VRIF Guidelines on Live VR Services

Ozgur Oyman (Intel), Mauricio Aracena (Ericsson), Tom De Koninck (TNO), Igor D.D. Curcio (Nokia Technologies), Thierry Fautier (Harmonic), Mick O'Doherty (Irdeto)

1. Introduction

Founded at the Consumer Electronics Show in 2017, the Virtual Reality Industry Forum (VRIF) (www.vr-if.org) is chartered with the mission to further the widespread availability of high quality audiovisual Virtual Reality (VR) experiences. Its charter includes delivering guidelines and advocating industry consensus around common technical standards for end-to-end VR interoperability from content production through distribution to consumption. VRIF delivered the first release of its Guidelines [1] in early 2018 detailing an initial set of best practices and recommendations for realizing high quality, interoperable 360 degree video experiences. Its latest Guidelines version 2.1 (2020) includes new features such as Live VR services, High Dynamic Range (HDR) support, use of text and fonts and watermarking for theatrical VR. Currently VRIF is working on new features around volumetric content production and distribution, cloud VR and edge computing, enhanced 360 video support and enabling social VR experiences.

For this year's invited contribution to the SMPTE journal, VRIF chose to address the theme of Live VR services. This has been a key area of focus in Guidelines development work in VRIF [1], which is contained in Section 2, with considerations around guiding use cases, end-to-end reference workflows, technical enablers around production and distribution, as well as industry deployment information based on TiledMedia and Intel Sports Live VR services. In Section 3, we highlight a few new features relevant for the future evolution of Live VR services and provide our conclusions in Section 4.

2. VRIF Guidelines on Live VR

2.1. Guiding Use Case

Live VR provides an immersive experience that certainly comes close to "being there". For many people, it is not possible to experience some live sports events or artistic performances, either because the "best seats in the house" are already sold or due to geographic constraints. A well-planned Live VR event places omnidirectional cameras at the prime viewing locations allowing the same viewing experience to be sold multiple times.

2.2. Reference Workflows

The Live VR workflow is depicted in Figure 1. It is considered that there are one or more supplementary live VR feeds for home delivery in addition to the main live feed that it is broadcast by traditional means such as terrestrial, satellite, broadband or cable delivery. The Live VR feed is captured using additional cameras dedicated for this purpose (besides ordinary cameras). The output of these cameras maybe already stitched or may require stitching in a secondary process. The stitching process may be supervised by a director at the outside broadcast (OB) truck or at the central studio at the main broadcast facility. In the venue, there might be several locations where the VR content is captured. The director may choose which Live VR feeds are distributed to home delivery, by producing one single VR feed, by giving the choice to the end user to select which camera, or by offering both a produced feed and individual camera positions.

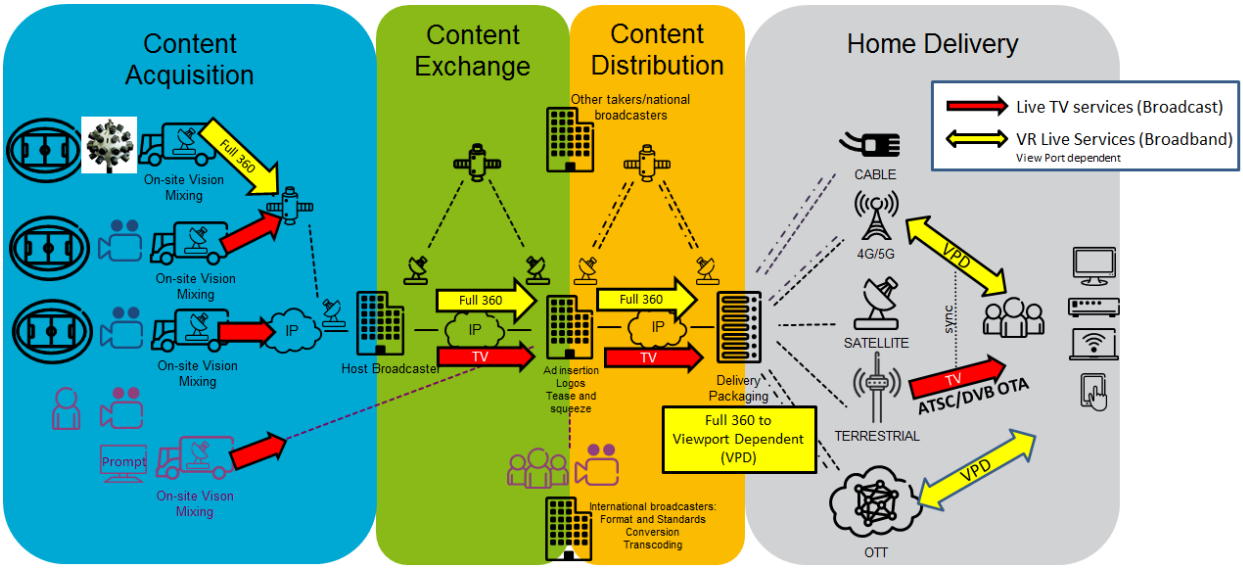

Figure 1: Live workflow for Television augmented with VR.

As in traditional Live TV, the content may be exchanged between other affiliates and even distributed to other regions including international destinations. This exchange occurs using the full 360° video content and is normally at very high quality, at least at 8K resolution and possibly even higher. The broadcaster has support for hybrid delivery of live TV content, that is, the main Live TV event is broadcast while the VR content is delivered via unicast-streaming service in parallel with the broadcast content.

Figure 2 depicts the comparison between a traditional live (linear) TV workflow and a Live VR workflow. It can be observed that similar workflows are utilized for traditional Live TV and Live VR. This should not be unexpected as the same operations of acquisition, production, processing and delivery apply for all entertainment experiences although the nature of the functions may differ.

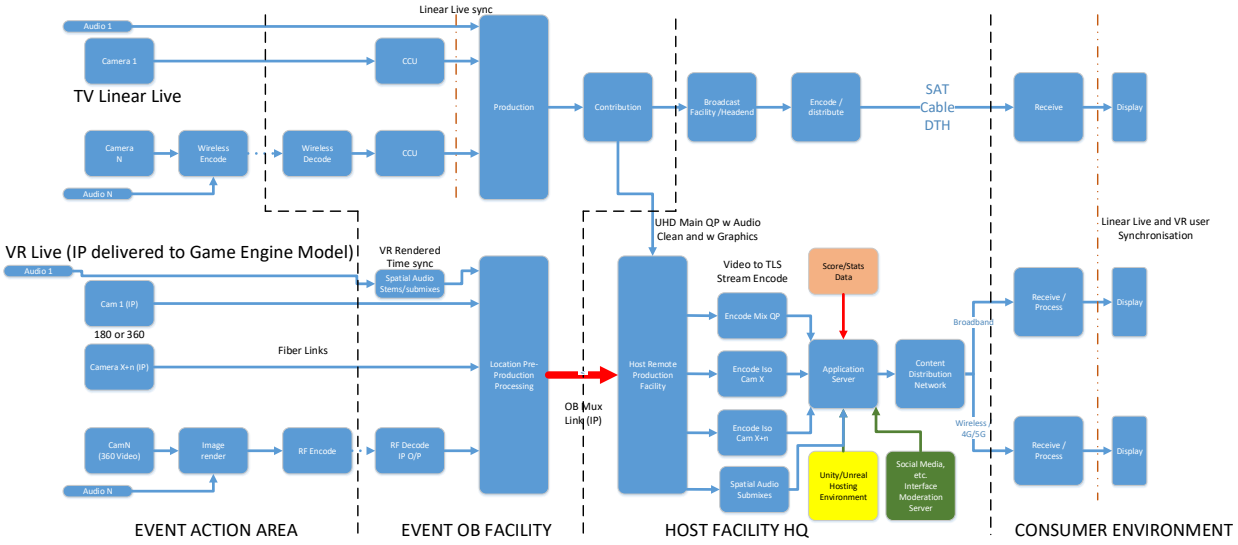

Figure 2: Linear TV and VR Live with - IP connected omnidirectional camera

Current broadcasting production workflows can be used for VR production by making use of an equirectangular projection in a vision mixer, for instance a 4K broadcast infrastructure will work if the full 360 VR video is captured at 4K resolution. If video is captured at 8K resolution or even higher, then VR and traditional TV Linear live workflows will start to differ.

2.3. Technical Enablers

2.3.1. Production Systems

From a Live VR 360 content production perspective, VRIF Guidelines [1] provide recommendations and best practices on several aspects such as camera placement, and impact of moving cameras, multi-camera shoots, LED Light Flicker and Direct sunlight. Different live VR 360 content capture scenarios are described including omnidirectional 360 camera with inbuilt stitching, omnidirectional 360 camera without inbuilt stitching and multiple omnidirectional 360 cameras with/without inbuilt stitching.

2.3.2. Distribution Systems

The 1st edition of MPEG's Omnidirectional Media Format (OMAF) specification in ISO/IEC 23090-2 [2] finalized in October 2017 has been the basis of the profiles addressed in VRIF Guidelines 1.0 and 2.0 [1]. These OMAF-based profiles, composed of a viewport-independent and a viewport-dependent profile, are equally applicable for both on-demand and live 360 VR content distribution.

2.3.3. Security

VRIF Guidelines [1] include security recommendations, the majority of which apply across both live and on-demand VR use cases. For example, traditional media security mechanisms, including encryption and watermarking, can be applied in the Live VR domain also.

2.4. Industry Deployment Information on Live VR Services

2.4.1. TiledMedia

A "VR camera" is usually a set of linked conventional video cameras that all record a piece of the environment. Stitching software then combines these video images into a single spherical video in the form of an equirectangular projection (ERP) (a flat map of the earth is also an ERP). A VR user only sees a small part of that sphere at any one point in time in their VR headset: about 1/8th of the complete picture. Since the video is played right in front of the user's eye, magnified by special lenses, the resolution needs to be very good. The industry standard in VR is to use 4K video for the entire sphere. 4K resolution in VR is 4 096 x 2 048 pixels, but the user will only see, approximately, a mere 1K x 1K per eye, right in front of them – not a good experience! To improve the quality-bandwidth tradeoff, TiledMedia uses the viewport aware delivery method in which only a portion of the 360 video corresponding to the viewport is delivered in high quality and this is achieved via tile-based delivery of the 360 video.

TiledMedia's 360 Live distribution system demonstrated at IBC 2019 and depicted in Figure 3 gives consumers an 8K experience while transmitting and decoding much less information; the platform sends only what a user actually sees. To make this possible, the image is cut up into around 100 high-quality tiles and only the actual tiles in view are sent, decoded and displayed. The real time nature of the experience, including the need to instantly respond to a user's head motion, makes this approach particularly challenging. When the user turns their head, the system retrieves new high resolution tiles, decodes and displays them – all within a tenth of a second. This happens so fast that users will hardly notice the low resolution background layer that is always present to prevent black areas from appearing in their field of view. The technical specification of the TiledMedia system is depicted in Figure 4 [1].

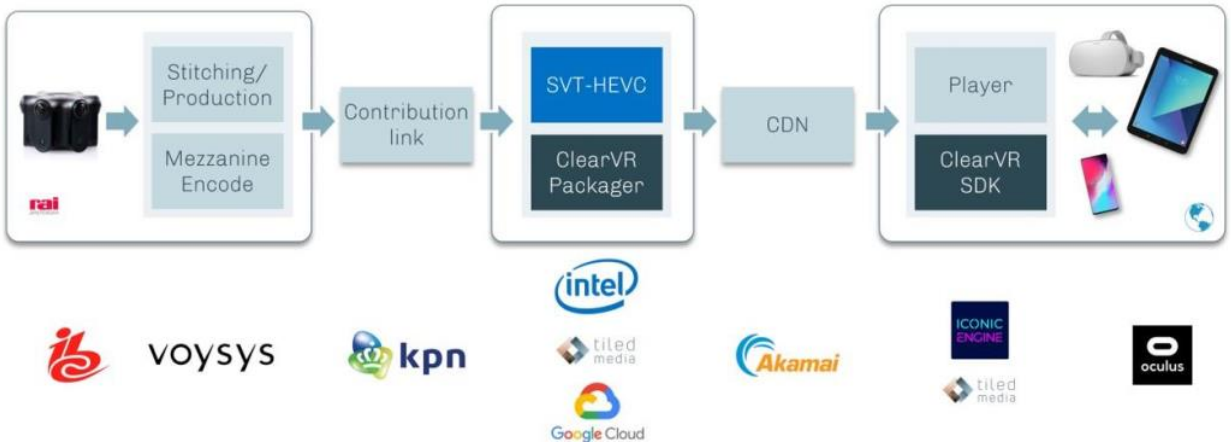

Figure 3:TiledMedia 360 live distribution system [1].

Number of cameras:	*two, user-switched*
Event Duration:	*5 days; 8+ hours per day*
Capture resolution:	*8 092 x 4 046*
Contribution bitrate:	*150 Mbit/s per camera (HEVC)*
Distribution resolution	*8 192 x 4 096*
Distribution Encoder:	*SVT-HEVC*
Distribution Format	*MP4-based ClearVR packaging*
Distribution bitrate:	*12 – 15 Mbit/s*
Streaming Protocol	*standard http/2 with multipart byterange requests*
User device decoder:	*HEVC Main level 5.1*
Glass-to-glass latency:	*~30 seconds*
Supported devices:	*Oculus headsets, iOS and Android tablets and phones*
Cloud processing:	*well over 1 000 Intel cores in Google Cloud Platform for the two streams (dynamically managed)*

Figure 4: Technical specs for TiledMedia 360 live distribution system [1].

2.4.2. Intel Sports

Intel's immersive media platform provides partners an end-to-end solution for processing immersive media content and delivering immersive media experiences to fans worldwide. During the recording of a sporting event, the capture systems, Intel® True View and Intel® True VR work independently or side-by-side at the stadium.

Intel® True VR depicted in Figure 5 is a transportable stereoscopic camera solution that can be set up in stadiums on game day. The stereoscopic camera pods are placed close to the action, and often coincide with the placements of traditional broadcast cameras. The cameras are placed in specific locations to enhance immersive media experiences. Intel® True VR outputs panoramic video from the stadium and sends it through the immersive media processing pipeline.

Figure 5: Capture system for Intel TrueVR [1].

Intel® True View depicted in Figure 6 is comprised of a camera array that is built into the perimeter of the stadium. The high-resolution cameras are angled to capture the entire field of play. In parallel or independently, Intel® True View outputs volumetric video through the same immersive media processing pipeline as Intel® True VR.

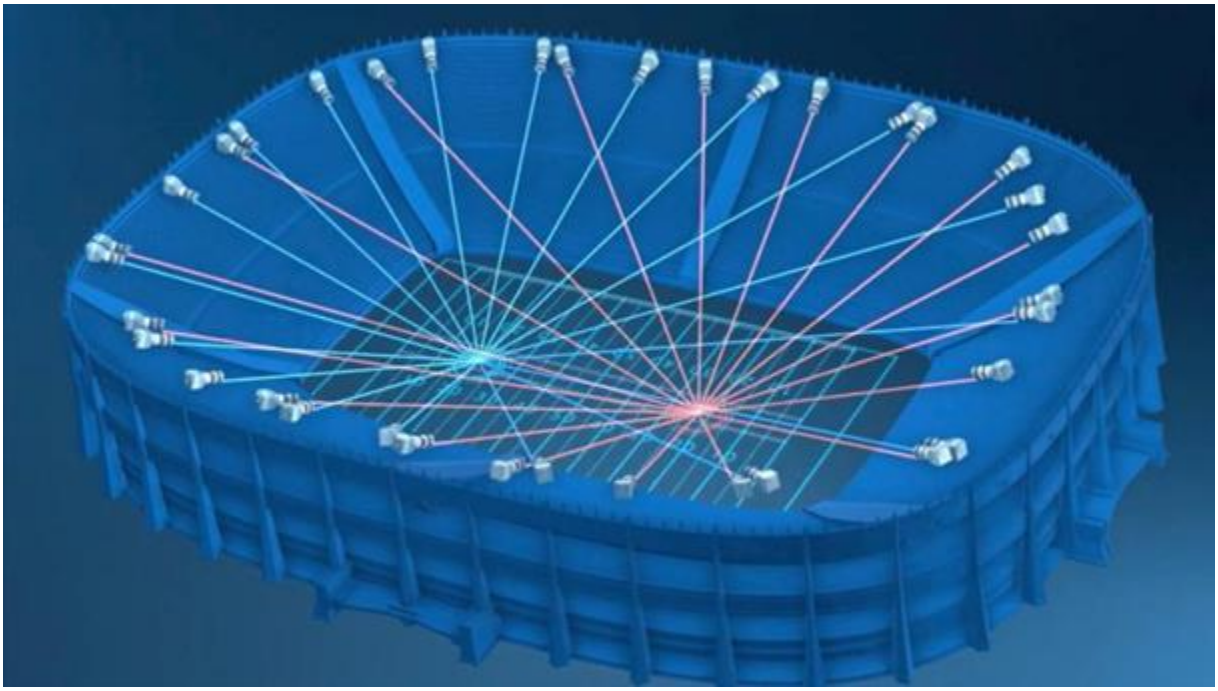

Figure 6: Capture system for Intel True View [1].

Once the content from either capture system is processed, the distribution pipeline enables delivering a catered stream for each supported device. The Intel Sports immersive media platform supports distributing both live and non-live content. The immersive media platform also supports 2D screen viewing experiences like mobile phones, traditional broadcast television, and web players as well as VR/AR/MR Head Mounted Devices (HMDs) and other immersive media streaming platforms. The corresponding immersive media platform architecture and workflow is depicted in Figure 7.

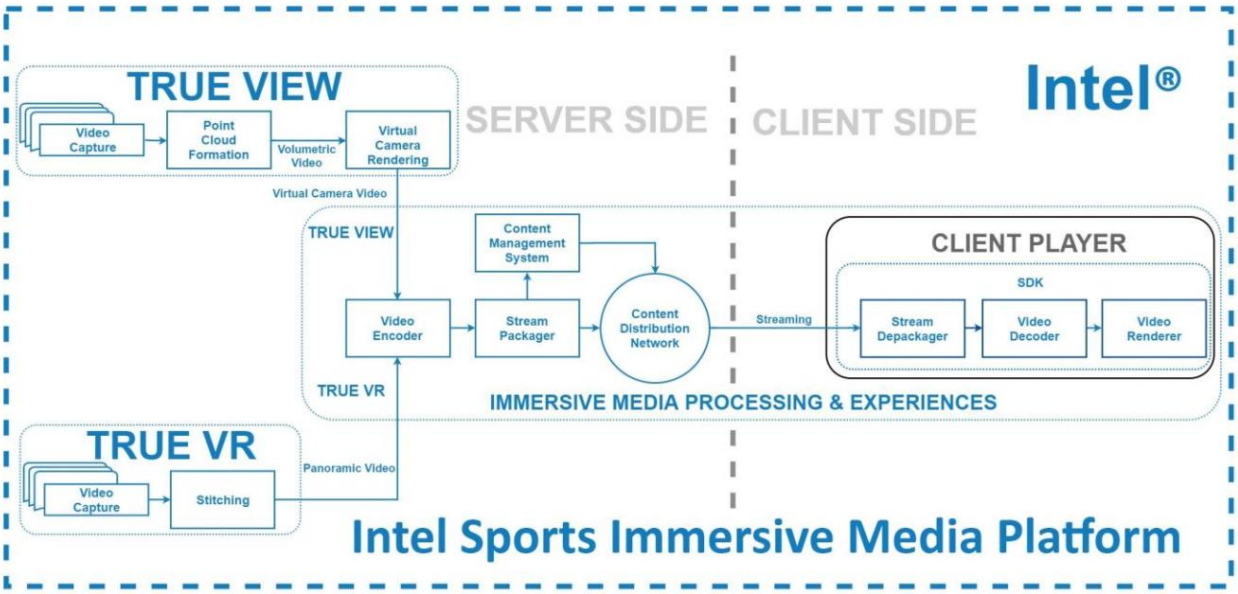

Figure 7: Intel Sports Immersive Media Platform [1].

3. Recent Advances on Live VR Services

3.1. Advances in Richer Live VR360 Experiences

MPEG will release the 2nd edition of their OMAF specification in July 2020 with additional features, such as multiple viewpoints, overlays and new tiling profiles [2]. Toward enabling richer live VR360 experiences, VRIF is working to develop recommendations around OMAF 2nd edition as part of its Guidelines 3.0 activity.

Multiple viewpoints can be thought as a set of 360º cameras which, for example, may be scattered around a basketball field (see Figure 8). The OMAF 2nd edition 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. This allows watching an event or object of interest from a different location and facilitates 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.

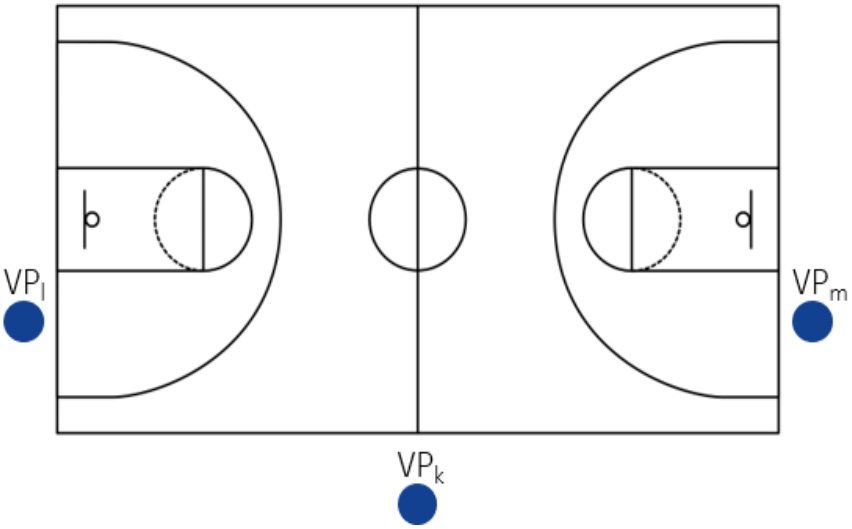

Figure 8: Example usage of multiple viewpoints in a basketball event.

Figure 8 illustrates an example of OMAF 2nd edition content with multiple viewpoints (VP_k, VP_l and VP_m). This allows the user to experience the action from different perspectives and facilitate interactive content consumption and creative storytelling [2].

Overlays are a way to enhance the information content of 360 video. They allow superimposing another piece of content (e.g., a picture, another video with news, advertisements, text or other) to be rendered on top of the main (background) omnidirectional video. Overlays also allow the creation of interactivity points or areas.

OMAF 2nd edition defines four different overlay types, based on their spatial position (see Figure 9, [3,4,5]):

- Overlays may be positioned on the users' viewing screen and always present on the users' viewport (viewport-relative or viewport-locked overlay);
- Overlays could be positioned at a depth from the user viewing position (sphere-relative 2D overlay);
- Overlays may be positioned over the background video without any gap between the two (sphere-relative omnidirectional overlay);
- 3D mesh at a given location within the unit sphere (sphere-relative 3D mesh overlay).

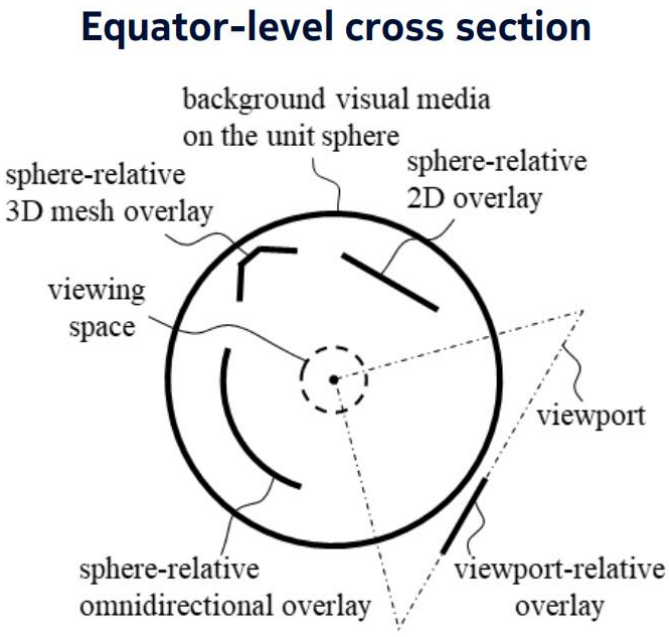

Figure 9: Different types of overlays defined in OMAF 2nd edition [4].

3.2. Volumetric Content and Point Clouds

Volumetric video has been recently gaining significant traction in delivering live VR experiences. Volumetric video contains spatial data and enables viewers to walk around and interact with people and objects, and hence it is far more immersive than 360 video footage because it captures the movements of real people in three dimensions. Users can view these movements from any angle by using positional tracking. 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. This is essentially the 3D equivalent of well-known pixels for representing 2D videos. These data points collectively describe the 3D geometry and texture of the scene or object. Such a volumetric representation leads to immersive forms of interaction and presentation with 6 degrees of freedom.

As part of its “Coded Representation of Immersive Media” (MPEG-I) project, MPEG is currently developing the “Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)” standard in ISO/IEC 23090-5 to compress point clouds using any existing or future 2D video codec, including legacy video codecs, e.g., HEVC, AV1, etc. [6]. An associated carriage format for V3C data is also developed by MPEG in specification ISO/IEC 23090-10 [7]. These standards are expected to be finalized and published before the end of this year. As part of its Guidelines 3.0 development, VRIF is currently considering these standards toward defining industry profiles for volumetric content compression and distribution.

3.3. Cloud VR and 5G

Higher bandwidths, lower latencies and support for edge computing enabled by 5G connectivity provide the desirable means to meet the high quality and interactivity needs of live VR experiences. 5G can support a wider range of QoS requirements addressing high bandwidth low latency needs of interactive Live VR applications, through a New Radio (NR) air interface as well as flexible QoS enabled via 5G core network architecture and network slicing. Moreover, the ability of the 5G system to leverage edge computing is essential to meet the performance requirements of Live VR, not only for better delivery performance via edge caching but also to offload some of the complex processing to the edge to perform various operations such as decoding, rendering, graphics, stitching, encoding, transcoding, etc., toward lowering the computational burden on the client devices. As part of its Guidelines 3.0 development, VRIF is developing the recommendations and best practices around such cloud VR capabilities.

In addition, relying on the higher bandwidths made available by 5G and the associated 5G HMD devices supporting 8K such as Qualcomm XR2, VRIF is developing a new distribution profile for viewport-independent delivery of 8K video content, that will match the same resolution per eye as achieved by 4K viewport-dependent OMAF profile today, while avoiding the complexities and costs associated with viewport-dependent processing and using off the shelf components such as DASH that has already a very wide industry support and its associated suite of features [8] . VRIF believes that this is a pragmatic means to increase the adoption of high-quality live VR experiences by leveraging 8K processing capabilities of new 5G mobile devices (phones and HMDs) expected to be available in the market by end of 2020.

3.4. Social VR

Communication, collaboration, teaching and remote support are new use cases for Live VR, which all have in common that multiple people have access to and interface with the virtual environment. Immersion and interaction are the main goals.

The virtual environment is most of the times a graphical 3D world, sometimes a real world captured in 360º. Current commercially available platforms use avatars to represent all users. Capture of gesture and pose is done using the VR controllers or sensors like the Leap Motion. In the case of volumetric representation (point cloud or mesh), a dedicated capture station is required, which uses one or several depth cameras like Microsoft’s Azure Kinect or Intel RealSense [9]. The platforms use high end game engines like Unity or web-based standards such as WebVR and WebXR [10], in combination with real-time communication technologies like WebRTC to ensure proper communication between users. Spatial audio is required to increase the realism in the virtual environment.

A specific use case, Immersive Teleconferencing and Telepresence for Remote Terminals (ITT4RT), where a remote participant joins a conference or meeting captured in real-time by 360 cameras, is developed by the Third Generation Partnership Project (3GPP) to specify

encapsulation and transport of VR360 content over RTP-based live streaming and real-time conversational applications [11].

3.5. Security Considerations

For Live VR services, which are sensitive to latency, meeting security targets while minimizing processing overhead and avoiding additional latency, both at the headend and in the playback client, requires careful engineering.

Typical Live VR examples include live sports, music and news events, and while some attacks may focus on capturing the entire 360-degree video, other attack vectors also need to be considered. For example, for a live sports event, there may be significant value in copying and redistributing only the user's field of view, if the user is following the main action in the event. Content protection technologies like encryption and watermarking must also work within the available latency budgets and may be tailored to protect either the entire 360 stream or just the user's field of view.

Client capabilities and delivery format are also important factors in assessing what security level a content provider can target in a given end to end VR Live solution. Understanding whether the playback client and device can support a secure media path, and whether that secure media path can support and, if necessary, manipulate the VR encrypted delivery stream(s) being delivered, again within the Live latency budget, is central to designing a secure Live VR deployment.

4. Conclusion

This paper presented recommendations and best practices on Live VR services based on VRIF Guidelines [1] along with two industry deployment examples. Furthermore, it addressed some of the recent advances toward future evolution of Live VR. VRIF continues to believe in eventual mainstream market adoption of Live VR services and products and is excited to develop new features as part of its Guidelines 3.0 development to further enhance Live VR experiences.

5. References

[1] VR Industry Forum Guidelines, available at: https://www.vr-if.org/guidelines/

[2] ISO/IEC 23090-2: "Information technology — Coded representation of immersive media — Part 2: Omnidirectional media format (OMAF)".

[3] I.D.D. Curcio, K. Kammachi-Sreedhar, S. Mate, "Multi-Viewpoint and Overlays in the MPEG OMAF Standard", ITU Journal: ICT Discoveries, Vol. 3(1), 18 May 2020.

[4] M. M. Hannuksela, Y.-K. Wang, "An overview of Omnidirectional MediA Format (OMAF)", submitted to Proceedings of the IEEE.

[5] K. Kammachi-Sreedhar, I.D.D. Curcio, A. Hourunranta, M. Lepistö, "Immersive Media Experience with MPEG OMAF Multi-Viewpoints and Overlays", ACM Multimedia Systems Conference, 8-11 June 2020, Istanbul, Turkey.

[6] ISO/IEC 23090-5: "Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)".

[7] ISO/IEC 23090-10: "Carriage of Visual Volumetric Video-Based Coding Data".

[8] DASH Industry Forum Guidelines: https://dashif.org/guidelines/

[9] VR Together: vrtogether.eu

[10] https://www.w3.org/TR/webxr/ ; https://labs.mozilla.org/projects/hubs/

[11] S4-200650: ITT4RT Permanent Document: https://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_108-e/Docs/S4-200650.zip