Introduction to the Special Section on Augmented Video

MERGING computer-generated content with real-world visual data is one of the main challenges in fields like augmented reality or visual effects and is increasingly important in broadcasting, gaming, medical, automotive, maintenance, and learning applications. Although augmented reality (AR) has been investigated for a long time, it has recently emerged as a hot topic, with significant commercial interest. One reason for that is the availability of new camera-equipped devices like smart phones or tablets that enable see-through capabilities. Combined with powerful graphics capabilities, sensors, and tracking methods, AR now becomes available to everyone. In addition, glasses-based systems, like Microsoft's HoloLens, allow for hands-free visualization, enabling many new applications.

However, high-quality augmented video still poses significant challenges concerning accurate tracking, registration, content adaptation, and object insertion. Therefore, many existing applications merely superimpose information and objects relevant to the displayed content instead of aiming at a seamless integration of rendered content. On the other hand, there is a large body of work targeting realistic insertion of computer-generated imagery content, e.g., among the visual effects community, but this mainly focuses on offline workflows with often intensive manual interaction. For an automatic and immersive fusion of real and synthetic video data, several issues still need to be solved.

This special issue aims at providing readers with the latest developments and emerging technologies that drive the convergence of these fields toward realistic composition of real and computer-generated video data. It addresses novel algorithms that target the generation and seamless fusion of content for high-quality applications within e.g. augmented reality, telepresence, gaming, or video rendering. For that purpose, 11 papers have been selected during the review process from 24 submissions, covering all aspects from individual components with respect to tracking or content creation to full augmented video systems.

The first paper, entitled "Video Stabilization for Strict Real-Time Applications," by Dong *et al.* addresses real-time tracking, which is an essential component for AR systems. The authors use short homography sequence estimates smoothed with Kalman filters, targeting video completion. The next paper, entitled "WELD: Weighted Low-Rank Decomposition for Robust Grayscale-Thermal Foreground Detection," by Li *et al.* also targets the image registration problem, but between grayscale and thermal video pairs. After fusion, objects in the scene can be accurately detected and tracked from the multimodal video data.

For accurate overlays fixed to particular surface points, a 3D scene model is required, which can be built from visual data. The next two papers therefore deal with 3D scene reconstruction. In "Light Field Depth Estimation Via Epipolar Plane Image Analysis and Locally Linear Embedding," Zhang *et al.* analyze a scene from a large number of images by means of epipolar plane images (EPIs). From the estimated line slopes, accurate depth can be determined while local analysis and consideration of reliability leads to increased speed and increased robustness. In contrast to using many images for estimation of geometry, Qui *et al.* estimate scene depth from a single view as described in their paper "DEPT: Depth Estimation By Parameter Transfer With a Lightweight Model for Single Still Images." Exploiting the correlation between color and depth, they extract image features and search for similar views in a database. Their parameters are then transferred to synthesize new depth maps.

The next three papers address content creation and animation, in particular the reconstruction and editing of human body and face models. In their paper entitled "Video-Based Outdoor Human Reconstruction," Zhu *et al.* propose a method for the estimation of 3D human body models from a set of images in outdoor environments with little constraints on capture conditions. The structure-from-motion approach is extended by silhouette information and a fusion and refinement step that can also tolerate small body motion. Human body animation is targeted in "SPA: Sparse Photorealistic Animation Using a Single RGB-D Camera" by Li *et al.* They follow a sample-based approach in which an actor is captured by Kinect, forming a database of motion samples that can be used to synthesize new performances. Similarly, Paier *et al.* propose in "A Hybrid Approach for Facial Performance Analysis and Editing" Paier *et al.* propose a new method for facial editing, animation, and expression retargeting. Based on a model-free surface tracking approach, temporally consistent dynamic texture sequences are extracted, which can be recombined in order to synthesize novel facial performances.

The last four papers combine several algorithmic components for setting up entire AR systems and applications. In "An Integrated Platform for Live 3D Human Reconstruction and Motion Capturing" by Alexiadis *et al.*, a live system for 3D human reconstruction from multiple Kinect inputs is presented targeting realistic tele-immersion. It consists of geometry estimation and motion tracking, as well as color and texture fusion with a particular focus on sports environments. The communication of multiple people in a telepresence environment is also

Digital Object Identifier 10.1109/TCSVT.2017.2681378

the topic of the paper “A Mixed Reality Telepresence System for Collaborative Space Operation” by Fairchild *et al.* Free viewpoint video is combined with immersive projection technology for a collaborative mixed-reality environment. Due to the large data sizes, the transmission of 3D human point cloud representations requires efficient encoding and streaming. This is addressed in the paper “Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-immersive Video” by Mekuria *et al.* Octree-based intra coding is combined with the inter prediction of rigid sub-blocks for real-time encoding of point cloud sequences. Finally, in their paper“Magic Glasses: From 2D to 3D,” Yuan *et al.* present a virtual try-on system for glasses. Based on a sample image, a 3D model of a pair of glasses is estimated by classifying pixels, refining the outline, and extrapolating depth. Using Kinect for face tracking, the final model can then be watched in a typical AR mirror-like environment.

PETER EISERT
Fraunhofer HHI and Humboldt University
Berlin, Germany

YEBIN LIU
Tsinghua University
Beijing, China

KYUONG MU LEE
Seoul National University
Seoul, South Korea

DIDIER STRICKER
University Kaiserslautern and DFKI
Kaiserslautern, Germany

GRAHAM THOMAS
BBC R&D
Salford, U.K.

Peter Eisert received the Dipl.Ing. degree in electrical engineering from Karlsruhe Institute of Technology, Karlsruhe, Germany, and the Dr.-Ing. degree from University of Erlangen-Nuremberg, Bavaria, Germany.

In 2001, he was a Post-Doctoral Fellow with Stanford University, Stanford, CA, USA, where he was involved in 3D image analysis and synthesis and facial animation, and computer graphics. He joined Fraunhofer Heinrich-Hertz-Institut (Fraunhofer HHI), Berlin, Germany, in 2002 and Humboldt University of Berlin (HU Berlin), Berlin, in 2009, where he is currently coordinating and initiating numerous national and international research projects. He is currently a Professor of Visual Computing with HU Berlin and the Head of the Vision and Imaging Technologies Department, Fraunhofer HHI. He has authored over 150 conference and journal papers on the subject of 3D reconstruction, facial expression analysis and synthesis, and image-based rendering. His research interests include 3D image analysis and synthesis, face processing, image-based rendering, computer vision, computer graphics, and virtual and augmented reality.

Dr. Eisert is an Associate Editor of *International Journal of Image and Video Processing*. He is on the Editorial Board of *Journal of Visual Communication and Image Representation*.

Yebin Liu received the B.E. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2002 and the Ph.D. degree from the Automation Department, Tsinghua University, Beijing, China, in 2009.

He was a Research Fellow with the Computer Graphics Group, Max Planck Institute for Informatics, Saarbrücken, Germany, in 2010. He is currently an Associate Professor with Tsinghua University. He has authored over 50 papers, including for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, ACM SIGGRAPH CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, INTERNATIONAL CONFERENCE ON COMPUTER VISION, IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, and EUROPEAN CONFERENCE ON COMPUTER VISION. His research areas include computer vision, computer graphics, and computational photography.

Dr. Liu received Second Prize for the National Technical Invention Award in 2008, First Prize for the National Technical Invention Award in 2012, the Young Academic Research Fellow Award of Tsinghua University in 2013, and the NSFC Excellent Yong Research Fellow in 2015.

Kyoung Mu Lee received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Seoul, Korea in 1984 and 1986, respectively, and the Ph. D. degree in electrical engineering from University of Southern California, Los Angeles, CA, USA, in 1993.

He was a Distinguished Lecturer of the Asia–Pacific Signal and Information Processing Association from 2012 to 2013. He is currently with the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea, as a Professor. His primary research interests include scene understanding, object recognition, low-level vision, visual tracking, and visual navigation.

Dr. Lee received the Most Influential Paper over the Decade Award from the IAPR Machine Vision Application in 2009, the ACCV Honorable Mention Award in 2007, and the Okawa Foundation Research Grant Award in 2006. He also has served as the Program Chair of ACCV2012, the Track Chair of ICPR2012, and the Area Chair of CVPR, ICCV, and ECCV many times. He will serve as a General Co-Chair of ACM MM2018, ACCV2018, and ICCV2019. He is currently serving as an Associate Editor-in-Chief of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and an Area Editor of *Computer Vision and Image Understanding Journal*. He has served as an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *Machine Vision Applications Journal*, *IPSJ Transactions on Computer Vision and Applications*, IEEE SIGNAL PROCESSING LETTERS, and *EURASIP Journal on Applied Signal Processing*.

Didier Stricker received the engineering diploma in electronics and signal processing from INPG, Grenoble, France, in 1992 and the Ph.D. degree from University Darmstadt, Germany, in 2002.

He is a Professor with University of Kaiserslautern and Scientific Director with Deutsches Forschungsinstitut für Künstliche Intelligenz (DFKI), Kaiserslautern, where he leads the Augmented Vision Research Department. From 2002 to June 2008, he led the Virtual and Augmented Reality Department with Fraunhofer Institute for Computer Graphics (Fraunhofer IGD), Darmstadt, Germany.

Prof. Stricker received the Innovation Prize from the German Society of Computer Science in 2006. He serves as a Reviewer for different European or national research organizations. He is a Regular Reviewer for the most important journals and conferences in the areas of virtual reality/augmented reality and computer vision.

Graham Thomas received the degree in physics from University of Oxford, Oxford, U.K., and the Ph.D. degree in motion estimation in video.

He joined BBC in 1983. He is a Visiting Professor with University of Surrey, Guildford, U.K. He currently leads the Immersive and Interactive Content Section, Research and Development Department, BBC, developing technology for new forms of content, with a focus on computer vision, graphics, and image processing. His work has led to many award-winning commercial products, including the Alchemist broadcast standards converter, the Piero sports graphics system, and the free-d camera tracking system for virtual studios. He recently led the EU REACT Project on multi-camera 3D capture of performers and the U.K. REFRAME Project on enhanced content production. His team is currently focused on video standards beyond high definition (higher frame rate and high dynamic range), interactive and augmented video, and content creation for 360 video, augmented reality and virtual reality applications. His team also covers audio work, including 3D and object-based audio.

Dr. Thomas is a fellow of the IET.