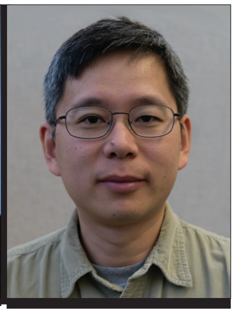
INTERACTIVE 3D VIDEO STREAMING







Ivan V. Bajic *Gene Cheung* *Pascal Frossard* *Joern Ostermann* *Wai-tian Tan*

Streaming of single-view video is now commonplace, so we expect to access any video content anywhere on any device. The delivery of 3D video streams, however, brings new technical challenges along with the promise of adding visual realism to the 3D scene by enhancing the perception of depth. The term "3D video" encompasses different types of video:

- *Stereoscopic vide*o, with two slightly shifted video views for the observer's left and right eyes
- *Multiview vide*o, where more than two video views are encoded for either an advanced multiview display terminal, or interactive applications where an observer can freely select any of the captured views for observation
- *Free viewpoint video*, where geometric information of the 3D scene (e.g., depth maps) are transmitted along with multiview videos so that novel images can be synthesized at virtual viewpoints via depth-image-based rendering (DIBR)

Coding and transport for each of these 3D video types require particular algorithms and specific optimization methods. For stereoscopic video, besides obvious interview correlation that can be exploited for coding gain, the human's quality perception of stereo images (where often the poor quality of one view image is "masked" by the better quality image of the other view) can also be leveraged for asymmetric bit allocation and transport. For multiview video, one unique challenge lies in encoding videos of different views in such a way that interactive view switching is facilitated, while good compression efficiency is preserved. View switching may be triggered by an observer's head movement, for example, and the viewpoint image corresponding to the new head position should be rendered fast enough to enable seamless and natural interaction with the 3D scene. For free viewpoint video, texture and depth images are fundamentally different kinds of signals and affect the image quality of the synthesized intermediate views in different ways during DIBR. Therefore, encoding and transport of texture and depth videos should be tailored to their respective signal characteristics and influence on synthesized image quality. In limited bandwidth applications, multiview and free viewpoint video put lower limits on latency than video telephony.

In this Feature Topic (FT) on interactive 3D video streaming, we have gathered a collection of articles that address different aspects of this important topic. The first article, "Adaptive Multiview Video Streaming: Challenges and Opportunities," provides an excellent overview of many of the technical challenges related to adaptive multiview video. The article discusses content representation, encoding, transmission issues in unicast and multicast, multipath and cloud-assisted delivery, client-side processing, as well as social aspects of multiview video distribution.

The second article, "Quality of Experience for 3D Video Streaming," provides an accessible account of several areas related to 3D quality of experience (QoE). It discusses how various artifacts occur and describes their effects on 3D QoE. It also presents different ways of estimating true 3D QoE, expressed in mean opinion scores, using measurable quantities in both full reference and reduced/no reference settings.

The third article discusses solutions for the delivery of multiview video in distributed architectures. It introduces a framework for 3D delivery in IP-based P2P systems and discusses the specific design of P2P overlays for these novel data. The article finally presents two representative scenarios that illustrate the benefits of these collaborative systems.

There are important issues that have not yet been sufficiently addressed in the literature. In interactive applications where a viewer periodically selects a subset of captured views for observation of the 3D scene, the question of enabling network-delay-free view switching while lowering transmission rate remains a relatively unexplored problem. Quality assessment of free viewpoint video has yet to be thoroughly investigated. In particular, the viewer's perception of 3D visual quality as the synthesized images are degraded due to lossy encoding or packet losses of texture and depth video, as well as smoothness and interaction delay in view switching, is not fully understood. Emerging 3D displays such as glasses-free multiview displays mean that multiple viewers can now interact simultaneously with the same visual content in the same physical space; the optimization of the coding and transport of multiview video content in such challenging situations is another interesting and open problem. It is our hope that this FT serves as an introductory resource to researchers who are new to the topic and as a springboard to future exploration into this fast growing field.

In closing, we would like to thank all the authors who submitted manuscripts to this FT, as well as all the reviewers who helped to assess the quality of the submitted manuscripts on a relatively tight schedule. Special thanks go to the past Editor-in-Chief, Dr. Steve Gorshe, the current Editor-in-Chief, Dr. Sean Moore, as well as the production staff, for making this issue possible.

BIOGRAPHIES

IVAN V. BAJIC [SM] (ibajic@ensc.sfu.ca) received his B.Sc.Eng. degree (summa cum laude) in electronic engineering from the University of Natal, South Africa, in 1998, and his M.S. degree in electrical engineering, his M.S. degree in mathematics, and his Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, New York, in 2000, 2002, and 2003, respectively. From 2003 to 2005, he was a visiting assistant professor with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, Florida. Since 2005, he has been with the School of Engineering Science, Simon Fraser University, Burnaby, British Columbia, Canada, where he is currently an associate professor and co-director of the Multimedia Communications Laboratory. His research interests include image and video processing and compression, perceptual aspects of visual information processing, and multimedia communications. He has served on the program committees of various conferences in the field, including GLOBECOM, ICC, ICME, and ICIP, and was Chair of the Media Streaming Interest Group of the IEEE Multimedia Communications Technical Committee, 2010–2012.

GENE CHEUNG [SM] (cheung@nii.ac.jp) received his B.S. degree in electrical engineering from Cornell University, Ithaca, New York, in 1995, and M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively. From August 2000 to October 2009, he was a senior researcher at Hewlett-Packard Laboratories Japan, Tokyo. He is currently an associate professor at the National Institute of Informatics, Tokyo. His research interests include robust media representation, single-/multiple-view video coding and streaming, and immersive communication. He served as an Associate Editor of *IEEE Transactions on Multimedia* from 2007 to 2011 and currently serves as Associate Editor of the DSP Applications column in *IEEE Signal Processing Magazine*, Associate Editor of *APSIPA Journal on Signal & information Processing*, and as an Area Editor for *EURASIP Signal Processing: Image Communication*. He also served as Area Chair for the IEEE International Conference on Image Processing (ICIP) 2010 and 2012, Technical Program Co-Chair of International Packet Video Workshop (PV) 2010, Track Co-Chair of the Multimedia Signal Processing track of the IEEE International Conference on Multimedia and Expo (ICME) 2011, Symposium Co-Chair for the CSSMA Symposium at IEEE GLOBECOM 2012, and Area Chair for ICME 2013. He is a co-recipient of the best student paper award at the IEEE Workshop on Streaming and Media Communications 2011 (in conjunction with ICME 2011), a best paper finalist at ICME 2011 and ICIP 2011, and a best paper runner-up award at ICME 2012.

PASCAL FROSSARD [M] (pascal.frossard@epfl.ch) received M.S. and Ph.D. degrees, both in electrical engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively. Between 2001 and 2003, he was a member of the research staff at the IBM T. J. Watson Research Center, Yorktown Heights, New York, where he worked on media coding and streaming technologies. Since 2003, he has been a professor at EPFL, where he heads the Signal Processing Laboratory (LTS4). His research interests include image representation and coding, nonlinear representations, visual information analysis, joint source and channel coding, multimedia communications, and multimedia content distribution. He was the General Chair of IEEE ICME 2002 and Packet Video 2007. He was the Technical Program Chair of EUSIPCO 2008, and a member of the organizing or technical program committees of numerous conferences. He has been an Associate Editor of *IEEE Transactions on Multimedia* (2004–) and of *IEEE Transactions on Circuits and Systems for Video Technology* (2006–). He is an elected member of the IEEE Image and Multidimensional Signal Processing Technical Committee (2007–), the IEEE Visual Signal Processing and Communications Technical Committee (2006–), and the IEEE Multimedia Systems and Applications Technical Committee (2005–). He has served as Vice-Chair of the IEEE Multimedia Communications Technical Committee (2004–2006) and as a member of the IEEE Multimedia Signal Processing Technical Committee (2004–2007). He received the Swiss NSF Professorship Award in 2003, and IBM Faculty Award in 2005, and the IBM Exploratory Stream Analytics Innovation Award in 2008.

JOERN OSTERMANN [F] (ostermann@tnt.uni-hannover.de) studied electrical engineering and communications engineering at the University of Hannover and Imperial College London, respectively. He received his Dipl.-Ing. and Dr.-Ing. from the University of Hannover in 1988 and 1994, respectively. In 1994 and 1995 he worked at AT&T Bell Labs. He was with AT&T Labs — Research from 1996 to 2003. Since 2003 he has been a full professor and head of the Institut für Informationsverarbeitung at the Leibniz Universität Hannover, Germany. Since 2011 he has been dean of the Department of Electrical Engineering and Computer Science. From 1993 to 1994, he chaired the European COST 211 sim group coordinating research in low-bit-rate video coding. He chaired the Ad Hoc Group on Coding of Arbitrarily-shaped Objects in MPEG-4 Video. Since 2008, he has been Chair of the Requirements Subgroup of MPEG (ISO/IEC JTC1 SC29 WG11). He was a scholar of the German National Foundation. In 1998, he received the AT&T Standards Recognition Award and the ISO award. He served as a Distinguished Lecturer of the IEEE CAS Society. He has published more than 100 research papers and book chapters. He is coauthor of a graduate level textbook on video communications. He holds more than 30 patents. His current research interests are video coding and streaming, 3D modeling, face animation, and computer-human interfaces.

WAI-TIAN TAN (wai-tian.dat@hp.com) received his B.S. degree in electrical engineering from Brown University in 1992, his M.S.E.E. degree from Stanford University in 1993, and his Ph.D. degree in electrical engineering and computer science from the University of California, Berkeley, in 2000. He has been with Hewlett Packard Laboratories, Palo Alto, California, since 2000, and is currently a principal research scientist. His research interests are in developing adaptive streaming systems under different practical constraints, and more generally in coding, communications, and new applications of video. He has over 50 research papers, holds 14 patents, and has developed production multimedia software for mobile devices and telecommunication networks.