MULTI-VIEWPOINT AND OVERLAYS IN THE MPEG OMAF STANDARD

Igor D.D. Curcio, Kashyap Kammachi Sreedhar, Sujeet S. Mate
Nokia Technologies, Tampere, Finland

Abstract *– Recent developments in immersive media have made possible the rise of new multimedia applications and services that complement the traditional ones, such as media streaming and conferencing. Omnidirectional video (a.k.a. 360-degree video) is enabling one of such new services that are progressively made available also by large media distribution portals (e.g., YouTube). With the aim of creating a standardized solution for 360-degree video streaming, the Moving Picture Experts Group (MPEG) has developed the Omnidirectional MediA Format (OMAF) second edition, or version 2, which is close to completion. The major new features of OMAFv2, compared to the first version, include (but are not limited to) the capability of using overlays and multiple omnidirectional cameras situated at different physical points (i.e., viewpoints). This paper focuses on the description of two of the new OMAFv2 features, the overlays and the multi-viewpoints, including the 360-degree video use cases enabled by these two features.*

Keywords – Immersive media, MPEG OMAF, multimedia streaming, multi-viewpoints, omnidirectional video, overlays.

1. INTRODUCTION

Immersive media is one of the current buzzwords in media technologies. It refers to the capability of making the user feel immersed in the audio, video, and other media, at the same time increasing the interactivity level. The Reality-Virtuality continuum [1], allows a wide spectrum of immersion and interactivity levels, more towards the real environment or the virtual environment, depending on the actual application or service considered.

Watching 360-degree videos is one way to consume immersive media. 360-degree video content is typically played back in a virtual environment using a Head Mounted Display (HMD). Whenever the user is enabled to explore the content only by changing the HMD orientation by varying the yaw, pitch and roll of the head (i.e., rotational movements), this is defined as a 3 Degrees of Freedom (3DoF) media [2]. YouTube already offers omnidirectional video in their portal, and this type of medium is becoming more and more popular. If the consumer is also allowed to move in the 360-degree space and navigate, walk, see behind the objects in the scene (i.e., translational movements), this is typically defined as 6 Degrees of Freedom (6DoF) media [2].

The Moving Picture Experts Group (MPEG) has defined the first standard for an Omnidirectional MediA Format (OMAF) [3] to enable the easy deployment of interoperable standardized streaming services for 360-degree video. OMAF is also the basis of the technology adopted by the Third Generation Partnership Project (3GPP) in their specification for omnidirectional video streaming since Release 15 [4]. OMAF defines the basic storage format as well as the transport over Dynamic Adaptive Streaming over HTTP (DASH) [23] and MPEG Media Transport (MMT) [24] for audio, video, image, and timed text. Lately, MPEG has been working on the second version of the OMAF standard [5] with the aim of extending the functionalities already enabled by the first version, and make its adoption more appealing for service providers and the media industry in general.

The major features specified in OMAFv2 are overlays, multi-viewpoints, sub-pictures and new tiling profiles for viewport-dependent streaming. This paper will focus on the first two ones. *Overlays* are a way to enhance the information content of 360-degree video. They allow us to superimpose another piece of content (e.g., a picture, another video with news, advertisements, text or other) on top of the main (background) omnidirectional video. Overlays also allow the creation of interactivity points or areas. The content captured by an omnidirectional capture device or an omnidirectional media corresponding to one omnidirectional camera is called a viewpoint in the OMAFv2 terminology. *Multi-viewpoint* is a set of capture devices which, for example, may be scattered around a stadium. The OMAFv2 specification enables a streaming format with multiple viewpoints to allow, for example, switching from one viewpoint to another, as done by multi-camera directors for traditional video productions.

Overlays have been researched in the past years in several application areas but, to the best of the authors' knowledge, minimally in the area of 360-degree video. In [6], the authors present an image overlay system to aid procedures in computerized tomography. The system in [7] shows a way to display dynamic image overlays during surgical operations using a stereo camera and augmented reality visualization. The authors of [8] describe a method for real-time overlay insertion in a predefined location of a pre-encoded ITU-T H.264/AVC video sequence. The detection and extraction of an overlaid text on top of a complex video scene and news are studied in [9, 10]. Augmented reality-based image overlays on optical see-through displays mounted on the front glass of a car have been studied by [11]. A similar research, but performed using a Virtual Reality (VR) HMD with a video-based see-through display, is presented in [12]. Here the authors use an accelerometer to compensate for the motion-to-photon delay between the image overlay and the reality displayed on the HMD's screen, and give a method to improve the registration among the two. An implementation of a system that displays overlays in unmanned aerial vehicles is presented in [13]. An intelligent video overlay system for advertisements is presented in [14]. Here the overlays are not placed in fixed positions, but are located such that the intrusiveness to the user is minimized and is done by detecting faces, text and salient areas in the video.

Multi-viewpoint 360-degree video streaming is a relatively new area. For traditional mobile 2D video, multi-camera video remixing has been extensively researched by some of the authors of this paper. See [15, 16, 17], for example. The work [18] presents streaming from multiple 360-degree viewpoints where these are capturing the same scene from different angles. A challenge described by the authors is viewpoint switching and how to minimize disruption after a switch. The authors also emphasize the importance of switching prediction in order to minimize the impact on the Quality of Experience (QoE). The research in [19] focuses on low latency multi-viewpoint 360-degree interactive video. The authors use multimodal learning and a deep reinforcement learning technique to detect events (visual, audio, text) and predict future bandwidths, head rotation and viewpoint selection for improving media quality and reducing latency.

The present paper focuses on two of the new main features included in the second edition of the MPEG OMAF standard, namely overlays and multi-viewpoints. The new enabled use cases are also introduced.

The structure of the remaining parts of this paper is as follows. Section 2 shows the OMAF system architecture. Section 3 focuses on the overlay capabilities and functionalities in OMAFv2. Section 4 describes how multiple omnidirectional viewpoints can be utilized in OMAFv2. Finally, section 5 concludes the paper.

2. MPEG OMAF SYSTEM ARCHITECTURE

This section introduces the general MPEG OMAF system architecture. This is depicted in Fig. 1, which is extracted from the draft OMAFv2 standard specification [5]. The figure shows the end-to-end content flow process from acquisition up to display/playback for live and on-demand streaming use cases. The specification applies to projected omnidirectional video (equirectangular and cube map) as well as to fisheye video. It defines media storage and metadata signaling in the ISO Base Media File Format (ISOBMFF) [25] (i.e., interfaces F and F' in Fig. 1). It also defines media encapsulation and signaling in DASH and MMT.

OMAF specifies also audio, video, image and timed text media profiles, i.e., the interfaces E'a, E'v, E'i. All other interfaces depicted in the figure above are not normatively specified. Additionally, OMAF defines different presentation profiles for viewport-independent and viewport-dependent streaming. For further details on these two concepts, the reader may refer to [22].

Following Fig. 1, media content is initially captured. Audio is encoded using 3D Audio with the MPEG-H [26] audio low complexity Profile at level 1/2/3 or the MPEG-4 High Efficiency AACv2 at Level 4 codec [27]. Visual content is first stitched, possibly rotated, projected and packed. Subsequently, it is encoded using the MPEG High Efficiency Video Codec (HEVC) Main 10 profile at Level 5.1 [28] or the MPEG Advanced Video Codec (AVC) Progressive/High profile at Level 5.1 [29]. Images are encoded using the HEVC image profile Main 10 at Level 5.1 or Joint Pictures Experct Group (JPEG) images [30]. The encoded streams are then placed into an ISOBMFF file for storage or encapsulated into media segments for streaming. The segments are delivered to the receiver via the DASH or MMT protocols. At the receiver side (player), the media is decapsulated, and then decoded with the respective decoder(s), and subsequently rendered on a display (e.g., a HMD) or loudspeakers. The head/eye tracking orientation/viewport metadata determine the user viewing orientation within

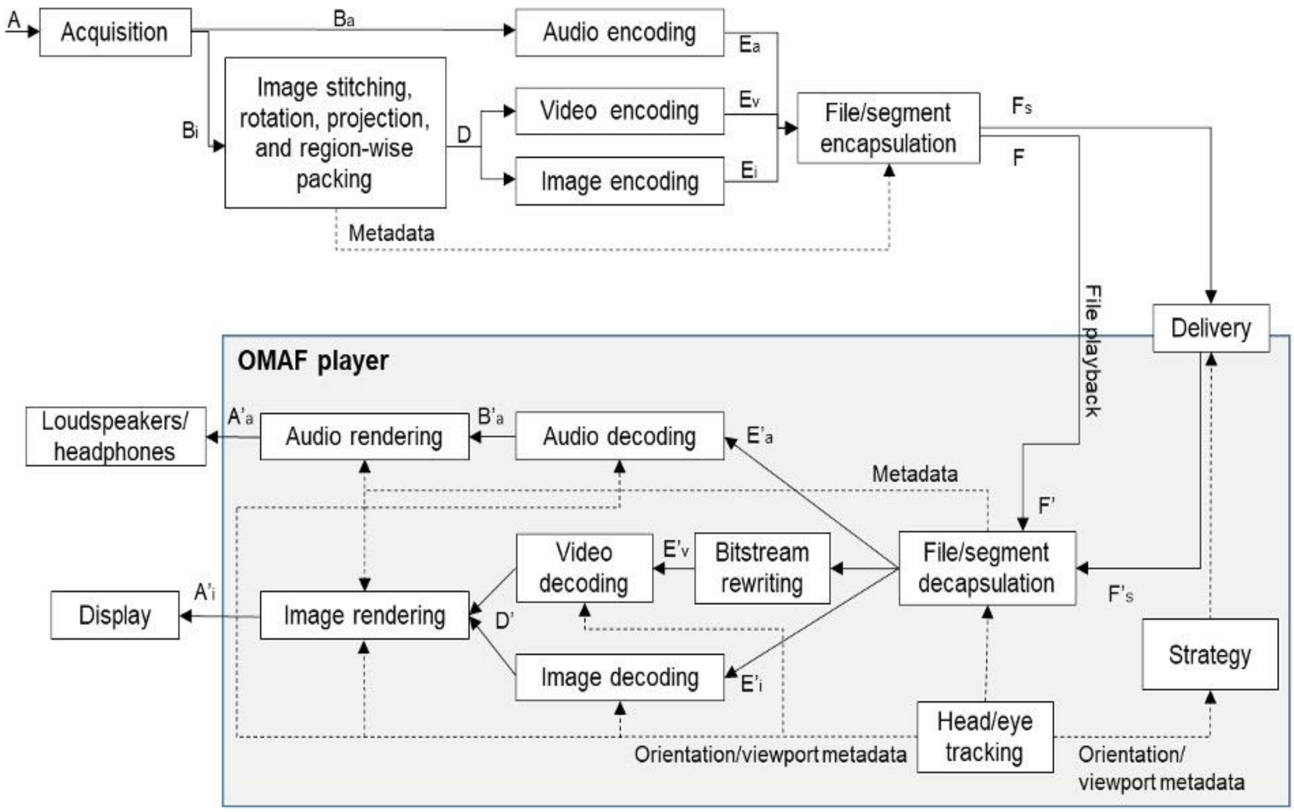

Fig. 1 – Content flow process for an MPEG OMAF system [5].

the omnidirectional space and is used to guide the player to download and render the appropriate part of the media streams. The viewport defines the horizontal and vertical user field of view within the omnidirectional space.

The new OMAFv2 functionalities discussed in this paper, namely overlays and multi-viewpoints, are defined both in the ISOMBFF, where the storage format syntax and semantics are described in detail, and at DASH and MMT transport levels [5]. The mentioned new functionalities are purposely described in the following two sections without mentioning the ISOBMFF and DASH syntax and semantics details.

3. OVERLAYS IN THE MPEG OMAF STANDARD

Overlays are conventionally used to provide complementary information about the video scene and enhance the overall user experience. Recent studies on the use of overlays in a VR application [20] have hinted that the presence of an overlay within an omnidirectional video increases the immersion experience of the user. Typical examples of overlay include, but are not limited to, advertisements, logos, subtitles, thumbnails of a map, etc.

Some of the omnidirectional video use cases in which overlays could be used are the following[1].

I. Omnidirectional sports videos in which the background video consists of the live match that is being played, and the overlay is a video of the scorecard.
II. Other use of overlay videos in the context of sports may include an omnidirectional background video of the match and an overlay video of the highlights package or an overlay video of a commentary box and interview places.
III. Overlays could also be part of an immersive session about a museum where the background video covers the museum near a well-known monument, and the overlay video tells about the historical facts of the monument.

[1] It is noted that the list of use cases presented here is not exhaustive and is limited only for briefness.

Fig. 2 shows an example of an overlay in an omnidirectional system. The background omnidirectional video is rendered on the spherical surface and the overlay video is rendered on a plane inside the spherical volume. The user viewing the omnidirectional video is assumed to be present at the center of the spherical volume.

MPEG OMAFv2 defines an overlay as a visual media (video, image, or timed text) that is rendered over an omnidirectional video or image or over a viewport. In the presence of an overlay, the visual media on which it is overlaid is referred to as *background visual media*. In the present context, the background visual media can be considered as OMAF omnidirectional video or image content. The next section describes the different features of overlays as defined in OMAFv2.

3.1 Overlay features in OMAFv2

OMAFv2 provides a framework to indicate the number of overlays present and also active in the omnidirectional system. Additionally, it provides the capability and flexibility to control different overlay features. These features can be broadly categorized into the following four types:

a) Spatial property;

b) Temporal property;

c) Interactivity property;

d) Inherent property.

3.1.1 Spatial property

Every overlay in the OMAFv2 system has a defined spatial position. The standard defines the following three possible positions for an overlay:

1) the overlay could be positioned at a depth from the user viewing position, as shown in Fig. 2;

2) the overlay may be positioned over the background video without any gap between the two, as shown in Fig. 3;

3) the overlay may be positioned on the users' viewing screen. This is also called viewport-relative or viewport-locked overlay; it is always present on the users' viewing screen and is independent of the users' current viewport. An example is shown in Fig. 4.

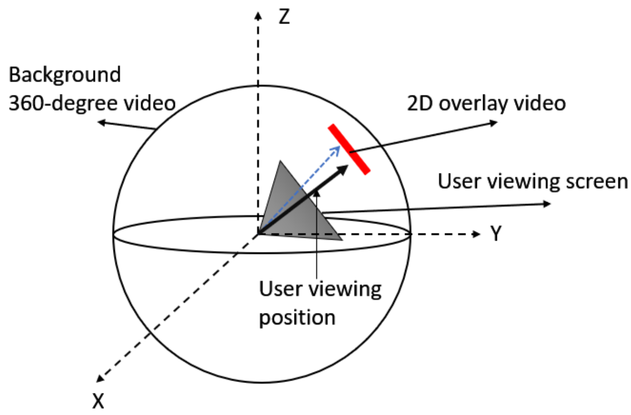

Fig.2 – Example of a 2D overlay at a depth from the user viewing point in an OMAFv2 omnidirectional system.

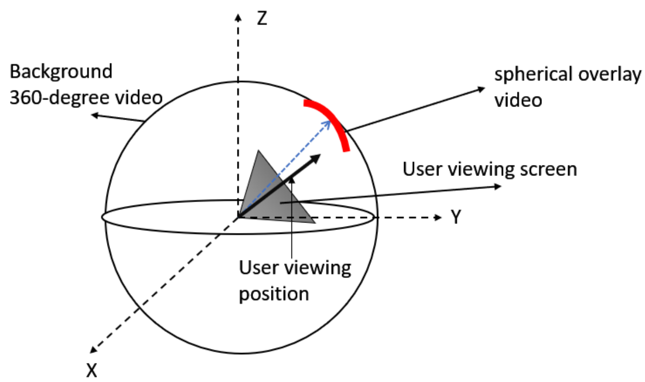

Fig. 3 – Example of a spherical overlay on the background video in an OMAFv2 omnidirectional system.

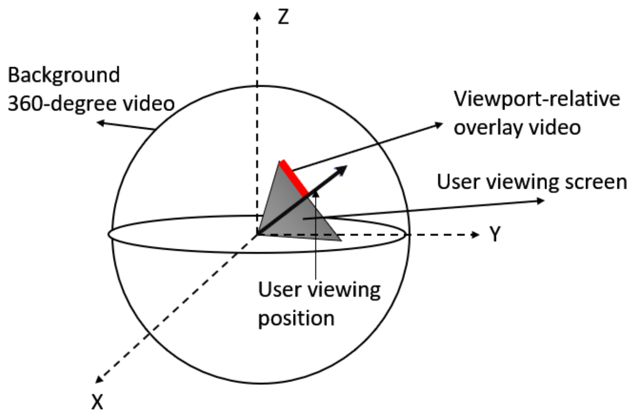

Fig. 4 – Example of a viewport-relative overlay in an OMAFv2 omnidirectional system.

The standard also enables the flexibility to signal the rotation angles (yaw, pitch, roll) (for 2D overlays only) and the size of each overlay. In addition, it also allows us to signal the layering order of overlays when multiple overlays are present at the same depth in the omnidirectional system.

3.1.2 Temporal property

OMAFv2 provides tools that enable altering the temporal properties of the overlay based on user interaction and the background timeline. A user may choose to switch to a new overlay from the currently viewing overlay, potentially altering the temporal property of the overlay relative to the background video (see also section 3.1.3). The standard also allows a content provider to control the playback of the overlay based on the user's viewing direction. This feature enables the control of the temporal property of the overlay based on the spatial direction. For example, the content provider may choose to pause the playback of an overlay when the overlay is not visible in the user's viewport, and once the user returns to the viewport where the overlay is visible, the overlay is played back from the pause state. Such a scenario can be realized in case of both 2D and spherical overlays of Fig. 2 and Fig. 3, respectively.

3.1.3 Interactivity property

The OMAFv2 standard allows the flexibility to enable user interaction with an overlay. Some of the possible interactions that could be performed on an overlay are the following:

1) rotate an overlay;
2) resize an overlay;
3) switch on-off an overlay;
4) change the spatial position within the omnidirectional system;
5) switch to a new overlay.

3.1.4 Inherent property

OMAFv2 provides for the following inherent properties to be either explicitly signaled or implied by the usage of an overlay:

1. **Source**: an overlay could be a separate bit stream from the background bit stream or could be part of the same stream as the background stream. The latter indicates that the overlay video spatially coexists with the background video.
2. **Representation**: an overlay could be a 2D video/image or a spherically projected video/image.
3. **Priority**: the standard allows us to signal a priority bound to an overlay. Signaling the priority is helpful for the omnidirectional system in the presence of a large number of overlays, as the system could take smart decisions based on its available network bandwidth and resources, such as memory and system buffers.
4. **Opacity and Alpha Channel**: the standard allows us to signal the opacity and the alpha values of an overlay. This helps in controlling the transparency of the overlay.

4. MULTI-VIEWPOINT IN THE MPEG OMAF STANDARD

Omnidirectional cameras typically capture subjects with sufficient details if they are close to it. Subjects which are further away from the camera appear with lower details. In addition, OMAF (v1) content enables a single viewpoint to have three degrees of freedom. 3DoF allows perusing content in all directions around a single location. However, a single viewpoint does not allow watching a person, event or object of interest from a different location. Hence, OMAFv2 has incorporated support for multiple viewpoints to address the need to enable high quality content capture as well as provide the possibility of experiencing any subject or event of interest from a different perspective. Furthermore, multiple viewpoints facilitate leveraging the well-established cinematic rules for multi-camera directors that make use of different shot types, such as wide-angles, mid-shots, close-ups, etc. [21].

The standard supports enablers which provide more freedom for content creators to design content for diverse scenarios and effective storytelling.

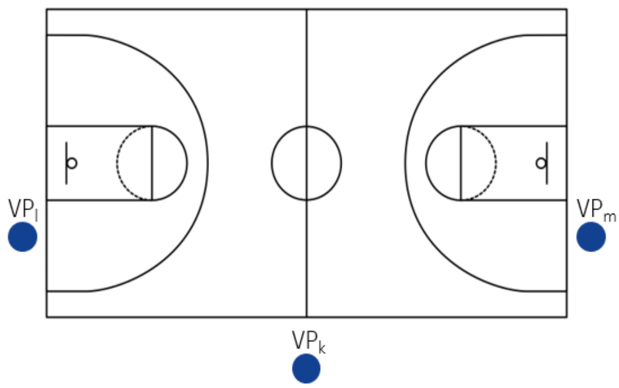

Fig. 5 – Example of a basketball game with multiple viewpoints.

Fig. 5 illustrates an example of OMAFv2 content with multiple viewpoints (VP_k, VP_l and VP_m). This allows the user to experience the action close to where it happens, as well as from different perspectives. In the following, the key concepts and features for supporting multiple viewpoints in the OMAFv2 standard are described. These facilitate interactive content consumption and creative storytelling.

4.1 Spatial relationship between multiple viewpoints

Each viewpoint is represented by a spatial position with respect to a common reference coordinate system with a granularity of 10^{-1} millimeters in the 3D space. This allows creating content such as the example given in Fig. 5. Having multiple viewpoints in a contiguous space (such as a sport game or a performance on a stage) benefits by having a common reference coordinate system to ensure the individual viewpoint positions correspond to the events in the content.

However, there are also scenarios where the content may consist of some non-contiguous spaces. For example, several viewpoints which are on a sport field, while other viewpoints which cover the space outside the stadium or the team locker room. This is achieved by defining multiple viewpoint groups, where only viewpoints in a group share a common reference coordinate system. Fig. 6 illustrates such a scenario that comprises of four viewpoint groups (Stadium, Locker Room 1, Locker Room 2, Stadium Entrance). Furthermore, the standard also supports viewpoints which have a dynamic position, such as first-person view from a race car or a flying drone.

The standard enables also the specifying of the geographical position and rotation offset between the geomagnetic North and the common reference coordinate system. This allows any OMAFv2 player with geolocation tracker and magnetic compass to be located in the real world.

Fig. 6 – Example content with 4 viewpoint groups (Stadium, Locker Room 1, Locker Room 2, Stadium Entrance).

4.2 Viewpoint switching

The OMAFv2 standard provides a versatile tool to leverage the possibility of switching between the different viewpoints within the same viewpoint group or across different viewpoint groups. Consequently, there is a possibility to create content based on the content creator storyline paradigm, as well as the user preference driven switching behavior.

4.2.1 Content creator specified switching

OMAFv1 supports the signaling of the default behavior information for a viewpoint, such as specifying the initial viewing orientation. This information can be used while starting with a viewpoint or after switching to a new viewpoint. However, this information may not be sufficient if the VR content is complex and comprises of multiple viewpoints where not all of them may be available for switching at a given time, depending on the storyline needs of the content creator. For example, in a scene where there are four viewpoints (VP_1, VP_2, VP_3 and VP_4), there may be a scenario where switching from VP_1 to VP_3 would be possible only after switching to VP_2. To handle such a situation, OMAFv2 provides the content creator with the tools to specify the switching candidates at any given time. In addition, the content creator can specify if a viewpoint switch occurs only as a temporally seamless switch or with a temporal offset. Furthermore, a preferred transition effect to be effective while switching between the two viewpoints may also be specified by the content creator. This enables creative and interactive storytelling for VR content.

4.2.2 User-preference-driven switching

The content creator intent is essential for consistent experience and storytelling. However, the user choice and exploration freedom are key aspects of a truly immersive content experience. A scenario is illustrated in Fig. 7 to describe the case of user-driven viewpoint switching metadata. The content is comprised of viewpoints VP_1, VP_2 and VP_3, with default initial viewing orientations V_{invo}-VP_i, where i=1, 2, 3. In addition, there are three objects of interest O_1, O_2 and O_3. The orientation of the user after switching from one viewpoint to another viewpoint depends on the user's viewport orientation before performing the viewpoint switch. The viewport orientation depends on the object of interest the user is following. The standard supports signaling of orientations of the object or person of interest while switching from a particular viewpoint to the other viewpoint. The content creator can analyze the content from all the viewpoints in order to determine the orientation of a particular person or object of interest for obtaining their orientation in each viewpoint at a given time. In Fig. 7, VO_1-VPi represent the viewport orientations of the object of interest O_1 from viewpoints VP_1 and VP_2. Similarly, for other objects of interest O_2 and O_3. This information can be utilized by the player application to select the object of interest in the current viewport and its orientation in the viewpoint where the user wishes to switch. The selection of the object of interest can be performed interactively by the user or based on a preset application implementation criterion, e.g., the selection of the object which is in the center of the viewport. This type of switching facilitates performing a viewpoint switch such that the object or person of interest remains in the viewport after the switch.

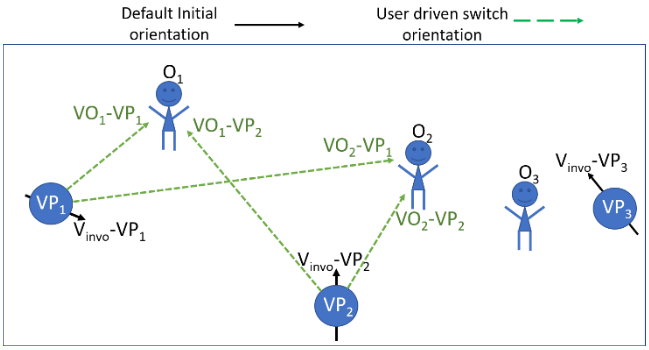

Fig. 7 – Illustration of user preference driven switching. Default initial viewing orientations V_{invo}-VP_i, where i=1,2,3 are indicated by the black arrow in each viewpoint. The green dotted lines represent the point of interest dependent viewport orientations (VOi-VPi). The user's viewport after the viewpoint switch depends on the viewport orientation before the switch.

4.2.3 Fully automatic viewpoint switching

The standard enables also the concept of fully automatic viewpoint switching with the help of recommended viewport timed metadata tracks which indicate the viewport changes for one or more viewpoints over a time interval. Consequently, the user experiences the change in viewport over time without any interaction or head motion. This is well suited for providing replays of sports highlights, or any other events which can benefit from different perspectives.

5. CONCLUSION

This paper presented an overview of two of the new features included in the second edition of the MPEG Omnidirectional MediA Format (OMAF) standard, namely, overlays and mutli-viewpoints. The MPEG committee is about to finalize the second edition of the OMAF standard specification. We introduced the main functionalities of these two features and highlighted the new enabled use cases. These make the new OMAFv2 standard even more appealing for content producers, service providers and device manufacturers. Other features in the MPEG standardization pipeline include the support of "limited" 6DoF content (a.k.a. 3DoF+), which enhances even further the spectrum of possible applications.

ACKNOWLEDGEMENT

The authors would like to thank all colleagues in the MPEG committee who have contributed to develop the OMAF version 2 standard specification.

REFERENCES

[1] I.D.D. Curcio, A. Dipace and A. Norlund (2016). Virtual Realities and Education, Research on Education and Media, 8(2), 60-68.

[2] Six degrees of freedom, Wikipedia, Accessed 04 May 2020.

[3] ISO/IEC (2018). Information technology – Coded representation of immersive media (MPEG-I) – Part 2: Omnidirectional media format, Final Draft International Standard.

[4] 3GPP (2018). 3GPP Virtual reality profiles for streaming applications, Technical Specification 26.118, v. 15.1.0.

[5] ISO/IEC (2020). DIS Draft Text for 2nd Ed. of 23090-2, MPEG document m52943.

[6] G. Fichtinger, A. Deguet, K. Masamune, E. Balogh, G.S. Fischer, H. Mathieu, R.H. Taylor, S.J. Zinreich and L.M. Fayad (2005). Image Overlay Guidance for Needle Insertion in CT Scanner, IEEE Transactions on Biomedical Engineering, 52(8), 1415-1424.

[7] N. Zevallos, R.A. Srivatsan, H. Salman, L. Li, J. Qian, S. Saxena, M. Xu, K. Patath and H. Choset (2018). A surgical system for automatic registration, stiffness mapping and dynamic image overlay, IEEE International Symposium on Medical Robotics, Atlanta, GA, U.S.A.

[8] E. Kaminsky, D. Grois and O. Hadar (2010). Efficient Real-Time Video-in-Video Insertion into a Pre-Encoded Video Stream for the H.264/AVC, IEEE International Conference on Imaging Systems and Techniques, Thessaloniki, Greece.

[9] W. Kim and C. Kim (2009). A New Approach for Overlay Text Detection and Extraction From Complex Video Scene, IEEE Transactions on Image Processing, 18(2), 401-411.

[10] S. Lee and K. Jo (2017). Automatic Person Information Extraction Using Overlay Text in Television News Interview Videos, IEEE 15th International Conference on Industrial Informatics, Emden, Germany, 583-588.

[11] T. Nakatsuru, Y. Yokokohji, D. Eto and T. Yoshikawa (2003). Image Overlay on Optical See-through Displays for Vehicle Navigation, IEEE/ACM 2nd International Symposium on Mixed and Augmented Reality, Tokyo, Japan.

[12] Y. Yokokohji, Y. Sugawara and T. Yoshikawa (2000). Accurate Image Overlay on Video See-Through HMDs Using Vision and Accelerometers, IEEE Virtual Reality Conference, New Brunswick, NJ, U.S.A.

[13] D. Madden, A. Scanlon, Y. Zhou, T.E. Choe and M. Smith (2014). Real Time Video Overlays, ACM SIGGRAPH, Vancouver, BC, Canada.

[14] J. Guo, T. Mei, F. Liu and X.-S. Hua (2009). AdOn: An Intelligent Overlay Video Advertising System, ACM 32nd International Conference on Research and Development of Information Retrieval, Boston, MA, U.S.A.

[15] S. Vihavainen, S. Mate, L. Seppälä, F. Cricri' and I.D.D. Curcio (2011). We Want More: Human-Computer Collaboration in Mobile Social Video Remixing of Music Concerts, ACM Conference on Human Factors in Computer Systems (CHI), Vancouver, Canada.

[16] S. Vihavainen, S. Mate, L. Liikkanen and I.D.D. Curcio (2012). Video as Memorabilia: User Needs for Collaborative Automatic Mobile Video Production, ACM Conference on Human Factors in Computer Systems (CHI), Austin, TX, U.S.A.

[17] S. Mate and I.D.D. Curcio (2017). Automatic Video Remixing Systems, IEEE Communications Magazine, 55(1), 180-187.

[18] X. Corbillon, F. De Simone, G. Simon and P. Frossard (2018). Dynamic Adaptive Streaming for Multi-Viewpoint Omnidirectional Video, ACM Multimedia Systems Conference, Amsterdam, The Netherlands, 237-249.

[19] H. Pang, C. Zhang, F. Wang J. Liu and L. Sun (2019). Towards Low Latency Multi-viewpoint 360° Interactive Video: A Multimodal Deep Reinforcement Learning Approach, IEEE Conference on Computer Communications, Paris, France, 991-999.

[20] M. Salmimaa, J. Kimmel, T. Jokela, P. Eskolin, T. Järvenpää, P. Piippo, K. Müller and J. Satopää (2018). Live delivery of neurosurgical operating theater experience in virtual reality, Journal of the Society for Information Display, 26(2), 98-104.

[21] S. Mate, I.D.D. Curcio, A. Eronen and A. Lehtiniemi (2015). Automatic Multi-Camera Remix from Single Video, ACM 30th Symposium on Applied Computing, Salamanca, Spain, 1270-1277.

[22] I.D.D. Curcio, H. Toukomaa and D. Naik (2018). 360-Degree Video Streaming and its Subjective Quality, SMPTE Motion Imaging Journal, 127(7), 28-38.

[23] ISO/IEC 23009-1, Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats.

[24] ISO/IEC 23008-1, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media transport (MMT).

[25] ISO/IEC 14496-12, Information technology – Coding of audio-visual objects – Part 12: ISO base media file format.

[26] ISO/IEC 23008-3, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio.

[27] ISO/IEC 14496-3, Information technology – Coding of audio-visual objects – Part 3: Advanced audio coding.

[28] ISO/IEC 23008-2, Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding.

[29] ISO/IEC 14496-10, Information technology – Coding of audio-visual objects – Part 10: Advanced video coding.

[30] ISO/IEC 10918-1:1994, Information technology – Digital compression and coding of continuous-tone still images – Part 1: Requirements and guidelines.