Introduction to the Issue on Interactive Media Processing for Immersive Communication

INTER-PERSONAL COMMUNICATION is an integral part of human society. For education, we communicate to teach and to learn; for work, we communicate to exchange and refine ideas; for recreation, we communicate to share presence and socialize as a community. While much of inter-personal communication today is conducted face-to-face in person, the recent advances in different facets of information technologies (e.g., depth sensing cameras, large high-resolution 3D displays, ubiquitous high-speed wireless networks) offer promises of sophisticated communication systems that can create immersive media experience for participants physically separated by large distance but connected via high-speed data networks. By "immersive," we mean participants, through real-time interaction with the presented media on their respective terminals, would look and feel as if they are sitting in the same room conversing with their partners naturally. Unlike fixed, non-interactive presentations of camera-captured videos as done in today's conference tools such as Skype and Google Hangout, it would mean participants can maintain eye contact as they converse (regardless of the locations of the capturing cameras). It would mean a participant can observe different viewpoints of the 3D scene as he/she shifts his head—a phenomenon called *motion parallax* that is the strongest cue in human's perception of depth. In some cases, it would even mean the remote sharing of human's touch sensation, also called haptics. In this issue, we focus on state-of-the-art technologies that enable such immersive communication.

With guidance from Editor-in-Chief Fernando Pereira, assistance from Publications Administrator Rebecca Wollman and recommendations from a team of dedicated reviewers, we selected 14 papers for this special issue from a competitive field of high-quality submissions and categorized them into the following categories. In the first category are papers that are concerned with the processing of 3D geometric data: how the geometry of the 3D scene can be efficiently acquired, denoised and compressed at the sender. Transmitted geometric information can enable a range of applications at the receiver, such as virtual view synthesis via depth-image-based rendering (DIBR). Towards better 3D geometry estimation for dynamic scene, Li *et al.* ("Single Shot Dual-Frequency Structured Light Based Depth Sensing") propose using a mixture of periodic waves, where the ambiguity is resolved using a novel application of number theory. Cong *et al.* ("Accurate Dynamic 3D Sensing with Fourier-Assisted Phase Shifting") propose a new phase-shifting method: motion vulnerability of multishot phase shifting profilometry can be overcome through single-shot Fourier transform profilometry, while still preserving high accuracy. Moreover, an efficient parallel spatial unwrapping strategy is proposed to solve the phase ambiguity of complex scenes without additional images. Turner *et al.* ("Fast, Automated, Scalable Generation of Textured 3D Models of Indoor Environments") propose scalable methods for modeling indoor environment from large-scale 3D point cloud data. The proposed methods can be applied to virtual walk-throughs of environments, gaming entertainment, augmented reality, indoor navigation, and energy simulation analysis.

Post-processing of acquired 3D geometric data is often then introduced to make low-level data more accessible and useable for applications. For compression, Ahn *et al.* ("Large-Scale 3D Point Cloud Compression Using Adaptive Radial Distance Prediction in Hybrid Coordinate Domains") propose to compress large-scale 3D point cloud via efficient coding of projected range images. To improve quality of acquired depth images, Rana *et al.* ("Probabilistic Multiview Depth Image Enhancement Using Variational Inference") propose an enhancement technique that weights contributions of depth pixels from different viewpoint images differently, given depth images are often estimated inaccurately due to imperfect stereo matching algorithms at the encoder. For foreground extraction, Zhao *et al.* ("Real-time and Temporal-coherent Foreground Extraction with Commodity RGBD Camera") propose a fully automatic, temporally coherent approach for real-time foreground object extraction from an RGBD video stream captured by a commodity depth camera. Temporal coherence of object extraction is achieved by using a closed-form matting formulation.

Advanced data sensing goes beyond 3D geometry processing, and Chaudhari *et al.* ("Perceptual and Bitrate-scalable Coding of Surface Textures in Remote Haptic Interaction") study haptics for immersive communication, showing that perceptual masking—a well-known phenomenon in audio—is also observed in haptics, and propose an efficient haptics codec based on a novel masking model.

In the second category are papers that are concerned with the transmission/streaming of immersive media data; compared to traditional single-view video streaming, the volume of immersive media captured from multiple cameras/microphones tend to be much larger and in higher dimension (e.g., multiple views), and the delay requirements tend to be more stringent for real-time media interactivity. Chakareski *et al.* ("View-Popularity-Driven Joint Source and Channel Coding of View and Rate Scalable Multi-view Video") study the scenario of multicasting multi-view video content in video-plus-depth format to a collection of heterogeneous clients. They propose a popularity-aware joint source-channel coding optimization framework that allocates source and channel coding rates to the captured content, such that the aggregate video quality of the reconstructed content across the client population is max-

Digital Object Identifier 10.1109/JSTSP.2015.2406092

1932-4553 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

imized. In an interactive multiview video streaming (IMVS) scenario where a user requests only one view at a time from a streaming server, De Abreu *et al.* ("Optimizing Multiview Video plus Depth Prediction Structures for Interactive Multiview Video Streaming") propose an algorithm to select optimal inter-view prediction structures (PSs) and quantization parameters (QPs) for coding of texture and depth maps so that visual distortion is minimized, subject to storage and transmission rate constraints. Badr *et al.* ("Streaming Codes with Partial Recovery over Channels with Burst and Isolated Erasures") study forward error correction codes for low-delay, real-time streaming communication over packet erasure channels. They consider "streaming codes" and propose a simplified class of erasure channels that introduce both burst and isolated erasures within the same decoding window. In the context of interactive network games, Lu *et al.* ("Cloud Mobile 3D Display Gaming User Experience Modeling and Optimization by Asymmetric Graphics Rendering") propose a mobile gaming platform, where 3D game video are rendered in cloud servers and streamed to mobile users in optimized bitrates.

In the third category, we selected three papers related to the system aspects of immersive communication focusing on end user experience. For images rendered on stereo or auto stereo 3D displays, Lee *et al.* ("3D Perception based Quality Pooling: Stereopsis, Binocular Rivalry and Binocular Suppression") propose a comprehensive and complete human 3D perception model for stereo images. They also design a stereo image quality pooling (3DPS) technique that classifies segment units in a stereo image into binocular or monocular vision regions. This technique is particularly useful to optimize perceptual quality of stereo images distorted by coding and transmission errors. Telepresence percept is a desirable feature for successful immersive communication from a user experience point of view. We selected two papers illustrating such systems that tend to attain this goal and illustrate how technical solutions can be integrated with human factors considerations. Nguyen *et al.* ("ITEM: Immersive Telepresence for Entertainment and Meetings—A Practical Approach") propose an end-to-end immersive telepresence system called ITEM that focuses on maximizing user experience. The system contains key technologies optimized from a system perspective, such as fast object-based video coding, spatialized audio capture and 3D sound localization. Roberts *et al.* ("Withyou—an experimental end-to-end telepresence system using video-based reconstruction") present a telepresence research platform, where non-verbal cues such as interpersonal distance, gaze, poster and facial expressions are considered in evaluation of user experience, so that distribution of computing and network resources can be optimized accordingly.

There remain many challenges in immersive communication. From a data sensing point of view, while a great deal of effort has been made to acquire/denoise/compress 3D geometric data, how to best utilize the geometric data at the receiver to enable more natural human-centric visual experience (e.g., better quality or more geometry-consistent virtual view synthesis) is still an open problem. For haptics, how to best incorporate haptic signals into a multi-modal media experience for end user (e.g., how a user perceives haptics in the presence of other strong video/audio signals) is understudied. From a network transmission point of view, in an interactive media streaming scenario where a user periodically requests a subset of media (e.g., a subset of captured viewpoint images) for image rendering at end terminal, what is the best high-dimensional visual data representation—one that is tailored specifically for interactive streaming of media subsets as per user request—is still not fully addressed. From a user experience point of view, despite recent advances in quality assessment of multimedia, how to best evaluate end-to-end immersive communication systems with multiple sensory interactions still requires some efforts in line with the emerging science of Quality of Experience (QoE). Encompassing issues beyond perceptual image quality such as higher cognitive concepts (e.g., naturalness, immersiveness, audiovisual experience), such systems appear to be challenging use cases for the QoE community from a subjective evaluation point of view. We hope that this special issue serves as a useful starting point for further investigation into these challenges in immersive communication.

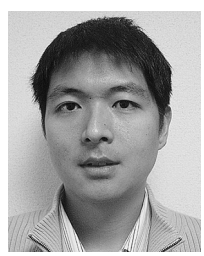
GENE CHEUNG, *Lead Guest Editor*
National Institute of Informatics
101-8430 Tokyo, Japan

DINEI FLORENCIO, *Guest Editor*
Microsoft Research
Redmond, 98052 WA USA

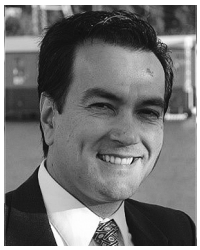
PATRICK LE CALLET, *Guest Editor*
University of Nantes
44306 Nantes, France

CHIA-WEN LIN, *Guest Editor*
National Tsing-Hwa University
30013 Nantes, Taiwan

ENRICO MAGLI, *Guest Editor*
Politecnico di Torino
10129 Torino, Italy



Gene Cheung (M'00–SM'07) received the B.S. degree in electrical engineering from Cornell University in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California, Berkeley, in 1998 and 2000, respectively. He was a Senior Researcher with Hewlett-Packard Laboratories Japan, Tokyo, from 2000 till 2009. He is now an Associate Professor in National Institute of Informatics in Tokyo, Japan. He is an Adjunct Associate Professor in the Hong Kong University of Science and Technology (HKUST) since 2015. His research interests include image and video representation, immersive visual communication and graph signal processing. He has served as associate editor for IEEE TRANSACTIONS ON MULTIMEDIA (2007–2011) and currently serves as associate editor for DSP Applications Column in IEEE *Signal Processing Magazine*, *APSIPA Journal on Signal and Information Processing* and *SPIE Journal of Electronic Imaging*, and as area editor for *EURASIP Signal Processing: Image Communication*. He is a co-author of best student paper award in IEEE Workshop on Streaming and Media Communications 2011 (in conjunction with IEEE International Conference on Multimedia & Expo (ICME) 2011), best paper finalists in ICME 2011 and IEEE International Conference on Image Processing (ICIP) 2011, best paper runner-up award in ICME 2012, and best student paper award in ICIP 2013.



Dinei Florencio (M'97–SM'05) received the B.S. and M.S. degrees from the University of Brasilia, Brasilia, Brazil, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, USA, all in electrical engineering. He has been a Researcher with Microsoft Research, Redmond, WA, USA, since 1999. Before joining Microsoft, he was a Member of the Research Staff at the David Sarnoff Research Center from 1996 to 1999. He was also a Co-Op Student with the AT&T Human Interface Laboratory (now part of NCR) from 1994 to 1996, and a Summer Intern at Interval Research Corporation, Palo Alto, CA, USA, in 1994. He has authored over 70 papers, and holds 55 granted U.S. patents. Dr. Florencio was General Co-Chair of MMSP 2009, WIFS 2011, Hot3D 2010 and 2013, and Technical Co-Chair of WIFS 2010, ICME 2011, and MMSP 2013. He is the Chair of the IEEE SPS Technical Committee on Multimedia Signal Processing (from 2014 to 2015), and a member of the IEEE SPS Technical Directions Board.



Patrick Le Callet (M'07–SM'14) received both an M.Sc. and a Ph.D. degree in image processing from Ecole polytechnique de l'Université de Nantes. He was also a student at the Ecole Normale Superieure de Cachan where he sat the "Aggrégation" (credentialing exam) in electronics of the French National Education. He worked as an Assistant Professor from 1997 to 1999 and as a Full Time Lecturer from 1999 to 2003 at the Department of Electrical Engineering of Technical Institute of the University of Nantes (IUT). Since 2003 he has taught at Ecole Polytechnique de l'Université de Nantes (Engineering School) in the Electrical Engineering and the Computer Science departments where is now a Full Professor. Since 2006, he has been the head of the Image and Video Communication lab at CNRS IRCCyN, a group of more than 35 researchers. He is mostly engaged in research dealing with the application of human vision modeling in image and video processing. His current centers of interest are 3D image and video quality assessment, watermarking techniques, and visual attention modeling and applications. He is co-author of more than 200 publications and communications and co-inventor of 13 international patents on these topics. He also co-chairs within the VQEG (Video Quality Expert Group) the "HDR Group" and "3DTV" activities. He is currently serving as associate editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS AND VIDEO TECHNOLOGY and IEEE TRANSACTIONS ON IMAGE PROCESSING.



Chia-Wen Lin (S'94–M'00–SM'04) received his Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000. He is currently a Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. He is also an Adjunct Professor with Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, during 2000–2007. Prior to joining academia, he worked for the Information and Communications Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, during 1992–2000. His research interests include image and video processing and video networking. Dr. Lin has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE MULTIMEDIA, and the *Journal of Visual Communication and Image Representation*. He has also recently served as guest editors for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, and *EURASIP Signal Processing: Image Communication*. He is currently Chair of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He served as Technical Program Co-Chair of the IEEE International Conference on Multimedia & Expo (ICME) in 2010, and Special Session Co-Chair of the IEEE ICME in 2009. His paper won the top 10% paper award presented by IEEE MMSP 2013, and Young Investigator Award presented by VCIP 2005. He received the Young Faculty Awards presented by CCU in 2005 and the Young Investigator Awards presented by National Science Council, Taiwan, in 2006.



Enrico Magli (S'97–M'01–SM'07) received the Ph.D. degree in electrical engineering from Politecnico di Torino, Torino, Italy, in 2001. He is currently an Associate Professor with Politecnico di Torino, where he leads the Image Processing Laboratory. His research interests are in the field of compression of satellite images, multimedia signal processing and networking, compressive sensing, distributed source coding, and image and video security. Dr. Magli was a General Cochair of IEEE International Workshop on Multimedia Signal Processing (MMSP) 2013 and IEEE International Conference on Multimedia and Expo (ICME) 2015. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, of the IEEE Transactions on Multimedia, and of the *EURASIP Journal on Image and Video Processing*. He was the recipient of the 2010 Best Reviewer Award of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING and the 2010 and 2014 Best Associate Editor Award of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and a corecipient of the IEEE Geoscience and Remote Sensing Society 2011 TRANSACTIONS Prize Paper Award.