha_u}{m-1}\right\} \cdot \sin\left\{(\gamma_0 - \alpha_v) + v \frac{2\alpha_v}{n-1} + d\right\} \end{aligned}$$

while we have the following mapping from \mathbb{W} to \mathbb{I} :

$$\begin{aligned} u(x, 0, z) &= \frac{\left[\arctan\left\{\frac{h \sin \gamma(x, 0, z)}{z-d}\right\} - (\theta_0 - \alpha_u)\right] \cdot (m-1)}{2\alpha_u} \\ v(x, 0, z) &= \frac{\left[\arctan\left\{\frac{z-d}{x-l}\right\} - (\gamma_0 - \alpha_v)\right] \cdot (n-1)}{2\alpha_v} \end{aligned}$$

An example of bird's-eye view using the above two methods is shown in Fig. 3. Both of them are using the planar ground plane assumption, and parameters from calibration. However, there are some difference between these two methods. First, WPM will not directly use camera's intrinsic or extrinsic parameters as IPM does. That means the calibration will be much easier for WPM. Second, changing a bird's-eye view generated from WPM means selecting another set of four points, while for IPM, the parameters of the camera need to be changed. Also, the quality of the generated bird's-eye views is different, which can be seen in the small (also enlarged) window in Fig. 3.

3 Edge Detection and Denoising

We recall an edge detection method as introduced in [11]. Black-white-black edges in vertical direction are detected in the bird's-eye image by a specially designed simple algorithm. Every pixel in the bird's-eye image, with value $b(x, y)$, is compared to values $b(x-m, y)$ and $b(x+m, y)$ of its horizontal left and right neighbors at a distance $m \geq 1$ as follows:

$$\begin{aligned} B_{+m}(x, y) &= b(x, y) - b(x+m, y) \\ B_{-m}(x, y) &= b(x, y) - b(x-m, y) \end{aligned}$$

Finally, using a threshold T , the edge map value will be

$$r(x, y) = \begin{cases} 1, & \text{if } B_{+m} \geq 0, B_{-m} \geq 0, \text{ and } B_{+m} + B_{-m} \geq T \\ 0, & \text{otherwise} \end{cases}$$

This edge detection method has the following properties. First, m can be adjusted to fit various widths of lane marks. Second, pixels within a lane mark are all labeled as being edge pixels, which is different from gradient-based edge operators (e.g., Sobel, Canny). This greatly improves the robustness in detecting points at lane marks. Third, shadows on the road surface do not influence edge detection at lane marks. Thus, the edge detection method can be used under various illumination conditions. Finally, horizontal edges are not detected. For an example of edge detection, see Figure 4.

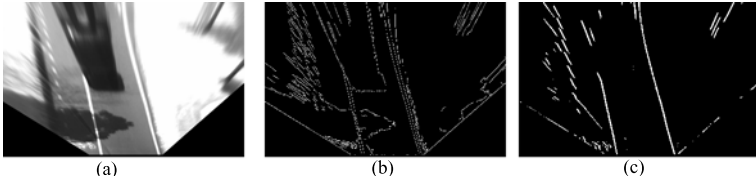


Fig. 4. Edge detection. (a) Bird’s-eye image. (b) Edge detection using the Canny operator. (c) Edge detection following [1].

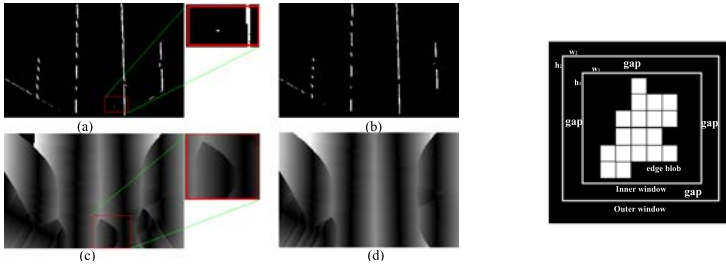


Fig. 5. Effect of an isolated noisy edge pixel for the distance transform. (a) Noisy edge map with an isolated edge point in the middle of the lane. (b) Denoised edge map. (c) RODT based on (a). (d) RODT based on (b). The operator to remove blobs is shown on the right.

The edge detection scheme as discussed above may generate some isolated small blobs (including single pixels) beside edges of real lane marks. These noisy blobs will greatly affect the result of the subsequent distance transform (see Fig. 5; the distance transform is discussed in Section 4). In order to remove such noise, a specified operation is applied. The general idea is first to find such isolated blobs, and then set them to zero (i.e., to the non-edge value). Two small windows (inner and outer) are used (see Fig. 5) at the same reference pixel, but the outer window is slightly larger than the inner one in width and height. The isolated blobs can be detected by moving at these two windows through the whole edge map, and comparing the sums of edge values within them. If two sums are equal, that means that the gap between two windows contains no edge points, then the edge blobs in the inner window are detected as being isolated, and set to zero. For computation efficiency, an integral image of the edge map is used for calculating the sum in the windows.

4 Distance Transform

The distance transform applied to the binary edge map labels each pixel with the distance to the nearest edge pixel (see [8] for details). Edge pixels are obviously labeled by 0, and this is shown as black in the generated distance map. Pixels “in the middle of a lane” are supposed to receive large labels, shown as bright pixels in the distance map (see Figure 6).

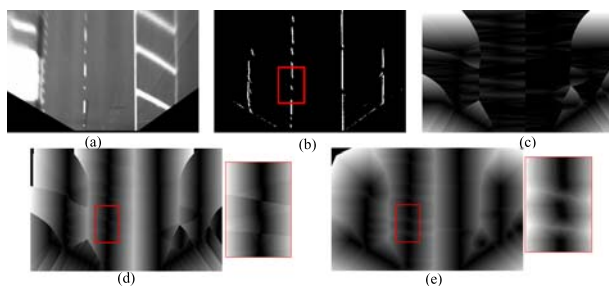


Fig. 6. EDT and ODT on a bird's-eye road image. (a) Bird's-eye image. (b) Edge map (the area in the rectangle is a discontinuous lane mark). (c) EDT. (d) Real part of ODT (absolute value). (e) Imaginary part of ODT. (c)(d)(e) have been contrast adjusted for better visibility.

The Euclidean distance transform (EDT) is, in general, the preferred option; using the Euclidean metric for measuring the distance between pixels. [3] proved that a 2D EDT can be calculated by two 1D EDTs, and this greatly improves the computational efficiency. A modified EDT was proposed in [9], called the *orientation distance transform* (ODT). This divides the Euclidean distance into a contributing component in the row and column direction. (Note that the use of a 4- or 8-distance transform would *not* lead to the same row and column components; however, practically there should not be a big difference with respect to the given context.) A complex number is assigned to each pixel by the ODT, with the distance component in the row direction as the real part, and the distance component in the column direction as the imaginary part. Then the magnitude and the phase angle of such a complex number, at a non-edge pixel, represent the Euclidean distance and the orientation to the nearest edge pixel, respectively. Note that the distance component in row direction is signed, with a positive value indicating that the nearest edge point lies to the right, and a negative value if it is to the left. See Figure 6 for an example. The imaginary part is mostly dark because the nearest edge pixels are, in general, in the same row; due to the applied edge detection method. Thus, we decided to ignore the imaginary part.

This paper uses only the Euclidean distance in row direction, and we call this the *real orientation distance transform* (RODT). The RODT of our edge map offers various benefits. First, every pixel indicates the nearest edge pixel, which will greatly facilitate the edge-finding procedure. Secondly, discontinuous lane marks will make almost no difference with continuous ones in the RODT (and this is different to the EDT of the edge map), as illustrated in Figure 6. Third, more information about the lane is provided by the distance transform compared to the edge map. For example, there's nothing provided by the edge map for the (virtual) centerline of a lane. While in distance map, pixels on the centerline will be assigned a large distance value.

The distance transform is sensitive to some isolated points or blobs in lane detection situation. As indicated in Fig. 5, an edge point on the middle of lane will greatly change the distance value for the surrounding pixels. So a denoising method on the edge map as introduced in Section 3 is necessary, and proves to be useful.

5 Experiments

Experimental results (using the test vehicle HAKA1) for lane detection (see [10]) are shown in Fig. 7 and for lane tracking in Fig. 8. The size of images used is 752×480 , recorded at 25 Hz.

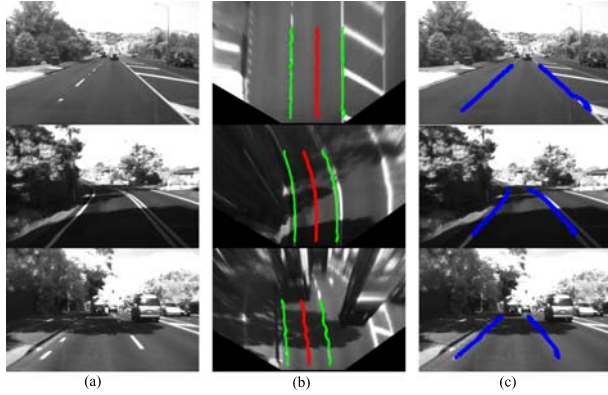


Fig. 7. Experimental results for lane detection. (a) Input images. (b) Lanes detected in the bird's-eye image. (c) Lanes detected.

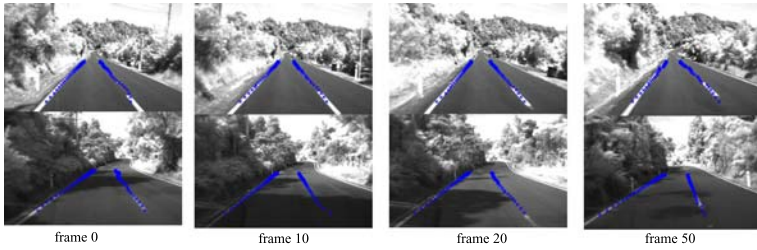


Fig. 8. Experimental results for using the robust lane tracking technique

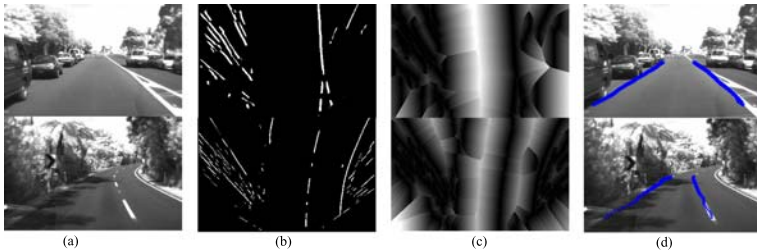


Fig. 9. Two examples of usefulness of RODT in detection of a lane boundary. (a) The input image. (b) The edge map. (c) RODT of (b). (d) Lane detection result.

The usefulness of distance transform can be seen from examples in Fig. 9. Lane marks are not apparent (or occluded) on the left hand side in these two situations, but the distance map provides sufficient information for identifying the lane boundaries.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grant 50875169.

References

1. Bertozzi, M., Broggi, A.: GOLD - A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Trans. Image Processing*, 7, 62–81 (1998)
2. Bertozzi, M., Broggi, A., Fascioli, A.: Stereo inverse perspective mapping: theory and applications. *Image and Vision Computing* 16, 585–590 (1998)
3. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transform of sampled functions. Technical report, Cornell Computing and Information Science (2004)
4. Gaspar, J., Winters, N., Santos-Victor, J.: Vision-based navigation and environmental representation with an omnidirectional camera. *IEEE Trans. Robotics and Automation* 16, 890–898 (2000)
5. Kim, Z.: Robust lane detection and tracking in challenging scenarios. *IEEE Trans. Int. Trans. Sys.* 9, 16–26 (2008)
6. Mallot, M.A., Hulthoff, H.H.B., Little, J.J., Bohrer, S.: Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics* 64, 177–185 (1991)
7. McCall, J.C., Trivedi, M.M.: Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *IEEE Trans. Int. Trans. Sys.* 7, 20–37 (2006)
8. Klette, R., Rosenfeld, A.: *Digital Geometry*. Morgan Kaufmann, San Francisco (2004)
9. Wu, T., Ding, X.Q., Wang, S.J., Wang, K.Q.: Video object tracking using improved chamfer matching and condensation particle filter. In: *SPIE-IS & T Electronic Imaging*, vol. 6813, pp. 04.1–04.10 (2008)
10. Jiang, R., Klette, R., Wang, S., Vaudrey, T.: Lane detection and tracking using a new lane model and distance transform. Technical Report, The University of Auckland (2009)

Lane Detection on the iPhone

Feixiang Ren¹, Jinsheng Huang¹,
Mutsuhiro Terauchi², Ruyi Jiang³, and Reinhard Klette¹

¹ The University of Auckland, Auckland, New Zealand

² Hiroshima International University, Hiroshima, Japan

³ Shanghai Jiao Tong University, Shanghai, China

Abstract. A robust and efficient lane detection system is an essential component of Lane Departure Warning Systems, which are commonly used in many vision-based Driver Assistance Systems (DAS) in intelligent transportation. Various computation platforms have been proposed in the past few years for the implementation of driver assistance systems (e.g., PC, laptop, integrated chips, PlayStation, and so on). In this paper, we propose a new platform for the implementation of lane detection, which is based on a mobile phone (the iPhone). Due to physical limitations of the iPhone w.r.t. memory and computing power, a simple and efficient lane detection algorithm using a Hough transform is developed and implemented on the iPhone, as existing algorithms developed based on the PC platform are not suitable for mobile phone devices (currently). Experiments of the lane detection algorithm are made both on PC and on iPhone.

Keywords: Intelligent transportation system, driver assistance, lane detection, iPhone, Hough transform.

1 Introduction

Intelligent Transportation Systems (ITS) are developed to improve road safety, reduce transportation times, or save fuel. Basic ITS technologies include traffic signal control systems, security CCTV systems, speed cameras, or automatic number plate recognition. Driver assistance systems (DAS) are advanced technologies, mainly designed for improving traffic safety. Various computing platforms are used for the application of DAS, such as PC, laptop, play station, integrated chips and so on. With the wide spread of mobile phones, many functionalities of DAS are also build into those phones, for example, GPS positioning or a video camera.

Lane detection is a core technology in DAS, as it can help to estimate the geometry of the road ahead, as well as the lateral position of the ego-vehicle on the road [6,8,10]. Lane detection is used in intelligent cruise control systems, for lane departure warning, road modeling, and so on. Typically, lane detection is used for localizing lane boundaries in given road images, and has been widely studied on various scenarios [9]. Recently, some commercial lane detection systems appeared, for example, based on embedded chips integrated into the car (e.g. mobileye [1]).

We demonstrate in this paper how to implement computer vision-based lane detection on a mobile phone, making good use of those mobile platforms with cameras. An iPhone [4], a product of Apple Inc, is used in this paper due to its integrated system components. Also, the lane detection method used in this paper covers various lane marks such as solid lines, dashed lines, road shoulders, or dotted lines. Due to the restriction of memory and computing power of an iPhone, we focus on a straight and flat lane model, and use the original Hough transform [2] (for straight lines). The aim of this paper is not to provide a top performance lane detection method, but a promising application situation on a mobile phone with restricted resources while a fairly robust detection result. For a brief review about algorithms for detecting and tracking lanes in general, see [5].

The paper is structured as follows. In Section 2, we give a general introduction to the iPhone, and the difficulties caused by iPhone's restricted resources. In Section 3, we describe the lane detection algorithm as implemented on the iPhone. In Section 4, experimental results from iPhone, and some comparison with PC-based lane detection are provided. Some conclusions are presented in Section 5.

2 iPhone

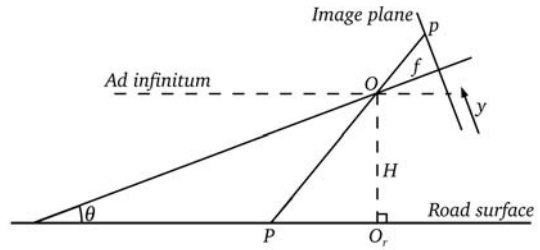
The iPhone is an evolutionary phone, it features a multi-touch screen, GPS, Wi-Fi access, an internal motion sensor, a build-in camera, a UNIX-based operating system that operates inside of the phone (but is normally not accessible for a programmer), and more, but what is not needed for our purposes (e.g., a web browser).

Generally, the implementation of computer vision-based lane detection on an iPhone has its difficulties in two aspects: hardware and software. As a mobile phone, iPhone has its limitation on computing power and memory, as well as other peripheral components. The comparison between iPhone and iMac (or, say, a typical PC) in hardware is conducted in Table 1. As indicated in this table, the shortage of physical memory and the slowness of CPU define the two main challenges. Typically, the same application running on an iPhone will be over 100 times slower than on an iMac. Furthermore, programming on an iPhone is another great challenge, as the iPhone is a closed system with limited reference documentation and functions, having no multi-tasking, no virtual memory, and less support in programming library.

For the installation of an iPhone in a car, see the left of Fig. 1. We mount the iPhone on the windscreen by an iPhone-specific car mount holder. This allows that the iPhone's camera is not blocked by the holder and also provides a stable mounting. (Of course, there is still the egomotion of the car, such as a changes in tilt and roll angles with respect to the road surface.) The user may place the phone either horizontally or vertically, and may freely move the holder to a different position. The system will automatically detect the placement of the phone using its internal sensor. We also use an USB car charger (to avoid any battery issue).

Table 1. A comparison between iPhone and iMac

	iPhone 3G	iMac
CPU	412 MHz	Intel Core 2 Duo 2.93 GHz
Memory	128 MB	4 GB 1066 MHz SDRAM
Screen Size	3.5-inch	24-inch
Screen Resolution	320 × 480	1920 × 1200
Power	3.7 V battery	200 W power

**Fig. 1.** Left: installation of iPhone on an ordinary car. Right: the geometry of road and camera.

The geometric model of the camera and lane using the above installation is shown on the right of Fig. 1. Below are the parameters:

- f : focal length
- O : focal center
- H : height of the camera above ground manifold
- θ : tilt angle of the camera

3 Lane Detection

With the restriction of the application platform, the developed lane detection algorithm must be simple and efficient. Our lane detection algorithm is based on the assumptions of planar ground manifold and locally straight lane boundaries. For lane detection in general situations, see [5]. The main flow chart of our algorithm is as shown in Fig. 2. After acquiring an image from the iPhone's camera, smoothing is applied in order to remove noise. Then a Canny edge detector is used to obtain a binarized edge map. Furthermore, in order to detect dotted lane marks, some morphologic operations are also adopted. The Hough

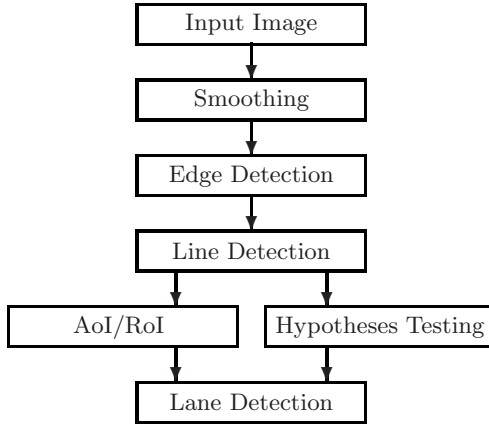


Fig. 2. Flowchart of lane detection

transform, applied to the edge map, detects lines. As there are many lines besides lane boundaries, a hypotheses testing module has been implemented for these candidate lines using multiple cues. An AoI/RoI will be defined in order to improve the computation efficiency, as the road will only appear in the bottom part of an input image. A real example of lane detection, using this algorithm, is shown in Fig. 3.

3.1 Preprocessing

The preprocessing of the input image includes three steps: smoothing, edge detection and morphologic operations. For the smoothing operation we may choose one of various filters, such as mean, median, or Gaussian. We decided for a 3×3 Gaussian filter. Similarly, the Canny edge detector is selected among other edge operators to generate a binarized edge map. For some kind of lane marks (e.g., dotted lane marks, which are common in New Zealand but not possible in areas having snow in winter), combinations of erosion and dilation (the two basic morphologic operations) are applied on the edge map in order to improve line detection results in the subsequent Hough transform. The effect of the applied morphologic operator is illustrated in Fig. 4.

3.2 Lane Detection Using a Hough Transform

Hough transforms are widely used for detecting lines and circles, and have also been applied in lane detection situation [7]. The selection of a Hough transform as the main processing step for lane detection is due to its robustness to noise and occlusion. This is because each edge point is considered independently of the others in a Hough transform. That means, the removal of some edge points, or the introduction of some more noise will not affect the result of Hough transform

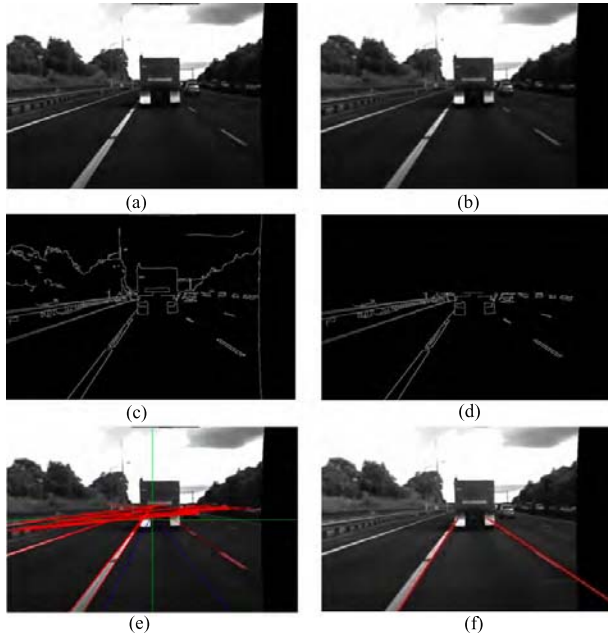


Fig. 3. An example of lane detection with intermediate results. (a) An input image. (b) Smoothing. (c) Edge detection. (d) AoI/RoI. (e) Hough transform. (f) Lane detection result.

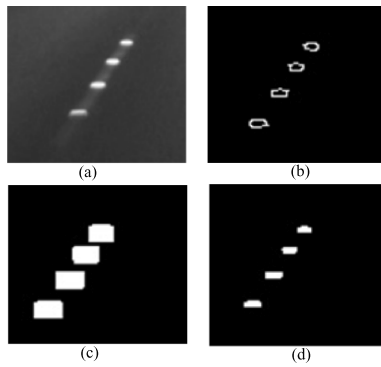


Fig. 4. Morphologic operator applied to the edge map of dotted lane marks. (a) An input image of dotted lane marks. (b) Edge detection. (c) Lane marks after repeated dilation. (d) Lane marks after repeated erosion.

very much. With this advantage, broken line patterns used in road marks, or lane marks partially occluded by obstacles (e.g., vehicles) can be easily detected using a Hough transform [7]. Another important reason for selecting a Hough transform compared to other lane detection methods is in its simplicity and

computation efficiency, which can be seen from comparisons of time needed for lane detection using a Hough transform or other algorithms, see [5] (discussion in Section 4).

3.3 Hypotheses Testing Using Multiple Cues

We assume that the lane width is between 3 to 5 meters and that an ego-vehicle is typically in the middle of the lane. We calculate prioritized lane border lines in the image, then we define for each line an area of interest (AoI) by a width parameter, for example, selected to be one meter.

By allowing different locations of the camera in the car, we define adaptively different areas of interest in images. Because the camera's optic axis is basically parallel to the forward direction, and the road surface is assumed to be planar (i.e., close to the car), the top part of the image is not used in the Hough transform or lane detection.

We apply the following multiple cues to decide whether lines are lane borders or not; candidate lines are found as a result of the Hough transform:

- AoI: A lane border must proceed in the AoI from the bottom towards the middle of the image.
- Lane width: The difference of a lane's width must be less than a constant (for example, one meter) when measured either at the bottom or the middle of the image.
- Parallel lines: Lane borders should be approximate parallel (taking perspective projection into account).
- Vanishing point: All parallel lines meet at the same point (i.e., the vanishing point). All lane borders should 'nearly intersect' the vanishing point.
- Enlargement of the AoI: If no lane border is found in an AoI, then we enlarge it (to the borders of the image; for example, if there is only a middle line marking in a local road, then the enlarged AoI should allow us to find the road shoulders).

Following the planar-ground and straight-lane assumptions used in this paper, experiments proved that, if there are lane borders detected by the Hough transform, then they will be robustly selected within the set of all detected lines according to the above cues, even in some complicated environments (see Fig. 5).

4 Experiment

The Environment Perception and Driver Assistance project (*.enpeda.*), see [3], at The University of Auckland provides calibrated multi-view video cameras in the test vehicle HAKA1 (High Awareness Kinematic Automobile 1). These data have been used for evaluating the lane detection algorithm in this paper, with results as illustrated in Fig. 5. It can be concluded from these experiments that the lane detection algorithm developed in this paper works robust for various lane types if local road planarity and piecewise straightness of lane borders is

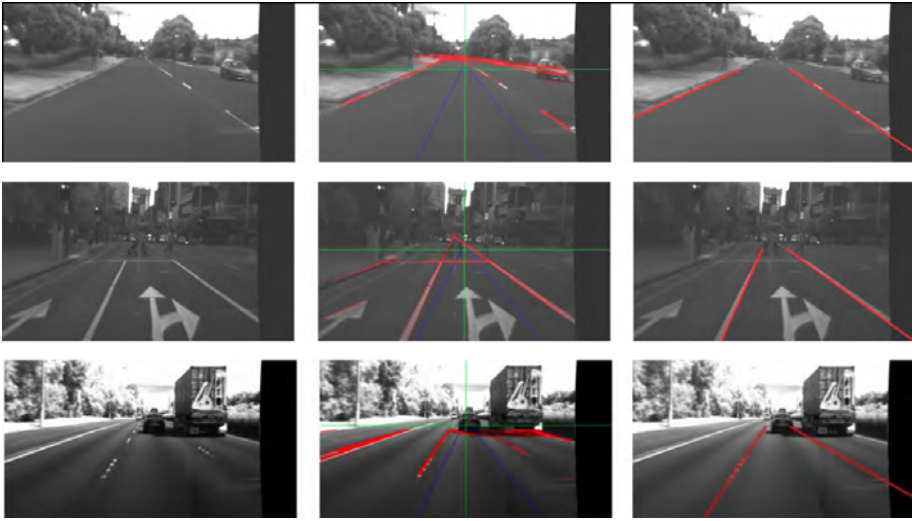


Fig. 5. Experiments of lane detection using recorded sequences. Left: input image. Middle: detected lines using the Hough transform. Right: Final lane detection result.



Fig. 6. Experimental results of lane detection on an iPhone at night

satisfied (as on highways, or on roads in cities which do not have many hilly areas).

Compared with another lane detection method [5] developed in the same research group, this lane detection algorithm computes much faster and consumes less memory, and fits well to the mobile platform.

The performance of the lane detection algorithm in the iPhone can reach one frame per second with optimization only in an experimental implementation. Figure 6 shows a lane detection result on the iPhone. Experiments in Auckland showed that the rate of successful lane detection is above 90 percent with clearly detected straight lane borders.

5 Conclusions

In this work, we reported about an implementation of lane detection on the iPhone, which is a promising mobile solution for driver assistance systems. A

simple and efficient lane detection method using a Hough transform was developed. We discussed the performance of this lane detection algorithm on a PC and on an iPhone. Experiments show that our scheme is practical, and the next generation of iPhones will allow to implement even more advanced techniques.

References

1. <http://www.mobileye.com/>
2. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. *Comm. ACM* 15, 11–15 (1972)
3. The .enpeda. project, <http://www.mi.auckland.ac.nz/enpeda>
4. iPhone: official website by Apple, <http://www.apple.com/iphone/>
5. Jiang, R., Klette, R., Wang, S., Vaudrey, T.: Lane detection and tracking using a new lane model and distance transform. Technical report, MITech-TR-39, The University of Auckland (2009)
6. Kim, Z.: Realtime lane tracking of curved local roads. In: *Proc. IEEE Conf. Intelligent Transportation Systems*, pp. 1149–1155 (2006)
7. John, B.M.: Application of the Hough transform to lane detection and following on high speed roads. In: *Proc. Irish Signals and Systems Conference* (2001)
8. Kim, Z.: Robust lane detection and tracking in challenging scenarios. *IEEE Trans. Intelligent Transportation Systems* 9, 16–26 (2008)
9. McCall, J., Trivedi, M.M.: Video based lane estimation and tracking for driver assistance: Survey, algorithms, and evaluation. *IEEE Trans. Intelligent Transportation Systems* 7, 20–37 (2006)
10. Wang, Y., Teoh, E., Shen, D.: Lane detection and tracking using B-snake. *Image Vision Computing* 22, 269–280 (2004)

Traditional Culture into Interactive Arts: The Cases of Lion Dance in Temple Lecture

Wen-Hui Lee, Chih-Tung Chen, Ming-Yu He, and Tao-i Hsu

Shih Hsin University, Dept. of Information Management,
Shih Hsin University, Dept. of Digital Multimedia Arts,
#1 Lane17 Sec.1, Mu-Cha Rd. Taipei, Taiwan, R.O.C
{s97660011, s97660010, s97660025, taoi}@cc.shu.edu.tw

Abstract. The lion dance in Chinese culture is one of profound arts. This work aims to bridge traditional culture and modern multimedia technology and application of network cameras for the interactive tool to design a set of activities to promote the lion as the main body. There consists of the imaging systems and interactive multimedia applications.

Keywords: Interactive Art, Lion Dance, Traditional Culture.

1 Introduction

1.1 Motivation for the Research

Lion dance activities in the Chinese community has been thousands of years. Lion dance from the Han Dynasty, peaked in the Southern and Northern Dynasty. It is a cultural exchange between China and the Western Regions. However, lion dance activities are not only unique to the Chinese people, but also Japan, Korea, Southeast Asian countries.

Traditional culture in a modern society faces a bottleneck for keeping vivid to people. The study will explore about the creative interaction design, lion dance culture, and the relationship between interactive media and traditional culture. To deepen the experience for visitors and promote the traditional culture – lion dance effectively.

1.2 The Objectives of the Research

This work attempts to promote traditional lion dance via multimedia interface design. It consists of through multimedia interface design interact promotion of the traditional lion dance:

- (1) Enhance people's awareness of traditional culture.
- (2) Use of the richness of the interaction interface to entertains in lion dance.
- (3) Retaining long-standing cultural tradition prevent it from disappearing.

2 Literature Review

Literature review in this chapter will be divided into three parts. Part I: Interactive design and interactive interface. With the basic concept of interaction to develop interactive works. Part II: To understand lion dance culture and history. Part III: To investigate the interaction between the media and traditional culture.

2.1 The Interaction Design and Interface

"Interaction design" combines the aesthetics, the humanities and technology. The use of "consciousness" to the definition of conduct between the object to the application of sensory perception of the interaction with the environment [4]. By means of "interaction", the user has a better grasp of absorption of ownership study to demonstrate the finished product is also easier to follow the idea of each person to adjust. And finished products through the manipulation of different people, can allow users to make easily and produce different results. Then change yours mind. That is a new visual experience [8]. Through the body and limbs of human action with computer "interactive" and trigger the body participate in fun activities. Through effective human-computer interface design pattern reach motivation. To make users enjoy the power of choice that is success in interaction [5].

Digital Art, communication and interactive has close relationship. Stressed person works through the body and the design of the conduct in interactive. Through human interaction and user interface, play an "interactive" and focus on conduct of participants to lead the most direct participants of the visual experience. The use of "multi-mode" environment interaction and the "perception" in management capability to understand the user's action (such as touch, pick up, move goods). Then create body immersion into the performance of the actual situation and enhance the interaction [1].

Dinkla have distinguished six important implications of interactivity:

(1) Power and Play; (2) Participation versus Interaction; (3) Proximity and Manipulation; (4) Strategies of Seduction; (5) Nonlinear Narration; (6) Remembering, Forgetting, and Reconstructing [10]. Through these qualities can view between interactive and user and develop more comfortable application [1].

Base on the user's interactive and display method; the interaction design can be roughly classified three main forms: (1) Interactive Wall (2) Interactive Tabletop (3) Interactive Floor. Detail description of Table 1 [1]. Some forms of artistic expression can be designed with computer vision technology inter actively.

2.2 Lion Dance's History

In China, the lion culture combining traditional culture. It contains a wealth of civilian life, religious beliefs and showing multi-cultural patterns of life. It is a symbol of ethnic consciousness and the characterization of groups in the region [2]. Tang Dynasty, Xinjiang lion dance figurines (see Figure 1), make Human and lion's image vivacity. Southern Song, "Hundred of people play in spring" (see Figure 2), depicted the case of children playing lion so vividly. Lion dance in the New Year become to the culture and have more festive atmosphere [2].

Table 1. Computer Vision interactive technology [1]

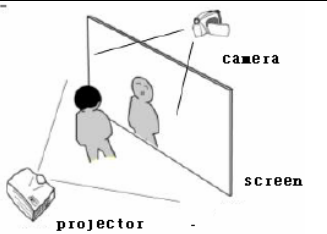
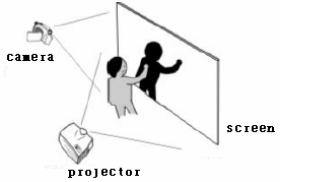
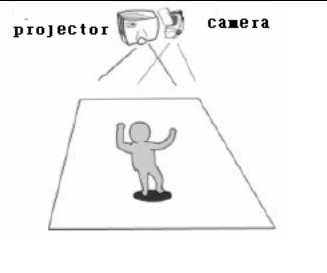
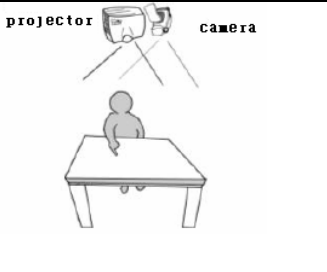
Interactive Wall		<p>Type 1: detect people's momentum Note: Use body gesture to interact</p> <p>In front of the camera body's motion interactive with projector.</p>
		<p>Type 2: detect shadow's motion Note: The outline of shadow to interact</p> <p>With backlight, make shadow in screen, camera set up in the back, analysis of the black shadow action.</p>
Interactive Floor		<p>Note: Focus on location move to make interactive</p> <p>Set up camera and video projector in the top of the user, analysis of the user standing location, and projection on bottom of location. Based on the location of objects arising interaction effect.</p>
Interactive Tabletop		<p>Note: Focus on hands to make interactive</p> <p>The camera is set on top of user to detect user's hand. And projector set up in top of user too, make images projector to desktop. Interaction can be accomplished by touch screen through hands.</p>



Fig. 1. Tang Dynasty, Xinjiang lion dance figurines



Fig. 2. Southern Song, Hundred of people play in Spring

2.3 Interactive Media and Traditional Culture

There are some issues in cultural industry. Typically, the enlisting and preserve the historical activities and performance is most important. With the help support of multimedia technology that allow people to interact with the culture. The results of digital media such as table 2 [6].

Table 2. Domestic digital media research result [6]

The field of digital media applications	Description
Humanities and Social	Combination of digital media technology, a game of chess to reproduce the Tang Dynasty Ladies, aristocrats playing polo, foreign countries into the gorgeous space-time Chang-an Example: Virtual Chang-an - Heritage Collection reproduce
Life and Entertainment	The use of 3D computer graphics, animation and other visual effects, and through the "3D wireless interactive technology" and the tip of the peripheral devices so that users can have an immediate and interactive media itself Example: 3D stereo photography services
Digital learning	Combination of broadband Internet and interactive new media technologies to address the needs of time and space, making scattered in remote teachers and students, and then a shared interactive virtual environment for interactive teaching and learning Example: Virtual classroom e-learning courses

3 Interactive Interface Design

3.1 Principles of Design

Visual information is captured by webcam. The data sets are manipulated with image processing techniques. Features are used to accomplish interactively with users.

3.2 Image Acquisition

“Processing” developed by MIT media lab is employed as an image processing tool. The image features consists of color information and shapes by means of processing.

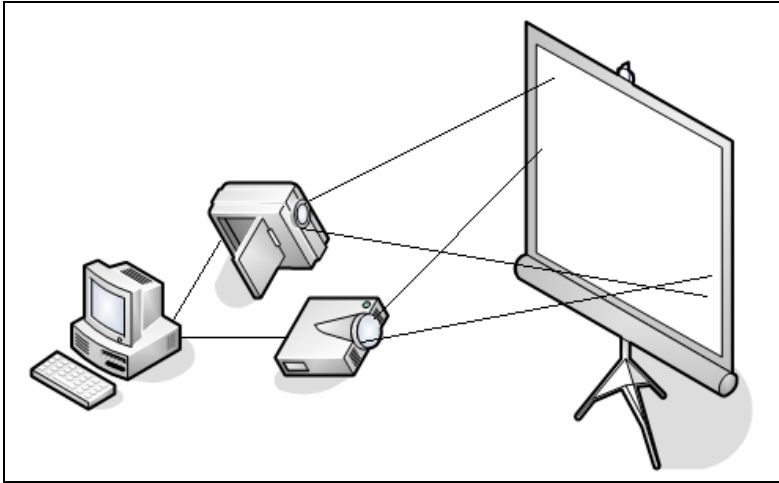


Fig. 3. Interact with diagram

3.3 Reference Creative Technology

The computer vision technology to the development of three forms: (1) Interactive Wall, (2) Interactive Tabletop, (3) Interactive Floor. Divided by the number of computer vision over the design of the computer as a Webcam capture color images to detect specific target object to move, So that objects can also move the location of the changes made compared to the data of the interaction as a reference input value, and the use of infrared cameras is to detect the infrared light, etc.

3.4 Work Design Characteristics

This study is designed to work real-time images of the interactive installation by the user to spherical objects moving in front of webcam, as the interactive mode.

Computer vision technology through the importation of video cameras into digital images, the number of digital images can be viewed as a two-dimensional matrix, in the matrix of each implementation with the following values as the color pixels.

By the algorithm can be related to processing analysis to guide the projection of the image generated simulation of the dynamic pursuit, that is, the integration of real and virtual contexts.

4 The Experimental Results

Lion dance is a traditional Chinese art performance. Chinese traditional lion dance that can avoid the evil ghosts so every festival, Spring Festival Tour will be occasions, such as lion dance groups will have to kick off the festive celebrations. In this paper, lion and the interaction between the red ball and use the information in the digital media source graphics with real-time computing (closed regional sphere of color and the outline of information), to achieve interactive visual feedback.

4.1 Design Concepts

When the red ball moves, the lion dance will follow the red ball to move to achieve the purpose of interaction with the user (See Figure 4).

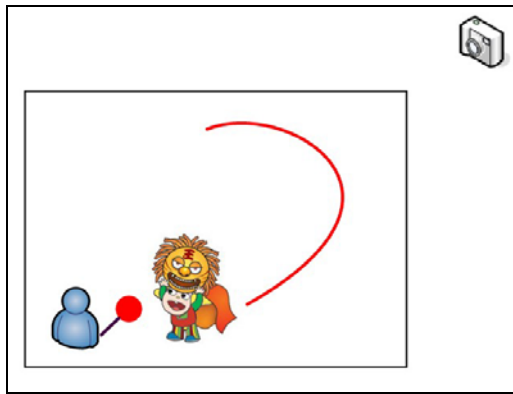


Fig. 4. The lion dance will follow the red ball.

4.2 Experiments

Judge the red ball to detect the use of the following basic steps to determine.

- (1) To catch up through the RGB color camera.

```
import processing.video.*
Capture video
R = red()
G = green()
B = blue()
```

- (2) Based on the prior definition of the scope of good red color, to obtain the location of the red color range (See Figure 5) .

```
video.loadPixels()
int index = 0;
for (int y = 0; y < video.height; y++) {
```

```

    for (int x = 0; x < video.width; x++) {
        // Get the color stored in the pixel
        pixelValue = video.pixels[index]
        get RGB (pixelValue) near the Red
        index++
    }
}

```

(3) Sphere obtained to use the above information, calculate the relative X, Y coordinates and combined with the main graphics.

```

image(lion.jpg, x, y)

```

(4) When the ball moves, so that graphics can have a direction of movement (See Figure 6).

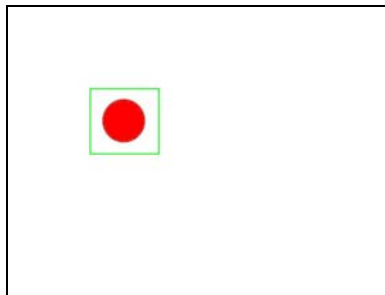


Fig. 5. Obtain the location of the red color range

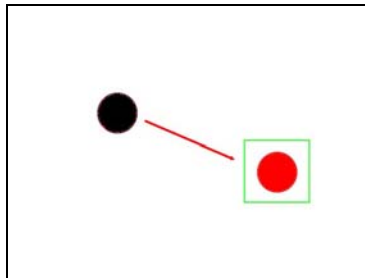


Fig. 6. The Ball movement traces

4.3 Preliminary Results of the Experiment

In this paper, the hand moves back and forth to enhance the impression that the lion dance. The red ball is tracked by the camera in terms of red color information. Experimental plan is shown as follows (See Figure 7 and Figure 4.3.2).

Obtain the location of red color range.

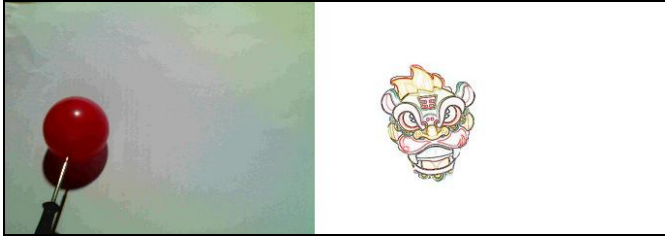


Fig. 7. Preliminary results of the experiment(original location)

4.4 Display of Results

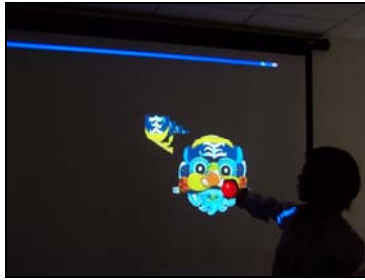


Fig. 8. Display the result (Original location)

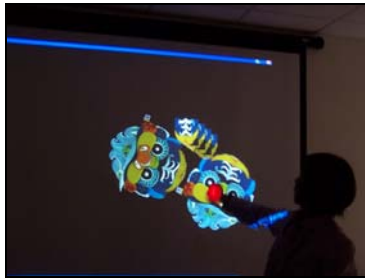


Fig. 9. Display the result (After moving the red ball)

5 Conclusion and Concluding Remarks

The study with a concept of modern multimedia technology, digital cameras and adding interactive elements to create a lion dance interactive multimedia platform. Concluding remarks:

(1) Visual Lively

With the festive lively combined with rhythm elements. Freshness of color to creates a fun platform to attract users. Players are able to experience to increase the user with the traditional cultural relations.

(2) Entertainment between platform and user

The study is Detected red-ball motion by the digital camera to make a interactive platform. User's action and reaction can be catches by digital camera. Generate what you see is what you get's effectively. Let tradition and modernity no distance and more closer.

Lion dance is a common activity. Make traditional culture be in people's hearts deeply. In order to promote the traditional activities, this work has demonstrated it's efficacy with respect to this aspects.

References

1. Wu, Z.-D., Wang, Z.-M.: Perceptual user interface to explore the use of digital interactive art research. *Journal of Informatics & Electronics* 2(1), 9–18 (2007)
2. Zhang, L.-G.: A study of Liu Dui Hakka lion dance; Tainan; National Tainan University Graduate Institute of Taiwan Culture; National Tainan University Graduate Institute of Taiwan Culture Master's thesis (2007)
3. Hua-ao Star; Lion Dance (2007),
<http://big5.sports.cn/b5/dragonlion.sport.org.cn/home/lsw/2007-06-16/123739.html>
4. Hsiao, Y.-Z., Ye, Z.-R., Zheng, Z.-M., Fan, B.-L.: Games Aesthetics in Art Education at the Digital Age (2005)
5. Zhang, T.-J.: Digital era aesthetics of educational games. *Digital Arts Education Network Journal* 5 (2004)
6. Xu, L.-Y., Lin, Q.-X.: Applying Aesthetics of Interaction to Interpersonal Communication Product Design. In: Cultural and creative industries Design Symposium (2003)
7. Ou, S.-H.: Applying Aesthetics of Interaction to Interpersonal Communication Product Design; Hsinchu; National Chiao Tung University Institute of Applied Arts; National Chiao Tung University Institute of Applied Arts Master Master's thesis (2003)
8. Zheng, M.-J.: Arts education in the new thinking. *Digital Arts Education Network Journal* 4 (2003)
9. Li, Y.-Y.: Research on ritual and performance of lion dance in Taiwan: The cases of Taiwanese Lion and Cantonese Lion; Taipei; National Taipei University Institute of Folk Art; National Taipei University Institute of Folk Arts Master Master's thesis (2003)
10. Dinkla, S.: The History of the Interface in Interactive Art (1994),
http://www.kenfeingold.com/dinkla_history.html

Improving Optical Flow Using Residual and Sobel Edge Images

Tobi Vaudrey¹, Andreas Wedel², Chia-Yen Chen³, and Reinhard Klette¹

¹ The *.enpeda.* Project, The University of Auckland, Auckland, New Zealand

² Daimler Research, Daimler AG, Stuttgart, Germany

³ National University of Kaohsiung, Taiwan

Abstract. Optical flow is a highly researched area in low-level computer vision. It is a complex problem which tries to solve a 2D search in continuous space, while the input data is 2D discrete data. Furthermore, the latest representations of optical flow use Hue-Saturation-Value (HSV) colour circles, to effectively convey direction and magnitude of vectors. The major assumption in most optical flow applications is the intensity consistency assumption, introduced by Horn and Schunck. This constraint is often violated in practice. This paper proposes and generalises one such approach; using residual images (high-frequencies) of images, to remove the illumination differences between corresponding images.

1 Introduction

Dense optical flow was first presented by Horn and Schunck [8]. Their approach exploited the intensity consistency assumption (ICA), coupled with a smoothness constraint. This was solved in a variational approach. Many more approaches have been proposed since this, most using this basic ICA and smoothness constraint. In recent years, the use of pyramids, warping and robust minimisation equations have improved results dramatically [3]. This has further been improved and computational enhancement in [21].

Previous studies have compared the results of optical flow algorithms against ground truth using various types of scenes [12,6,11]. The earlier works in [2,6,11] use synthetically rendered scenes, and calculate the ground truth via ray-tracing. The more recent work of [1] calculates ground truth using structured lighting for real scenes. All of the scenes in these papers have been made publicly available. They are of good quality, but have a very limited number of frames (under 20).

None of these scenes are very difficult for the latest optical flow algorithms. The *Yosemite* scene from [2] has varying illumination in the sky, therefore most people do not use the sky for their evaluations. This means that most approaches still rely heavily on the ICA, and if this is violated the results become much worse. This was formally highlighted in [18], and then experimentally in [14]. This violation of the ICA is a major issue in real-world scenarios, such as driver assistance and security video analysis. A sample of the scene used, with ground truth, is shown in Figure 1. This demonstrates how lighting can differ dramatically between two frames.

For dealing with illumination artifacts, there are three basic approaches: simultaneously estimate the optical flow matching and model brightness change within the optical



Fig. 1. Example frames from EISATS scene. Frame 1 (left) and 2 (middle) are shown with ground truth flow (right) also showing the color key (HSV circle for direction, saturation for vector length, max saturation at flow length 10).

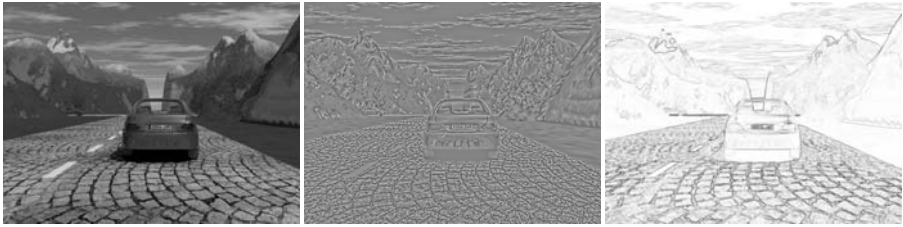


Fig. 2. Example for removing illumination artifacts due to different camera exposure in the frame 2 of EISATS set 2. Original (left) has its residual image (middle, computed using $TV-L^2$) and Sobel edge image (right) shown. Notice that the residual image retains more information than the Sobel image.

flow estimation [7], try to map both images into a uniform illumination model, or map the intensity images into images which carry the illumination-independent information (e.g., using colour images [12][20]).

Using the first option, only reflection artifacts can be modeled without major computational expense. From experiments with various unifying mappings, the second option is basically impossible (or, at least, a very big challenge). The third approach has more merit for research; we restrain our study to using the more common grey value images.

An example of mapping intensity images into illumination-independent images is the structure-texture image decomposition [15] (an example can be seen in Figure 2). More formally, this is the concept of *residuals* [9], which is the difference between an intensity image and a smoothed version of itself. One of the first approaches, that exploited the residual images of [15], is $TV-L^1$ improved optical flow [19], which is an improvement to the original $TV-L^1$ proposed in [21]. A residual is, in fact, an approximation of a high-pass filter, so only high frequencies remain present.

In this paper we generalise the residual operator by using *any* smoothing operator to calculate the low frequencies. Included in this study are three edge-preserving filters ($TV-L^2$ [15], median, bilateral [17]), two general filters (mean and Gaussian), and a gradient preserving filter (trilateral [4]). Furthermore, we use an edge detector as a reference (Sobel [16]). This paper shows experimentally that any residual image is better than the original image when illumination variance is causing issues.

2 Smoothing Operators and Residuals

Let f be any frame of a given image sequence, defined on a rectangular open set Ω and sampled at regular grid points within Ω .

f can be defined to have an additive decomposition $f(\mathbf{x}) = s(\mathbf{x}) + r(\mathbf{x})$, for all pixel positions $\mathbf{x} = (x, y)$, where $s = S(f)$ denotes the *smooth component* (of an image) and $r = R(f) = f - S(f)$ the *residual* (Figure 2 shows an example of the decomposition). We use the straightforward iteration scheme:

$$s^{(0)} = f, \quad s^{(n+1)} = S(s^{(n)}), \quad r^{(n+1)} = f - s^{(n+1)}, \quad \text{for } n \geq 0.$$

The concept of residual images was already introduced in [9] by using a 3×3 mean for implementing S . We apply the $m \times m$ mean operator and also an $m \times m$ median operator in this study. Furthermore, we use an $m \times m$ Gaussian filter, with σ for the normal approximation. The other operators for S are defined below.

TV- L^2 filter. [15] assumed an additive decomposition $f = s + r$ into a *smooth component* s and a *residual component* r , where s is assumed to be in $L^1(\Omega)$ with bounded TV (in brief: $s \in \text{BV}$), and r is in $L^2(\Omega)$. This allows one to consider the minimization of the following functional:

$$\inf_{(s,r) \in \text{BV} \times L^2 \wedge f=s+r} \left(\int_{\Omega} |\nabla s| + \lambda \|r\|_{L^2}^2 \right) \quad (1)$$

The TV- L^2 approach in [15] was approximating this minimum numerically for identifying the “desired clean image” s and “additive noise” r . See Figure 2. The concept may be generalized as follows: any *smoothing operator* S generates a *smoothed image* $s = S(f)$ and a *residuum* $r = f - S(f)$. For example, TV- L^2 generates the smoothed image $s = S_{TV}(f)$ by solving Equ. (1).

Sigma filter. This operator [10] is effectively a trimmed mean filter; it uses an $m \times m$ window, but only calculates the mean for all pixels with values in $[a - \sigma_f, a + \sigma_f]$, where a is the central pixel value and σ_f is a threshold. We chose σ_f to be the standard deviation of f (to reduce parameters for the filter).

Bilateral filter. This edge-preserving Gaussian filter [17] is used in the spatial domain (using σ_2 as spatial σ), also considering changes in the colour domain (e.g., at object boundaries). In this case, offset vectors \mathbf{a} and position-dependent real weights $d_1(\mathbf{a})$ define a local convolution, and the weights $d_1(\mathbf{a})$ are further scaled by a second weight function d_2 , defined on the differences $f(\mathbf{x} + \mathbf{a}) - f(\mathbf{x})$:

$$s(\mathbf{x}) = \frac{1}{k(\mathbf{x})} \int_{\Omega} f(\mathbf{x} + \mathbf{a}) \cdot d_1(\mathbf{a}) \cdot d_2[f(\mathbf{x} + \mathbf{a}) - f(\mathbf{x})] \, d\mathbf{a} \quad (2)$$

$$k(\mathbf{x}) = \int_{\Omega} d_1(\mathbf{a}) \cdot d_2[f(\mathbf{x} + \mathbf{a}) - f(\mathbf{x})] \, d\mathbf{a}$$

Function $k(\mathbf{x})$ is used for normalization. In this paper, weights d_1 and d_2 are defined by Gaussian functions with standard deviations σ_1 and σ_2 , respectively. The smoothed function s equals $S_{BL}(f)$. It therefore only takes into consideration values within a

Gaussian kernel (σ_2 for spatial domain, f for kernel size) within the colour domain (σ_1 as colour σ).

Trilateral filter. This gradient-preserving smoothing operator [4] (i.e., it uses the local gradient plane to smooth the image) only requires the specification of one parameter σ_1 , which is equivalent to the spatial kernel size. The rest of the parameters are self tuning.

It combines two bilateral filters to produce this effect. At first, a bilateral filter is applied on the derivatives of f (i.e., the gradients):

$$g_f(\mathbf{x}) = \frac{1}{k_{\nabla}(\mathbf{x})} \int_{\Omega} \nabla f(\mathbf{x} + \mathbf{a}) \cdot d_1(\mathbf{a}) \cdot d_2(\|\nabla f(\mathbf{x} + \mathbf{a}) - \nabla f(\mathbf{x})\|) \, d\mathbf{a} \quad (3)$$

$$k_{\nabla}(\mathbf{x}) = \int_{\Omega} d_1(\mathbf{a}) \cdot d_2(\|\nabla f(\mathbf{x} + \mathbf{a}) - \nabla f(\mathbf{x})\|) \, d\mathbf{a}$$

Simple forward differences $\nabla f(x, y) \approx (f(x+1, y) - f(x, y), f(x, y+1) - f(x, y))$ are used for the digital image. For the subsequent second bilateral filter, [4] suggested the use of the smoothed gradient $g_f(\mathbf{x})$ [instead of $\nabla f(\mathbf{x})$] for estimating an approximating plane $p_f(\mathbf{x}, \mathbf{a}) = f(\mathbf{x}) + g_f(\mathbf{x}) \cdot \mathbf{a}$. Let $f_{\Delta}(\mathbf{x}, \mathbf{a}) = f(\mathbf{x} + \mathbf{a}) - p_f(\mathbf{x}, \mathbf{a})$. Furthermore, a neighbourhood function

$$n(\mathbf{x}, \mathbf{a}) = \begin{cases} 1 & \text{if } \|g_f(\mathbf{x} + \mathbf{a}) - g_f(\mathbf{x})\| < A \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

is used for the second weighting. A specifies the adaptive region and is discussed further below. Finally,

$$s(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{k_{\Delta}(\mathbf{x})} \int_{\Omega} f_{\Delta}(\mathbf{x}, \mathbf{a}) \cdot d_1(\mathbf{a}) \cdot d_2(f_{\Delta}(\mathbf{x}, \mathbf{a})) \cdot n(\mathbf{x}, \mathbf{a}) \, d\mathbf{a} \quad (5)$$

$$k_{\Delta}(\mathbf{x}) = \int_{\Omega} d_1(\mathbf{a}) \cdot d_2(f_{\Delta}(\mathbf{x}, \mathbf{a})) \cdot n(\mathbf{x}, \mathbf{a}) \, d\mathbf{a}$$

The smoothed function s equals $S_{TL}(f)$. Again, d_1 and d_2 are assumed to be Gaussian functions, with standard deviations σ_1 and σ_2 , respectively. The method requires specification of parameter σ_1 only, which is at first used to be the radius of circular neighbourhoods at \mathbf{x} in f ; let $\bar{g}_f(\mathbf{x})$ be the mean gradient of f in such a neighbourhood. Let

$$\sigma_2 = 0.15 \cdot \left\| \max_{\mathbf{x} \in \Omega} \bar{g}_f(\mathbf{x}) - \min_{\mathbf{x} \in \Omega} \bar{g}_f(\mathbf{x}) \right\| \quad (6)$$

(Value 0.15 was recommended in [4]). Finally, also use $A = \sigma_2$.

Numerical Implementation. All filters have been implemented in OpenCV, where possible the native function was used. For the TV-L², we use an implementation (with identical parameters) as in [19]. All other filters used are virtually parameterless (except a window size) and we use a window size of $m = 3$ ($\sigma_1 = 3$ for trilateral filter [1]). For the bilateral filter, we use color standard deviation $\sigma_1 = I_r/10$, where I_r is the range of the intensity values (i.e., $\sigma_1 = 0.2$ for the scaled images). The default value of $\sigma = 0.95$ is

¹ The authors thank Prasun Choudhury (Adobe Systems, Inc.) and Jack Tumblin (EECS, Northwestern University), for their implementation of the trilateral filter.

used for the Gaussian filter. All images are scaled to the range $-1 < h(\mathbf{x}) < 1$ using normalisation.

In our analysis, we also use Sobel edge images [16]; this operator provides a normalised gradient function. This is another form of illumination invariant images.

3 Optical Flow on EISATS Dataset

One of the most influential evaluations of optical flow in recent years is from Middlebury Vision Group [1]. This dataset is used to evaluate optical flow in relatively simple situations. To highlight the effect of using residual images, we used a high ranking (see [13]) optical flow technique called TV- L^1 optical flow [21]. The results for optical flow were analysed on the EISATS dataset [5]; see [18] for Set 2 (see below for details). Numerical details of implementation are given in [19]. The specific parameters used were:

Smoothness:	35	Duality threshold θ :	0.2	TV step size:	0.25
# of pyramid levels:	10	# of iterations / level:	5	# of warps / iteration:	25

The flow field is computed using $U(h_1, h_2) = \mathbf{u}$. This is to show that a residual image r provides better data for matching than for the original image f . We computed the flow using $U(r_1^{(n)}, r_2^{(n)})$ with $n = 1, 10, 50,$ and 100 to show how each filter behaves. The results are compared to optical flow on the original images $U(f_1, f_2)$, and also for the Sobel-edge images. Figure 4 shows an example of this effect, obviously the residual image vastly improves optical flow results. In fact, the original image results are so noisy that they cannot be used.

EISATS Synthetic Dataset. This dataset was made public in [18] for Set 2 and is available from [5]. We are only interested in bad illumination conditions. We therefore use the altered data to resemble illumination differences in time, as performed in [14]; the differences start high between frames, then go to zero at frame 50, then increase again. For all t (frame number) we alter the original image f using a constant brightness, using $f(\mathbf{x}) = f(\mathbf{x}) + c$. The constant brightness change is defined by: even values of t , $c = t - 52$, and odd values of t , $c = 51 - t$. An example of the data used can be seen in Figure 1 and the brightness change over the sequence can be seen in Figure 3.

Results. To compare the results numerically, we calculated the end-point-error (EPE) as used in [1], which is basically a 2D-root mean squared error. The results can be seen in Figure 5. The zoomed out graph highlights that the results for the original image are unusable. The shape of the graph is appropriate as well, because the difference between intensities of the images gets closer together near the middle of the sequence,

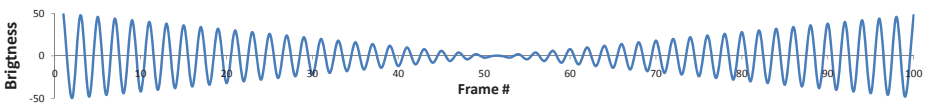


Fig. 3. Graph showing brightness change over sequence for EISATS dataset 2

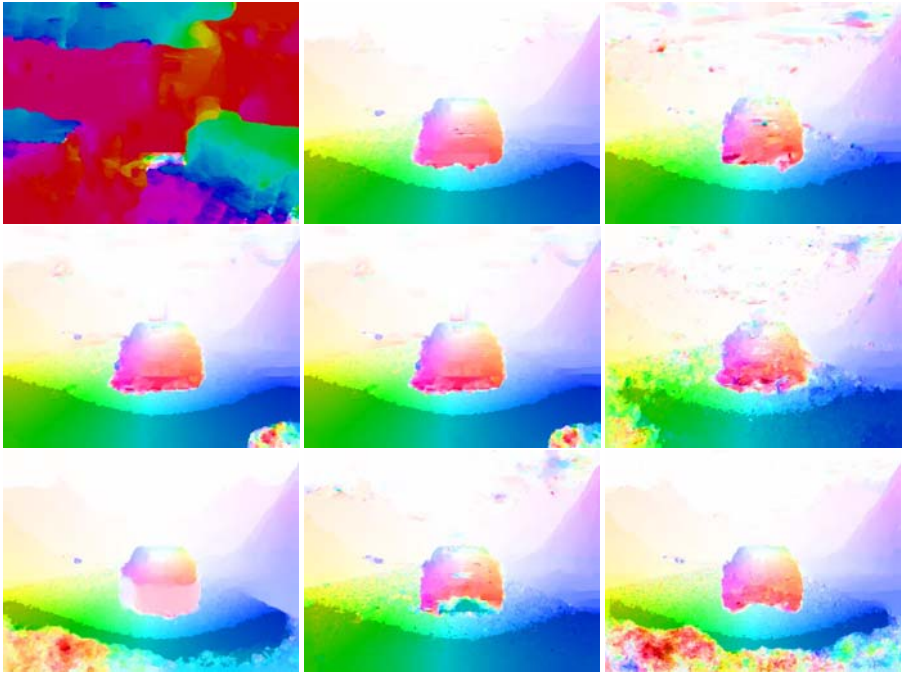


Fig. 4. Sample optical flow results on EISATS scene. Colour is encoded as in Figure 1. Top row (left to right): Using original images, Sobel edge images, and trilateral filter. Middle row (left to right): Gaussian, mean, and sigma filter. Bottom row (left to right): Median, bilateral, and TV-L² filter.

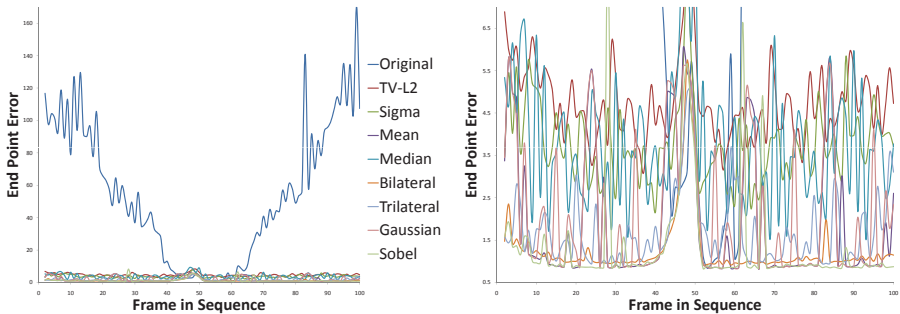


Fig. 5. End-Point-Error results over entire EISATS sequence. Filter iterations $r^{(r)}$ of $n = 100$ are shown. The left shows how different the magnitude is for the original sequence, and the right hand graph is zoomed in between 0.5 and 7.

Table 1. Results of TV-L¹ optical flow on EISATS sequence. Results are shown for different numbers n of iterations. Statistics are presented for the average (Ave.), zero-mean standard deviation (ZMSD), and the rank based on ZMSD.

n	TV-L ²	Sigma	Mean	Median	Bilateral	Trilateral	Gaussian	Original	Sobel	
1	Ave.	7.58	7.74	7.69	7.36	6.80	6.34	7.71	55.04	1.35
	ZMSD	7.59	7.76	7.71	7.38	6.83	6.36	7.72	69.41	1.84
	Rank	5	8	6	4	3	2	7	9	1
10	Ave.	6.88	7.45	5.63	4.73	3.30	1.72	6.66	-	-
	ZMSD	6.91	7.47	5.69	4.93	3.44	1.93	6.70	-	-
	Rank	7	8	5	4	3	2	6	9	1
50	Ave.	5.17	5.59	2.83	3.85	1.47	1.72	2.83	-	-
	ZMSD	5.24	5.67	3.27	4.16	1.75	1.93	3.23	-	-
	Rank	7	8	5	6	1	3	4	9	2
100	Ave.	4.76	3.78	1.95	3.84	1.26	1.72	2.19	-	-
	ZMSD	4.85	3.89	2.53	4.16	1.46	1.93	2.72	-	-
	Rank	8	6	4	7	1	3	5	9	2

and further away near the end. The zoomed graph shows the EPE values between 0.5 and 7.

A major point to highlight is that at different frames in the sequence, there are different rankings for the filters. If you look, for example, at the $n = 100$ graph at frame 25, the rank is (best to worst): Sobel, trilateral, bilateral, sigma, TV-L², median, Gaussian, then mean. But if you look at frame 75 (roughly the same difference in illumination) the rank is (best to worst): mean, Sobel, bilateral, trilateral, median, sigma, TV-L², with Gaussian coming last; a completely different order! From this it should be obvious that a smaller dataset will not pick up on these subtleties, so a large dataset (such as a long sequence) is a prerequisite for better understanding of the behaviour of an algorithm.

Since we have such a large dataset (99 results, 100 frames) we calculated the average and zero-mean standard deviation (ZMSD) for different iteration numbers $n = 1, 10, 50$, and 100. These results are shown in Table 1. Obviously, the original images are far worse than any residual image. From this table you can see that the order of the rankings shift around depending on the number of iterations for the residual image n . Another point to note is that the Sobel filter (which has only 1 iteration) is the best until 50 iterations; this is when bilateral filtering is the best. Simple mean filtering (which is much faster than any other filter) comes in at rank 5 after 50 iterations, and gets better around 100 iterations. It is notable that the difference between the average and ZMSD highlights how volatile the results are, the closer together the numbers, the more consistent the results.

4 Conclusions and Future Research

We have demonstrated how different residual images effect the results of optical flow. Any residual image is better than the original optical flow when using the illumination adjusted dataset. It turns out that bilateral filter may be the best, but a Sobel filter or mean residual image will work as well. So far, only simple residual images and Sobel images have been tested. Other smoothing algorithms and illumination invariant models

need to be tested, such as those exploiting phase information. Finally, a larger dataset can be used to further verify the illumination artifact reducing effects of residual images.

References

1. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M., Szeliski, R.: A database and evaluation methodology for optical flow. In: ICCV, pp. 1–8 (2007)
2. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. *Int. J. of Computer Vision* 12(1), 43–77 (1994)
3. Brox, T., Bruhn, A., Papenber, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J.G. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
4. Choudhury, P., Tumblin, J.: The trilateral filter for high contrast images and meshes. In: Proc. Eurographics Symp. Rendering, pp. 1–11 (2003)
5. .enpeda. dataset 2 (EISATS), <http://www.mi.auckland.ac.nz/EISATS>
6. Galvin, B., McCane, B., Novins, K., Mason, D., Mills, S.: Recovering motion fields: an evaluation of eight optical flow algorithms. In: Proc. 9th British Machine Vision Conf., pp. 195–204 (1998)
7. Haussecker, H., Fleet, D.J.: Estimating optical flow with physical models of brightness variation. *IEEE Trans. Pattern Analysis Machine Intelligence* 23, 661–673 (2001)
8. Horn, B.K.P., Schunck, B.G.: Determining optical flow. *Artificial Intelligence* 17, 185–203 (1981)
9. Kuan, D.T., Sawchuk, A.A., Strand, T.C., Chavel, P.: Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. Pattern Analysis Machine Intelligence* 7, 165–177 (1985)
10. Lee, J.-S.: Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing* 24, 255–269 (1983)
11. McCane, B., Novins, K., Crannitch, D., Galvin, B.: On benchmarking optical flow. *Computer Vision and Image Understanding* 84, 126–143 (2001)
12. Mileva, Y., Bruhn, A., Weickert, J.: Illumination-robust variational optical flow with photometric invariants. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 152–162. Springer, Heidelberg (2007)
13. Middlebury Optical Flow Evaluation, <http://vision.middlebury.edu/flow/>
14. Morales, S., Woo, Y.W., Klette, R., Vaudrey, T.: A study on stereo and motion data accuracy for a moving platform. In: Proc. Int. Conf. on Social Robotics, ICSR (to appear, 2009)
15. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
16. Sobel, I., Feldman, G.: A 3x3 isotropic gradient operator for image processing. *Pattern Classification and Scene Analysis*, 271–272 (1973)
17. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proc. IEEE Int. Conf. Computer Vision, pp. 839–846 (1998)
18. Vaudrey, T., Rabe, C., Klette, R., Milburn, J.: Differences between stereo and motion behaviour on synthetic and real-world stereo sequences. In: Proc. IEEE Image and Vision Conf., New Zealand (2008), doi:10.1109/IVCNZ.2008.4762133
19. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An improved algorithm for TV-L¹ optical flow. In: Post Proc. Dagstuhl Motion Workshop (to appear, 2009)
20. van de Weijer, J., Gevers, T.: Robust optical flow from photometric invariants. In: Proc. Int. Conf. on Image Processing, pp. 1835–1838 (2004)
21. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV-L¹ optical flow. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 214–223. Springer, Heidelberg (2007)

3 Case Studies: A Hybrid Educational Strategy for ART/SCI Collaborations

Elif Ayiter, Selim Balcisoy, and Murat Germen

Abstract. In this paper we report on a transdisciplinary university course designed to bring together fine art/visual communication design and computer science students for the creation and implementation of collaborative visual/audio projects that draw upon the specialized knowledge of both these disciplines. While an overview of the syllabus and the teaching methodologies is undertaken in the introduction, the focus of the paper concentrates upon an in-depth discussion and analysis of 3 specific projects that were developed by 3 distinct teams of students comprised of one artist/designer and one engineer each.

Keywords: Education, transdisciplinary, collaboration, design, artistic output, bio-data, virtual architecture, data visualization.

1 Introduction

The work discussed in this paper, is the student's output of a hybrid course positioned between visual arts/visual communication design and computer engineering, initially developed in the spring of 2003, by a computer scientist, specializing in computer graphics and a graphic designer, specializing in electronic applications. Both were practitioners in their fields as well as academicians and both had encountered collaborative projects in their separate practices, giving them an insider's knowledge and appreciation of the need to develop common languages; a transdisciplinary understanding of disparate areas of expertise as part of the curriculum, both for artist/designers as well as engineers [1].

Thus the design of the course focused on bringing students from separate fields of learning together; neither party relinquishing their areas of expertise and most importantly, neither party being required to learn the other's skills. In other words the aim was not to make artist/designers out of engineers and vice versa. What was aimed at was the formation of teams, comprised of one artist/designer and one computer engineer, each of whom would bring their individual skills, expertise, talent, creativity and viewpoints into the development of one single project, from its conception to its implementation to its final presentation.

1.1 Instructional Strategies

The course is a 3 credit undergraduate course, which 3rd and 4th year students from both the computer science as well as the art and design program can take, provided they have taken the prerequisite courses. For computer engineers this is an introductory course in computer graphics, for artists/designers basic design and typography.

The student body is limited to an even number – not to exceed 20 in total, comprised of an equal amount of students from each discipline since the methodology relies entirely upon the formation of equal numbered teams.

1.2 Developing a Transdisciplinary Syllabus

“A truly integrated curriculum entails the intellectual development of faculty as much as that of students, such that the two groups become reciprocal members of a shared, mutually self-critical learning community” [2].

It was inevitable that during the collation of the teaching material as well as the actual teaching of the course both instructors made major discoveries, not only about each others fields of expertise but also about one another’s thought processes and approach. This is felt by both to be an ultimately enriching experience, which has also spilled over into general teaching activity.

The first 5 weeks of teaching is devoted to two activities which run in tandem: The first of these is the supervision of the development of the individual projects which students embark upon during the second week of classes, as soon as team selection has been made. The second is the transmission of a rigorous survey of the areas of study. Students are given joint lectures, presented by both faculty members. The focus of these lectures is a presentation of a diverse range of innovative new media projects, such as the output of information visualisation, wearable computing, networked installations and performances, shape grammars to virtual worlds; approaching this content both from a programming as well as an aesthetic point of view. Students are encouraged to investigate the shortcomings as well as the successful aspects of this presented output, determining for themselves to what extent collaborative enterprise promises success in new media output.

Along with the presentation and discussion of the above material the instructors also give lectures on specific subjects, relating them to their own areas of expertise. Thus concepts such as “colour”, “light”, “space” are investigated both from a scientific as well as an artistic perspective. Thus, when broaching the subject of “light”, during a three hour class, diverse names and concepts, from Caravaggio to Gouraud Shading, from bioluminescence to the absence of light in Islamic miniature paintings are tossed about and speculated upon, constantly creating multiple associations and new pathways, compelling both students and faculty to make new connections, seek new means of expression and communication [1].

1.3 Project Development

The course depends entirely on successful teamwork and collaboration. To this end teams are formed by means of a lottery drawn during the second week of class, pairing one computer engineer with one artist/designer randomly, in a simulation of real work environments, where one is often compelled to collaborate with partners that one has no prior acquaintanceship with, that indeed one may not even particularly like and get along with.

The term project is then conceived of, formulated and developed jointly by the two team members. As stated above, it is not the intention of this course to make engineers out of artists and vice versa. Nevertheless during the conception and development of the

joint project both parties are compelled to think along the same lines, to empathize, to understand, to implement their unique as well as professional creative and problem solving abilities to the same end.

Engineers are urged to think creatively, while artists and designers are asked to be focused and pragmatic; both sides having to bring in considerable problem solving as well as problem formulating skills into the enterprise.

During the initial phases of development the mind maps are brought into class to be discussed by fellow students and faculty. Additions, subtractions and collations are made, culminating in a workable roadmap or ground plan upon which the work commences. At this juncture the two team partners start working on individual aspects of the project, the artist/designer focusing on the visual/audio language components, while the engineer starts to develop the programming environment. The close collaboration and the constant exchange of ideas, however, is expected to last during this crucial phase of operations and weekly class reports demonstrating the progression ensure that it is indeed maintained rigorously.

At the end of the 15 week semester the projects are handed in as fully operational computer applications/environments. Accompanying these is the delivery of a 5 minute long demonstrational video and a 3000 word documentation of the work. Students are expected to provide references as part of this documentation.

2 The Projects

2.1 Project 01: The Miro Avatar

This project originates from the desire to integrate an emotional presence into the Web by using EEG data to visualize personal emotions. However, since EEG data cannot have reliable or interpretable meanings concerning an emotional state, one may only speak of collecting the "idea" of an emotion rather than the accurate visual representation of an emotion. EEG is used to accumulate an individual's brain signals; signals that occur each moment, unconsciously, in response to the interaction with the immediate environment. Since it was desired to interpret these signals as the idea of one's emotional presence in virtual reality, a 3-D virtual world, Second Life became the platform to apply this metaphor of reality.

The collection of the EEG data has been done by an open source program, Brain-Bay, which outputs the biosignal as EDF files. These files are converted to ASCII text files with another open source program, Polyman. The content of these files are integers between -4000 and 4000. The ASCII files are uploaded to a website from which scripts in LSL read the files that contain the EEG data. The Avatar changes according to the incoming brain wave upon user activation: The script begins reading data from the server and the Avatar changes its shape according to the incoming integer values. Thus the user's brain waves form a virtual manifestation that represents his/her Avatar which can also be considered as a metaphor for the representation of one's mind; since, figuratively, what is visualized is the person's "thoughts".

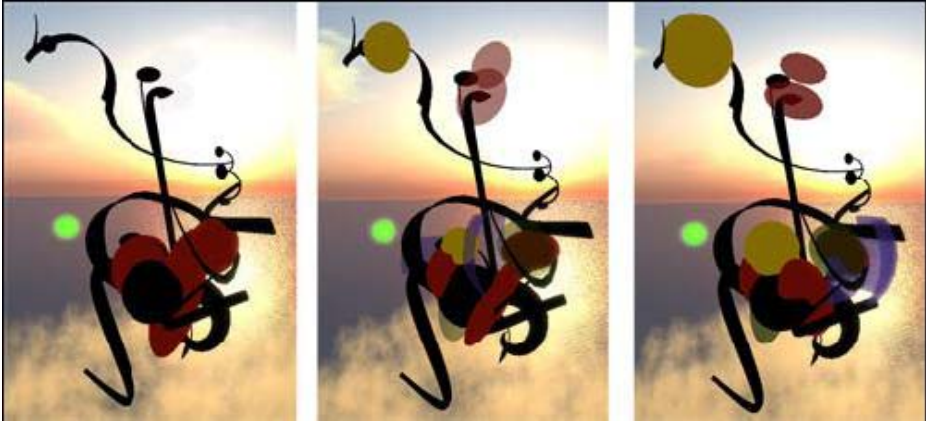


Fig. 1. The Miro Avatar: shown here changing its appearance through the data of the incoming brain wave

As far as the creative process is concerned, the visualization of one's emotional presence is inspired by the four dimensional painting ideas Miró had proposed in his late years. Thus, the avatar, composed of various visual elements featured in Miró's paintings, continuously changes its shape and is redrawn, transcending the two and three dimensionality of painting and sculpture. As expected, this representation contradicts with the regular tendencies of metaverse and MMORPG players who, usually, create a reflection of themselves by making an avatar corresponding to their actual appearance or giving it physical traits that they aspire to possess in real life.

The virtual spaces under consideration are simulations of real spaces which can solely be interacted and experienced through mental processes; via visual, audible, and cognitive stimulations in the brain. So, instead of creating an avatar based on actual physical traits, the output of the project offers an alternative visual entity, usable as an avatar, derived from the fact that users cannot have a real physical presence in virtual spaces and the fact that their mental input is the only factor that creates the illusion of presence in a virtual space. Other users can "see" them, not because they are physically there; but because there is an avatar that is shaped via their thoughts and desires with whom they may interact as if it were a real person. Thus, in terms of representation, virtual appearance may rely only on the output of unconscious thoughts mirroring the surrealist and expressionist approach of Miró's paintings [4].

2.2 Project 02: "Hairogram"

In Hairograph, the statistics of a nation, such as education, economy and population data are shown by using hair strands as the feature for representing multivariate data, creating an interactive tool for people who want to know about the statistics of a nation through the representation of a human head where each person or a group of people is shown by hair strands. In the area of representing multivariate data sets a commonly used method is combining differentiated volumes into one visualized/cohesive volume [5]. In Hairograph, a hair strand is capable of visualizing different variables by simply modifying its default parameters such as color and length.

For the study of visualizing national statistics, Worldmapper, uses morphing techniques to demonstrate how a country evolves for the given parameter such as migration, population or tourism [6]. In Hairograph, different hairstyles are generated for different countries regarding the selected statistics by using the combination of hair strands. Thus the motivation underlying the project is the similarity of the role of a single person or a group of people with homogenous characteristic within society and the similar contribution of a hair strand or a bunch of hair to the entire hair style. Although only one hair strand may not change the hairstyle, the combination of multiple hair strands may significantly affect it. Therefore, hair style is a powerful differentiating feature, which can be adapted to national differences. Moreover, hair is an effective feature for multivariable representation since every hair strand carries multiple properties such as branchiness, length and color.

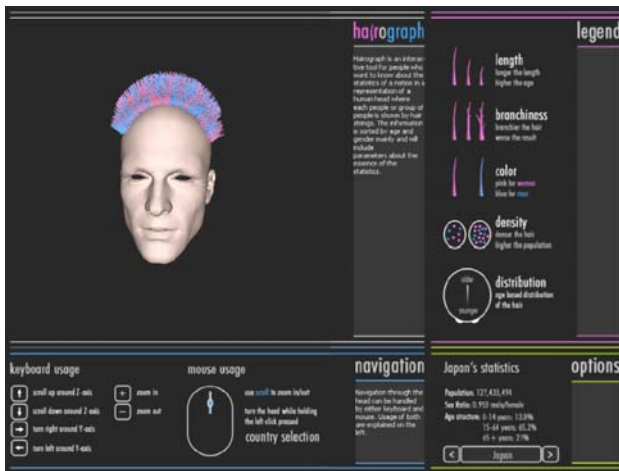


Fig. 2. Snapshot of “Hairogram” featuring the educational statistics of Japan

With Hairograph, a country is visualized as a human head where its hairstyle gives a brief and yet comprehensive overview information about the highlighted statistics. Realistic hair models were discarded in favor of an abstraction since a data visualization application would require to be comprised of abstractions for the optimum conveyance of the data.

The first variable used for hair strands is branchiness. Having more branches makes hair appear unhealthy and as such branchiness is a variable showing the corresponding statistics in such a way that when comparing two national economies, the country which has less branched hair strands is the country with the better economy. The second variable is length: For representing people’s ages, length was used as a variable where longer hair represents age and shorter hair stands for the young. Thus, length was used as a common variable which did not interfere with the information concerning well-being; whilst it gave information about overall statistics such as the distribution of old/young people in the country. The third variable used is color: Blue and pink were assigned to represent sex in accordance with the standard representational convention. The fourth parameter, density of hair, gives information about the population of the country.

Apart from these variables, the distribution of the hair is fixed in such a way that longer hairs are placed on the back of the head, whereas shorter hairs are towards the front, enabling users the comparison of two countries in regard to their younger and older population since during this location the strands are placed under a consideration of the percentage youth versus age within a population.

2.3 Project 03: The Reflexive Campus

The third project involves the creation of a generative, emergent architectural construct destined to become the virtual campus of a Real Life University on a dedicated island in Second Life®. Designed around a core spiral structure the virtual building re-configures itself based upon user demand, adding classrooms, exhibition and meeting areas as well as conference halls and auditoriums, as and when required. Thus, a user-centric, emergent, dynamic construct has been implemented for the creation of a virtual campus, which evolves and changes over time and based upon usage. The ensuing design system also attempts to configure itself based upon the nature of activity undertaken within it: Thus spaces required for survey based learning spread themselves out over the “shallow” horizontal axis, whereas areas of research and deeper levels of inquiry utilize the “deeper” vertical axis in a visual representation of different strategies of learning and levels of epistemological inquiry.

Inspired by two buildings, one real – the Great Mosque at Samarra, built in 847 AD, and one imaginary - the Tower of Babel as envisioned by Brueghel the Elder, in his so called painting from 1563; a spiral was decided upon as the building module around which the entire construct would be generated. However, it was not solely the fascination with these buildings that prompted this decision: When formal investigations commenced it was found that the continuity inherent in the circular and yet ever expanding shape of the spiral was uniquely suited to the demands of the task at hand, which was one of creating a construct capable of infinitely enlarging itself, capable of rendering parts of itself over and over again, and yet capable of maintaining its holistic structure throughout the generative process. An examination of Morphogenesis provided validation for the implementation of a spiral as the primary form of architectural growth, from which the campus would flourish, since the spiral is a shape also found in patterns of natural growth. [8]

Challenging Goethe's statement that “architecture is frozen music”, the creators of the Reflexive Campus built the construct around a core atrium, comprised of three half rings slowly rotating around a pivot. Each of these rings fulfills a separate function: Orienting the visitor, creating a catalyst for interaction and providing exhibition spaces for the university's activities. The complex emphasizes the tenets of dynamic/reflexive architecture [9] by expanding and shrinking, where the size is proportionate to the current occupant population and the shape corresponds to the usage that this population is demanding of the space at that time. Thus, while a day of intensive teaching of undergraduate courses will see a campus spread out over a wide horizontal axis accomplished through the rendering of dedicated modules/classrooms for each of the individual course taught; a day of intensive research activity or conferences will see a campus rising along a vertical axis through the rendering of dedicated auditoriums and meeting areas and seminar rooms. Finally, days of mixed activity and maximum usage will see a campus spreading on both axes.

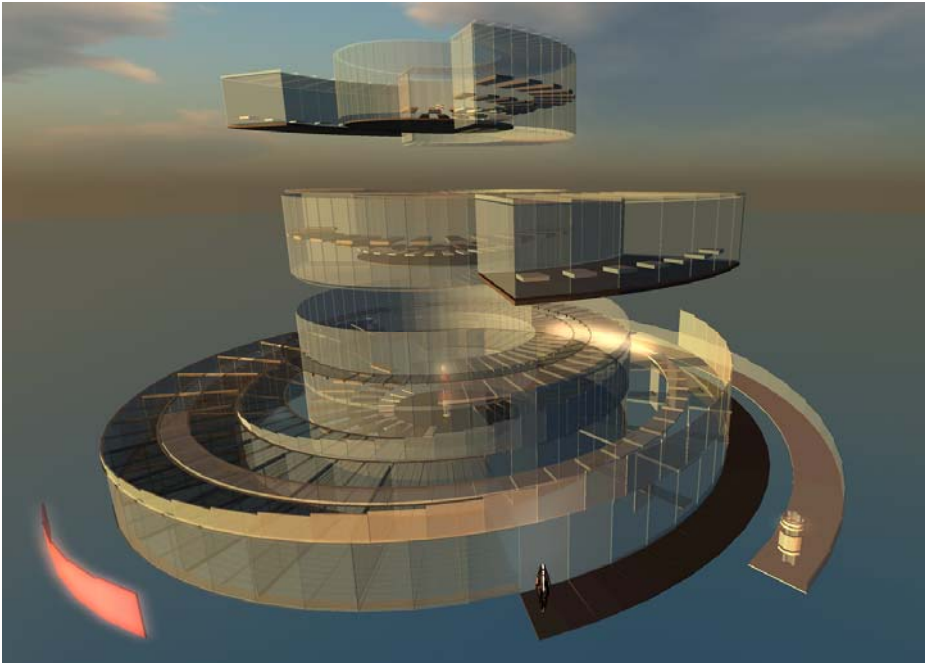


Fig. 3. The Reflexive Campus: Global view

Generating modules for classes or conferences is accomplished through clicking on a red sphere located in the central atrium by persons given the proper administrative authority to do so. The modules which create the spaces of learning/research activity can be categorized as the fully circular large auditoriums for 49 users each which render themselves one above the other directly aligned to the core of the construct, thus causing the building to rise vertically and secondly the pie shaped outer, smaller modules designated for 7 users, which spread outwards from the center in a sunray like manner, causing the construct to expand horizontally, allowing for a virtual, structural re-interpretation of the Platonic diagram of epistemology in the allocation of types of knowledge in relation to one another placed spatial axes as visual metaphors for the acquisition of knowledge.

3 Future Work

The course has proven to be a success in that a considerable number of student's projects have been written up and presented as research output by the students themselves at various international conferences. Thus, to utilize the findings, experience and output of the course in the creation of a transdisciplinary graduate program, as well as to integrate the results and findings of the course into the development of novel art and science curricula, both on the graduate as well as the undergraduate level, are other avenues pursued rigorously.

Acknowledgments. Heartfelt appreciation of, and thanks to the students, whose creative as well as academic rigor has formed the substance of this particular study: İsil Demir, Can Sen, Yigit Yuksel (The Miro Avatar); Berkay Kaya, Can Cecen (Hairograph); Bulut Sakcak, Fethi Ozdol (The Reflexive Campus).

References

1. Ayiter, E., Balcisoy, S., Germen, M.: Propagating Collaboration: An instructional methodology for artists and engineers. In: Proceedings of CC 2007: Seeding Creativity: Tools, Media and Environments, pp. 45–52. ACM, USA (2007)
2. Clark, M.E., Wawrytko, S.A. (eds.): Rethinking the Curriculum: Toward an Integrated, Interdisciplinary College Education. Greenwood Press, New York (1990)
3. Ascott, R.: Telematic Embrace: Visionary Theories of Art, Technology, and Consciousness, p. 38. University of California Press, Berkeley (2003)
4. Demir, I., Sen, C., Yuksel, Y.: Using EEG to visualize brain waves in second life, Engineering Reality of Virtual Reality. In: Proceedings of SPIE, vol. 7238 (2009)
5. Woodring, J., Shen, H.: Multi-variate, Time-varying, and Comparative Visualization with Contextual Cues. IEEE Transactions on Visualization and Computer Graphics 2006, 909–916 (2006)
6. Dorling, D., Barford, A., Newman, M.: WORLDMAPPER: the world as you’ve never seen it before. IEEE Transactions on Visualization and Computer Graphics 2006, 757–764 (2006)
7. Kaya, B., Cecen, C.: Hairograph: Synthesizing Statistics with Hair Style. InfoVis 2007, Poster Session, Sacramento, California, USA (2007)
8. Kemp, M.: Doing What Comes Naturally: Morphogenesis and the Limits of the Genetic Code. Art Journal 55 (1996)
9. Spiller, N.: Reflexive Architecture. Academy Press, UK (2002)

A Multimedia, Augmented Reality Interactive System for the Application of a Guided School Tour

Ko-Chun Lin¹, Sheng-Wen Huang¹, Sheng-Kai Chu¹, Ming-Wei Su¹,
Chia-Yen Chen¹, and Chi-Fa Chen²

¹ Department of Computer Science and Information Engineering,
National University of Kaohsiung, Taiwan

² Department of Electrical Engineering, I-Shou University, Taiwan
ayen@nuk.edu.tw

Abstract. The paper describes an implementation of a multimedia, augmented reality system used for a guided school tour. The aim of this work is to improve the level of interactions between a viewer and the system by means of augmented reality. In the implemented system, hand motions are captured via computer vision based approaches and analyzed to extract representative actions which are used to interact with the system. In this manner, tactile peripheral hardware such as keyboard and mouse can be eliminated. In addition, the proposed system also aims to reduce hardware related costs and avoid health risks associated with contaminations by contact in public areas.

Keywords: Augmented reality, computer vision, human computer interaction, multimedia interface.

1 Introduction

The popularity of computers has induced a wide spread usage of computers, as information providers, in public facilities such as museums or other tourist attractions. However, in most locations, the user is required to interact with the system via tactile means, for example, a mouse, a keyboard, or a touch screen. With a large number of users coming into contact with the hardware devices, it is hard to keep the devices free from bacteria and other harmful contaminants which may cause health concerns to subsequent users. In addition, constant handling increases the risk of damage to the devices, incurring higher maintenance cost to the providing party. Thus, it is our aim to design and implement an interactive system using computer vision approaches, such that the above mentioned negative effects may be eliminated. Moreover, we also intend to enhance the efficiency of the interface by increasing the amount of interaction, which can be achieved by means of a multimedia, user augmented reality interface.

In this work, we realized the proposed idea by implementing an interactive system that describes various school locations, both verbally and visually, according to the location pointed to by the user on a school map. The implemented system does not require the keyboard or the mouse for interaction; a camera and a printout of the map

are used to provide the necessary input instead. To use the system, the user is first given a printed school map, as often given out to visitors to the school, he/she can then check out the different locations by moving his/her finger across the paper and point to the marked locations on the map. The movement and location of the fingertip are captured by an overhead camera and the images are analyzed to determine the user's intended actions. In this manner, the user does not need to come into contact with anything other than the map that is given to him/her, thus eliminating health risks due to direct contact with harmful substances or contaminated surfaces.

The paper is organized as follows. Section 2 describes the various stages in the implementation of the system; section 3 demonstrates the usage of the implemented system, and finally in section 4, we discuss possible extensions and future applications of the system.

2 Background

Augmented reality (AR) has received a lot of attentions due to its attractive characteristics including real time immersive interactions and freedom from cumbersome hardware [10]. There have been many applications designed using AR technologies in areas such as medical applications, entertainment, military navigation, as well as many other new possibilities.

An AR system usually incorporates technologies from different fields. For example, technologies from computer graphics are required for the projection and embedding of virtual objects; video processing is required to display the virtual objects in real time; and computer vision technologies are required to analyse and interpret actions from input image frames. As such, an AR system is usually realized by a cross disciplinary combination of techniques.

Existing AR systems or applications often use designated markers, such as the AR encyclopedia or other applications written by ARToolkit [11]. The markers are often bi-coloured and without details to facilitate marker recognition. However, for the guidance application, we intend to have a system that is able to recognize colour and meaningful images of objects or buildings as printed on a brochure or guide book and use them for user interactions.

3 System Design and Implementation

The section describes how the system is designed and implemented. Issues that arose during the implementation of the system, as well as the approaches taken to resolve the issues are also discussed in the following.

3.1 System Design

To achieve the goals and ideas set out in the previous section, the system is designed with the following considerations.

Minimum direct contact. The need for a user to come into direct contact with hardware devices such as a keyboard, or a mouse, or a touch screen, should be minimized.

User friendliness. The system should be easy and intuitive to use, with simple interface and concise instructions.

Adaptability. The system should be able to handle other different but similar operations with minimum modifications.

Low cost. We wish to implement the system using readily available hardware, to demonstrate that the integration of simple hardware can have fascinating performance.

Simple and robust setup. Our goal is to have the system installed at various locations throughout the school, or other public facilities. By having a simple and robust setup, we reduce the chances of a system failure.

3.2 System Setup

In accordance to the considerations listed above, the system is designed to have the input and out interfaces as shown in Fig. 1. The system obtains input via a camera, located above and overlooking the map. The camera captures images of the user's hand and the map. The images are processed and analyzed to extract the motion and the location of the fingertip. The extracted information is used to determine the multimedia data, including text, pictures, sound files, and/or movie clips, to be displayed for the selected location on the map.

Based on the designed interface, the system's operations are implemented according to the following procedures. In preparation for the system's execution, the maps are scanned first, and image maps containing the feature points are constructed and stored in a database.

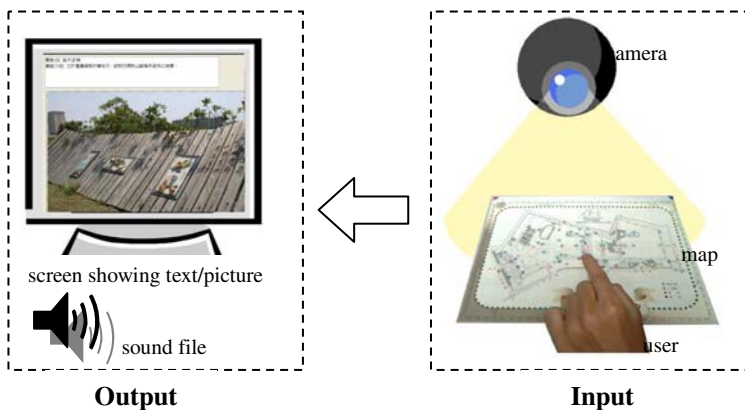


Fig. 1. Designed interface of the system

Upon the execution of the system, the camera is first adjusted according to the conditions of the operating environment. When a valid map is placed under the camera, the camera will acquire an image of the map and use it as a template for comparison. The system continues to monitor the map via the camera and checks for the presence and motion of the user's fingertip. Once the user's fingertip has stayed pointed to at a particular location for a given time interval, the system acknowledges that the user wishes to learn about the location. The system will then extract the coordinates of the fingertip and perform the required transformation to match the coordinates of the point with the image map. If the location is indeed a feature point, multimedia information associated with the location will be displayed on the screen and played through the speakers.

The flowchart in Fig. 2 describes the main stages in the system's operation.

In addition to the basic operations, the system is also able to perform in different modes and provide multi-lingual interactions. To demonstrate the ability, the system is implemented to recognize two different maps containing different feature points. There are also marked regions on the maps which allow the user to switch the supported language. Currently, we have implemented descriptions of the school in Mandarin, Taiwanese and English to provide a multi-lingual experience for school visitors.

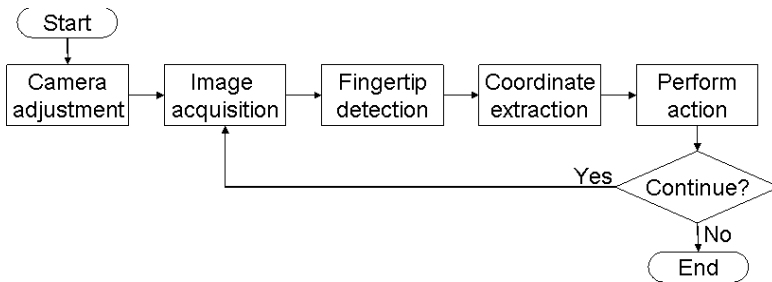


Fig. 2. Flowchart of the system's basic operations

3.3 Camera Adjustment

When the system is first setup, the camera needs to be adjusted for the surrounding environment. Parameters such as the distance between the camera and user's map, the direction of viewing, contrast and colours, need to be selected to obtain valid input images.

In our experiments, we used an USB camera placed vertically above and looking straight down at the map. The camera's resolution is set to 320 by 240 for image capture. It has been found from experiments that such a resolution is sufficient for a distance of 40 to 50 centimeters between the map and the camera. If necessary, the resolution of the captured images can be increased in other applications. The camera is also set to adjusted contrast and brightness automatically to avoid taking pictures that are too dull or too dark for processing.

3.4 Image Acquisition

The system starts by acquiring an image of a valid map that has been placed in the camera's field of view. The image is used as a template for background subtraction as long as the map is not moved too much. As the camera captures new images, each new image is compared with the background image to determine if the user has placed his/her finger onto the map. When there is a large difference between the background and the current captured image, the system checks for two conditions; user's finger over map, and movement of the map.

When the user's finger points to a location on the map and stays there for a given period of time (e.g. 500ms), the system acknowledges that the user wishes to know about the particular location and acquires the image for further processing.

In addition to checking for finger movements, the system also checks to see if the map is still in the same location. If the map has been moved beyond a certain threshold, the system will display a warning to the user and re-capture a background image. The current implementation requires that the map remains mostly in the same location during the execution of the program. Nevertheless, a more adaptable approach able to deal with map movements is considered for future versions of the system.

At this stage, we can also calculate the angle between the map in the input image and the horizontal axis, θ , as shown in Fig. 3. The value will be used in later process for coordinate transformation.

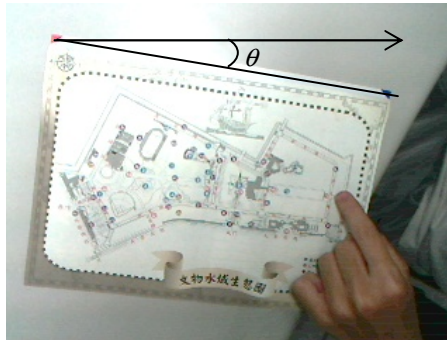


Fig. 3. An example input image and its rotation angle θ

3.5 Fingertip Detection

In the previous stage where the system detects that the user wishes to know about a particular location, a difference image between the background and the captured image would have been produced. The difference image is thresholded to produce a binary image. Morphological operations are performed on the binary image to determine location of the fingertip, (x_0, y_0) with respect to the map in the input image. The input image and the segmented image are shown in Fig. 4, note that the location of the detected fingertip is marked by a star, the boundaries of the detected map is shown as a rectangle.

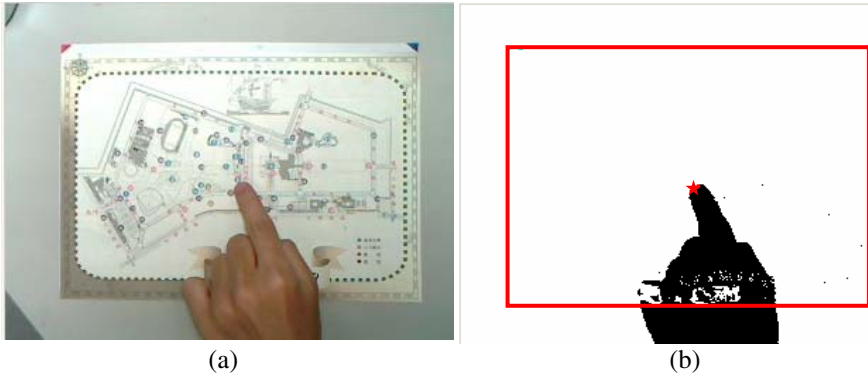


Fig. 4. (a) The input image with the user’s hand and (b) the located fingertip marked with a star after background subtraction and segmentation

3.6 Coordinate Extraction

The coordinates of the fingertip, (x_0, y_0) , with respect to the map, is transformed to the coordinates on the map stored in the database, (x, y) which are used to determine the information associated with the selected location. The coordinates (x, y) are obtained from (x_0, y_0) according to the following equation, in which θ is the map’s angle of rotation from the input image.

$$\begin{aligned} x &= x_0 * \cos \theta - y_0 * \sin \theta \\ y &= x_0 * \sin \theta + y_0 * \cos \theta \end{aligned} \tag{1}$$

Once the coordinates (x, y) are obtained, they are used to find a certain block number on the image map. An example of blocks on an image map is shown in Fig. 5, where the blocks have dimensions k by l and are numbered by integers. Eqn. 2 is used to find the block number of a set of coordinates.

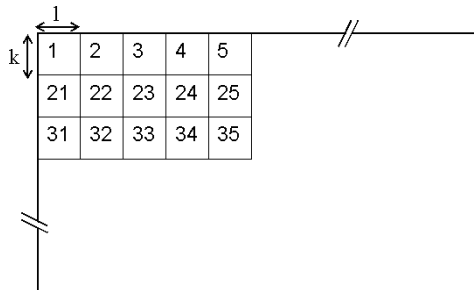


Fig. 5. Example of blocks on an image map

$$\text{block number} = W // l * (x \bmod k) + (y \bmod l) + 1 . \quad (2)$$

There are two reasons for dividing the image map into blocks. Firstly, the user's fingertip may not be very precise at pointing to a certain location on the map, therefore, there has to be a tolerance for selection, which is incorporated in the idea of a k by l block. Secondly, by converting a set of coordinates into integers, indexing can be performed more efficiently, since the feature points can each be given an associated block number for referencing.

3.7 Presentation of Selected Information

Each feature point on the map has its associated information, presented in the forms of text, images, and audio data. The data are displayed when the system determines that the user has selected the feature point. In addition, the system is also able to present the information in different languages. The default language is Mandarin, the official language in Taiwan. By pointing to marked regions on the side of the map, the user is able to interact with the system in English or Taiwanese, thus providing multi-lingual support for different users.

4 Discussion

The implemented system is operational and has received recognition in the departmental project competition. Nevertheless, the system is still in its prototype stage and there are many functions yet to be designed and implemented. For example, currently the system does not tolerate large displacement of the map during execution, since it will be unable to obtain a difference image for segmentation and fingertip detection. In the future, we plan to solve the problem by using a tracking algorithm for map and fingertip detection. In addition, we also wish to expand the system's database, such that it will be able to recognize more maps (and other informative materials) to provide more interactions. There are also plans to incorporate augmented reality functions into the system for more immersive interactions. Overall, the implemented system has shown promising potentials, as well as being fun to use. The system has also provided many ideas for future projects.

5 Conclusion

A multimedia, augmented reality interactive system has been designed and implemented in this work. In particular, the system is implemented to demonstrate its application in providing visitor information for the school.

The implemented system does not require the user to operate hardware devices such as the keyboard, mouse, or touch screen. Instead, computer vision approaches are used to obtain input information from the user via an overhead camera. As the user points to certain locations on the map with a finger, motions are detected and analysed by the system, and relevant information regarding the selected location are displayed. Hence, the user is able to operate the system without contacting any hardware device except for the printout of the maps.

The implementation of the system is hoped to reduce the cost of providing and maintaining peripheral hardware devices at information terminals. At the same time, eliminating health risks associated with contaminations by contact in public areas.

Work to enhance the system is ongoing and it is hoped that the system will be used widely in the future.

References

1. Liu, C.-H.: Hand Posture Recognition, Master thesis, Dept. of Computer Science & Eng., Yuan Ze University, Taiwan (2006)
2. Chen, C.-Y.: Virtual Mouse: Vision-Based Gesture Recognition, Master thesis, Dept. of Computer Science & Eng., National Sun Yat-sen University, Taiwan (2003)
3. Lai, J.C.: Research and Development of Interactive Physical Games Based on Computer Vision, Master thesis, Department of Information Communication, Yuan Ze University, Taiwan (2005)
4. Yeh, H.-C.: An Investigation of Web Interface Modal on Interaction Design - Based on the Project of Burg Ziesar in Germany and the Web of National Palace Museum in Taiwan, Master thesis, Dept. of Industrial Design Graduate Institute of Innovation & Design, National Taipei University of Technology, Taiwan (2007)
5. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall, Englewood Cliffs (2002)
6. Jain, R., Kasturi, R., Schunck, B.G.: Machine Vision. McGraw-Hill, New York (1995)
7. Klette, R., Schluns, K., Koschan, A.: Computer vision: three-dimensional data from images. Springer, Singapore (1998)
8. Brown, T., Thomas, R.C.: Finger tracking for the digital desk. In: First Australasian User Interface Conference, vol. 22(5), pp. 11–16 (2000)
9. Wellner, P.: Interacting with papers on the DigitalDesk. Communications of the ACM, 28–35 (1993)
10. Azuma, R.T.: A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6, 355–385 (1997)
11. Augmented Reality Network (2008), <http://augmentedreality.ning.com>

Reconstruction of Cultural Artifact Using Structured Lighting with Densified Stereo Correspondence

H.-J. Chien¹, C.-Y. Chen¹, and C.-F. Chen²

¹Computer Science & Information Engineering, National University of Kaohsiung, Taiwan

²Electrical Engineering, I-Shou University, Taiwan
ayen@nuk.edu.tw

Abstract. In recent years a variety of 3D scanning technologies has been applied to the field of Heritage Preserve. The paper describes an operational 3D scanning system that has been set up at Kaohsiung Museum of History for digital archival of cultural artifacts. Also, an approach that refines resolution of stereo correspondence to improve model quality obtained by structured lighting is proposed. The reconstructed models are shown to evaluate performance of improved scanner.

Keywords: 3D reconstruction, structured lighting, stereo correspondence, digital archival, heritage preservation.

1 Introduction

Computer vision based shape recovery techniques have been widely applied to preservation of cultural heritages in recent years [1][2][3][4]. These techniques are able to reconstruct 3D models in an accurate, non-tactile, and non-destructive manner without involving high energy projection, making them the suitable for digitalization of fragile objects. Structured lighting is a popular approach because it works on most surface materials, from perfect diffuse to specular and even under water, as long as the camera can sense change of intensity due to light projection. Furthermore, it does not rely on textures characteristics, thus it can be applied to featureless surface.

In a typical stereo vision setup, two or more cameras are placed to look toward the same object. The relationship between pixels associated with same 3D coordinates in different frames is known as the *stereo correspondence* between the images [5], which may be difficult to solve. Computation intensive techniques like block matching or epipolar analysis are required, and in the worst case only degenerated correspondence is available.

Structured lighting simplifies the correspondence problem by using an active projection device to replace one of two cameras in stereo vision [4], [6]. The accuracy of structured lighting is directly influenced by the accuracy of lens parameter and stereo correspondence. Theoretically, the range measurement can be quite accurate if the cameras are well-calibrated and the projector's correspondence is precisely determined. However in real applications, many practical issues exist. One of the most noticeable constraints is the heterogeneity of image dimension. An off-the-shelf projector is often capable to project a screen of up to 1024 by 768 pixels. However, the resolution of an

image captured by a typical digital camera could be at least two times the projector's resolution. We have observed that such a mismatch in resolution causes recovered surface to have a saw-toothed profile, as will be discussed later.

This paper describes an operational 3D scanning system that has been set up at Kaohsiung Museum of History for digital archival of cultural artifacts as well as an approach that improves model quality obtained by structured lighting. The rest of the paper is organized as follows. In section 2 an overview to our reconstruction system is stated, and the subsequent section 3 focuses on the establishment, evaluation and improvement of stereo correspondence. The performance of our proposal applied to real cases is studied in section 4, and finally the conclusion is given in section 5.

2 System Setup

This section describes how the various components in the system are placed to triangulate the 3D coordinates, as well as the calibration method adopted.

2.1 Physical Setup

The basic setup of the proposed structured light system consists of one camera and one projector placed in front of the object to be reconstructed. To acquire a complete 3D model the object is placed on a turntable. The position of camera is adjusted so that the rotation axis of the turntable is as close as possible to the camera's XZ plane. In this way, we can obtain not only meshes from different views but also volumetric model by the *Shape from Contour* technique [7]. The intersection of the camera and the projector's fields of view decides the system's effective view. Therefore the projector is arranged to make sure that space covered in the effective view is large enough to fit in the cultural artifact from all viewing directions. The setup of the system is shown in Fig. 1.

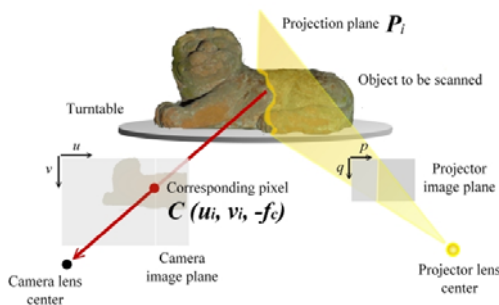


Fig. 1. 3D relationships between the camera, data projector and the object

The distance from the camera's optical center to object surface can be measured from the lens parameters and the correspondence. Assume these two prerequisites are available, the triangulation algorithm gives depth of pixel (u_i, v_i) by finding intersect

point of the back-projected ray L_i through the camera's center and the corresponding vertical projection plane P_i . In equation (1) and (2) the ray in camera-centered space and the plane in projector-centered space are defined respectively regarding to camera focal length f_c and angle between projection plane and optics axis of projector α_i . Note the coordinates (u_i, v_i) are shifted to match center of image plane.

$$L_i = s \left(\frac{\hat{u}_i}{f_c}, \frac{\hat{v}_i}{f_c}, -1 \right), s \in \Re \quad (1)$$

$$P_i = (p \cos \alpha_i, q, -p \sin \alpha_i), \quad p, q \in \Re \quad (2)$$

We obtain the transform from camera space to projector space from the extrinsic parameters. Let a 3 by 3 rotation matrix R and a 3 by 1 translation vector T denote that transform, equations (1) and (2) form a linear system as shown in equation (3).

$$\begin{pmatrix} p \cos \alpha_i \\ q \\ -p \sin \alpha_i \end{pmatrix} = s \cdot R \begin{pmatrix} \hat{u}_i / f_c \\ \hat{v}_i / f_c \\ -1 \end{pmatrix} + T \quad (3)$$

Although the linear system can be solved by computing its inverse matrix, dividing third row by first yields a straightforward solution of s , as shown by equation (4).

$$s = - \frac{T_x \tan(\alpha_i) + T_z}{[R_x \tan(\alpha_i) + R_z] \cdot \left(\frac{\hat{u}_i}{f_c}, \frac{\hat{v}_i}{f_c}, -1 \right)} \quad (4)$$

2.2 Lens Calibration

The main task of calibration is to find intrinsic and extrinsic parameters of the lens, and the correctness of calibration directly affects system accuracy. Once parameters of both camera and projector are acquired, the linkage between two devices can be constructed for triangulation. Typically a projector is treated as an inversed pinhole device as it uses backward perspective projection [8][9].

There are several mature and practical methods for camera calibration [10]. Our system is calibrated by a variation of Tsai's method. A checkerboard pattern is used as a reference object for Tsai's method. To start the calibration, the board is placed in the scene and an image is captured by the camera. The corners are extracted and used to compute the homography transform. Rectified control points are returned to facilitate the labeling procedure. By solving the constructed over-determined linear system all sixteen initial parameters are established. We then apply the Levenberg-Marquardt method for non-linear optimisation to minimize re-projection error. To preserve orthogonality and normality of rotation matrix, we augmented the objective function by adding significantly weighted penalty terms, since these two characteristics are important and should not be neglected during optimisation.

Calibration for the projector is similar, with the exception of being unable to capture images. Therefore, we need to find the correspondence first. Once found, the mapping is reversed to generate a partial view from the projector, since many pixels

may map to same pixel location in the projector’s image plane. Image enhancement and interpolation may be required to generate a complete view, as shown in Fig. 2. The calibrated parameters are shown in table 1. Despite the degeneracy of using planar calibration object instead of volumetric objects (e.g. a cube), multiple calibrations yield consistent result.

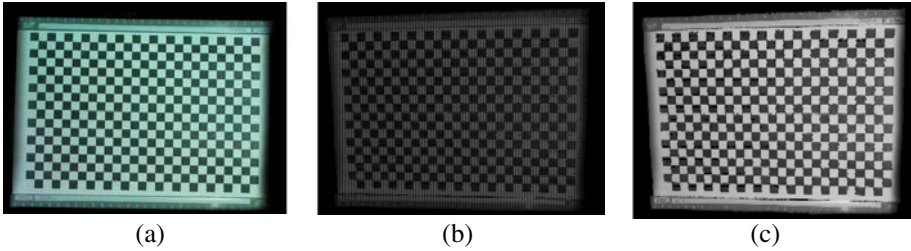


Fig. 2. (a) the camera’s view of a reference checkerboard, (b) the projector’s view of the same scene, obtained from correspondence, and (c) after image enhancements. Note (b) looks dark due to loss of noninvertible information caused by sparse correspondence.

Table 1. Calibrated parameters of camera and projector lens

	Camera			Projector		
Focal length	3566 pixels			2449 pixels		
Principle	(969.03, 643.93)			(509.29, 384.61)		
Pixel aspect ratio	1.0013			1.0086		
Translation	(12.50, 4.26, -86.87)			(18.22, 1.59, -126.04)		
Rotation matrix	-0.99	0.00	-0.12	-0.99	0.02	-0.11
	-0.01	-1.00	0.05	-0.02	-1.01	0.02
	-0.12	0.04	0.99	0.16	0.03	0.99

3 Stereo Correspondence

Stereo correspondence is a prerequisite to perform 3D measurement for all shape recovery methods based on multiple view geometry [5]. A correspondence states how pixels are related in two image planes. In practical applications, such pixel-wise relationships are too discrete and not dense enough to produce smooth surfaces. This section focuses on an approach that reliably improves density of correspondence.

3.1 Establishing Correspondence

Comparing to passive vision the structured lighting technique obtains the correspondence in relatively straightforward way. Gray-coded pattern projection has been considered a robust and fast technique to establish pixel-wise mapping between image plane of the camera and of the projector [7]. Since it uses a time-multiplexed binary coding scheme, a sequence of light patterns which consist of bi-colour stripes is projected onto the scene over time. The one-dimensional coordinate of each unit strip can be uniquely identified from decoded binary images.

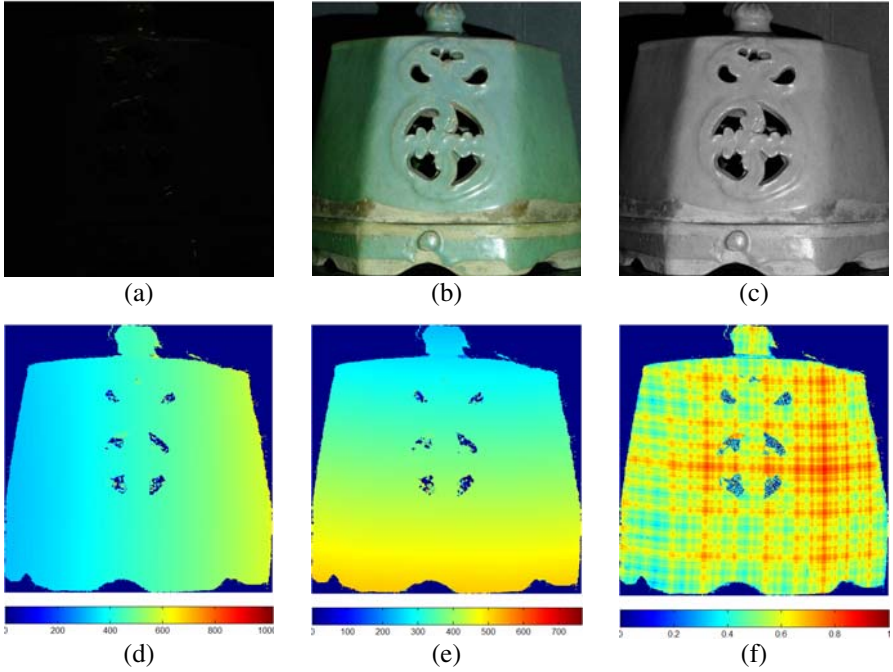


Fig. 3. First row: images of scene illuminated by uniform (a) black and (b) white patterns used to define the interval and (c) the denominator of intensity ratio R_k . Second row: (d) and (e) show stereo correspondences in x - and y -coordinates respectively, (f) shows the confidence for pixel correspondence.

Reflectance-based adaptive binarization is a commonly used method to decide whether a pixel is illuminated by a certain colour of a light pattern. Non-uniform thresholds are computed according to per-pixel intensity ratio given in equation (5). Two additional patterns are required to define range $[I_{min} I_{max}]$. Fig. 3(a-c) shows an artifact lit by patterns of different colour (black and white) and resultant intensity range.

$$R_k(u, v) = \frac{I_k(u, v) - I_{min}(u, v)}{I_{max}(u, v) - I_{min}(u, v)} \quad (5)$$

The number of patterns needed to establish dense correspondence is logarithmically bounded by number of unit strips. For a projector that has 1024 pixels in the horizontal direction, ten projections of light pattern are required to label all 1024 unit strips. However in practice, dense correspondence is often not feasible due to limits of heterogeneous image dimensions, placement of devices, boundary uncertainty, light interference, noises and difficulty in extraction of frequently interleaved stripe. Fig. 3(d-f) visualizes established vertical and horizontal 512-stripe correspondences, both increases monotonically due to strict ordering property.

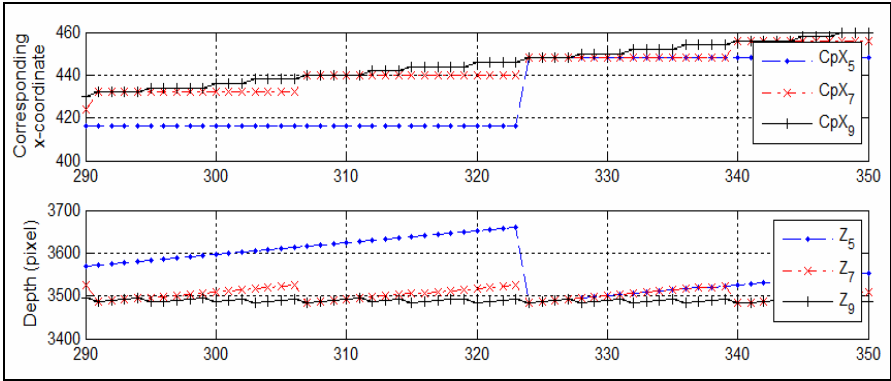


Fig. 4. Upper plot displays x-coordinate correspondences obtained using 5, 7, and 9 Gray-coded patterns respectively. The lower are triangulated depths with respect to each correspondence.

Although the visualized results may look smooth, the correspondences are discretely stepping, causing sudden changes in recovered surfaces, as shown by the profile in Fig. 4. By reducing number of unit stripes the saw-toothed effect becomes more significant. Our approach to solve the problem is to adjust the correspondence in each stepping interval such that the correspondence grows smoothly. An intuitive approach would be to use linear interpolation for correspondence adjustment.

3.2 Quantitative Confidence

A dual-peaked function is used to evaluate the reliability of decoded information. In particular, the value of this function depends on intensity ratio and is peaked at zero and one. As the ratio moves away from these peaks the confidence indicator decays gradually and approaches zero where the ratio becomes insignificant.

The modeling of the confidence indicator is reasonable as each surface patch is assumed to behave as being illuminated by either a white or a black stripe. Factors such as light inter-reflection and noise may cause the ratio to shift away from expected position, hence a tolerance is given. We model the function by summing a zero-mean and a one-mean Gaussian function. The deviations are adjustable to control the tolerable uncertainty. Fig. 3(c) shows the evaluated pixel confidence.

3.3 Correspondence Adjustment

More flexible adjustment on correspondence can be achieved by taking the quantitative confidence into account. The idea is to assign new values to unsure elements by referencing their reliable neighbors. Such model is actualized by equation (6), where f is the original correspondence, w is the confidence, d is a function for spatial proximity, and g is the new correspondence. In our implementation, d is a radial basis function such that the correspondence can be adjusted using convolution.

$$g(u, v) = \frac{\sum_{i=0}^n \sum_{j=0}^m d_{u,v}(i, j) \cdot w(i, j) \cdot f(i, j)}{\sum_{i=0}^n \sum_{j=0}^m d_{u,v}(i, j) \cdot w(i, j)} \quad (6)$$

4 Experiment

We have implemented the proposed approach and applied it several objects. We choose the cap of the cultural object (in Fig. 3) to demonstrate the quality of the reconstructed model. The system has a projector with 512 vertical stripes. Therefore each stripe occupied averagely 800/512 (1.5625) pixel width in the projector's image plane. The 512-level correspondence is further expanded to sub-pixel accuracy using the proposed method. Fig. 5 shows primordial-decoded, convoluted, as well as spline interpolated correspondences of a profile. The recovered surface using correspondence in sub-stripe resolution is shown in Fig. 6(a). In contrast, Fig. 6(b) looks rugged due to sparse correspondence.

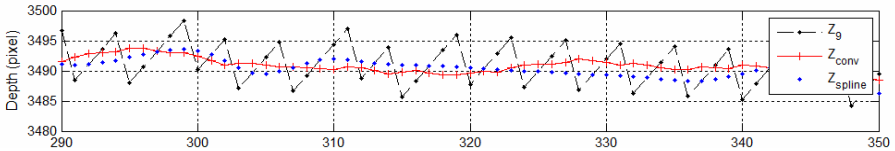


Fig. 5. Comparison between original, convoluted, and interpolated correspondences. It shows a significant reduction of depth ramping. The interpolation doesn't take confidence indicator into account, may result in loss of shape features.



Fig. 6. (a) shows the surface recovered from densified correspondence. Comparing to (b) the reconstruction using original correspondence, the saw-toothed effect is quite unobvious.

5 Conclusion and Future Work

This paper proposes an approach to improve structured light system accuracy. Results have been provided to demonstrate the improvement of the proposed approach. By applying confidence-based convolution on decoded correspondence the resolution is increased. As the result, the saw-toothed phenomenon is removed from surface. Our

method does not require smoothing and down sampling operation as a conventional structured lighting scanner may do.

Future research will focus on further improvement of the 3D scanning system. As observed during experiments, the calibration errors of projector's lens are much significant than camera's. Therefore, it is reasonable to design a projector-specific calibration method. Also, the approach of not calculating the parameters explicitly but reconstructing the projector from calibrated camera with established correspondence is considered for future work.

Acknowledgments. This research is sponsored by National Digital Archives Program, grant number NSC 97-2631-H-390 -001. Cultural artifacts, working place, and permission to access restricted warehouse are kindly provided by Kaohsiung Museum of History.

References

1. Guidi, G., Beraldin, J.A., Atzeni, C.: High-accuracy 3D modeling of cultural heritage: the digitizing of Donatello's Maddalena. *13th IEEE Transactions on image processing* 13(3), 370–380 (2004)
2. Anestis, K., Fotis, A., Christodoulos, C.: On 3D reconstruction of the old city of Xanthi A minimum budget approach to virtual touring based on photogrammetry. *Journal of Cultural Heritage* 8(1), 26–31 (2007)
3. 3D Modelling of Castles,
http://www.gim-international.com/issues/articles/id1073-D_Modelling_of_Castles.html
4. Rocchini, C., Cignoni, P., Montani, C., Pingi, P., Scopigno, R.: A Low Cost 3D Scanner based on Structured Light. In: *Eurographics 2001*, vol. 20(3), pp. 299–308 (2001)
5. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press, Cambridge (2004)
6. Salvi, J., Pags, J., Batlle, J.: Codification Strategies in Structured Light Systems. *J. PR* 37, 827–849 (2004)
7. Chen, C.-Y., Klette, R., Chen, C.-F.: *Shape from Photometric Stereo and Contours*. Springer, Berlin (2003)
8. Batlle, J., Mouaddib, E., Salvi, J.: Recent Progress in Coded Structured Light as a Technique to Solve the Correspondence Problem: A Survey. *J. PR* 31, 963–982 (1998)
9. Gühring, J.: Dense 3-D Surface Acquisition by Structured Light using Off-The-Shelf Components. In: *SPIE Photonics West, Videometrics VII*, vol. 4309, pp. 22–23 (2001)
10. Klette, R., Schlüns, K., Koschan, A.: *Computer Vision - Three-dimensional Data from Images*. Springer, Singapore (1998)

Motion Generation for Glove Puppet Show with Procedural Animation

Chih-Chung Lin¹, Gee-Chin Hou², and Tsai-Yen Li¹

¹ Department of Computer Science, National Chengchi University, Taiwan
{s9421, li}@cs.nccu.edu.tw

² Department of TV and Radio, National Chengchi University, Taiwan
phou@nccu.edu.tw

Abstract. Traditional Taiwanese glove puppet show is a unique form of performing art. In this work, we aim to use the aid of computer animation technologies to preserve this cultural heritage and create innovative ways of performance. By observing the demonstration of a puppet master, we analyze the characteristics of how a hand puppet is manipulated and design animation procedures mimicking the motion of a glove puppet. These procedures are implemented on a real-time animation platform for procedural animation. Through the system, we allow a user to perform glove puppet animation with high-level inputs. We hope that, with the help of this system, not only the art of manipulating a glove puppet can be systematically documented, but the entry barrier for learning it can also be greatly reduced.

Keywords: Character Animation, Digital Content, Glove Puppet Show, Procedural Animation, Traditional Art.

1 Introduction

Taiwanese glove puppet show is one of the most popular folk arts in Taiwan and is also an indispensable entertainment in the daily life of Taiwan society. It combines elements of different traditional drama, music, delicate costumes and props. Being an exquisite art of action, the manipulation of glove puppet requires ability of space perception and hand-eye coordination with high level of precision. Because of the nature of these psychomotor skills, repeated practice is a must in order to achieve a mastery level of performance. Young puppeteers are usually trained by oral instructions. With no written-down materials, a senior puppeteer let apprentices know a simplified storyline and have them observe closely when he/she performs. It usually took apprentices more than three years to be competent as the assistant of principal puppeteer. In addition, this kind of show relies heavily on teamwork. The puppets are usually manipulated by one or two puppeteers in a show. The principal puppeteer usually manipulates all puppets and performs the narratives, dialogs and singing all by himself/herself.

In recent years, the applications of computer animation are becoming prevailing due to the fast development of hardware and software for computer graphics. These applications can be found in entertainment, education, and commerce. In addition to

commercial films and TV, computer animation also serves as an effective way to preserve traditional performing arts. For example, Li and Hsu [1] have designed an authoring system that can create the animations of traditional shadow play with high-level user inputs. The goal of designing this type of animation system is not only to capture the art in a digital form but also to lower the entry barrier for learning and producing such a show. Furthermore, we hope that innovative forms of performing art can be invented to take advantage of digital technologies.

Puppet animation is a special type of character animation. Currently, the mainstream methods for generating character animations can be divided into three categories. The first type of method generates character animation by sampled data. For example, as low-level signals, sampled data can be obtained with the motion capture technologies when the motions are performed by a real actor [4]. Animations made with this technology are more plausible. However, without knowing the meaning and structure of the motions, it is also difficult to modify a captured motion to fit the constraints of a new environment. The second way is about using commercial 3D animation software to set the key frames of a character by animation professionals and interpolate between key frames to generate the final animations. However, even for good animators, producing an animation in this way is usually labor intensive and time consuming. The third way is about generating animations by simulation or planning procedures [1][8]. It is also called knowledge-based animation since it usually requires a dynamics model of the objects under simulation or motion-specific knowledge to plan and generate the motions. Due to the complexity of the underlying computational model, this type of animations usually is generated in an off-line manner and displayed in real time. Nevertheless, due to the increasing computing power of computers, it is becoming more feasible to generate this kind of animations in real time with appropriate model simplification.

In this paper, we use the approach of procedural animation to generate the animation of traditional glove puppet show. Procedural animation belongs to the third type of methods for animation generation since it uses the knowledge of glove puppet manipulation to design animation procedures. A motion usually is divided into multiple phases delimited by keyframes with distinct spatial constraints in each phase. Then the motion between keyframes is computed by appropriate interpolation functions capturing temporal or spatial constraints. Animations produced with this approach have the advantages of being flexible and computationally efficient for real-time environments. Nevertheless, it is a great challenge to design a procedure with appropriate parameters that can produce plausible motions.

2 Related Work

Most of previous work on traditional Taiwanese glove puppet show was conducted via the motion capture approach. For example, recently the X-Lab at National Chao-Tung University captured the motion of the hand playing a glove puppet with a data glove [9]. They also have implemented a system with a user interface allowing a user to create a puppet show by putting small pieces of motions together. In addition to displaying the motions with 3D graphics, they have also designed a puppet robot of the same size as the hand puppet that can mimic the captured motion. In our work, the

design goal is different from capturing and reproducing the hand motion of a puppet show. Instead, we hope to realize the animation by modeling the way that a hand puppet is manipulated in a procedural manner with parameters. By adjusting the parameters, we can produce a good variety of flexible puppet animations.

Puppet animations usually consist of two types of motions: intended primary motions and passive secondary motions. For example, the motions for the body, head and hands are primary motions manipulated by the animator while the motions of the legs, clothes, and hair belong to the secondary motions that move passively according to the primary motion, the gravity, and the environment. Previous work divided the problem into layers of abstraction and treated the secondary motions at the lower layers. For example, in [2], the main actions of a character were acquired by the motion capture technique and then the associated secondary motions were computed. Perlin also proposed to add noise to the primary motion as part of the secondary motion to make the generated motion look more realistic [8].

Li and Hsu [5] also distinguished these two types of motions for producing 2D traditional shadow play animation semi-automatically. The authoring tool allows a user to specify few key postures of the animated character in a 2D environment and then uses a motion planning algorithm to compute the path of primary motions. Secondary motions are then computed according to the obstacle constraints and the gravity force.

3 Animation System Design

3.1 Animation Platform

In this work, we aim to design a 3D animation system customized for real-time display of glove puppet animation. The system is constructed based on an experimental testbed for procedural animation, call IMHAP (Intelligent Media Lab's Humanoid Animation Platform), that was proposed in our previous work [7]. The architecture of the system is shown in Fig. 1. The system was designed with the MVC (Model, View,

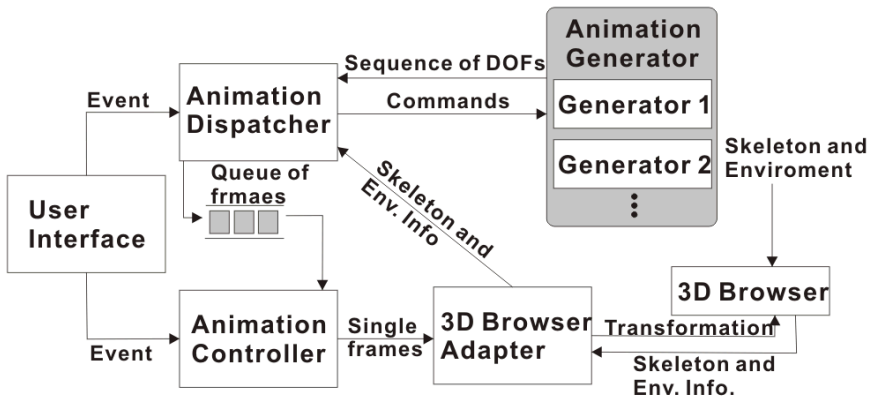


Fig. 1. System architecture of the puppet animation platform

and Controller) design pattern. Model is used for encapsulating data, view is used for interaction with the user, and controller is used for communication between model and view. In Fig. 1, every block represents a module, and the arrows between the modules stands for data flow. In these modules, animation dispatcher and animation generator are the kernel of this animation system defining how the animation generation problem is decomposed into elementary procedures implemented in the animation generators. In MVC, these two components play the role of *model* while the animation controller plays the role of *controller*. The last two components, 3D browser and user interface, play the role of *view* that enables interactive display. For the 3D display, we have adopted the open-source JMonkey package as our 3D browser [3].

3.2 Kinematics Model

The kinematics structure of a puppet is similar to a human figure with minor differences as shown in Fig. 2. For example, the major active joints of a puppet are at the neck, shoulder, thigh, and pelvis. The joints of elbows and knees are usually ignored. The motions of the legs are triggered either by another hand or by the gravity during swinging. The clothes are the passive part that moves with the legs as they swing. As shown in Fig. 2, we attach additional links and joints to the clothes at the front and back of the legs such that we can move the clothes when the legs are lifted.

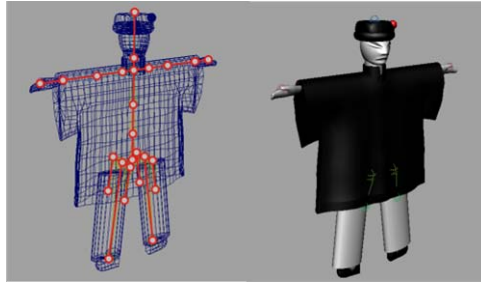


Fig. 2. Kinematics model of the hand puppet including the additional linkages for clothes

3.3 Motion Model

As mentioned in the previous subsection, in procedural animation, different motions require different animation generators. We divide the process of designing an animation procedure into three main steps. In the first step, we observe the motions performed by a master puppeteer and define appropriate motion parameters that can be used to describe the characteristics of the motion. For example, for the walk motion, the adjustable motion parameters include step length, body swing, and so on. In the second step, we decompose the motion into several phases by defining key frames separating these phases according to the motion parameters described above. In the last step, we define the procedure for interpolation between the key frames. Two types of interpolations are needed to produce plausible motions. One is to interpolate along a curve for trajectory following while the other defines how a motion is timed. In the current system, linear interpolation is used for simple motions while Bezier curves are used for more sophisticated motions.

To illustrate how the motions are generated in a procedural manner, we use three different types of motions as examples: *walk*, *run* and *roll-over*, as described below in more details.

3.3.1 Walk

There exist various ways of walking with different meanings under different scenarios. Advised by our puppet master, we have divided a normal walk into several phases. In the walk motion, the adjustable parameters include step length, step height and body swing angle. As shown in Fig. 3, we define four keyframes according to the given motion parameters.

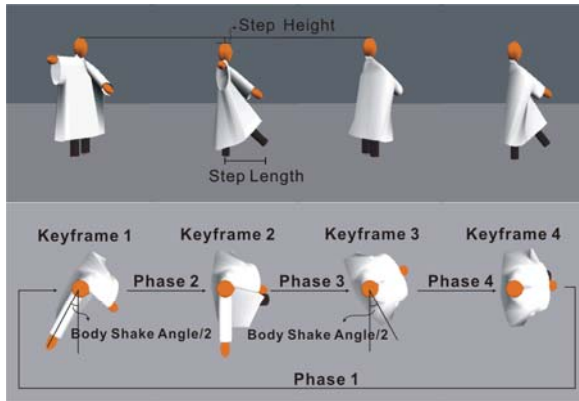


Fig. 3. The keyframes and the procedure parameters for the walk motion

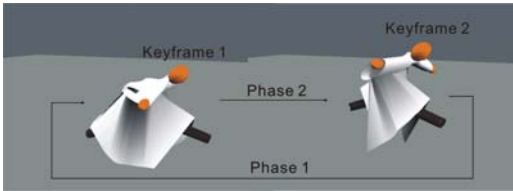


Fig. 4. The key frames of the run motion

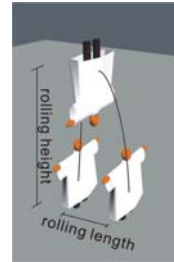


Fig. 5. A snapshot of the roll-over motion

3.3.2 Run

The key of creating this type of motions by a puppeteer is on a sudden swing of one leg upward and then moving the body up and down to swing both legs further to create the running motion as shown in Fig. 4. The swinging leg motions belong to the type of secondary motions since the legs move in accordance with the primary motion of the body. The speed and frequency of a body moving up and down determine the rotational speed of the legs swinging back and forth.

3.3.3 Roll-over

The roll-over motion is specified by two parameters: rolling height and rolling length. Two key frames are needed to define this motion. A Bezier curve is defined with two parameters and used as the trajectory of the puppet. The puppet also must make a complete revolution on the orientation along the trajectory. A snapshot of the generated animation for the roll-over motion is shown in Fig. 5.

4 Experimental Results

We have implemented a java-based animation system that allows a user to design the motions in a glove puppet show with an interactive graphical interface as shown in Fig. 6. At the upper-left corner, the system hosts a 3D browser that can display the animations generated by the user-designed procedures. The user can compile a sequence of motions (such as walk, run, and roll-over) by selecting them and setting their parameters from the graphical user interface on the right. The lowest part of the platform is the interface for play control that allows a user to view the animation continuously or frame-by-frame.

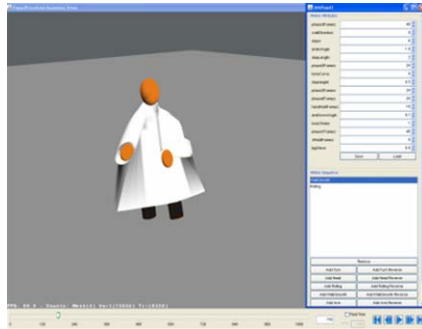


Fig. 6. Graphical user interface for interactive animation generation of puppet show

To illustrate that the motions generated by our animation procedures are rather flexible, we use different sets of parameters to generate animations. In Fig. 7(a), we show the walking motion generated with parameters of (swing_angle, step_length) = (0.3, 1.5) while the parameters are (1, 0.8) for Fig. 7(b). In Fig. 7(c), we show a running motion generated with the parameters of (step_length, step_height)=(0.5, 2) while the parameters are (1.5, 3) for Fig. 7(d). In Fig. 7(e), we show a rolling-over motion generated with the parameters of (rolling_height, rolling_length)=(10, 5) while the parameters are (20, 10) for Fig. 7(f). In Fig. 8, we show a composite example generated by compiling multiple motions into a final show. The puppet ran and then rolled over to make a complete revolution. The motions generated on the platform can be further exported into a MEL script for further rendering in animation packages such as Maya. An example of the motion rendered in Maya is shown in Fig. 9.

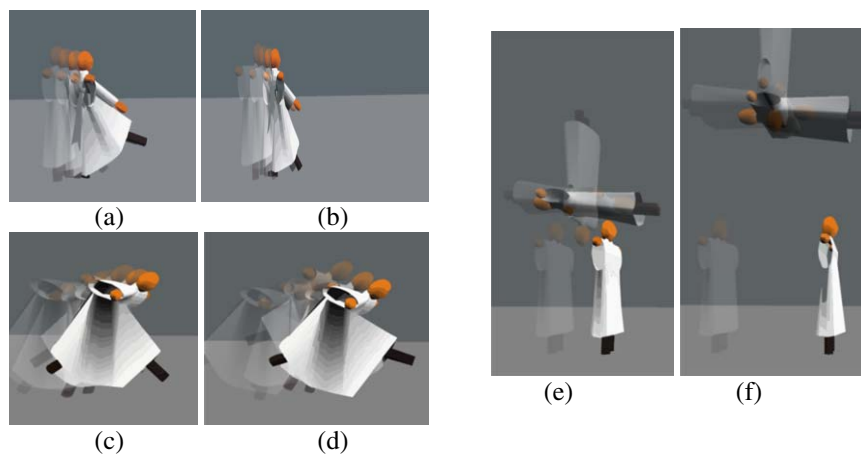


Fig. 7. Using different sets of motion parameters for walking (a)(b), running (c)(d), and rolling-over (e)(f)

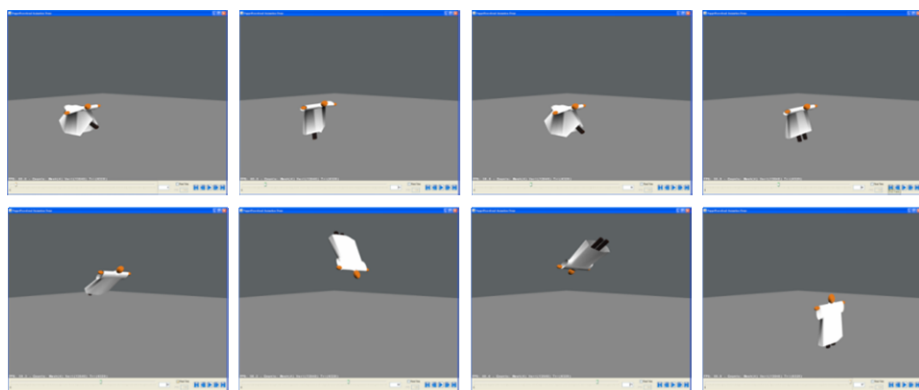


Fig. 8. An example of compiling different motions together. The puppet entered with a running motion and then did a roll-over motion at an end.



Fig. 9. An example of exported animation rendered in Maya

5 Conclusion and Future Work

With the help of computer animation techniques, we hope to preserve and promote the art of glove puppet show, a precious culture heritage in Taiwan. In this paper, we have made a first attempt to model the basic motions of a glove puppet by observing the performance of an expert puppeteer and implementing the motions with procedural animation. These animation procedures will work as reusable basic units for compiling a play in the future. Furthermore, we hope to design an animation scripting language that can be used to document the motions of glove puppets as well as the screenplay. We also hope that the music behind the script can be generated automatically according to the tempo of the show and synchronized with the cues from the puppeteer.

References

- [1] Chen, P.F., Li, T.Y.: Generating Humanoid Lower-Body Motions with Real-time Planning. In: Proceedings of 2002 Computer Graphics Workshop, Taiwan (2002)
- [2] Dontcheva, M., Yngve, G., Popović, Z.: Layered Acting For Character Animation. *ACM Transactions on Graphics* (2003)
- [3] jMonkey Engine, <http://www.jmonkeyengine.com/>
- [4] Kovar, L., Gleicher, M., Pighin, F.: Motion Graphs. In: Proceedings of ACM SIGGRAPH (2002)
- [5] Li, T.Y., Hsu, S.W.: An Authoring Tool for Generating Shadow Play Animations with Motion Planning Techniques. *International Journal of Innovative Computing, Information, and Control* 36(B), 1601–1612 (2007)
- [6] Li, T.Y., Liao, M.Y., Tao, P.C.: IMNET: An Experimental Testbed for Extensible Multi-user Virtual Environment Systems. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3480, pp. 957–966. Springer, Heidelberg (2005)
- [7] Liang, C.H., Tao, P.C., Li, T.Y.: IMHAP – An Experimental Platform for Humanoid Procedural Animation. In: Proceedings of the third International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Tainan, Taiwan (2007)
- [8] Perlin, K.: Real Time Responsive Animation with Personality. *IEEE Transactions on Visualization and Computer Graphics* (1995)
- [9] X-lab of National Chao-Tung University, http://xlab.cn.nctu.edu.tw/modules/x_movie/x_movie_view.php?cid=1&lid=35

Can't See the Forest: Using an Evolutionary Algorithm to Produce an Animated Artwork

Karen Trist, Vic Ciesielski, and Perry Barile

RMIT University

GPO Box 2476V, Melbourne, Vic, Australia

{karen.trist,vic.ciesielski,perry.barile}@rmit.edu.au

Abstract. We describe an artist's journey of working with an evolutionary algorithm to create an artwork suitable for exhibition in a gallery. Software based on the evolutionary algorithm produces animations which engage the viewer with a target image slowly emerging from a random collection of grey-scale lines. The artwork consists of a grid of movies of eucalyptus tree targets. Each movie resolves with different aesthetic qualities, tempo and energy. The artist exercises creative control by choice of target and values for evolutionary and drawing parameters.

Keywords: Evolutionary programming, evolved art, new media art, software art, animation, algorithmic art, AI, genetic programming, genetic art, non photorealistic rendering.

1 Introduction

This paper describes the construction of an artwork using a software system for evolutionary art [1] based on genetic programming [5]. The artwork, 'Can't see the forest', consists of a grid of animated drawings of trees created by using a software program we call 'The Shroud' (after the Shroud of Turin), for its capacity to produce a sort of ghost-like rendering or trace of a target image. We explore this evolutionary art process by describing the created work and the artistic and technical considerations that we encountered in its production. The goal of this collaboration between artist and computer scientist was ultimately to produce an engaging animation suitable for exhibiting in an art gallery.

The Shroud software produces visually engaging animated movies drawing grey level straight-line strokes at various angles and of finite lengths that give a variety of likenesses to a photographic or drawn target image. Figure 1 shows a sequence from a run showing the target with a series of frames of the subject gradually evolving¹. The movies are of two main kinds: (1) start with a canvas on which a large number of lines are drawn and keep redrawing the canvas until a recognisable subject emerges (see Figure 1), and (2) start with a blank canvas and keep adding lines until a recognisable subject emerges (see Figure 2). In both cases the animations engage the

¹ Animations of some of the panels are at http://evol-art.cs.rmit.edu.au/npr_theshroud/

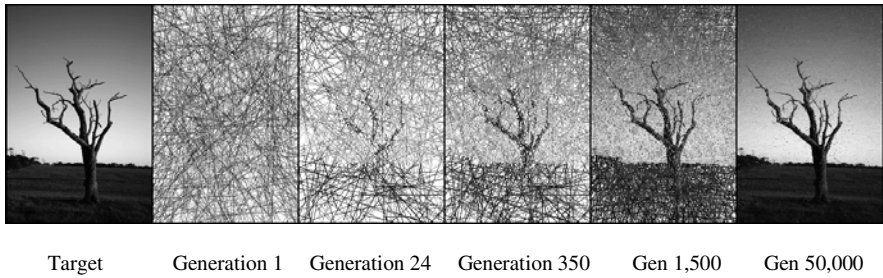


Fig. 1. Sequence from a run of the evolutionary algorithm

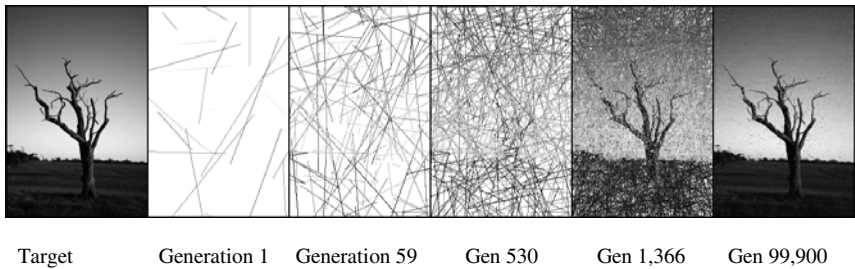


Fig. 2. Sequence from a run of the evolutionary algorithm starting from a blank canvas

viewer’s attention as the viewer waits to see what will emerge. The challenge for the artist is to use the programs to produce animations that are not too fast or too slow, that have interesting targets, that keep the viewer in suspense for long enough, but not too long, and to assess to what level of clarity the target will be revealed.

Having artists working with computer scientists produces some interesting cultural and technical issues. For example, in genetic programming from a computer science perspective, the desire would generally be to find an optimal result quickly and utilising the fewest resources. However, because we are here concerned with generating perhaps unusual results based on subjective, aesthetic criteria, the focus is more on ways for the programmer to give the artist increased creative control, rather than just finding the best level of fitness to a target image as quickly as possible.

The finished artwork consists of a seemingly infinite loop of animated ‘drawings’, with no clear start, middle or end. Each run converges more or less towards the target image with its own starting and ending points on the main loop. The duration of the individual loops also vary (being factors of the main loop). The overall effect is to produce a dynamic piece with a balance of resolved and semi-resolved renderings of the target images over time. See Figure 3 for an example of a random sequence.

2 Related Work

A wide variety of evolutionary approaches have been used to generate art works, for example genetic algorithms, genetic programming and ant colony optimisation. A number of these art works have been on a botanical/flora/plant growth theme. In Panspermia

[9] a simulated botanical environment was built using a genetic model of growth. Sims later expanded this work with Genetic Images [8] in which viewers were able to interactively re-sequence genetic markers used in the production of evolving images. In [6], a survey is presented of various evolutionary approaches to adapting L-Systems – simple languages that allow rules to be generated that model plant growth.

Many contemporary artists are producing work (installations, painting, photography etc) that responds to issues of sustainability in the environment. Patricia Piccinini's work 'Signs of Life' from the 'Plasticology' series consists of '50 monitors with computer generated plants constantly swaying on the wind' [7], inspired by the artist's ongoing exploration of 'bio-scientific practices of manipulation and alteration of living beings, of creating new worlds.' [2]. Lyndal Jones's 'Avoca Project' responds to the impacts of climate change in a novel way through intervention in the environment and a series of site-inspired artworks [4]. Our artwork was also inspired by the pressing problems evident today such as global warming, deforestation and habitat destruction.

Repetition is an important creative practice for artists, whether used in the process of working up ideas, (a painter might do many small studies of a subject before the final, full-size painting), or as an essential feature of the finished work. Impressionist painter Claude Monet produced many studies of repetition of a theme over his long career: of haystacks, cathedrals and of the water lilies in his garden at Giverny, 'in order to display minute discriminations of perception, the shift of light and colour from hour to hour on a haystack, and how these could be recorded by the subtlety of eye and hand' [3,p359].

Pop artist Andy Warhol also used repetition extensively in his many series of popular culture and iconic figures. However, his silk screen print grids are not about recording changes in light but his repetitions instead 'mimic the condition of mass advertising' [3,p359] speaking 'eloquently about the condition of image overload in a media-saturated culture'[3,p540]. Warhol produced variations on a single image, repeating 'these images until repetition is magnified into a theme of variance and invariance, and of the success and failures of identicalness' [10].

3 Artistic Inspiration

The artwork interrogates concepts of environmental sustainability through creating a digital dying 'forest' (through repetition in a grid format, which suggests a continuation beyond the frame) of evolving and devolving animated movies of dead eucalypt trees. The idea of using genetic programming to create this work seemed an appropriate way to explore the effects of deforestation and global warming on our fragile and continually evolving environment in Australia. It also seemed apt that the programs were represented as tree structures, while the targets are also literally trees. The eucalypts photographed are naturally hardy and long-lived, however many have succumbed to drought in the last few years and can be seen in paddocks on farms and in forests, weathered silver and ghost-like.

The artist lived in Tasmania during the heyday in the 1970s of clear felling for wood-chipping in the old growth forests. The artist had been dismayed to see windrows of the leftover trees bulldozed into heaps for eventual burning. In the Shroud

program, when the tree size is small there are very few lines drawn and the aesthetic effect is very like a game played by the artist when she was a child called ‘Pick Up Sticks’. It was played by letting the sticks fall like spaghetti into a saucepan - holding them upright and then letting them fall into a heap. The trick was to pick up as many sticks as you could without making the others move. The simple, straight lines drawn in the Shroud software reminded the author of this childhood pastime, only then for her consciousness to shift back to the tragedy of the piles of logs in the Tasmanian forests. In ‘Pick Up Sticks’ the black sticks earned more points. In the wind-rows of bulldozed trees they would all be burnt to black.

Artistic possibilities were suggested during the process of software testing. Normally an artist drawing would use a rough sketch as a beginning, which would then usually be worked up into a more finished piece. However, as a result of doing a large series of runs and then playing them alongside each other on the computer screen as they evolved differently and at different rates, the artist was inspired by the creative and narrative possibilities inherent in combining both the ostensible ‘failures’ and the ‘successes’. Thus the software experimentation drove the idea of using a grid of multiple animations for the final artwork with some not fully resolving.

The artist was not interested in simply reproducing the target images in another form – animated drawings. She wanted to extend the creative possibilities by utilising the well-established art practice of repetition. The software was ideal for creating a digital dying ‘forest’ as it allowed for almost limitless versions of the target. However, the concept of regeneration is also present in the process, as the animations on a loop continually evolve towards the tree targets and then dissolve into random chaos, only to be reborn again.

4 Shroud Program

4.1 The Evolutionary Algorithm

Evolutionary algorithms are characterized by a population of potential solutions that improve over the course of the evolutionary process. In our case an individual is a program that draws an image on a canvas and the measure of improvement, or fitness, is resemblance to a target image. See [1] for a complete technical description.

In the first generation the population of programs is created randomly. Each program is able to generate a sequence of greyscale brush strokes on the canvas. The program is represented as a tree structure (not to be confused with the tree images that are the targets of the animations). The parameters of brush strokes such as line length, intensity and angle are determined at this time. The basic evolutionary process proceeds in the following manner:

- a) Wipe the canvas clean and draw the output of one program on to the canvas.
- b) Compare this rendered canvas to the target image. A rendering with a greater similarity to the target image is assigned a high measure of ‘fitness’, while a rendering with lesser similarity to the target image is assigned a lower measure of ‘fitness’.
- c) Do this rendering, perform comparison to the target image and assign fitness to each program, wiping the canvas clean prior to each rendering.

At the end of this process the programs are listed in order of fitness. The programs now enter a phase called the mating phase. Programs are combined with one another in a process similar to crossover in biological reproduction. Parts of two programs are combined and two 'offspring' are produced and then placed into the population of the next generation. The selection of mating partners is based on the fitness of individual programs. Offspring are also generated by a process of mutation in which a random change is made to a single parent. Programs with higher fitness have a greater probability of being selected for mating. This is a process similar to Darwinian evolution through natural selection. Parts of each parent go into the offspring. The theory is that the building blocks (the parts of programs that render brush strokes) of programs with higher fitness will be propagated into a new generation, while building blocks that do not promote high fitness will disappear from the 'gene pool'. The evolutionary process is terminated when a pre-specified number of generations has been completed or a pre-specified level of similarity to the target image is reached. Whenever a program is found that has a higher fitness than the best program in previous generations, a frame of the final animation is generated. There are a number of parameters which control the evolutionary and drawing processes. These are shown in Table 1.

Table 1. Parameters of the evolutionary and drawing processes

Program Tree Size	The maximum number of strokes that can be drawn. Larger programs draw more strokes.
Number of Generations	The number of generations to execute before stopping.
Fitness Target	Stop when the fitness reaches this value.
Write Nth Best	Rather than rendering every best individual discovered as a frame, only write every Nth.
Fitness Write Threshold	Only write a frame if the fitness has improved this amount since the last frame written.
Accelerated Convergence	In the basic program a pixel on the canvas is always changed based on values in a draw function. If this parameter is set a pixel is only changed if the new value would be closer to the target value.
Incremental Rendering	Rather than starting with a blank canvas the rendering starts with the best canvas from the previous generation.

5 Producing the Work with 'Shroud'

The artist exercises creative control by choice and preprocessing of targets and by choosing values for the evolutionary and drawing parameters described in Table 1. Some combinations of parameters lead to unusual and unexpected effects.

5.1 Selection and Preparation of Suitable Target Images

The trees were photographed so as to include minimum background distraction so the tree stands out clearly against a light-coloured sky. The artist tried, where possible, to

keep the scale and the horizon line reasonably constant for continuity between individual animations within the final piece. The target images chosen needed enough contrast and strong outlines so that even in the small tree size animations we would be able to make out the beginnings of a ‘drawing’ of a tree, even though very few lines are drawn.

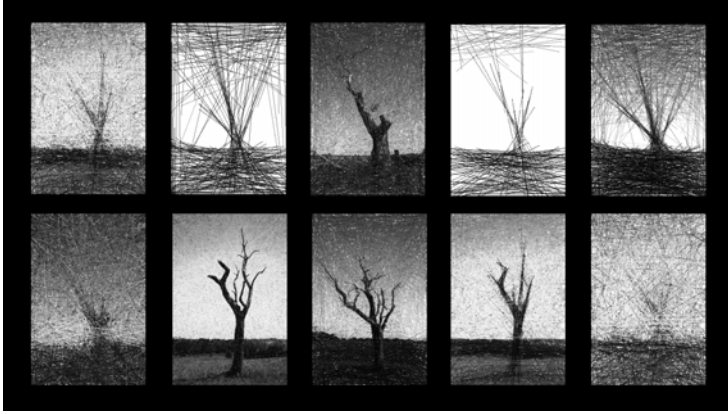


Fig. 3. Random sequence from artwork

5.2 Choices for the Artist

- a) How close to the target should the final image be? In some situations photorealism is desired. In other situations a fuzzy, suggestive abstraction is desired. The level of photorealism is controlled by a number of parameters acting in concert. Large values for *Max Generations* and *Program Tree Size* and turning on *Incremental Rendering* will give more photorealism. Beyond a certain number of generations there is often minimal improvement. However, too few generations will often not produce anything resembling the target, particularly with smaller tree sizes.
- b) At what rate should the target emerge? Usually the artist’s goal is to have the target emerge at a rate that will keep the viewer’s interest. If the target emerges too fast there is no mystery or anticipation. If the target emerges too slowly the viewer loses interest. The rate at which the target emerges is primarily controlled by the parameters *Write Nth Best* and *Fitness Write Threshold*. Smaller values of these parameters will give longer animations or animation segments and more frames to be viewed before the target is revealed. Setting *Accelerated Convergence* and *Incremental Rendering* will generally decrease the time before the target is revealed. Setting a large *Program tree size* and turning on *Accelerated Convergence* can result in the target being revealed in the very first frame.
- c) Should the target be built up from a blank canvas adding more and more lines or should it emerge slowly from a large number of randomly drawn lines? A small tree size and *Incremental Rendering* will result in the former effect while a large tree size without *Incremental Rendering* will give the latter effect.

- d) How much energy should be in the animation? Some animations exhibit a considerable energy in that lines are continually being drawn and redrawn. In other animations the drawing of the lines is not as obvious and the target slowly emerges without such vigorous line drawing on the canvas. Generally small program sizes and *Incremental Rendering* turned off give the former effect while larger program sizes and *Incremental Rendering* turned on give the latter effect.

The individual images in Figure 3 are snapshots from animations that have been created with a variety of choices for the parameters described in this section.

6 Reflection on Challenges and Opportunities

The artist's main practice is still photography. Working with series of images to form a narrative is common practice in photography. However, the opportunity to work with subtle repetitions of an image and then animating the results into video format for presentation was creatively stimulating.

A challenge for programmer is that the evaluative process is done on subjective and aesthetic (rather than efficiency) criteria. It can be challenging for a programmer to evaluate success without using objective and clear criteria. Ideally, in evolutionary programming, the animation would evolve to closely resemble the target image in as few generations as possible. But for the artist planning a final artwork using many different animations in a grid or on multiple screens, it might be more desirable to be able to utilise the steps along the way in the artwork, as opposed to using only the final or 'best' result. An artist, for whom the process is often the most engaging part of the work, will often react creatively to accidents rather than having a pre-defined set of criteria for successful outcomes. The programmer saw himself in this process as a mediator between the artist and the machine.

Combinations of different choices can give quite a wide range of different effects, some desirable and others not. For example, large tree sizes and incremental rendering can give a too obvious resolving of the target in the first generation. Conversely, turning off *Accelerated Convergence* and *Incremental Rendering* can produce many thousands of generations and still not reveal the target.

7 Conclusions and Future Work

'Can't see the forest' is a work that has been created by exploratory and systematic use of a software system based on evolutionary algorithms. It adds creatively to the conversation around issues of environmental sustainability. The feedback loop established between the artist and the programmer functioned to quickly solve technical problems and to provide artistic inspiration through experimenting with the different iterations of the software as development progressed. This iterative process was very satisfying and fruitful creatively.

A major frustration for the artist with the use of the Shroud software is that some of the runs take hours or even days. In future work we plan to develop algorithms that will run much faster and give the artist much earlier feedback on the conse-

quences of parameter choices. We also plan to give the artist more control over the length and thickness of the lines drawn, to work with coloured lines and images, and to look at the possibility of changing the parameters during the course of a run.

Acknowledgments. We thank our colleagues Dr. M. Berry, D. Keep and Dr. D. D'Souza for suggestions relating to this paper and RMIT's Design Research Institute for seed funding.

References

1. Barile, P., Ciesielski, V., Trist, K., Berry, M.: Animated Drawings Rendered by Genetic Programming. In: GECCO 2009: Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation, pp. 939–946. ACM, Montreal (2009)
2. Haraway, D.: Speculative Fabulations for Technoculture's Generations: Taking Care of Unexpected Country, Essay (2007), <http://www.patriciapiccinini.net> (accessed 24-April-2009)
3. Hughes, R.: American Visions: The Epic History of Art in America. Harvill Press (1997)
4. Jones, L.: The Avoca Project (2009), <http://www.avocaproject.org/> (accessed 24-April-2009)
5. Koza, J.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge (1992)
6. McCormack, J.: Evolutionary L-systems. In: Hingston, P.F., Barone, L.C., Michalewicz, Z. (eds.) Design by Evolution: Advances in Evolutionary Design, pp. 168–196. Springer, Berlin (2008)
7. Piccinini, P.: Signs of Life (1997), <http://www.patriciapiccinini.net> (accessed 24-April-2009)
8. Sims, K.: Artificial Evolution for Computer Graphics. In: SIGGRAPH 1991: Proceedings of the 18th Annual Conference on Computer graphics and Interactive Techniques, pp. 319–328. ACM, New York (1991)
9. Sims, K.: Panspermia (1990), <http://www.karlsims.com/panspermia.html> (accessed 27-April-2009)
10. Wilson, W.: Prince of Boredom: The Repetitions and Passivities of Andy Warhol (1968), <http://www.warholstars.org/andywarhol/articles/william/wilson.html> (accessed 24-4-2009)

Automatic Generation of Caricatures with Multiple Expressions Using Transformative Approach

Wen-Hung Liao and Chien-An Lai

Department of Computer Science, National Chengchi University,
Taipei, Taiwan
{whliao, g9202}@cs.nccu.edu.tw

Abstract. The proliferation of digital cameras has changed the way we create and share photos. Novel forms of photo composition and reproduction have surfaced in recent years. In this paper, we present an automatic caricature generation system using transformative approaches. By combining facial feature detection, image segmentation and image warping/morphing techniques, the system is able to generate stylized caricature using only one reference image. When more than one reference sample are available, the system can either choose the best fit based on shape matching, or synthesize a composite style using polymorph technique. The system can also produce multiple expressions by controlling a subset of MPEG-4 facial animation parameters (FAP). Finally, to enable flexible manipulation of the synthetic caricature, we also investigate issues such as color quantization and raster-to-vector conversion. A major strength of our method is that the synthesized caricature bears a higher degree of resemblance to the real person than traditional component-based approaches.

Keywords: Caricature generation, facial animation parameters, image morphing, facial feature localization.

1 Introduction

Caricatures are exaggerated, cartoon-like portraits that attempt to capture the essence of the subject with a bit of humor or sarcasm. These portraits are usually drawn by professional cartoonists/painters based on their personal judgment or representative style. Recently, interests in creating systems that can automatically generate 2D caricatures using photo as the input have surged due to the proliferation of image capturing devices as well as advances in facial feature analysis. Some system (e.g., Mii editor in Wii) employs a *component-based* approach, in which a collection of facial components are pre-defined and then selected to assemble a cartoon face either manually or by similarity analysis of individual facial parts. A major issue with component-based approach is that the composite cartoons exhibit limited varieties in their styles due to the nature of template selection. Even worse, the resulting caricature sometimes bear little resemble to the person even when the individual parts are matched flawlessly. In *example-based* approaches [1][2], a shape exaggeration model is learned from a labeled example set. At runtime, the facial shape is extracted using

active shape model (ASM) and then classified into one of the exaggeration prototypes for further deformation. It appears that shape plays a primary role (compared to texture) in this type of approach. As a result, the computer-generated caricatures are usually monochrome, lacking details in textural elements.

To address the issues described above, we have developed and implemented an automatic caricature generation system using a *transformative framework* [3]. In this type of approach, a new caricature is synthesized by transferring the style from a reference cartoon drawing via matching of the two sets of pre-defined control points in the input photo and the reference image, as depicted in Fig. 1. In [4], we further improve the precision of the feature localization result using active appearance model (AAM). We also propose to create a 3D caricature model by merging two planar caricatures using principles of binocular vision.

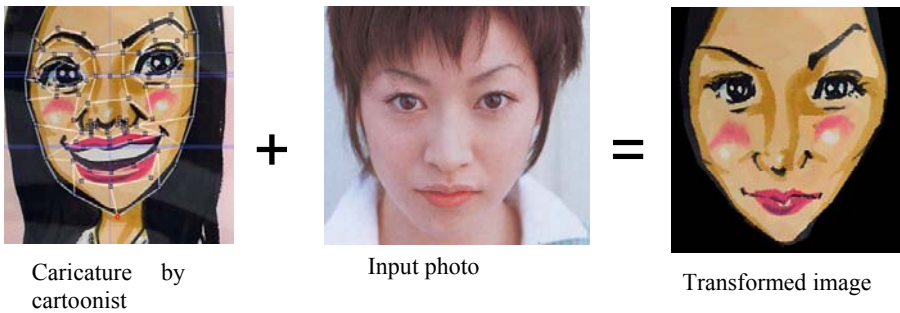


Fig. 1. Transformative approach for generating caricature from a single reference drawing [3]

The research described in this paper marks an extension to our previous efforts. Specifically, we incorporate the MPEG-4 facial animation parameters into the shape model to enable of automatic generation of caricature with multiple expressions using a single reference drawing. We also exploit the idea of creating new drawing styles by blending features from multiple sources using polymorph technique [5]. As the collection of reference drawings becomes larger, it will be convenient to select/recommend a proper template based on shape similarity measure. Finally, to facilitate efficient storage and flexible manipulation of the synthesized caricature, we look into issues regarding color reduction and raster-to-vector conversion.

The rest of this paper is organized as follows. In Section 2, we first review the facial mesh model designed for our application. We then present two general approaches for creating caricatures from an input photo and a reference cartoon. The second part of Section 2 is concerned with template selection using shape information, as well as style blending using polymorph technique. Section 3 defines the 2D FAP set employed in our system and the associated facial component actions, which are further combined to derive archetypal expression profiles, i.e., different emotion types. Results of synthesized images with multiple expressions are presented as Emoticons. Section 4 concludes this paper with a brief summary.

2 Caricature Generation

The overall procedure for generating caricature from input photos is illustrated in Fig. 2. In the first step, a frontal-view photo is presented as the source image. Using active appearance model (AAM), we can locate important facial components and the corresponding mesh, as shown in Fig. 3. The mesh is comprised of 8 components, with a total of 66 control points and 114 constrained conforming Delaunay triangles (CCDT) for defining the facial configuration.

The caricature can then be generated using two different approaches: image morphing or rotoscoping. In image morphing, a collection of reference cartoons with different styles needs to be gathered. Control points as well as the corresponding mesh of these cartoons will be labeled manually. The user will choose a template to transfer the style from. We then proceed to identify the corresponding triangular meshes between the reference cartoon and the input image. After the correspondence has been established, it is a simple matter to apply image morphing technique to transfer the texture from the reference drawing to the input photo. In our current implementation, barycentric coordinate is employed to speed up the morphing process. Details regarding the coordinate transformation on CCDT can be found in [6]. Typical morphing results are shown in Fig. 4.

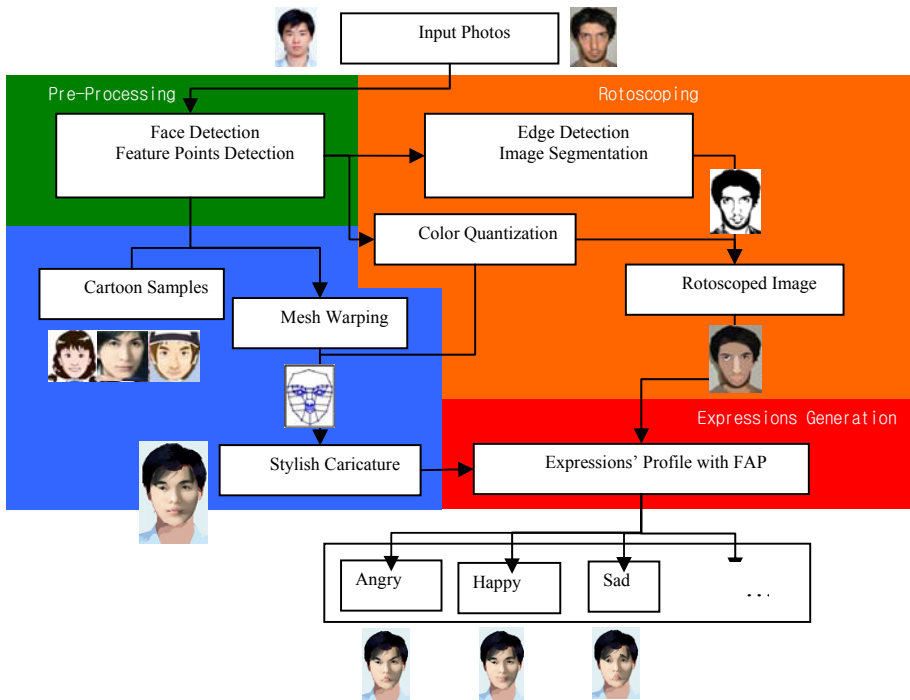


Fig. 2. Proposed framework for generating caricatures with multiple expressions

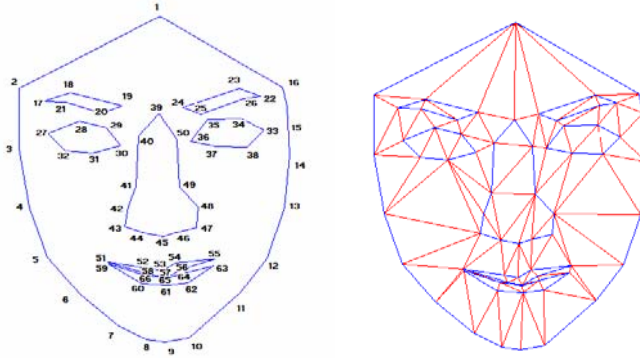


Fig. 3. (a) Control points for the 2D face model, (b) triangular facial mesh

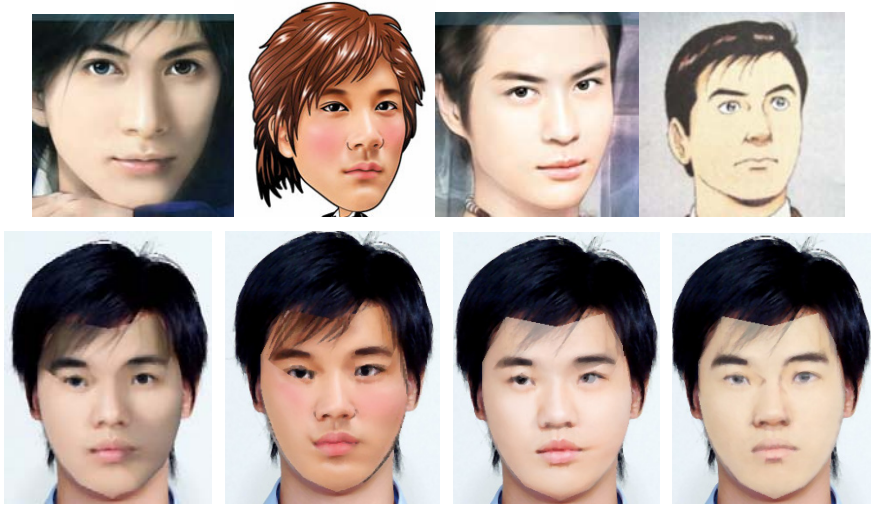


Fig. 4. Caricatures generated from different reference drawings

The results depicted in Fig. 4 (bottom row) are generated using different reference cartoons (top row). Generally speaking, each synthesized image bears certain degree of likeness to the original photo, yet at the same time exhibits the drawing style originated from the reference cartoon.

As the number of available templates increases, it becomes convenient for the system to select or recommend best references based on shape similarity. Since each component in our face model forms a closed contour, Fourier descriptor can be utilized to represent the individual parts. Let $(x_i(n), y_i(n))$ denote the n -th contour point of the i -th facial component, and let $u_i(n) = x_i(n) + jy_i(n)$, the facial shape descriptors (FSD) can then be defined according to:

$$a_i(k) = F\{u_i(n)\} = \sum_{n=0}^{N_i-1} u_i(n) \cdot e^{-j2\pi kn/N_i} \quad (1)$$

where N_i indicates the total number of control points in the i -th component.

Using Eq. (1), the difference between two facial shapes P (extracted from photo) and Q_s (template in the database) can be expressed according to:

$$D(P, Q_s) = \sum_{i=1}^M \text{Dist}(FSD_i^P, FSD_i^{Q_s}) \quad (2)$$

where

$$\text{Dist}(FSD_i^P, FSD_i^{Q_s}) = \left[\frac{1}{L_i} \sum_{k=0}^{L_i} \left(|a_i^P(k)| - |a_i^{Q_s}(k)| \right)^2 \right]^{0.5} \quad (3)$$

M denotes the number of facial components, and L_i denotes the number of Fourier coefficients retained to represent the i -th component.

Fig. 5 presents the top 5 choices for photos of different gender using FSD distance. An additional benefit of defining FSD arises from the blending process. It is known that blending of styles can be easily accomplished using polymorph technique. With FSD, the weight for each template can be set to be inversely proportional to the distance between the input shapes and the respective reference caricatures, as shown in Fig. 6.

The second method to generate caricature is roto-scoping. In roto-scoping, we combined contour detection, mesh generation, histogram specification and mean-shift algorithm to obtain a simplified region-based representation of the face with user-defined

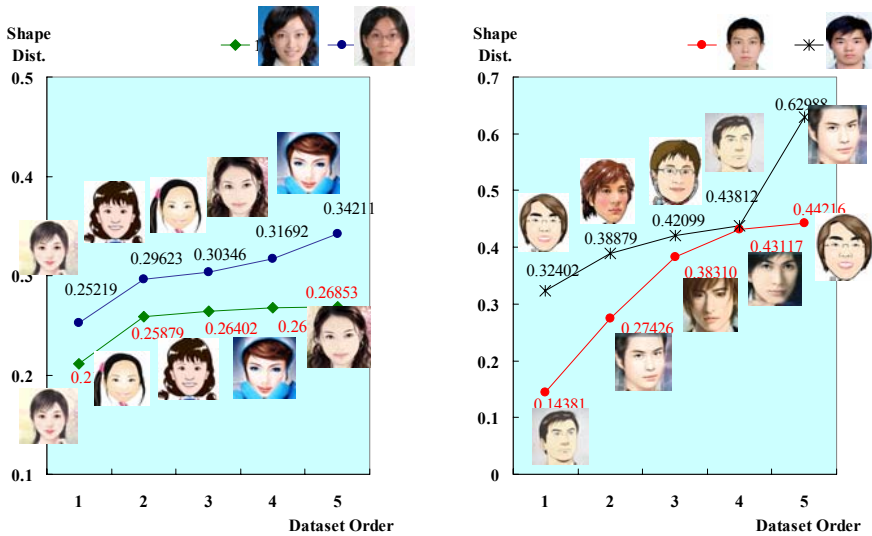


Fig. 5. Top template recommendations using shape similarity analysis

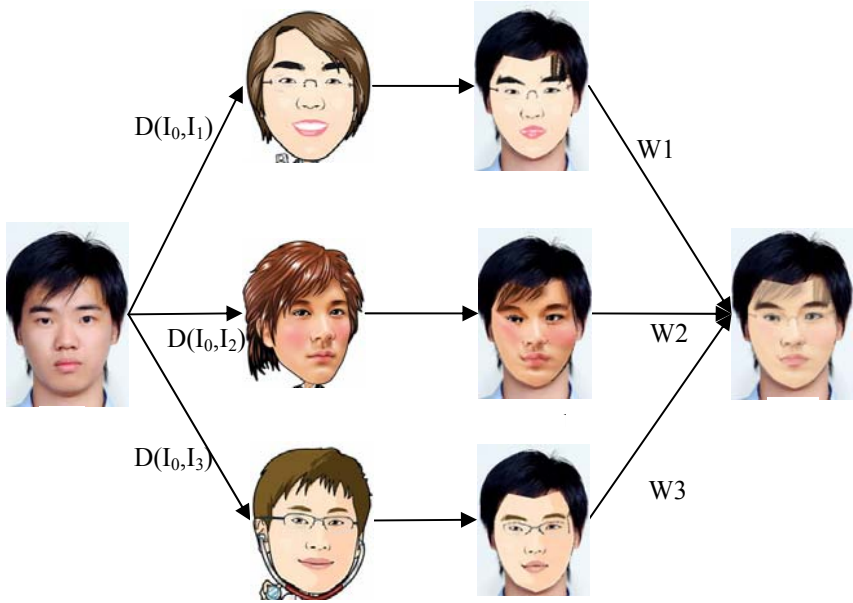


Fig. 6. Blending of styles using polymorph with FSD-controlled weights

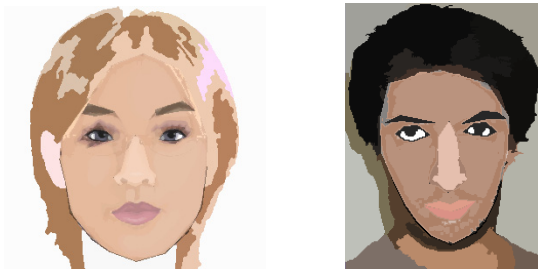


Fig. 7. Rotoscoped images

color palette. Details regarding the rotoscoping process can be found in [6]. Fig. 7 illustrates some rotoscoped images. Due to the nature of mean-shift segmentation, some triangular meshes are merged and contain the same color. These simplified regions are readily subject to raster-to-vector conversion to obtain vector-based representation such as Scalar Vector Graphics (SVG).

It should be noted that the color quantization stage will effectively reduce the number of colors used in the resulting picture, whereas the histogram specification stage will attempt to preserve color information from the original photo. These steps can also be incorporated into the image morphing approach to retain the color tone from the input photo.

3 Synthesis of Multiple Expressions



The FAP in MPEG-4 defines the set of parameters for 3D facial animation. Here we are mainly concerned with the synthesis on a 2D model. Consequently, we decompose the archetypal expression profile defined in [7] to arrive at a reduced set of facial animation parameters (FAP) for 2D caricatures, as shown in Table 1.

Using facial component action definition, the FAP can be set to manipulate the 66 facial definition points (FDP) in our model directly to synthesize different expressions. We have also defined 12 profiles according to popular emoticons. An example is illustrated in Table 2, where the emotion type (surprise) and its corresponding facial component action definition (FCAD), applied FDP and adjusted FAP are tabulated. Fig. 8 depicts the cartoons generated using the emoticon settings.

Table 1. Extracted 2D FAP Set

Group	Description	FAP No.	No. of Points	No. of Points Included in Reduced 2D Set
1	Visemes, Expressions	1~2	2	0
2	Jaw, chin, inner-lowerlip, corner-lips, midlip	3~18	16	11 (69%)
3	Eyeballs, pupils, eyelids	19~30	12	4 (33%)
4	Eyebrow	31~38	8	8 (100%)
5	Cheeks	39~42	4	0
6	Tongue	43~47	5	0
7	Head rotation	48~50	3	0
8	Outer-lip positions	51~60	10	8 (80%)
9	Nose	61~64	4	0
10	Ears	65~68	4	0
	Total		68	31 (45.6%)

Table 2. An example of emoticon profile setting

Emoticon	FCAD	Applied FDP	Related FAP
 Surprise	Brow: up Eye: wide Jaw: oh		$F_{31,33,35,32,34,36}(1), F_{37,38}(1.2)$ $F_{19,21}(1.2), F_{20,22}(1.2) F_3(0.6), F_{4,8,9}(0),$ $F_{5,10,11}(-0.6)F_{12,59}(-0.8),$ $F_{13,60}(-0.8)F_{6,53}(0.5),$ $F_{7,54}(0.5)F_{55,56,57,58}(1.3)$

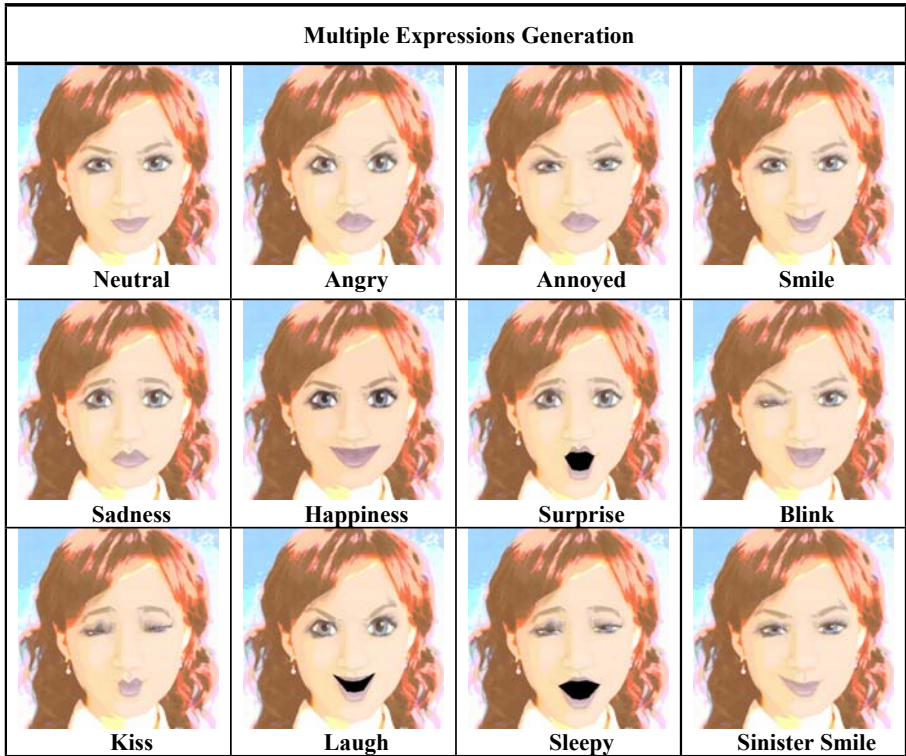


Fig. 8. Synthesized cartoons with multiple expressions

4 Conclusions

We have demonstrated an automatic caricature generator that is capable of synthesizing personalized cartoons with multiple expressions using only one reference image. New styles can be easily created by blending existing samples. The incorporation of FSD facilitates selection of templates in a cartoon database, as well as weight calculation in polymorph process.

Compact representation of the synthesized caricatures remains an issue for further investigation. Current model cannot strike a balance between fidelity and color quantization/region segmentation factors. Raster-to-vector conversion usually generates over-segmented patches. Novel types of vector representation such as diffusion curves may be applied to tackle this problem.

References

1. Liang, L., Chen, H., Xu, Y.Q., Shum, H.Y.: Example-Based Caricature Generation with Exaggeration. In: 10th Pacific Conference on Computer Graphics and Applications (PG 2002), pp. 386–393 (2002)

2. Mo, Z., Lewis, J.P., Neumann, U.: Improved Automatic Caricature by Feature. Normalization and Exaggeration. In: Proceedings of ACM SIGGRAPH Conference on Computer Graphics and Interactive Technique (2004)
3. Chiang, P.Y., Liao, W.H., Li, T.Y.: Automatic Caricature Generation by Analyzing Facial Features. In: Proc. of 2004 Asian Conference on Computer Vision (2004)
4. Chen, Y.L., Liao, W.H., Chiang, P.Y.: Generation of 3D Caricature by Fusing Caricature Images. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 1, pp. 866–871. IEEE Press, Los Alamitos (2006)
5. Lee, S., Wolberg, G., Shin, S.Y.: Polymorph: Morphing Among Multiple Images. IEEE Computer Graphics and Applications 18, 58–71 (1998)
6. Lai, C.-A.: Automatic Generation of Caricatures with Multiple Expressions Using. Transformative Approach. Master's Thesis, National Chengchi University (2008)
7. Perlin, K.: Layered Compositing of Facial Expression. In: ACM SIGGRAPH Technical Sketch (1997)

The Autonomous Duck: Exploring the Possibilities of a Markov Chain Model in Animation

Javier Villegas

Media Arts & Technology University of California Santa Barbara
Vivonets Lab, Experimental Visualization Lab
jvillegas@umail.ucsb.edu

Abstract. This document reports the construction of a framework for the generation of animations based in a Markov chain model of the different poses of some drawn character. The model was implemented and is demonstrated with the animation of a virtual duck in a random walk. Some potential uses of this model in interpolation and generation of in between frames are also explored.

1 Introduction

The process of generation of all the different drawings that make up an animation is a time consuming task. Some seconds of animation can demand a lot of drawings from a team of skilled animators, however, the use of computer had affected this practice significantly.

In the early days of animation the famous and experienced animators could not handle the amount of drawings that they have to paint and they develop different techniques to optimize the process. They concentrate their work in the most important frames (the ones that define the behavior of the characters) and they mark the moments of time where in between frames must be drawn. Then, they hire good drawers (even if they had no skills with the mechanics of motion) to fill the in-between frames [1].

Everything changed when computers appeared, but many concepts are still valid. The animators now use computers to design what is call the keyframes and the animation software is used to calculate the in-between frames [2]. In simple objects, animation attributes like color, size, shape or position can be easily parametrized in few numbers and a linear or spline interpolation can be used to generate the intermediate frames [3]. In more complex and articulated personages the interpolation of parameters describing the position of points in the figure must be followed by techniques like inverse kinematics to ensure that the in-between frame is a valid pose of the character [4].

If all the frames are generated automatically, the character will be autonomous. A lot of work has been done in the design of autonomous virtual characters that can behave based in a set of rules [5]. In this document, the movement of a character is represented with a Markov chain and the possibilities of this representation for autonomous characters and interpolation of frames are explored. The next section presents the preliminary concepts of Markov processes and Markov chains; the third part explains the model that is going to be used with a duck character. The results of the model for autonomous behavior will be commented in part 4. The possibilities of this model in the generation of

smooth and natural in-between frames will be presented in part 5, and finally, some results and possibilities for future development will be discussed.

2 Technical Background

2.1 Markov Random Processes

A discrete value, discrete time Markov random process can be defined as:

$$P(x_{n+1} = j | x_0 = i_0, x_1 = i_1, \dots, x_n = i_n) = P(x_{n+1} = j | x_n = i_n) \tag{1}$$

The value of the random variable x in some time n determines the conditional probabilities for the future values of the variable, and there is not dependence with the previous values. In short words, in a first order Markov process the future is independent of the pass if the present is known.

As this process is discrete value, it is a common practice to use the word “state” to reference any of the countable values of the random variable and represent the process in diagrams of nodes and arrows where each node represents a state and each arrow illustrate what is the conditional probability of going from the source state to the destiny state.

Fig. 1 shows the Markov chain representation of the very well know process call “the random walk”. In this process a particle chooses randomly the direction of movement for the next time, so if the current position and the transition probabilities are known, the probability of being in any state (position) in the next time can be calculated.

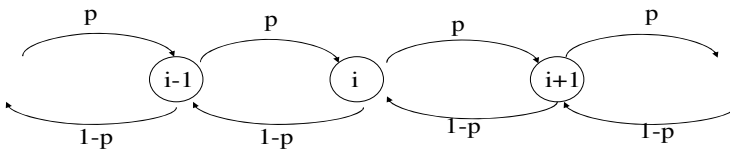


Fig. 1. Markov chain representing the random walk process, each node represents one state and each labeled arrow shows the value of the probability of going from one state to another

3 Description

3.1 General Scheme

The proposed framework is illustrated in Fig. 2. A model of the movement of a character using a Markov chain is designed and a 2D or 3D figure for each state is created. If the system is in “random walk” (autonomous) mode, the Markov model is used to calculate the next state to be drawn based in the current one. If the system is in interpolation mode, the Markov chain is used to find the intermediate frames between two given states. The render stage uses the information of the state calculation block and the current camera position (if applicable) to generate a new frame in the animation.

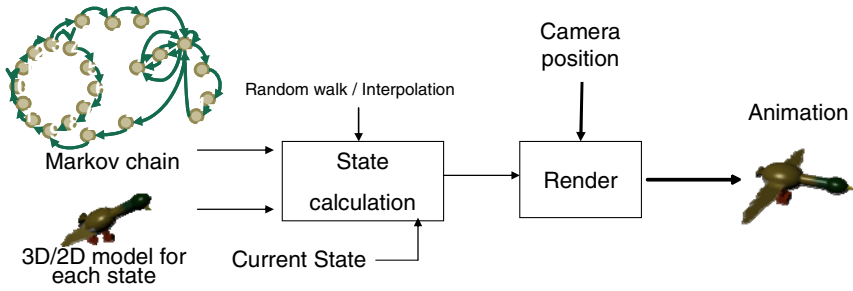


Fig. 2. General Scheme

3.2 Set of States

The framework was explored using the animation of a duck . A discrete time, discrete value random model was constructed using different poses of the duck (flying, walking, feeding) as the states of the Markov chain and defining a set of possible transitions to animate the movement of the duck. Fig 3 shows the full Markov chain.

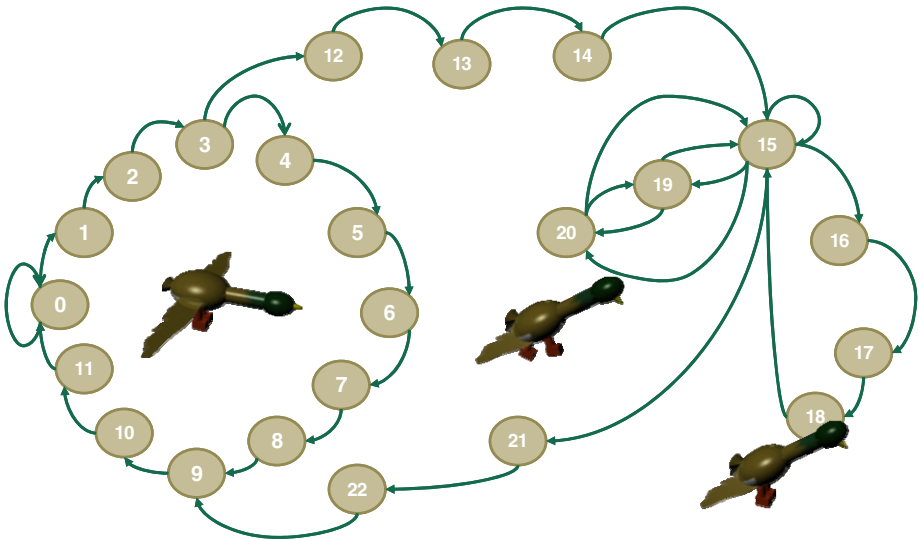


Fig. 3. The Markov chain

The set of states from 0 to 11 shows the flying cycle (Fig. 4). Duplicate positions appear (e.g., 2,4) to ensure that a first order model can represent the cycle adequately (i.e., sequence 0 - 1 - 0 is not valid).

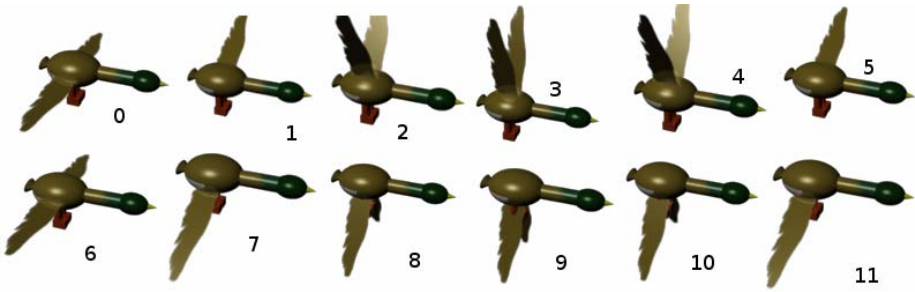


Fig. 4. States of the flying cycle

States 15 – 20 represent the land cycle, 15 to 18 model a feed cycle and 15,19 and 20 an oversimplified walk cycle. The landing and take off is represented in states 12,13,14 and 21, 22.

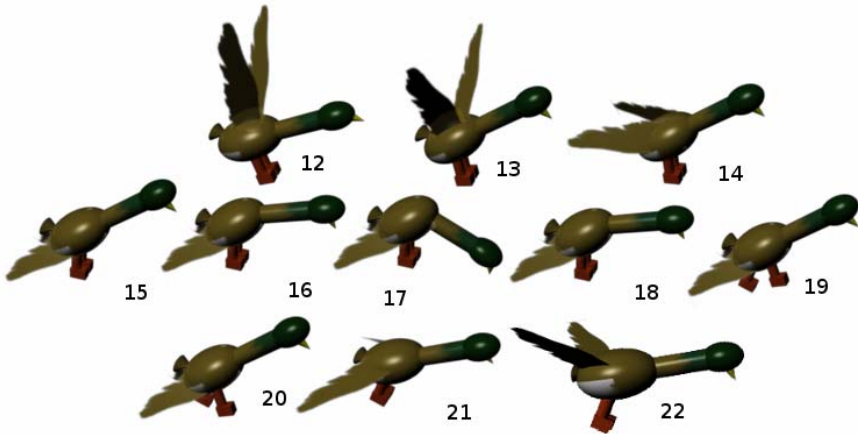


Fig. 5. States of terrestrial movements; landing and taking off

3.3 Continuous Parameters

The position of the bird is updated randomly in each time but with dependence on the previous position. This is accomplished modeling only the orientation as a continuous random variable and updating the position with constant jumps in that direction. The angle of azimuth for the next frame is calculated with a Gaussian random variable centered in the current direction, so this time the parameter can be modeled like a discrete time continuous value Markov process.

Assuming that the bird will tend to fly horizontally and only sporadically will change its angle of elevation to gain or loss altitude, the elevation angle is modeled as a triangular distribution centered in zero, so in this direction the orientation is independent of the orientation in previous states

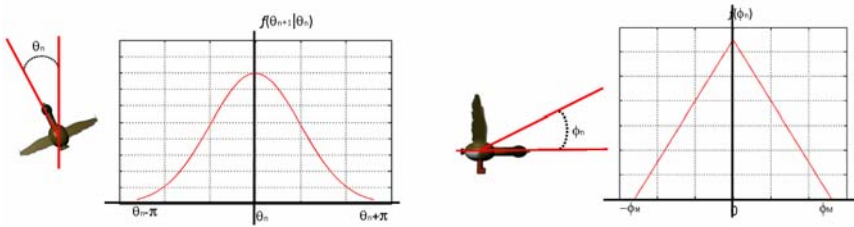


Fig. 6. The azimuth and elevation models

4 Autonomous Mode

The model was implemented using Maya Embedded language and some videos of the duck were generated using different parameters. The videos can be downloaded from (<http://www.mat.ucsb.edu/~jvillegas/ARTSIT2009/ArtsITVVideos.html>). Changes in the variance of the Gaussian and in the transition probabilities were introduced and different behaviors were observed. With bigger variance the duck tend to move more erratically and the transition parameters also conditioned the duck to pass more time flying or in the floor

5 Interpolation

The possibility to generate smooth and meaningful images between keyframes is also promissory and the general idea is going to be explained next.

5.1 Interpolation of States

The trellis diagram of the state transition can be traced and used as way to visualize all the possible routes between two states in different times. Fig. 7 shows all the possible transitions from time n and time $n+4$ for going from state 15 and returning to that state. The more probable path can be found with the adaptation of a computational efficient technique know as the Viterbi algorithm [9]. The basic concept behind the Viterbi algorithm is to store only one survival path for each state in each time iteration. This is consistent with the Markov assumption since once in a particular state the behavior is independent of the previous states so it makes sense to store only the best path (the most probable) that arrives to that state. As illustration, for a particular set of transition probabilities, the most probable path that arrives to each node is shown with a continuous line in Fig. 7. In this example the sequence of intermediate states would be: 15- 16-17-18-15.

5.2 Interpolation of Continuous Parameters

The probabilistic model allow us to find a set of intermediate values of orientation angles, if the initial and last position are known. Defining the joint distribution of the angle in time k and knowing that it was in angle θ_0 at time 0 we get:

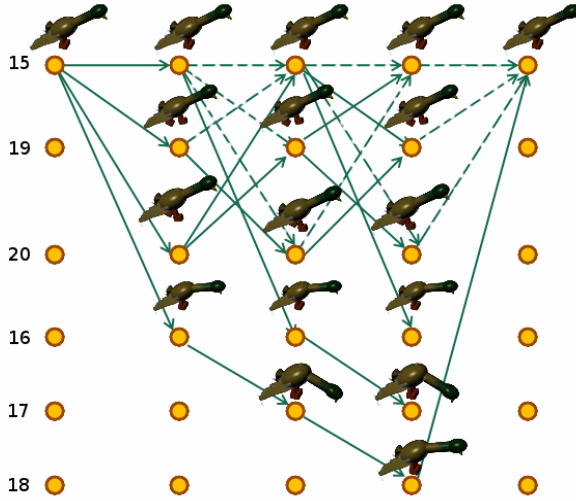


Fig. 7. Trellis representation of all the transitions from state 15 in 5 different consecutive times. Continuous lines represent the most probable path that arrives to each node.

$$f(\theta_k | \theta_0) = \prod_{n=0}^{k-1} f(\theta_{n+1} | \theta_n) \tag{2}$$

As it is mentioned before, the azimuth angle is modeled as a Gaussian random variable centered in the previous angle.

$$f(\theta_k | \theta_0) = \prod_{n=0}^{k-1} (2\pi\sigma^2)^{k/2} e^{-\left(\frac{\theta_{n+1}-\theta_n}{2\sigma^2}\right)^2} \tag{3}$$

To find out what is the choice for the intermediate angles that maximizes the probability of having θ_0 and θ_k in the extremes, we use a maximum likelihood estimation approach [6]. Finding the set of angles that maximize the expression in (3) is equivalent to find the set of angles that minimizes:

$$\sum_{n=1}^{k-1} \theta_{n+1}^2 - 2\sum_{n=0}^{k-1} \theta_{n+1}\theta_n + \sum_{n=1}^{k-1} \theta_n^2 \tag{4}$$

Taking derivatives and equating the expression to zero, we get:

$$\theta_i = \frac{\theta_{i-1} + \theta_{i+1}}{2} \quad i = 1, 2, k-1 \tag{5}$$

Equation (5) shows that if the conditional distributions are Gaussian (as in the case of azimuth angle), the best possible choice of intermediate states must satisfy that every angle is in the middle point of the adjacent angles, this is trivial accomplished using linear interpolation between the initial and final orientation.

Now for the elevation angle since:

$$f(\theta_k | \theta_0) = \prod_{n=0}^{k-1} f(\theta_{n+1} | \theta_n) = \prod_{n=0}^{k-1} f(\theta_{n+1}) \quad (6)$$

And since the extremes are fixed we want to find the set of choices for θ_1 to θ_{k-1} that make the expression (6) maximum. Since we have no control over the extremes and all the angles are chosen independently, the better choice will be to select the maximum possible value for each intermediate step. In the case of elevation angle with the distribution shown in Fig. 6, this will be to select zero (horizontal fly) for each intermediate angle.

6 Results and Future Development

The use of a Markov chain to represent the different poses of an animated character was explored. Also a discrete time continuous value Markov process was used to model the displacement of the same character. Although using this model, the interpolation has some limitations since the number of poses is always finite, the results can be improved if the Markov chain model is rich enough.

One of the main advantages of this alternative over the standard interpolation techniques is its versatility. Simple, articulated, elastic or even hand made draws can be used to define the poses of the character with no need for additional parameterizations. The Viterbi search for the intermediate state can be slightly modified using positive power of the transition probabilities to generate different interpolation paths. It can also be combined with traditional interpolation techniques of the continuous parameters instead of using the MLE approach for intermediate angles.

The autonomous mode can produce different outputs depending on the transition probabilities used, but also having a good model designed for a particular character will allow the exploration of interesting options. External signals like audio and video can be used to drive the personage across its possible states. If music is used it can be thought as a virtual dancer, and the dance routine can be always different. If video is used virtual puppetry applications can be designed.

References

- [1] Richard, W.: *The Animator's Survival Kit*. Faber & Faber (2002)
- [2] Dan, L.: *Maya Manual*. Springer, Heidelberg (2003)
- [3] Rick, P.: *Computer Animation: Algorithms and Techniques*. Morgan-Kaufmann, San Francisco (2001)
- [4] Allan, W., Mark, W.: *Advanced animation and rendering techniques*. ACM Press, New York (1991)
- [5] Reynolds Craig, *Boids*, <http://www.red3d.com/cwr/boids>
- [6] Henry, S., John, W.: *Probability and Random Processes with Applications to Signal Processing*. Prentice Hall, Englewood Cliffs (2002)
- [7] Steves, R.: *Character Animation*. Focal Press (2007)
- [8] Chris, W.: *Animation, The Mechanics of Motion*. Focal Press (2005)
- [9] Lou, H.: *Implementing the Viterbi Algorithm*. *Signal Processing Magazine* (September 1995)

MixPlore: A Cocktail-Based Media Performance Using Tangible User Interfaces

Zune Lee¹, Sungkyun Chang², and Chang Young Lim¹

¹ Graduate School of Culture Technology,
Korea Advanced Institute of Science and Tehcnology, Korea

² Department of Music, Seoul National University, Korea
zuuune@gmail.com, raynor1@hanmail.net, cylim@kaist.ac.kr

Abstract. This paper presents MixPlore, a framework for a cocktail-based live media performance. It aims to maximize the pleasure of mixology, presenting the cocktail as a plentiful art medium where people can fully enjoy new synesthetic contents by the integration of bartending and musical creation. For this, we fabricated Tangible User Interfaces (TUIs): tin, glass, muddler, and costume display, etc. The basic idea of performance and music composition is to follow the process of making cocktails. At the end of every repertoire, the performer provides the resultant 'sonic cocktails' to audience.

Keywords: Cocktail, taste, computer music, tangible user interface, live media performance, recipe.

1 Introduction

Food and the sense of taste are very much intertwined and they are important elements in human lives. Essentially, everyday we eat food to survive, every food has its own taste, and taste gives us great pleasure when eating food. Also taste and food influence social interactions and communication in our lives; food brings people together in its preparation and in its consumption [1]. Therefore, taste and food play pivotal roles in human lives and cultures.

In the area of art, the word, 'taste' has been traditionally mentioned to appreciate art pieces and express artistic styles, but we never 'literally' tasted them with our mouth and stomach since artworks are not edible. In modern art, there have been several artists who introduced food (but not taste) to art territories. They considered food as a performance medium dissociated from eating, taste, and nourishment [1]. In *Futurist Cookbook* (1932), F.T. Marinetti, the Italian Futurist, asserted the importance of food and cooking as Futurist aesthetics and philosophy, and his *Cucina Futurista* (Futurist Cuisine) was the first systematic approach to performance-oriented aesthetics of food [2]. Since then, artists such as Ann Hamilton, Alice Rio, Jana Sterbak, et al. were inherited from Marinetti's aesthetics in their happenings.

On the other hand, the field of Human-Computer Interaction (HCI) has studies related to food and drink in some different ways. iBar shows how a table-top interface can enhance novel interactions between people in the context of bar [3].

Lover's Cups presents an example of how drinking interfaces improve social interactions and personal communication channels [4]. There have been other research branches that are about interactive cookbook systems and interactive kitchens in ubiquitous computing environments. Studies [5, 6 and 7] in interactive cookbooks present novel cooking interfaces enabling users to easily search food recipes and prepare for food. Augmented Reality with computer vision, display, and sensor systems has been applied to interactive kitchen [8]. However, these works do not focus on media art and performances but on the introduction of computing to cooking areas. HCI has been rarely combined with food and taste in order to provide new user experiences to people. There seems to be only one example for using food and technology in media performance; Sound Kitchen presents how chemical reaction can make a musical performance [9]; it is not directly related to 'taste'; it is hard to be drinkable.

From the observation, though sound was not a central part in Futurist Cookbook, Marinetti inspires us to introduce cooking procedures to music and media art. Kirshenblatt-Gimblett proposed that food and performance converge conceptually in the three aspects [1]: First, to perform is to do, which counterpoints to make and serve food. Second, to perform is to behave; it means social practices in food. Third, when doing and behaving are shown, taste becomes a sensory experience and taste as aesthetic faculty converge [1]. HCI projects mentioned above also show potential that HCI techniques contribute to creating new user experiences in media art. Recently, Grimes and Harper proposed new visions of human-food-interaction in HCI and mentioned that making food is a way of expressing creativity [10]. They point out smelling, preparing, touching, and tasting foods, and even remembering past food experiences can evoke emotional responses. Finally they asserted potential in exploring systems that couple music with food in a variety of ways [10].

Previous artworks [Futurist Cookbook, Sound Kitchen] and studies [1, 10] motivated us to strongly introduce the sense of taste and the context of cooking to media performance. Here, we propose that cocktail a new medium for a performing art. First, cocktail, in itself, is a synesthetic medium; it has recipes, visuals, sounds and taste. Second, bartending, in itself, is performative; it has gestures and shows. Third, we think that the procedure of making cocktail is analogous to that of making sound. Fig.1 shows relationships between cocktail and music. Fourth, cocktail has appropriate attributes for us to formulate it as a media performance: simple recipes, range of ingredients, etc. compared with other food. Fifth, it is easier for us to design tangible interfaces from general cocktailware.

Therefore, our goals are as follows: For achieving high level artistic pleasure and plentiful expressions, we try to make further integration of human's senses such as taste, audio, visual in MixPlore. We explore new aesthetic territories by linking the factor of taste to musical creation; the taste becomes a primary element when interacting with sound elements. We attempt to design appropriate tangible user interfaces for edible performance. This paper will show how cocktail, taste, sound, and technology can be coupled to produce a new synesthetic media performance.

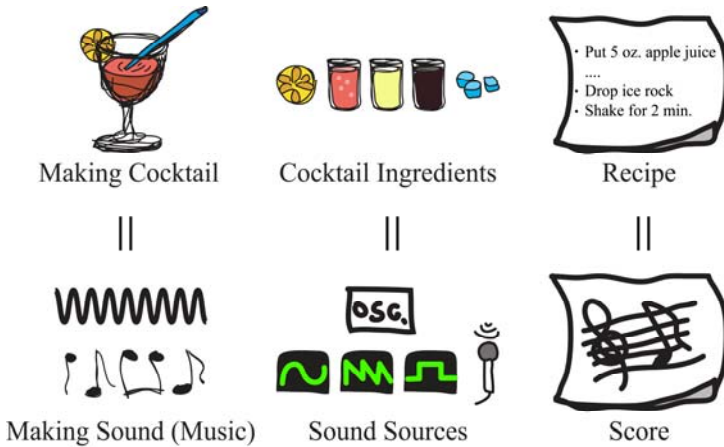


Fig. 1. The analogies between cocktail and music

2 Scenario and Approach

The primary design strategy with this framework has been to build an interactive system for simultaneously making cocktail and sound, and to propose suitable performance scenarios and contents. Early in our design stage, three video interviews were conducted with full-time bartenders to transfer more ideas directly from the field of bartending to our research. In the proceeding of interview, we observed and recorded their behavior and found the procedure of cocktail generally have preparing, mixing, serving, drinking cocktails and other actions. These actions were used to extract atomic gestures for our performance design. Additionally, we observed general cocktail instruments and recipes from the interviews and articles [11, 13]. Atomic gestures and interface design will be described in 3.

2.1 Recipe Observation

Most of all, we started our research with observing existing cocktail recipes. By analyzing them, we realized that recipes are deeply related to their own mixing methods and tools. If we could find a few formalized procedures of making cocktail from recipes, it would be ideal to design the entire performance. Considering a great number of existing recipes, we came to briefly categorize them into two types: The shaking-based type (scenario #1) and the layering-based type (scenario #2). For example, in the case of Mojito, the bartender muddles mint leaves, puts them and other ingredients in a shaker, and shakes (stirs) it [11]. In the case of B-52, s/he carefully pours liqueurs, which have different densities, with a stirrer to make several visible layers [13]. Our observation allows us to have criteria for necessary actions and interfaces in our performance design.

2.2 Possible Scenarios

Thus, we have two possible scenarios in this research as seen in Fig. 2. In the preparing step, the performer is ready for the performance with setting the table. Generally, the act of preparation is a part of performance but sometimes it can be done in advance of the performance, which means the step would be excluded in the real performing. The step of mixing, the core part of performance, has two kinds of act: shaking and layering. With the act of shaking, the bartender mixes up cocktail ingredients and sound sources by using the shaker/tin interface. Differently from shaking, the act of layering enables her/him to make multiple layers of cocktails and sounds (tones) by using the glass-laying interface. Then, the resultant sonic cocktail might be served to audience in the step of serving. Finally, audience can drink it.

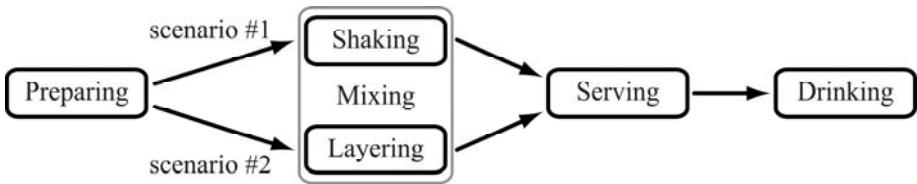


Fig. 2. Two scenarios based on the procedure of making cocktail

Therefore, this whole procedure becomes the process of performance. All performance repertoires of Mixlore followed those two scenarios. We will describe them in 4.

2.3 Design Approach

Interface Design approach. Based on our observation of recipes and scenarios, the motive of interfaces came from the ordinary cocktailware, tableware, or glassware to provide good affordances for usability to performers (as users) and audience. Interfaces should have proper controllability, visibility, mobility, feedback, and waterproofing for guaranteeing the performance's spectacle and stability. Also, interfaces should display audio-visual effects to audience in the performance. Besides, they should also make it possible to easily be installed and uninstalled in the stage at every repertoire.

Performance Design (Recipes) and Sound Design. Since they can dominate the entire performance design, cocktail recipes and mappings have the highest priority. We have three strategies for performance design and mapping between elements of performance: sound, taste, and interactions. The first strategy is to use original cocktail recipes and then devise the appropriate sonic recipes. This can guarantee us tastes and visuals of cocktail to some extents. The second one is to first consider the music recipes (or music composition) and then adjust the cocktail recipes to the sound factors. This method lets sound designers relatively free from the conflict between the two recipes. The last one is to intuitively consider two recipes at the same time. In the early design stage, we chose the only first strategy, but we currently follow the second and third one.

3 MixPlore System

Through our observation on recipes, MixPlore system has a set of tangible user interfaces: Shaker/tin, cup, muddler, mat, costume display, glass-layering interface and auxiliary interfaces. These interfaces can capture the bartender’s simple gestures and brain part facilities her/his mixing of cocktail and sound. All components are connected via a wireless/wired network as shown in Fig.3.

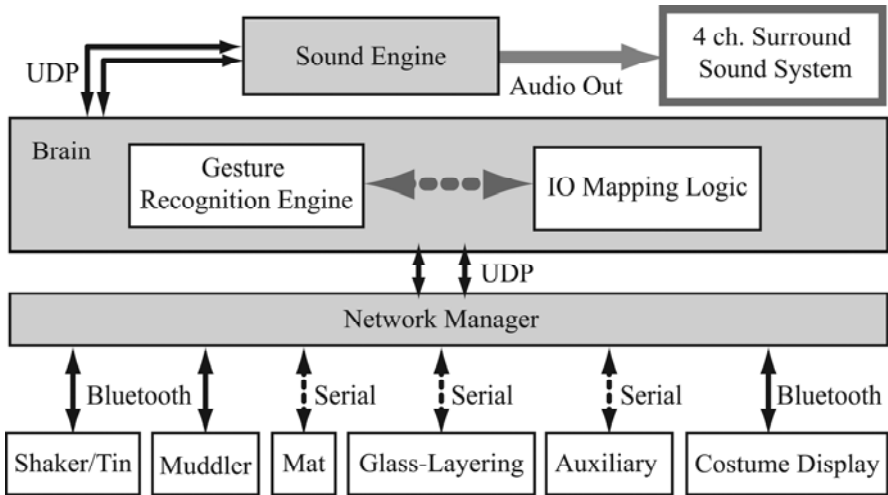


Fig. 3. Overview of MixPlore system

3.1 Interfaces and System Design

Tin and Cup. These are for mixing sound and cocktail ingredients. These capture gesture values of the performer, transmit the values to the network manager to generate sound and make a visual feedback. As shown in Fig.4a and 4b, the interfaces consist of sensing, control, communication, visual feedback (LEDs), and battery (3.3V). A three-axis accelerometer is to capture the motion data of the performer and measure the slope of the tin and cup. A switch is a trigger. A micro-controller (Atmega168) is the core for sensing and interfacing. For the wireless communication, we use Bluetooth (SENA ESD100).

The tin mainly functions as a shaker to mix sonic and cocktail ingredients in the step of mixing. In the step of preparing, cups are used for containing each ingredient and just modulating sound before mixing. Some special tins and cups incorporate a microphone for recording samples to use them as live sound sources.

Muddler. This interface, which also has the almost same architecture of the tin, functions as a stirrer that mixes and mashes the ingredients (e.g. mint leaves or grapes) and sounds (Fig.4c). The LED display on the upper part gives us several visual feedbacks to enhance performance spectacle (e.g. shifting and blinking).



Fig. 4. The tin (a), cup (b), and muddler (c). Tin (95x 185mm), cup (85 x 130mm), and muddler (35 x 330mm).

Mat. The mat with switches is a trigger for starting and ending sound. We put cups and tins on the mats. The mat with FSRs (Force-Sensing Resistor) measures the variation of weight. We use it for detecting resultant cocktails. The simple LED display is embedded in it.

Glass-layering. It is specially designed for the novel integration of visual, sound, and taste by simultaneously layering liquors with different alcohol proofs (or drinks with different specific gravities) and sound. The glass has a liquid level detector and two dials for sound modulation. We designed a capacitive sensor for detecting the liquid level. By adjusting the liquid level and layering other liquors, the performer can express an intriguing change of sound as well as imaging a merged taste.

The costume display. The display, as a wearable media, receives gesture information from tin and muddler via Bluetooth and displays visual patterns on the specially designed a 16 x 16 LED grid in the costume. The display responds to the bartender's gestures. It shows different patterns depending on gesture patterns of tin and muddler.



Fig. 5. The glass-layering interface (a) (50 x 240mm), costume display with graphic patterns (b, c), (260 x 180mm)

Network Manager. We used a 1:N Bluetooth network based on the polling communication model. For this we used SENA Parani MSP100, a Bluetooth Access Point directly connected to a brain PC and SENA ESD100 Bluetooth modules of each agents: the shakers, tins, muddlers and costume displays.

Brain. The brain functions as a central control logic consisting of a gesture recognition engine and IO mapper. The brain is built by Max/MSP [14]. The brain works in following ways. After initialization, it first observes the status of mats: The status

indicates if the mat is triggered or not. It is important since the information represents which cup/tin is held up or put down. According to it, 1) The IO mapper determines whether to go to the next state or stay at the current state. The procedure list is similar to Fig.2, but pre-defined for each repertoire. 2) The IO mapper properly routes data of each interface to targets such as the sound engine or costume display. In this process, the mapper needs a specified mapping table between sound properties and cocktail ingredients for each repertoire.

While using the muddler or tin, the recognition engine receives acceleration parameters to discriminate the performer's gestures. We used Gesture Follower (GF) [12], a HMM-based gesture recognition external for Max/MSP. In MixPlore, we defined several atomic gestures such as up-down, front-back, rotating, left-right and right-left. It is observed that it takes 1000 ~ 1500 ms for GF to receive the reliable recognition result and return it to the mapper. Such amount of time could become significant delay in the real-time performance.

4 Sound Engine and Performance Design

The performance idea is to have a performer making a cocktail in the process that s/he controls or triggers various musical sounds. As mentioned in 2.3, early in our design stage, we had to first decide the cocktail recipe and then determine the corresponding music recipe to the cocktail: It is because there are many possibilities to determine sound design and mappings, which can make us confused in our design processes. The two methods of making sound, considered in the early design stage, were roughly a step sequencer and the combination of sound synthesis. Currently, we have basically 2 types of sound engine based on scenario #1 and #2. The sound engines were written as a patch in Max/MSP [14]. The main advantage of this way was that we could fix and modify the engine instantly with practices of applying many recipes.

4.1 Sound Engines Based on 2 Scenarios

Sound Engine 1 (SE1). Based on Scenario #1 (see 2.2, shaking), our interfaces were mats, cups, tin and muddler. Sound Engine 1 was designed to synthesize sound by shaking, the main concept of the scenario. As we illustrated in system overview (Fig.3) and brain part (3.1), SE1 receives data of each interface and gesture type. Fig. 6 simply shows the schema of SE1.

This patch works as follows. In the case of cups mapped to sinusoidal sounds, we can modulate the sounds through shaking the cup. The type of modulation is selected by the gesture type received (e.g. up/down, rotation, etc.). In another case of cups mapped to a sampled sound such as human voice, the gesture type routes it into granular synthesis or convolution filter. Therefore, authors conclude SE1 suits for scenario #1. It is because the type of shaking gesture directly selects the method of sound synthesis as we intended. Finally, we wrote some presets for SE1 in our music composition stage. Additionally, we used some VST (Virtual Studio Technology) plug-ins from Pluggo [15] for sound effects.

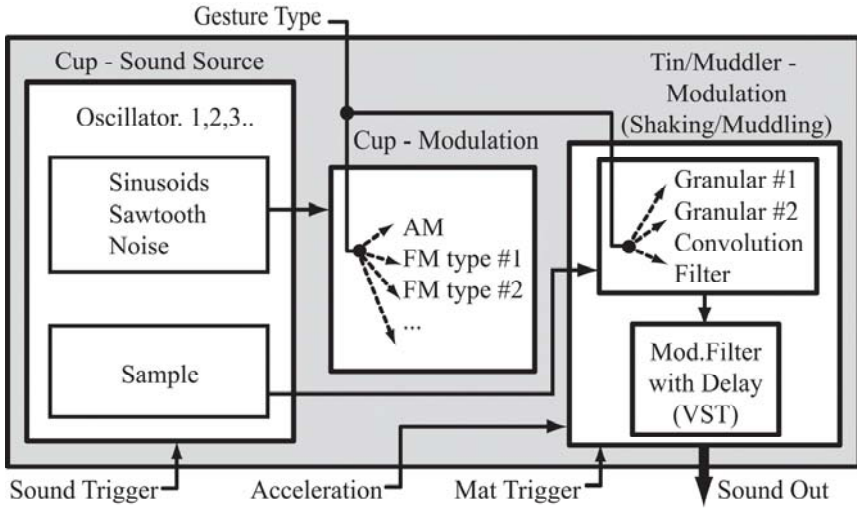


Fig. 6. The sound schema of Sound Engine1, based on Scenario #1

Sound Engine 2 (SE2). Based on Scenario #2 (see 2.2, layering), interfaces to be used are mats, layering-glass, and cups. SE2 was designed to synthesize sound by layering. In Fig.7, the central block represents a note modulation with a step sequencer, which modulates actual note pitches. One step has one note. Firstly picking up a cup starts the sequencer to play a step repeatedly. Picking up a new cup then simply generates a new step. By in/decreasing the level of liquors in the glass, the performer tunes up/down the pitch of the selected note in real-time. To convert the level into the pitch of a note, the sequence rounds off to the nearest pitch based on a chromatic scale. To control the tempo of the sequencer and filter modulation, we added 2 auxiliary potentiometers to this engine. Finally, the performer can end up the performance with a finish shot cup. Then the result sound is processed by the convolution of the sequenced music and the finish sample sound.

4.2 Composition

Based on 2 scenarios and each coupled sound engine, 7 repertoires (recipes) has been composed and performed by authors. In our composition, a very experimental mind was needed. Basically it was crucial to make recipes meeting both good tastes and plentiful musical expressions. On the other hand, each recipe needed its own performance duration. Therefore, we experimentally composed recipes minimum 2 minutes up to maximum 18 minutes long for playing.

4.3 Performance Repertoires

Example #1. Here, we show an example of performance, based on scenario#1.

Ingredients: Muscat (suitable amount), Blue liquor (1/2 Oz), Orange extract (1/2 Oz), Gin (1/2 Oz), Ice rocks, Alice’s voice (taken from audience) (Duration: 15 minutes).

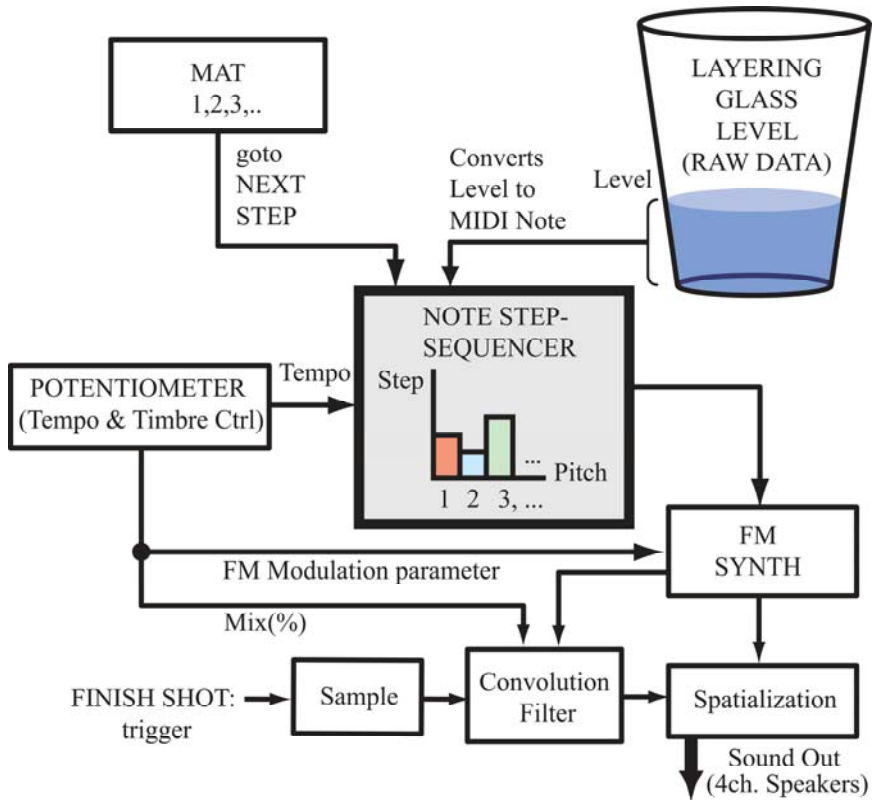


Fig. 7. The sound schema of Sound Engine2, based on Scenario #2



Fig. 8. Setting for an example of performance: cups, tins, mats and tins

- 1) Put each ingredient in the tin interface in a given order. Such ingredients are pre-mapped to the particular sound (sinusoids, sample sounds). Or, with the cup interfaces, the performer can record the voice of a selected audience during the performance. It could be one of good sonic ingredients.

- 2) Mix all the ingredients into a cocktail by bartending gestures. Sometimes, the performer can partially mix up some ingredients instead of mixing all ingredients simultaneously. Then put one of other ingredients and mix them. Iterate this process until the last ingredient is inserted, and then mix up all things. This procedure is necessary to create a long enough performance and music. Here, we suggest 10 minutes.
- 3) Those gestures are pre-mapped to particular sound properties to make effects on them; up-down: ring modulation; front-back: adding an echo effect to the particular sound; right-left: controlling panning of the stereo sound; shaker rotation: filtering the entire sound (dissolve filter)
- 4) Pour the mixed cocktail into a glass to generate a resultant sound.
- 5) Serve the cocktail with sound to an audience and let her/him drink the cocktail and the sound.

Example#2. This example shows that we make the integration of cocktail layering and note layering (arpeggio) with simple unique modulations like tempo and timbre. It is based on scenario#2.

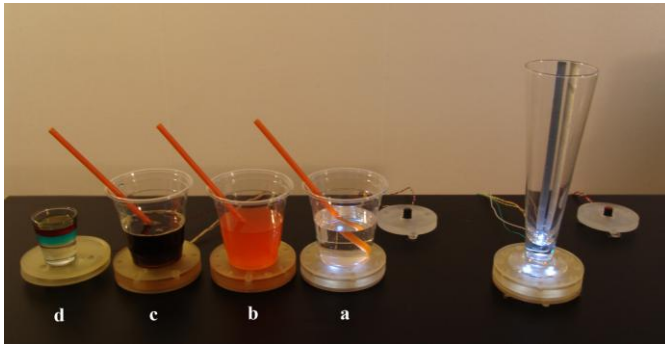


Fig. 9. Setting for example #2: cups, mats and layering-glass

- a. Light syrup (1st cup): Pick up the 1st cup and pour its content into the glass with a level detector (capacitive sensor). The glass tunes up the pitch according to the level of the ingredient.
- b. Mixed orange extract (2nd cup): Pick up the 2nd cup and pour its content into the glass. Then freely modulate the tempo and timber with auxiliary knobs.
- c. Black liquor (3rd cup): Pick up the 3rd cup and pour its content into the glass. Then improvise with controlling the knobs as you wish.
- d. A layered cocktail in a small glass (4th cup): A finish shot. Pick up the last small cup and drop it into the glass for layering to create the result sound.

5 Discussion and Future Works

The first performance of Mixplore was showcased in People, Art, and Technology 2007, an international network performance festival, Seoul, Korea, 2007 [16]. We

also had two workshops with experts and general audience after the festival. Before and after the festival, we had meetings with professional bartenders. These all gave us precious feedbacks with which we could discuss our interfaces and performance.

5.1 System and Interfaces Aspects

Tin and others. When using the tin, cups and muddler, they are coupled with the gesture-recognition engine to discriminate several gestures. For more exact recognition, it is very important that the engine knows when the performer starts and finishes gestures. Since it is not usable to return the time data by those interfaces, we let the performer use pedals to return each start and end time at every gesture.

Glass-layering. The capacitive sensor is used for the glass-layering interface to measure the height of liquid and reflect it for sound modulation. Since human fingers have capacitance and it can give noises to the capacitive sensor, the performer must carefully handle the interface. This fact decreases in the usability of the layering interface and we will solve this problem in future.

Latency. The entire latency of our system is about 20 ~ 40 ms, which is ignorable for real-time performances. However, it is observed that the costume display has the 400 ~ 800 ms, which is not suitable for a musical performance. It is caused by the delay of Bluetooth communication and that of processing LED patterns in the costume display.

Visibility. We devised the costume display for audience to concentrate on the performer and it was effective as a visual feedback. Yet, it was suggested that some visual devices be added to MixPlore for more visual spectacles. It would be visually improved if a table-top interface is together with this work.

5.2 Performance Aspects

Performance repertoires. When performing with new interfaces, audience cannot help being unfamiliar with it. This fact influenced our performance design. It was very difficult for us to decide how many portions of a performance we should assign to the interface demonstration as a performance since we thought performance and demonstration are in different areas. We still have questions on the issue, but in MixPlore, we divided our performance repertoires to two categories: the demonstration-based repertory and the art-based repertory.

Performing duration. The duration is important since it is directly related to cocktail tastes. Actually, it takes less than 2~3 minutes for a bartender to make a cocktail in the bar and it is not enough time to make a performance. Thus the main issue is how we could increase the performance duration up to 10~15 minutes without losing original tastes. The performance duration prolonged can influence cocktail tastes, so too much long shaking and muddling can make tastes worse. Such a thing occurred in our performance and we solved this problem by inserting other performing elements into the performance without changing cocktail-making duration.

Interaction with audience. Early in design stages, we considered audience' participating in our performance. We planned that the performer provided audience with

chances to shake the tin or muddler during the performance, or to drink the resultant sonic cocktail at the end. These situations were very welcome.

6 Conclusion

We have described MixPlore, a framework for designing a cocktail-based media performance with tangible interfaces. MixPlore has shown that cocktail is an attractive and plentiful medium for new performative expressions by introducing the sense of taste and synesthetic interactions into the circle of media performance. In the future, we plan to couple MixPlore with more visual expressions. Table-top interfaces could be one of the best solutions for enhancing synesthetic expressions and visual feedbacks. Another consideration is to install an LCD display and camera module into the tin and cup interface to mix sound, image and taste. It would make them fully capable interfaces for synesthetic performance.

References

1. Kirshenblatt-Gimblett, B.: Playing to the Senses: Food as a Performance Medium. *Performance Research* 4, 1–30 (1999)
2. Delville, M.: Contro la Pastasciutta: Marinetti's Futurist Lunch. *Intervals* 1, 25–37 (2007); Printemps
3. A System for the Interactive Design of any Bar-Counter, <http://www.i-bar.ch/en/info>
4. Chung, H., Lee, C.J., Selker, T.: Lover's Cups: Drinking Interfaces as New Communication Channels. In: *ACM SIGCHI 24th Conference on Human Factors in Computing systems*, pp. 375–380. ACM Press, New York (2006)
5. Ju, W., Hurwitz, R., Judd, T., Lee, B.: CounterActive: An Interactive Cookbook for the Kitchen Counter. In: *ACM SIGCHI 19th Conference on Human Factors in Computing systems*, pp. 269–270. ACM Press, New York (2001)
6. Chi, P., Chen, J., Chu, H., Chen, B.: Enabling Nutrition-aware Cooking in a Smart Kitchen. In: *ACM SIGCHI 25th Conference on Human Factors in Computing systems*, pp. 2333–2338. ACM Press, New York (2007)
7. Terrenghi, L., Hilliges, O., Butz, A.: Kitchen stories: sharing recipes with the Living Cookbook. *Personal and Ubiquitous Computing* 11(5), 409–414 (2007)
8. Bonanni, L., Lee, C., Selker, T.: Attention-based design of augmented reality interfaces. In: *CHI 2005 extended abstracts on Human factors in computing systems*, pp. 1228–1231. ACM Press, New York (2005)
9. Shiraiwa, H., Segnini, R., Woo, V.: Sound Kitchen: Designing a Chemically Controlled Musical Performance. In: *2003 New Interfaces For Musical Expression*, pp. 83–86. McGill University, Montreal (2003)
10. Grimes, A., Harper, R.: Celebratory technology: new directions for food research in HCI. In: *Proceeding of the 26th annual SIGCHI conference on Human factors in computing systems*, pp. 467–476. ACM Press, New York (2008)
11. Regan, G.: *The Joy of Mixology: The Consummate Guide to the Bartender's Craft*. Clarkson Potter, New York (2003)
12. http://ftm.ircam.fr/index.php/Gesture_Follower
13. http://en.wikipedia.org/wiki/B-52_cocktail_recipe
14. <http://www.cycling74.com>
15. <http://www.nabi.or.kr/party2007>

Author Index

- Ahmed, Salah Uddin 40
Ayiter, Elif 223
- Balcisoy, Selim 223
Barile, Perry 255
Bhowmick, Partha 181
Bißmann, Michael 105
- Chan, Li-Wen 112
Chang, Sungkyun 279
Chang, Yin-Wei 56
Chen, C.-F. 239
Chen, C.-Y. 239
Chen, Chi-Fa 231
Chen, Chia-Yen 165, 215, 231
Chen, Chih-Tung 206
Chen, Chih-Wei 157
Chen, Lieu-Hen 72
Chen, Ying-Chung 25
Chien, H.-J. 239
Christou, Maria 97
Chu, Sheng-Kai 231
Ciesielski, Vic 255
Clay, Alexis 148
Correa, Carlos D. 136
Couture, Nadine 148
- Delord, Elric 148
Domenger, Gaël 148
- Erzigkeit, Hellmut 33
- Fuh, Chiou-Shann 157
- Germen, Murat 223
Gottschalk, Stefan 105
Graziosi, Fabio 128
- Hamlyn, Jim 17
He, Ming-Yu 206
Her, Jiun-Jhy 17
Ho, Yi-Hsin 72
Hoenig, Florian 33
Hornegger, Joachim 33
Hou, Gee-Chin 247
- Hsiao, Chuan-Heng 112
Hsieh, Chun-Ko 112
Hsieh, Wen-Chieh 72
Hsu, Su-Chu 25
Hsu, Tao-i 206
Hsu, Wei-Hsien 136
Huang, Da-Yuan 157
Huang, Fay 56
Huang, Jinsheng 198
Huang, Sheng-Wen 231
Huang, Shiang-Yin 80
Huber, Linda 105
Hung, Yi-Ping 112
- Jaccheri, Letizia 40
Jiang, Ruyi 190, 198
- Keck, Alexander 105
Klette, Reinhard 9, 165, 190, 198, 215
Köppen, Mario 64
Kornhuber, Johannes 33
Kumar, Gautam 181
- Lai, Chien-An 263
Lai, Mei-Yi 80
Lee, Wen-Hui 206
Lee, Wen-Pin Hope 48
Lee, Zune 279
Li, Tsai-Yen 247
Liang, Feng-Ju 48
Liao, Wen-Hung 263
Lim, Chang Young 279
Lin, Che-Yu 80
Lin, Chih-Chung 247
Lin, Jiun-Shian 25
Lin, Ko-Chun 231
Lin, Quo-Ping 112
Liu, Che-Wei 80
Liu, I-Ling 112
Liu, Ting-Yu 72
Luciani, Annie 97
- M'kadmi, Samir 40
Ma, Kwan-Liu 136
Maier, Andreas 33

- Mei, Jianqiang 136
Mombaur, Katja 120

Ohnishi, Kei 64
Oshita, Masaki 88

Radke, Alexander 105
Raschke, Michael 120
Ren, Feixiang 198
Rinaldi, Claudia 128
Russell, James C. 165

Schenzel, Peter 173
Schönefeldt, Andreas 105
Schubert, Alexander 120
Seidenglanz, Marko 105
Sharma, Naveen Kumar 181
Soutschek, Stefan 33
Stoke, Norbert 105
Steidl, Stefan 33
Stork, David G. 1

Stuermer, Michael 33
Stussak, Christian 173
Su, Ming-Wei 231
Suzuki, Kenji 88

Tarquini, Francesco 128
Terauchi, Mutsuhiro 190, 198
Torii, Akihiko 9
Trist, Karen 255
Tsai, Chi-Hung 80

Vaudrey, Tobi 190, 215
Villegas, Javier 272

Wang, Hsi-Chun 48
Wang, Shigang 190
Wedel, Andreas 215

Yeh, Jeng-Sheng 80
Yoshida, Kaori 64