Immersive Media Delivery: Overview of Ongoing Standardization Activities

Christian Timmerer

Abstract

More and more immersive media applications and services are emerging on the market, but lack international standards to enable interoperability. This article provides an overview about ongoing standardization efforts in this exciting domain and highlights open research and standardization issues.

Introduction

Universal media access (UMA) [1], as proposed in the late 1990s and early 2000s, is now reality. It is very easy to generate, distribute, share, and consume any media content, anywhere, anytime, and with/on any device. These kinds of real-time entertainment services — specifically, streaming audio and video — are typically deployed over the open, unmanaged Internet. Interestingly, these services now account for more than 70 percent of the evening traffic in North American fixed access networks, and it is assumed that this number will reach 80 percent by the end of 2020 [2]. A major technical breakthrough and enabler was certainly adaptive streaming over HTTP, resulting in the standardization of MPEG-DASH [3].

One of the next steps toward a truly immersive media experience is most likely related to virtual reality (VR) applications and, specifically, omnidirectional (360°) media delivery, which is currently built on top of the existing adaptive streaming ecosystems [4].

Omnidirectional video (ODV) content allows the user to change her/his viewing direction in multiple directions while consuming the video, resulting in a more immersive experience than consuming traditional video content with a fixed viewing direction. Such video content can be consumed using different devices ranging from smartphones and desktop computers to head-mounted displays (HMDs) like Oculus Rift, Samsung Gear VR, and HTC Vive, among others. When using an HMD to watch such content, the viewing direction can be changed by head movements. On smartphones and tablets, the viewing direction can be changed by touch interaction or by moving the device around thanks to built-in sensors. On a desktop computer, the mouse or keyboard can be used for interacting with the omnidirectional video.

In the past, more and more standards development organizations (SDOs) (including support organizations) started working on different aspects in this domain. In this article, we highlight major interfaces within an immersive media delivery ecosystem. We provide an overview of ongoing standardization activities in this domain. The article concludes with a list of open issues that could be understood as a research and standardization roadmap for the (near) future.

Major Interfaces in Immersive Media Delivery

The basic system architecture, including major interfaces of an immersive media delivery ecosystem, is shown in Fig. 1.

Media capture comprises multiple audio-video tracks (e.g., from several microphones and cameras) including various types of metadata ①, which are *fused/stitched* together and further *edited* before entering the subsequent processing step ②. This processing step considers projection and metadata, and encodes the media content (potentially in various versions for adaptive delivery). Furthermore, it *encapsulates* the media content utilizing appropriate *storage* and/or *delivery* formats (possibly including *encryption*) ③ before it is decoded on the end user device. After *decoding* (including *decapsulation* and possibly decryption), various *projection* and metadata ④ will guide the *rendering* process, which interacts with the corresponding input/output technology (e.g., HMDs) eventually enabling the *consumption* of the immersive media content ⑤. The role of metadata in this context is a crucial aspect and needs to be selected very carefully. The focus of this article is on interfaces 2-4 as these interfaces are mainly concerned with interoperability.

Standardization Overview

Figure 2 depicts an overview of SDOs, including support/related organizations, that have started new projects or are currently working in the immersive media delivery area grouped into three clusters (bottom up):

1. *Data representation and formats* providing basic tools to be adopted directly by industry or referenced by other SDOs (or parts thereof)
2. *Guidelines*, *system standards*, and *application programming interfaces* (APIs) typically providing so-called system specifications including end-to-end aspects
3. Quality of experience (QoE) addressing the perceived quality as experienced by the end users of such applications and services

The author is with Alpen-Adria-Universität Klagenfurt/Bitmovin Inc.

Digital Object Identifier:
10.1109/MCOMSTD.2017.1700038

The current focus of the VRIF is on an end-to-end guideline adopting OMAF and DASH which primarily targets on-demand services and later will be extended to live services. Future guidelines may also include other distribution models such as broadcasting. Interestingly, the VRIF will also address security and privacy issues related to such services.

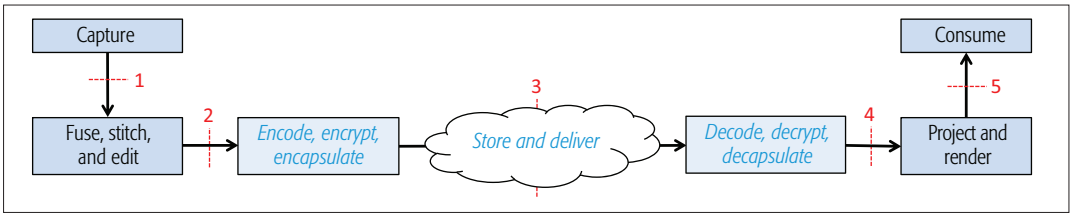

Figure 1. Basic system architecture and major interfaces for immersive media delivery.

Data Representations and Formats

JPEG[1] started an initiative called Pleno [5], which aims to define a standard framework for capture, representation, and exchange of images related to omnidirectional, depth-enhanced, point cloud, light field, and holographic modalities. Please note that a call for proposals might have a limited scope. Additionally, the JPEG XS requirements document references VR applications. JPEG XS is about a low-latency lightweight image coding system, which aims to support increasing resolutions (e.g., 8K) and frame rates in an efficient manner. Finally, JPEG recently created an Ad Hoc Group (AhG) on JPEG360 with the mandates to collect and define use cases for 360° image capture applications, develop requirements for such use cases, solicit industry engagement, collect evidence of existing solutions, and update descriptions of needed metadata.

MPEG[2] recently approved MPEG-I (or MPEG-i) as a new work item (ISO/IEC 23090 — coded representation of immersive media) that targets future immersive applications. MPEG-I will enable various forms of audio-visual immersion including panoramic video with 2D and 3D audio with various degrees of true 3D visual perception. It currently foresees five parts:

1. The first part will be a technical report describing the scope of this new standard and a set of use cases and applications from which actual requirements can be derived. Such technical reports are usually publicly available for free.
2. The second part specifies the omnidirectional media format (OMAF) addressing the urgent need of the industry for a standard in this area. The scope of OMAF comprises the coding, storage, delivery, and rendering of omnidirectional images and video and the associated audio. The standard will also specify media encapsulation and signaling with existing delivery formats such as MPEG-DASH [3] and MMT [7].
3. Part three will address immersive video coding.
4. Part four defines immersive audio coding.
5. Finally, the fifth part will contain a specification for point cloud compression, for which a call for proposals is currently available.

Additionally, MPEG established an AhG related to immersive media quality evaluation with the goal to review and document existing methods to assess human perception and reaction to immersive media stimuli. The mandate of the AhG also includes developing immersive media quality metrics and investigating their measurability in immersive media services. Finally, it will develop guidelines for evaluating quality of experience of immersive media services. It is expected that this activity will have strong links to the activities within the QoE cluster of SDOs.

IEEE[3] has started IEEE P2048 as a standard for virtual reality and augmented reality, specifically P2048.2/9 on immersive video/audio taxonomy and quality metrics — to define different categories and levels of immersive video — and P2048.3/10 on immersive video/audio file and stream formats — to define formats of immersive video files and streams, and the functions and interactions enabled by these formats. Interestingly, P2048.2 seems to be related to QoE aspects, which calls for interaction with other SDOs interested in such taxonomies and quality metrics. P2048.3 could benefit from MPEG standards, specifically OMAF but also extensions thereof. IEEE P2048 currently foresees eight parts where P2048.8 will define interoperability between virtual objects and the real world, which is probably closely related to MPEG-V (ISO/IEC 23005; media context and control) [6], which defines an architecture and specifies associated information representations to enable the *interoperability between virtual worlds*, such as the digital content provider of a virtual world, (serious) gaming, and simulation; and *with the real world*, including sensors, actuators, vision and rendering, robotics (e.g., for revalidation), (support for) independent living, social and welfare systems, banking, insurance, travel, real estate, rights management, and many others. Additionally, IEEE P3333.3 defines a standard for HMD-based 3D content motion sickness, reducing technology to resolve VR sickness caused by the visual mechanism set by the HMD-based 3D content motion sickness through the study of:

1. Visual response to the focal distortion
2. Visual response to the lens materials
3. Visual response to the lens refraction ratio
4. Visual response to the frame rate

Guidelines, System Standards, and APIs

The VR industry forum[4] has been established with the aim "to further the widespread availability of high quality audiovisual VR experiences, for the benefit of consumers" comprising working groups related to requirements, guidelines, interoperability, communications, and liaison. The current focus of the VRIF is on an end-to-end guideline adopting OMAF and DASH that primarily targets on-demand services and later will be extended to live services. Future guidelines may also include other distribution models such as broadcasting. Interestingly, the VRIF will also address security and privacy issues related to such services. The

[1] https://jpeg.org/

[2] http://mpeg.chiariglione.org/

[3] https://www.ieee.org/

[4] http://www.vr-if.org/

[5] http://dashif.org/

VRIF has some commonalities to the DASH-IF[5] but does not mandate a specific technology or standard.

3GPP[6] has completed its work on a technical report, "Virtual Reality (VR) Media Services over 3GPP," which provides a general introduction to VR, various use cases, audio/video quality evaluation, latency and synchronization aspects, and concludes with a gap analysis as well as various recommendations for future work. Additionally, 3GPP started new work items related to VR profiles for streaming media, enhanced voice services (EVS) codec extension for immersive voice and audio services, test methodologies for the evaluation of perceived listening quality in immersive audio systems, a new study item on QoE metrics for VR, and a new study item on 3GPP codes for VR audio. In general, 3GPP documents and work are publicly available.

DVB[7] started a so-called commercial module (CM) study mission group on virtual reality that has been promoted to an official group, CM-VR. The goal of this group is to deliver commercial requirements for relevant technical module (TM) groups that eventually will work on technical specifications to deliver VR contents over digital video broadcast (DVB) networks. The current focus of CM-VR is on use cases related to panoramic/3DoF+ as outlined within MPEG-I but continues exploring 6DoF uses within its study mission group. DVB will consider work done within other organizations such as MPEG, VRIF, and 3GPP.

The Khronos[8] group announced a VR standards initiative that resulted in OpenXR (cross-platform, portable, virtual reality) defining two levels of APIs for VR and AR, that is, an application interface and device interface, which assumes a vendor-specific runtime system in between the two interfaces. It again could benefit from MPEG standards in terms of codecs, file formats, and delivery formats. In this context, the World Wide Web Consortium (W3C)[9] WebVR already defines an API that provides support for accessing virtual reality devices, including sensors and head-mounted displays, on the web. Utilizing MPEG-DASH and WebVR currently allows the deployment of 360° services within HTML5 environments but requires further optimization in terms of delivery and consumption, including support for encrypted immersive media content.

Quality of Experience

QUALINET[10] is a European network concerned with QoE in multimedia systems and services. In terms of VR/360, it runs the Immersive Media Experiences (IMEX) task force to identify use cases and datasets/tools, and to develop a framework, methodology, and best practices for immersive media experiences. QUALINE*t al.*so coordinates standardization activities in this area and thus is considered as a kind of support organization for SDOs. In particular, it can help organize and conduct formal QoE assessments in various domains. For example, QUALINET members have conducted various experiments during the development of MPEG-H high efficiency video coding (HEVC).

The International Telecommunication Union — Telecommunication Standardization Sector (ITU-T)[11] started a new work program referred to as "G.QoE-VR," but no further information is available at the time of writing. In this context, it is worth mentioning that the Video Quality Experts Group (VQEG)[12] has an Immersive Media Group (IMG) with the mission of "quality assessment of immersive media, including virtual reality, augmented reality, stereoscopic 3DTV, and multiview." Interestingly, QUALINET and VQEG recently established a joint QUALINET-VQEG team on immersive media (JQVIM) with the following goals:

- Collecting and producing open source immersive media content and data sets
- Establishing and recommending best practices and guidelines
- Collecting and producing open source immersive media tools
- Surveying of standardization activities

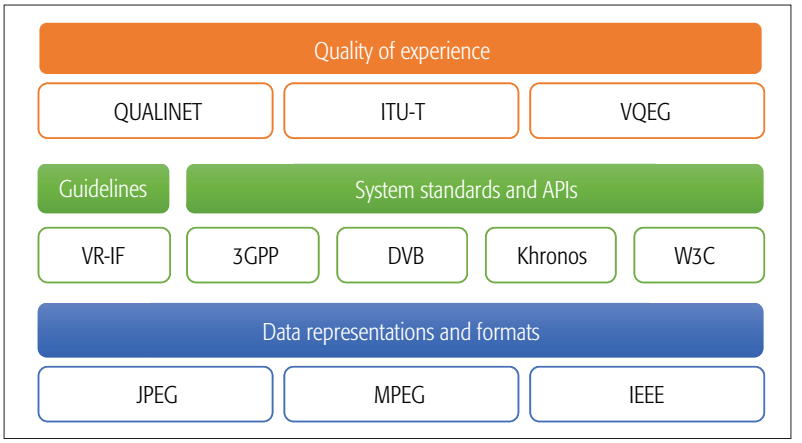

Figure 2. Overview of standardization activities.

Open Research and Standardization Issues

JPEG and MPEG seem to be slightly ahead in terms of standard development, which is appreciated in general as these SDOs typically provide the basic tools to be adopted by others (e.g., DVB, VRIF). It is expected that IEEE and others will catch up and hopefully benefit from what JPEG and MPEG is producing, leading to vital liaison activities.

Current deployments in this domain — mainly 360° video — rely on a single projection format adopting equirectangular projection, which is primarily considered within the first batch of standards to be published soon (i.e., MPEG OMAF). However, equirectangular projection — despite its simplicity and widespread adoption — is also known to be inefficient, specifically at the pole regions of images and videos, which results in bandwidth actually being wasted when delivering immersive 360° video content over the Internet. Thus, there is a need for more efficient projection formats offering the best or at least acceptable trade-off in terms of efficiency and usability for actual deployments.

Another open issue is related to the encoding and encapsulation for adaptive delivery. Approaches like viewport-adaptive streaming or tile-based streaming have been proposed in the

[6] http://www.3gpp.org/

[7] https://www.dvb.org/

[8] https://www.khronos.org/

[9] https://www.w3.org/

[10] http://www.qualinet.eu/

[11] http://www.itu.int/

[12] https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx

The user and quality of experience, including its assessment methodology, is not yet defined and many challenges needs to be addressed. An important aspect is the (non-) existence of a common dataset to be used for various dimensions including processing (fusing, stitching, editing), encoding, delivery, and rendering/consumption.

past, both offering pros and cons, which hampers actual deployments thereof. The optimal encoding and encapsulation options, including appropriate codec selection (e.g., HEVC vs. AV1), are still subject to research with the goal of optimizing network resource/bandwidth utilization while maintaining both user and QoE. In terms of encryption, the MPEG common encryption provides a vital tool that has been adopted widely, for example, within HTML5 environments with media source extensions (MSE) and encrypted media extensions (EME). However, when combined with WebVR, which is required for current 360° immersive media content, EME cannot be used as such as WebVR requires full access to the unencrypted media content for the actual rendering. This is a relevant business use case that needs to be addressed to enable the deployment of encrypted 360° immersive media content within HTML5 environments.

Finally, the user experience and QoE, including its assessment methodology, are not yet defined, and many challenges need to be addressed. An important aspect is the (non-) existence of a common dataset to be used for various dimensions including processing (fusing, stitching, editing), encoding, delivery, and rendering/consumption. The joint QUALINET/VQEG efforts are a first and necessary step toward addressing this issue.

In general, the overall goal of standards, as well as for immersive media delivery, is to specify the minimum required to enable interoperability and at the same time maximize flexibility to allow for innovative and competitive products and services (which is often a conflicting goal per se). In the end, and specifically for immersive media delivery applications and services, we should avoid finding us in a "situation"[13] or where the following quote becomes true: "The nice thing about standards is that you have so many to choose from" — Andrew S. Tanenbaum, *Computer Networks*

Acknowledgments

The author would like to acknowledge Touradj Ebrahimi (EPFL), Thomas Stockhammer (Qualcomm), Ludovic Noblet (b-com), Gilles Teniou (Orange), Rob Koenen (TNO), and Raimund Schatz (AIT) for providing additional hints and/or material on JPEG, DVB, 3GPP, VRIF, and ITU-T.

References

[1] R. Mohan, J. R. Smith, and Chung-Sheng Li, "Adapting Multimedia Internet Content for Universal Access," *IEEE Trans. Multimedia*, vol. 1, no. 1, Mar 1999, pp. 104–14.

[2] Sandvine, "2016 Global Internet Phenomena Report: Latin America & North America," 2016; http://sandvine.com/

[3] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming over the Internet," *IEEE MultiMedia*, vol. 18, no. 4, Apr. 2011, pp. 62–67.

[4] C. Timmerer, M. Graf, and C. Mueller, "Adaptive Streaming of VR/360-Degree Immersive Media Services with High QoE," *Proc. 2017 NAB Broadcast Engineering and IT Conf.*, Las Vegas, NV, Apr. 2017.

[5] T. Ebrahimi *et al.*, "JPEG Pleno: Toward an Efficient Representation of Visual Reality," *IEEE MultiMedia*, vol. 23, no. 4, Oct.-Dec. 2016, pp. 14–20. DOI: 10.1109/MMUL.2016.64.

[6] C. Timmerer *et al.*, "Interfacing with Virtual Worlds," *Proc. 2009 NEM Summit*, St. Malo, France, Sept. 28–30, 2009.

[7] Y. Lim *et al.*, "MMT: An Emerging MPEG Standard for Multimedia Delivery over the Internet," *IEEE MultiMedia*, vol. 20, no. 1, Jan.–Mar. 2013, pp. 80–85.

Biography

Christian Timmerer (christian.timmerer@itec.uni-klu.ac.at) received his M.Sc. (Dipl.-Ing.) in January 2003 and his Ph.D. (Dr. techn.) in June 2006 (for research on the adaptation of scalable multimedia content in streaming and constrained environments), both from the Alpen-Adria-Universitaët (AAU) Klagenfurt. He is currently an associate professor at the Institute of Information Technology (ITEC) within the Multimedia Communication Group. His research interests include immersive multimedia communications, streaming, adaptation, quality of experience, and sensory experience. He was the General Chair of WIAMIS 2008, QoMEX 2013, and MMSys 2016, and has participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, SocialSensor, COST IC1003 QUALINET, and ICoSOLE. He also participated in ISO/MPEG work for several years, notably in the area of MPEG-21, MPEG-M, MPEG-V, and MPEG-DASH, where he also served as a Standard Editor. In 2012 he cofounded Bitmovin (http://www.bitmovin.com/) to provide professional services around MPEG-DASH, where he holds the position of chief innovation officer (CIO). Subscribe to his blog (blog.timmerer.com) or follow him on Twitter (@timse7).

[13] https://xkcd.com/927/