MPEG-4 Facial Animation

The Standard, Implementation and Applications

Edited by

Igor S. Pandzic and Robert Forchheimer

Linköping University, Sweden

JOHN WILEY & SONS, LTD

MPEG-4 Facial Animation

The Standard, Implementation and Applications

MPEG-4 Facial Animation

The Standard, Implementation and Applications

Edited by

Igor S. Pandzic and Robert Forchheimer

Linköping University, Sweden

JOHN WILEY & SONS, LTD

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 1UD, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Pappelallee 3, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-470-84465-5

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India
Printed and bound in Great Britain by Antony Rowe Limited, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Contents

Appendices

List of Contributors

Jörgen Ahlberg
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping
Sweden
ahlberg@isy.liu.se

Michele Cannella
TAU Tecnologia Automazione Uomo s.c.r.l.
Via XX Settembre 3/6
16121 Genova
Italy
michele.cannella@tetralab.it
michele.cannella@tau-online.it

Roddy Cowie
Queen's University of Belfast
Belfast, N. Ireland
r.cowie@qub.ac.uk

Ellen Douglas-Cowie
Queen's University of Belfast
Belfast, N. Ireland
e.douglas-cowie@qub.ac.uk

Franck Davoine
Université de Technologie de Compiegne
Laboratoire Heudiasyc, BP20529
60205 France
Franck.Davoine@hds.utc.fr

Robert Forchheimer
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping
Sweden
Robert@isy.liu.se

Pengyu Hong
1614 Beckman Institute
Urbana IL61801
hong@ifp.uiuc.edu

Thomas S. Huang
2039 Beckman Institute
Urbana IL61801
huang@ifp.uiuc.edu

Stefanos Kollias
Image, Video and Multimedia Systems Laboratory
National Technical University of Athens
Electrical & Computer Engineering Department
Computer Science Division
ECE Building – 1st Floor – Room 11.23
Athens, Greece
Stefanos@cs.ntua.gr

Fabio Lavagetto
Università degli Studi di Genova
Dipartimento di Informatica, Sistemistica e Telematica
Via all'Opera Pia 13
16145 Genova, Italy
fabio@dist.unige.it

Haibo Li
Digital Media Lab
Umeå University
SE-901 87 Umeå
Haibo.li@tfe.umu.se

Sotiris Malassiotis
ITI, Greece
malasiot@iti.gr

Andrew Marriott
Senior Lecturer
School of Computing
Curtin University of Technology
Hayman Road, Bentley
Western Australia, 6102
raytrace@smtp.cs.curtin.edu.au

Jörn Ostermann
AT&T Labs-Research
Rm A5–4F36
200 Laurel Ave South
Middletown, NJ 07748, USA
Joern.Ostermann@ieee.org

Igor S. Pandzic
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping
Sweden
igor@isy.liu.se

Department of Telecommunications
Faculty of Electrical Engineering and Computing
University of Zagreb
Unska 3
HR-10000 Zagreb
Croatia
Igor.Pandzic@fer.hr

Montse Pardas
Universitat Politècnica de Catalunya
Barcelona, Spain
montse@gps.tsc.upc.es

Catherine Pelachaud
Universita di Roma "La Sapienza"
Dipartimento di Informatica e Sistemistica
via Buonarroti, 12
00185 Roma, Italy
cath@dis.uniroma1.it

Dr. Eric Petajan
Chief Scientist and Founder
face2face animation, Inc.
2 Kent Place Blvd
Summit, NJ 07901
eric@f2f-inc.com

Roberto Pockaj
Eptamedia s.r.l.
c/o Dipartimento di Informatica, Sistemistica e Telematica
Via all'Opera Pia 13
16145 Genova
Italy
roberto.pockaj@eptamedia.com

Amaryllis Raouzaiou
Image, Video and Multimedia Systems Laboratory
National Technical University of Athens
Electrical & Computer Engineering Department
Computer Science Division
ECE Building – 1st Floor – Room 11.23
Athens, Greece
araouz@image.ece.ntua.gr

Nicolas Tsapatsoulis
Image, Video and Multimedia Systems Laboratory
National Technical University of Athens
Electrical & Computer Engineering Department
Computer Science Division
ECE Building – 1st Floor – Room 11.23
Athens, Greece
ntsap@image.ntua.gr

Gael Sannier
W Interactive SARL, France
gael@winteractive.fr

Zhen Wen
1614 Beckman Institute
Urbana IL61801
zhenwen@ifp.uiuc.edu

Liwen You
Linköping University Linköping
Sweden
liwenyou@ieee.org

Author Biographies

Jörgen Ahlberg was born in Karlstad (Sweden) in 1971. He got his M.Sc. in Computer Science and Engineering from Linköping University (Sweden) in 1996, and joined the Image Coding Group, also at Linköping University, as a Ph.D. student the same year. In the Image Coding Group and at Université de Technologie de Compiègne (France, 1999) he has since then undertaken research in different aspects of model-based coding, such as facial parameter compression, texture coding, detection and tracking of faces and facial features and evaluation of face models. He was also an active participant in the development of MPEG-4 Face Animation and is currently working in the European InterFace project. He has seen Andrew Marriott wear shoes. J. Ahlberg can be reached at *ahlberg@isy.liu.se*.

Michele Cannella was born in Genoa on January 14, 1968. He got the 'laurea' degree in Electrical Engineering at DIST, University of Genoa, in 1992. From September 1992 to March 1993 he had a post-degree scholarship from Marconi, working on applications based on Oracle DB. In the period from September 1994 to June 1995, he was a consultant with Modis spa on C applications with Unix-Motif user interface. From April 1996 to May 1999 he worked with Elmer spa, with specific responsibility for the European ACTS project Vidas, with reference to software porting on real-time platform. From January 2000 he has been working with TAU scrl for the European project Interface, with reference to design and implementation of 'real world' applications based on the innovative technologies developed by the partners of the project. In September 2001 he founded Tetralab srl (*www.tetralab.it*), a software development and consulting company based at Genoa. His interests are in the field of software design using object-oriented technologies (UML), the applications of markup languages (XML) and java-based web applications. Michele can be reached at *michele.cannella@tetralab.it*

Roddy Cowie works in the School of Psychology at Queen's University, Belfast. His core interest is in the relationship between the way humans form their impressions of the world and algorithms that it is natural to implement on computers. He has argued that superficially appealing models of perception and cognition tend to be oversimplified, and that systematic attention to the character of human experience can be a key to finding less obvious alternatives. He has used anomalous experiences and illusions to highlight the role that shape plays in human vision and to derive algorithms that are similarly shape-oriented. With Ellen Douglas-Cowie, he has highlighted the way speech conveys impressions of the speaker alongside the overt message, and has developed programs for extracting relevant attributes of the speech signal. He has studied areas

where aspects of experience that are profoundly subjective, and difficult to externalize, nevertheless have a major effect on everyday behavior, particularly acquired deafness and more recently religion. Several of these interests link to the area of emotion, and it has been his main focus for the past five years, developing techniques for measuring subjective impressions of emotion and the speech variables that convey them. He can be reached at *r.cowie@qub.ac.uk*

Ellen Douglas-Cowie works in the School of English at Queen's University, Belfast, and she is currently the head of the school. Her research is on the information that speech carries about the speaker, with emphasis on collecting and using natural data. Her Ph.D. research was a widely cited sociolinguistic study, which revealed the complexity of the factors that influence people's choice of speech style. With Roddy Cowie, she carried out seminal research on the way speech is affected when people lose their hearing and the impact of the changes on listeners. This project led them to develop automatic techniques for extracting features of prosody that affect the impression a speaker creates, and it has provided a basis for a broader attack on prosodic speech styles, ranging across clinical varieties, the 'phone voice', skilled and unskilled reading and so on. Recently, their main focus has been on the signs of emotion, vocal and visual, and Ellen has led the assembly of a substantial audio-visual database of emotional speech extracted from real interactions. In 2000, they organized a pioneering workshop on speech and emotion, and they are editing a special issue of Speech Communication arising from it. Ellen can be reached at *e.douglas-cowie@qub.ac.uk*

Franck Davoine received a Ph.D. in Signal, Image and Speech Processing from the Institut National Polytechnique de Grenoble, France, in 1995. He was a visiting researcher at the Division of Image Coding of the University of Linköping, Sweden, from 1996 to 1997 and worked on very low bit rate video representation and coding. He joined the laboratory HEUDIASYC of the University of Technology of Compiègne, France, in 1997 as an assistant professor, and he is currently a CNRS researcher of the same laboratory. His research interests include facial image analysis for human interaction to virtual and augmented environments, and digital image watermarking for content protection, authentication and control. Franck Davoine can be reached at *Franck.Davoine@hds.utc.fr*

Robert Forchheimer received the M.S. degree in electrical engineering from the Royal Institute of Technology, Stockholm (KTH) in 1972 and the Ph.D. degree from Linköping University in 1979. During the academic year 1979 to 1980, he was a visiting research scientist at University of Southern California where he worked in the areas of image coding, computer architectures for image processing and optical computing.

Dr Forchheimer's research areas have involved data security, packet radio communication, smart vision sensors and image coding. He has authored and coauthored papers in all of these areas and also holds several patents. He is the cofounder of several companies within the university science park. Dr Forchheimer is currently in charge of the Image Coding Group at Linköping University. His main work concerns algorithms and systems for image and video communication.

Pengyu Hong received the B. Engr. and M. Engr. degree, both in computer science, from Tsinghua University, Beijing, China, in 1995 and 1997, respectively. He received

his doctorate from the Department of Computer Science at the University of Illinois at Urbana-Champaign in December 2001. In 2000, he received the Ray Ozzie fellowship for his research work on face modeling, facial motion analysis and synthesis. He is now a postdoc in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign. He is conducting research in the areas of human–computer interaction, multimedia information processing, computer vision and pattern recognition, data mining and machine learning. His home page is *www.ifp.uiuc.edu/~hong*.

Thomas S. Huang received his B.S. Degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, China and his M.S. and Sc.D. Degrees in Electrical Engineering from the Massachusetts Institute of Technology, Cambridge, Massachusetts. He was on the faculty of the Department of Electrical Engineering at MIT from 1963 to 1973 and on the faculty of the School of Electrical Engineering and director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt, distinguished professor of Electrical and Computer Engineering, and research professor at the Coordinated Science Laboratory, and head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology and cochair of the Institute's major research theme Human Computer Intelligent Interaction.

Dr. Huang's professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 12 books and more than 400 papers in Network Theory, Digital Filtering, Image Processing and Computer Vision.

He is a member of the National Academy of Engineering; a foreign member of the Chinese Academy of Engineering and a fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of America, and has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior US Scientist Award, and a fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award in 1987 and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. Also in 2000, he received the Honda Lifetime Achievement Award for 'contributions to motion analysis'. In 2001, he received the IEEE Jack S. Kilby medal.

Stefanos Kollias was born in Athens in 1956. He obtained a Diploma in Electrical Engineering from the National Technical University of Athens (NTUA) in 1979, an M.Sc. in Communication Engineering from the University of Manchester Institute of Science and Technology in England in 1980 and a Ph.D. in Signal Processing from the Computer Science Division of NTUA in 1984. In 1974 he obtained an honorary diploma in the Annual Panhellenic Competition in Mathematics. In 1982 he was given a COMSOC Scholarship from the IEEE Communication Society.

Since 1986 he has served as lecturer, assistant and associate professor of the Department of Electrical and Computer Engineering of NTUA. From 1987 to 1988 he was a visiting research scientist in the Department of Electrical Engineering and the Center for Telecommunications Research of Columbia University in New York, USA, on leave from NTUA. Since 1997 he has been a Professor of NTUA and Director of the Image, Video and Multimedia Systems Lab.

His research interests include image and video processing, analysis, coding, storage, retrieval, multimedia systems, computer graphics and virtual reality, artificial intelligence, neural networks, human–computer interaction and medical imaging. Fifteen graduate students have completed their doctorate under his supervision; another ten are currently doing their Ph.D. He has published more than 140 papers, 60 of them in international journals. In the last few years, he and his team have been leading or participating in forty-five projects, both European and national.

Fabio Lavagetto was born in Genoa, Italy, on August 6, 1962. He received the Masters degree in electrical engineering from the University of Genoa, Genoa, Italy, in March 1987 and the Ph.D. degree from the Department of Communication, Computer and System Sciences (DIST), University of Genoa, in 1992. He was a visiting researcher with AT&T Bell Laboratories, Holmdel, NJ, during 1990 and a contract professor in digital signal processing at the University of Parma, Italy, in 1993. Presently, he is an associate professor with DIST, University of Genoa, where he teaches a course on radio communication systems and is responsible for many national and international research projects. From 1995 to 2000, he coordinated the European ACTS project VIDAS, concerned with the application of MPEG-4 technologies in multimedia telecommunication products. Since January 2000, he has been coordinating the IST European project INTERFACE, which is oriented to speech/image emotional analysis/synthesis. He is the author of more than 70 scientific papers in the area of multimedia data management and coding. He can be reached at *fabio@dist.unige.it*.

Haibo Li is a full professor in Signal Processing in the Department of Applied Physics and Electronics (TFE), Umeå University, Sweden. He received the Technical Doctor degree in Information Theory from Linköping University, Sweden, in 1993. His doctoral thesis dealt with advanced image analysis and synthesis techniques for low bit rate video. Dr Li got the 'Nordic Best Ph.D. Thesis Award' in 1994. In 1997, Dr. Li was awarded the title of 'Docent in Image Coding'. During his period at Linköping University, he developed advanced image and video compression algorithms, including human face image analysis, extremely low bit rate video compression and 3-D video transmission. After joining Umeå University, he is now directing the Digital Media Lab, Umeå Center for Interaction Technology (UCIT), Umeå University, and working on advanced Human, Thing and Information interaction techniques. Prof. Li has been chairing sections at relevant international conferences and has been actively involved in MPEG activities in low bit rate video compression. He has contributed to several EU projects, such as VIDAS, SCALAR, INTERFACE and MUCHI. He has published more than 90 technical papers including chapters in books and holds several patents. Haibo Li can be reached at *haibo.li@tfe.umu.se*.

Sotiris Malassiotis received the B.S. and Ph.D. degrees in Electrical Engineering from the Aristotle University of Thessaloniki in 1993 and 1998, respectively. From 1994 to 1997 he was conducting research in the Information Processing Laboratory of Aristotle University of Thessaloniki. He is currently a senior researcher in the Informatics and Telematics Institute, Thessaloniki. He has participated in several European and National research projects. He is the author of more than ten articles in refereed journals and more than twenty papers in international conferences. His research interests include image

analysis, image coding, virtual reality and computer graphics. Dr. Malassiotis may be reached by e-mail at *malasiot@iti.gr* (*http://www.iti.gr/people/malasiot/en/index.html*)

Andrew Marriott is a senior lecturer in the School of Computing at Curtin University of Technology, Perth, Western Australia. His research interests include facial animation, unnatural terrain environments and, of course, pretty pictures. In 1988 he formed the Computer Animation Negus (CAN), a research and development group at Curtin whose aim is to provide a sophisticated environment for animation work at the undergraduate, postgraduate and commercial level. Fax, a facial animation system, was first released into the public domain in 1992 and has been used by many researchers and has 'starred' in a few films. He is developing the Mentor System – a large-scale Java-based graphical mentoring system. He is the principal developer of VHML – the Virtual Human Markup Language (*www.vhml.org*). He is also a full partner in a 5th Framework European Union project called Interface. He has been known to wear shoes. You may find out more about him at *http://www.computing.edu.au/~raytrace* or you may email him via *raytrace@cs.curtin.edu.au*.

Jörn Ostermann studied Electrical Engineering and Communications Engineering at the University of Hannover and Imperial College London, respectively. He received Dr.-Ing. from the University of Hannover in 1994. From 1988 to 1994, he worked as a research assistant at the Institut für Theoretische Nachrichtentechnik, conducting research in low bit rate and object-based analysis–synthesis video coding. In 1994 and 1995 he worked on Visual Communications Research at AT&T Bell Labs. He has been a member AT&T Labs – Research since 1996. He is working on multimodal human–computer interfaces with talking avatars, streaming, video coding and standardization.

From 1993 to 1994, he chaired the European COST 211 sim group coordinating research in low bit rate video coding. Within MPEG-4, he chaired the Adhoc Group on Coding of Arbitrarily shaped Objects in MPEG-4 Video. Jörn was a scholar of the German National Foundation. In 1998, he received the AT&T Standards Recognition Award and the ISO Certificate of Appreciation. He is a senior member of IEEE, the IEEE Technical Committee on Multimedia Signal Processing, chair of the IEEE CAS Visual Signal Processing and Communications (VSPC) Technical Committee and a distinguished lecturer of the IEEE CAS Society. He has contributed to more than 50 papers, book chapters and patents. He is coauthor of the textbook Video Processing and Communications.

Igor S. Pandzic is currently a visiting scientist at the University of Linköping, Sweden, as well as a visiting professor at the University of Zagreb, Croatia, where he obtained an Assistant Professor position at the time when this book was being published. Formerly he worked as a senior assistant at MIRALab, University of Geneva, Switzerland, where he obtained his Ph.D. in 1998. The same year he worked as visiting scientist at AT&T Labs, USA. Igor received his B.Sc. degree in Electrical Engineering from the University of Zagreb in 1993, and M.Sc. degrees from the Swiss Federal Institute of Technology (EPFL) and the University of Geneva in 1994 and 1995, respectively. His current research interests focus on virtual characters for the Internet and mobile platforms, and include Networked Collaborative Virtual Environments, facial analysis and synthesis, computer-generated film production and parallel computing. He has published one book

and around 50 papers on these topics. Igor was one of the key contributors to the Facial Animation specification in the MPEG-4 International Standard for which he received an ISO Certificate of Appreciation in 2000. He is involved in undergraduate and postgraduate teaching activities at Linköping and Zagreb Universities. He served in Program Committees of numerous conferences. Igor can be reached at *igor@isy.liu.se* and *Igor.Pandzic@fer.hr*.

Montse Pardàs received the MS degree in telecommunications and the Ph.D. degree from the Polytechnic University of Catalonia, Barcelona, Spain, in July 1990 and January 1995, respectively. Since September 1994 she has been teaching communication systems and digital image processing at this University, where she is currently associate professor. From January 1999 to December 1999 she was a research visitor at Bell Labs, Lucent Technologies, New Jersey. Her main research activity deals with image and sequence analysis, with a special emphasis on segmentation, motion and depth estimation, mathematical morphology and face analysis for synthetic model extraction. Montse Pardàs can be reached at *montse@gps.tsc.upc.es*.

Catherine Pelachaud received a Ph.D. in Computer Graphics at the University of Pennsylvania, Philadelphia, USA, in 1991. Between 1993 and 1994 she was involved in a project funded by an NSF grant, which implemented a system that automatically generates and animates conversations between multiple humanlike agents, with Prof. Badler, Prof. Cassell and Prof. Steedman. In 1993 she was part of the organization of a workshop, sponsored by NSF, on standards for facial animation held at the University of Pennsylvania. Between 1993 and 1996 she was a postdoctorate, with a Human Capital and Mobility grant, in the computer science department at the University of Rome. In 1998 she worked on the EU EAGLES project, which aims to promote standards and distribution of resources in the spoken language field. Since 2000 she has been working in MagiCster, a EU project, whose goal is to build a believable conversational agent. She is also part of the Natural Interaction and Multimodality Working Group of the EU project ISLE. Her research interest includes language standard for agent, conversational agent and human behavior simulation. Since 2002 she has been a research associate at the University of Rome 'La Sapienza'. Catherine Pelachaud can be reached at *cath@dis.uniroma1.it*

Eric Petajan is chief scientist and founder of face2face animation, inc, 2 Kent Place Blvd, Summit, NJ 07901, USA, and chaired the MPEG-4 Face and Body Animation (FBA) group. Prior to forming face2face, Eric was a Bell Labs researcher, where he developed facial motion capture, HD video coding and interactive graphics systems. Starting in 1989, Eric was a leader in the development of HDTV technology and standards leading up to the US HDTV Grand Alliance. He received a Ph.D. in EE in 1984 and an MS in Physics from the University of Illinois, where he built the first automatic lipreading system. Eric is also associate editor of the IEEE Transactions on Circuits and Systems for Video Technology. He can be reached at *eric@f2f-inc.com*.

Roberto Pockaj was born in Genoa, Italy, in 1967. He received the masters degree in Electronic Engineering in 1993 from the University of Genoa, Genoa, Italy, and the Ph.D. degree in computer engineering from the Department of Communication, Computer and System Sciences (DIST), University of Genoa, in 1999. From June 1992

to June 1996 he was with the Marconi Group, Genoa, Italy, working in the field of real-time image and signal processing, for optoelectric applications (active and passive laser sensors). Between 1996 and 2001 he collaborated on the management of the European projects ACTS-VIDAS and IST-INTERFACE, and participated in the definition of the new standard MPEG-4 for the coding of multimedia contents within the Ad Hoc Group on Face and Body Animation. In 2001 he co-founded a start-up company EPTAMEDIA srl, carrying on business in the area of facial animation software. He is currently a contract researcher at DIST. He has authored many papers on image processing and multimedia management. He can be reached at *roberto.pockaj@eptamedia.com* and *pok@dist.unige.it*.

Amaryllis Raouzaiou was born in Athens, Greece, in 1977. She graduated from the Department of Electrical and Computer Engineering, the National Technical University of Athens in 2000 and is currently pursuing her Ph.D. degree at the same University. Her current research interests lie in the areas of synthetic–natural hybrid video coding, human–computer interaction, machine vision and neural networks. She is a member of the Technical Champer of Greece. She is with the team of IST project ERMIS (Emotionally Rich Man-Machine Interaction Systems, IST-2000-29319). Amaryllis Raouzaiou can be reached by e-mail at *araouz@image.ece.ntua.gr*.

Nicolas Tsapatsoulis was born in Limassol, Cyprus, in 1969. He graduated from the Department of Electrical and Computer Engineering, the National Technical University of Athens in 1994 and received his Ph.D. degree in 2000 from the same University. His current research interests lie in the areas of human–computer interaction, machine vision, image and video processing, neural networks and biomedical engineering. He is a member of the Technical Champers of Greece and Cyprus and a member of IEEE Signal Processing and Computer societies. Dr. Tsapatsoulis has published eight papers in international journals and more than 20 in proceedings of international conferences. He served as Technical Program Cochair for the VLBV'01 workshop and as a member of the Technical Program Committee for the ICANN'02 conference. He is a reviewer of the *IEEE Transactions on Neural Networks* and *IEEE Transactions on Circuits and Systems for Video Technology* journals. Since 1995 he has participated in seven research projects at Greek and European level. Dr. Tsapatsoulis can be reached at *ntsap@image.ntua.gr* or at *http://www.image.ntua.gr/~ntsap/*.

Gael Sannier is a computer scientist who published several papers in the field of virtual humans. He worked in MIRALab, University of Geneva, for several years, where his research interests were focused on realistic texture mapping as well as on improving interactions with the virtual humans. He is a cofounder of W Interactive SARL, a French company that provides solutions for creation and animation of virtual characters on the Internet. More information on *http://www.winteractive.fr*, or email *gael@winteractive.fr*

Zhen Wen received the B. Engr degree from Tsinghua University, Beijing, China, and the MS degree from University of Illinois at Urbana Champaign, Urbana, Illinois, USA, both in computer science. Currently he is a Ph.D. student in the Department of Computer Science at University of Illinois at Urbana Champaign. His research interests are face modeling, facial motion analysis and synthesis, image-based modeling and rendering.

Liwen You received her B.Sc. in Electrical Engineering from Nanchang University of China, in 1997 and M.Sc. in Communication and Information Systems from Dalian

University of Technology of China in 2000. Now she is an international master student of Communication and Interactivity Program in Linköping University, Sweden. She is also a student member of IEEE computer society and communication society. She can be reached by e-mail at *liwenyou@ieee.org*.

Foreword

Ever since it became possible to drive the electron beams of cathode ray tubes (CRT) or the coils of loudspeakers with a computer, the number of applications that have been enabled has been staggering. The practical exploitation can be broadly classified as follows: (1) the computer can take the role of directly presenting some audio and visual samples acquired from the natural world and stored on a peripheral device or (2) it can present the result of computations based on some internal computer program that produces audio or visual information perceivable by human senses.

The former case has been driven by the telecommunication and audiovisual entertainment industries – even though these industries used to shy away from the idea of using a computer as the device driving the presentation – while the latter case has been driven by the Information Technology (IT) industry and by the hybrid mixture of the IT and entertainment industries called *Video games*.

Each of the industries involved used to have their own paradigms. The telecommunication and broadcast industries were obsessed with the idea of representing high-quality information with the smallest number of bits. This implied that the information – and it is a lot of bits for a two-hour movie – had to have a binary form. Further, processing could be virtually anything, provided that it could be implemented in special-purpose VLSI operating in real time and providing the given audio and video quality. On the other hand, the IT industry gave little importance to information representation. It sufficed that it be 'human readable' and understood by the specific computing device at hand. Real-time transmission was not an issue and the need to provide real-time pictures and audio was mitigated by the fact that information was local.

In July 1993, MPEG started its third standard project: MPEG-4. The title of the project 'Very low bit rate audiovisual coding' betrayed the original goal of serving the new unstructured world of ubiquitous fixed and mobile digital networks (not to mention portable storage devices). In a sense it was the continuation of its work done in its preceding two standards that largely targeted at the telecommunication and entertainment industries. Two years later, however, MPEG made the bold decision to try and integrate synthetic audio and visual information as well. This decision signaled the start of the Synthetic–Natural Hybrid Coding and was a major reason that prompted the change of the title of the MPEG-4 project six months later to 'Coding of audiovisual objects'.

This decision posed considerable new technical challenges. To provide what it claimed, MPEG-4 needed technologies to compose heterogeneous (natural and synthetic) audiovisual objects in a 3-D space. The starting point was provided by the

technology developed by the Virtual Reality Modeling Language (VRML) Consortium (now Web3D) with which a fruitful collaboration was started in 1997. VRML 97 provided a purely textual representation, but the BInary Format for Scene description (BIFS) provided a bit-efficient representation. In addition, MPEG added the ability to extend the static VRML composition technology with real-time updates. Lastly, MPEG provided a technology to compose objects in a 2-D space as a special case.

This book is particularly concerned with a special part of the MPEG-4 SNHC world: the capability to animate humanlike faces and bodies. MPEG-4 SNHC could draw from a wealth of research results that had been waiting for a conversion into commercial exploitation opportunities and succeeded in striking a balance between accuracy and efficient coded representation of the information needed to animate faces and bodies.

This book, authored by some of the major contributors to MPEG-4 Face and Body Animation (FBA), is recommended to those who want to have an in-depth knowledge of standard face and body animation technologies and to open a window on the wide world of applications enabled by MPEG-4 FBA.

Leonardo Chiariglione
Telecom Italia Lab

Preface

In the recently released MPEG-4 International Standard, the Moving Pictures Expert Group extended its focus from traditional audio and video coding to a much broader multimedia context including images, text, graphics, 3-D scenes, animation and synthetic audio. One of the more revolutionary parts of the new standard is the Face and Body Animation (FBA) – the specification for efficient coding of shape and animation of human faces and bodies. This specification is a result of collaboration of experts with different backgrounds ranging from image coding/compression to video analysis, computer graphics as well as speech analysis and synthesis, all sharing a common interest in computer simulation of humans. The result is a specification that is suitable for a wide range of applications not only in telecommunications and multimedia but also in fields like computer animation and human–computer interfaces. This is the first standard for animation of human faces and bodies, and it has already gained a wide acceptance within the research community. As practical applications emerge, this acceptance is spreading into the commercial areas as well.

In this book we concentrate on the animation of faces. The authors are some of the leading practitioners in this field, and include most of the important contributors to the MPEG-4 Facial Animation (FA) specification. In part one we put the MPEG-4 FA specification against the historical background of research on facial animation and model-based coding and provide a brief history of the development of the standard itself. In Part 2 we provide a comprehensive overview of the FA specification with the goal of helping the reader understand how the standard works, what is the thinking behind it and how it is intended to be used. Part 3, forming the bulk of the book, covers the implementations of the standard on both the encoding and decoding side. We present several face animation techniques for MPEG-4 FA decoders, as well as architectures for building applications based on the standard. While the standard itself actually specifies only the decoder, for applications it is interesting to look at technologies for producing FA content, and we treat a wide range of such technologies including speech analysis/synthesis and video analysis. The last part of the book brings a collection of applications using the MPEG-4 FA specification.

In the appendix, we propose a standard benchmark method, with publicly available data, for assessing the quality of MPEG-4 FA decoders.

We hope that this book will be a valuable companion for practitioners implementing applications based on the MPEG-4 FA specification or for those who simply wish to understand the standard and its implications.

Igor S. Pandzic, Robert Forchheimer
Editors