# Guest Editorial
# Immersive Video Coding and Transmission

## I. Introduction

Immersive media are gaining in popularity today, and significant efforts are being undertaken in academia and industry to explore its immanent new scientific and technological challenges. There are significant activities in industry and standardization to provide enablers for production, coding, transmission, and consumption of this type of media and the new user experiences it enables. In terms of standardization, the topic has triggered multiple activities in the areas of systems, 3D graphics, audio, image, and video. The technological roadmap foresees an evolution from consumption of the visual media with three degrees of freedom (3DoF), i.e. the ability to look around at a fixed viewing position in an observed scene) to 3DoF+, enabling further limited modifications of the viewing position, and to different variants of six degrees of freedom (6DoF), allowing the user not only to look around but also to move around in the observed scene. Different terminology is used in various communities, referring to immersive or omnidirectional media, virtual reality (VR), or 360° video. While the coded representation of audio-visual media for 6DoF remains a field of active research, 3DoF technologies have sufficient maturity to progress towards specification in near-term standards and recommendations. At the video codec level, this includes coding of 2D and 3D virtual reality (VR) / 360° content using the HEVC standard (Rec. ITU-T H.265 | ISO/IEC 23008-2) as the initial step, with new Supplemental Enhancement Information (SEI) messages for omnidirectional video. For storage and delivery, the Omnidirectional Media Format (OMAF) (ISO/IEC 23090-2) provides a set of consistent enablers for download and over-the-top streaming of 3DoF content. At the same time, immersive still image coding formats are currently developed in JPEG Pleno.

With respect to video, ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11) are currently developing the new video coding standard Versatile Video Coding (VVC) with an expected compression capability that significantly exceeds that of the HEVC standard and its current extensions. The scope of this joint activity includes consideration of a variety of video sources and video applications, including camera-view content, screen content, consumer generated content, high dynamic range content, and also explicitly virtual reality/360° content. The standardization timeline foresees the finalization of the specification by the end of 2020.

On the systems side, and beyond the initial 3DoF activities, the new MPEG project on immersive media (referred to as MPEG-I) attempts to provide a more consistent view on immersive media and support new experiences in the mid and long-term. Based on use cases beyond 3DoF and architectural considerations, audio, video and 3D graphics aspects are evaluated, including new representation formats such as point clouds and light fields. Orthogonal aspects such as improved delivery, consistently reported metrics as well as quality evaluation of immersive media are also in scope.

Also outside of MPEG, the first set of enabling specifications for immersive media with 3DoF are cornerstones for evolving systems and related standardization and interoperability activities. Among others, 3GPP recognizes the value of MPEG 3DoF technologies for the work on 5G VR Streaming which has been released in October 2018. The VR Industry Forum promotes the MPEG enablers for full end-to-end operability, combining them with production, security, distribution and rendering centric activities. On the latter, in particular the Khronos/OpenXR and the W3C WebVR groups target to provide interoperability for platform APIs.

This Special Issue aims at a capture of the status of this emerging technology with respect to latest scientific progress, to corresponding standardization efforts, to subjective assessment of immersive media and also with respect to the impact of this technology on regular users.

## II. Organization and Overview

This special issue accepted 17 papers in several key areas, namely (1) streaming architecture and systems design for immersive media, (2) compression of omnidirectional video (3DoF), (3) compression of volumetric video (6DoF), and (4) human interaction and perceived quality.

We begin with an overview paper, “Standardization Status of Immersive Video Coding,” by the Guest Editors. The paper strives at providing a comprehensive overview on the standardization activities in the field of immersive video coding and transmission. It presents the components of the emerging MPEG-I standards suite from the general architecture, to systems, omnidirectional video, Versatile Video Coding, point cloud compression, and schemes for 3DoF+ and 6DoF video compression. The selected and accept papers in this special issue partially pick up and deepen on technology aspects raised in this overview while others present interesting alternative approaches or important related or complementary aspects.

### *A. Streaming Architectures and Systems Design*

Four papers are grouped around architectures for streaming 360°, i.e. omnidirectional, video. The paper “Delay Impact on MPEG OMAF's Tile-Based Viewport-Dependent 360 Video Streaming” by Sanchez De La Fuente et al. presents the approach of tile-based streaming which is pursued in the MPEG OMAF specification in comparison to streaming of viewport-

specific projections. The paper deliberates on the trade-off between end-to-end delay and visual quality of these schemes. The following paper "An Optimal Tile-based Approach for Viewport-adaptive 360-degree Video Streaming" by Nguyen *et al.* takes up on this line. It proposes an adaptation approach for tile-based view-port adaptive streaming, dedicatedly taking into account viewport estimation errors and head movements and discusses the impact of segment duration and buffer sizes on the system performance. A different approach is pursued in the next paper "A Two-Tier System for On-Demand Streaming of 360 Degree Video over Dynamic Networks" by Sun *et al.* Here, two components are streamed, including a lower quality version of the full 360° scene and a high quality version of the viewport area which is currently active. The last paper in this category, "OmniCast: Wireless Pseudo-Analog Transmission for Omnidirectional Video" by Xiong *et al.*, proposes a different transmission concept for omnidirectional video, which reportedly provides graceful degradation under changing wireless channel conditions at low complexity and low latency.

### *B. Compression of Omnidirectional Video*

Five papers are arranged under the topic of omnidirectional video compression. The first two papers, "360-degree Video Coding based on Projection Format Adaptation and Spherical Neighboring Relationship" by Hanhart *et al.* and "Efficient Projection and Coding Tools for 360° Video" by Lee *et al.* present technology which has been submitted as proposals for standardization of projections formats and compression tools for omnidirectional video in the emerging video coding standard VVC. They detail the proposed concepts and analyze the performance impact of their respective design choices in comparison to state-of-the-art video coding. In the next paper, "Spherical Coordinates Transform-Based Motion Model for Panoramic Video Coding", Liu *et al.* propose a solution to encompass geometrical distortions in the projected 2D video when the original 3D motion is captured. The paper extends the applicable motion model to take these effects into account. The following paper "Intra-prediction Mode Propagation for Video Coding" by Zhang *et al.* specializes on advanced intra prediction for enhanced compression efficiency in a versatile video coding scheme. The paper "An Assigned Color Depth Packing Method with Centralized Texture Depth Packing Formats for 3D VR Broadcasting Services" by Yang *et al.* focusses on the transmission of 3D video which is represented by texture and depth information. The proposed scheme has been adopted for distribution of 3D VR services in Taiwan.

### *C. Compression of Volumetric Video and Light Fields*

With the subsequent set of papers, the focus is turned from omnidirectional video from a fixed viewpoint (3DoF) to schemes allowing for the modification of the viewing position (3DoF+ and 6DoF). The first paper summarizes the ongoing standardization development in this field for volumetric data which is represented by point clouds. Under the title "Emerging MPEG Standards for Point Cloud Compression," Schwarz *et al.* detail the two technological solutions under development. Video-based point cloud compression (V-PCC) relies on existing video compression technology applied to projected scenes while geometry-based point cloud compression (G-PCC) directly addresses the efficient compression of sparse point clouds. In the next paper, "Rate-Utility Optimized Streaming of Volumetric Media for Augmented Reality" by Park *et al.*, the aspect of data partitioning is taken up. The authors propose a scheme using 3D-tiles and rate-utility-based bit-allocation to optimize the user experience for this data. The following paper "Surface Light Field Compression using a Point Cloud Codec" by Zhang *et al.* proposes to represent the object surfaces by view maps, capturing the color of the points from a given viewing direction. The coefficients of the B-spline wavelets which are used to represent the view maps are compressed by PCC algorithms. The paper "Light Field Image Compression Using Generative Adversarial Network Based View Synthesis" by Jia *et al.* presents a scheme for compression of light field images using a learned generation of sub-aperture images to provide enhanced compression efficiency. "Benchmarking Open-Source Static 3D Mesh Codecs for Live Immersive Media Streaming" by Doumanoglou *et al.* is the final paper in this category. The authors compare open-source 3D intra coding schemes under the perspective of suitability for tele-immersive applications.

### *D. Human Interaction and Perceived Quality*

Three papers are grouped in the category on human interaction and perceived quality. As the consumption of immersive media is strongly affected by a match of the system reaction with the users expectation with respect to latency, consistency, and visual quality, these aspects play a crucial role with regard to the acceptance of the technology by the user. The first paper, "Prediction of the Influence of Navigation Scan-path on Perceived Quality of Free-Viewpoint Videos" by Ling *et al.* propose the concept of hypothetical rendering trajectories to develop a reliable video quality assessment measure for the evaluation of free-viewpoint video. The authors base this development on results of subjective viewing tests. The following paper "Visual Attention-Aware Omnidirectional Video Streaming Using Optimal Tiles for Virtual Reality" by Ozcinar *et al.* proposes a quality metric for omnidirectional video which is based on visual attention maps. In a second step, a method using visual attention maps for rate allocation for tile-based coding of omnidirectional video is presented. Last but not least, the paper "Real-time 3D 360-Degree Telepresence with Deep Learning-Based Head Motion Prediction" by Aykut *et al.* presents a delay-compensating 3D 360° vision system. The core aspect of the scheme is a reliable prediction of user head motion in order to cope with latency issues at the receiver side.

## III. Paper Selection Process and Acknowledgement

This special issue has a highly selective review process, in order to comply with the high-quality standards of the IEEE Journal on Emerging and Selected Topics in Circuits and Systems. In total, we received 32 submissions, of which 17 were finally accepted. All accepted papers went through two revisions to meet the quality requirements. The review of each paper involved at least 3 independent reviewers, with criteria emphasizing equally on quality and fitness to the scope of this issue.

The Guest Editors would like to thank everyone who made submissions to this issue and committed much time to revisions to meet our quality expectations. We are also deeply indebted

to anonymous reviewers for their time and effort to help us make difficult decisions, especially because they had to work under high pressure to ensure the timely publication of this issue.

Last but not least, we are grateful to the Editor-in-Chief, Prof. Eduard Alarcon, the Deputy Editor-in-Chief, Prof. Massimo Alioto, as well as the Senior Editorial Board of this journal for their support and advice throughout the editing process.

We hope you enjoy delving into the contents of this special issue and will follow up with contributing more to advance this wide and evolving field.

MATHIAS WIEN, *Guest Editor*
Lehrstuhl für Bildverarbeitung
RWTH Aachen University
Germany

JILL BOYCE, *Guest Editor*
Intel Corp.
USA

THOMAS STOCKHAMMER, *Guest Editor*
Qualcomm Incorporated
USA

WEN-HSIAO PENG, *Guest Editor*
Department of Computer Science
National Chiao Tung University
Taiwan