Standardization Status of Immersive Video Coding

Mathias Wien, *Member, IEEE*, Jill M. Boyce, *Fellow, IEEE*, Thomas Stockhammer, *Senior Member, IEEE*, and Wen-Hsiao Peng, *Senior Member, IEEE*

***Abstract*— Based on increasing availability of capture and display devices dedicated to immersive media, coding, and transmission of these media has recently become a highest-priority subject of standardization. Different levels of immersiveness are defined with respect to an increasing degree of freedom in terms of movements of the observer within the immersive media scene. The level ranges from three degrees of freedom allowing the user to look around in all directions from a fixed point of view to six degrees of freedom, where the user can freely alter the viewpoint within the immersive media scene. The moving pictures experts group (MPEG) of ISO/IEC is developing a standards suite on "Coded Representation of Immersive Media," called MPEG-I, to provide technical solutions for building blocks of the media transmission chain, ranging from architecture, systems tools, coding of video and audio signals, to point clouds and timed text. In this paper, an overview on recent and ongoing standardization efforts in this area is presented. While some specifications, such as high efficiency video coding or version 1 of the omnidirectional media format, are already available, other activities are under development or in the exploration phase. This paper addresses the status of these efforts with a focus on video signals, indicates the development timelines, summarizes the main technical details, and provides pointers to further points of reference.**

***Index Terms*— Joint video experts team, MPEG, omnidirectional video, standardization, versatile video coding.**

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 for new user experiences.

In terms of standardization, the topic has triggered multiple activities in the areas of systems tools, 3D graphics support, as well as coded representation of immersive audio, image, and video signals. In ISO/IEC MPEG, a new project was launched, referred to as "Coded Representation of Immersive Media", and abbreviated as MPEG-I [1]. At the stage of writing this paper, it comprises 8 parts, including specifications for architectures, systems, video, audio, point clouds, as well as metrics, metadata and interfaces for network-based processing of immersive media content [2]. The technological roadmap foresees an evolution from consumption of the visual media with three degrees of freedom (3DoF), the ability to change orientation to look around at a fixed viewing position in an observed scene, i.e. 360° video, to 3DoF+ which enables limited modifications of the viewing position, to different variants of six degrees of freedom (6DoF), allowing the user not only to change orientation but also to change position to move around in the observed scene.

While the coded representation of audio-visual media for 6DoF is a field of very active research, 3DoF technologies have reached sufficient maturity for specification in near-term standards and recommendations. In terms of existing standards for the video coding 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) [3] as the initial step, which is complemented with new Supplemental Enhancement Information (SEI) messages for handling of omnidirectional video. In order to be able to improve the compression efficiency specifically for very high resolution 2D and 3D virtual reality (VR)/360° content, the Joint Video Experts Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11) has just started the standardization activity for Versatile Video Coding (VCC), to be published as Rec. ITU-T H.266 | ISO/IEC 23090-3. 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. A Joint Call for Proposals on Video Compression with Capability beyond HEVC was evaluated in April 2018, leading to a first VVC Working Draft already at that event [4], with updates provided at each subsequent meeting in [5]–[7].[1] The standardization timeline foresees the finalization of the specification by the end of 2020.

For audio, the already published ISO/IEC 23008-3 MPEG-H Audio standard provides all enablers for 3D audio including channel, object and scene-based representations as well as 3D rendering. The emerging MPEG-I part 4 Immersive Audio standard, to be published as ISO/IEC 23090-4 [8], will extend the capabilities of audio rendering to 6DoF.

Manuscript received October 3, 2018; accepted January 29, 2019. Date of publication February 13, 2019; date of current version March 11, 2019. This paper was recommended by Guest Editor E. Alarcon. *(Corresponding author: Mathias Wien.)*

M. Wien is with the Institute for Imaging and Computer Vision, RWTH Aachen University, 52056 Aachen, Germany (e-mail: wien@lfb.rwth-aachen.de).

J. M. Boyce is with Intel, Hillsboro, OR 97225 USA (e-mail: jill.boyce@intel.com).

T. Stockhammer is with Qualcomm, 81671 Munich, Germany (e-mail: tsto@qti.qualcomm.com).

W.-H. Peng is with the Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan (e-mail: wpeng@cs.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JETCAS.2019.2898948

[1] JVET documents are publicly available at http://phenix.int-evry.fr/jvet/

Last but not least, the Omnidirectional MediA Format (OMAF) [9] provides a set of tools for storage and delivery of 3DoF content. In the present version 1 of OMAF, a selected set of projection formats is supported enabling transmission of monoscopic and stereoscopic videos and the corresponding audio, including the required media encapsulation and signaling for streaming and over-the-top (OTT) delivery. With these specifications at hand, the foundations for coding, transmission, and presentation of 3DoF audio-visual are readily available.

At the same time, immersive still image coding formats are currently under development in JPEG Pleno. JPEG Pleno intends to provide a standardized framework to facilitate the capture, representation, and exchange of omnidirectional, depth-enhanced, point-cloud, light-field, and holographic imaging modalities [10]. JPEG Pleno is going to be published as ISO/IEC 21794, with four parts being defined as work items at the time of writing this paper[2]: Part 1 - Framework, targeted for completion by the end of 2019; Part 2 - Light Field Coding, targeted for completion by the end of 2019; Part 3 - Conformance Testing, targeted for completion by the end of 2021; Part 4 - Reference software, targeted for completion by the end of 2021.

Industrial forums and consortia provide supplemental enabling specifications for immersive media with 3DoF that are cornerstones for evolving systems and related standardization and interoperability activities. Among others, 3GPP recognizes the value of MPEG 3DoF technologies: the first release of 5G includes streaming enablers based on OMAF, but with a very clear focus on interoperability based on existing mobile platforms. 3GPP TS26.118 [11] defines three video media profiles: (i) a simple legacy version based on H.264/AVC, (ii) a version that permits streaming and download based on existing DASH and file format clients, and (iii) an advanced profile that permits decoding on existing chipsets, but requires innovative processing and streaming technologies to enable viewport-adaptive streaming. Audio is based on MPEG-H audio and follows the recommendation from MPEG-I part 2 OMAF. The VR Industry Forum promotes the MPEG specifications for full end-to-end operability, combining them with production, security, distribution and rendering centric activities. On the latter, in particular the work in Khronos/OpenXR and W3C WebVR groups is targeted at providing interoperability for platform APIs [13].

This paper focusses on the status of standardization with respect to coding and transmission of immersive video signals, including an overview of already adopted specifications and a summary of the larger picture of the specifications to come in this field. The paper is structured as follows. In Section II, an overall system overview for the delivery of immersive audio-visual signals is provided and the terminology in the context of immersive audio-visual media is set. Section III outlines the technology for transmission of omnidirectional video based on HEVC. In Section IV, the current status and the perspectives of the emerging Versatile Video Coding standard are laid out. Sections V and VI provide a summary of the developments for point cloud coding and light field coding, respectively, which are enabling technologies for extended degrees of freedom in immersive media consumption. Section VII concludes the paper.

[2]See the ISO/IEC Secretariat 29 Programme of Work for Working Group 1 (JPEG), https://www.itscj.ipsj.or.jp/sc29/29w42901.htm#JpegPLENO.

II. Omnidirectional 3D Media Architecture

A. Degrees of Freedom

For the consumption of omnidirectional image and video signals and accompanying 3D audio, it is assumed that the user is able to observe a surrounding scene, e.g., by presentation using a head-mounted display (HMD). The portion of the scene which is observed by the user is referred to as a *viewport*, representing the region of the scene which is covered by the user's field of view (which may be limited by the given viewing device). A monoscopic presentation of the viewport presents the same view to both eyes of the user while a stereoscopic presentation additionally allows for a 3-dimensional (3D) impression of the scene based on depth cues available from the pictures presented to the two eyes. The degrees of freedom in the context of immersive media consumption are illustrated in Fig. 1.

This ability to look around and listen from a centre point in 3D space is defined as 3 degrees of freedom (*3DoF*). According to Fig. 1:

- tilting side to side on the x-axis is referred to as *Roll*ing,
- tilting forward and backward on the y-axis is referred to as *Pitch*ing,
- turning left and right on the z-axis is referred to as *Yaw*ing.

Yaw and pitch movements can be represented by the shift of a point on the sphere by the *azimuth* angle Φ_d and the *elevation* angle Θ_d, as illustrated in Fig. 1a. The presentation of an omnidirectional image or video and 3D audio allows the user to freely alter Φ_d and Θ_d and thereby assess the surrounding virtual scenery without changing the location of the view point.

In MPEG-I terminology, the term 3DoF refers to omnidirectional media which allow for rotational and un-limited movements around the x, y and z axes (respectively roll, pitch, and yaw), creating a 3D impression of the observed scene (see Fig. 1b). Note that in the context of 3D presentation, the roll head movement may impose a specific challenge for most conventionally captured omnidirectional video as in many cases, capture is realized in a stereoscopic fashion assuming a strictly horizontal disparity between the views [14]. As a next step, the term 3DoF+ refers to the case where the system additionally allows for (restricted) relocation of the view point, e.g., by limited translational movements of the user's head around the original view point (Fig. 1c).

The *6DoF* scenario is split into three categories, representing media with increasing flexibility features. *Windowed 6DoF* (Fig. 1d) describes the case with constrained rotational movements around the x and y axes and additionally constrained movements along the z axis. This could be illustrated as the case that the user would observe the scene by watching through a window. *Omnidirectional 6DoF* (Fig. 1e) further

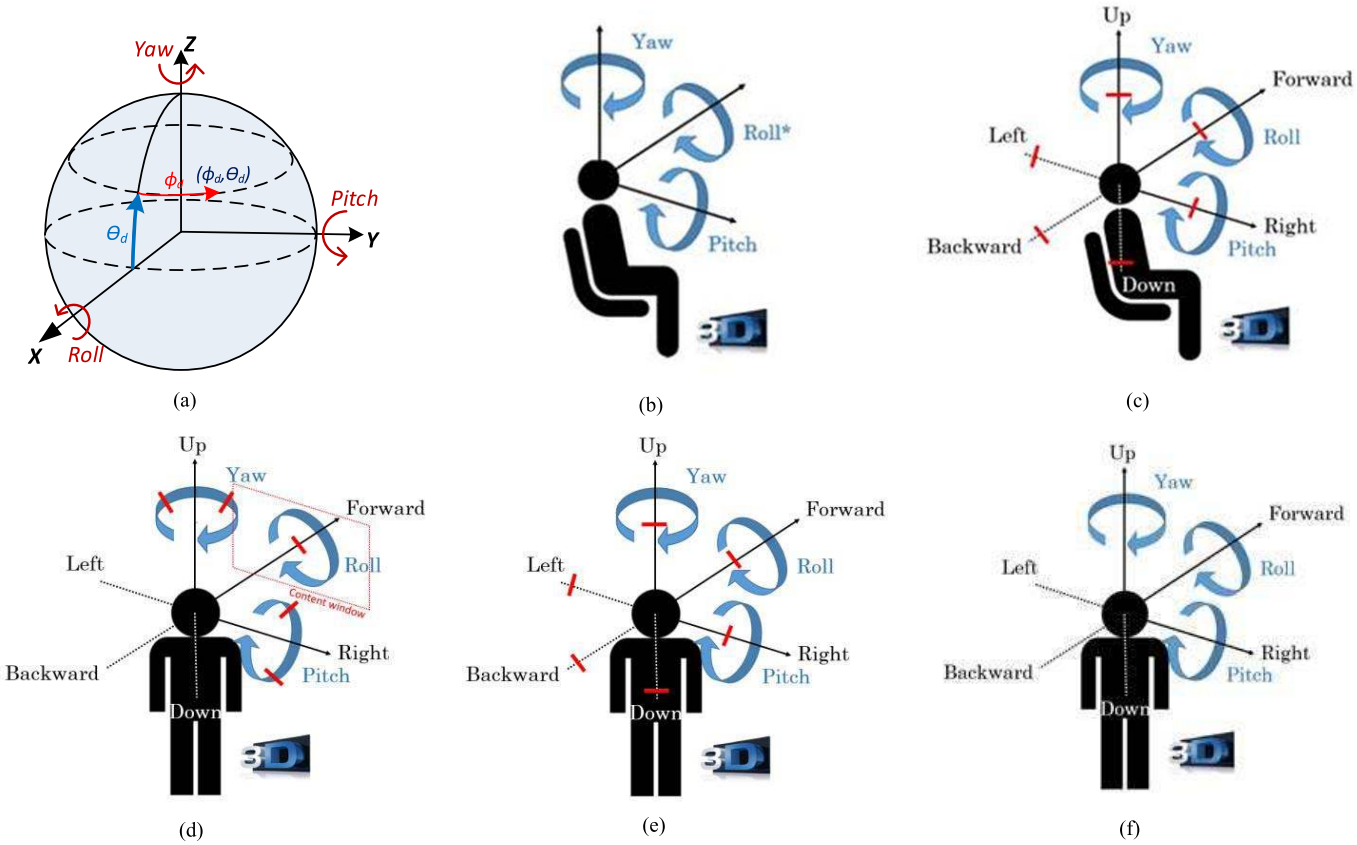

Fig. 1. Illustration of the degrees of freedom as used in MPEG-I [2]. a) Definition of degrees of freedom [9]. b) 3DoF. c) 3DoF+. d) Windowed 6DoF. e) Omnidirectional 6DoF. f) 6DoF.

extends the freedom to enabling the user to freely move around in a scene within a restricted range, and finally, unconstrained *6DoF* specified the cases with full translational movement along the x, y and z axes.

B. MPEG-I: Coded Representation of Immersive Media

At the end of 2016, MPEG initiated a survey along its members and to outside industry on their perspective on use cases, requirements, technologies, timelines and finally standardization needs in the context of virtual reality and immersive media. Based on the results of this survey as well as dispersed MPEG internal activities around virtual reality, MPEG streamlined the activities and initiated a new project on "Coded representation of immersive media", referred to as MPEG-I.

The primary focus of this project is the enabling of coding, transmission, and presentation of immersive media. Several key aspects were identified to be addressed - ranging from the architectures, systems technologies, video and audio. The ISO/IEC 23090 MPEG-I standards collection is currently developed to serve this purpose.

At the time of writing this paper, the whole area is in a phase of significant development. In order to cope with the urgent demand for standardized solutions from industry, a multi-phase approach is pursued towards completion of this standards collection. In the first phase, which was completed end of 2017, streaming of 360° video (i.e., 3DoF) with 3D audio is enabled, based on existing MPEG compression technologies. The first version of the Omnidirectional MediA Format (OMAF) adds rendering centric features to the existing audio and video specifications (primarily MPEG-H 3D-Audio and HEVC) and enables storage based on the ISO Base Media File Format and DASH-based streaming. MPEG is currently working on an extension of OMAF that is expected to include enablers for overlay, interactivity as well as a 3DoF+ extension with a finalization target in 2020. In a second track, full support for 6DoF media is developed. The timeline foresees this project to last at least until about the year 2022. This project includes Point Cloud coding, rendering centric interactive 6DoF, as well as natural 6DoF representation of content such as light fields. Whereas some aspects have already progressed quite far, for others early discussions on use cases, sample content and market matureness are necessary in order to add more details to the work and time plan.

At this stage, the foreseen parts of the standards collection are briefly presented:

- **Part 1 - Immersive Media Architecture (23090-1)** [2]. This part sets terminology and definitions to be available for the following parts of the standards collection. It further includes a general system overview, use-case descriptions, as well as considerations regarding relevant media quality metrics, including aspects of audio and video signal quality, resolution, latency, and impression

consistency. At the time of writing, this technical report is at the Committee Draft level in MPEG. It is expected to be further adapted and updated by the time phase 2 of MPEG-I standardization is started.

- **Part 2 - Omnidirectional Media Format (OMAF, 23090-2)** [9]. This part specifies the omnidirectional media format for coding, storage, delivery, and rendering of omnidirectional media. It provides the systems interface for the encapsulated media, including video, images, audio, and timed text. The first edition of OMAF has been completed and the work on a second edition of OMAF towards the support of 3DoF+ is in progress. OMAF is further described in Section III.
- **Part 3 - Versatile Video Coding (VVC, 23090-3)** [7]. This part represents the latest progression in standardized visual coding technologies. It has been launched as a joint project of ISO/IEC JTC1 and ITU-T Q6/16 to support the demand for higher efficiency video compression capabilities and improved functionality, specifically with respect to coding of ultra-high resolution video as required in the context of 360° video transmission. The contents, features, and development status of VVC are detailed in Section IV.
- **Part 4 - Immersive Audio (23090-4)**. The goal of this part is to provide audio coding for the 6DoF scenario [2]. The work on this technology in MPEG is in the exploration phase with a call for proposals targeted for the end of 2019 [15], 16].
- **Part 5 – Video-Based Point Cloud Compression (V-PCC, 23090-5)**. This part addresses compression of 3D visual media represented by point clouds sequences captured over time with support of random access to subsets of the point cloud. The contents, features, and development status of V-PCC are further detailed in Section V.
- **Part 6 - Immersive Media Metrics (23090-6)**. This part specifies immersive media metrics and a measurement framework for evaluating the immersive media quality and experience. It also includes a client reference model for collection of the metrics which defines observation points ranging from network access to media playback [17]. Completion of this part is planned for mid-2020.
- **Part 7 - Immersive Media Metadata (23090-7)**. This part specifies immersive media metadata that can be consistently used in different application and system environments. For image and video media, the metadata includes definition of coordinate systems, projection formats, texture-to-sphere mappings, coverage definitions, or rotation parameters [18]. Completion of this part is planned for mid-2020.
- **Part 8 - Network-Based Media Processing (NBMP, 23090-8)**. This part is supposed to specify the workflow for upload of media data to network-based processing entities, the processing operations, and access to the processed media data and corresponding metadata, including optional real-time access [19]. Completion of this part is planned for the end of 2019.
- **Part 9 - Geometry-Based Point Cloud Compression (G-PCC, 23090-9)**. This part addresses efficient compression of sparse point clouds. The contents, features, and development status of G-PCC are further detailed in Section V.

As future extensions of MPEG-I, technologies for 3DoF+ and 6DoF are studied in MPEG exploration activities. The exploration is denoted as MPEG-I Visual, and includes light-field coding. A Call for Proposals for 3DoF+ video was issued in January 2019 [20], with responses due in March 2019. It is expected that the MPEG-I solution for the coding of 3DoF+ video will be built on the existing HEVC standard for video and depth information while 3DoF+ metadata will be standardized in MPEG-I part 7 and referenced at the systems level in OMAF. The status of this activity is reported in Section VI.

C. System Overview

A system overview for the immersive media transmission chain is presented in Fig. 2. It presents the required building blocks for the realization of 3DoF audio and video consumption. This architecture is the baseline reference architecture for OMAF. For extension towards 3DoF+ and 6DoF scenarios, the general structure is expected to be preserved, based on enhanced or correspondingly extended building blocks. However, especially for 6DOF, it is expected that rendering will play a more central role in the media consumption. For more details, refer to Section VII.

We will focus on the aspects concerned with visual media by the example of video in the following.

The processing chain starts with the acquisition of the 3DoF media. Commonly, multiple cameras are required to capture the 360° scene around the observer. Therefore, an image stitching, projection and mapping step has to be performed. The output of this building block is a 2D representation of the original 360° scene which is then fed to the video encoder. The encoding process of the video has to take into account aspects of local random access on top of established bitstream properties in order to enable viewport dependent selective access to the portions of the encoded video stream representing the currently 'active' portion of the scene which is surrounding the current viewport. Accordingly, the video must be partitioned into segments representing local areas of the scene. The video bitstream further needs to be segmented into chunks along the time axis for transmission.

Generally, the video data is segmented and encapsulated according to the transmission requirements of the given application scenario. This may imply, for example, direct file playback or streaming of the media from a remote server. The portions of the complete encoded bitstream which are to be delivered to the receiver side are controlled by the head- and eye-tracking building block at the receiver side. This block determines and specifies the currently required viewport parameters which are sent as metadata to the delivery building block.

As can be seen from Fig. 2, the orientation and viewport metadata are also required as input to all building blocks at the receiver side. The processing starts with the decapsulation

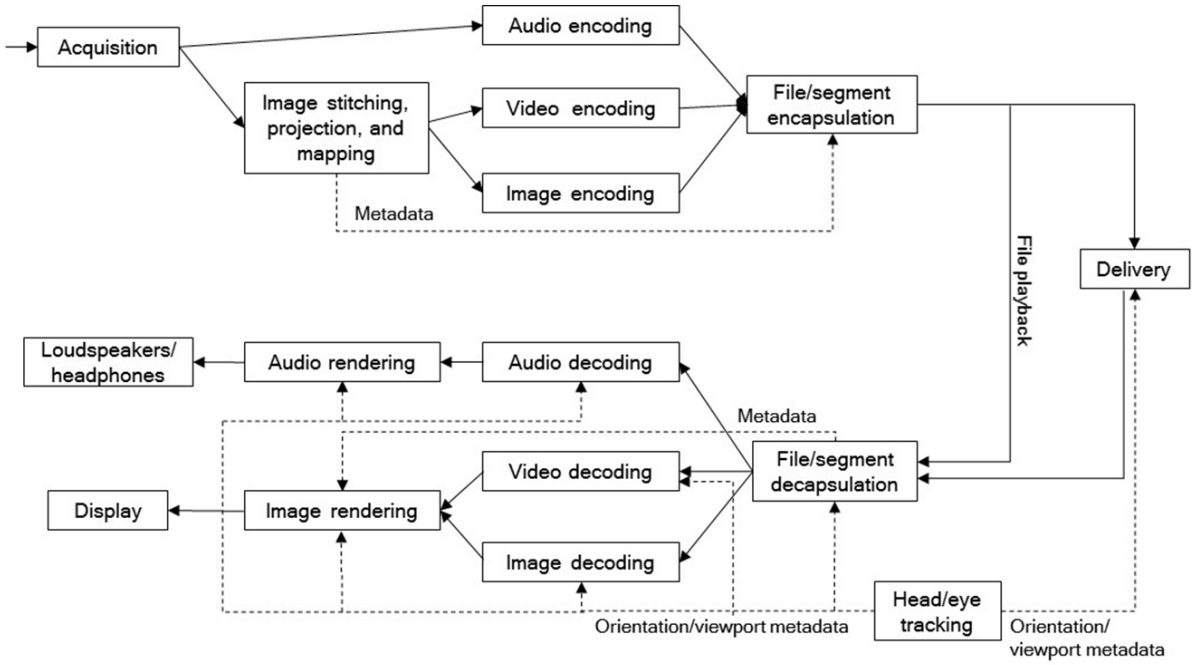

Fig. 2. MPEG-I architecture block diagram for 3DoF media content [2].

of the received video bitstream, extracting the video data itself and the associated metadata which describe the features of the encapsulated media. The video data and all required metadata are then fed to the decoding stage. The decoder passes the reconstructed video to the renderer which processes the media according to the needs of the given playout device, e.g., a head-mounted display.

D. Rendering and Display

In contrast to conventional images or video, in the case of immersive media with 3DoF and beyond, the signal is not coded in a format ready for display. Rendering the encoded media into a format tailored to the target display system is required. Depending on the head movements and the directions that the user looks at, a portion of the observed scene is to be presented. The *field of view* (FOV) which can be covered by head-mounted display devices for this task is in the range of up to 100° horizontally and up to 113° vertically [2]. The 'path' which is chosen by the user by looking around during the consumption of the immersive media is called a *viewport* of the media. The viewport generally depends on the user's intentions and is determined by tracking their movements. For testing purposes, defined viewports may be generated which then can be displayed on conventional screens. Thereby it is possible to expose multiple observers to a consistent viewport of the media when conducting a quality assessment experiment.

The quality of experience during consumption is largely impacted by the available media and display resolutions, the display frame rate, and the latency of the system response to user interaction. In [2], a resolution of 40 pixels per degree at a frame rate of 90 Hz and a maximum latency below 20 ms is considered to provide high quality immersive experience. A further aspect to be considered is the appropriate handling of stitching errors in the composition of the 360° scene. In the case of 3DoF+ and beyond, a consistent parallax between the foreground and the background is required.

In HMDs, there are specific display problems to be addressed [22]. These devices require dedicated optics to present the rendered viewport to the eyes of the observer at a very short distance. The required wide-angle biconvex lenses create barrel distortion which has to be compensated. Another aspect is chromatic aberration: after passing through the lens, colors are focused at different positions in the focal plane. Image processing within the HMD is required to compensate for these artifacts.

III. Omnidirectional Video Transmission Using HEVC

Studies within the Joint Collaborative Team on Video Coding (JCT-VC) and JVET have shown that HEVC can be used to carry 360° video without modification of the core codec, by first mapping the omnidirectional spherical representation to a rectangular picture using projection mapping. It is necessary to provide metadata with the compressed bitstream to describe the projection mapping used. Projection mapping metadata can be carried at the elementary stream layer, using SEI messages, and/or at the systems layer using OMAF [9]. It was considered necessary to standardize both elementary stream and systems layer mechanisms to provide flexibility for a wide variety of use cases.

A. Projection Formats

At the elementary stream layer, JCT-VC developed SEI messages to define projection formats for equirectangular

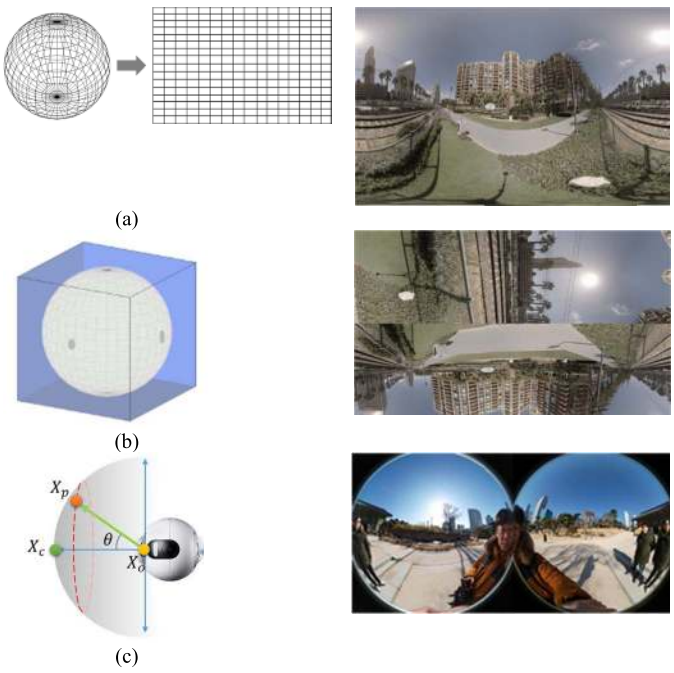

Fig. 3. Projection formats supported in OMAF version 1. (a) Equirectangular projection [81]. (b) Cubemap projection [81]. (c) Fisheye projection [9].

projection and cube map projection, which were standardized in HEVC version 5 [3]. An SEI message to define a fisheye projection format is under development, and planned for inclusion in a new HEVC version. A draft version of the fisheye SEI message is available in [23]. At the systems layer, OMAF equivalently includes support for equirectangular projection, cube map projection, and fisheye projection formats. The three projection formats are illustrated in Fig. 3.

For development and testing purposes, JVET developed a software package called 360Lib [24] to perform conversions between projection formats and to calculate objective video quality metrics consistently across the projection formats under consideration. The JVET output document JVET-G1003 provides a description of the formats implemented in the 360Lib software [25]. It is noted that the software package includes more formats than are standardized for HEVC. A comparison of the compression performance impact of the projection formats implemented in 360Lib can be found in [26].

The experimental conditions and evaluation procedures to be used in the context of standardization for 360° video is described in a common test conditions document JVET-H1030 [27]. These were defined by JVET to compare the coding efficiency of different projection formats. Reference [28] provides an overview of how the JVET group performed the comparisons.

B. Omnidirectional Media Format

The Omnidirectional MediA Format (OMAF), was developed to provide a standardized format for omnidirectional media applications. Compared to traditional media application formats, the end-to-end technology for omnidirectional video (from capture to playback) is more easily fragmented due to various capturing and video projection technologies. From the capture side, there exist many different types of cameras capable of capturing 360° video, and on the playback side there are many different devices that are able to playback 360° video with different processing capabilities. Version 1 of the OMAF specification has been technically frozen by the end of 2017.

OMAF specifies an architecture as sown in Fig. 2 for omnidirectional media with projected video as shown in Fig. 3. The media are encapsulated in the ISO Media Base File Format for which specific constraints and requirements are specified, which are grouped in presentation profiles. The media themselves are encoded using respective standards such as HEVC as described above. Constraints and requirements on the encapsulated media are expressed by the specification of the applicable media profiles in OMAF. In this paper, we only focus on aspects of the OMAF video profiles.

The specification supports both monoscopic video and stereoscopic video. As mentioned before, OMAF supports the equirectangular projection, cube map projection, and fisheye projection formats which are also specified in the corresponding HEVC SEI messages. Thereby, information on the projection format is consistently available at the elementary bitstream level as well as the systems level.

The core elements of OMAF include the specification of a coordinate system for omnidirectional media, the precise definition of a conversion process between sets of coordinate axes of this coordinate system as well as the specification of the projection between the 360° sphere and the projected 2D representation of the video. Region-wise packing formats are defined to allow for region-wise re-arrangement of the projected video to be encoded, including rescaling and the introduction of guard bands around the regions in order to prevent the introduction of cross-region artifacts. Thereby, regions which are in focus of a current viewport may be transmitted with a higher resolution than the remaining regions.

OMAF video profiles exist in two variants, which are either viewport-independent or viewport-dependent. The viewport-independent variant is the more general case, where no information of a viewport is considered and the video signal is packed in a pre-defined way. Viewport-dependent video profiles allow for a flexible arrangement of the packed regions, exploiting adaptive rescaling to optimize the transmission rate. An example application of a viewport-dependent video profile is illustrated in Fig. 4. By splitting the encoded picture into tiles with restrictions on the applicable motion vectors not to point outside of the tile region across reference pictures, the picture region can be decomposed and recomposed with tile set of different quality and different resolution in HEVC bitstreams. Such tile sets are referred to as motion-constrained tile sets (MCTS).

The applicable mapping of the presentation and media profiles to transport, such as DASH, are provided as Annexes to the specification.

As mentioned before, at the time of writing, version 1 of OMAF has been finalized, including the features described above. Extensions of OMAF are planned to encompass the provision of 3DoF+ and 6DoF media streams [16].

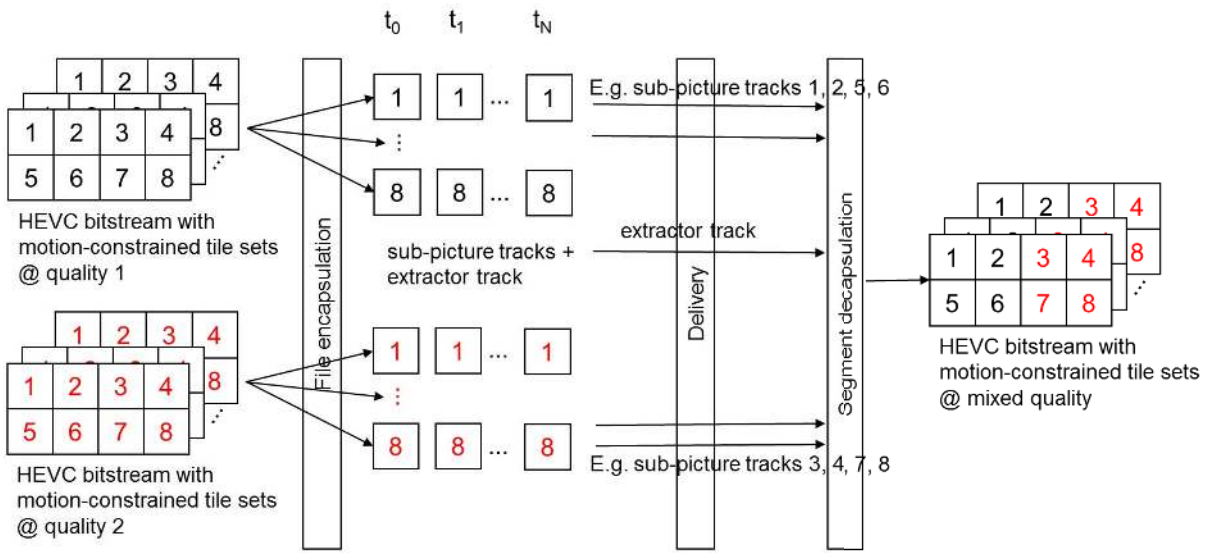

Fig. 4. Example of merging of HEVC MCTS-based sub-picture tracks of the same resolution [9].

IV. Versatile Video Coding

Versatile Video Coding (VVC) is the name of the latest video coding standardization project of ITU-T and ISO/IEC. The focus of the project specifically includes very high resolution video and omnidirectional video compression test cases to embrace this emerging application space [29]. This section provides a brief summary of the current state of this project. Furthermore, aspects specific to omnidirectional (or 360°) video are addressed.

A. Timeline and History

The tentative project timeline for the Versatile Video Coding (VVC) standard targets for technical completion by the end of 2020. The origins of the VVC project date back to 2015, when the Joint Video Exploration Team, also referred to as JVET, was formed by ITU-T and MPEG, to study future video coding. This joint exploration team defined an initial Joint Exploration Model (JEM) for collaborative study, which included both reference software [30] and a test model document [30]. The JEM was revised multiple times, culminating in JEM 7 [32].

During this exploration phase, a large number of coding tools were added to the JEM, using more permissive selection criteria than would be applied for adoption of tools into a draft standard. In some cases, multiple tools with similar focus were included in the JEM. Thereby, performance impact and tool interaction could be studied and evaluated. Some tools added to JEM had significant implementation complexity. The JEM software [30] requires significantly more computational resources for encoding and decoding than the HEVC HM reference software [33]. Fig. 5 shows a comparison of the coding efficiency of each iteration of the JEM, along with encoder runtimes, relative to the HM, using the JEM common test conditions [34]. JEM7 showed an average 28.5 % bitrate reduction compared to the HM, at a cost of about 10× encoder runtime and about 7× decoder runtime.

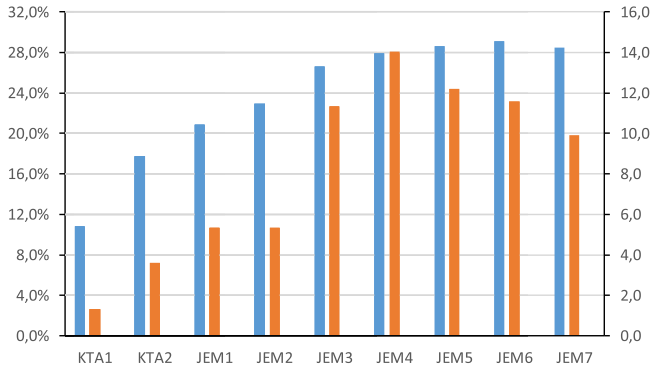

Fig. 5. Bjøntegaard bitrate savings (percent, blue bars, left axis) and encoder run-time comparison (factor compared to the HM software, orange, right axis) for the different versions of the JEM software. Data from JVET AhG report [33].

A Call for Evidence (CfE) was issued in April 2017 [35] and responses were reviewed in July 2017. Three categories were defined for different types of content, including Standard Dynamic Range (SDR), High Dynamic Range (HDR), and 360° video. Reference [36] summarized the responses to the call. Objective quality improvements in the SDR category showed about 35% bitrate reduction compared to HEVC for the test set, which is a bit higher than has been observed for the JEM using the common test conditions, as mentioned above. The difference is based upon the specific sequences and rate point differences between the CfE test set and the common test conditions, which had different JEM performance.

Following the CfE, the "future video coding" exploration activity was formalized into an official project, and the Joint Video Exploration Team evolved into the Joint Video Experts team, also called JVET [37].

B. Call for Proposals and Test Model

A Call for Proposals (CfP) was issued in 2017 [38] with responses reviewed in April 2018. As in the CfE, there were

SDR, HDR, and 360° video categories. Over all categories, 46 submissions by 32 institutions and organizations were registered, including 12 submissions in the 360° video category. The evaluation included objective performance measurements (using PSNR and Bjøntegaard metrics), information on encoder and decoder memory requirements and software runtimes and the results of a subjective evaluation using double-stimulus tests in all categories. A report on the results of the Call for Proposals is available as JVET-J0080 [39].

In April 2018, the project was named Versatile Video Coding (VVC). Based upon the results of the CfP, two software models were defined for study by JVET, the VVC test model (VTM [4], [31] and the BenchMark Set (BMS) [40]. Although the VTM and BMS can be used with HDR and 360° video, their definition was focused on the SDR use case. The initial test model version, VTM1 [31], is based on HEVC with some tools removed, including strong intra smoothing, sign data hiding in transform coding, most high-level syntax, tiles and wavefronts, and quantization weighting. VTM1 also added a multi-type tree partitioning (quad, triple, binary), similar to that first proposed in [41], and imposed a maximum Coding Unit size of 128 × 128 luma samples, and limited transform size blocks in the range of 4×4 to 64×64. The BMS1 included a wider selection of coding tools from the JEM. A list of tools in the BMS1 can be found in [42]. Experimental results of "tool on" and "tool off" tests using the VVC common test conditions [43] for the VTM and BMS tools are available in [44]. Thirteen core experiments were defined to study coding tools, including the tools included in the BMS [45].

In July 2018, the second version of the VTM and VMS were defined, which added tools to both. Many of the new tools added to the VTM2 test model [5], [46] were variants of the April 2018 BMS tools that were studied in core experiments. A list of the tools in the VTM and BMS is included in Table I.

TABLE I
CODING TOOLS ADOPTED INTO VERSATILE VIDEO CODING TEST MODEL (VTM) AND BENCHMARK SET (BMS) (STATUS DRAFT 2)

Model	Level	Tools
VTM	Partitioning	QuadTree/TrenaryTree/BinaryTree
	Intra prediction	87 Intra prediction modes
		Cross component linear model
	Inter prediction	Affine motion coding
		Advanced temporal motion vector prediction (TMVP)
	Residual coding	Multiple transform selection
		Dependent quantization
		Sign data hiding
	Loop filtering	Deblocking filter
		Adaptive loop filter
	Entropy coding	CABAC
	High-level syntax	Basic HLS
BMS	Inter prediction	Decoder motion vector refinement
		Generalized bi-prediction
		Current picture referencing
		Bi-directional template matching
		Bi-directional optical flow (BIO)
	Residual coding	4x4 Non-separable secondary transform (NSST)

Note: VTM tools are also included in the BMS.

C. 360° Video Category

As mentioned above, the 360° video category of the CfP was targeting at the omnidirectional video application scenario. A brief overview is provided here. A summary of the tools included in responses in the 360° video category is provided in [47]. Overall, 12 submissions were received in the 360° category, with 3 parties contributing 2 submissions each. Overall, 9 different projection formats were used in the submissions. Most CfP responses in this category used variations on the cubemap projection format, including two using the Equi-Angular Cubemap (EAC) format as included in [25]. Four submissions did not include any specific 360° video coding tools but just applied their proposed SDR tool set on the 360° video test set. The other submissions included tools which modified the coding loop. The proposed modifications mainly included awareness of the decoder with respect to the projection format and thereby, with respect to the location of face boundaries in the encoded video. Conventional video tools were modified to either consider the correct spherical neighborhood at these boundaries or to simply deactivate coding tools across 'wrong' neighborhoods. Such modifications included tools for intra prediction, loop filtering, or context selection for entropy coding. For motion compensation, geometry-corrected face extensions were proposed in multiple submissions. This allows for more precise motion compensation across face boundaries.

Evaluation of the test sequences in the 360° video category took into account the omnidirectional characteristics of the test data. Here, the objective and subjective quality assessment were carried out a bit differently compared to the conventional SDR category. Two objective quality metrics were to be provided with the submissions. These were calculated using the full coded frames, representing the full sphere, as described in [38]. The WS-PSNR metric provides a modified PSNR metric with the contribution of each sample location in the projected plane relative to the spherical area covered by the sample. The S-PSNR-NN is calculated by accumulating the error at sample locations in the projected plane which correspond to 655,362 uniformly distributed positions on the sphere. Details of the metric calculations can be found in [25].

For subjective evaluation, the immersive use-case was taken into account. Although it would be common for 360° video to be viewed on a head-mounted display, with the viewer selecting the region of the sphere to view based on their motion, this approach was not used during subjective testing. Test sequences were only ten seconds long, and it was thought that the subjective test viewers would all look at different regions of the sphere, so the results would not directly correspond. Instead, for subjective viewing, rectilinear dynamic viewports of size 78.1° × 49.1° were generated from the sequences and were viewed on a normal 2D monitor. The dynamic viewports were generated based on a pre-determined dynamic path through the sphere, simulating view head motion. In order to avoid the possibility of respondents to optimize video quality of just the region that was visible in the dynamic viewport path at the expense of other regions within the sphere, the selected dynamic path per sequence was not published until after the

proponents had submitted their responses to the [48]. This method was developed and tested in the context of the CfE with a different set of dynamic viewport paths [49]. The test revealed that viewers are very sensitive to dynamic viewport selected by a means other than user-controlled motion. As a consequence, slower speeds of motion were applied for the CfP view paths than in the CfE case.

Compression gains over HEVC were reported to be similar to but slightly lower than the gains for SDR [39], although it is difficult to do a direct comparison because of differing test sequences. In the ongoing development phase of VVC, coding tools specific to 360° video are considered in a dedicated Ad-hoc Group [50], [51], and in core experiments [52]. At the time of writing, decisions regarding the inclusion of specific 360° video coding tools within the VVC design have not yet been made, but may be drawn later. This would require sufficient impact on VVC compression efficiency to justify the increase in implementation complexity.

V. Point Cloud Compression

Point clouds are a volumetric representation for describing 3D objects or scenes. A point cloud comprises a set of unordered data points in a 3D space, each of which is specified by its spatial (x, y, z) position possibly along with other associated attributes, e.g., RGB color, surface normal, and reflectance. These data points collectively describe the 3D geometry and texture of the scene or object. Such a volumetric representation lends itself to immersive forms of interaction and presentation with 6DoF.

Efficient compression of point clouds for realistic rendering is much needed to reduce the required data rate or storage. In January 2017, MPEG issued a Call for Proposals (CfP) for Point Cloud Compression [53] targeting an international standard for three major categories of applications, namely, category 1: static point clouds for representing static objects and scenes; category 2: dynamic, time-varying point clouds for immersive video and VR applications; and category 3: dynamically acquired point clouds for autonomous navigation purposes. Typical use cases and capture mechanisms of these point clouds are illustrated in [54].

In October 2017, a total of 13 responses (3 for category 1, 9 for category 2, and one for category 3) to the call were received and evaluated both objectively and subjectively [55], [56]. The objective evaluation involved computing PSNR of geometric and attribute errors using the nearest neighbor correspondence between decoded and reference data points. In particular, two types of PSNR, point-to-point and point-to-plane, were defined for geometric errors. Moreover, the peak value used for PSNR computation varies according to the emphasis on local geometric error or global shape error as indicated by the geometry precision of the tested point cloud. Subjective evaluation was conducted for 6 sequences selected from categories 1 and 2 by rendering their decoded point clouds along pre-defined virtual view camera paths not known in advanced to proponents. For comparison, the CfP anchor uses a point cloud compression method proposed in [57], [58].

Substantial gain in compression performance over the anchor was reportedly achieved by the winning proposal in each category [55], [56]. Remarkably, the CfP results showed that static and dynamically acquired point clouds (in categories 1 and 3) can be efficiently compressed with octree-based geometry coding while compression of dynamic point clouds (in category 2) can largely benefit from conventional video codecs, e.g., HEVC and AVC, by converting time-varying 3D point clouds into 2D video frames. The latter allows rapid deployment of point cloud services and products. Based on these observations, it was decided to pursue further standardization of point cloud compression on two tracks: Geometry-based point cloud compression (G-PCC), which relies on the former concept and is targeted to be standardized as ISO 23090-9, and video-based point cloud compression (V-PCC) which is pursues the latter approach and is to be standardized as ISO 23090-5.

At the time of writing, the MPEG-I 3D Graphics group is progressing towards publishing this new international standard in early 2020. In July 2018, two test models were released, TMC13 [59] and TMC2 [60], corresponding to the geometry-based point cloud compression for categories 1 and 3 and video-based point cloud compression for category 2, respectively, along with the Working Drafts for G-PCC [61] and for V-PCC [62]. The G-PCC test model and working draft were a product of a consolidation effort to combine the respective test models developed earlier for categories 1 and 3. The working group also started Core Experiments on tile and slice based coding, inter-prediction for geometry coding, lossless coding, projection plan selection and reordering, lossy attribute coding, etc., to improve on the current design under the Common Test Conditions [63]. When finalized, G-PCC is expected to support, among others, both lossless and lossy compression of geometry and attribute information, coarse-to-fine progressive compression of 3D point clouds, spatial random access for view dependent decoding, and temporal random access [64].

By January 2019, the test models reached version 5 [65], [66] and a corresponding Working Draft for G-PCC and a study text of a Committee Draft for V-PCC were drafted [67], [68].

VI. MPEG-I Visual

The MPEG-I Visual workgroup with MPEG Video is studying approaches to meet requirements to provide video playback for 6DoF, as described in Section II.A. These requirements include virtual walkthroughs within a bounded volume, from 3DoF+ with slight body and head movements in a sitting position to 6DoF allowing walking steps from a central position. The work includes the capture and rendering with dedicated cameras and displays, typically referred to as Light Field devices, targeting dense Light Field representations and their dedicated codecs [69], [70].

A. 3DoF+

The 3DoF+ effort targets adding "motion parallax" to 360° video, where the relative positions of objects move, based on viewer motion, e.g., changes in both (yaw, pitch, roll) orientation and spatial (x, y, z) position. The complexity of 3DoF+ is reduced from that of full 6DoF, because of the

limited range of motion. The 3DoF+ project is intended to utilize legacy HEVC decoder hardware, and hence requires the use of single layer HEVC profiles, e.g., Main or Main 10 profiles, but may contain multiple single layer bitstreams.

3DoF+ Common Test Conditions (CTC) were initially defined in April 2018 [71] and revised in July 2018 [72]. A Call for Test Materials was issued in April 2018 [73] and revised and reissued in July 2018 [74]. The test video sequences are to contain texture and depth for a scene simultaneously captured from many different camera positions, along with metadata describing the camera positions. Each camera may either capture omnidirectional (360°) video in Equirectangular Projection (ERP) format or ordinary 2D rectilinear video.

A Draft Call for Proposals (CfP) on 3DoF+ were issued in July 2018 [21] , which was updated in October 2018 [75], and the Final Call for Proposals issued in January 2019 [20], with responses to be evaluated at the March 2019 meeting. The test sequences included in the CfP used both synthetic sequences, using computer rendering to generate texture and depth, and camera captured content. The CfP anchors use simulcast HEVC to code multiple texture views and depth views. For each of the test sequences, two anchors are provided, with the first anchor containing coded versions of all available source views, and the second anchor selecting a subset of the available source views, with both anchors meeting the same target bitrates. Responses to the CfP are free to select any subset (including the full set) of the available source views, and may form combined views. However, the combined total pixel rate of all bitstreams is an important criterion for evaluation of the responses.

Evaluation of the responses to the call will consider objective and subjective quality. Objective quality metrics based on PSNR will be calculated at each source view position, and at a pre-determined set of intermediate view positions, to represent playback for a viewer position which does not correspond to a source camera position. Subjective quality tests will be conducted using dynamic "pose traces", representing a moving viewer playback position. The effect of compression on the texture and depth pictures on the 3DoF+ viewer experience cannot be fully determined without studying the impact on interpolated views.

The Reference View Synthesizer software (RVS) [76] for view synthesis has been developed by the MPEG-I Visual group. The anchors will use RVS to interpolate intermediate views. Responses to the call may use RVS or may provide their own view interpolation approach.

The timeline for 3DoF+ standardization is not finalized, but it is estimated to be in mid 2020. When the 3DoF+ technical solution is standardized, it is expected to define metadata for use with HEVC. Metadata can be signaled at the elementary bitstream in SEI messages for HEVC, or in the systems layer, where it is planned to be included in OMAF v2 (phase 1b). View interpolation/synthesis will be required in decoder-end implementations, but the view synthesis method is not within the scope of what will be standardized. View synthesis can be considered akin to how a video encoders are treated in the standardization, e.g., video encoder reference software is provided and studied, but encoders are not fully standardized and left flexible for implementations.

B. 6DoF

Within the 6DoF related activities in MPEG-I Visual, multiple different types of input content are studied, including Windowed 6 DoF, Omnidirectional 6 DoF, and Dense Representation of Light Fields. These are exploration activities, without specific standardization timelines. Unlike the 3DoF+ effort, there is no restriction that the 6DoF activities use the HEVC single-layer profiles.

Windowed 6DoF is a restricted form of 6DoF where the user virtually views the scene from behind a (virtual) window, with any position allowing to still see at least part of the scene. Omnidirectional 6DoF is a restricted form of 6DoF, or an extended form of 3DoF allowing both (yaw, pitch roll) rotation and small translational movements of the body within a restricted volume, typically a person taking a few steps from a central position, with the ability to look all around. Omnidirectional 6 DoF is similar to 3DoF+, but allows a larger range of view position. Because of the similarities between Omnidirectional 6DoF and 3DoF+, a single Common Test Conditions document defines tests for both activities. Windowed 6DoF content is captured from ordinary 2D rectilinear cameras, while Omnidirectional 6DoF content is captured from omnidirectional (360° video) cameras.

Exploration experiments for Omnidirectional 6DoF have been defined in [77] and for Windowed 6DoF in [78]. For both types of content, depth estimation and view synthesis are studied. In the depth estimation experiments, the PSNR-based objective quality metrics are calculated with synthesized views generated using the estimated depth. The DERS depth estimation software is studied by the group [79]. Experiments have also been conducted on compression, using 3D-HEVC [3], [80], which is an extension of HEVC to code multiview texture and depth, to code the 6DoF content. The VVC codec may also be considered for use in this activity.

An exploration experiment on "Dense Representation of Light Fields" studies formats for dense light field content, lenslet format and Multiview format, and their impact on compression. The dense light field content is texture views with multiple cameras that are very densely packed, such as by lenslet cameras. The lenslet format is a single rectangular video composed of a 2D regular grid of micro-images, each of the same resolution. The multiview format (Dense Regular Multiview) is a 2D grid of views, each represented as a rectangular video of the same resolution.

VII. System Extensions for Immersive Media

Beyond the classical media coding and compression efforts, immersive media also calls for enhancements on system layers. Due to the extremely rich data sets, it is not expected that the entire 6DoF scene is available to the device at all times. It is expected, however, that by inclusion of rendering engines, only the data related to the current viewport will be downloaded from the network. This requires smart and efficient storage and streaming technologies that extend the current formats for the ISO BMFF and DASH, in particular to include cloud

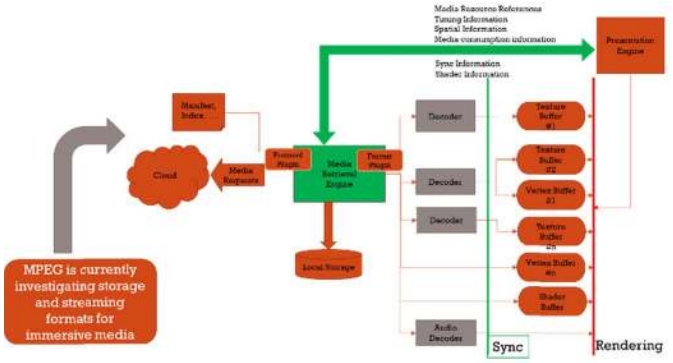

Fig. 6. Cloud Media Access of Immersive Media.

access of immersive media as shown in Fig. 6. It is also generally believed that scene description will play an important role for immersive media. However, rather than developing its own scene description, MPEG expects to rely on existing formats, but wants to ensure that MPEG immersive media can be integrated in such environments, as done similarly today for Web-based media for which HTML-5 media elements are used for the playback and control of interactive scenes.

VIII. Conclusion

Coding and transmission of immersive media is an emerging and very active field of technical development. This paper provides an overview on recent and ongoing standardization efforts in this area with a focus on the technical aspects related to video. MPEG-I "Coded Representation of Immersive Media" is the ISO/IEC standards suite addressing all aspects of coding and transmission of immersive media, ranging from architecture, over systems tools, coding of video and audio, point cloud coding, metadata, and metrics to interfaces for network-based media processing. It is developed over multiple phases with the specification of standards for coding and transmission of 3DoF video being available (phase 1a), standards for 3DoF+ emerging (phase 1b), and those suitable for 6DoF being under exploration (phase 2). The standards suite is planned to be completed by the end of 2022.

References

[1] L. Chiariglione. *MPEG-I Website*. Accessed: Feb. 8, 2019. [Online]. Available: https://mpeg.chiariglione.org/standards/mpeg-i

[2] M.-L. Champel, R. Koenen, G. Lafruit, and M. Budagavi, *Draft 1.0 of ISO/IEC 23090-1: Technical Report on Architectures for Immersive Media*, document N17685, ISO/IEC JTC1 SC29/WG11 MPEG, 123rd Meeting, Ljubljana, Slovenia, Jul. 2018.

[3] *High Efficiency Video Coding*, document ITU-T H.265, 5th ed., 2018.

[4] B. Bross, *Versatile Video Coding (Draft 1)*, document JVET-J1001, Joint Video Experts Team (JVET) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, 10th Meeting, San Diego, CA, USA, Apr. 2018.

[5] B. Bross, J. Chen, and S. Liu, *Versatile Video Coding (Draft 2)*, document JVET-K1001, Joint Video Experts Team (JVET) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Ljubljana, Slovenia, 11th Meeting, Jul. 2018.

[6] B. Bross, J. Chen, and S. Liu, *Versatile Video Coding (Draft 3)*, document JVET-L1001, Joint Video Experts Team (JVET) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Macao, 12th Meeting, Oct. 2018.

[7] B. Bross, J. Chen, and S. Liu, *Versatile Video Coding (Draft 4)*, document JVET-M1001 and N17506, Joint Video Experts Team (JVET) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Marrakech, Morocco, 13th Meeting, Jan. 2019.

[8] C. Timmerer, *MPEG Standardisation Roadmap*, document N17506, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.

[9] *Information Technology–Coded Representation of Immersive Media (MPEG-I)—Part 2: Omnidirectional Media Format*, document N17563, ISO/IEC FDIS 23090-2:201x, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.

[10] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, Oct./Dec. 2016.

[11] M. Pope, *3GPP Virtual Reality Profiles for Streaming Applications*, ETSI, document TS26.118, Sep. 2018.

[12] *Guidelines*, VR Industry Forum, Fremont, CA, USA, 2018.

[13] Khronos.org. *OpenXR Architecture*. Accessed: Feb. 8, 2019. [Online]. Available: https://www.khronos.org/openxr

[14] K. Lackner, A. Boev, and A. Gotchev, "Binocular depth perception: Does head parallax help people see better in depth?" in *Proc. 3DTV-Conf., True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, Jul. 2014, pp. 1–4.

[15] *Draft MPEG-I Audio Requirements*, document N17848, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, MPEG Audio Subgroup, Ljubljana, Slovenia, Jul. 2018.

[16] L. Chiariglione, *MPEG Time Line*, document N17701, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, Ljubljana, Slovenia, Jul. 2018.

[17] Y. He and Y. Sanchez, and Y.-K. Wang, *WD 3 of ISO/IEC 23090-6 Immersive Media Metrics*, document N17587, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.

[18] S. Deshpande, Y.-K. Wang, and M. Hannuksela, *Text of Working Draft ISO/IEC 23090-7 Immersive Media Metadata*, document N17587, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.

[19] J. Bae, *WD of ISO/IEC 23090-8 Network-Based Media Processing*, document N17872, ISO/IEC JTC1 SC29/WG11 MPEG, 123rd Meeting, Ljubljana, Slovenia, Jul. 2018.

[20] B. Kroon, *Call for Proposals on 3DoF+ Visual*, document N18145 ISO/IEC JTC1 SC29/WG11 MPEG, Marrakesh, Morocco, 125th Meeting, Jan. 2019.

[21] J. Boyce, M.-L. Chapel, Z. Deng, B. Kroon, and V. M. Vadakital, *Draft Call for Proposals on 3DoF+*, document N17724, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, Ljubljana, Slovenia, 2018.

[22] B. A. Watson and L. F. Hodges, "Using texture maps to correct for optical distortion in head-mounted displays," in *Proc. Virtual Reality Annu. Int. Symp.*, vol. 95, Mar. 1995, pp. 172–178.

[23] J. Boyce, H.-M. Oh, G. J. Sullivan, A. Tourapis, and Y.-K. Wang, *Additional Supplemental Enhancement Information for HEVC (Draft 2)*, document JCTVC-AE1005, Joint Collaborative Team on Video Coding (JCT-VC) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 31st Meeting, Apr. 2018.

[24] JVET. *360Lib*. (2017). [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib

[25] Y. Ye, E. Alshina, and J. Boyce, *Algorithm Descriptions of Projection Format Conversion and Video Quality Metrics in 360Lib Version 4*, document JVET-G1003, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Torino, 7th Meeting, Jul. 2017.

[26] Y. He, K. Choi, and V. Zakharchenko, *JVET AHG Report: 360 Video Conversion Software Development*, document JVET-I0006, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Gwangju, South Korea, 9th Meeting, Jan. 2018.

[27] J. Boyce, E. Alshina, A. Abbas, and Y. Ye, *JVET Common Test Conditions and Evaluation Procedures for 360° Video*, document JVET-H1030, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Macau, 8th Meeting, Oct. 2017.

[28] P. Hanhart, Y. He, Y. Ye, J. Boyce, Z. Deng, and L. Xu, "360° video quality evaluation," in *Proc. Int. Picture Coding Symp. (PCS)*, San Francisco, CA, USA, 2018, pp. 328–332.

[29] MPEG Requirements, *Requirements for a Future Video Coding Standard v5*, document N17074, ISO/IEC JTC1 SC29/WG11 MPEG, Torino, 119th Meeting, Jul. 2017.

[30] JVET. *JEM*. Jul. 21, 2017. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/

[31] J. Chen and E. Alshina, *Algorithm Description for Versatile Video Coding and Test Model 1 (VTM 1)*, document JVET-J1001, Joint Video Experts Team (JVET) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 10th Meeting, Apr. 2018.

[32] J. Chen, E. Alshina, G. Sullivan, and J.-R. Ohm, *Algorithm Description of Joint Exploration Test Model 7 (JEM 7)*, document JVET-G1001, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Torino, 7th Meeting, Jul. 2017.

[33] M. Karczewicz and E. Alshina, *JVET AHG Report: Tool Evaluation (AHG1)*, document JVET-H0001, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Macau, CN, 8th Meeting, Oct. 2017.

[34] K. X. Suehring and Li, *JVET Common Test Conditions and Software Reference Configurations*, document JVET-G1010, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Torino, 7th Meeting, Jul. 2017.

[35] M. Wien, V. Baroncini, J. Boyce, and A. Segall, and T. Suzuki, *Joint Call for Evidence on Video Compression With Capability Beyond HEVC*, document JVET-F1002, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, 6th Meeting, Hobart, TAS, Australia, Jul. 2017.

[36] V. Baroncini, P. Hanhart, M. Wien, J. Boyce, and A. Segall, and T. Suzuki, *Results of the Joint Call for Evidence on Video Compression With Capability Beyond HEVC*, document JVET-G1004, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Torino, 7th Meeting, Jul. 2017.

[37] *Terms of Reference of the Joint Video Experts Team (JVET) for Video Coding Standard Development*, document J. ITU-T und ISO/IEC, Macau, 2017.

[38] A. Segall, V. Baroncini, J. Boyce, and J. Chen, and T. Suzuki, *Joint Call for Proposals on Video Compression With Capability Beyond HEVC*, document JVET-H1002, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Macao, 8th Meeting, Oct. 2017.

[39] V. Baroncini, *Results of Subjective Testing of Responses to the Joint CFP on Video Compression Technology With Capability Beyond HEVC*, document JVET-J0080, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, 10th Meeting, San Diego, CA, USA, Apr. 2018.

[40] J. Boyce, *BoG Report on Benchmark Set Tool Selection*, document JVET-J0096, Joint Video Experts Team (JVET) ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 10th Meeting, Apr. 2018.

[41] X. Li, H.-C. Chuang, M. Chen, L. Zhang, X. Zhao, and A. Said, *Multi-Type-Tree*, document JVET-D0117, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Chengdu, China, 4th Meeting, Oct. 2016.

[42] W.-J. Chien, B. Alshina, J. Chen, E. François, Y. He, and Y.-W. Huang, *Methodology and Reporting Template for Tool Testing*, document JVET-J1005, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 10th Meeting, Apr. 2018.

[43] J. Boyce, K. Suehring, and X. Li, and V. Seregin, *JVET Common Test Conditions and Software Reference Configurations*, document JVET-J1010, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 10th Meeting, Apr. 2018.

[44] W.-J. Chien *et al.*, *JVET AHG Report: Tool Reporting Procedure (AHG13)*, document JVET-K0013, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Ljubljana, Slovenia, 11th Meeting, Jul. 2018.

[45] G. Sullivan and J.-R. Ohm, *Meeting Report of the 10th Meeting of the Joint Video Experts Team (JVET)*, document JVET-J1000, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 10th Meeting, Apr. 2018, pp. 10–20.

[46] J. Chen, Y. Ye, and S. H. Kim, *Algorithm Description for Versatile Video Coding and Test Model 2 (VTM 2)*, document JVET-K1002, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Ljubljana, Slovenia, 11th Meeting, Jul. 2018.

[47] J. Boyce, *BoG Report on 360° Video*, document JVET-J0085, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, 10th Meeting, San Diego, CA, USA, Apr. 2018.

[48] J. Boyce and Z. Deng, *Dynamic Viewports for 360◦ Video CfP Subjective Testing*, document JVET-J0073, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, San Diego, CA, USA, 10th Meeting, Apr. 2018.

[49] M. Wien, J. Boyce, and M. Zhou, *Viewpaths for the CFE VR Sequences*, document JVET-G0066, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Torino, 7th Meeting, Jul. 2017.

[50] G. Sullivan and J.-R. Ohm, *Meeting Report of the 11th Meeting of the Joint Video Experts Team (JVET)*, document JVET-K1000, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Ljubljana, Slovenia, 11th Meeting, Jul. 2018, pp. 10–18.

[51] J. Boyce, G. Auwera, K. Choi, and P. Hanhart, *JVET AHG Report: 360◦ Video Coding Tools and Test Conditions (AHG8)*, document JVET-K0008, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Ljubljana, Slovenia, 11th Meeting, Jul. 2018.

[52] P. Hanhart and J.-L. Lin, *CE13: Summary Report on Projection Formats*, document JVET-K0033, Joint Video Experts Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, Ljubljana, Slovenia, 11th Meeting, Jul. 2018.

[53] R. Schaefer, *Call for Proposals for Point Cloud Compression V2*, document N16763, ISO/IEC JTC1 SC29/WG11 MPEG, 117th Meeting, Hobart, TAS, Australia, Apr. 2017.

[54] C. Tulvan, R. Mekuria, Z. Li, and S. Laserre, *Use Cases for Point Cloud Compression*, document N16331, ISO/IEC JTC1 SC29/WG11 MPEG, 115th Meeting, Geneva, Switzerland, Jun. 2016.

[55] M. Preda, C. Tulvan, and C. Cao, *Merged Results of PCC CFP*, document M41501, ISO/IEC JTC1 SC29/WG11 MPEG, 120th Meeting, Macao, Oct. 2017.

[56] V. Baroncini, P. Cesar, E. Siahaan, I. Reimat, and S. Subramanyam, *Report of the Formal Subjective Assessment Test of the Submission Received in Response to the Call for Proposals for Point Cloud Compression*, document M41786, ISO/IEC JTC1 SC29/WG11 MPEG, 120th Meeting, Macao, Oct. 2017.

[57] R. Mekuria, K. Blom, and P. Cesar, *Point Cloud Codec for Tele-Immersive Video*, document M38136, ISO/IEC JTC1 SC29/WG11 MPEG, 114th Meeting, San Diego, CA, USA, Feb. 2016.

[58] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," in *Proc. IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 828–842, Apr. 2017.

[59] K. Mammou, P. A. Chou, D. Flynn, and M. Krivokuća, *PCC Test Model Category 13 v3*, document N17762, ISO/IEC JTC1 SC29/WG11 MPEG, 123th Meeting, Ljubljana, Slovenia, Jul. 2018.

[60] V. Zakharchenko, *PCC Test Model Category 2*, document N17767, ISO/IEC JTC1 SC29/WG11 MPEG, 123th Meeting, Ljubljana, Slovenia, Jul. 2018.

[61] O. Nakagami, *PCC WD G-PCC (Geometry-Based PCC)*, document N17770, ISO/IEC JTC1 SC29/WG11 MPEG, 123th Meeting, Ljubljana, Slovenia, Jul. 2018.

[62] K. Mammou, *PCC WD V-PCC (Video-Based PCC)*, document N17771, ISO/IEC JTC1 SC29/WG11 MPEG, 123th Meeting, Ljubljana, Slovenia, Jul. 2018.

[63] S. Schwarz, G. Martin-Cocher, D. Flynn, and M. Budagavi, *Common Test Conditions for Point Cloud Compression*, document N17766, ISO/IEC JTC1 SC29/WG11 MPEG, 123th Meeting, Ljubljana, Slovenia, Jul. 2018.

[64] R. Mekuria, C. Tulvan, and Z. Li, *Requirements for Point Cloud Compression*, document N16330, ISO/IEC JTC1 SC29/WG11 MPEG, 115th Meeting, Geneva, Switzerland, Jun. 2016.

[65] D. Flynn, *G-PCC Test Model V5*, document N18174, ISO/IEC JTC1 SC29/WG11 MPEG, 125th Meeting, Marrakech, Morocco, Jan. 2019.

[66] J. Rickard, *V-PCC Test Model V5*, document N18176, ISO/IEC JTC1 SC29/WG11 MPEG, 125th Meeting, Marrakech, Morocco, Jan. 2019.

[67] O. Nakagami, *PCC WD G-PCC (Geometry-Based PCC)*, document N18179, ISO/IEC JTC1 SC29/WG11 MPEG, 125th Meeting, Marrakech, Morocco, Jan. 2019.

[68] A. Tourapis, *Study Text of ISO/IEC CD 23090-5 Video-Based Point Cloud Compression*, document N18180, ISO/IEC JTC1 SC29/WG11 MPEG, 125th Meeting, Marrakech, Morocco, Jan. 2019.

[69] M. P. Tehrani *et al.*, *Exploration Experiments for MPEG-I: Compression of Dense Representation of Light Fields*, document N17723, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, Ljubljana, Slovenia, Jul. 2018.

[70] G. Lafruit and B. Kroon, *Summary on MPEG-I Visual Activities*, document N17717, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, Ljubljana, Slovenia, Jul. 2018.

[71] J. Jung, B. R. Doré, G. Lafruit, and J. Boyce, *Common Test Conditions on 3DoF+ and Windowed 6DoF*, document N17618, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.

[72] J. Jung, B. R. Doré, G. Lafruit, and J. Boyce, *Common Test Conditions on 3DoF+ and Windowed 6DoF V2*, document N17726, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, Ljubljana, Slovenia, Jul. 2018.
[73] B. R. Doré and Lafruit, *Call for Test Materials for 3DoF+ Visual*, document N17617, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.
[74] G. Lafruit, B. R. Doré, G. Bang, B. Kroon, and J. Jung, *Call for MPEG-I Visual Test Materials on 3DoF+ and 6DoF*, document N17720, ISO/IEC JTC1 SC29/WG11 MPEG, 123nd Meeting, Ljubljana, Slovenia, Jul. 2018.
[75] B. Kroon, *Draft Call for Proposals on 3DoF+ Visual*, document N18097, ISO/IEC JTC1 SC29/WG11 MPEG, Macao, 124th Meeting, Oct. 2018.
[76] B. Kroon and G. Lafruit, *Reference View Synthesizer (RVS) 2.0 Manual*, document N17759, ISO/IEC JTC1 SC29/WG11 MPEG, 123rd Meeting, Ljubljana, Slovenia, Jul. 2018.
[77] G. Bang, B. Kroon, M. P. Tehrani, K. Wegner, and G. Lafruit, *Exploration Experiments for MPEG-I: Omnidirectional 6DoF (V4)*, document N17722, ISO/IEC JTC1 SC29/WG11 MPEG, 123rd Meeting, Ljubljana, Slovenia, Jul. 2018.
[78] E. Juárez and D. Doyen, *Exploration Experiments for MPEG-I: Windowed-6DoF*, document N17721, ISO/IEC JTC1 SC29/WG11 MPEG, 123rd Meeting, Ljubljana, Slovenia, Jul. 2018.
[79] T. Senoh, K. Yamamoto, N. Tetsutani, and H. Yasuda, *MPEG-I-Visual: Enhanced DERS for Quad Reference Views (eDERS)*, document M41955, ISO/IEC JTC1 SC29/WG11 MPEG, 121st Meeting, Gwangju, South Korea, Jan. 2018.
[80] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.
[81] K. P. Choi, V. Zakharchenko, M. Choi, and E. Alshina, *Test Sequence Formats for Virtual Reality Video Coding*, document JVET-C0050, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP3 and ISO/IEC JTC1 SC29/WG11, 3rd Meeting, Geneva, Switzerland, May 2016.
[82] M.-L. Champel, R. Koenen, G. Lafruit, and M. Budagavi, *Proposed Draft 1.0 of TR: Technical Report on Architectures for Immersive Media*, document N17685, ISO/IEC JTC1 SC29/WG11 MPEG, 122nd Meeting, San Diego, CA, USA, Apr. 2018.

Mathias Wien received the Diploma and Dr.Ing. degrees from Rheinisch-Westfälische Technische Hochschule Aachen (RWTH Aachen University), Aachen, Germany, in 1997 and 2004, respectively. In 2018, he achieved the status of the habilitation, which makes him an independent scientist in visual media communication. He was with the Institut für Nachrichtentechnik, RWTH Aachen University (head: Prof. Jens-Rainer Ohm) as a Researcher from 1997 to 2006, and as a Senior Researcher and the Head of Administration from 2006 to 2018. Since 2018, he has been with Lehrstuhl für Bildverarbeitung, RWTH Aachen University (head: Prof. Dorit Merhof) as a Senior Researcher, a Leader of the Visual Media Communication Group, and the Head of Administration. He has published more than 60 scientific articles and conference papers in the area of video coding and has co-authored several patents in this area. He has further authored and co-authored more than 100 standardization documents. He has published the Springer textbook *High Efficiency Video Coding: Coding Tools and Specification*, which fully covers Version 1 of HEVC. His research interests include image and video processing, immersive, space-frequency adaptive and scalable video compression, and robust video transmission. He has been an Active Contributor to VVC, HEVC, and H.264/AVC. He has participated and contribute to ITU-T VCEG, ISO/IEC MPEG, the Joint Video Team, the Joint Collaborative Team on Video Coding, and the Joint Video Experts Team of VCEG and ISO/IEC MPEG. He is a member of the IEEE Signal Processing Society and the IEEE Circuits and Systems Society. He is a member of the VSPC Technical Committee of the IEEE CAS Society. He has served as a Co-Editor of the scalability amendment to H.264/AVC. In the aforementioned standardization bodies, he has co-chaired and coordinated several AdHoc groups and tool- and core experiments. Since 2018, he has been serving as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.

Jill M. Boyce (S'85–M'90–SM'10–F'19) received the B.S. degree in electrical engineering from the University of Kansas in 1988 and the M.S.E. degree in electrical engineering from Princeton University in 1990.

She was formerly the Director of algorithms with Vidyo, Inc., where she led video and audio coding and processing algorithm development. She was formerly a Vice President of Research and Innovation Princeton for Technicolor, formerly Thomson. She was formerly with Lucent Technologies Bell Labs, AT&T Labs, and Hitachi America. She is currently an Intel Fellow and the Chief Media Architect with Intel, responsible for defining media hardware architectures for Intel's video hardware designs. She represents Intel at the Joint Collaborative Team on Video Coding and Joint Video Exploration Team of ITU-T SG16 and ISO/IEC MPEG. She serves as an Associate Rapporteur of ITU-T VCEG, and was an Editor of the Scalability High Efficiency Video Coding Extension. She is the inventor of more than 150 granted U.S. patents, and has published more than 40 papers in peer-reviewed conferences and journals. She was an Associate Editor from 2006 to 2010 of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.

Thomas Stockhammer received the Dipl.Ing. and Dr.Ing. degrees from the Munich University of Technology, Munich, Germany. He was a Visiting Researcher with the Rensselear Politechnical Institute, Troy, NY, USA, and the University of California at San Diego, San Diego, CA, USA. After acting as a Co-Founder and a CEO of Novel Mobile Radio Research for ten years, he joined Qualcomm as Director Technical Standards in 2014. In his different roles, he has co-authored more than 200 research publications and more than 150 patents. In his day job, he is the active and has leadership and rapporteur positions in 3GPP, DVB, MPEG, IETF, ATSC, CTA, VR Industry Forum, and the DASH-Industry Forum in multimedia communication, TV-distribution, content delivery protocols, virtual reality, and adaptive streaming.

Wen-Hsiao Peng received the Ph.D. degree from National Chiao Tung University (NCTU), Taiwan, in 2005. He was with the Intel Microprocessor Research Laboratory, USA, from 2000 to 2001, where he was involved in the development of ISO/IEC MPEG-4 fine granularity scalability. Since 2003, he has been actively participating in the ISO/IEC and ITU-T video coding standardization process and contributed to the development of SVC, HEVC, and SCC standards. From 2015 to 2016, he was a Visiting Scholar with the IBM Thomas J. Watson Research Center, USA. He is currently a Professor with the Computer Science Department, NCTU. He has authored more than 60 journal/conference papers and more than 60 ISO/IEC and ITU-T standards contributions. His research interests include video/image coding, deep/machine learning, multimedia analytics, and computer vision. He is a Member of the VSPC and MSA Technical Committees of the IEEE CAS Society. He was a Technical Program Co-Chair for 2011 IEEE VCIP, 2017 IEEE ISPACS, and 2018 APSIPA ASC; a Publication Chair for 2019 IEEE ICIP; an Area Chair for IEEE ICME and VCIP; and a Review Committee Member for IEEE ISCAS. He served as a Lead Guest Editor/Guest Editor/SEB Member for IEEE JETCAS. More recently, he was an Elected Distinguished Lecturer of APSIPA and a Chair Elect of IEEE CASS VSPC Technical Committee.