Understanding MPEG-I Coding Standardization in Immersive VR/AR Applications

By Gauthier Lafruit, Daniele Bonatto, Christian Tulvan, Marius Preda, and Lu Yu

Abstract

After decennia of developing leading-edge 2D video compression technologies, the Moving Picture Expert Group (MPEG) is currently working on the new era of coding for immersive applications, referred to as MPEG-I, where "I" refers to the "Immersive" aspects. It ranges from 360° video with head-mounted displays to free navigation in 3D space with head-mounted and 3D light field displays. Two families of coding approaches, covering typical industrial workflows, are currently considered for standardization—MultiView + Depth (MVD) Video Coding and Point Cloud Coding—both supporting high-quality rendering at bitrates of up to a couple of hundreds of megabits per second. This paper provides a technical/historical overview of the acquisition, coding, and rendering technologies considered in the MPEG-I standardization activities.

Keywords

3/6 Degrees of Freedom (3/6-DoF), depth image-based rendering (DIBR), Moving Picture Expert Group immersive (MPEG-I) video compression, point cloud coding (PCC)

This paper provides a technical/historical overview of the acquisition, coding, and rendering technologies considered in the MPEG-I standardization activities.

Introduction

The Moving Picture Expert Group (MPEG) standardization committee is currently working on MPEG-I coding technologies to support immersive applications,[1] where multimedia content can be viewed from various viewpoints that are different from the camera acquisition viewpoints, therefore supporting free navigation around regions of interest in the scene, for example, circling around a player in a sports event, similar to *The Matrix* bullet time effect.[2]

MPEG-I supports 360° video on head-mounted displays [extension of existing video codecs with supplemental enhancement information (SEI) messaging for the projection format in the so-called *Omnidirectional Media Format* (OMAF)], covering only 3 Degrees of Freedom (3DoF) head rotations. Extensions supporting motion parallax within some limited range around the central viewing/camera position, referred to as *3DoF+*, are expected to be standardized in mid/end-2020. This will allow larger ranges of freedom of movement, eventually achieving full 6DoF that allows any user-viewing position in 3D space, with standards to be accepted by industry beyond 2020.[3]

Competitive coding technologies for advanced virtual reality (VR)/augmented reality (AR) and light field display devices are under study, encompassing equirectangular video projection (ERP), MultiView + Depth (MVD) Coding, and Point Cloud Coding (PCC), where the former two are familiar to video-based production workflows (e.g., 3D film production) and the latter to 3D graphics-based workflows (e.g., 3D game production), both steadily evolving toward cinematic VR/AR.

MPEG has issued several Calls for Test Material, Exploration, and Core Experiments for comparing the relative merits of technologies from industrial proponents around the world, supporting 3D extensions of High-Efficiency Video Coding (HEVC),[4] MVD Coding in video production, and Octree- and kd-based 3D data representations used for PCC in early versions of Lidar devices.[5]

The plan is that 3DoF+ will be supported in a short term by market-existing, 2D video codec devices adding supplementary metadata, while 6DoF may need enhanced coding tools in a longer term to handle even larger volumes of data. In that respect, the maturity of existing technologies for PCC, assessed after a Call for Proposal issued by MPEG in 2017, conducted the committee to start building the technical specifications for this coding approach with the aim to publish the final standard by early 2020.

The MVD video coding technologies for MPEG-I are under exploration in the MPEG Video Group, while PCC technologies are studied in MPEG 3D Graphics

Digital Object Identifier 10.5594/JMI.2019.2941362
Date of publication: XX XXXX 2019

Group (3DG). Both types of technologies are grouped under the MPEG-I umbrella since they contribute to the common goal of addressing immersive applications. The two subgroups, however, have historically started their activities independently of each other, using their own data sets and common test conditions (CTC), but we will see in the remainder of the paper that cross-fertilization has led to technologies showing stunning similarities, eventually leading to a common bitstream format. Nevertheless, the software tools to create the meta-data in 3DoF+ and PCC remain specialized for the application domain.

Both MPEG-I Video and MPEG-I Graphics coding technologies are even expected to reach similar bitrates of around a couple of hundreds of megabits per second for high-end cinematic VR/AR productions, irrespective of the technological specificities of the proposed coding approaches.

MPEG-I Processing and Coding Pipeline

Figure 1 shows the generic processing and coding pipeline in a typical MPEG-I immersive application, seamlessly integrating video- and graphics-based approaches. The input data consist of multiple camera views of the scene and associated depth (hence the name *MVD*) and/or point clouds from RGB color + Depth (RGBD) depth sensing devices.

The input camera feeds are preprocessed for color correction, distortion removal, depth estimation, and/or point cloud extraction, before being compressed and transmitted, eventually accompanied by some metadata. The decoder [**Fig. 1** (right)] unpacks, decodes, and extracts the data in the video- or graphics-based data representation formats, and, finally, a renderer does additional post-processing to obtain an animated image sequence that is displayed on the screen or head-mounted device (HMD).

In contrast to classical 2D video coding, the renderer does much more than placing the decoded data as pixels on the screen. For instance, in MPEG-I Video, the images decoded from the bitstream will be interpolated by a view synthesis (VS) process to create any virtual view to the scene, hence providing the 3DoF+ or 6DoF immersive experience to the user. In MPEG-I graphics, however, a point cloud is created from the decoded bitstream—a collection of colored points in 3D space—which are projected on the screen through a typical OpenGL 3D graphics pipeline. Since the points are not connected and may possibly leave gaps, they are enlarged to disks with splatting[6] through the rendering (post-processing) module shown in **Fig. 1**.

The following sections provide more details on the various modules shown in **Fig. 1** for the MPEG-I video and MPEG-I graphics processing pipelines, indicating their differences and commonalities.

MPEG-I Video Multiview + Depth Coding

In the MPEG-I video pipeline, the various color camera views are transmitted with mild preprocessing (e.g., distortion removal and color correction) to the Metadata for Immersive Video (MIV) coder, and processed after decoding at the renderer side for creating any virtual viewpoint in response to the user's spatial viewing position. For the latter, typically, VS requires a depth map per camera input for synthesizing any intermediate view with depth image-based rendering techniques.[7] Consequently, all camera feeds and their corresponding depth maps are transmitted through the network, as in the example shown in **Fig. 2** for the Technicolor Painter

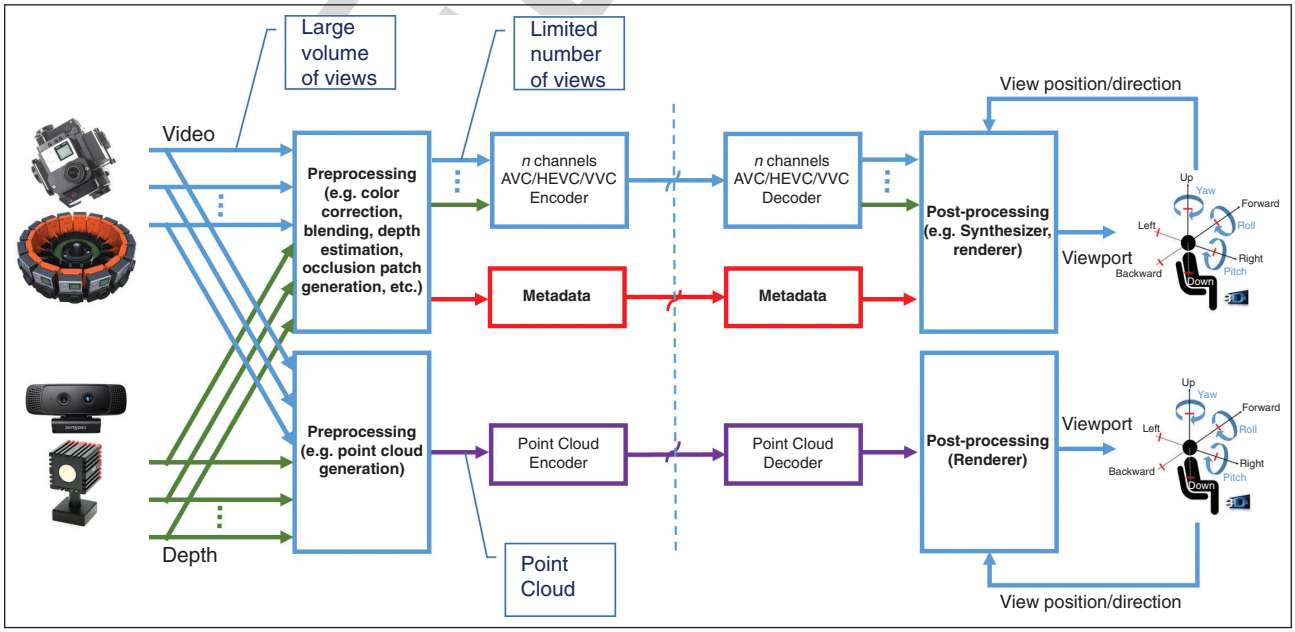


FIGURE 1. Video and graphics-based workflow of MPEG-I.

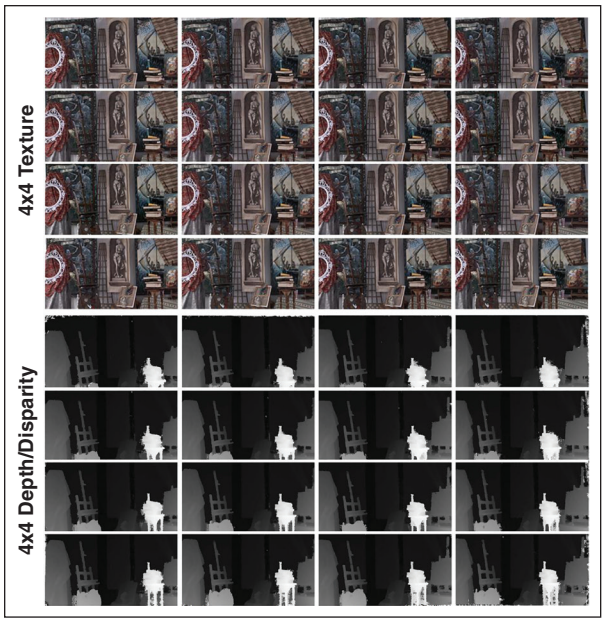


FIGURE 2. Technicolor painter multiview capture (top) with estimated depth (bottom).

test sequence, which is one of the many MVD video test sequences used in MPEG-I.[8]

The creation of these depth maps in the preprocessing module is not part of the coding standard and is the sole responsibility of the content provider, who may use active depth sensing or passive depth estimation techniques (e.g., stereo matching). MPEG-I video recommends using its Depth Estimation Reference Software (DERS),[9] with a recent extension to enhanced DERS (eDERS),[10] and refactored versions for better widespread use, even outside the MPEG community.[11]

Strictly speaking, the VS module, if not used at the encoder (see the next paragraph as a counter example), is also not part of the coding standard, though it has a huge impact on the final rendering quality (similar to the depth estimation/sensing) and all benchmarking decisions. Therefore, it has been extensively studied over the past years, starting with VS Reference Software (VSRS),[9] which was originally developed for horizontal-parallax-only (HPO) autostereoscopic displays with small disparity ranges (hence subject to improvement), its extensions to enhanced VSRS (eVSRS),[12] and recently more advanced implementations[13,14] that surpass VSRS and are now part of the Test Model for Immersive Video (TMIV) reference software.[15] This software will eventually become the reference implementation of the upcoming MIV standard.[16]

Although the preprocessing and post-processing modules shown in **Fig. 1** are not part of the coding standard, they are considered in all MPEG-I experiments, since they impact the Quality-Bitrate performance figure of the coding system. Moreover, the redundancies between the multiview images shown in **Fig. 2** might be exploited for better coding, using VS as a view prediction, where a camera view is predicted from adjacent camera inputs, and only the difference image (the residual) is actually coded and transmitted through the network. This core idea is further extended in MIV by creating some reference views (e.g., the bottom stitched view in **Fig. 3** using the top input views[17]) and disocclusions: these are regions that are originally not visible from some camera viewpoints, but become visible for virtual viewpoints in between the transmitted camera feeds. This creates a collection of disocclusion patches, cf. the red box in the middle row of **Fig. 3**, that are packed together as supplemental metadata into a so-called *atlas*. Eventually, reference views and supplemental metadata/atlases are coded with existing video codecs and transmitted through the network. With this information, the MIV decoder and VS can reconstruct any virtual viewpoint to the scene.

Note that the input camera views do not have to be parallel, in contrast to what is suggested in **Fig. 2**. Indeed, the system works perfectly well with convergent/divergent camera feeds as in the example of the Technicolor Museum test sequence[18] shown in **Fig. 3**, with its disocclusion patches shown in the middle row. In these conditions, MIV outperforms unaltered MVD coding techniques based on HEVC.[19] With a ballpark figure of 0.04 bits per refreshed pixel (including the depth maps) for HEVC,[20] a typical setup of 16–25 camera feeds in ultrahigh-definition (UHD) (3840 × 2160 pixels) would lead to a maximum of 150–240 Mbits/sec for 30 frames/s. In applications with HMDs requiring much higher frame rates (at least 90–120 frames/s, i.e., three to four fold), the total bitrate will increase, but probably less than the corresponding frame rate ratio (expected to be a factor 2).

MPEG-I Graphics Point Cloud Coding

In MVD coding as presented in the previous section, raw RGBD data (color + depth) is used in the preprocessing modules shown in **Fig. 1** to create the metadata in **Fig. 3** to be transmitted. In contrast, PCC starts from another data representation where all the input data has been thoroughly processed/filtered to create 3D objects that are specifically meant to be viewed from any direction. This filtering often involves heavy processing that goes far beyond the video preprocessing of the previous section to ensure high-quality 3D point clouds and/or 3D object meshes.[21,22]

Since 3DG uses point clouds as input data representation, the early coding activities of the 3DG group were oriented toward Octree- and kd-based coding used in the very first Lidar devices.[5] The basic principle is that the points are grouped into a hierarchical structure of branches and leaves that allows for better difference/residual coding between a representative point and its direct neighbors in a group [**Fig. 4** (bottom)]. This method yields compression performances of one order

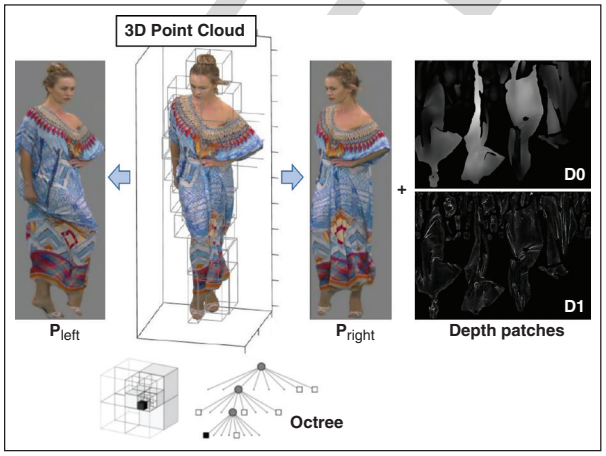
Courtesy: Technicolor/InterDigital.

FIGURE 3. MIV coding of input views (top) to one or more reference views (bottom) and disocclusion patches collected into an atlas (middle).

of magnitude for static scenes, but it was very difficult to further extend its performances to the temporal axis with leaves that jump from one branch to another in the octree, even after a simple translation of an unaltered object in space.

Specifically, for dynamic point clouds, it was therefore proposed to find existing codecs that could already well exploit the temporal changes of the data, leading to the usual suspect: the video codec. The point cloud (typically for a single object) is first segmented in subsets—called *patches*—each with a smoothly varying depth profile. Each patch is projected onto different planes in space with respect to its local orientation (**Figs. 4** and **5**), together with its depth maps (i.e., the distance from each point to the projection plane, called D0). The so-obtained images are then coded with already widely accepted 2D video codecs [e.g., Advanced Video Coding (AVC) or HEVC].

One may object that it makes little sense to start from a multicamera acquisition providing images, out of which a point cloud is typically created by photogrammetry,[21] which in turn is projected back into an MVD projection. However, be aware that, in practice, the extraction of a point cloud of natural scenery from images (e.g., the preprocessing module shown in **Fig. 1**) requires many different viewpoints to be acquired, typically in the order of hundred(s) of images; while, once a high-quality point cloud is extracted, a lower number of well-chosen projection directions (e.g., one order of magnitude less) may be sufficient to well-code the point cloud in its whole.



Courtesy: 8i.

FIGURE 4. The 3D Point Cloud, its octree, and projections (P_{left}, P_{right}, …) with depth patches D0 and D1.

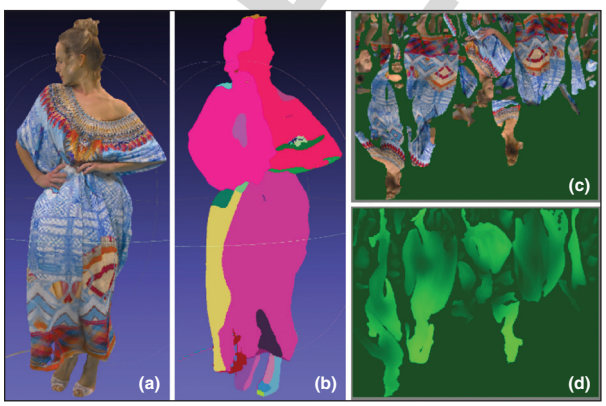
Nevertheless, note that, in this point cloud projection process, there may be some occlusions that cannot be handled properly—for example, when two points in space are projected on the same point in the projection plane, such as under the arms of the persons. For this case, a second depth map (D1) is introduced that encodes the delta between the two points along with the projection axis (**Fig. 4**). One may observe that the 2D distribution of pixels in the patch image is not compression-friendly, that is, the 2D space is not uniformly occupied. To handle this situation, an occupancy map that consists of a binary mask of useful pixels is also encoded and transmitted.

This patch concept is actually extended over all regions of the object—similar to the texture UV mapping of 3D graphics objects[23]—even where there are no occlusions, leading to the typical structure of **Fig. 5(c)**, which corresponds to the metadata shown in **Fig. 1**. This has the advantage that traditional video codecs can be used, making MPEG PCC[24] straightforward to be supported by a huge set of devices already available on the market.

Experiments are still under consideration to best distribute the patches temporally to keep the highest coherence over time, and hence the best exploitation of redundancy in the codec for higher coding gains.[25]

With respect to the coding performances in PCC, a bit rate of 10–20 Mbits/sec at 30 frames/s has been observed, per object, on the extensive point cloud animation test set used in MPEG-I graphics[24] and MPEG 3DG,[26] rendered on a UHD display. It is important to indicate that these figures are obtained for single-object PCC coding, hence the total bitrate for scenes with multiple objects is increased accordingly with the number of objects.

For simple scenes with a dozen objects, 120–240 Mbits/sec at 30 frames/s is therefore required, which is the same performance figure as reported with MVD coding in MPEG-I Video. As a result, not only do the coding approaches of MPEG-I Video and MPEG-I Graphics exhibit a lot of similarities, they also yield comparable coding performances.



Courtesy: Apple Inc.

FIGURE 5. (a) One projection of a point cloud, and (b) its segmentation in (c) texture and (d) depth patches.

Consequently, MPEG-I Video and MPEG-I Graphics share a lot of technologies, with one noteworthy difference: while MPEG-I Video takes great care in the VS during the post-processing in **Fig. 1**, MPEG-I Graphics heavily relies on a proper point cloud extraction in the preprocessing module of **Fig. 1**. As this difference is not part of the standard, there are no strict boundaries between the two approaches. The MPEG committee has therefore taken actions to create a single data coding format, both for MPEG-I video and MPEG-I PCC,[27] somehow merging the three central arrows of **Fig. 1** into a single transmission format. Only the pre- and post-processing modules of **Fig. 1** remain different.

Future MPEG-I Experiments

Ever since the first working draft issued after the Call for Proposals in July 2017, MPEG-I PCC continued to evolve by integrating new tools to increase the coding efficiency: the lossless mode is now supported by grouping the misprojected points into a special patch, alternative approaches for encoding the occupancy map were proposed and time-consistent packing is under construction. Although the activity is still ongoing, it is expected that an additional 20%–30% of the coding gain will be obtained before issuing the final standard by early 2020.

On the other hand, MPEG-I video has set up CTC since April 2018, for 3DoF+ and 6DoF scenarios with test material, anchor definitions, objective and subjective evaluation methods, and so on. Gradual improvements have been reached in the TMIV software,[15] clearly showing that packing information of the video streams provides the user with an interactive experience of motion parallax in a 3D scene (3DoF+). After its final standardization in mid/end-2020, it is expected that a Call for Proposals will be issued for long-term 6DoF activities, which provide larger freedom of movement in the scene beyond the capabilities of 3DoF+.

Conclusion

The year 2020 will be an important milestone for immersive VR/AR applications in the range of 3DoF+ to 6DoF, thanks to the new coding standards of MPEG-I. Two MPEG-I approaches (i.e., video- and 3D graphics-based) have been studied and provide similar compression technology and coding performances, reaching a couple of hundreds of megabits per second at 30 frames/sec. A common data compression and transmission format will be released in 2020 within the frameworks of MPEG-I Visual and PCC.

Acknowledgments

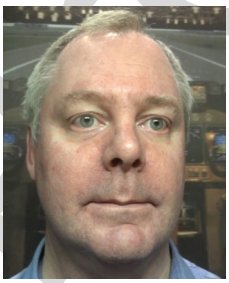
We would like to thank all of the MPEG-I experts for their contributions to this work, as well as Innoviris, the Brussels Institute for Research and Innovation,

Belgium, for supporting the work with respect to multicamera acquisitions, processing, and rendering (contract number 2015-DS-39a, 3DLicorneA). **Figs. 2–5** are altered reproductions with permission from the MPEG-I contributors, that is, Technicolor/InterDigital, 8i, and Apple Inc.

References

1. Moving Picture Experts Group, MPEG-I, "Coded Representation of Immersive Media." *ISO/IEC 23090*. [Online]. Available: https://mpeg.chiariglione.org/standards/mpeg-i, 2018.
2. K. C. Karthikeyan, "How the Matrix Bullet Time Works?" Geekswipe, Aug. 12, 2017. [Online]. Available: https://geekswipe.net/art/films/how-matrix-bullet-time-works/
3. R. Koenen, "MPEG Standardization Roadmap," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/N17506, San Diego, CA, Apr. 2018.
4. G. J. Sullivan et al., "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circ. Syst. Video Techn.*, 22(12):1649–1668, Dec. 2012.
5. R. Schnabel and R. Klein, "Octree-Based Point-Cloud Compression," *Proc. Symp. Point-Based Graphics, Eurographics*, pp. 111–121, Jul. 2006.
6. M. Botsch, M. Spernat, and L. Kobbelt, "Phong Splatting," *Proc. First Eurographics Conf. Point-Based Graphics*, Switzerland, pp. 25–32, 2004.
7. W. Sun et al. "An Overview of Free Viewpoint Depth-Image-Based Rendering (DIBR)," *Proc. Second APSIPA Annu. Summit Conf.*, Singapore, pp. 1023–1030, Dec. 2010.
8. M. Panahpour Tehrani et al., "Overview of MPEG-I Visual Test Materials," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/N17606, San Diego, CA, Apr. 2018.
9. D. Mieloch, G. Lafruit, and B. Kroon, "Summary on MPEG-I Visual Activities," *ISO/IEC JTC1/SC29/WG11*. MPEG2019/N18560, Gothenburg, Sweden, Jul. 2019.
10. T. Senoh et al., "Updated eDERS to Higher Precision," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/m42525, San Diego, CA, Apr. 2018.
11. S. Rogge et al., "Refactoring DERS 8.0 to RDE," *ISO/IEC JTC1/SC29/WG11*. MPEG2019/M48075, Gothenburg, Sweden, Jul. 2019.
12. T. Senoh et al., "Enhanced VSRS to Four Reference Views," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/m42526, San Diego, CA, Apr. 2018.
13. R. Doré et al. "3DoF+ Intermediate View Synthetizer Proposal," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/m42486, San Diego, CA, 2018, April 2018.
14. S. Fachada et al., "Depth Image-Based View Synthesis with Multiple Reference Views for Virtual Reality." *3DTV-CON*, Jun. 2018.
15. B. Salahieh et al., "Test Model 2 for Immersive Video," *ISO/IEC JTC1/SC29/WG11*. MPEG2019/N18577, Gothenburg, Sweden, Jul. 2019.
16. J. Boyce, R. Doré, and V. K. Malamal Vadakital, "Working Draft 2 of Immersive Video," *ISO/IEC JTC1/SC29/WG11*. MPEG2019/N18576, Gothenburg, Sweden, Jul. 2019.
17. J. Fleureau et al., "Description of Technicolor Intel Response to MPEG-I 3DoF+ Call for Proposal," *ISO/IEC JTC1/SC29/WG11*. MPEG2019/m47445, Geneva, Switzerland, Mar. 2019.
18. R. Doré, G. Briand, and T. Tapie, "Technicolor 3DoFPlus Test Materials," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/m42349, San Diego, CA, Apr. 2018.
19. J. Jung, "Workshop on Coding Technologies for Immersive Audio/Visual Experiences: How to Achieve 6DoF Compression," *ISO/IEC JTC1/SC29/WG11*. MPEG2019/N18559, Gothenburg, Sweden, Jul. 10, 2019.
20. A. Hinds et al., "Toward the Realization of Six Degrees-of-Freedom with Compressed Light Fields," *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2017.
21. B. Blizard, "The Art of Photogrammetry: Introduction to Software and Hardware," Tested, Feb. 11, 2014. [Online]. Available:http://www.tested.com/art/makers/460057-tested-dark-art-photogrammetry/
22. M. Dou et al., "Fusion4D: Real-Time Performance Capture of Challenging Scene," *ACM Trans. Graphics*, 35(4), Jul. 2016, pp. 114:1–114:13.
23. G. Aguiar, "Blender Beginner Tutorial - UV Mapping (Part 4)," YouTube, Sep. 17, 2017. [Online]. https://www.youtube.com/watch?v=Vj0k4-I33lQ
24. M. Preda, "Report on PCC CfP Answers," *ISO/IEC JTC1/SC29/WG11*. MPEG2017/N17251, Macao, China, Oct. 2017.
25. M. Budagavi et al., "PCC Core Experiments for Category 2," *ISO/IEC JTC1/SC29/WG11*. MPEG2018/N17346, Gwangju, Korea, Jan. 2018.
26. MPEG 3DG, "Call for Proposals for Point Cloud Compression V2." *ISO/IEC JTC1/SC29/WG11*. MPEG2017/N16763, Hobart, Australia, Apr. 2017.
27. S. Schwarz et al., "Emerging MPEG Standards for Point Cloud Compression," *IEEE J. Emerg. Select. Topics Circ. Syst.*, 9(1):133–148, Mar. 2019.

About the Authors



Gauthier Lafruit received an ME degree in electromechanics from the Free University of Brussels, Belgium, in 1989, where he also obtained his doctoral dissertation in 1995 in the field of wavelet imaging, for which he was awarded the Barco Prize. In 1996, he joined imec, Leuven, Belgium, the microelectronics research institute, specializing in compression and image analysis for space applications and broadcasting. This gradually led him to follow standardization committees of European Space Agency (Consulative Committee for Space Data Systems) as well as audiovisual and multimedia standards in general (JPEG and MPEG). After a year at the University of Hasselt, Belgium, taking advantage of a specialization scholarship in research, he joined the Université Libre de Bruxelles, Brussels, the French wing of the Free University of Brussels, in 2014, where he is currently an associate professor in 3D imaging in all its forms, including stereoscopy, multicamera acquisitions, 3D video games, virtual reality, and digital holography.

Daniele Bonatto received a computational intelligence software engineering degree in applied sciences from the Université Libre de Bruxelles, Brussels, Belgium, in 2016. He is currently enrolled in a PhD program in realtime 3D computing, jointly between the Université Libre de Bruxelles and the Vrije Universiteit Brussel, respectively the French and

Dutch wing of the Free University of Brussels, Brussels, Belgium. He works on realtime free-viewpoint rendering of natural scenery with sparse multicamera acquisition setups. Jointly with the Moving Picture Experts Group (MPEG), Bonatto developed the reference view synthesis software in 2018 and two high-density natural scene data sets for static and dynamic content explorations.

Christian Tulvan received an engineering degree in electronics and a master's degree in embedded systems from the University of Transylvania, Braşov, Romania. He is currently a research engineer at the Institut Mines-Télécom, Paris, France. Tulvan has been the chairman of the Reconfigurable Media Coding Ad hoc Group (AhG) and Moving Picture Expert Group (MPEG) Assets AhG, International Organization for Standardization (ISO) MPEG. He contributes to various ISO standards with technologies in the field of 3D graphics. He is part of a research team focusing on 3D graphics, cloud architectures, and interactive media. Tulvan received several ISO certifications of appreciation.

Marius Preda received an engineering degree in electronics from the Politehnica University of Bucharest, Bucharest, Romania, in 1998, and a PhD degree in mathematics and informatics from Paris V University - René Descartes, Paris, France, in 2001. Preda has been actively involved in Moving Picture Expert Group (MPEG)-4 since 1998, focusing in particular on coding for video and face and body animation (FBA). He was the main contributor of the new animation tools dedicated to generic synthetic objects, promoted by MPEG-4 as part of Animation Framework eXtension (AFX) specifications and currently leads the Point Cloud Coding activities in the MPEG-I standardization. Preda's research interests include generic virtual character animation, rendering, low-bitrate compression and transmission of animated data, multimedia composition, and multimedia standardization.

Lu Yu is a professor at the College of Information Science and Electronic Engineering at Zhejiang University, Hangzhou, China. She received a BEng degree (Hons.) in radio engineering and a PhD degree in communication and electronic systems from Zhejiang University in 1991 and 1996, respectively. Her research interests include visual perception, visual perceptual quality assessment, visual signal representation and coding, multimedia processing, and related architecture design. Yu has published more than 160 peer-reviewed papers and was granted more than 80 patents. She acted as video subgroup cochair and chair of the Audio and Video Coding Standardization Working Group of China (AVS) for 16 years and has been ISO/IEC JTC 1/SC 29/WG11 (MPEG) video subgroup chair since January 2018.

Presented at IBC2018, Amsterdam, The Netherlands, 13–17 September 2018. This paper is published here by kind permission of the IBC.