Introduction to the Special Issue on MPEG-7

TODAY, with the explosive growth in availability of digital multimedia content from a variety of sources, such as scanners, CDs, DVDs, cameras, digital audio and video recorders, broadcast, and libraries on the Internet, finding what is needed has become increasingly harder. To address this challenge, with the help of advances in the areas of annotation, indexing, and natural language processing, more and more sophisticated search engines are being developed to facilitate search for multimedia content. However, with a few exceptions, much of this search is text-centric, and thus dependent on the keywords used to index the multimedia content, as well as nuances of specific search engines. Even with an intimate knowledge of a search engine, users still typically face difficulty in locating that specific piece of content that they are looking for, or may receive many hits due to potential mismatches, requiring considerable time for browsing to narrow down the search for multimedia content. Furthermore, learning the intricacies of different search engines so that effective searches can be conducted using them can require considerable effort. A better solution that allows multimedia-centric search is needed.

Search/retrieval and browsing are essentially pull applications. Filtering of multimedia content, such as from broadcast, is another type of application, called a push application. Here, the term *broadcast* is used loosely and is taken to include over-the-air TV, cable TV, satellite TV, and other applications where multimedia content is pushed to the consumer. Television guides with brief textual description or brief visual previews of certain programs exist. However, while consumers may have access to several hundred channels of TV content, the amount of information available to help filter and select the content of interest is very limited, and mainly requires manual effort. Even if users can locate what they are looking for and record the content for later consumption, the size of stored information may grow rapidly. Effective yet simple-to-use tools would be needed for managment and extraction of content as needed. Relatively few tools that are effective—yet simple to use—are currently available commercially.

To facilitate effective search or filtering, multimedia content needs to be described effectively. To this end, "meta-data" descriptors are needed to express various features of the multimedia content. Furthermore, standardization of features becomes important to allow devices of different vendors to interoperate with each other. Over the last 15 years, a number of organizations have attempted to standardize features and descriptors within the context of specific multimedia applications. Even when applications from different vendors are somewhat similar, the features and descriptors they support can be quite different due to insufficient standardization. This reduces interoperability due to difficulty in using databases from a different source with a particular search engine. It also requires a user to be aware of such intricacies to be effective in search or filtering. Thus, an interoperable solution that reduces the burden on the user is needed.

Realizing the aforementioned situation, and acknowledging the industry demand and the availability of adequate technology, the Moving Picture Experts Group (MPEG) of the International Standards Organization (ISO) initiated work on MPEG-7, its fourth standard, in 1997. MPEG-7 was termed the "Multimedia Content Description Standard," quite different to earlier MPEG standards such as MPEG-1, MPEG-2, and MPEG-4 which address video coding. MPEG-7 specifies standardized Descriptors and Description Schemes for audio and video, as well as integrated multimedia content. Also standardized is a Description Definition Language that allows new Descriptors and Description Schemes to be defined. Furthermore, MPEG-7 specifies a systems layer that allows synchronization and access of MPEG-7 description meta data with or without the corresponding multimedia data.

MPEG-7 has ignited much interests in industry with broad interests in media, content management, content provision, broadcasting, computers, and telecommunications, as well as researchers and academia. Currently, the MPEG-7 standard is in a fairly mature stage and is on its way to final standardization later this year. The pace of progress in development of MPEG-7 has been quite frantic. This Special Issue provides a snapshot of the status of MPEG-7 technology as of around April 2001, and contains only invited papers from experts who are involved in MPEG-7 standardization. It is designed to be useful to beginners, as well as experts alike, as it covers various topics at different depths.

The first paper of this issue, "Overview of the MPEG-7 Standard" by S.-F. Chang, T. Sikora, and A. Puri, introduces the goal, context, terminology, components, and open issues of the MPEG-7 standard, and its relationships with other meta-data standards.

Next, a set of four papers addresses the visual part of the MPEG-7 standard. The first, "The MPEG-7 Visual Standard for Content Description—An Overview" by T. Sikora, presents a high-level overview of the organization and components of the MPEG-7 Visual Standard. The next paper, "Color and Texture Descriptors" by B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, discusses specific Descriptors to describe color and texture in visual scenes. This is followed by "MPEG-7 Visual Shape Descriptors" by M. Bober, which discusses the Descriptors available to describe shapes in visual scenes. The last of the set of visual papers, "MPEG-7 Visual Motion Descriptors" by S. Jeannin and A. Divakaran, discusses the representation of motion in visual scenes.

The next set of three papers addresses the audio part of the MPEG-7 standard. The first paper, "Overview of MPEG-7 Audio" by S. Quackenbush and A. Lindsay, presents a high-

Publisher Item Identifier S 1051-8215(01)05723-8.

level overview of the organization and components of the MPEG-7 Audio Standard. The next paper, "SpokenContent Representation in MPEG-7" by J. P. A. Charlesworth and P. N. Garner, introduces tools for recognition of spoken content in MPEG-7 Audio. The last paper of this set, "MPEG-7 Sound Recognition Tools" by M. Casey, discusses the tools for recognition of sounds included in MPEG-7 Audio.

The paper "MPEG-7 Multimedia Description Schemes" by P. Salembier and J. R. Smith addresses the description schemes for multimedia content of the MPEG-7 Standard.

The last set consists of two papers that, respectively, address systems and Description Definition Language (DDL) parts of the MPEG-7 Standard. The first, "MPEG-7 Systems: Overview" by O. Avaro and P. Salembier, presents an overview of progress in MPEG-7 Systems. The second, "An Overview of the MPEG-7 Description Definition Language (DDL)" by J. Hunter, introduces DDL, a language standardized in MPEG-7 to describe Descriptors and Description Schemes.

We hope that you enjoy this collection of outstanding papers on MPEG-7.

SHIH-FU CHANG, *Guest Editor*
Department of Electrical Engineering
Columbia University
New York, NY 10027 USA

ATUL PURI, *Guest Editor*
AT&T Labs–Research
Red Bank, NJ 07701 USA

THOMAS SIKORA, *Guest Editor*
Department of Interactive Media—Human Factors
Heinrich–Hertz Institute
Berlin D-10587, Germany

HONGJIANG ZHANG, *Guest Editor*
Microsoft Research
5F Beijing Sigma Center
Beijing 100080, China

Shih-Fu Chang (S'89–M'93–SM'01) is an Associate Professor of Electrical Engineering at Columbia University, New York, where he currently leads the ADVENT university–industry consortium and the digital video and multimedia group. The group conducts research in searching, filtering, manipulation, transmission, and security of digital video and multimedia information, in addition to having played an active role in developing the ISO MPEG-7 Standard. He has actively participated in several large-scale cross-disciplinary projects, including Columbia's Digital Library Project, Digital News System, and Art Image Education for K–12 Schools. He and his group have developed several well-known systems and algorithms for video searching, like VideoQ, WebSEEk, and compressed video manipulation, like WebClip. He has also served as a Consultant in several new media technology companies.

Prof. Chang is currently a Distinguished Lecturer of the IEEE Circuits and Systems Society in the area of multimedia systems and technology. He was a General Co-Chair of the ACM 8th Multimedia Conference 2000 and an Associate Editor for several journals. He was awarded the Navy ONR Young Investigator Award, a CAREER Award from the the National Science Foundation, a Faculty Development Award from IBM, a Service Recognition Award from ACM, and has received three Best Paper Awards from IEEE, ACM, and SPIE.

Atul Puri (S'86–M'87) received the M.S. degree from the City College of New York in 1982, and the Ph.D. degree a from the City University of New York in 1988, both in electrical engineering.

He was a Consultant in the Visual Communications Research Department, Bell Labs, Holmdel, NJ, which he joined as a member of Technical Staff in 1988. Since 1996, he has been a Technology Consultant in image processing research, AT&T Labs, Middletown, NJ, having previously been a Principal Technical Staff Member. He has been involved in video-coding research for videoconferencing, video on digital storage media, HDTV, and stereoscopic TV applications. His current research interests are in the areas of flexible video and multimedia systems and services for the Web/Internet. He has represented AT&T at ISO MPEG audio–visual coding standards, such as MPEG-1, MPEG-2, and MPEG-4. He is a Technical Editor of the Video part of the MPEG-4 standard and is also involved in the MPEG-7 multimedia content description standard, currently in progress. He holds over 24 patents and has published over 40 technical papers. He is a coauthor of the book *Digital Video: An Introduction to MPEG-2* (London, U.K.: Chapman & Hall, 1997) and co-editor of the book *Multimedia Systems, Standards and Networks* (New York: Dekker, 2000).

Dr. Puri is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. He has received a number of awards from AT&T, including the Exceptional Contribution Award, the Individual Performance Award, and the Standards Recognition Award. He is also a recipient of the ISO Prize. He is a member of IEEE Communications Society and the IEEE Signal Processing Society.

Thomas Sikora (M'93–SM'96) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Bremen University, Germany, in 1985 and 1989, respectively.

In 1990, he joined Siemens Ltd., Melbourne, Australia, and Monash University, Melbourne, Australia, as a Project Leader, responsible for video compression research activities in the Australian "Universal Broadband Video Codec" consortium. Between 1994 and 1998, he directed video-coding research activities at Heinrich–Hertz Institute (HHI), Berlin, Germany. He is currently Head of the Interactive Media—Human Factors Department at HHI, in addition to being an Adjunct Professor in the Electrical Engineering Department, Technical University Berlin, Berlin, Germany. He is co-founder of $2SK$ Media Technologies and Perspective Technologies GmbH, two recent Berlin-based start-up companies involved in audio and video signal processing, compression technology and 3-D visualization. He has been involved in ITU and ISO standardization activities, as well as in several European research activities for a number of years. As the chairman of the ISO-MPEG video group, he is responsible for the development and standardization of the MPEG video-coding algorithms. Between 1995–1998, he also served as the chairman of the European COST 211^{ter} video compression research group. He is frequently an industry consultant on issues related to interactive digital audio and video.

Dr. Sikora was the recipient of the 1996 German ITG Award (German Society for Information Technology). He is an Associate Editor for a number of international journals, including the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *IEEE Signal Processing Magazine*, the *EURASIP Signal Processing Journal*, and the *EURASIP Signal Processing: Image Communication Journal*. He is a member of ITG.

HongJiang Zhang (S'88–M'91–SM'97) received the B.S. degree from Zhengzhou University, China, and the Ph.D. degree from the Technical University of Denmark, both in electrical engineering, in 1982 and 1991, respectively.

From 1992 to1995, he was with the Institute of Systems Science, National University of Singapore, where he led several of the earliest projects in the area of video and image content analysis and retrieval. He was also with the MIT Media Lab, Cambridge, MA, during 1994 as a Visiting Researcher. From 1995 to 1999, he was a Research Manager at Hewlett-Packard Labs, where he was responsible for research and technology transfers in the areas of multimedia management, intelligent image processing, and Internet media. In 1999, he joined Microsoft Research, Beijing, China, where he is currently a Senior Researcher and the Assistant Managing Director, mainly in charge of media computing and information processing research. He has co-authored three books, over 130 referred papers and book chapters, seven special issues of international journals in multimedia processing, content-based media retrieval, and Internet media, as well as numerous patents.

Dr. Zhang currently serves on the Editorial Boards of five professional journals, including IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and a dozen committees of international conferences.