A Survey on Adaptive 360° Video Streaming: Solutions, Challenges and Opportunities

Abid Yaqoob, *Graduate Student Member, IEEE*, Ting Bi, *Member, IEEE*, and Gabriel-Miro Muntean, *Senior Member, IEEE*

***Abstract*—Omnidirectional or 360° video is increasingly being used, mostly due to the latest advancements in immersive Virtual Reality (VR) technology. However, its wide adoption is hindered by the higher bandwidth and lower latency requirements than associated with traditional video content delivery. Diverse researchers propose and design solutions that help support an immersive visual experience of 360° video, primarily when delivered over a dynamic network environment. This paper presents the state-of-the-art on adaptive 360° video delivery solutions considering end-to-end video streaming in general and then specifically of 360° video delivery. Current and emerging solutions for adaptive 360° video streaming, including viewport-independent, viewport-dependent, and tile-based schemes are presented. Next, solutions for network-assisted unicast and multicast streaming of 360° video content are discussed. Different research challenges for both on-demand and live 360° video streaming are also analyzed. Several proposed standards and technologies and top international research projects are then presented. We demonstrate the ongoing standardization efforts for 360° media services that ensure interoperability and immersive media deployment on a massive scale. Finally, the paper concludes with a discussion about future research opportunities enabled by 360° video.**

***Index Terms*—360° video streaming, virtual reality, HTTP adaptive streaming, MPEG-DASH, video tiling, viewport prediction, quality assessment, standards.**

I. Introduction

THERE has been a considerable increase in the production and use of omnidirectional or 360° videos, supported by both recent technological advancements in networking and computing and users' increasing interest to enrich their experience. Major video platforms, such as YouTube, Facebook, ARTE, and Vimeo, have put many efforts into promoting 360° video services. Although at the moment, most content relates to gaming and entertainment, an increasing number of 360° videos cover content from other applicability areas, including education, immersive telepresence, infotainment, documentaries, and sports among others [1].

A typical 360° video viewing arrangement involves a user interacting with the scene through a head-mounted display (HMD) device. Fig. 1[1] illustrates the virtual 360° environment surrounding a user and the current region (denoted as viewport), which the user sees at any moment in time. There are various HMD types, including Samsung Gear VR,[2] Oculus Rift,[3] HTC VIVE,[4] Google Cardboard,[5] Daydream,[6] and PlayStation VR,[7] etc. These HMDs differ in terms of their field of view (FoV): 100° (i.e., Samsung Gear, Google Cardboard, and Daydream) or 110° (i.e., Oculus Rift and HTC VIVE). Moreover, they can be standalone/mobile VR that do not require any external components (i.e., Samsung Gear VR, Google Cardboard, and Daydream) or tethered VRs that require a PC or a PlayStation (i.e., Oculus Rift and HTC VIVE). The modern HMDs can manage real-time immersive content due to the combination of powerful sensors and advanced display features. VR HMDs are likely to see broad adoption shortly, as forecast worldwide shipment is expected to be about 100 million HMD units by 2021, and about 50% of them are anticipated to be mobile headsets [3].

Related to the delivery of 360° video, there are viewport-independent solutions that stream all 360° content to users regardless of the viewport position, often wasting network resources. Viewport-dependent or human FoV-based schemes transmit the content within or surrounding the viewport, but use intensive computational resources to enable this. For all streaming types, 360° video requires a very high video resolution (i.e., larger than 4K) to provide high user Quality of Experience (QoE) levels. Netflix's recommended connection speed for streaming ultra HD video is 25 Mbps [4], but only less than 10% of global network connections have bandwidths higher than 25 Mbps [5]. Thus, the 360° video delivery to end terminals through diverse variable networks is highly challenging. The bandwidth requirements become even harder to meet when the same content is streamed to multiple VR clients.

Manuscript received February 18, 2020; revised May 5, 2020; accepted June 13, 2020. Date of publication July 3, 2020; date of current version November 20, 2020. This work was supported by the European Regional Development Fund through the Science Foundation Ireland (SFI) Research Centres Programme under Grant SFI/12/RC/2289_P2 (Insight Centre for Data Analytics) and Grant 16/SP/3804 (ENABLE). *(Corresponding author: Gabriel-Miro Muntean.)*

Abid Yaqoob is with the Insight Centre for Data Analytics, Dublin City University, D09 W6Y4 Dublin, Ireland (e-mail: abid.yaqoob2@mail.dcu.ie).

Ting Bi and Gabriel-Miro Muntean are with the Performance Engineering Laboratory, Dublin City University, Dublin, Ireland (e-mail: ting.bi@dcu.ie; gabriel.muntean@dcu.ie).

Digital Object Identifier 10.1109/COMST.2020.3006999

[1]The "Harbor" image has been sourced from [2] and has been used throughout the paper.

[2]Samsung Gear VR, http://www.samsung.com/global/galaxy/gear-vr.

[3]Oculus Rift, http://www.oculus.com/rift/.

[4]HTC VIVE, https://www.vive.com/.

[5]Google Cardboard, https://www.google.com/get/cardboard/.

[6]Daydream, https://www.google.com/get/daydream/.

[7]PlayStation VR, https://www.playstation.com/en-ca/explore/playstation-vr/.

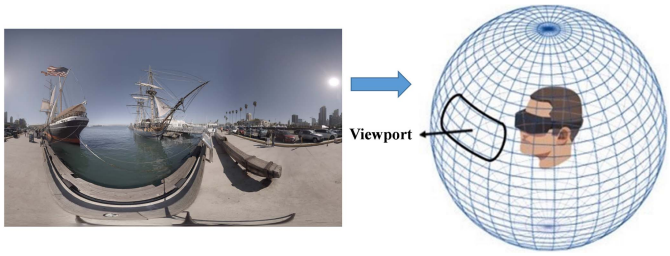

Fig. 1. 360° video viewing on a VR headset.

TABLE I
SUMMARY OF RELATED SURVEYS

Works	Year	Scope
[23]	2019	Considers both 360°image/video with a focus on processing, quality assessment, and compression techniques
[24]	2019	Deals with the different technologies and challenges considering 360° data model
[25]	2019	Focuses on visual distortion in 360° videos
[26]	2018	Targets underlying network-related issues to AR/VR systems instead of display and streaming strategies
[27]	2016	Mainly focuses on streaming of 360, panoramic, and multi-view videos

Following the standardization of MPEG Dynamic Adaptive Streaming over HTTP (DASH) [6], [7], 360° video adaptation can be performed similarly to the traditional video content. The adaptation is performed in a closed client-server loop by leveraging information reflecting the user's viewport position [8]. The 360° video is spatially partitioned into the desired number of non-overlapped rectangular regions known as tiles [9]–[11]. The tiled video is further split temporally into many fixed-duration segments. The client employs a smart algorithm to decide which tiles to fetch and at which quality level following variable head movements and network conditions. The requests for video segments have to be performed in advance to enable timely content delivery for the remote playout. The client performs estimations of the request-response time, which is a function of many factors, including network-related (e.g., latency, throughput, etc.) and content-dependent (e.g., segment duration, encoding rate, etc.). The client also estimates the future viewport position to pre-fetch the segments to be displayed next. As some discrepancy between the actual and predicted viewport positions is expected [12], some solutions [13]–[15] also stream tiled videos outside the predicted viewport area, including some schemes [8], [16]–[18] which use a lower resolution for those tiles to save bandwidth. Tiled videos are encoded and decoded independently. They provide the best rate-distortion (RD) performance in a high-level parallel environment.

When a user wearing an HMD moves his/her head, the viewport position changes, and the updated content should be displayed in real-time. Latency to performing the viewport switching severely influences the usage of 360° video display devices. This may cause problems for immersive video adoption, as users may experience degraded video quality and low responsiveness of their HMDs [19]. Quick response to viewer head movement has a substantial impact on viewer satisfaction when exposed to rich media content in an immersive environment than when presented with a traditional video. This is particularly important in a VR environment, where an inaccurate viewport delivery can cause motion sickness [20], making the 360° video streaming experience undesirable to the users [21].

Supporting high QoE levels is not trivial, and challenges include monitoring and responding in real-time to user's head movements, viewport identification and selection, dynamic content quality adjustment, employing efficient delivery protocols, etc. Several useful surveys have already focused on 360° video systems. Table I includes some of the recent surveys on state-of-the-art 360° videos. Xu *et al.* [22] surveyed various aspects of 360° content perception, coding, and quality assessment. The authors covered human visual attention based datasets for 360° content and presented several heuristic-based and data-driven approaches for attention modeling. Moreover, they overviewed different quality assessment and coding techniques for 360° content. Zink *et al.* [23] provided a survey on 360° video streaming systems with a focus on content creation, storage, distribution, and rendering techniques. The authors provided a general review of the QoE evaluation and edge-based 360° data distribution model. Azevedo *et al.* [24] reviewed the most common visual artifacts found in 360° streaming applications and characterized them as: spatial, temporal, stereoscopic, and head movement artifacts. The authors focused on quality assessment of 360° videos by considering the existing tools to find the significant reasons for 360° video distortion from an end-user perspective. He *et al.* [25] described some network-related challenges in AR/VR streaming and focused on the potential of underlying network issues. El-Ganainy and Hafeeda [26] discussed the different representation formats, QoE models, and streaming solutions to overcome the bandwidth limitations for VR content. However, despite their merits, none of these surveys focuses on the adaptive streaming of 360° video content.

This paper presents solutions proposed to support the adaptive streaming of 360° video content to enable an interactive and immersive user experience. It surveys several works on traditional video streaming in general and 360° video streaming in particular, outlining challenges and research opportunities. The major contributions of this survey paper are as follows:

1) presents major 360° video streaming stages, including content creation, projection, encoding, packaging, transmission, and rendering.
2) discusses relevant adaptive video streaming schemes and then focuses on adaptive 360° video approaches that dynamically adjust the size and quality of the viewport.
3) details the network-assisted transmission of high-resolution content to single or multiple users.
4) investigates main research challenges in adaptive 360° video streaming, including addressing viewport prediction, QoE assessment, and low latency streaming for handling both the on-demand and live 360° video streaming.
5) discusses important technical standardization efforts to allow interoperability and flexibility for immersive media services.

6) presents international projects which develop technologies or employ 360° videos.
7) identifies future avenues for 360° video research and development.

The rest of the paper is organized as follows: Section II reminds the reader about the evolution of video streaming standards. Section III illustrates the general 360° video streaming framework. Section IV provides major adaptive streaming schemes for traditional video content. Section V surveys various streaming schemes for 360° video, including viewport-independent, viewport-dependent, and tile-based solutions. Section VI covers the network-related solutions that offload the computational and storage tasks from end terminals to the nearby edges. Section VII identifies the challenges of performing adaptive streaming with omnidirectional video content. Section VIII illustrates the standardization and technological efforts in immersive media workspace. Section IX summarizes the potential of 360° videos in different sectors. Section X hosts the paper's final discussions and conclusions. Finally, Section XI summarizes the future directions related to multi-dimensions being surveyed.

II. Video Streaming Overview

The Internet, including most of the networks it consists of, relies on a best-effort data delivery approach. This approach is suitable for most content distribution services except for high-quality multimedia streaming applications. The real-time requirements are critical differences between multimedia and other data network traffic and require special attention. The concept of streaming has achieved considerable attraction due to the advancements in both network and video compression technologies. Both industrial and academic research development efforts have focused on proposing solutions for streaming multimedia from dedicated servers to remote users. As the goal was achieving high Quality of Service (QoS) levels, diverse Standards Developing Organizations (SDOs) such as International Telecommunication Union-Telecommunications (ITU-T), Internet Engineering Task Force (IETF), 3rd Generation Partnership Project (3GPP), European Telecommunications Standards Institute (ETSI), etc., have increased their activities to propose new technologies and protocols to support not only multimedia streaming; but also QoS-aware streaming.

Some of the early protocols which would support QoS-aware multimedia delivery designed on top of the Internet Protocol (IP) were Integrated Services (IntServ) [27] and Differentiated Services (DiffServ) [28]. The IntServ architecture includes models for representing service types and quantifying resource management. Resources are explicitly reserved for meeting application specifications and are carried out by a signaling protocol known as Resource Reservation Protocol (RSVP) [29]. IntServ uses RSVP to represent the QoS requirements of an application's traffic along with the end devices in the network. The IntServ/RSVP model ensures guaranteed and predictive services based on a quantitative specification of resource reservations. However, the IntServ per-flow QoS-support approach is challenging to scale. In contrast to IntServ, DiffServ is a straightforward and lightweight solution that performs differentiation per class of traffic. DiffServ architecture aims to facilitate higher than best-effort QoS and scales well to extended networks. Nevertheless, service providers' (SPs) QoS customization may affect the fine-grained end-to-end QoS.

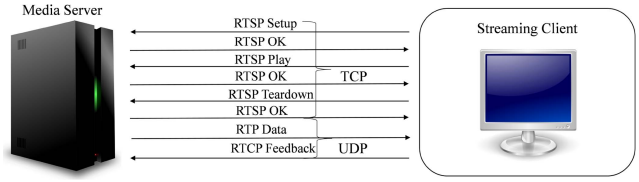

Fig. 2. Traditional RTSP streaming system.

Real-time Transport Protocol (RTP) [30] based on the simple User Datagram Protocol (UDP) [31] was proposed to standardized packet format and enabling real-time multimedia streaming. Two fundamental principles used to design RTP are application-layer framing and integrated layer processing. RTP Control Protocol (RTCP) [32] is also a UDP based protocol and monitors the transmission statistics and QoS levels. It achieves the synchronization across multiple streams. The multimedia servers in traditional streaming systems are controlled by a standard protocol known as Real-time Streaming Protocol (RTSP) [33], which first establishes a client-server connection before downloading the desired video file. The switching between multiple representations requires a connection re-establishment to download the desired quality from the beginning. Fig. 2 illustrates the RTSP streaming system where the state information of the streaming session is maintained throughout the session. It does not deliver the actual media that is the task of RTP. Contrarily, such protocols have some problems in traversing firewalls and Network Address Translations (NATs), and require a dedicated network infrastructure, increasing the additional complexity, and implementation costs.

Transmission Control Protocol (TCP) [34] allows for efficient data transmission but suffers from variable delays in return for its reliability. TCP proved to be beneficial by the researchers for delay-tolerant video transmission in the early 2000s. The rate fluctuation of TCP was compensated by introducing an application layer playout buffer. In early implementations, the design of HTTP over top of TCP allows the progressive download, where a constant quality video file is downloaded as quickly as TCP enables. The client plays the video before completing the download. A significant limitation is that different types of clients receive the same video quality over various network connections, which may cause rebuffering or unwanted stalls. This technique is unsuitable for mobile environments, where bandwidth varies more than in static environments [35]. The progressive download does not support live streaming [36] unless the video file is sliced into segments and the manifest supports it. This situation motivated the researchers towards the development of HTTP Adaptive Streaming (HAS). Fig. 3 shows a HAS communication system where the client uses the TCP as transport and HTTP as the application-layer protocol to pull the multimedia from a server, which is the host of the media content. With HAS, the video

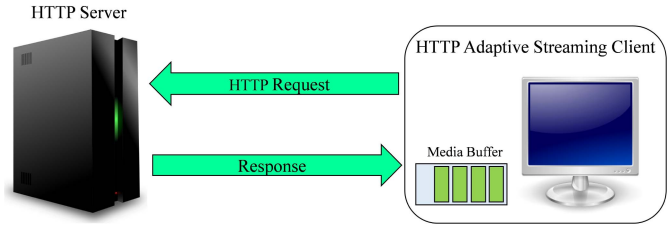

Fig. 3. Adaptive streaming over HTTP.

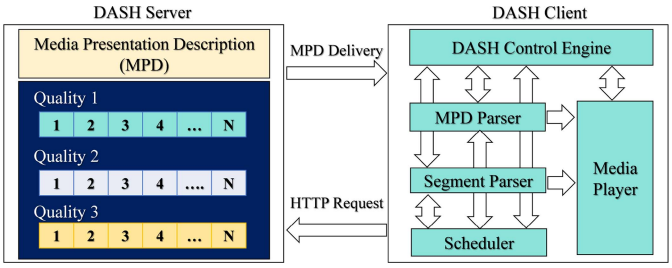

Fig. 4. MPEG-DASH client-server model.

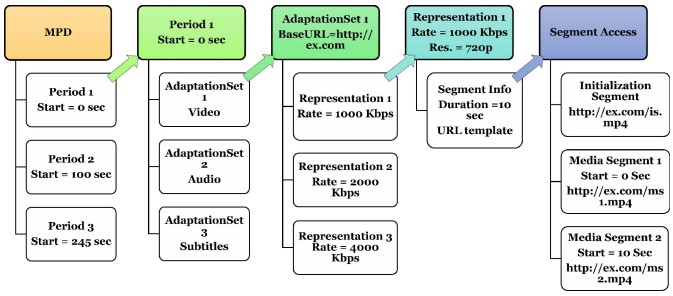

Fig. 5. Hierarchical data model of MPD.

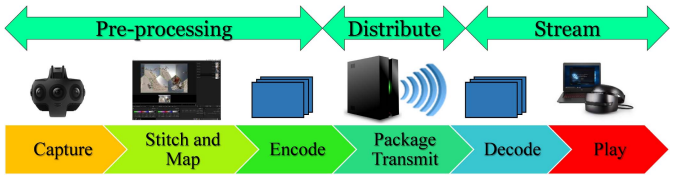

Fig. 6. End-to-End 360° video streaming framework.

bitrate of each chunk can be adapted according to the present network settings.

In 2012, the joint efforts of 3GPP and Moving Picture Experts Group (MPEG) [37] resulted in the emergence of codecs agnostic DASH standard. DASH uses an application layer adaptation algorithm to provide a seamless on-demand or live streaming to a wide range of devices over heterogeneous networks. The adaptation algorithms try to prevent playback interruptions while improving the streaming experience by adapting the segment bitrate in line with the ongoing network conditions. MPEG-DASH is well supported on existing HTTP infrastructure and minimizes the network load on the server-side, which was not the case with the previous protocols [38]. When employing a DASH-based approach, any video is prepared in several representations, encoded at different bitrates [39]. Each video representation is divided into several equal duration segments. Different bitrate segments are aligned such that the client can choose the appropriate video bitrate according to different network situations. The server stores the multimedia content information, and a client-side adaptation algorithm decides the bitrate of the next segment to be delivered [40]. The adaptation algorithm is not part of the DASH standard.

Fig. 4 shows an example of DASH-based client-server architecture, connecting a DASH video server to a DASH client. The server is responsible for pre-processing (e.g., encoding, segmentation, etc.), and storing of video segments and an XML file called Media Presentation Description (MPD). The MPD file contains the content information (e.g., metadata, mimeType codecs [41], [42], IP addresses, video profiles, and download URLs). The video information is described as one or more *Periods* in MPD. Each *Period* has one or multiple *AdaptationSets*, each including several video versions called *Representations*. Each *Representation* is characterized by bitrate, height, and width of the video components and includes multiple video *Segments*. The *Segment* is the primary content unit and can be accessed and presented to the end-user. The data model of MPD is demonstrated in Fig. 5.

The DASH standard does support not only the tiled video streaming but also allows the MPD elements association to the non-timed information. Spatial Relationship Description (SRD) is one of such syntax that streams the spatial sub-regions of a video with adaptive multi-rate streaming supported by DASH [43]. The MPD of MPEG-DASH is extended for SRD because of the spatial relationship between associated video areas. It allows the DASH client to fetch the video streams at suitable bitrates related to the user interest. The SRD syntax considers MPD authoring requirements and hybrid deployments with both legacy and SRD-aware DASH clients. This aspect supports multiple advanced use-cases outside the conventional DASH, i.e., interactive high definition streaming to mobile devices, high-quality zooming and navigation features, and streaming of wide-angle panoramic and 360° videos.

III. 360° Video Streaming Framework

Lately, there is a definite market move towards different forms of immersive video, including omnidirectional or 360° video. Cameras with high-resolutions are available for close-to and professional creation of 360° movies. The increased efficiency of stitching software allows for better content preparation than ever. Networks deliver high bitrate content to remote end-users devices. Devices themselves (e.g., HMDs, mobile phones, tablets, etc.) are equipped with robust sensors, processing, and display components that enable them to receive, process, and display rich media content. However, supporting real-life streaming of 360° videos is still very challenging because they are associated with vast amounts of data, and its handling is highly time-sensitive. Fig. 6 illustrates the major stages in 360° video streaming. Next, these stages are discussed in turn.

Fig. 7. Different projection patterns of 360° video (a) equirectangular, (b) cubemap, (c) pyramid, and (d) offset cubemap projections.

A. Capturing and Stitching

A multidimensional camera is used to capture the 360° scene of an environment. Several companies, including GoPro,[8] Insta360,[9] Ricoh Theta,[10] Samsung Gear,[11] etc., have started manufacturing Commercial-Off-The-Shelf (COTS) portable 360° cameras. In order to generate a full 360° view of the environment, there is a need for multiple inputs from different cameras (e.g., frame rate and aspect ratio). The aspect ratio is recommended to be 4:3 to capture the maximum possible spatial area. Modern applications are used to attach different views to get a full 360° view. The captured 360° view is then represented as a three-dimensional (3D) sphere.

B. Projection Schemes

After capturing and stitching steps; the 360° sphere representation is transformed into a planer format. Two main projection techniques are employed: (i) viewport-independent projection and (ii) viewport-dependent projection.

In *Viewport-Independent Projection*, the full 360° video is projected to a 2D plane in uniform quality, independent of the viewport. Examples include equirectangular projection [44] and cubemap projection [44]. An equirectangular projection (ERP) is the most known mapping technique for 360° videos. Fig. 7a shows an equirectangular projection frame. The world map is the most common example of this projection technique. Equirectangular projection can be represented as flattening a sphere around the viewer on to a two-dimensional surface using yaw and pitch values. The yaw values range from −180° left to 180° right, while the pitch values range from 90° top to −90° bottom. Several 360° video streaming services, including YouTube, Youku, and Jaunt VR, use the equirectangular projection. However, with this projection, the poles are represented with more pixels compared to the equator, possibly consuming the limited bandwidth of the user in less interesting regions. Furthermore, the compression efficiency is inadequate for this projection because of the image distortion.

A cubemap projection (CMP) is another popular mapping scheme. A six-sided cube combination is used to map the pixels of a sphere to the relevant pixels on the cube, as shown in Fig. 7b. This projection is widely used in gaming applications. Cubemap projection is more space-efficient and decreases the video size by 25% compared to an equirectangular approach [45]. Facebook widely used it for streaming 360° content and released the open-source code[12] to transform an equirectangular mapped video to the cubemap representation. A significant inefficiency of this scheme is that a limited user's FoV is rendered. Thus, wasting the associated transmitted data.

In *Viewport-Dependent Projection*, the viewing areas are represented with higher fidelity than other areas. Examples of such projections are the pyramid projection [45], truncated square pyramid projection (TSP) [46], offset cubemap projection [47], etc. In a pyramid projection, the 360° sphere is projected onto a pyramid. Fig. 7c represents a pyramid projection where the base part is considered as the viewing region and coded with the highest available quality. Most of the projected area belongs to the user's viewing direction. This approach decreases the size of the video by 80% [45]. The main drawback associated with this projection is that if users move their heads by 120°, the quality decreases aggressively

[8]https://gopro.com/
[9]https://www.insta360.com/
[10]https://theta360.com/
[11]https://www.samsung.com/global/galaxy/gear-360/
[12]https://github.com/facebook/transform360

as they rotate their heads by 180°. On the other hand, TSP projection reduces the amount of data over the edges and improves the streaming performance for high bitrate content. However, this projection approach involves sharp edges.

The offset cubemap projection is shown in Fig. 7d. It is similar to a conventional cubemap technique where the spherical pixels are projected to six faces of the cube, e.g., left, right, front, back, top, and bottom. However, it has an orientation where the viewing region associated with an offset orientation is represented in higher quality. It offers smooth quality variations. This projection has a strong storage overhead. With offset cubemap projection, multiple quality representations, e.g., 88 versions, are required to serve different bandwidth profiles [47].

C. Encoding of 360° Video Content

The compression efficiency has improved significantly following the development of next-generation coding solutions. Currently, traditional and 360° videos use the same encoders, including Advanced Video Coding (AVC)/H.264, High-Efficiency Video Coding (HEVC)/H.265, VP9, VP10, AV1, etc. The current generation 4K 360° video requires 10-50 Mbps bitrate, while next-generation and 6DoF 360° videos require 50-200 Mbps and 200-1000 Mbps bitrates, respectively [48]. Therefore, the efficient compression of 360° video is essential to be able to deliver it over the best-effort networks. AVC/H.264 standard uses a 16x16 macroblock structure for frame encoding. The size of encoded data is reduced through the motion prediction feature of the encoder. On the other side, HEVC/H.265 saves nearly 50% video bitrate compared to the AVC/H.264 with the same subjective quality. HEVC/H.265 supports the tiling feature for efficient video streaming [49], [50], where a 64x64 Coding Tree Unit (CTU) structure is used to encode each tile and achieves a higher compression ratio than AVC/H.264. The video tiles are physically separated and concatenated in a regular stream to decode them by a single decoder. The next-generation, Versatile Video Coding (VVC) standard [51] is expected to increase compression efficiency by up to 30% compared to HEVC/H.265.

D. Packaging and Transmission

The latest broad access to capturing and storage devices has shifted the production of rich media from studios to the large public. Regular users are not only consuming videos on smart devices, but they are also involved in creating and distributing the content. DASH is a modern standard to facilitate such over-the-top (OTT) services because of its compatibility and simplicity in the context of current network infrastructure. Fig. 8 illustrates the DASH-based adaptive streaming of 360° tiles. In this context, a client requires a high-quality subset of tiles according to the viewport position (red rectangle) and available bandwidth. The quality at which the client requests the tiles is dependent on network delivery.

360° videos are bandwidth-hungry streaming applications; demanding low response and buffering time. This is not always feasible, especially when there are far-located data

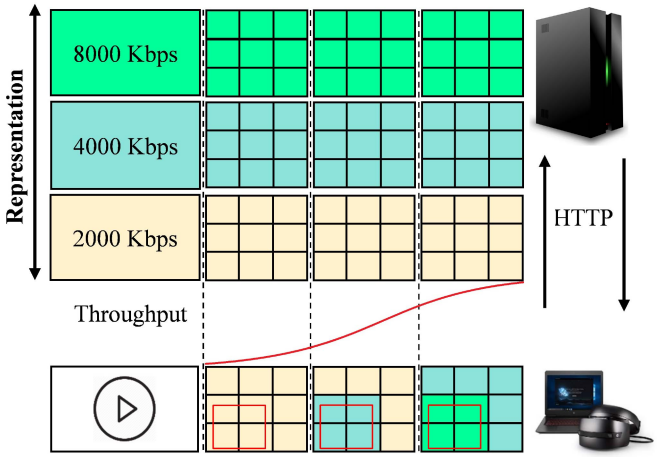

Fig. 8. Adaptive streaming of 3x3 tiles of 360° video with three segments and each containing one frame.

service points. Recently, several new solutions, i.e., micro data centers [52], fog computing [53], and mobile edge computing [54], [55], have emerged, intending to minimize the distance and number of network components between users and far located large processing centers (i.e., cloud). These technologies augment and extend the storage and computational resources from the current cloud model to the edges of the networks with objectives of low latency, efficient mobility, and real-time interactive support. Thus, the network transmission performance improves the user immersive experience and is considered an essential step towards achieving an experience-centric 360° video delivery.

E. 360° Video Rendering and Displaying

360° video rendering and displaying are computationally expensive due to the considerable processing power requirements. The client-based rendering solution is the most commonly used option in Web-based players like YouTube and Facebook, where the viewport is constructed from decoded frames and then presented to the end-user. A major problem is the wastage of computational resources at the client-side to process a big part of the 360° video content, which does not belong to the FoV of the user. Another 360° video rendering option is cloud-based rendering [56]. There is no need for content processing on the client-side. The video data is on the cloud, and only the requested FoVs are streamed to the end-users with no additional requirements of bandwidth or hardware resources.

360° video has become an integral part of multimedia applications. The consumer is always interested in the interactive and immersive streaming experience. The term “streaming” refers to the continuous delivery of audio-visual content to the user. Several standard organizations and industry forums are contributing many helpful insights for the future of 360° videos. In general, the architecture of the 360° video is still under progress. Security goals are nearly the same for both traditional and 360° content. These goals are targeted to protect unauthorized distribution, modification, resolution rights, and output control.

IV. Adaptive Video Streaming Schemes

The usefulness of multimedia services is strongly dependent on how efficiently the client can manage the available network resources. Adaptive video delivery solutions help to improve the streaming experience by dealing with different objectives, including video quality, energy consumption, load balancing, etc. on mobile, wireless, and wired access networks [57]. They adapt the transmitted video bitrate to match both the network conditions and the quality objectives. Based on the location of the adaptive controllers, the adaptive schemes from the literature can be divided into server-side and client-side adaptive schemes. Most server-side adaptive solutions require the client to send system and/or network information. Muntean *et al.* [58] proposed the Quality-oriented Adaptive Scheme (QOAS) to provide excellent perceptual quality of streamed multimedia content to end-users. QOAS includes a client-server architecture where the quality-oriented adaptation decisions are taken at the server-side. QOAS involve the adjustment of the streamed quality level of the multimedia streams according to the feedback on the perceived quality received from the clients. Yuan and Muntean [59] presented an intelligent Prioritized Adaptive Scheme (iPAS) for video delivery over IEEE 802.11 networks. A Stereotype-based Bandwidth Allocation module on the iPAS server is used to combine the QoS-related parameters and video content characteristics for priority classification of content and bandwidth share allocation. By differentiating the multimedia streams, the solution provides a prioritized allocation of the available wireless channel bandwidth. Zou *et al.* [60] introduced DOAS, a Device-Oriented Adaptive multimedia Scheme for LTE networks. This scheme was built on top of the LTE downlink scheduling mechanism. In contrast to existing server-based solutions, DOAS performs the adaptation specifically based on device characteristics t0 provide superior QoE to the multi-screen end-users.

Recently, client-side adaptive schemes have attracted essential consideration as the client-driven HAS streaming increases scalability by removing the session maintenance from the server-side [61]. The architecture seamlessly utilizes the existing HTTP delivery infrastructure (e.g., HTTP caches and servers). Significant efforts have been put in designing diverse adaptation schemes over the last several years, such as throughput-based, buffer-based, and hybrid adaptation approaches. Meanwhile, the extensive proliferation of wireless network access technologies and multiple network interfaces on modern devices prompt the network transmission performance over various access networks. Multipath based adaptive video streaming can dramatically improve QoE by providing additional communication support [62]. Next, the most representative adaptive streaming approaches are discussed.

A. Throughput-Based Adaptive Solutions

The schemes in this category select the video bitrates from the server based on the estimated network throughput. For video bitrate selection, HTTP clients evaluate the network throughput from previous observations [63]. Liu *et al.* [64] determined the throughput variations in a streaming session by measuring the segment fetch time (SFT), which represents the period between starting and receiving instants of HTTP GET request. The decision engine performs the adaptation decisions solely based on the measured throughput. In another work [65], the same authors considered both parallel and sequential segment fetching approaches in content distribution networks. The authors compared both the actual and expected SFTs to select the quality levels of future segments. However, the bitrate adaptation unit implements conservative bitrate increase and aggressive bitrate decrease policies, which significantly lowers the end-user satisfaction.

Jiang *et al.* [66] identified some bitrate selection issues when several HAS clients share a common congested bandwidth link. They studied the design space of adaptive algorithms related to three metrics (i.e., efficiency, stability, and fairness) and introduced an algorithm named FESTIVE, which explores a robust mechanism for segment scheduling, throughput estimation, and bitrate selection. FESTIVE contains a randomized scheduler to schedule the downloading of the next video chunks. The authors identified that a practical adaptive bitrate approach must try to avoid three main aspects when multiple clients share a full bandwidth link with capacity *W*, and every client x plays video bitrate $b_{x,t}$ at time t:

- *Inefficiency*: The multiple HAS clients must be able to select the highest possible representations to improve their experience. The inefficiency is defined as:

$$Inefficiency = \frac{\left|\sum_x b_{x,t} - W\right|}{W} \tag{1}$$

  A lower inefficiency value shows that multiple clients sharing a bottleneck link have the highest possible bitrates for effective bandwidth utilization.
- *Unfairness*: The available bandwidth should be equally distributed in a multi-client streaming environment. The unfairness is given as $\sqrt{1 - JainFair}$ [67]. Ideally, a small value of unfairness is desired which indicates that multiple clients have similar bitrates.
- *Instability*: Unnecessary bitrate switches can negatively influence the streaming experience. An instability metric is defined as:

$$\frac{\sum_{d=0}^{k-1}\left|b_{x,t-d} - b_{x,t-d-1}\right| * w(d)}{\sum_{d=1}^{k} b_{x,t-d} * w(d)} \tag{2}$$

Li *et al.* [68] introduced the Probe AND Adapt (PANDA) algorithm to examine the network state considering an average target data bitrate for future bitrate selection. PANDA aims to minimize bitrate oscillations by correctly probing the network when several HAS clients share a congested bandwidth channel. The performance evaluation against FESTIVE [66] and Microsoft Smooth Streaming [69] player shows that PANDA has the best adaptive behavior among these solutions, achieving the highest efficiency, fairness, and stability under different bandwidth and player settings.

Xiao *et al.* [70] analyzed that the overall streaming quality depends not only on the local measurements of the throughput, but also on the network capacity of the server. The authors utilized a server-initiated push mechanism to stream

the DASH content to mobile clients to lower the end-to-end latency. They also leveraged HTTP/2's stream termination feature to perform intermediate quality adjustments. The segment scheduling based on the estimated user's QoE, energy cost, and available resources offer improved streaming experience to users. While there is evidence that the performance is improved, this work was evaluated via simulations in a controlled LAN environment only. Sun *et al.* [71] proposed the Cross Session Stateful Predictor (CS2P), a data-driven throughput estimation scheme to overcome the inaccurate HAS traffic prediction problem. The authors first applied clustering approaches to the streaming sessions sharing similar characteristics, and then predicted corresponding throughput samples for each cluster using different Hidden Markov Models. An experimental evaluation with a large scale dataset reveals that CS2P efficiently estimates the available network throughput to improve the overall video bitrate adaptation. Similar to CS2P, some other solutions such as CFA [72] and Pytheas [73] also involve a data-driven controller to estimate the available throughput. However, these works do not support system heterogeneity and involve additional training complexity, making them less attractive.

A critical challenge in adaptive streaming is to estimate the available network throughput accurately [74]. An erroneous throughput estimation results in fetching incorrect quality segments and limits the overall QoE [75]. For example, an underestimation of throughput may lead to bringing lower quality segments, while an over-estimation can result in rebuffering. Employing throughput-based adaptation for 360° video without sophisticated throughput estimation mechanisms may lead to instability and poor QoE, especially under highly dynamic wireless and cellular networks [76].

B. Buffer-Based Adaptive Solutions

Buffer-based adaptive clients request the upcoming segments based on the present buffer occupancy during the video playback. To overcome the limitations of incomplete network information, Mueller *et al.* [77] presented a buffer based approach in combination with a client metrics toolset and compensation algorithm in a multi-client cache-enabled environment. The client-centric model effectively detects the bitrate switching period and compensates these switches by choosing appropriate video bitrates, resulting in up to 20% media bitrate increases. Despite the limited simulation analysis, the oscillation compensation approach described is very interesting. Huang *et al.* [78] proposed the Buffer Based Adaptation (BBA) approach for Netflix clients that drop the rebuffering events by up to 20% against the default algorithm. However, BBA considered a large buffer size, usually in minutes. Therefore, it does not guarantee the same promising performance for short videos. Spiteri *et al.* [79] proposed the Buffer Occupancy-based Lyapunov Algorithm (BOLA) that considers bitrate adaptation as an optimization problem associated with playback quality and rebuffering time. BOLA aims to avoid rebuffering by maintaining the buffer close to a set target level. For a sudden drop in buffer level, BOLA avoids the frequency of stalling events by requesting the lowest available video bitrate. The authors implemented BOLA using an open-source DASH reference player[13] and showed that it offers an adequate enhancement to the video quality with less probability of rebuffering.

To ease the buffer utilization for adaptation decisions, Beben *et al.* [80] proposed an enhanced Adaptation and Buffer Management Algorithm (ABMA+) that determines the download time of future representations based on the probability of rebuffering events. ABMA+ ensures a smooth playback by selecting the maximum bitrates based on pre-computed buffer size and segment download time. This buffer-based strategy results in a fine deployment with a low computational cost. Sieber *et al.* [81] proposed a Scalable Video Coding (SVC)-based approach, namely Bandwidth Independent Efficient Buffering (BIEB), to improve the video representations selection. BIEB fetches the video chunks based on layers distribution and thus maintains a stable buffer level to avoid frequent interruptions. However, BIEB does not consider stalls or quality switches in the QoE model. Furthermore, SVC-based streaming approaches involve additional coding and processing overheads. While video quality variation rate and playback stalls negatively impact the user's satisfaction, Tian and Liu [82] proposed a control-theoretic approach using a PID controller that enforces a buffer set-point to keep buffer to an optimum level. The algorithm decreases the video bitrate by a small margin to prevent the adjustments of unnecessary video bitrates.

Buffer-based adaptation approaches try to keep the buffer level to the stable state to avoid the risk of buffer underflow/underflow. DASH flows can experience high queuing delays (i.e., up to 1 second) and severe congestions, leading to buffer bloat problem [83]. This one-way queuing delay of 1s rigorously diminishes QoE of real-time multimedia services. This is a considerable problem because the active queue management (AQM) policies that aim to reduce network congestion do not adequately reduce this unwanted delay [83]. This concern can be even more critical for 360° videos due to their larger size than traditional flows, and the influence of different aggressive 360° HAS clients. The dynamic adjustment of the receive window size of the DASH client according to the queue size of the network device can significantly reduce the buffer bloat effect [84]. Moreover, an ample buffer space could not be feasible for a smooth 360° video streaming because of the high uncertainty of long-term viewport prediction. Usually, a small buffer (<3s) is reasonable under short-term viewport prediction [85]. However, there is a high chance of the playback stalling with short buffer space. Therefore, short duration segments can also be used for tile-based streaming to lower the risks of playback buffering. However, short segments have lower coding efficiency compared to long segments, especially for tile-based streaming [18].

C. Hybrid Adaptive Solutions

In this category of adaptive approaches, the client determines the video bitrate of upcoming segments considering both the throughput and playback buffer signals. Yin *et al.* [86] presented a control-theoretic approach called Model Predictive

[13]https://reference.dashif.org/dash.js/

Control (MPC) that utilizes the set of well-defined parameters for estimating the available network and buffer resources to optimally adjust the bitrate decisions for high QoE. The proposed QoE model employs the average video quality R_k, average bitrate switches, rebuffering events, and initial delay T_s components.

$$QoE_1^K = \sum_{k=1}^{K} q(R_k) - \lambda \sum_{k=1}^{K-1} |q(R_{k+1}) - q(R_k)| - \mu \sum_{k=1}^{K} (d_k(R_k)/C_k - B_k)_+ - \mu_s T_s \quad (3)$$

where C_k and B_k represent the available bandwidth and buffer occupancy for the kth chunk, respectively. The components' weights (i.e., λ, μ, and μ_s) depend on user interest and can be adjusted accordingly. MPC considers throughput estimation using a harmonic mean approach and can explicitly manage the complex control objectives. This work studied only a single-player, so there was no fairness consideration.

Yaqoob *et al.* [87] proposed a throughput and buffer occupancy-based adaptation (TBOA) approach to select the suitable video bitrates to achieve enhanced streaming experience in single and multiple-client environments. TBOA increases the bitrate aggressively to make efficient use of the available bandwidth. It also waits for the buffer to cross a certain level before decreasing the bitrate to obtain a steady performance. Miller *et al.* [88] proposed a hybrid approach that employs three thresholds for the buffer level, such that $0 < B_{min} < B_{low} < B_{high}$. The target interval B_{tar} is between B_{low} and B_{high}. However, the algorithm tries to stay at the optimum interval $B_{opt} = \frac{B_{low}+B_{high}}{2}$. By controlling B_{low} and B_{high} thresholds, the proposed solution tries to stable the buffer and bitrate variations in response to the unknown TCP throughput. The algorithm exhibited smooth and fair behavior but did not involve any user satisfaction metrics.

Vergados *et al.* [89] proposed a fuzzy logic-based DASH solution to control the rebuffering events and video streaming quality. The proposed solution considers the average throughput estimation approach and achieves higher video bitrates and fewer number of quality fluctuations against [64], [65], [82], [88]. However, unlike [86], this work does not consider QoE metrics. Sobhani *et al.* [90] addressed the shortcomings of existing throughput estimation approaches by employing the Kaufman's Adaptive Moving Average (KAMA) [91]. The authors proposed a fuzzy logic-based adaptation approach for bitrate adjustments using KAMA-based throughput measurements and Grey Prediction Model (GPM) [92] based buffer level estimations. The emulation performance under competing flows reveals that the proposed system has better fairness (50% on average) and better-perceived quality (17% maximum) compared to four alternative baseline approaches. Similarly, Wang *et al.* [93] introduced a Spectrum-based Quality Adaptation (SQUAD) algorithm to solve the inconsistencies of throughput prediction and buffer level estimation. Both throughput and buffer level feedback signals were used for appropriate quality selection. Initially, SQUAD fetches the lowest quality segments to lower the start-up time. The authors showed that SQUAD offers significantly improved performance regarding video quality switching frequency and magnitude. Unfortunately, none of the solutions discussed provides good balancing between video quality and bandwidth utilization.

While video quality variation rate and playback stalls negatively impact the user's satisfaction, Tian and Liu [82] proposed a control-theoretic approach using a PID controller that enforces a buffer set-point to keep buffer to an optimum level. The algorithm decreases the video bitrate by a small margin to prevent the adjustments of unnecessary video bitrates. However, the proposed solution does not ensure fairness among multiple competing clients, resulting in lower perception levels. Zhou *et al.* [94] proposed the Throughput Friendly DASH (TFDASH) to achieve fairness, stability, and efficiency among competing clients. The proposed model achieves the maximum and fair bandwidth utilization by avoiding OFF periods. In addition, the dual-threshold buffer ensures stable playback.

Adaptive video solutions require smart mechanisms for throughput estimation, fairly and efficiently utilizing the available network resources for quality adjustments, and maintaining sufficient playback buffer occupancy to avoid playback interruptions, etc. In a single-client environment, adaptive algorithms work reasonably well. However, multiple clients competing for the bandwidth quickly choke the entire network. When the client buffer reaches a maximum level, the client enters an ON-OFF phase, during which it may not correctly estimate the available bandwidth because every client will adjust the video bitrate without respecting others. This leads to bandwidth under-utilization and unequal bandwidth distribution among competing clients [95].

D. Multipath-Based Adaptive Solutions

The desire to deliver an increasing amount of high-resolution content across the existing heterogeneous networks (HetNets) has fuelled research in the field of rich multimedia transmission. Nowadays, multiple network interfaces in the user equipment (e.g., WiFi and LTE) can be leveraged for enhanced performance of time-sensitive applications (e.g., multimedia streaming, video conferencing, etc.), and for increasing the wireless availability and communication from always-connected state to always best-connected state [96].

For all innovative rich media solutions, an important challenge in the current network environment remains the delivery of an increased amount of content. A good solution is to employ multipath content delivery, as illustrated in Fig. 9. In this context, the direct employment of the Multipath Transmission Control Protocol (MPTCP) [97] helps but is not ideal because it requires kernel stack modification at both the sender and receiver terminals. Additionally, the MPTCP traffic may not pass through middleboxes as it is restricted by several network operators [98]. Other solutions such as CMT-QA employs specifically multiple network technologies, including cellular (e.g., LTE) and wireless broadband (e.g., WiFi), in order to enable concurrent multipath content delivery [99].

Chen *et al.* [100] proposed a multi-source player, called MSPlayer, to achieve high-quality video transmission over

TABLE II
360° VIDEO STREAMING TECHNIQUES

	Viewport-Independent Streaming	Viewport-Dependent Streaming	Tile-based Streaming
Definition	Transmit whole 360° data frame in equal quality	Viewing orientation based appropriate quality selection from the server	Streaming of 360° rectangular tiles in same or different qualities
Adaptation	Network-based bitrate adaptation	Network- and viewport-based bitrate adaptation, viewport size adaptation	Network- and viewport-based bitrate adaptation, Tiles/viewport size adaptation
Bandwidth Requirements	High	Medium	Medium
Latency Requirements	Similar to traditional videos	<20 ms	<20 ms
Projections	CMP, ERP	Pyramid, TSP, Offset cubemap, multi-resolution ERP/CMP [108], etc.	CMP, ERP, Pyramid, TSP, Offset cubemap, multi-resolution ERP/CMP, etc.
Cache efficiency	High	Low	Medium
Storage and Processing	Low	High	Medium
Streaming Extensions	Projection	Projection, viewing orientation	Projection, viewing orientation, DASH-SRD

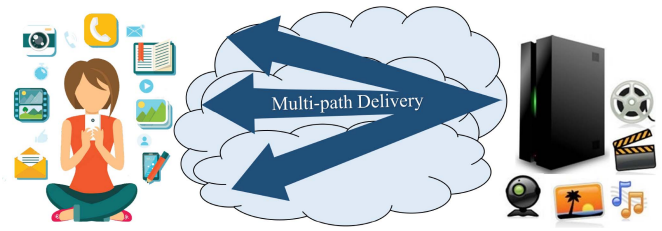

Fig. 9. Multipath wireless delivery in the heterogeneous network environment.

multiple links and resiliency in the case of failure. The client-driven bitrate allocation of future video segments depends on the estimated network conditions. After bitrate selection, the video segments are alternatively downloaded over the two available networks. However, downloading segments over different paths may cause out-of-order delivery. Xu *et al.* [101] analyzed the real-time quality of the data transmission paths by exploring the interaction between the data link layer and the transport layer and proposed a fairness-driven efficient flow control mechanism. The performance evaluation of the cross-layer fairness solution has been carried out against CMT-QA [99] considering average throughput, fairness, and PSNR metrics. Simulation results show that the cross-fairness solution attains higher fairness levels, but obtains lower average throughput and PSNR in comparison to CMT-QA. Kim and Chung [102] employed both the WiFi and LTE network interfaces to download partial segments from multiple video sources. The aggregated bandwidth of the paths is smoothed to avoid bandwidth fluctuations. The authors implemented a partial segment request strategy to avoid out-of-order-related problems. The partial segments transmitted over various paths are combined before they are presented to users.

Go *et al.* [103] considered networking cost constraints to schedule all the segments in a block with the same selected video bitrates across multiple networks. The experimental evaluation of the MPEG-DASH based streaming strategy under WiFi and cellular networks provide seamless video playback with low energy consumption for mobile devices. However, they did not analyze the impact of perceived video quality. Evensen *et al.* [62] extended the HTTP based streaming system, called DAVVI, to achieve multi-channel support over 3G and WiFi networks. The video segments are dynamically divided into subsegments based on the quality of each channel so that the maximum load could be applied to each channel. Using multiple network interfaces for multimedia delivery requires sophisticated mechanisms for path quality measurements and data scheduling to avoid packet loss and out-of-order delivery issues that can adversely affect user QoE. However, the existing solutions are limited in terms of measuring the real-time information of the paths due to the highly dynamic and complex nature of wireless heterogeneous networks.

Many solutions, capable of delivering high-quality video content, have been proposed to date. Elgabli *et al.* [104] considered two paths for SVC-based prioritized adaptive video delivery. The segments belonging to each layer can be transferred from one of the available routes based on the quality, chunk deadlines, and path preferences. However, the proposed work did not consider applying maximum contribution on any path. Zhang *et al.* [96] presented a priority-aware two stream-based adaptive solution that uses different video bitrates for each stream. The proposed scheme implements an integrated bandwidth approach to enable a higher video bitrate for the high-priority stream and terminates the low-priority stream when there is not enough bandwidth available. Yun and Chung [105] proposed a DASH-based streaming framework for multi-view videos, which includes a buffer-based server-push scheme and a parallel transmission mechanism to lower the switching time between different transmitted views. However, only a single path configuration is adopted in these works. Unlike [96], Rahman and Chung [106] introduced a HAS-based multi-view conference streaming solution where multiple streams of the presenter, audience, and presentation screen are transmitted concurrently over multiple paths. The proposed scheme assigns equal priority levels to all three streams. It employs a unified bandwidth approach so that a unified quality could be assigned to the segments of all the streams. For each segment of the multiple streams, the path is decided by considering network throughput and bitrate of the segments. Unfortunately, this work does not consider the influence of multiple channels, which may decrease the overall performance.

Taking advantage of multipath network characteristics and priority features [96], [104] for 360° tiled video streaming can provide improved streaming performance. All adaptive solutions presented in this section are generic, targeting standard video delivery. Although they can be employed

for omnidirectional video delivery, however, they were not designed to consider any specific aspects related to 360° video content.

V. Adaptive 360° Video Streaming Schemes

Streaming techniques for 360° videos have progressed from a full-view equal quality (viewport-independent) mode to viewport-only equal quality or full-view non-equal quality mode (viewport-dependent) and tile-based approaches. A 360° video is encoded as a full omnidirectional scene compared to a regular video encoding. Adaptive 360° video streaming leverages the DASH streaming framework for bitrate adaptation. Next, the most representative streaming schemes for 360° videos are discussed. They are also summarized in Table II.

A. Viewport-Independent Streaming

Viewport-independent streaming is the most straightforward way to stream 360° content because the whole frame is streamed in an equal quality similar to the traditional videos. The 360° sphere is projected/mapped using viewport-independent projection formats, e.g., ERP or CMP. The projected video after encoding is transmitted to the 360° client that does not require any orientation information from HMD sensors. The client should be able to support projection formats. Accordingly, the DASH client performs bitrate adaptation similar to a traditional video, i.e., the representations of the same projection format for upcoming segments are requested based on network characteristics. Hence, a minimal DASH extension (e.g., projection related metadata) is required to support equal quality streaming. Afzal *et al.* [108] performed an experimental characterization of thousands of 360° YouTube videos by directly fetching the complete frame regardless of the current viewing direction of a user. The authors found that 360° videos has about six times higher bitrates, multiple resolution formats, and reduced motions compared to regular videos. Viewport-independent streaming is mostly applied to stream sports [109], education [110], and tourism [111] content.

The implementation simplicity has become a pleasant introduction to viewport-independent streaming. However, it has 30% less coding efficiency compared to viewport-dependent streaming [112]. Moreover, it requires extensive bandwidth and decoding resources for invisible areas. By employing viewport-dependent streaming, these resources could be saved and adequately used for visible content.

B. Viewport-Dependent Streaming

In viewport-dependent streaming, the end devices receive only the certain video frame areas, which contain the visual information equal or greater angle of the viewport. The end devices should detect the related viewport in response to the user head movement and send the interaction signals to the cloud or edge servers to precise the player information. Such solutions are adaptive as they dynamically perform their area selections and quality adjustments, reducing the transmitted bitrate during 360° video streaming. In this regard, several adaptation sets associated with the user's orientation are prepared at the server-side. A client decides which adaptation set to fetch according to the network and estimated viewing position. However, these adaptive solutions require smart mechanisms of viewing region identification, synchronization with user head movement, and quality adjustment, etc., to keep providing smooth playback experience. Several works [47], [107], [113] focus on providing better coding efficiency and resource management without affecting the viewport quality.

Sreedhar *et al.* [107] implemented and compared multi-resolution variants of ERP and CMP, and the existing variants of pyramid projection, e.g., rhombic pyramid, square pyramid, and a truncated pyramid. The authors showed that the proposed multi-resolution variants for viewport-dependent projection schemes give the best RD performance compared to the pyramid formats. Zhou *et al.* [47] analyzed different projection schemes for viewport adaptive streaming using Oculus HMD. The authors showed that the Offset cubemap projection strategy results in a 5.6% to 16.4% average gain in visual quality. The proposed framework adapts the size and quality of the viewport based on the available network resources and future viewing position. This two-dimensional adaptation strategy could download over 57% additional chunks spending 20% extra network bandwidth when compared to an ideal downloading procedure.

The high-quality streaming of the whole omnidirectional sphere is not a smart idea due to limited network resources. Corbillon *et al.* [113] described a practical approach to produce differentiated quality segments for viewport-dependent streaming. They proposed the Quality Emphasized Regions (QERs) strategy to scale the resolution of certain regions when a limited number of representations are available to stream. In order to improve the viewport quality, He *et al.* [114] performed a network response based joint adaptation of the viewport size and bitrate under congested network conditions. The simulation results based on NS-3[14] show that dynamic viewport coverage offers better picture quality when compared with transmitting the full 360° view. Moreover, the network response-based rate adaption also ensures improved video quality when adjusted based on overall traffic variations. Naik *et al.* [115] performed two subjective experiments by smartly employing asymmetric qualities for both background and foreground views of stereoscopic videos. The authors demonstrate that the proposed strategy could save up to 15% and 41% bitrate for both background and foreground tiles, respectively.

In viewport-dependent adaptive streaming, the client performs adaptation based on the network characteristics as well as the viewing orientation of the user. Therefore, the DASH manifest should also include the viewing position information in addition to the projection metadata. These approaches have substantial storage requirements because several different content versions are stored at the server-side to support suitable viewport adaptation. Moreover, these approaches seem to be less cache-efficient and need resource-intensive encoding, which can be expensive, particularly for commercial and live

[14] https://www.nsnam.org/

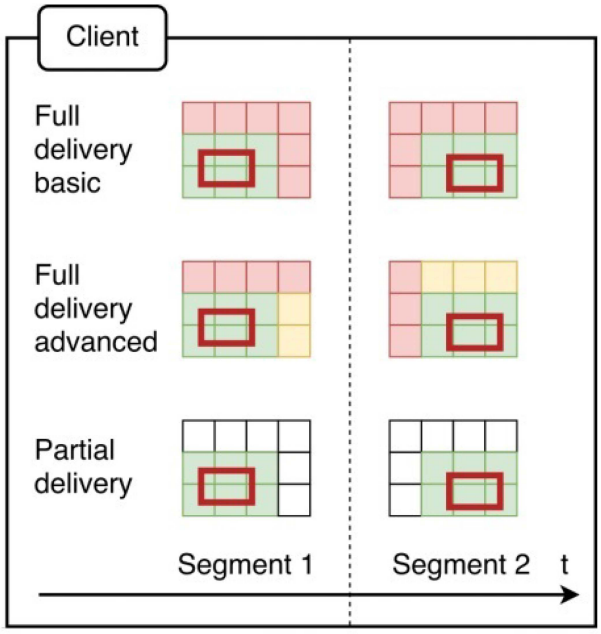

Fig. 10. Different Tiling options for adaptive 360° video streaming [8].

streaming services. When the user changes his/her viewing position, the viewport resolution can be adjusted by replacing the old variant with a new high-resolution option only at the next available random access point (RAP). Accordingly, there is a need to optimize the RAP period to optimize the viewport resolution adaptation. Furthermore, low latency and active viewport switching are necessary when considering viewport-dependent streaming.

C. Tile-Based Streaming

In traditional HAS, any video is segmented into small chunks for adaptive transmission to utilize the available bandwidth wisely. In 360° videos, the chunks are further partitioned into equal/non-equal rectangular tiles to precisely adjust the quality of the viewing tiles. Graf *et al.* [8] investigated three tiling strategies, i.e., full delivery basic, full delivery advanced, and partial delivery, using state-of-the-art video codecs to efficiently save the resources for unviewed part of 360° videos. The selection of different tiling strategies for two consecutive segments based on different viewing regions is shown in Fig. 10. The evaluation of different tiling patterns, e.g., 1x1, 3x2, 5x3, 6x4, and 8x5, against monolithic streaming shows that a 6x4 tiling scheme implements a useful trade-off between bandwidth consumption and coding efficiency. Furthermore, the full delivery basic streaming under different bandwidth settings achieves a bandwidth saving of around 65% compared to full delivery advanced and partial delivery strategies. Zhou *et al.* [116] proposed a cluster-based approach, namely ClusTile, where the tiles with the minimal bandwidth requirements are streamed to overcome the coding efficiency and computation overheads. ClusTile achieves a bandwidth saving of up to 72% and 52% compared to the traditional and advanced tile-based streaming approaches, respectively. The disadvantage of this approach is that cluster-based tiles selection may lead to inappropriate tiles selection when a discrepancy occurs between actually viewed and downloaded tiles during the streaming session.

Ghosh *et al.* [117] proposed to download the surrounding and far away tiles at the minimum available quality. The predicted viewport quality was adaptively upsampled. Experiments confirmed that variable qualities for both the viewport and surrounding regions boost QoE levels by 20% in comparison to other algorithms. Ozcinar *et al.* [14] introduced an adaptive 360° video streaming framework that utilizes the visual attention metric to compute the optimal tiling patterns for each of the frames. Then for each of the selected patterns, a non-uniform bitrate is assigned to the tiles belonging to different regions. The bitrate selection entirely depends on the estimated viewport and network conditions. However, the proposed framework struggles to optimize the viewport quality as a large portion of bandwidth is utilized for transferring non-viewport tiles.

Xie *et al.* [85] proposed an optimization framework for tile-based streaming to minimize the pre-fetched tile error and improve the smoothness of tile borders with different associated bitrates. Two QoE functions are defined with the objectives to minimize the expected quality distortion $(\Phi(X))$ and spatial quality variance $(\Psi(X))$ of the viewport when considering the viewing probability of the tiles. These functions are defined as follows:

$$\Phi(X) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} D_{i,j} \cdot x_{i,j} \cdot p_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{M} x_{i,j} \cdot s_i} \tag{4}$$

$$\Psi(X) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} x_{i,j} \cdot p_i \cdot \left(D_{i,j} - s_i \cdot \Phi(X)\right)^2}{\sum_{i=1}^{N}\sum_{j=1}^{M} x_{i,j} \cdot s_i} \tag{5}$$

A target-buffer based adaptation approach is applied for a smooth playback under a small buffer with the need for short-term viewport prediction. At adaptation step k, when the kth set of segments are downloaded completely, the buffer occupancy b_k is given by:

$$b_k = b_{k-1} - \frac{R_k \cdot T}{C_k} + T \tag{6}$$

To prevent running out of chunks, the buffer occupancy is controlled by setting a target buffer level B_{target}, that is, $b_k = B_{target}$. The average spatial quality variance is 0.97, which is smaller than other tile-based strategies. The proposed probabilistic adaptive framework achieves around 39% gains on perceptual quality with 46% on average lower spatial quality variance.

Van der Hooft *et al.* [118] divide the 360° frame into viewport and non-viewport regions. The proposed solution first selects the lowest quality level for both regions and then increases the quality of viewport tiles. If the bandwidth is still available, the quality allocation is repeated for the remaining tiles. Unfortunately, the presented works [14], [117], [118] do not consider the viewport prediction errors when adjusting the viewport bitrate. These heuristics attempt to aggressively increase the viewport quality based on the available bandwidth. Instead of completely relying on bandwidth estimations, Nguyen *et al.* [119] proposed a new adaptation mechanism that dynamically decides the viewport bitrate considering both the

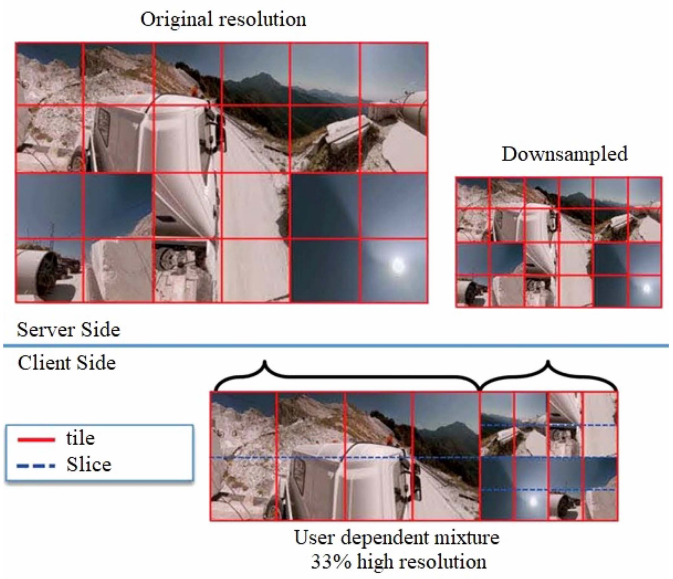

Fig. 11. Different resolutions of HEVC motion-constrained tiles [120].

predicted head movements and viewport errors during each segment duration. Unlike [8], [14], [117], which cover extension tiles in all directions, their proposed scheme jointly adapts the coverage and bitrate of the extension tiles. The experimental evaluation under diverse recorded user head movements demonstrates an increase in the viewport quality without acquiring excessive bandwidth utilization for non-viewport regions.

The SRD extension of DASH provides an association between various versions of the tiles to achieve higher bitrate savings. Le Feuvre and Concolato [11] employed the MPEG-DASH SRD feature. They introduced different priority setups for both independent and motion-constrained HEVC tiles to enable an efficient implementation of the tile-based approach. The authors developed a DASH client using the GPAC Open Source multimedia framework [121] to perform tiled streaming with configurable adaptation parameters. Influenced by the bandwidth problem for the interactive videos, D'Acunto *et al.* [122] proposed an MPEG-DASH SRD approach to facilitate smooth streaming of zoomable and pannable videos. The low-resolution tiles are always downloaded to avoid rebuffering when the user navigates the view. The current viewing region is upsampled and presented to the user to support a high-quality zooming feature. The authors implemented their design in a JavaScript-based SRD video player.[15] Hosseini and Swaminathan [15] proposed SRD-based prioritized streaming of viewport, neighboring (maximum 8 tiles), and the rest of the tiles. They constructed a 3D geometry with six 3D meshes to smoothly represent the tiles in 3D space. The authors showed that differentiated quality streaming results in a bandwidth saving of 72% in comparison to a baseline approach. Kim and Yang [123] used an enhanced version of MPEG-DASH SRD to choose between quality variable tile layers. The researchers designed and implemented a multilayer rendering enabled 360° VR player based on their previous work [124] to support high-resolution and low latency streaming for highly unpredictable head motion data.

In HEVC, the motion-constrained tileset (MCTS) [125] is an adjacent division of the whole frame represented as sub-videos and provides decoding support for a freely-selected tileset. Zare *et al.* [16] employed the MCTS concept for panoramic video streaming. They partitioned two different video quality versions to tiles and streamed the viewport tiles in original captured resolution and remaining tiles in a lower resolution. It has shown that variable bitrate for selected tiles reduces 30% to 40% bitrate. Similarly, Skupin *et al.* [120] presented a tile-based variable resolution streaming system using an HEVC encoder. The cubemap 360° video is tiled into 24 grids; each representing a separate bitstream. Two different quality versions are streamed to the client, i.e., eight tiles in high-quality and sixteen tiles in low-quality, as shown in Fig. 11. Son *et al.* [126] implemented the MCTS-based HEVC and scalable HEVC (SHVC) codecs for independent tile extraction and transmission in viewport-based mobile VR streaming. The proposed method achieves more than 47% bandwidth saving. However, the proposed design underperforms in comparison with the original HM and SHM encoders because MCTS restricts the temporal motion information. Lee *et al.* [127] encoded the 360° tiles with the MCTS technique and streamed the mixed quality video tiles to the end-user using a saliency detection network. The improved usage of MCTS with the saliency model enables flexible decoding support for the region of interest tiles without any added complexity.

Scalable video coding is an alternative strategy to achieve viewport adaptation. The base layer is always required and can be prefetched at the client-side to avoid rebuffering events. Enhancement layers increase the viewport quality and can be requested when sufficient bandwidth is available. Additionally, SVC facilitates an efficient in-network caching support to reduce the distribution cost when multiple clients request the same content [128]. Nasrabadi *et al.* [76] used a scalable coding scheme to solve the rebuffering issues for 360° video streaming. However, this method suffers from severe quality fluctuations because it does not involve any mechanism to deal with viewport prediction errors. Nguyen *et al.* [129] suggested using SVC by incorporating viewport prediction to overcome the randomness of both the network channels and head movements. The proposed tile layer updating and late tile termination features can improve the viewport quality by 17% as demonstrated by the experiments.

Reinforcement learning (RL) [130] for traditional video streaming [131], [132] adjusts efficiently the video bitrate and achieves long-term QoE rewards. Different from the traditional video content, 360° video includes several new aspects such as tiles size [133], viewport prediction, etc. Applying existing RL adaptation policies directly to 360° video streaming may lower streaming performance. Fu *et al.* [134] proposed a sequential reinforcement learning approach for 360° video streaming, called 360SRL, that makes adaptation decisions based on the rewarded QoE of previous decisions instead of estimated bandwidth. 360SRL uses a tile-based streaming simulator to boost

[15]https://github.com/tnomedialab/dash-srd.js

the training phase. The trace-driven evaluation demonstrates that 360SRL outperforms baseline adaptation approaches by achieving 12% QoE improvement.

Jiang *et al.* [135] also leveraged RL for the bitrate selection of viewport and non-viewport tiles based on historical bandwidth, buffer space, tile sizes, and viewport prediction errors, etc. The architecture of the proposed system consists of state buffer, viewport prediction (VPP), and tiles bitrate selection (TBS) agents. The state buffer provides the user viewing patterns and network states to the VPP and TBS agents. The VPP agent then estimate the next viewport position by employing an LSTM model. The TBS agent is trained by the Asynchronous Advantage Actor-Critic (A3C) [136] algorithm to perform suitable bitrate decisions. Quan *et al.* [137] analyzed user QoE by extracting pixel-wise motion through a Convolution Neural Network (CNN) and used it to group tiles dynamically to provide an important balance between the video quality and encoding efficiency. Next, the authors used a RL-based adaptation agent which intelligently adapts the quality of each tile to the dynamic environment. The validation of the proposal using real LTE bandwidth traces demonstrates superior performance in terms of perceived quality while also saving bandwidth resources.

Deep learning enables RL to optimize an aggregated reward further using multi-faceted state and action spaces [138]. Kan *et al.* [139] and Xiao *et al.* [140] designed a deep reinforcement learning (DRL) framework that adaptively adjusts the streaming policy based on exploration and exploitation of environmental factors. Both solutions perform the bitrate decision with the A3C algorithm of DRL due to its effectiveness in making agents more and more intelligent. The performance evaluation reveals that the proposed systems balance various QoE metrics, including average visual quality, average quality fluctuations, and rebuffering, among others. Similarly, Zhang *et al.* [141] proposed a DRL model that dynamically learns to adapt the bitrate allocation using the LSTM-based ACTOR-CRITIC (AC) network considering viewport prediction accuracy and network conditions. Real-word and trace-driven evaluations show that the proposed scheme adapts well to a broad set of dynamic features and offers a 20% to 30% improved QoE reward compared to the legacy methods.

Tile-based streaming requires a low number of content versions at the server-side. It incorporates lower storage and processing overhead compared to viewport-dependent streaming. Most proposed schemes [16], [76], [117], [120], use different resolutions for viewport and adjacent tiles. This can reduce the bandwidth cost for efficient streaming. However, the different resolution tiles can significantly lower the perceived video quality in case of wrong viewport prediction. In a subjective experiment with 50 users, Wang *et al.* [142] showed that most of the users observed significant quality degradation when mixing 1920x1080 resolution tiles with 960x540 resolution tiles. However, the users noticed a small difference when mixing 1920x1080 resolution tiles with 1600x900 resolution tiles. This mixing effect leads to even severe quality degradation for high motion content. Therefore, in addition to dynamically perform the tiles selection [14], [133], and DRL-based bitrate adaptation [139]–[141], there should be an appropriate choice of streaming resolutions to attain a perfect trade-off between the streaming quality, spatial quality variance, viewport prediction errors, and bandwidth efficiency.

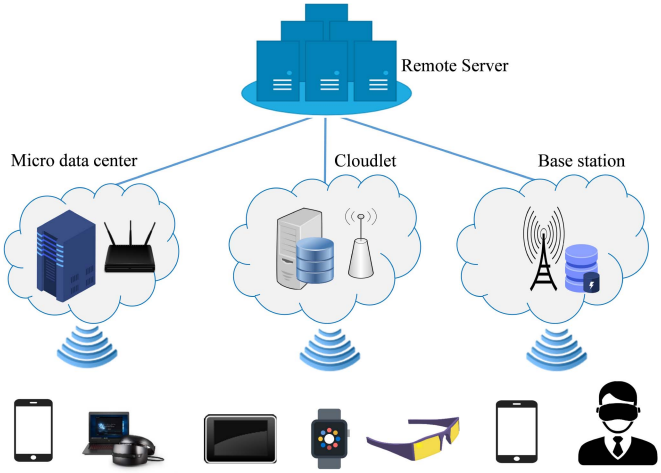

Fig. 12. Architecture of mobile edge-assisted streaming of 360° video.

VI. Network-Related Solutions

Emerging immersive and interactive user applications with higher bandwidth, QoS, and computing requirements, are among the applications which would benefit the most from the 5th generation (5G) networks [143]. The traditional network architecture with centralized cloud-based computing and storage is not adequate for real-time high bitrate content delivery. Edge caching and mobile edge computing (MEC) are regarded as pivotal enablers for 360° video services [144]. Next, we discuss some of the most recent edge-assisted and cooperative transmission approaches for 360° video.

A. Edge-Assisted Streaming of 360° Video

The massive video content can be transferred to the edge nodes and to downstream clients to meet the high-resolution levels and stringent latency requirements by the management of short-range communication. In edge computing [145], [146], the processing and storage tasks are shifted from the core network to the edge nodes such as base stations (BSs), cloudlets, micro data centers, set-top boxes, headsets, etc., with significant advantages in comparison to traditional networks. Fig. 12 provides an architecture of edge computing and edge cache-enabled wireless networks for 360° video streaming.

Hou *et al.* [147] investigated some fundamental trade-offs between local devices, remote-edges, and cloud-servers for the rendering of viewport-only video, full 360° video, and 6-Degrees of Freedom (6DoF) video. The authors proposed that edge/cloud servers rendering can make the computations to be lighter and enables the feasibility and portability for wireless VR/AR experiences. Zhang *et al.* [148] proposed a hybrid edge cloud infrastructure for VR multiplayer gaming, where central cloud updates the global game events and edge cloud manages view updating and massive frame rendering tasks to

support a large number of online gamers with low end-to-end delay. They further presented a server selection algorithm that ensures fairness among VR players based on the QoS and mobility impact of players. In contrast to [147], [148], Lo *et al.* [149] considered device heterogeneity for the edge-assisted rendering of 360° video. The edge server transcodes the HEVC tiles stream into the viewport video stream and transmits to multiple clients. Their optimization algorithm dynamically decides which client should be served by edge nodes according to the video quality, HMD types, and bandwidth constraints for the enhanced QoE of VR users.

Caching solutions for conventional video content [150], [151] cannot be readily deployed for caching of 360° video. To facilitate the transmission of 360° video in an edge-cache enabled network, a proxy cache between the two transmission ends is deployed to make the content available near the user. Edge caching can substantially decrease duplicate transmissions and make the content servers more scalable [152]. Mahzari *et al.* [153] introduced a popular content (e.g., FoV) caching policy for 360° video based on the watching behavior of other users. Experimental evaluation with open-source head movement traces of 156 users [154] shows superior performance in terms of cache usage when compared to Least Frequently Used (LFU) and Least Recently Used (LRU) caching policies with at least 40% and 17% improvements, respectively. Similarly, Papaioannou and Koutsopoulos [155] proposed a tile resolution and demand statistics-based caching policy to improve the viewport coverage with the minimum error between requested and cached tiles versions. The experimental evaluation with different caching and transmission delays results in improvements of the cache hit ratio, especially for the layered-encoded tiles.

Edge caching can be performed at the Evolved Packet Core (EPC), which might cause a suboptimal performance because the packet size is very small. An alternative way is to cache data at the Radio Access Network (RAN). However, it is more complicated due to the tunneling and packaging of data. Liu *et al.* [156] deployed a tile-caching approach for mobile networks at both RAN and EPC to save transmission bandwidth subject to the constraint of video streaming latency. The cache nodes for EPC and each RAN are deployed in the Packet Data Network Gateway (P-GW) and eNodeBs, respectively. The content controller entity in EPC is responsible for improved cache utilization for tiles content. This joint tile-caching design can significantly reduce the bandwidth pressure for the backhaul network with excellent scalability.

To leverage the collaborative transmission opportunities, Maniotis *et al.* [157] proposed a tile-level video popularity-aware caching and transmission in cellular networks containing a Macro-cell Base Station (MBS) and multiple Small Base Stations (SBS). They employed advanced coding schemes to create a flexible encoded tile structure and enabled cooperative caching of tiles in each SBS. This cooperation allows storing only the likely to be watched tiles at SBSs, while the other tiles can be fetched over the backhaul link. Chen *et al.* [158] proposed an echo-liquid state DRL model for joint caching and distribution in a scenario where the captured content is transmitted from Drone Base Stations (DBS) to small BSs using high-frequency millimeter wave (mmWave) communication technology. To meet the instantaneous delay target, BSs can cache some popular content from the data. However, extensive deployment of small BSs consumes a substantial amount of energy. In contrast to the computation-constrained MEC architecture, Yang *et al.* [159] exploited caching and computing resources in a communication-constrained MEC architecture to lower the requirements of communication-resources. However, this kind of architecture needs a resource-intensive task scheduling to balance the communication cost and the delay. Chakareski [160] explored the state-of-the-art caching, computing, and communication (3C) for VR/AR applications in a multi-cell network environment. The proposed framework allows BSs to exploit appropriate computation and cache resources to maximize aggregate reward. However, the authors focused on caching/rendering only, without considering the user's viewing experience and processing time, which can significantly reduce VR quality of experience.

Sun *et al.* [161] took advantage of both FoV caching and necessary computing operations ahead of time at the end terminals to save communication bandwidth without sacrificing the response time. For homogeneous FoVs, the joint caching and computing framework perform the best decision about caching and post-processing steps. For heterogeneous FoVs, the authors applied a concave-convex expression to obtain attractive results. Rigazzi *et al.* [162] proposed a three-tier (i.e., 3C) solution based on an open-source project, Fog05, to distribute the intensive tasks (e.g., coding/decoding and frame reconstruction) across cloud, constrained fog, and edge nodes. The 3C solution facilitates system scalability, interoperability, and lifecycle maintenance of 360° video streaming services. Experimental evaluation shows a significant reduction in bandwidth, energy consumption, deployment cost, and terminal complexity. Elbamby *et al.* [163] presented a joint framework for interactive VR game applications by employing proactive computing and mmWave transmission under latency and reliability constraints. The proposed framework precomputes and stores video frames to reduce the VR traffic volume. The evaluation demonstrates that the proposed joint policy can reduce up to 30% end-to-end delay.

Edge computing provides some important benefits to support high-resolution and high interactive VR video delivery over limited bandwidth networks, including:

1) *Latency Reduction:* In general, the cloud alone cannot satisfy the requirements of all latency-sensitive applications, as it is usually far away from user devices. Edge computing enables collaborative computing where users can access a shared pool of servers. This design enables meeting the latency requirements of 360° video applications [21], [146], [164].
2) *Lower Energy Consumption:* Computation offloading to distributed computing clusters according to network architecture and resource provisioning improves significantly the energy performance of mobile devices [165], [166].
3) *Load Management:* Edge caching provides means to store the content near the users, i.e., BSs, small

TABLE III
SUMMARY OF EDGE-ASSISTED SOLUTIONS FOR 360° VIDEOS

Works	Design Type	Design Objective	No. of UEs	Application Domain
[148]	Cloud, remote edge, and local edge based computation, Hybrid-casting	Bitrate, Latency	Multiple	FoV only, full 360° video, model for 6DoF
[149]	Central and edge cloud-based computing	Bandwidth, Latency, Scaling	Multiple	Massively Multiplayer Online Games
[150]	Device heterogeneity based edge computing	Bandwidth, Energy, Quality	Multiple	360° video
[154]	FoV popularity-based caching	Cache optimization	Single	360° video
[156]	Tiles popularity-based caching	Cache Optimization, Accurate tiles coverage	Single	360° video
[158]	Popularity aware caching and transmission	Improved Cache-hit ratio, Transmission rate	Multiple	360° video
[157]	Joint EPC and RAN caching	Bandwidth, Latency	Single	360° video
[159]	Wireless VR network architecture with echo-liquid state deep learning model	Reliability of VR Users	Multiple	360° content transmission with unmanned aerial vehicles (UAVs)
[160]	Communication-constrained MEC framework	Low communication cost	Single	360° video
[161]	3C based cooperative multi-cell network	Improved transmission of VR content	Single	VR video
[162]	3C enabled mobile VR network	Bandwidth, Latency	Single	360° Video
[163]	3C architecture based on 5G-CORAL system	Bandwidth, Energy, Deployment cost	Single	360° video
[164]	Proactive computing and mmWave communication	Latency	Multiple	VR Gaming

cells, or end terminals, lowering the load on the core network [167].

Table III provides a summary of edge-assisted solutions for 360° and VR videos. Most of the task-offloading MEC solutions [147]–[149], [161]–[163], focus on optimizing bandwidth, energy, or latency only. Developing solutions that focus on many other important objectives (e.g., reliability, mobility, QoS, deployment cost, security) at the same time could support a promising VR experience. Leveraging the power of edge computing with the caching can boost mobility, portability, location-awareness, effective data distribution, network context understanding, and safety for service provisioning, etc., [168]. The hierarchical edge-cloud architecture [148] is necessary to accommodate the fast dynamic transmission of 360° video. In contrast to single static cache [155], multiple dynamic cache models can help to manage the abrupt viewport or network changes to improve the viewport hit ratio for multiple users. Regardless of the environment, the proactive caching [163] can increase the perceptual quality by employing prediction mechanisms to prefetch and cache parts of the video.

B. Cooperative Transmission of 360° Video

In the present information era, there is an increasing use of 360° video streaming due to both user demand and advancements in supporting networking and computing techniques. However, streaming redundant information outside the viewport wastes significant network bandwidth. The bitrate requirements become even harder to meet when the same 360° content is streamed to multiple users over the bandwidth-constrained networks. Several approaches employ cooperative transmission of 360° video for serving a group of viewers to improve transmission efficiency. Ahmadi *et al.* [169] introduced a DASH-based weighted tile approach to optimize the coding performance of tiles requested by the subgroups of users. The proposed multicast streaming solution assigns appropriate weights to the tiles based on the probability to be watched by the users. It then selects the bitrate of tiles for each subgroup respecting the available bandwidth and tile weights. However, there could be substantial spatial quality variations due to the different quality of adjacent tiles, leading to poor streaming performance. Additionally, the discrete optimization problem is needlessly large and may not guarantee positive performance.

Bao *et al.* [170] proposed a multicast framework based on motion prediction and channel conditions of concurrent viewers to deliver only the likely to be watched blocks of 360° video. However, the proposed solution does not consider to optimize resource allocation in wireless multicasting. Different from [170], Guo *et al.* [171] envisioned random motion patterns and erratic channel conditions for each user and exploited multicast opportunities to avoid the redundant data transmissions. The authors considered two non-convex problems: (i) under the given video quality constraints, minimization of average transmission time and energy, (ii) under the given transmission time and energy budget, maximization of video quality for each user. Similarly, Long *et al.* [172] considered the transmission time, video quality smoothness, and power constraints to optimize the aggregated utility of multiple users in a single-server multi-user wireless network environment. To reduce the transmission complexity, the authors prepared the tiles in multiple qualities and divided the tiles set into disjoint subsets for each group of users.

The transmission of high-resolution content to multiple users should try to balance the expensive bandwidth, minimal latency, and high transmission reliability requirements.

Zhang *et al.* [128] introduced a cooperative streaming scheme using the SVC quality adaptation methodology to improve the bandwidth sharing among multiple users watching the 360° content in a proximity MANET environment. The proposed heuristic approach selects the optimal subsets of tiles based on the probability of being watched and the aggregated group-level preference while meeting the constraints of the available network resources. Kan *et al.* [173] proposed a server-side hybrid multicast-unicast cooperative streaming scheme to deliver quality variable 360° video tiles to multiple users. The clustering mechanism groups the users based on their watching behaviors to ease the sharing of the same video content. The proposed system then jointly selects the transmission mode and the apt bitrate for each tile to enhance the overall QoE.

For large scale VR deployment, Huang and Zhang [174] devised a MAC scheduling approach in MIMO networks. The resource allocation scheme is based on three main functions: i) motion-to-photon (MTP) [175] latency-based VR frame weight calculation, ii) maximum Aggregate Delay-Capacity Utility (ADCU)-based user selection, and iii) a link adaptation method to balance the ultra-high requirements of VR data transmission. Li and Gao [176] proposed the MultiUser Virtual Reality (MUVR) framework, where an edge cloud adaptively memorizes and reuses the redundant VR frames to reduce the computation and transmission load. MUVR provides a two-level cache design, such as a small local cache at each user-end and a sizeable central cache at the edge. This cache design reduces the memory requirements by generating the background view for all users, reusing frames whenever possible. The empirical evaluation using the Android platform and Unity VR application engine demonstrates that the proposed framework reduces the frame-associated data and computation load with more than 95% and 90%, respectively.

Sharing popular content such as 360° video is a natural choice for live streaming to multiple adjacent users. However, non-cooperative users competing for bandwidth quickly choke the entire network. Therefore, in order to achieve improved QoE for multiple users, researchers have put efforts towards i) identifying the likely demands of multiple users to equitably distribute the available network resources, ii) analyzing cross-users watching behavior to accurately transmit the required sub-frames to the end-user [170], [177], and iii) securing the VR frames transmission to multiple end-users due to the side-channel attacks [176].

VII. Adaptive 360° Video Streaming Challenges

The level of satisfaction of a user viewing 360° video content is more sensitive to disturbance when using a headset than when a traditional display is employed. The immersive experience is negatively influenced by imperfect viewport prediction and highly dynamic network conditions. For instance, poor network conditions introduce extended round trip latency which strongly affects the perceived quality. Several challenges, as illustrated in Fig. 13, need to be addressed in order to create and maintain a strong immersive and engaging user experience with 360° video. This section discusses viewport prediction, quality assessment aspects, and the impact of network conditions on on-demand as well as live 360° video streaming.

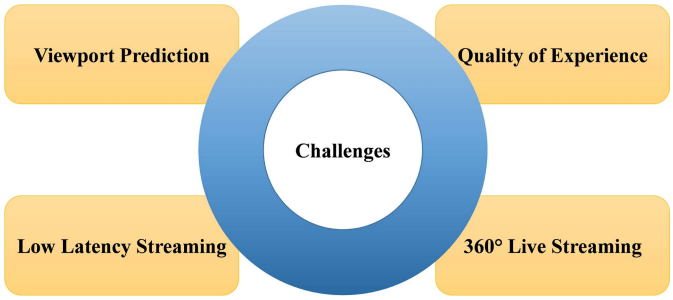

Fig. 13. Adaptive 360° video streaming challenges.

A. Viewport Prediction

One of the essential characteristics of HMD is to respond quickly to the viewer's head movement. The HMDs process the interaction signals when users change their viewport and can detect the related viewport to precise the player information so that a view becomes available to the user from a normal visual angle. Viewport prediction performs an essential role in the optimized streaming of 360° video. Wearable HMDs equipped with position sensors allow the clients to update a viewing scene corresponding to their viewing orientations. Viewport prediction approaches are often classified into *content-agnostic approaches* that predict the future viewing position based on the historical information, and *content-aware approaches*, that require video content information to anticipate the future viewports.

1) Content-Agnostic Approaches: Several existing content-agnostic approaches predict future viewing position using various prediction methods such as, average [17], linear regression (LR) [17], [178], Dead Reckoning (DR) [179], clustering [177], [180], [181], straightforward machine learning (ML) [182]–[184], and encoder-decoder architecture [183], [185]. Qian *et al.* [17] used average, linear regression, and weighted linear regression models for viewport prediction and then entirely streamed those tiles that will overlap with the estimated viewport. They showed that weighted linear regression performs better than average and simple linear regression methods when predicting viewport for the next 0.5s, 1s, and 2s. Petrangeli *et al.* [186] divided the tiles of the equirectangular frames into three regions (i.e., viewport, adjacent, and outside) and assigned variable bitrates to the visible and non-visible regions by incorporating the viewer head movements. The authors used a linear extrapolation of the recent (100ms) watching the history of the user to predict the future fixation point. Different from the LR model, Mavlankar and Girod [179] performed viewing region prediction using motion vectors, i.e., speed and acceleration, of the viewer for a pan/tilt/zoom streaming system. La Fuente *et al.* [187] considered two prediction variants: angular velocity and angular acceleration, for estimation of the future head orientation of the user from his/her previous orientation data. According to the prediction results, different quantization parameter (QP) values are assigned to each

tile. Unfortunately, these methods have limited prediction accuracy when predicting the viewport further in the future (e.g., beyond a 2s interval) [17]. Consequently, if video tiles are requested based on a wrong prediction, the user's actual viewport may be covered by black tiles for which no content was requested.

The cross-users watching behavior can improve the prediction performance. Ban *et al.* [180] exploited the cross-users watching history using the K-Nearest-Neighbors (KNN) algorithm and user's personalized behavior using the LR model. The absolute and relative improvement achieved in terms of viewport prediction accuracy is about 20% and 48%, respectively. Liu *et al.* [181] proposed using a data fusion approach to estimate the future viewing position by taking into account several features, i.e., the engagement level of users, behavior of the users watching the same video, the behavior of a single user watching multiple videos, end-user device, mobility-level, etc. Based on the concept of vehicle trajectory prediction, Petrangeli *et al.* [177] considered similar trajectories form a cluster to predict future viewports. The trajectory-based clustering approach results in improved prediction accuracy for longer horizons. However, they examined three different trajectories for Euler angles (θ, ϕ, ψ), which might lead to unsatisfactory performance. Rossi *et al.* [188] proposed a clustering method to identify clusters of users based on meaningful viewports overlap in spherical space. The clustering algorithm based on the Bron-Kerbosch (BK) [189] algorithm recognizes substantial groups of users who are watching the same 60% of 3s long chunks of spherical video. This method provides clusters with compatible and significant geometric viewports overlap compared to the benchmarks.

LR methods result in poor prediction accuracy for a long-term prediction horizon [182]. Long-Short Term Memory (LSTM) [190] is a Recurrent Neural Networks (RNN) architecture that is suitable for sequence modeling and patterns exploiting. In order to achieve higher precision than LR in FoV prediction, Jiang *et al.* [135] developed a viewpoint prediction method using a LSTM model with 128 neurons. The authors analyzed the 360° dataset [191] and observed that users have fast head turns in the horizontal direction, but almost stable motions in the vertical direction. The experimental comparison with LR and average approaches reveals that the LSTM-based viewport predictor generates lower prediction errors considering both horizontal and vertical head movements.

Bao *et al.* [182] conducted a subjective experiment with 150 users to analyze their viewing behavior across 16 video clips. They showed that the angles (θ, ϕ, ψ) representing the user motion in 3D space have a strong auto-correlation and negligible cross-correlation. Therefore, these angles can be measured separately. The authors developed two separate LSTM neural network models to predict θ and ϕ, separately. These prediction results are then used for a targeted area streaming to utilize the available network resources efficiently. Hou *et al.* [192] proposed a deep learning-based view generation method to extract and stream only the predicted viewport tiles in advance for 360° videos and 3 Degrees of Freedom (3DoF) VR applications. The authors trained their model using a real large-scale dataset, i.e., about 36,000 head movement traces for 19 videos. In another work [193], the same researchers introduced a new predictive approach involving multilayer perceptron (MLP) and LSTM models to predict the head (i.e., viewing direction) as well as body (i.e., standing position) movements in a 6DoF VR environment. The predictive view is pre-rendered to enable low-latency VR experience.

In several cases, the movement of the user is highly volatile during specific parts of the video. This adds pressure to the training of machine learning approaches. To reduce the impact of user movements, Heyse *et al.* [194] proposed a contextual agent based on the RL model which first detects the significant movement of the user and then predicts the direction of the movement. The layered self-learning executor outperforms a spherical trajectory extrapolation approach [118] which models the user movements as a fraction of trajectory rather than a full trajectory on a unit sphere. Qian *et al.* [12] proposed an algorithm, called Flare, to minimize the mismatch between actual and predicted viewport. The researchers employed an ML approach to perform frequent viewport predictions concerning four intervals across 1300 head movement traces collected from 130 users. With the viewport trajectory prediction, Flare enables an incorrect prediction to be replaced by the latest prediction.

LSTM networks have a time-consuming sequential training nature. Encoder-decoder LSTM models parallelize the training process resulting in improved prediction accuracy compared to LR and LSTMs. Yu and Liu [185] used the attention-based LSTM encoder-decoder network architecture to avoid expensive recursion and to capture the viewport changes better. The proposed architecture achieves a higher prediction accuracy, lower training complexity, and faster convergence compared to the traditional RNNs. Jamali *et al.* [183] proposed using LSTM encoder-decoder network architecture for the long-term, i.e., up to 3.5s, viewport prediction. The authors collected cross-users orientation feedback over low latency heterogeneous networks to adjust the prediction performance for target users on high-latency networks.

2) Content-Aware Approaches: Content-aware viewport prediction is considered to be a vital enabler for 360° video because it can improve the prediction efficiency. Aladagli *et al.* [201] proposed a saliency-driven model to improve the prediction accuracy. However, this work did not consider the user's viewing behavior for 360° videos. Viewport prediction errors could be minimized by understanding the user's unique visual attention for 360° videos. Most existing methods [196], [197], [202] focus on considering saliency patterns as well as positional information in 360° display to achieve better prediction results. The general architecture of saliency and positional information based fixation prediction model is shown in Fig. 14. Instead of using traditional saliency models, Nguyen *et al.* [202] proposed PanoSalNet to capture the unique visual attention of the users in 360° frames to improve the saliency detection performance. The fixation prediction solution with both HMD sensor features and saliency maps results in a measurable gain. Xu *et al.* [196] proposed two DRL models for the viewport prediction network

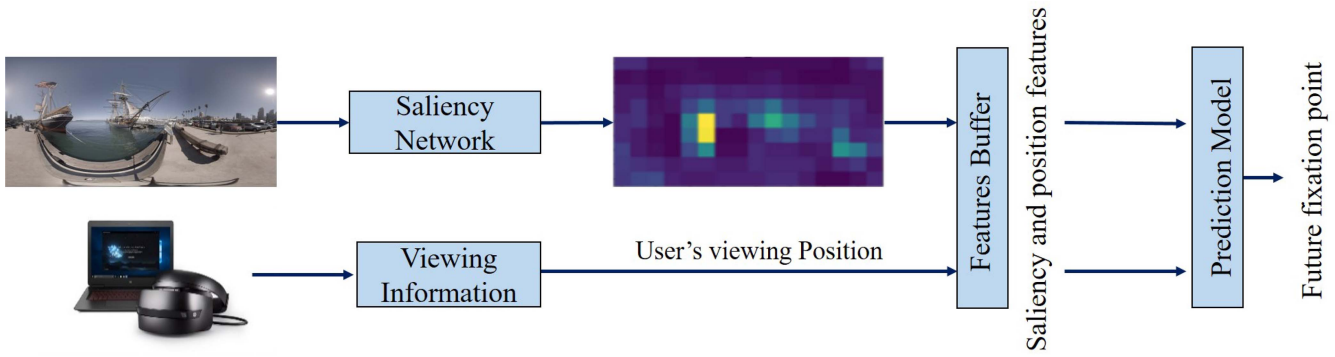

Fig. 14. Saliency and viewing position based fixation prediction network.

TABLE IV
SUMMARY OF VIEWPORT PREDICTION APPROACHES FOR 360° VIDEOS

Works		Design Type	Prediction Horizon	Dataset	Display
Content-Agnostic Approaches	[17]	Average, LR, WLR	2s	4 videos, 5 users	Google Cardboard
	[187]	Linear extrapolation	1 segment	1 video, 10 users	Samsung Gear VR
	[188]	Angular velocity, Angular acceleration	1 segment	10 videos, 17 users	OculusRift CV1
	[181]	Users clustering	6s	[155]	HTC Vive
	[189]	Clique clustering	3s	[192]	Razer OSVR HDK2
	[178]	Trajectory-based clustering	10s	[183]	Oculus DK2
	[136]	LSTM model	3s	[192]	Razer OSVR HDK2
	[183]	Neural network model	1s	16 videos, 150 users	Oculus DK2
	[193]	LSTM model	1 segment	19 videos, 36000 users	Samsung Gear VR
	[195]	Contextual bending learning	1s	1 video, 5 users	HMD
	[12]	Online machine learning	3s	10 videos, 130 users	Samsung Gear VR
	[186]	Attention-based encoder-decoder	3s	[155], [192], [196]	Razer OSVR HDK2, Monitor, HTC Vive
	[184]	LSTM encoder-decoder LSTM + guide users	3.5s	[155]	HTC Vive
Content-Aware Approaches	[197]	DRL models	30ms	76 videos, 58 users	HTC Vive
	[198]	CNN-LSTM model	1s	208 videos, 45 users	HTC VIVE
	[199]	LSTM encoder-decoder ConvLSTM	10s	[155], [198]	HTC Vive
	[200]	RNN+CFVT	1s	[183]	Oculus DK2
	[201]	Sensor and content features	1s	10 videos, 25 users	Oculus Rift
	[202]	Saliency-driven	2s	[183]	Oculus DK2
	[203]	Sensor and content features	2.5s	[155], [192]	Razer OSVR HDK2 HTC Vive

considering both motion trajectories and visual features. The offline model detects the saliency in each frame based on content popularity. The online model then predicts the viewport direction and magnitude based on the obtained saliency maps from the offline model as well as the previous viewport information. However, the network aims to predict the next viewport position for only 30ms, i.e., one future frame. Xu *et al.* [197] collected a large scale dynamic dataset of 208 360° videos viewed by 45 subjects with an eye-tracking-capable HTC VIVE headset. The authors proposed predicting the gaze displacement based on the history scan path and image features. They performed saliency computation at three spatial scales related to the current gaze point, the viewport, and the whole image. The possible image features are extracted by feeding the images, and the corresponding saliency maps to a CNN, while the LSTM model captures the history information. Then both the LSTM and CNN peculiarities are coupled for gaze prediction over the next second.

Since the users are more attracted to the moving objects. Therefore, in addition to saliency maps, Fan *et al.* [200] also considered motion maps of content using pre-trained CNN to estimate the future fixation points of the user. However, the motion maps exploitation needs further investigations since there could be multiple motions, which makes the prediction not reliable enough. Yang *et al.* [199] predicted the single viewport based on the history viewing angles information using the CNN model. Next, the authors considered a viewport trajectory prediction using a fusion layer that combines the results of the content-agnostic and content-aware prediction models such as RNN and CFVT (correlation filter-based viewport tracker) models. The incorporation of the fusion model enables both models to support better prediction and improves

accuracy by 40%. Ozcinar *et al.* [14] transformed the viewport trajectories to the viewport based visual attention maps and then dynamically streamed the tiles of variable sizes to ensure higher coding efficiency.

Current prediction models are limited regarding how much they can predict in the future. Li *et al.* [198] proposed two models for viewport prediction in viewport-dependent and tile-based streaming systems. The first model employs an LSTM encoder-decoder network architecture based on the trajectories of the users. The second model employs a convolutional LSTM encoder-decoder architecture using the heatmaps of the sequences to predict the future orientation of the user.

Accurate orientation prediction enables the 360° client to download the most relevant tiles at high-resolution. Table IV summarizes the content-agnostic and content-aware head movement prediction approaches. Current neural network models [196], [197], [200]–[202] that employ both the saliency and position information perform poorer than a simple no-motion baseline which directly utilizes the current viewing position for future viewport position estimation [203]. The noise level in the estimated saliency limits the prediction accuracy of these models [203]. Besides, these models involve additional computational complexity. For the reliable prediction of attention points in 360° video and understanding the relationship between the user's viewing likelihood and saliency maps, the saliency models must be improved and well fitted by training on large scale datasets, especially captured with different camera rotations [204]. On the other hand, the convLSTM encoder-decoder [198] and trajectory-based prediction approaches [177], [180] are suitable for long-term prediction and can bring considerable QoE improvement, especially in a cooperative streaming environment.

B. Quality of Experience Assessment

As the omnidirectional video is highly prevalent, it is imperative to determine the user's specific quality aspects with this type of video distribution. QoE plays a crucial role in video streaming applications [205]. In traditional video streaming, QoE is mostly affected by network load and delivery performance [206], [207]. These solutions employe existing sub-optimal objective metrics, including QoS metrics, structural similarity (SSIM), and Peak Signal to Noise Ratio (PSNR), to access QoE levels. However, these metrics may not be most appropriate for assessment of omnidirectional video quality, which is strongly affected by both network conditions and user's viewing behavior.

1) Subjective Quality Assessment: Since users are the ultimate consumer of videos, subjective quality assessment is the actual and surest way to estimate the quality of 360° video streaming. Several subjective quality assessment approaches exist in the literature. Upenik *et al.* [208] performed a subjective test to experience the quality of 360° images by using a MergeVR[16] HMD. The experimental data, including subjective scores, view tracking, and time spent on each stimulus was obtained through a custom software application. The viewing direction information is then used to compute saliency maps. Unfortunately, this study did not consider the subjective assessment of 360° videos.

The significant difference between regular and 360° videos is that only viewport content is displayed in 360° videos. To cover up the performance gap of quality evaluation metrics between 360° videos and regular videos, Zhang *et al.* [210] proposed a subjective assessment method for panoramic videos called SAMPVIQ. In their experimental setup, twenty-three participants were allowed to view four impaired videos and rate between 0 to 5 for the overall video quality experience. In this work, a comparatively high rating variance was observed between participants. Xu *et al.* [212] proposed two subjective quality evaluation metrics, namely overall differential mean opinion score (O-DMOS) and vectorized DMOS (V-DMOS), to assess the quality loss in 360° videos. The O-DMOS metric computes the total differential score of the subjective test sequences and is similar to the DMOS metric for regular videos. Schatz *et al.* [214] studied the impact of stalling events when consuming 360° content on HMD compared to traditional 2D display devices. The authors found that subjective quality assessment for immersive content is not trivial and could lead to more open issues than actual recommendations. In general, the expectations would be similar to those for the traditional HAS, i.e., no stalls at all if possible.

Several open-source tools for 360° videos are already available. For example, AVTrack360 [229], OpenTrack, and 360player [191], which captures the head traces of users viewing 360° videos, or VRate [230], which is a Unity-based tool to provide subjective questionnaires in a VR environment. In addition, Pérez and Escobar [231] proposed a full-fledged Android-based application, MIRO360, to support the development of guidelines for the future VR subjective tests. MIRO360 facilitates test patterns considering both short and long sequences to assess the visual quality during the playback following the ITU-R BT.500-13 [232] recommendations. The application has a user-friendly interface and is provided in GitHub.[17]

Cybersickness is a potential barrier to achieve higher QoE levels and can cause fatigue, nausea, discomfort, and aversion [233]. Singla *et al.* [216] conducted two subjective experiments with limited resolution and bandwidth options under different delay settings, e.g., 0ms, 12ms, 47ms, and 112ms. Their significant contributions include the development of subjective testbed, testing methods, and metrics to evaluate the video perception level and cybersickness in viewport adaptive 360° video streaming. The authors revealed that tile-based streaming performs well under limited bandwidth conditions. They also found that 47ms delay does not substantially affect the perceived quality. Tran *et al.* [219] considered several influencing factors such as spatial complexity of content, quantization parameter and resolution characteristics, and rendering models to evaluate the cybersickness, quality rating, usability, and presence of the user. It has shown that the fast-moving content highly promotes cybersickness in a VR environment. Moreover, with high usability and presence ratings, the cybersickness of the user may also become elevated.

[16]https://mergevr.com/

[17]https://github.comlzerepolbap/miro360

TABLE V
SUMMARY OF SUBJECTIVE QUALITY ASSESSMENT APPROACHES FOR 360° VIDEOS

Works	QoE Aspects	Dataset	Rendering mode	Recommendations/Rating Criterion	Application
[210]	Perception	6 images, 48 users	MergeVR headset	Absolute Category Rating (ACR) [211], ACR with hidden reference (ACR-HR) [211], ITU-R BT500-13,	360° images
[212]	Perception	16 videos (10s), 23 users	HTC Vive	SAMPVIQ recommended in ITU-R BT.1788 [213]	360° video
[214]	Perception	48 videos (20-60s), 40 users	HTC Vive	O-DMOS, V-DMOS based on DMOS [215]	360° video
[216]	Perception, Presence	2 videos (60s), 22 users	VR static, VR move TV sets	ITU-R BT.500-13, ITU-T P.913 [217]	360° video
[218]	Cybersickness, Perception	30 videos (10s), 28 users 42 videos (10s), 25 users	Samsung Gear VR	ACR, ITU-R Rec. BT.500-13, Ishihara charts [219], Snellen charts [220]	360° video
[221]	Cybersickness, Perception, Usability, Presence	60 videos (30s), 36 users	Samsung Gear VR Samsung Galaxy S6	-	360° video
[222]	Cybersickness, Perception	12 videos (60-65s), 28 users	HTC Vive, Oculus Rift	ACR	360° video
[223]	Perception, Presence, Sensory realism	8 videos (20s), 25 users	HMD, PC	ITU-T P.913	360° video
[224]	Presence, Usability, Emotions	1 VR game (20-25 minutes) 22 users	Oculus Rift, PC	System Usability Scale (SUS) [225], Differential Emotions Scale (DES) [226], Presence Questionnaire (PQ) [227]	VR Gaming
[228]	Perception, Presence, Usability, Sensory realism	1 video (2 minutes) 33 users	Oculus Rift, PC	Sensorial guidelines [229], ACR, Ishihara charts, Snellen chart	VR video
[230]	Perception, Presence, Cybersickness	4 videos (20s), 12 users	HTC Vive Pro, PC	Pair Comparison (PC) [211]	360° video

Singla *et al.* [220] assessed the viewing discomfort of twenty-eight subjects watching six YouTube videos in full HD and Ultra HD resolutions on Oculus Rift and HTC Vive headsets. The authors reported that the HMD type slightly affects the perceived quality, while the resolution and content type strongly influence the personal experience. Additionally, more female users experienced simulator sickness compared to male users.

The spatial presence increases the sense of immersion in a VR scene. Zou *et al.* [221] presented a subjective framework to measure the spatial presence of the twenty-five subjects experiencing 360° videos on HMD and monitor. The proposed system framework consists of three layers from top to bottom, i.e., spatial presence layer, perception layer, and technical influencing factors layer. The psychological spatial presence aspects form the spatial presence layer. The perception layer characterizes video realism, audio realism, and interactive elements. Finally, the technical influencing factors layer consists of several modules linked to the perception layer to reflect sensory realism. Hupont *et al.* [222] applied generic perceptual principals to study the spatial presence of users playing VR games on Oculus HMD and traditional 2D display screens. The subjective evaluations with twenty-two participants show that 3D virtual realism points to higher amazement, immersion, presence, usability, and excitement compared to 2D displays.

Recent efforts aim to explore alternative ways of quality assessments. Salgado *et al.* [234] intended to capture various physiological metrics, e.g., heart rate (HR), electrodermal activity (EDA), body surface temperature, electrocardiographic signal (ECG), respiration rate, blood volume pressure (BVP), and electroencephalography signals (EEG) using wearable sensors to evaluate the quality of the immersive wheelchair simulator. Egan *et al.* [226] hired thirty-three volunteers to evaluate the quality scores in VR and non-VR rendering modes based on HR and EDA signals. It has shown that EDA has a strong influence on quality scores compared to HR.

Different technical and perceptional features such as distortions, sharpness, colorfulness, contrast, flickering, etc., are used for assessment of perceived video quality. Fremerey *et al.* [228] identified that visual quality strongly depends on the employed motion interpolation (MI) algorithms and the video characteristics, e.g., camera rotation and movements of the objects. In a subjective experiment, 12 video experts reviewed four video sequences interpolated to 90 fps using FFmpeg blend, FFmpeg MCI (Motion Compensated Interpolation), and butterflow. The authors found that MCI provides excellent improvements in QoE compared to other algorithms.

Subjective tests are directly associated with human eyes and shed light on the impact of different aspects of 360° video quality assessment. Among these aspects, spatial presence and cybersickness caused by viewing 360° video via VR headsets are most important since these effects do not occur if a user watches regular videos on a desktop screen. Table V details the subjective approaches with respect to *Perceptual*, *VR sickness*, *Presence*, *Usability*, and *Sensor-based* QoE aspects. Subjective quality assessment needs comprehensive manual efforts and is thus expensive, time-consuming, and error-prone. Contrarily, objective quality assessment is more manageable and practical.

2) Objective Quality Assessment: It is natural to employ conventional objective assessment approaches for 360° content due to the similar encoding structure and 2D plane projection formats. However, the sampling density in existing projection methods is not uniform at each pixel position. Yu *et al.* [235] introduced S-PSNR and L-PSNR for sphere-based PSNR calculation. The S-PSNR calculates PSNR by equally weighting all positions of the pixels on a spherical surface. By utilizing interpolation algorithms, S-PSNR can achieve an objective quality measurement of 360° videos supporting various projections schemes. The L-PSNR measures PSNR by weighting pixels based on their latitudes and access frequency. L-PSNR could measure the average viewport PSNR without specific

TABLE VI
SUMMARY OF OBJECTIVE QUALITY ASSESSMENT APPROACHES FOR 360° VIDEOS

Works	Metrics/Methods	Dataset	Distortions	Projection	Resolution
[237]	L-PSNR, S-PSNR, Viewport-PSNR	10 videos (10s)	4 QPs	ERP, CMP, Dyadic, Cylindrical Equal-area	4096x2048
[238]	WS-PSNR	2 videos (1s)	QP (30-37)	ERP, CMP	3840x1920
[239]	S-SSIM	8 videos (10s)	QP (22, 27, 32, 37, 42)	ERP	3600x1800
[240]	WS-SSIM	2 images	QP (30-37)	ERP	3840x1920
[241]	ProbGaze-PSNR, ProbGaze-SSIM	[155]	CRF (15, 20, 25, 30, 35)	ERP	1920x1080
[214]	NCP-PSNR, CP-PSNR	48 videos (20-60s)	QP (27, 37, 42)	ERP	2880x1440 7680x3840
[242]	S-PSNR, WS-PSNR [238], CPP-PSNR [243]	4 images	Bitrates (0.25, 0.50, 0.75, 1.00) bits per pixel	ERP, CMP	3000x1500
[244]	PSNR, S-PSNR [237], S-PSNR with interpolation, WS-PSNR [238], CPP-PSNR [243]	3 videos (30s)	QP (22, 28, 32, 36, 40)	ERP	3840x1920, 2880x1440, 2160x1080, 1440x720
[245]	Video quality, Quality variances, Stalling time, Startup delay	[155]	Bitrates(1.8 Mbps for video 1, 2.7 Mbps for video 2)	ERP	1920x1080
[246]	Head and eye movement-based objective quality assessment	600 videos (10-23s)	QP (27, 37, 42)	ERP, TSP, RCMP	3840x1920 7680x3840
[247]	Pixel-wise quality metric	16 videos (10s)	6 Bitrates	ERP	4096x2048

head movement trajectories. The quality metrics designed for 360° videos include remapping based on the corresponding projection format. Zakharchenko *et al.* [241] proposed a Craster Parabolic Projection-PSNR (CPP-PSNR) metric to compare various projection schemes by remapping the pixels to CPP projection without changing the spatial resolution and calculating the PSNR at actual pixel locations. With CPP, the resolution of a 360° video could hardly be decreased. Sun *et al.* [236] proposed a quality measurement metric, called weighted-to-spherically-uniform PSNR (WS-PSNR), to measure the quality variance between original and impaired content. The authors considered weights according to the position of pixels on the sphere for quality assessment of 360° content.

Different from PSNR, SSIM is another quality evaluation metric that reflects the image distortion with three factors, including luminance, contrast, and structure [246]. Chen *et al.* [237] analyzed the SSIM results for 2D and 360° videos and introduced a spherical-structural similarity (S-SSIM) metric to compute the similarity between impaired and original 360° videos. In S-SSIM, a reprojection is incorporated to calculate the similarity between the two extracted viewports. Zhou *et al.* [238] proposed WS-SSIM metric, by considering the similar weights, as in [236], to measure the similarity between the windows on the projected area. The performance evaluation reveals that WS-SSIM is closer to human perception compared to other quality evaluation metrics.

Van der Hooft *et al.* [239] proposed the ProbGaze metric, based on the spatial dimension of tiles and the gaze point within the viewport. ProbGaze considered the weight of peripheral tiles to provide a suitable quality measurement. The ProbGaze versions of the objective metrics, e.g., SSIM and PSNR, were able to estimate the quality changes on peripheral tiles when a user suddenly change the viewing position compared to the center-based (which considers only the weight of the viewport tiles) and average-based (which considers the average weight of all the tiles) versions of PSNR and SSIM metrics. Xu *et al.* [212] introduced two objective quality evaluation metrics, i.e., content-based perceptual PSNR (CP-PSNR) and non-content-based perceptual PSNR (NCP-PSNR), for encoded 360° videos. The first metric weighs the pixels distortion based on the spherical panorama content, while the second metric considers the human preference statistics to estimate the quality loss.

Although several objective quality metrics, namely S-PSNR, L-PSNR, CPP-PSNR, WS-PSNR, S-SSIM, WS-SSIM, etc., have been extensively used for 360° video evaluation. However, they do not truly capture the perceived quality, especially when HMDs are employed to watch the videos [247]. Upenik *et al.* [240] considered a subjective experiment with four high fidelity 360° images watched by forty-five participants to evaluate and compare the performance of objective quality metrics under different encoding settings. The experimental comparison with subjective ground-truth data reveals that the current objective metrics (e.g., WS-PSNR, CPP-PSNR, etc.) have a lower correlation with subjectively perceived quality. Tran *et al.* [242] demonstrated a higher correlation between objective and subjective results compared to [240]. However, this work considers a limited dataset (e.g., three videos watched by 18 users). Therefore, an optimal quality metric specially designed for 360° content is strongly needed.

Machine learning-based quality assessment approaches could bridge the gap between subjective and objective quality assessments. Da Costa Filho *et al.* [243] proposed a two-stage model for QoE assessment of VR content. First, the playout performance of adaptive VR video is determined by using machine learning techniques. In the second step, the model utilizes the estimated metrics including video quality, quality variances, stalling time, and startup delay in determining the user's QoE. Li *et al.* [244] introduced a DRL-based quality assessment model that considers both head and eye movements during a streaming session. The 360° video is divided into several patches. The patches with low viewing probability are eliminated. Both reference and impaired video sequences are inputted into a deep learning executable to calculate the quality score of the patches. Next, these scores are weighted and added

to get the final quality scores. Yang *et al.* [245] considered multi-level quality features and fusion models for objective quality assessment. The quality features are computed with region of interest (ROI) maps and include pixel-level, region-level, object-level, and equator bias features. The fusion model is built by a backpropagation neural network to combine the multiple quality features for obtaining the overall quality score.

Accurate QoE assessment is a significant factor in optimizing the 360° video streaming service and is a fundamental one for adaptive delivery solutions. More structured research towards designing widely acceptable accurate QoE assessment models and metrics for 360° video is highly needed, and this is challenging. Most of the literature on quality assessment considers limited camera motions or content characteristics to capture different quality attributes, as described in Table VI. The subjects are usually regular viewers but not the expert reviewers, as considered in [228]. Measuring visual quality alone in VR, however, is not sufficient for a complete QoE framework. It is also essential to find the impact of other factors, e.g., cybersickness, physiological symptoms, user discomfort, HMD weight, usability, VR audio, viewport degradation ratio, network characteristics (e.g., delay, jitter, bandwidth, etc.), content characteristics (e.g., camera motion, frame rate, encoding, projection [248], etc.), and streaming characteristics (e.g., viewport deviations, playback rebuffering, spatial, and temporal quality variations, etc.) to get scientific community acceptance.

C. Low Latency Streaming

Rich video services such as 360° and VR videos require a low response delay. This response delay is a combination of sensor delay, cloud/edge processing delay, network delay, requests overhead, buffering delay, rendering delay, and feedback delay. The low delay requirements are even more stringent for cloud-based VR gaming, immersive telepresence, and video conferencing, although some of these services are not very common yet. However, instant audio and video updates are expected by user brain when they change their viewing angles. Therefore, ultra-low terminal processing latency, fast edge/cloud computation, and very low network latency are required to ensure such level of responsiveness to user head movements. With modern HMD devices, sensor delay has been decreased to an amount that is unnoticeable by users [249]. Moreover, the transmission delay is significantly reduced by the new 5G mobile and wireless communication technologies [250]. However, work towards reducing processing, buffering, and rendering delay are essential to minimize the motion-to-photon delay. Many immersive applications target an MTP latency less than 20ms [251]; ideal is to achieve a less than 15ms delay, which makes it nearly imperceptible by the users.[18]

A simple strategy for minimizing the startup time in adaptive streaming is to decrease the data needed to initiate the playback. Usually, small download and startup time is observed with short video segments [18], [252]. Van der Hooft *et al.* [253] considered the streaming of news-related content using: (i) server-based encoding, (ii) server-based user profiling, (iii) server push strategy and (iv) proactive storage of the video data at client-side, to lower the end-to-end system latency. The proposed framework lowers startup time and allows fast content switching under different network settings. Similar to traditional videos, long response delay degrades the performance of viewport adaptive streaming schemes. Nguyen *et al.* [254] analyzed the influence of adaptation interval delay and buffering delay on viewport-dependent adaptive streaming. The authors proposed a server-side bitrate computation strategy to minimize the impact of response delay. The proposed system estimates the available network throughput and the future viewport position following the client's response. The decision engine on the server-side then streams the suitable tiles to meet the delay constraints. The real-world experiments reveal that small adaptation and buffering delays are inevitable for viewport-dependent adaptive streaming.

Spatially splitting a video frame into rectangular tiles increases the network overhead because of separate requests for each tile in HTTP/1.1. The request explosion problem leads to a longer response delay and can be resolved using HTTP/2's server push feature that enables a Web server to multiplex messages upon a single HTTP request. This approach was previously introduced to overcome latency in Web-based transmission, but now is also being used in video streaming applications. Wei and Swaminathan [255] utilized HTTP/2 protocol to facilitate low latency HTTP adaptive streaming. The proposed server-push scheme attempts to avoid multiple GET requests by sending several segments (e.g., k) upon a single request. Similarly, Petrangeli *et al.* [186] used HTTP/2 server-push feature in combination with the specific request parameters to facilitate 360° video streaming. The client sends a single call for a segment, and the server transmits the k tiles using a first come first served (FCFS) policy. With HTTP/2's priority feature, the high-priority tiles could be fetched on an urgent basis to improve the performance in a high round trip time (RTT) network environment.

Xu *et al.* [256] employed a k-push scheme for 360° videos to push k number of tiles to the client that compose a single temporal segment. The proposed method along with the QoE-aware bitrate adaptation algorithm improves the video quality by 20% and reduces the network transmission delay by up to 30%, under different RTT settings. Yahia *et al.* [257] used the priority as well as the multiplexed features of HTTP/2 to organize the controlled adaptive transmission of urgent video tiles between two consecutive viewport predictions, i.e., before and during the delivery of the same segment. Instead of using HTTP/2, Yen *et al.* [258] developed a QUIC enabled architecture that utilizes stream priority and the multiplexing feature for secured and low latency transmission of 360° video. When a viewport change occurs, QUIC enables to quickly stream regular tiles at a low priority and viewport tiles at a high priority over a single QUIC connection to reduce the missing ratio of viewport tiles. The authors showed that the QUIC protocol based adaptive 360° streaming outperforms the traditional HTTP/1.1 and HTTP/2 solutions.

[18]http://blogs.valvesoftware.com/abrash/latency-the-sine-qua-non-of-ar-and-vr

Ultra-low network latency must be assured to timely deliver the viewport because of the user's continuous interaction through end devices. The low latency streaming of 360° video in cellular networks is supported by the deployment of mobile edge computing architecture. Mangiante *et al.* [48] proposed an explicit edge processing-based viewport rendering solution to reduce the latency as well as battery utilization and computational load on end terminals. However, the authors fail to provide any effective algorithm or establish a practical execution platform. Liu *et al.* [259] employed remote rendering techniques to hide the network latency by achieving high refresh frequency for an untethered VR system. This system utilizes high-end GPUs supported by a 60Ghz wireless link to accelerate computation speed and 4K rendering with reduced display latency. Although the proposed work actively provides high-quality and low latency streaming, it is noteworthy that excessive bandwidth connection (60Ghz) is utilized for offloading, which is not commonly available. Viitanen *et al.* [260] introduced an end-to-end VR gaming system to lower the latency, energy, and computation cost by performing the edge-based rendering where the FoV frames are transmitted as HEVC encoded bitstreams to the end-devices. The authors achieved a low end-to-end system delay (30ms) for a stereo 1080p 30fps format. However, this work focuses on a scenario where sufficient bandwidth is available and powerful gaming laptops are used for processing the video instead of smartphones. Shi *et al.* [21] considered the high-quality rendering of 360° video without focusing much on viewport prediction. The proposed MEC-VR system employs a remote server to dynamically adapt the viewport coverage by using an adaptive cropping filter that adds some extra area outside the viewport according to the observed system latency. The latency based viewport coverage adjustment allows the client to accommodate and compensate for abrupt head movements.

The latency of each user in a shared VR environment depends on the locality of users and the distribution of edge resources within the physical network space. Park *et al.* [261] proposed a bandwidth allocation strategy in a linear cellular topology by considering the two-way communications between multiple users and edge servers to minimize the end-to-end system latency. The authors determined that streaming latency strongly relies on the processing performance of the edge servers and the relationship between multiple interacting users in physical and virtual space. Perfecto *et al.* [262] integrated deep neural network and mmWave multicast transmission to depreciate the streaming latency in a cooperative VR environment. The neural network model estimates the upcoming viewports of users. The users are grouped based on predicted correlations and locality to optimize the correct viewport admission. The proactive multicast resource scheduling is then performed to minimize the latency and traffic volume for VR.

Edge-assisted solutions are predominant in taming the latency in single- [21], [48] and multi-user [261], [262] environments. Besides, the support for server-based viewport computation [254], server-push mechanisms [186], [256], [257], and remote rendering [21], [259] also enable low latency streaming over current wireless networks. The current 4G networks are enough for early adopter immersive multimedia. However, the emerging 5G networks are expected to satisfy the ultra-high requirements of immersive content [167].

D. 360° Live Video Streaming

Traditional broadcast TV channels are a popular source of live streaming of events. Nowadays, personalized as well as 360° live video streaming on video-sharing portals such as Twitch, Periscope, YouTube, and Facebook is witnessing a massive growth. This trend is fueled by the high-resolution 360° capturing cameras and increased efficiency of stitching or post-processing software to improve content preparation. However, live 360° video streaming is more delay-sensitive because of the cloud-based transcoding operation between the content producer and consumer [263]. The existing processing devices are limited in terms of real-time processing tasks such as transcoding, rendering, etc. Hu *et al.* [264] proposed an edge-based live streaming system, called MELiveOV, that enables a capillary distribution of processing tasks of high-resolution omnidirectional content to the 5G enabled edge servers. The end-to-end live streaming system includes a content creation module, transmission module, and viewport prediction module. The mobile edge-assisted streaming design reduces the bandwidth requirement by 50%. Griwodz *et al.* [265] developed a FoV-optimized prototype for live 360° streaming that combines RTP with DASH-based pull-patching to transmit two quality levels of 360° video to both a Huawei IPTV set-top-box and a Gear VR headset with Samsung Galaxy S7. The authors implemented the idea of a collective decoder by multiplexing several decoders on a single H.265 hardware decoder to reduce the switching time.

Video transcoding and adaptive transmission are some of the key factors in media compress/decompress, changing bitrates, or up-sampling/down-sampling of 360° videos. Liu *et al.* [266] showed that only transcoding the viewport has the potential to cut the computational requirements of high-performance transcoding significantly. Baig *et al.* [267] developed fast encoding schemes to deliver live 4K videos to commodity devices. The proposed system employs a layered video coding approach to deliver quality variable chunks over highly dynamic and unpredictable WiGig and WiFi links. Le *et al.* [268] proposed a real-time transcoding and encryption system for live 360° CCTV stream using the RTSP network control protocol. The proposed transcoding method is based on ARIA crypto library, Intel media SDK (Software Development Kit), and FFmpeg library for high-performance transcoding of live 360° CCTV content. The proposed system could manage parallel transcoding operations and achieves high-speed transcoding performance (up to 200% improvement) against libx265 FFmpeg.

Stitching plays a critical role in deciding the overall streaming quality compared to other factors such as capturing, transmission speed, decoding, and rendering, etc. Chen *et al.* [269] proposed an event-driven stitching approach that considers different types of semantic information in 360° frames as events to optimize the stitching time budget. Based on the semantic

information in a VR frame, the tile actuator module selects the suitable tiling design. The stitcher module then performs tile-based stitching such that event tiles will have higher stitching quality based on the available resources. The evaluation reflects that the proposed system adapts well to different sets of events and timing constraints by achieving 89.4% of the timing budget.

Compared to on-demand streaming, 360° live video streaming presents several challenges, such as handling the user navigations without prior knowledge, the first time streaming of the video, and transcoding the live video on the fly. These challenges become even more problematic in a multi-user scenario. In regards to handling the viewing patterns of multiple users, scalable multicast [270] can be employed to serve multiple users with quality levels approaching to on-demand streaming over both low and high bandwidth networks. Besides, ROI based stitching of tiles [269] and transcoding [266] can significantly reduce the latency requirements of delay-sensitive interactive applications.

VIII. Standards and Technologies

Standardization is a key issue for proper technical interoperability. The overall goal of standardization efforts is to specify the minimum essential to enable creative and competitive technologies and services. Several standardization efforts for omnidirectional videos are currently gaining momentum. To ensure universal media access and interoperability for production, distribution, sharing, and consumption [271], media is often encapsulated and signaled using standardized file formats and transport protocols. In this context, MPEG has developed several standards including MPEG-2 Transport Stream (TS) [272], MPEG-4 Part 14 (MP4) [273], MPEG-DASH, MPEG Media Transport (MMT) [274], etc., to ensure the media interoperability. The standardization efforts for immersive media applications are described by considering the immersive media formats, 6DoF+ streaming, immersive audio and video standards, and QoE metrics in this section.

A. Immersive Media Formats

MPEG has developed a standard for immersive media, called MPEG-I (MPEG Immersive media), that includes an Omnidirectional Media Format (OMAF) [275] focusing on the specifications of omnidirectional media applications. OMAF is the first international standard on immersive media format and describes the means to allow the coding, presentation, and consumption of 360° videos. OMAF is compatible with existing standards, including coding (e.g., HEVC), file format (e.g., ISOBMFF), delivery signaling (e.g., DASH, MMT), and includes metadata information of encoding, projection, packing, and viewport orientation.

In an OMAF workflow, the media is captured via one or more fish-eye lenses where stitching and projection steps are shifted towards the capturing side. OMAF considers equirectangular projection and cubemap projection due to their effectiveness. OMAF also relies on region-wise packing (RWP) [276], where the projected frame can be scaled, resampled, rotated, and mirrored according to the streaming

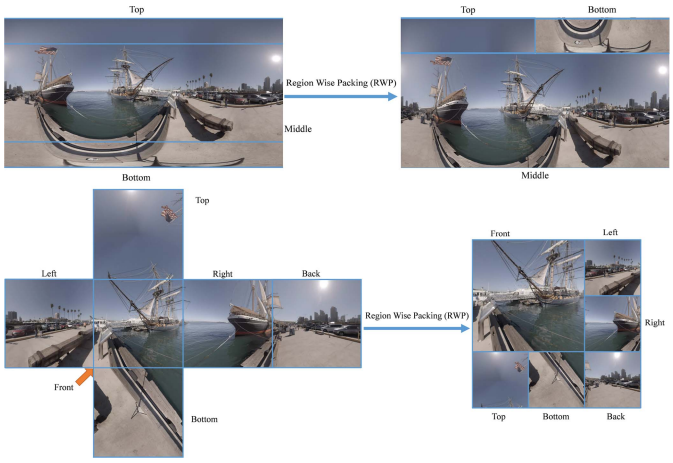

Fig. 15. Region-wise resampling and positioning for the ERP (top) and the CMP (bottom) projections.

requirements. The region-wise packing can be used to circumvent the weaknesses of the projection schemes and to reduce the computational complexity by downsampling the lesser important regions, i.e., top and bottom regions in case of ERP, and all the regions except the front face in case of CMP, as shown in Figure 15. The encoded frame contains the resulting packed regions. After encoding, the content is processed using the existing media file format and transport protocols together with some metadata to facilitate additional signaling for 3DoF navigation and selective viewport delivery to DASH clients [277].

The integration of OMAF with DASH comes with additional property descriptors included in the MPD to inform the client about 360° media properties. Several newly defined omnidirectional metadata are added in MPD, including Projection Format (PF) descriptor, Region-Wise Packing (RWPK) descriptor, Content Coverage (CC) descriptor, Spherical Region-wise Quality Ranking (SRQR) descriptor, 2D Region-Wise Quality Ranking (2DQR) descriptor, and Fisheye OMnidirectional Video (FOMV) descriptor. OMAF specifies nine media profiles, including three video profiles such as HEVC-based viewport-independent, HEVC-based viewport-dependent, and AVC-based viewport-dependent video profiles [278]. OMAF provides a consistent quality to the whole frame regardless of the viewport position for viewport-independent streaming. The regular HEVC codec and DASH streaming format can be used for viewport-independent streaming. However, the adaptive viewport-based operation using HEVC/AVC codec is a technical development of OMAF that allows unconstrained use of rectangular RWP for enhanced quality of viewport regions.

In 2016, MPEG approved the Common Media Application Format (CMAF) [279] that aims to provide a uniform encoding format and media profiles to be used across multiple applications and devices. The CMAF makes it possible to request lower latency segments. ISO Base Media File Format (ISOBMFF) [280] is the most popular file format for timed data exchange, management, and presentation. The ISOBMFF files consist of a series of compliant and extensible file-level boxes. Each box represents a data structure comprised of four

printer characters code. The ISOBMFF media and metadata streams kept in a track are distributed separately. The media data includes coded audio and video data. The metadata, similar to the conventional formats, represents media type, codec properties, timestamps, and size, etc. ISOBMFF specifies additional metadata for omnidirectional content such as projection format, rotation, frames packaging, encoding, and delivery. The track describes the format or content of samples, including the coding and packaging format. The video tracks that need post-processing after decoding for proper display are marked with the 'resv' entry type. ISOBMFF ensures the flexible aggregation of valuable information for easy access and inclusion in a transport manifest that supports efficient media consumption over the dynamic network.

B. Towards 6DoF+

MPEG-I divides the standardization of immersive media into three phases [281]. The phase 1a of MPEG-I aimed to complete essential monoscopic and stereoscopic 360° video services, also known as 3DoF, by 2017. In a 3DoF scenario, the user can freely move his head along three axes, i.e., yaw, pitch, and roll. If a user fastly moves his/her head while watching the stereoscopic video, the parallax error indicating the poor VR experience is observed. The phase 1b of the MPEG-I standard aims to support commercialized 3DoF+ services by the end of 2020. In 3DoF+, the user can freely look in any direction along with limited head movements in front/back, up/down, and left/right directions. 3DoF+ will advance the viewport quality to strengthen the sense of realism. If the view is missing in the original video, 3DoF+ headsets will be able to employ the reference intermediate view synthesizer [282] to synthesize the view accordingly. In March 2019, MPEG announced a Call for Proposals (CfPs) on 3DoF+ videos to develop advanced coding solutions and 3DoF+ metadata standardization. The purpose of the final phase, phase 2 of MPEG-I, is to support more elaborated services such as 6DoF by 2022. 6DoF visual system will provide full support for orientation and position tracking. The user will have free movements, similar to the real-world environment. Fig. 16 represents different possible movements of a user with respect to 3DoF, 3DoF+, and 6DoF.

Digital Video Broadcasting (DVBen)[19] created the Commercial Module VR (CM-VR) group to support commercial requirements for efficient delivery of VR media over digital video broadcast networks. Currently, the CM-VR research group targets the panoramic/3DoF+ content as highlighted within MPEG-I. Furthermore, DVB CM-VR Study Mission Group (DVB-VR-SMG) also aims to explore exceptional experiences, such as 6DoF. CM-VR group is also considering the work done by other organizations, e.g., MPEG, VRIF, and 3GPP, to ensure it is in synchronization with the latest technological developments.

C. Immersive Audio and Video Standards

Immersive audio and video is an enabling technology behind the media and entertainment industries. The VR Industry Forum (VRIF)[20] is composed of a wide range of participants from different sectors to provide high-quality audiovisual VR media experience to consumers. The main focus of the VRIF is to perceive 360° VR content accompanied by 3D spatial audio by pursuing VR guidelines. VRIF is focusing on building a content library that could benefit the industry to examine and promote the VR implementations.

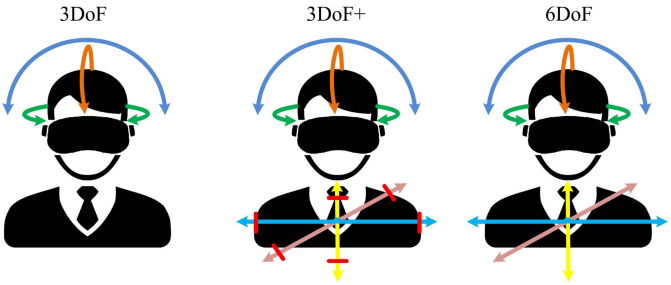

Fig. 16. Degree of freedom based viewing arrangements.

3GPP has documented the impact of VR by evaluating the relevance and potential of VR services for the industry. The document includes the audio-video media formats, delivery procedures, subjective tests as well as the latency and synchronization aspects [283]. 3GPP has put several efforts on immersive audio services with the primary intention to support codecs for VR and 3D Audio with enhanced voice services (EVS) extensions. The study items include codecs for VR audio (CODVRA) and codec for immersive voice and audio services (IVAS) to support potential standardization in this domain. The Streaming Video Alliance (SVA)[21] association covers the video ecosystem for developing best practices and specifications to promote the online video value chain. The SVA study group on VR and 360° videos focus on understanding the market potential, existing VR technologies such as players and use-cases, and cataloging standardization efforts. The Video Electronics Standards Association (VESA)[22] responsible for a lot of digital display standards, including VGA, DisplayPort, etc. has formed a special working group to explore the standardization of the emerging AR/VR applications. The group is responsible for: (i) establishing communication connections and data transfer protocols for AR/VR services, (ii) analyzing existing VESA standards to suggest any modifications, and (iii) collaborating with other standards organizations [284].

IEEE P2048 is one of the largest working group that involves participants from over 200 companies and institutions for the immersive media standardization. IEEE P2048.10/P2048.3 standards define the immersive audio/video file and streaming formats. IEEE P2048.9/P2048.2 deals with the taxonomy and quality metrics for the several variants of immersive audio/video services. These standards are required to reduce the confusion among variants of immersive media services. By categorizing these variants, IEEE

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

[20]https://www.vr-if.org/

[21]https://www.streamingvideoalliance.org//

[22]https://www.vesa.org/

TABLE VII
360° VR VIDEOS ON DIFFERENT MEDIA NETWORKS

Country	Media Networks	Section	URL
USA	CNN	CNN VR	https://edition.cnn.com/vr
	NBC news	News VR	https://www.nbcnews.com/storyline/nbc-news-vr
UK	BBC	VR	https://canvas-story.bbcrewind.co.uk/vr/
		360 VIDEO & VR-IMMERSIVE NEWS	https://bbcnewslabs.co.uk/projects/360-video-and-vr/
	Sky	Sky VR	https://www.sky.com/pages/vr/
Germany	WDR	360° -Videos	https://www1.wdr.de/virtual-reality-uebersicht-100.html
	ZDF	360° videos ǀ Virtual Reality on ZDF	https://vr.zdf.de/
France	France24	Virtual reality	https://www.france24.com/en/tag/virtual-reality/
Russia	Russia Today	RT360	https://www.rt.com/360/
Qatar	Al Jazeera	Contrast VR	https://contrastvr.com/

P2048.9/P2048.2 facilitate the development process and support the robust growth of the industry. IEEE P2048.4 standard deals with methods for maintaining a person's meaningful representation in a VR environment. IEEE P2048.6 deals with designing and developing different prototypes and techniques to strengthen fully immersive user experience. IEEE P2048.8 is closely related to MPEG-V and specifies different categories and levels of interoperability among the virtual and real worlds [285]. Furthermore, the IEEE P3333.3 standard provides technical guidance to reduce cybersickness or 3D sickness to protect the viewer's health and develop a healthy ecosystem.

JPEG Pleno [286] is an initiative by the Joint Photographic Experts Group (JPEG)[23] that targets to determine a standard system for recording, packaging, and transmission of immersive media content. It aims to provide efficient tools to support advanced content representation functionalities with limited resource utilization. The supported features include data and metadata manipulation, low latency, scalability, editing, JPEG compatibility, random access, privacy protection, and security. JPEG XT [287] is an image format standard for coding of 360° images. JPEG XT is backward compatible with the existing JPEG standards and involves multi-part specifications. The further effort by JPEG is JPEG360 Ad Hoc Group that was established in 2017 to ensure full interoperability for enabling interactions with multi-sensor images captured using omnidirectional cameras. The main objectives include defining use cases and requirements for 360° applications, updating metadata descriptions, gathering evidence of existing solutions, and facilitating the processes leading to the evaluation of new proposals. The functional metadata requirements include storing multiple images within a single file, stitching software, projection type, coding format, pixel format, and orientation, etc.

D. QoE Metrics

With the rapid development of immersive video technology, there is much need for advancements in precise quality assessment methodologies. Towards this cause, MPEG Joint Video Exploration Team (JVET) has defined some common test conditions and reference configurations for performance evaluation of 360° video [288]. MPEG-JVET has further investigated several quality assessment metrics by considering different aspects of 360° video. The 3GPP work item, FS_QoE_VR, aims at the identification of QoE metrics for VR content and device conditions and capabilities. QUALINET [289] focuses on QoE of multimedia systems and services. The Immersive Media Experience (IMEx) group of QUALINET focuses on a broad range of services including, mulsemedia, 360° VR, 3D audio, and future video coding solutions to develop methodologies and best practices for immersive media experience. The IMEx aims to identify application domains, use cases, software and hardware tools, and subjective quality assessment methodologies. It further supports standardization activities and liaison with other SDOs for collaborative research and mutual benefits.

Video Quality Experts Group (VQEG) has initiated Immersive Media Group (IMG) [290], which targets the quality assessment of immersive media applications. The collaboration of the IMG group of VQEG and the IMEx group of QUALINET leads to the development of the Joint QUALINET-VQEG team on Immersive Media (JQVIM). This joint venture aims to collect and produce immersive media content, tools, data sets, as well as maintaining standardization procedures and research activities for QoE assessment of AR/VR immersive media [271].

IX. POTENTIAL FOR IMMERSIVE VIDEO CONTENT

The potential of immersive video content concerning different use cases and projects is presented in this section. The universal applicability of 360° VR video has extended to many commercial sectors. Consumer adoption of the technology remains still in its early stages but is proving well popular in almost every field.

A. Applications

1) News Production and Broadcasters: 360° video, also called cinematic VR, is a format with an increasing presence in the news production due to its low-cost production [291]. Many public service media organizations have already seen the potential of this format, evidenced by the creation of applications or simply by the rate of consumption of 360° videos. 360° news production has evolved from its initial experimental

[23]https://jpeg.org/

TABLE VIII
OPEN SOURCE TOOLS FOR 360° VIDEO

Aspect	Tools	Description	Availability
Download	Youtube-dl	Command-line program to download videos from YouTube	https://github.com/ytdl-org/youtube-dl
Projection	360tools_conv	Convert 360° videos into different projection formats	https://github.com/Samsung/360tools
	360Lib	Projection conversion tool for 360° video	https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/
	360Transformations	Projection tool for viewport-dependent 360° video streaming	https://github.com/xmar/360Transformations
Transcode	FFMPEG	A multi-purpose tool to record, transcode, mux, demux, encode and decode the 360° video	https://ffmpeg.org
Codec	Kvazaar	Cross-platform HEVC encoder with tiling support	https://github.com/ultravideo/kvazaar
	HM	The HEVC reference software includes both encoder and decoder functionalities as well as the tiling support	https://hevc.hhi.fraunhofer.de/HM-doc/index.html
Packager	MP4Box	DASH packaging software for manipulating different media extensions. It divides tiled videos into equal duration video segments.	https://gpac.wp.imt.fr/mp4box/
	Video2DASH	Converts a 360° video (.mp4) to tiles and then to DASH segments (.m4s)	https://github.com/confiwent/video2DASH
Player	OMAF.js	HTML5 based 360° tiled player supporting HEVC-based viewport-dependent OMAF video profile.	https://github.com/fraunhoferhhi/omaf.js
	eleVR-Web-Player	The web player allows watching 360° video on Oculus Rift or mobile VR HMD	https://github.com/hawksley/eleVR-Web-Player
	VRClient	Python-based VR client without HMD support	https://github.com/jvdrhoof/VRClient
	WebVR	JavaScript API that provides web-based support for VR experience by using WebGL API	https://webvr.info/samples/
Head tracking	AVTrack360	Tracks the head posture data of users viewing 360° video	https://github.com/acmmmsys/2018-AVTrack360
	OpenTrack	Tracks the head posture data of users viewing 360° video	https://github.com/opentrack/opentrack
	VRTracker	Records head position data and eye gazing data	https://github.com/Archer-Tatsu/VRTracker
Saliency	V-BMS360	Model to predict saliency maps based on head movements	https://github.com/Telecommunication-Telemedia-Assessment/V-BMS360
	PanoSaliency	Model to generate saliency maps	https://github.com/phananh1010/PanoSaliency
	PanoSalNet	Saliency prediction model for improved head movement prediction	https://github.com/phananh1010/PanoSalNet
	Saliencymetric360	Saliency based quality evaluation metric for 360° video	https://github.com/mmspg/saliencymetric360
Quality assessment	Testbed360-android	A testbed for subjective quality assessment of 360° video	https://github.com/mmspg/testbed360-android
	Evaluation_VR	Subjective quality assessment tool for HTC Vive	https://github.com/Archer-Tatsu/Evaluation_VR-onebar-vive
	V-CNN	Convolutional neural network-based visual quality evaluation of 360° viewport	https://github.com/Archer-Tatsu/V-CNN
	360tools_metric	Supports various objective quality metrics for 360° video evaluation	https://github.com/Samsung/360tools
	Omnieval	Compute the quality difference between two 360° videos	https://github.com/mattcyu1/omnieval

phase to become a more critical part of several news organizations [292]. At the same time, the availability of capable cameras and their use in the fast-paced news environment has lightened the 360° news production, so it is natural to note that it is increasingly employed in Europe, America, and Asia. Table VII describes the 360° video sections on different public service media websites.

Many digital broadcasters are focusing on creating and publishing 360° VR content on different platforms. The basic aim of this prospective process for almost all companies is to identify the available techniques and learn about their potential. A major U.K. based broadcaster Sky[24] has shown much interest in this field. Yle[25] from Finland is actively testing next-generation media experiences. In this regard, Yle has a future media incubator called Yle Beta, which is lacking hardware but provides different ways of storytelling, such as iteration and pivoting. However, the progress of these ongoing efforts towards success needs to be further investigated.

2) Entertainment: The new immersive possibilities are endless with 360° video. The user can virtually attend 360° view of live sport with a favorite rotating seat, enjoy a live concert, watch movies or visit relatives in far-away locations. Small town classrooms will be able to virtually tour the amazing world sights, famous science laboratories, gigantic theme parks, and industries. Nowadays, TV cartoons also employ 360° video applications [293]. Regular videos are being replaced by 360° video to provide more creative and entertaining opportunities.

Theme parks create a tangible impact that can extend the user's sense of perception to the next levels [294]. Theme parks implement special audio effects and scenery to produce an appealing experience. Combining VR with powered roller coasters can increase the ridership of less popular attractions. The consumer gets 360° VR experience by wearing the VR goggles or glasses synchronized with the moving roller coaster. For instance, Canada's Wonderland[26] theme park offers Thunder Run VR attraction. Similarly, Alpenexpress Enzian[27] roller coaster located in the Europa-Park is a famous ride attraction where the visitors can experience the virtual environment with Samsung Gear VR headsets. The most

[24]https://www.sky.com/pages/vr/
[25]https://yle.fi/
[26]https://www.canadaswonderland.com/play/rides/thunder-run
[27]https://www.europapark.de/en/attractions/alpenexpress-enzian

modern VR fascination is "Star Wars: Secrets of the Empire" offered in Disney resorts where the users can touch, talk, and interact using dimensional set pieces. Similarly, the virtual walk-through tours facilitate a full 360° panorama by capturing the videos or still images of walkways and structures to allow an engaging experience.

3) Sports: 360° video is viral in worldwide sports. 360° video adds numerous advantages to almost every sport. The omnidirectional capturing around an athlete or sports ground brings the most beneficial experiences by providing the users with new viewing-angles. Gänsluckner *et al.* [109] presented blended learning based 360° climbing course to provide the participants with an interactive learning experience. The authors showed that for learning climbing techniques, 360° videos are much better than regular videos. Hebbel-Seeger [295] considered the use of 360° videos for education and training processes in sports. The authors determined that 360° video has high potential in the athletic training process, and there is a need to adopt modern storytelling means to attract user attention.

4) Medical Domain: Applications of 360° VR videos are not limited to entertainment or sports only. There are many more uses ranging from academic research to engineering, design, business, and arts, etc. The apparent and most practical applicability is in the medical field [296]. Immersive videos can be used to train physicians, neurosurgeons, and paramedics as a hopeful solution to the opioid epidemic and can effectively reduce patient pain. Doctors can analyze tumors and phobias without any scalpel. Rare syndromes can be reconstructed virtually for practicing purposes to minimize human error in a real environment. These advancements will inevitably lead to achieving significant time and cost-saving practices in both the training and teaching processes [297].

5) Education: In education, 360° video helps to present the scenarios that are complex to describe with conventional videos, words, or even images [298]. 360° recording of a classroom is thought to be a compelling tool for pre-service teachers in exploring different activities performed by students. A study of physical education using 360° video found that it helps the students to reconstruct the classroom situation and its meaning [110]. It has been assumed that complex human interactions captured by 360° videos can be played as many times as possible [299]. Furthermore, 360° video supports innovative activities in curricula. A case study presented by Kavanagh *et al.* [300] highlighted the impact of 360° videos from an education perspective. They identified different challenges concerning the video quality, direction, and handheld shooting, causing the "giant-hands" effect.

B. Research Projects

In parallel with infotainment, omnidirectional videos have intense activities regarding research. For instance, these videos expedite a non-invasive opportunity to collect the data for research group collaboration by using the observational schemes [301]. Several large-scale research projects have already considered using AR, VR, and omnidirectional video content. There are three important avenues which include proposing solutions to enhance the classic multimedia content and make it more immersive, creating tools and mechanisms for exchanging and displaying such rich media content at high-quality, and just generating and using innovative content in various contexts and societal areas. As can be expected, due to the complexity of the targeted task, most projects focus on the later, whereas the first two avenues have attracted less effort so far.

The NEWTON project[28] (ICT-20-2015) is a large-scale research and development initiative to design, develop, integrate, and disseminate innovative solutions in technology-enhanced learning (TEL). NEWTON project solutions include employing AR and VR content to increase learner quality of experience and targets both classic educational stream learners and students with special educational needs.

The REVEAL project[29] (ICT-24-2016) employs state-of-the-art VR technologies designed for gaming in education. The Reveal project develops solutions to use the PlayStation VR technologies for innovative educational applications to engage worldwide audiences and make them aware of European historical and scientific heritage.

The H-Reality project[30] (FETOPEN-01-2016-2017) aims to integrate commercial pioneers of ultrasonic haptics, state-of-the-art vibrotactile actuators, novel modeling of the skin and mechanics of touch, and psychophysical rendering of sensation to create a new sensorial experience involving digital 3D shapes and textures.

The Hyper360 project[31] (ICT-19-2017) is set to offer a complete end-to-end production toolset for enriched 360° videos, including new 3D storytelling elements, while also leveraging the powerful implicit preference extraction means that omnidirectional viewing offers, (i.e., the viewing direction), to build a personalization framework on top of this format and enable increased immersion.

Due to the highly limited number of large international projects in this space, there is increasing effort in this space driven by both market (e.g., 360° videos playout on PlayStation 4)[32] and diverse national and international funding agencies (e.g., EU Horizon 2020 ICT-25-2018-2020 call on Interactive Technologies).[33]

C. 360° Video Tools and Datasets

Several open-source tools for omnidirectional video preprocessing (i.e., downloading, mapping, transcoding, coding, packaging), playing, viewer head tracking, saliency computation, and quality assessment are presented in Table VIII.

[28]Networked Labs for Training in Sciences and Technologies for Information and Communication (NEWTON), http://newtonproject.eu.

[29]Realizing Education through Virtual Environments and Augmented Locations (Reveal), https://revealproject.eu.

[30]Mixed Haptic Feedback for Mid-Air Interactions in Virtual and Augmented Realities (H-Reality), https://cordis.europa.eu/project/rcn/216340_en.html.

[31]Enriching 360° media with 3D storytelling and personalization elements (Hyper360), http://www.hyper360.eu.

[32]"Viewing 360° images and videos on PlayStation VR" Blog, Oct. 2016, https://community.eu.playstation.com/t5/PS-VR-Support/Viewing-360°-images-and-videos-on-PS-VR/m-p/24599350.

[33]ICT-25-2018-2020 call, https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-25-2018-2020.html.

TABLE IX
360° VIDEO HEAD MOVEMENT DATASETS

Dataset	Participant	Size	Resolution	Duration	Description & Availability
David et al. [304]	57 (25 female, 32 male, age: 19 to 44, mean: 25.7)	19 videos	3840x1920 pixels	20s	HMD: HTC VIVE; Eye-tracker: SensoMotoric Instrument; All 19 videos were observed by all participants for their entire duration. https://salient360.ls2n.fr/datasets/
Wu et al. [155]	48 (24 female, 24 male, 64.6% are 20 to 25 years)	18 videos	1920x960-3840x1920 pixels	164s - 655s	HMD: HTC VIVE; Participants are free to look around in the first experiment with half videos and instructed with the question for the rest 9 videos in the second experiment. https://wuchlei-thu.github.io/
Lo et al. [305]	48 (age: 20 to 48)	10 videos	3840x2160 pixels	60s	HMD: Oculus Rift DK2; videos were observed by all participants for their entire duration.
Li et al. [306]	95 (56 female, 39 male, age: 18 to 24)	73 videos	Up to 3840x2160 pixels	37s - 640s	HMD: Oculus Rift CV1; http://vhil.stanford.edu/360-video-database/
Qian et al. [12]	130 (55 female, 75 male, age: 18 to 68))	10 videos	3840x2048 pixels	117s - 293s	HMD: Samsung Gear VR; A total of 1300 views from the participants with 4420 minutes of collected viewport traces.
Nasrabadi et al. [206]	60 (17 female, 43 male, age: 18 to 40+)	28 videos	Up to 3840x2160 pixels	60s	HMD: Oculus Go; Each participant watched 14 videos plus an introductory video, with each video being watched by 30 viewers. https://github.com/afshin-aero/360dataset
Bao et al. [183]	153 (57 female, 96 male, age: 10 to 60)	16 videos	Up to 3840x2048 pixels	30s	HMD: Oculus DK2; A total of 985 views recorded with total duration of 492 minutes. http://360videoexp.com/
Xu et al. [183]	45 (20 female, 20 male, age: 20 to 24)	208 videos	3840x2048 pixels	20-60s	HMD: HTC VIVE; Fixation and head positions are captured for a total of 210,000 frames. https://github.com/xuyanyu-shh/VR-EyeTracking/
Corbillon et al. [192]	59 (12 female, 47 male, age: 6 to 62)	5 videos	3840x2048 pixels	70s	HMD:Razer OSVR HDK2; Head movements are recorded in unit Hamiltons quaternions at the sampling rate of 30Hz. http://dash.ipv6.enstb.fr/headMovements/
Xu et al. [307]	DS1: 130 (54 female, 76 male) DS2: 54 (20 female, 34 male) DS3: 91 (45 female, 46 male)	10 videos	3840x2048 pixels	117-2930s	HMD:Samsung Gear VR; Three different datasets recorded on HMD, PC, and smartphone.

VR HMDs are likely to see broad adoption in the future due to the combination of innovative advanced hardware and improved viewing experience.The viewers can freely navigate to any part of the video by rotating their head. Several public datasets [12], [154], [182], [182], [191], [204], [302]–[305] provide the head movement traces of different participants watching 360° videos on popular HMDs. A summary of those datasets is listed in Table IX.

X. DISCUSSIONS AND CONCLUSION

The immense popularity of 360° video is due to the immersive user experience. 360° videos have different requirements related to frame rate, resolution, image quality, and user viewing behavior in comparison to regular videos. Further, there are increasing demands in terms of ultra-high network capacity and ultra-low response delay to support a high-quality streaming experience on existing, highly dynamic networks. This opens up novel avenues that have attracted the attention of many researchers.

Based on the latest developments, this work considered a streaming architecture for 360° video compatible with MPEG-DASH, from an end-user perspective. In this context, this survey presented a study of the streaming solutions, with a focus on 360° videos with details on content preparation, processing, and transmission to the end-user display devices such as VR headsets, smartphones, and high-resolution monitors, etc. Different adaptive streaming approaches were discussed, including viewport-independent, viewport-dependent, and tile-based schemes. Tile-based streaming has lower caching, storage, and processing complexities compared to viewport-dependent streaming. The edge-based adaptive and cooperative transmission of 360° tiles in the context of prefetching, caching, and fair distribution with the recorded user interactions such as head movements and eye-tracking are also presented. Edge nodes can reduce the influx of data to the core network, while the cloud can ease backup operation and resource coordination.

Several research challenges and research opportunities are presented with a focus on viewport prediction approaches, QoE assessment, and the impact of other constraints on 360° video network delivery. An uninterrupted 360° video should be displayed to the user in order to maintain high perception levels. Therefore, a proper QoE framework that focuses on all features of omnidirectional content is required because the existing 2D video models may suffer from critical limitations for 360° video. The perceived quality, spatial presence, as well as the technical parameters and viewer preferences, should be considered for the improved on-demand as well as 360° live video streaming. The survey also presents the latest international efforts in terms of large research projects, and standardization focused on the improvement of 360° video streaming and that of related media such as VR. Additionally, the growing potential of immersive rich media content in diverse application domains, which include medicine, infotainment, sports, and education, etc., has been discussed.

The development of new applications for diverse domains is natural; however, of concern is the need for handling other issues related to immersive media content delivery, including support for long-term accurate viewport prediction, QoE

assessment, and QoE-oriented delivery solutions. In conclusion, despite the amount of research done on 360° video streaming, several research challenges still exist and provide opportunities for further investigations in order to reach the goal of having a viable real-life implementation which provides users with the best experience.

XI. Future Research Directions

A detailed literature review included in this paper shows that despite the recency of the topic, many research efforts have already focused on 360° videos to provide smooth playback when delivered over bandwidth-limited highly dynamic networks. Even though the improvements introduced by the research and development community, high-quality 360° content creation, distribution, and streaming remain a major issue. Next, several potential research avenues of very much interest in this area are indicated.

For all innovative immersive multimedia, an important challenge remains the **projection/mapping** of an increased amount of content. In this context, the direct employment of the existing projection schemes help as they are supported by the existing graphics hardware. However, oversampling/undersampling may cause information loss in extracted viewports and degrades the performance of a VR system. A possible direction is developing techniques that enhance the projection processing functions and bandwidth utilization by natively allocating more pixels to the viewport.

In a receiver-centric architecture, multiple users are expected to watch different parts of the same content. It is necessary to upgrade the resolutions without further increasing the system latency. For low latency interactivity, adequate compression of 360° data units is required. Many issues related to **encoding** are open for further research. Certainly, new video coding techniques must be developed to attain higher compression efficiency and fast representation switching to offer lower latency and computational cost. Of particular interest are methods that enable intra-prediction and motion vector (MV) prediction across different quality zones. Moreover, by leveraging the tile-based design of HEVC, high-level parallelism for both encoder and decoder can be achieved.

Most of the existing literature on **edge-assisted** 360° video streaming ignore mobility and scalability in their fundamental design. A possible direction is to ensure scalability in design decisions. In addition, how the support for high-speed users will impact the overall service provisioning, task offloading, scheduling, and resource allocation needs further investigations. Edge nodes placed near to users are more vulnerable to security attacks than cloud data centers. One can design secure edge node sites, strong access-control policies, and privacy-aware offloading and load balancing schemes to support secure VR communication. Similarly, designing energy-aware computation offloading schemes are also a promising direction.

The high fidelity 8K, 12K, or 16K media services with a high degree of freedom support add pressure to the overall capacity of existing networks. It is the next-generation **mmWave wireless technology** with several-Gbps communication capabilities that are expected to offer ultra-high capacity, ultra-high transmission reliability, ultra-low latency, ultra-high mobility, and massive communication support for rich media immersive applications over short or long ranges [306]. In the future, transmission of the uncompressed part (viewport only) to multiple mobile users with different viewing patterns and preferences over 5G links can improve the performance of QoS-sensitive applications. Studies of adaptive computing offloading using mmWave communication can support synchronized 360° streaming in both indoor and outdoor environments. In addition, dynamically resources allocation in edge-clouds assisted HetNets is fundamental to secure a fully immersive and high-quality user experience.

The number of proposals handling **tiling**-related challenges in 360° video streaming has witnessed intense attention from researchers. They empower differentiated streaming of quality-variable tiles based on user interest to maintain a necessary balance between viewport quality and bandwidth utilization. Many factors affect this balance and additional work is still needed to improve QoE, cost, and bandwidth consumption. Therefore, improved design insights should be considered for interactive selection of tiles along with optimum bandwidth allocation for tiles in the context of prioritized 360° video distribution. Machine-learning-based bandwidth prediction approaches can be utilized for instance to capture the actual bandwidth patterns for bitrate allocation. Different quality levels for tiles in a frame introduces artifacts, especially at the boundary of tiles. Besides, the influence of temporal quality variations in the viewing region on user QoE requires further investigations [307]. Most of the existing works focus on bandwidth-efficient streaming using a fixed tiling pattern. How the variable size tiling impacts the overall streaming performance and at which cost is another interesting question needs to investigate. Moreover, the multipath-based transmission of high-resolution 360° tiles can bring better performance and more flexibility by delivering the high-priority tiles through the best available paths. Care should be taken however to avoid out-of-order delivery.

A fundamental challenge to 360° video streaming is **viewport prediction**. Despite the abundant research history of saliency detection, the existing approaches may provide inaccurate viewport estimation. A research direction, hence, is to identify situations where the saliency approaches are accurate for all users by precisely determining the prediction accuracy. Motion-based saliency estimation [308] and dynamic modeling of the user interest, especially for multiple VR users are useful for long-term viewport prediction. A recent work [193] is a potential starting point for both head and body movement prediction. Besides, one can improve the performance of viewport-dependent streaming by dynamically deciding the coverage of the viewport and peripheral regions based on user head movements and prediction errors.

Naturally, the **quality assessment** of 360° videos relates to a broad range of technical research. Although the growing interest in both subjective and objective quality assessment in recent years, research is still in its early stages. Challenges include learning how to define the test protocols needed during the subjective assessment for 360° videos, how to aggregate the quality measured by different users under different rating

scales, how to assess the different network and content characteristics impacts on the overall quality ratings, how to plan the test sessions to minimize the simulator sickness scores, how to assess the quality loss by considering the eye movements within the viewport, how to develop objective metrics that consider the heterogeneous quality of omnidirectional video and how to statistically analyze the effectiveness of objective quality metrics in correlation with MOS considering large-scale datasets.

Live broadcasting raises numerous concerns such as handling user interactivity, transcoding decisions, estimation of the bandwidth, achieving fairness, and smooth quality streaming, which are especially critical in an immersive multimedia context. Detailed studies for live VR streaming should be performed focusing on diverse aspects such as inter-stream fairness, user scheduling, taming latency, network traffic balancing, user's feedback, and QoS consideration in a mobile environment. Modifying the workflows of existing CNNs to predict the viewport for interactive VR live broadcasting is a direction worth exploring.

REFERENCES

[1] C. Westphal, "Challenges in networking to support augmented reality and virtual reality," in *Proc. Int. Conf. Comput. Netw. Commun. (ICNC)*, 2017, pp. 26–29.

[2] J. Boyce and E. Alshina, *JVET Common Test Conditions and Evaluation Procedures for 360° Video*, document JVET-D1030, JVE Team of ITU-T VCEG and ISO/IEC MPEG, Chengdu, China, 2016.

[3] "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021," Cisco, San Jose, CA, USA, White Paper, 2017.

[4] Netflix. (2014). *Internet Connection Speed Recommendations*. [Online]. Available: https://help.netflix.com/en/node/306

[5] J. Thompson, J. Sun, R. Möller, M. Sintorn, and G. Huston, "Q1 2017 state of the internet-connectivity report," Akamai, Cambridge, MA, USA, Rep., 2017.

[6] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the Internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011.

[7] *Dynamic Adaptive Streaming Over HTTP (DASH): Media Presentation Description and Segment Formats*, ISO/IEC Standard 23009-1:2014, 2014.

[8] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. 8th ACM Multimedia Syst. Conf. (MMSys)*, 2017, pp. 261–271. [Online]. Available: http://doi.acm.org/10.1145/3083187.3084016

[9] Y. Sánchez, R. Skupin, and T. Schierl, "Compressed domain video processing for tile based panoramic streaming using HEVC," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2015, pp. 2244–2248.

[10] S. Heymann *et al.*, "Representation, coding and interactive rendering of high-resolution panoramic images and video using MPEG-4," in *Proc. Panoramic Photogrammetry Workshop (PPW)*, 2005, pp. 29–38.

[11] J. Le Feuvre and C. Concolato, "Tiled-based adaptive streaming using MPEG-DASH," in *Proc. ACM 7th Int. Conf. Multimedia Syst.*, 2016, p. 41.

[12] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan, "Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 99–114.

[13] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360 video streaming using tiles for virtual reality," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 2174–2178.

[14] C. Ozcinar, J. Cabrera, and A. Smolic, "Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 217–230, Mar. 2019.

[15] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, 2016, pp. 107–110.

[16] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," in *Proc. 24th ACM Int. Conf. Multimedia*, 2016, pp. 601–605.

[17] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proc. ACM 5th Workshop All Things Cellular Oper. Appl. Challenges*, 2016, pp. 1–6.

[18] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017, pp. 1–7.

[19] H. Pang, C. Zhang, F. Wang, J. Liu, and L. Sun, "Towards low latency multi-viewpoint 360° interactive video: A multimodal deep reinforcement learning approach," in *Proc. IEEE INFOCOM IEEE Conf. Comput. Commun.*, 2019, pp. 991–999.

[20] K. M. Stanney, R. R. Mourant, and R. S. Kennedy, "Human factors issues in virtual environments: A review of the literature," *Presence Teleoper. Virtual Environ.*, vol. 7, no. 4, pp. 327–351, 1998. [Online]. Available: https://doi.org/10.1162/105474698565767

[21] S. Shi, V. Gupta, M. Hwang, and R. Jana, "Mobile VR on edge cloud: A latency-driven design," in *Proc. 10th ACM Multimedia Syst. Conf.*, 2019, pp. 222–231.

[22] M. Xu, C. Li, S. Zhang, and P. Le Callet, "State-of-the-art in 360 video/image processing: Perception, assessment and compression," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 5–26, May 2020.

[23] M. Zink, R. K. Sitaraman, and K. Nahrstedt, "Scalable 360° video stream delivery: Challenges, solutions, and opportunities," *Proc. IEEE*, vol. 107, no. 4, pp. 639–650, 2019.

[24] R. G. D. A. Azevedo, N. Birkbeck, F. De Simone, I. Janatra, B. Adsumilli, and P. Frossard, "Visual distortions in 360-degree videos," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jul. 8, 2019, doi: 10.1109/TCSVT.2019.2927344.

[25] D. He, C. Westphal, and J. J. Garcia-Luna-Aceves, "Network support for AR/VR and immersive video application: A survey," in *Proc. ICETE*, 2018, pp. 525–535.

[26] T. El-Ganainy and M. Hefeeda, "Streaming virtual reality content," Dec. 2016. [Online]. Available: arXiv:1612.08350.

[27] Y. Bernet *et al.*, "A framework for integrated services operation over Diffserv networks," IETF, RFC 2998, 2000.

[28] S. Blake, D. L. Black, M. A. Carlson, E. B. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," IETF, RFC 2475, 1998.

[29] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource reservation protocol (RSVP)—Version 1 functional specification," IETF, RFC 2205, 1996.

[30] H. Schulzrinne, S. L. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," IETF, RFC 1889, 1996.

[31] J. Postel, "User datagram protocol," IETF, RFC 768, 1980.

[32] T. Friedman, R. Cáceres, and A. Clark, "RTP control protocol extended reports (RTCP XR)," IETF, RFC 3611, 2003.

[33] H. Schulzrinne, A. Rao, and R. Lanphier, "Real time streaming protocol (RTSP)," IETF, RFC 2326, 1998.

[34] J. Postel, "Transmission control protocol," IETF, RFC 793, 1981.

[35] J. Kua, G. Armitage, and P. Branch, "A survey of rate adaptation techniques for dynamic adaptive streaming over HTTP," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1842–1866, 3rd Quart., 2017.

[36] T. Stockhammer *et al.*, "Dynamic adaptive streaming over HTTP-design principles and standards," in *Proc. 2nd Annu. ACM Conf. Multimedia Syst.*, 2011, pp. 133–144. [Online]. Available: https://doi.org/10.1145/1943552.1943572

[37] D. Le Gall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, no. 4, pp. 46–59, 1991.

[38] J. Summers, T. Brecht, D. Eager, and B. Wong, "Methodologies for generating HTTP streaming video workloads to evaluate Web server performance," in *Proc. 5th Annu. Int. Syst. Storage Conf.*, 2012, pp. 1–12.

[39] T. Stockhammer, "Dynamic adaptive streaming over HTTP—Standards and design principles," in *Proc. 2nd Annu. ACM Conf. Multimedia Syst.*, 2011, pp. 133–144.

[40] T. Bi, A. Pichon, L. Zou, S. Chen, G. Ghinea, and G.-M. Muntean, "A DASH-based multimedia adaptive delivery solution," in *Proc. ACM 10th Int. Workshop Immersive Mixed Virtual Environ. Syst.*, 2018, pp. 1–6.

[41] R. Castagno and D. Singer, "MIME type registrations for 3rd generation partnership project (3GPP) multimedia files," IETF, RFC 3839, 2004.

[42] H. Garudadri, "MIME type registrations for 3GPP2 multimedia files," IETF, RFC 4393, 2006.

[43] O. A. Niamut, E. Thomas, L. D'Acunto, C. Concolato, F. Denoual, and S. Y. Lim, "MPEG DASH SRD: Spatial relationship description," in *Proc. 7th Int. Conf. Multimedia Syst.*, 2016, p. 5.

[44] D. Salomon, *Transformations and Projections in Computer Graphics*. London, U.K.: Springer, 2007.

[45] E. Kuzyakov and D. Pio. (2016). *Next-Generation Video Encoding Techniques for 360 Video and VR*. [Online]. Available: https://code.facebook.com/posts/1126354007399553/nextgeneration-video-encoding-techniques-for-360-video-and-vr

[46] G. V. der Auwera, H. M. Coban, and M. Karczewicz. *Truncated Square Pyramid Projection (TSP) for 360 Video*, document JVET D0071, Joint Video Experts Team, 2016.

[47] C. Zhou, Z. Li, and Y. Liu, "A measurement study of oculus 360 degree video streaming," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 27–37.

[48] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "VR is on the edge: How to deliver 360 videos in mobile networks," in *Proc. ACM Workshop Virtual Real. Augmented Real. Netw.*, 2017, pp. 30–35.

[49] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[50] V. Sze, M. Budagavi, and G. J. Sullivan, "High efficiency video coding (HEVC)," in *Integrated Circuit and Systems, Algorithms and Architectures*, vol. 39. Cham, Switzerland: Springer, 2014, p. 40.

[51] G. J. Sullivan, "Video coding standards progress report: Joint video experts team launches the versatile video coding project," *SMPTE Motion Imag. J.*, vol. 127, no. 8, pp. 94–98, 2018.

[52] V. Avelar, "Practical options for deploying small server rooms and micro data centers," Schneider Elect., Rueil-Malmaison, France, White Paper, 2015.

[53] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput.*, 2012, pp. 13–16.

[54] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," ETSI, Sophia Antipolis, France, White Paper, 2015.

[55] A. Manzalini and N. Crespi, "An edge operating system enabling anything-as-a-service," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 62–67, Mar. 2016.

[56] Fraunhofer FOKUS. (2016). *Cloud-Based 360° Video Playout*. [Online]. Available: https://www.fokus.fraunhofer.de/go/360

[57] G.-M. Muntean and N. Cranley, "Resource efficient quality-oriented wireless broadcasting of adaptive multimedia content," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 362–368, Mar. 2007.

[58] G.-M. Muntean, P. Perry, and L. Murphy, "A new adaptive multimedia streaming system for all-IP multi-service networks," *IEEE Trans. Broadcast.*, vol. 50, no. 1, pp. 1–10, Mar. 2004.

[59] Z. Yuan and G.-M. Muntean, "A prioritized adaptive scheme for multimedia services over IEEE 802.11 WLANs," *IEEE Trans. Netw. Service Manag.*, vol. 10, no. 4, pp. 340–355, Dec. 2013.

[60] L. Zou, R. Trestian, and G.-M. Muntean, "DOAS: Device-oriented adaptive multimedia scheme for 3GPP LTE systems," in *Proc. IEEE 24th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, 2013, pp. 2180–2184.

[61] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann, "A survey on bitrate adaptation schemes for streaming media over HTTP," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 562–585, 1st Quart., 2019.

[62] K. Evensen, D. Kaspar, C. Griwodz, P. Halvorsen, A. Hansen, and P. Engelstad, "Improving the performance of quality-adaptive video streaming over multiple heterogeneous access networks," in *Proc. 2nd Annu. ACM Conf. Multimedia Syst.*, 2011, pp. 57–68.

[63] T. C. Thang, Q.-D. Ho, J. W. Kang, and A. T. Pham, "Adaptive streaming of audiovisual content using MPEG DASH," *IEEE Trans. Consum. Electron.*, vol. 58, no. 1, pp. 78–85, Mar. 2012.

[64] C. Liu, I. Bouazizi, and M. Gabbouj, "Rate adaptation for adaptive HTTP streaming," in *Proc. 2nd Annu. ACM Conf. Multimedia Syst.*, 2011, pp. 169–174.

[65] C. Liu, I. Bouazizi, M. M. Hannuksela, and M. Gabbouj, "Rate adaptation for dynamic adaptive streaming over HTTP in content distribution network," *Signal Process. Image Commun.*, vol. 27, no. 4, pp. 288–311, 2012.

[66] J. Jiang, V. Sekar, and H. Zhang, "Improving fairness, efficiency, and stability in HTTP-based adaptive video streaming with festive," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 326–340, Dec. 2014.

[67] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, *A Quantitative Measure of Fairness and Discrimination*, Eastern Res. Lab., Digit. Equip. Corporat., Hudson, MA, USA, 1984.

[68] Z. Li *et al.*, "Probe and adapt: Rate adaptation for HTTP video streaming at scale," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 719–733, May 2014.

[69] A. Zambelli, *IIS Smooth Streaming Technical Overview*, Microsoft Corporat., Albuquerque, NM, USA, 2009.

[70] M. Xiao, V. Swaminathan, S. Wei, and S. Chen, "DASH2M: Exploring HTTP/2 for Internet streaming to mobile devices," in *Proc. ACM Multimedia Conf.*, 2016, pp. 22–31.

[71] Y. Sun *et al.*, "CS2P: Improving video bitrate selection and adaptation with data-driven throughput prediction," in *Proc. ACM SIGCOMM Conf.*, 2016, pp. 272–285.

[72] J. Jiang, V. Sekar, H. Milner, D. Shepherd, I. Stoica, and H. Zhang, "{CFA}: A practical prediction system for video QoE optimization," in *Proc. 13th Symp. Netw. Syst. Design Implement. (NSDI)*, 2016, pp. 137–150.

[73] J. Jiang, S. Sun, V. Sekar, and H. Zhang, "PytheAS: Enabling data-driven quality of experience optimization using group-based exploration-exploitation," in *Proc. 14th Symp. Netw. Syst. Design Implement. (NSDI)*, 2017, pp. 393–406.

[74] Z. Yuan, H. Venkataraman, and G.-M. Muntean, "iBE: A novel bandwidth estimation algorithm for multimedia services over IEEE 802.11 wireless networks," in *Proc. IFIP/IEEE Int. Conf. Manag. Multimedia Netw. Services*, 2009, pp. 69–80.

[75] W. ur Rahman and K. Chung, "A novel adaptive logic for dynamic adaptive streaming over HTTP," *J. Vis. Commun. Image Represent.*, vol. 49, pp. 433–446, Oct. 2017.

[76] A. T. Nasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using scalable video coding," in *Proc. ACM Multimedia Conf.*, 2017, pp. 1689–1697.

[77] C. Mueller, S. Lederer, R. Grandl, and C. Timmerer, "Oscillation compensating dynamic adaptive streaming over HTTP," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2015, pp. 1–6.

[78] T.-Y. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson, "A buffer-based approach to rate adaptation: Evidence from a large video streaming service," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 187–198, 2015.

[79] K. Spiteri, R. Urgaonkar, and R. K. Sitaraman, "BOLA: Near-optimal bitrate adaptation for online videos," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2016, pp. 1–9.

[80] A. Beben, P. Wiśniewski, J. M. Batalla, and P. Krawiec, "ABMA+: Lightweight and efficient algorithm for HTTP adaptive streaming," in *Proc. ACM 7th Int. Conf. Multimedia Syst.*, 2016, p. 2.

[81] C. Sieber, T. Hoßfeld, T. Zinner, P. Tran-Gia, and C. Timmerer, "Implementation and user-centric comparison of a novel adaptation logic for DASH with SVC," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manag. (IM)*, 2013, pp. 1318–1323.

[82] G. Tian and Y. Liu, "Towards agile and smooth video adaptation in dynamic HTTP streaming," in *Proc. ACM 8th Int. Conf. Emerg. Netw. Exp. Technol.*, 2012, pp. 109–120.

[83] J. Gettys, "Bufferbloat: Dark buffers in the Internet," *IEEE Internet Comput.*, vol. 15, no. 3, p. 96, May/Jun. 2011.

[84] A. Mansy, B. Ver Steeg, and M. Ammar, "Sabre: A client based technique for mitigating the buffer bloat effect of adaptive video flows," in *Proc. 4th ACM Multimedia Syst. Conf.*, 2013, pp. 214–225.

[85] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, "360ProbDASH: Improving QoE of 360 video streaming using tile-based HTTP adaptive streaming," in *Proc. ACM Multimedia Conf.*, 2017, pp. 315–323, doi: 10.1145/3123266.3123291.

[86] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A control-theoretic approach for dynamic adaptive video streaming over HTTP," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 325–338, 2015.

[87] A. Yaqoob, T. Bi, and G.-M. Muntean, "A DASH-based efficient throughput and buffer occupancy-based adaptation algorithm for smooth multimedia streaming," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 643–649.

[88] K. Miller, E. Quacchio, G. Gennari, and A. Wolisz, "Adaptation algorithm for adaptive streaming over HTTP," in *Proc. 19th Int. Packet Video Workshop (PV)*, May 2012, pp. 173–178.

[89] D. J. Vergados, A. Michalas, A. Sgora, D. D. Vergados, and P. Chatzimisios, "FDASH: A fuzzy-based MPEG/DASH adaptation algorithm," *IEEE Syst. J.*, vol. 10, no. 2, pp. 859–868, Jun. 2016.

[90] A. Sobhani, A. Yassine, and S. Shirmohammadi, "A video bitrate adaptation and prediction mechanism for HTTP adaptive streaming," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 2, p. 18, 2017.

[91] P. Kaufman, *Smarter Trading*, vol. 22. New York, NY, USA: McGraw-Hill, 1995.

[92] S. Liu and J. Y. L. Forrest, *Grey Systems: Theory and Applications*. Berlin, Germany: Springer-Verlag, 2010.

[93] C. Wang, A. Rizk, and M. Zink, "SQUAD: A spectrum-based quality adaptation for dynamic adaptive streaming over HTTP," in *Proc. 7th Int. Conf. Multimedia Syst.*, 2016, p. 1.

[94] C. Zhou, C.-W. Lin, X. Zhang, and Z. Guo, "TFDASH: A fairness, stability, and efficiency aware rate control approach for multiple clients over DASH," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 198–211, Jan. 2019.

[95] S. Q. Jabbar, D. J. Kadhim, and Y. Li, "Proposed an adaptive bitrate algorithm based on measuring bandwidth and video buffer occupancy for providing smoothly video streaming," *Technology*, vol. 9, no. 2, pp. 191–195, 2018.

[96] W. Zhang, S. Ye, B. Li, H. Zhao, and Q. Zheng, "A priority-based adaptive scheme for multi-view live streaming over HTTP," *Comput. Commun.*, vol. 85, pp. 89–97, Jun. 2016.

[97] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "TCP extensions for multipath operation with multiple addresses," IETF, RFC 6824, 2013.

[98] B. Hesmans, F. Duchene, C. Paasch, G. Detal, and O. Bonaventure, "Are TCP extensions middlebox-proof?" in *Proc. ACM Workshop Hot Topics Middleboxes Netw. Function Virtualization*, 2013, pp. 37–42.

[99] C. Xu, T. Liu, J. Guan, H. Zhang, and G.-M. Muntean, "CMT-QA: Quality-aware adaptive concurrent multipath data transfer in heterogeneous wireless networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 11, pp. 2193–2205, Nov. 2013.

[100] Y.-C. Chen, D. Towsley, and R. Khalili, "MSPlayer: Multi-source and multi-path video streaming," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2198–2206, Aug. 2016.

[101] C. Xu, Z. Li, J. Li, H. Zhang, and G.-M. Muntean, "Cross-layer fairness-driven concurrent multipath video delivery over heterogeneous wireless networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 7, pp. 1175–1189, Jul. 2015.

[102] Y. Kim and K. Chung, "Multipath-based transmission scheme for improving the QoE of HTTP adaptive streaming," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 12–20, Aug. 2018.

[103] Y. Go, O. C. Kwon, and H. Song, "An energy-efficient HTTP adaptive video streaming with networking cost constraint over heterogeneous wireless networks," *IEEE Trans. Multimedia*, vol. 17, no. 9, pp. 1646–1657, Jul. 2015.

[104] A. Elgabli, K. Liu, and V. Aggarwal, "Optimized preference-aware multi-path video streaming with scalable video coding," *IEEE Trans. Mobile Comput.*, vol. 19, no. 1, pp. 159–172, Jan. 2020.

[105] D. Yun and K. Chung, "DASH-based multi-view video streaming system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 8, pp. 1974–1980, Aug. 2018.

[106] W. U. Rahman and K. Chung, "A multi-path-based adaptive scheme for multi-view streaming over HTTP," *IEEE Access*, vol. 6, pp. 77869–77879, 2018.

[107] K. K. Sreedhar, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, 2016, pp. 583–586.

[108] S. Afzal, J. Chen, and K. Ramakrishnan, "Characterization of 360-degree videos," in *Proc. ACM Workshop Virtual Real. Augmented Real. Netw.*, 2017, pp. 1–6.

[109] M. Gänsluckner, M. Ebner, and I. Kamrat, *360 Degree Videos Within a Climbing MOO*, Int. Assoc. Develop. Inf. Soc., Lisbon, Portugal, 2017.

[110] L. Roche and N. Gal-Petitfaux, "Using 360 video in physical education teacher education," in *Proc. Soc. Inf. Technol. Teacher Educ. Int. Conf.*, 2017, pp. 3420–3425.

[111] C. Kelling, H. Väätäjä, and O. Kauhanen, "Impact of device, context of use, and content on viewing experience of 360-degree tourism video," in *Proc. ACM 16th Int. Conf. Mobile Ubiquitous Multimedia*, 2017, pp. 211–222.

[112] R. Skupin, Y. Sanchez, D. Podborski, C. Hellge, and T. Schierl, "Viewport-dependent 360 degree video streaming based on the emerging omnidirectional media format (OMAF) standard," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, p. 4592.

[113] X. Corbillon, A. Devlic, G. Simon, and J. Chakareski, "Optimal set of 360-degree videos for viewport-adaptive streaming," in *Proc. ACM Multimedia Conf.*, 2017, pp. 943–951.

[114] D. He, C. Westphal, and J. Garcia-Luna-Aceves, "Joint rate and FoV adaptation in immersive video streaming," in *Proc. ACM Sigcomm Workshop AR/VR Netw.*, 2018, pp. 27–32.

[115] D. Naik, I. D. Curcio, and H. Toukomaa, "Optimized viewport dependent streaming of stereoscopic omnidirectional video," in *Proc. ACM 23rd Packet Video Workshop*, 2018, pp. 37–42.

[116] C. Zhou, M. Xiao, and Y. Liu, "ClusTile: Toward minimizing bandwidth in 360-degree video streaming," in *Proc. IEEE INFOCOM IEEE Conf. Comput. Commun.*, 2018, pp. 962–970.

[117] A. Ghosh, V. Aggarwal, and F. Qian, "A rate adaptation algorithm for tile-based 360-degree video streaming," Apr. 2017. [Online]. Available: arXiv:1704.08215.

[118] J. van der Hooft, M. T. Vega, S. Petrangeli, T. Wauters, and F. De Turck, "Optimizing adaptive tile-based virtual reality video streaming," in *Proc. IFIP/IEEE Symp. Integr. Netw. Service Manag. (IM)*, 2019, pp. 381–387.

[119] D. V. Nguyen, H. T. T. Tran, A. T. Pham, and T. C. Thang, "An optimal tile-based approach for viewport-adaptive 360-degree video streaming," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 29–42, Mar. 2019.

[120] R. Skupin, Y. Sanchez, D. Podborski, C. Hellge, and T. Schierl, "HEVC tile based streaming to head mounted displays," in *Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, 2017, pp. 613–615.

[121] J. Le Feuvre, C. Concolato, and J.-C. Moissinac, "GPAC: Open source multimedia framework," in *Proc. 15th ACM Int. Conf. Multimedia*, 2007, pp. 1009–1012.

[122] L. D'Acunto, J. van den Berg, E. Thomas, and O. Niamut, "Using MPEG DASH SRD for zoomable and navigable video," in *Proc. ACM 7th Int. Conf. Multimedia Syst.*, 2016, p. 34.

[123] H.-W. Kim and S.-H. Yang, "Region of interest-based segmented tiled adaptive streaming using head-mounted display tracking sensing data," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 12, pp. 1–24, 2019.

[124] H. W. Kim, J. W. Yang, J. Y. Yang, J. H. Jang, and W. C. Park, "MPEG-DASH SRD based 360 VR tiled streaming system for foveated rendering," in *Proc. IEEE Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, 2018, pp. 587–591.

[125] R. Skupin, Y. Sanchez, K. Suehring, T. Schierl, E. Ryu, and J. Son, "Temporal MCTS coding constraints implementation," in *Proc. 120th MPEG Meeting ISO/IEC JTC1/SC29/WG11*, vol. 120, 2017, Art. no. m41626.

[126] J. Son, D. Jang, and E.-S. Ryu, "Implementing motion-constrained tile and viewport extraction for VR streaming," in *Proc. 28th ACM SIGMM Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2018, pp. 61–66.

[127] S. Lee, D. Jang, J. Jeong, and E.-S. Ryu, "Motion-constrained tile set based 360-degree video streaming using saliency map prediction," in *Proc. 29th ACM Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2019, pp. 20–24.

[128] X. Zhang, X. Hu, L. Zhong, S. Shirmohammadi, and L. Zhang, "Cooperative tile-based 360-degree panoramic streaming in heterogeneous networks using scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 217–231, Jan. 2020.

[129] D. V. Nguyen, H. Van Trung, H. L. D. Huong, T. T. Huong, N. P. Ngoc, and T. C. Thang, "Scalable 360 video streaming using HTTP/2," in *Proc. IEEE 21st Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2019, pp. 1–6.

[130] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 285–286, Jan. 2005.

[131] H. Mao, R. Netravali, and M. Alizadeh, "Neural adaptive video streaming with pensieve," in *Proc. Conf. ACM Special Interest Group Data Commun.*, 2017, pp. 197–210.

[132] J. van der Hooft, S. Petrangeli, M. Claeys, J. Famaey, and F. De Turck, "A learning-based algorithm for improved bandwidth-awareness of adaptive streaming clients," in *Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manag. (IM)*, 2015, pp. 131–138.

[133] D. V. Nguyen, H. T. Tran, and T. C. Thang, "Adaptive tiling selection for viewport adaptive streaming of 360-degree video," *IEICE Trans. Inf. Syst.*, vol. 102, no. 1, pp. 48–51, 2019.

[134] J. Fu, X. Chen, Z. Zhang, S. Wu, and Z. Chen, "360SRL: A sequential reinforcement learning approach for ABR tile-based 360 video streaming," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2019, pp. 290–295.

[135] X. Jiang, Y.-H. Chiang, Y. Zhao, and Y. Ji, "Plato: Learning-based adaptive streaming of 360-degree videos," in *Proc. IEEE 43rd Conf. Local Comput. Netw. (LCN)*, 2018, pp. 393–400.

[136] V. Mnih *et al.*, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.

[137] W. Quan, Y. Pan, B. Xiang, and L. Zhang, "Reinforcement learning driven adaptive VR streaming with optical flow based QoE," Mar. 2020. [Online]. Available: [Online]. Available: arXiv:2003.07583.

[138] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Aug. 2017.

[139] N. Kan, J. Zou, K. Tang, C. Li, N. Liu, and H. Xiong, "Deep reinforcement learning-based rate adaptation for adaptive 360-degree video streaming," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 4030–4034.

[140] G. Xiao, X. Chen, M. Wu, and Z. Zhou, "Deep reinforcement learning-driven intelligent panoramic video bitrate adaptation," in *Proc. ACM Turing Celebration Conf. China*, 2019, pp. 1–5.

[141] Y. Zhang, P. Zhao, K. Bian, Y. Liu, L. Song, and X. Li, "DRL360: 360-degree video streaming with deep reinforcement learning," in *Proc. IEEE Conf. Comput. Commun. (IEEE INFOCOM)*, 2019, pp. 1252–1260.

[142] H. Wang, V.-T. Nguyen, W. T. Ooi, and M. C. Chan, "Mixing tile resolutions in tiled video: A perceptual quality assessment," in *Proc. ACM Netw. Oper. Syst. Support Digit. Audio Video Workshop*, 2014, p. 25.

[143] S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, "Edge caching and computing in 5G for mobile AR/VR and tactile Internet," *IEEE MultiMedia*, vol. 26, no. 1, pp. 21–30, Nov. 2018.

[144] K. Bilal and A. Erbad, "Edge computing for interactive media and video streaming," in *Proc. IEEE 2nd Int. Conf. Fog Mobile Edge Comput. (FMEC)*, 2017, pp. 68–73.

[145] M. Erol-Kantarci and S. Sukhmani, "Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G," in *Ad Hoc Networks*, Y. Zhou and T. Kunz, Eds. Cham, Switzerland: Springer Int., 2018, pp. 169–177.

[146] L. Lin, X. Liao, H. Jin, and P. Li, "Computation offloading toward edge computing," *Proc. IEEE*, vol. 107, no. 8, pp. 1584–1607, Jun. 2019.

[147] X. Hou, Y. Lu, and S. Dey, "Wireless VR/AR with edge/cloud computing," in *Proc. IEEE 26th Int. Conf. Comput. Commun. Netw. (ICCCN)*, 2017, pp. 1–8.

[148] W. Zhang, J. Chen, Y. Zhang, and D. Raychaudhuri, "Towards efficient edge cloud augmentation for virtual reality MMOGS," in *Proc. 2nd ACM/IEEE Symp. Edge Comput.*, 2017, p. 8.

[149] W.-C. Lo, C.-Y. Huang, and C.-H. Hsu, "Edge-assisted rendering of 360° videos streamed to head-mounted virtual reality," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, 2018, pp. 44–51.

[150] K. Poularakis, G. Iosifidis, A. Argyriou, I. Koutsopoulos, and L. Tassiulas, "Caching and operator cooperation policies for layered video content delivery," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (IEEE INFOCOM)*, 2016, pp. 1–9.

[151] J. Poderys, M. Artuso, C. M. O. Lensbøl, H. L. Christiansen, and J. Soler, "Caching at the mobile edge: A practical implementation," *IEEE Access*, vol. 6, pp. 8630–8637, 2018.

[152] G. Li *et al.*, "Data-driven approaches to edge caching," in *Proc. ACM Workshop Netw. Emerg. Appl. Technol.*, 2018, pp. 8–14.

[153] A. Mahzari, A. T. Nasrabadi, A. Samiei, and R. Prakash, "FoV-aware edge caching for adaptive 360 video streaming," in *Proc. ACM Multimedia Conf.*, 2018, pp. 173–181.

[154] C. Wu, Z. Tan, Z. Wang, and S. Yang, "A dataset for exploring user behaviors in VR spherical video streaming," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 193–198.

[155] G. Papaioannou and I. Koutsopoulos, "Tile-based caching optimization for 360° videos," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2019, pp. 171–180.

[156] K. Liu, Y. Liu, J. Liu, A. Argyriou, and Y. Ding, "Joint EPC and RAN caching of tiled VR videos for mobile networks," in *Proc. Int. Conf. Multimedia Model.*, 2019, pp. 92–105.

[157] P. Maniotis, E. Bourtsoulatze, and N. Thomos, "Tile-based joint caching and delivery of 360° videos in heterogeneous networks," in *Proc. IEEE 21st Int. Workshop Multimedia Signal Process. (MMSP)*, 2019, pp. 1–6.

[158] M. Chen, W. Saad, and C. Yin, "Echo-liquid state deep learning for 360° content transmission and caching in wireless VR networks with cellular-connected UAVs," *IEEE Trans. Commun.*, early access.

[159] X. Yang, Z. Chen, K. Li, Y. Sun, and H. Zheng, "Optimal task scheduling in communication-constrained mobile edge computing systems for wireless virtual reality," in *Proc. IEEE 23rd Asia–Pac. Conf. Commun. (APCC)*, 2017, pp. 1–6.

[160] J. Chakareski, "VR/AR immersive communication: Caching, edge computing, and transmission trade-offs," in *Proc. ACM Workshop Virtual Real. Augmented Real. Netw.*, 2017, pp. 36–41.

[161] Y. Sun, Z. Chen, M. Tao, and H. Liu, "Communications, caching, and computing for mobile virtual reality: Modeling and tradeoff," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7573–7586, Nov. 2019.

[162] G. Rigazzi, J. Kainulainen, C. Turyagyenda, A. Mourad, and J. Ahn, "An edge and fog computing platform for effective deployment of 360 video applications," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshop (WCNCW)*, Apr. 2019, pp. 1–6.

[163] M. S. Elbamby, C. Perfecto, M. Bennis, and K. Doppler, "Edge computing meets millimeter-wave enabled VR: Paving the way to cutting the cord," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2018, pp. 1–6.

[164] D. Antonogiorgakis *et al.*, "A view on edge caching applications," Jul. 2019. [Online]. Available: arXiv:1907.12359.

[165] S.-Y. Lien, S.-C. Hung, H. Hsu, and D.-J. Deng, "Energy-optimal edge content cache and dissemination: Designs for practical network deployment," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 88–93, May 2018.

[166] M. Yan *et al.*, "Assessing the energy consumption of 5G wireless edge caching," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2019, pp. 1–6.

[167] Q.-V. Pham *et al.*, "A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art," Jun. 2019. [Online]. Available: arXiv:1906.08452.

[168] Z. Tao *et al.*, "A survey of virtual machine management in edge computing," *Proc. IEEE*, vol. 107, no. 8, pp. 1482–1499, 2019.

[169] H. Ahmadi, O. Eltobgy, and M. Hefeeda, "Adaptive multicast streaming of virtual reality content to mobile users," in *Proc. ACM Thematic Workshops Multimedia*, 2017, pp. 170–178.

[170] Y. Bao, T. Zhang, A. Pande, H. Wu, and X. Liu, "Motion-prediction-based multicast for 360-degree video transmissions," in *Proc. 14th Annu. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, 2017, pp. 1–9.

[171] C. Guo, Y. Cui, and Z. Liu, "Optimal multicast of tiled 360 VR video," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 145–148, Feb. 2018.

[172] K. Long, C. Ye, Y. Cui, and Z. Liu, "Optimal multi-quality multicast for 360 virtual reality video," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2018, pp. 1–6.

[173] N. Kan, C. Liu, J. Zou, C. Li, and H. Xiong, "A server-side optimized hybrid multicast–unicast strategy for multi-user adaptive 360-degree video streaming," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 141–145.

[174] M. Huang and X. Zhang, "MAC scheduling for multiuser wireless virtual reality in 5G MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2018, pp. 1–6.

[175] K. Mania, B. D. Adelstein, S. R. Ellis, and M. I. Hill, "Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity," in *Proc. 1st Symp. Appl. Perception Graph. Visual.*, 2004, pp. 39–47.

[176] Y. Li and W. Gao, "MUVR: Supporting multi-user mobile virtual reality with resource constrained edge cloud," in *Proc. IEEE/ACM Symp. Edge Comput. (SEC)*, 2018, pp. 1–16.

[177] S. Petrangeli, G. Simon, and V. Swaminathan, "Trajectory-based viewport prediction for 360-degree virtual reality videos," in *Proc. IEEE Int. Conf. Artif. Intell. Virtual Real. (AIVR)*, 2018, pp. 157–160.

[178] Y. Hu, Y. Liu, and Y. Wang, "VAS360: QoE-driven viewport adaptive streaming for 360 video," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, 2019, pp. 324–329.

[179] A. Mavlankar and B. Girod, *Video Streaming With Interactive Pan/Tilt/Zoom*. Berlin, Germany: Springer, 2010, pp. 431–455. [Online]. Available: https://doi.org/10.1007/978-3-642-12802-8_19

[180] Y. Ban, L. Xie, Z. Xu, X. Zhang, Z. Guo, and Y. Wang, "Cub360: Exploiting cross-users behaviors for viewport prediction in 360 video adaptive streaming," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2018, pp. 1–6.

[181] X. Liu, Q. Xiao, V. Gopalakrishnan, B. Han, F. Qian, and M. Varvello, "360° innovations for panoramic video streaming," in *Proc. 16th ACM Workshop Hot Topics Netw.*, 2017, pp. 50–56.

[182] Y. Bao, H. Wu, T. Zhang, A. A. Ramli, and X. Liu, "Shooting a moving target: Motion-prediction-based transmission for 360-degree videos," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2016, pp. 1161–1170.

[183] M. Jamali, S. Coulombe, A. Vakili, and C. Vazquez, "LSTM-based viewpoint prediction for multi-quality tiled video coding in virtual reality streaming," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2020, pp. 110–117.

[184] F. Duanmu, E. Kurdoglu, S. A. Hosseini, Y. Liu, and Y. Wang, "Prioritized buffer control in two-tier 360 video streaming," in *Proc. ACM Workshop Virtual Real. Augmented Real. Netw.*, 2017, pp. 13–18.

[185] J. Yu and Y. Liu, "Field-of-view prediction in 360-degree videos with attention-based neural encoder-decoder networks," in *Proc. 11th ACM Workshop Immersive Mixed Virtual Environ. Syst.*, 2019, pp. 37–42.

[186] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, "An HTTP/2-based adaptive streaming framework for 360 virtual reality videos," in *Proc. ACM Multimedia Conf.*, 2017, pp. 306–314.

[187] Y. La Sanchez, G. S. Bhullar, R. Skupin, C. Hellge, and T. Schierl, "Delay impact on MPEG OMAF's tile-based viewport-dependent 360° video streaming," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, early access.

[188] S. Rossi, F. De Simone, P. Frossard, and L. Toni, "Spherical clustering of users navigating 360° content," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 4020–4024.

[189] C. Bron and J. Kerbosch, "Algorithm 457: Finding all cliques of an undirected graph," *Commun. ACM*, vol. 16, no. 9, pp. 575–577, 1973.

[190] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[191] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 199–204.

[192] X. Hou, S. Dey, J. Zhang, and M. Budagavi, "Predictive view generation to enable mobile 360-degree and VR experiences," in *Proc. ACM Morning Workshop Virtual Real. Augmented Real. Netw.*, 2018, pp. 20–26.

[193] X. Hou, J. Zhang, M. Budagavi, and S. Dey, "Head and body motion prediction to enable mobile VR experiences with low latency," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2019, pp. 1–7.

[194] J. Heyse, M. T. Vega, F. de Backere, and F. de Turck, "Contextual bandit learning-based viewport prediction for 360 video," in *Proc. IEEE Conf. Virtual Real. 3D User Interfaces (VR)*, Mar. 2019, pp. 972–973.

[195] F. Duanmu, Y. Mao, S. Liu, S. Srinivasan, and Y. Wang, "A subjective study of viewer navigation behaviors when watching 360-degree videos on computers," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2018, pp. 1–6.

[196] M. Xu, Y. Song, J. Wang, M. Qiao, L. Huo, and Z. Wang, "Predicting head movement in panoramic video: A deep reinforcement learning approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2693–2708, Nov. 2019.

[197] Y. Xu *et al.*, "Gaze prediction in dynamic 360° immersive videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5333–5342.

[198] C. Li, W. Zhang, Y. Liu, and Y. Wang, "Very long term field of view prediction for 360-degree video streaming," in *Proc. IEEE Conf. Multimedia Inf. Process. Retrieval (MIPR)*, 2019, pp. 297–302.

[199] Q. Yang, J. Zou, K. Tang, C. Li, and H. Xiong, "Single and sequential viewports prediction for 360-degree video streaming," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2019, pp. 1–5.

[200] C.-L. Fan, J. Lee, W.-C. Lo, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "Fixation prediction for 360 video streaming in head-mounted virtual reality," in *Proc. ACM 27th Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2017, pp. 67–72.

[201] A. D. Aladagli, E. Ekmekcioglu, D. Jarnikov, and A. Kondoz, "Predicting head trajectories in 360 virtual reality videos," in *Proc. IEEE Int. Conf. 3D Immersion (IC3D)*, 2017, pp. 1–6.

[202] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction," in *Proc. ACM Multimedia Conf.*, 2018, pp. 1190–1198.

[203] M. F. R. Rondon, L. Sassatelli, R. A. Pardo, and F. Precioso, "Revisiting deep architectures for head motion prediction in 360° videos," Nov. 2019. [Online]. Available: arXiv:1911.11702.

[204] A. T. Nasrabadi *et al.*, "A taxonomy and dataset for 360° videos," in *Proc. 10th ACM Multimedia Syst. Conf.*, 2019, pp. 273–278.

[205] Z. Yuan, G. Ghinea, and G.-M. Muntean, "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 104–117, Jan. 2015.

[206] B. Ciubotaru, G.-M. Muntean, and G. Ghinea, "Objective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 202–212, Jun. 2009.

[207] B. Ciubotaru, G. Ghinea, and G.-M. Muntean, "Subjective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Trans. Broadcast.*, vol. 60, no. 1, pp. 50–60, Mar. 2014.

[208] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *Proc. IEEE Picture Coding Symp. (PCS)*, 2016, pp. 1–5.

[209] "Subjective video quality assessment methods for multimedia applications," IETF, Fremont, CA, USA, ITU Recommendation P. 910, 2009.

[210] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang, and Z. Fei, "Subjective and objective quality assessment of panoramic videos in virtual reality environments," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2017, pp. 163–168.

[211] "Methodology for the subjective assessment of video quality in multimedia applications," IETF, Fremont, CA, USA, ITU Recommendation BT.1788, 2007.

[212] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan, "Assessing visual quality of omnidirectional videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3516–3530, Dec. 2019.

[213] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "A subjective study to evaluate video quality assessment algorithms," in *Proc. Human Vis. Electron. Imag. XV*, vol. 7527, 2010, Art. no. 75270H.

[214] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, "Towards subjective quality of experience assessment for omnidirectional video streaming," in *Proc. 9th Int. Conf. Qual. Multimedia Exp.(QoMEX)*, 2017, pp. 1–6.

[215] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, ITU, Geneva, Switzerland, 2014.

[216] A. Singla, S. Göring, A. Raake, B. Meixner, R. Koenen, and T. Buchholz, "Subjective quality evaluation of tile-based streaming for omnidirectional videos," in *Proc. 10th ACM Multimedia Syst. Conf.*, 2019, pp. 232–242.

[217] I. Shinobu, *Ishihara's Tests for Colour Deficiency: The Series of Plates Designed as a Test for Colour Deficiency*. Tokyo, Japan: Kanehara, 1999.

[218] H. B. Peters, "Vision screening with a Snellen chart," *Optometry Vis. Sci.*, vol. 38, no. 9, pp. 487–505, 1961.

[219] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on QoE of 360 video for VR communication," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, 2017, pp. 1–6.

[220] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays," in *Proc. IEEE 9th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2017, pp. 1–6.

[221] W. Zou, F. Yang, W. Zhang, Y. Li, and H. Yu, "A framework for assessing spatial presence of omnidirectional video on virtual reality device," *IEEE Access*, vol. 6, pp. 44676–44684, 2018.

[222] I. Hupont, J. Gracia, L. Sanagustin, and M. A. Gracia, "How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift," in *Proc. IEEE 7th Int. Workshop Qual. Multimedia Exp. (QoMEX)*, 2015, pp. 1–6.

[223] J. Brooke *et al.*, "SUS—A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.

[224] C. Izard, "Emotions, personality, and psychotherapy," in *The Psychology of Emotions*. New York, NY, USA: Springer, 1991.

[225] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[226] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, "An evaluation of heart rate and electrodermal activity as an objective QoE evaluation method for immersive virtual reality environments," in *Proc. IEEE 8th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2016, pp. 1–6.

[227] *Sensory Analysis—General Guidance for the Design of Test Rooms*, Int. Org. Stand., Geneva, Switzerland, 1988.

[228] S. Fremerey, F. Hofmeyer, S. Göring, and A. Raake, "Impact of various motion interpolation algorithms on 360° video QoE," in *Proc. 11th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, Jun. 2019, pp. 1–3.

[229] S. Fremerey, A. Singla, K. Meseberg, and A. Raake, "AVtrack360: An open dataset and software recording people's head rotations watching 360° videos on an HMD," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 403–408.

[230] G. Regal, R. Schatz, J. Schrammel, and S. Suette, "VRate: A Unity3D asset for integrating subjective assessment questionnaires in virtual environments," in *Proc. IEEE 10th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2018, pp. 1–3.

[231] P. Pérez and J. Escobar, "MIRO360: A tool for subjective assessment of 360 degree video for ITU-T P. 360-VR," in *Proc. 11th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, Jun. 2019, pp. 1–3.

[232] "Methodology for the subjective assessment of the quality of television pictures BT series broadcasting service," Int. Telecommun. Union, Geneva, Switzerland, Rep. BT.500-13, 2012.

[233] A. S. Fernandes and S. K. Feiner, "Combating VR sickness through subtle dynamic field-of-view modification," in *Proc. IEEE Symp. 3D User Interfaces (3DUI)*, 2016, pp. 201–210.

[234] D. P. Salgado *et al.*, "A QoE assessment method based on EDA, heart rate and EEG of a virtual reality assistive technology system," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 517–520.

[235] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. IEEE Int. Symp. Mixed Augmented Real.*, Sep. 2015, pp. 31–36.

[236] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1408–1412, Sep. 2017.

[237] S. Chen, Y. Zhang, Y. Li, Z. Chen, and Z. Wang, "Spherical structural similarity index for objective omnidirectional video quality assessment," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2018, pp. 1–6.

[238] Y. Zhou, M. Yu, H. Ma, H. Shao, and G. Jiang, "Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video," in *Proc. 14th IEEE Int. Conf. Signal Process. (ICSP)*, 2018, pp. 54–57.

[239] J. van der Hooft, M. T. Vega, S. Petrangeli, T. Wauters, and F. De Turck, "Quality assessment for adaptive virtual reality video streaming: A probabilistic approach on the user's gaze," in *Proc. 22nd Conf. Innov. Clouds Internet Netw. Workshops (ICIN)*, 2019, pp. 19–24.

[240] E. Upenik, M. Rerabek, and T. Ebrahimi, "On the performance of objective metrics for omnidirectional visual content," in *Proc. IEEE 9th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2017, pp. 1–6.

[241] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," in *Proc. Opt. Photon. Inf. Process. X*, vol. 9970, 2016, Art. no. 99700C.

[242] H. T. Tran, N. P. Ngoc, C. M. Bui, M. H. Pham, and T. C. Thang, "An evaluation of quality metrics for 360 videos," in *Proc. IEEE 9th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, 2017, pp. 7–11.

[243] R. I. T. da Costa Filho *et al.*, "Predicting the performance of virtual reality video streaming in mobile networks," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 270–283.

[244] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the gap between VQA and human behavior on omnidirectional video: A large-scale dataset and a deep learning model," Jul. 2018. [Online]. Available: arXiv:1807.10990.

[245] S. Yang *et al.*, "An objective assessment method based on multi-level factors for panoramic videos," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[246] X. Jin *et al.*, "Adaptive image quality assessment method based on structural similarity," *J. Optoelectron. Laser*, vol. 25, no. 2, pp. 378–385, 2014.

[247] F. Lopes, J. Ascenso, A. Rodrigues, and M. P. Queluz, "Subjective and objective quality assessment of omnidirectional video," in *Proc. Appl. Digit. Image Process. XLI*, vol. 10752, 2018, Art. no. 107520P.

[248] S. Yao, C. Fan, and C. Hsu, "Towards quality-of-experience models for watching 360° videos in head-mounted virtual reality," in *Proc. 11th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, Jun. 2019, pp. 1–3.

[249] J. Orlosky, K. Kiyokawa, and H. Takemura, "Virtual and augmented reality on the 5G highway," *J. Inf. Process.*, vol. 25, pp. 133–141, Feb. 2017.

[250] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[251] M. Fiedler, H.-J. Zepernick, and V. Kelkkanen, "Network-induced temporal disturbances in virtual reality applications," in *Proc. IEEE 11th Int. Conf. Qual. Multimedia Exp. (QoMEX)*, 2019, pp. 1–3.

[252] J. Van Der Hooft, S. Petrangeli, T. Wauters, R. Huysegems, T. Bostoen, and F. De Turck, "An HTTP/2 push-based approach for low-latency live streaming with super-short segments," *J. Netw. Syst. Manag.*, vol. 26, no. 1, pp. 51–78, 2018.

[253] J. van der Hooft, D. Pauwels, C. De Boom, S. Petrangeli, T. Wauters, and F. De Turck, "Low-latency delivery of news-based video content," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 537–540.

[254] D. V. Nguyen, H. T. Tran, and T. C. Thang, "Impact of delays on 360-degree video communications," in *Proc. IEEE TRON Symp. (TRONSHOW)*, 2017, pp. 1–6.

[255] S. Wei and V. Swaminathan, "Low latency live video streaming over HTTP 2.0," in *Proc. ACM Netw. Oper. Syst. Support Digit. Audio Video Workshop*, 2014, p. 37.

[256] Z. Xu *et al.*, "Tile-based QoE-driven HTTP/2 streaming system for 360 video," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2018, pp. 1–4.

[257] M. B. Yahia, Y. Le Louedec, G. Simon, and L. Nuaymi, "HTTP/2-based streaming solutions for tiled omnidirectional videos," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2018, pp. 89–96.

[258] S.-C. Yen, C.-L. Fan, and C.-H. Hsu, "Streaming 360° videos to head-mounted virtual reality using DASH over QUIC transport protocol," in *Proc. 24th ACM Workshop Packet Video*, 2019, pp. 7–12.

[259] L. Liu *et al.*, "Cutting the cord: Designing a high-quality untethered VR system with low latency remote rendering," in *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2018, pp. 68–80.

[260] M. Viitanen, J. Vanne, T. D. Hämäläinen, and A. Kulmala, "Low latency edge rendering scheme for interactive 360 degree virtual reality gaming," in *Proc. IEEE 38th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, 2018, pp. 1557–1560.

[261] J. Park, P. Popovski, and O. Simeone, "Minimizing latency to support VR social interactions over wireless cellular systems via bandwidth allocation," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 776–779, Oct. 2018.

[262] C. Perfecto, M. S. ElBamby, J. Del Ser, and M. Bennis, "Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework," 2018. [Online]. Available: arxiv.abs/1811.07388.

[263] J. Yi, S. Luo, and Z. Yan, "A measurement study of YouTube 360° live video streaming," in *Proc. 29th ACM Workshop Netw. Oper. Syst. Support Digit. Audio Video*, 2019, pp. 49–54.

[264] X. Hu, W. Quan, T. Guo, Y. Liu, and L. Zhang, "Mobile edge assisted live streaming system for omnidirectional video," *Mobile Inf. Syst.*, vol. 2019, May 2019, Art. no. 8487372.

[265] C. Griwodz *et al.*, "Efficient live and on-demand tiled HEVC 360 VR video streaming," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, 2018, pp. 81–88.

[266] X. Liu, B. Han, F. Qian, and M. Varvello, "LIME: Understanding commercial 360° live video streaming services," in *Proc. 10th ACM Multimedia Syst. Conf.*, 2019, pp. 154–164.

[267] G. Baig *et al.*, "Jigsaw: Robust live 4k video streaming," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1–16.

[268] T. T. Le, J. Jeong, and E.-S. Ryu, "Efficient transcoding and encryption for live 360 CCTV system," *Appl. Sci.*, vol. 9, no. 4, p. 760, 2019.

[269] B. Chen, Z. Yan, H. Jin, and K. Nahrstedt, "Event-driven stitching for tile-based live 360 video streaming," in *Proc. 10th ACM Multimedia Syst. Conf.*, 2019, pp. 1–12.

[270] R. Aksu, J. Chakareski, and V. Swaminathan, "Viewport-driven rate-distortion optimized scalable live 360° video network multicast," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, 2018, pp. 1–6.

[271] C. Timmerer, "Immersive media delivery: Overview of ongoing standardization activities," *IEEE Commun. Stand. Mag.*, vol. 1, no. 4, pp. 71–74, Dec. 2017.

[272] *Generic Coding of Moving Pictures and Associated Audio Information, Part 1: Systems*, ISO/IEC Standard 13818-1:2018, 2018.

[273] D. Singer, W. Belknap, and G. Franceschini, *Information Technology Coding of Audio-Visual Objects Part 14: MP4 File Format*, ISO/IEC Standard 14496-14, Feb. 2004.

[274] Y. Lim, K. Park, J. Y. Lee, S. Aoki, and G. Fernando, "MMT: An emerging MPEG standard for multimedia delivery over the Internet," *IEEE MultiMedia*, vol. 20, no. 1, pp. 80–85, Jan.–Mar. 2013.

[275] B. Choi, Y. Wang, M. Hannuksela, Y. Lim, and A. Murtaza, "Study of ISO/IEC DIS 23000-20 omnidirectional media format," ISO/IEC JTCl/SC29/WG11, Torino, Italy, Rep. N16950, 2017.

[276] S. Oh and S. Hwang, "OMAF: Generalized signaling of region-wise packing for omnidirectional video," in *Proc. 118th MPEG Meeting ISO/IEC JTC1/SC29/WG11 MPEG2017/m40423*, 2017, pp. 1–11.

[277] R. Skupin, Y. Sanchez, Y.-K. Wang, M. Hannuksela, J. Boyce, and M. Wien, "Standardization status of 360 degree video coding and delivery," in *Proc. Vis. Commun. Image Process. (VCIP)*, 2017, pp. 1–4.

[278] M. M. Hannuksela, Y. Wang, and A. Hourunranta, "An overview of the OMAF standard for 360° video," in *Proc. Data Compression Conf. (DCC)*, Mar. 2019, pp. 418–427.

[279] K. Hughes and D. Singer, *Information Technology—Multimedia Application Format (MPEG-A)—Part 19: Common Media Application Format (CMAF) for Segmented Media*, ISO/IEC Standard 23000-19, 2017.

[280] D. Singer, *Information Technology—Coding of Audio-Visual Objects—Part 12: ISO Base Media File Format*, ISO/IEC Standard 14496-12, 2005.

[281] J. Jung, B. Kroon, R. Doré, G. Lafruit, and J. Boyce, "Update on N17618 v2 CTC on 3DoF+ and Windowed 6DoF," Moving Picture Experts Group, Belém, Brazil, Rep. ISO/IEC JTC1/SC29/WG11, 2018.
[282] R. Doré, J. Fleureau, B. Chupeau, and G. Briand, "3DoF plus intermediate view synthesizer proposal," Moving Picture Experts Group, Belém, Brazil, Rep. ISO/IEC JTC1/WG11, 2018, p. 137.
[283] "Virtual reality (VR) media services over 3GPP," 3GPP, Sophia Antipolis, France, Rep. TR 126 918, 2018.
[284] M. S. Brennesholtz, "VR standards and guidelines," in *SID Symp. Dig. Tech. Papers*, vol. 49, 2018, pp. 1–4.
[285] Y. Trivedi, "Standards for the virtual world!" *IEEE Commun. Stand. Mag.*, vol. 1, no. 2, pp. 8–12, Jun. 2017.
[286] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, Oct.–Dec. 2016.
[287] T. Richter, "On the standardization of the JPEG XT image compression," in *Proc. IEEE Picture Coding Symp. (PCS)*, 2013, pp. 37–40.
[288] J. Boyce, E. Alshina, A. Abbas, and Y. Ye, *JVET Common Test Conditions and Evaluation Procedures for 360 Video*, document N16701, ITU, Geneva, Switzerland, 2017.
[289] QUALINET. *European Network on Quality of Experience in Multimedia Systems and Services*. Accessed: Apr. 10, 2019. [Online]. Available: http://www.qualinet.eu/
[290] Video Quality Experts Group (VQEG). *Immersive Media Group*. Accessed: Apr. 10, 2019. [Online]. Available: https://www.its.bldrdoc.gov/vqeg/projects/immersive-media-group.aspx
[291] E. Domínguez, "Going beyond the classic news narrative convention: The background to and challenges of immersion in journalism," *Front. Digit. Humanities*, vol. 4, p. 10, May 2017.
[292] N. De la Peña *et al.*, "Immersive journalism: Immersive virtual reality for the first-person experience of news," *Presence Teleoper. Virtual Environ.*, vol. 19, no. 4, pp. 291–301, 2010.
[293] A. S. Alqahtani, L. F. Daghestani, and L. F. Ibrahim, "Environments and system types of virtual reality technology in STEM: A survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 77–89, 2017.
[294] D. A. Guttentag, "Virtual reality: Applications and implications for tourism," *Tourism Manag.*, vol. 31, no. 5, pp. 637–651, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0261517709001332
[295] A. Hebbel-Seeger, "360 degrees video and VR for training and marketing within sports," *Athens J. Sports*, vol. 4, no. 4, pp. 243–261, 2017.
[296] S. G. Izard *et al.*, "Virtual reality as an educational and training tool for medicine," *J. Med. Syst.*, vol. 42, no. 3, p. 50, 2018.
[297] C. Cox, "The use of computer graphics and virtual reality for visual impact assessments," Ph.D. dissertation, School Chem. Environ. Mining Eng., Univ. Nottingham, Nottingham, U.K., 2003.
[298] J. R. Zeballos, "The potential of 360-degree videos for teaching, learning and research," in *Proc. 12th Annu. Int. Technol. Educ. Develop. Conf. (INTED)*, 2018, pp. 54–61.
[299] R. Pea, M. Mills, J. Rosen, K. Dauber, W. Effelsberg, and E. Hoffert, "The diver project: Interactive digital video repurposing," *IEEE MultiMedia*, vol. 11, no. 1, pp. 54–61, Jan.–Mar. 2004.
[300] S. Kavanagh, A. Luxton-Reilly, B. Wüensche, and B. Plimmer, "Creating 360 educational video: A case study," in *Proc. ACM 28th Aust. Conf. Comput. Human Interact.*, 2016, pp. 34–39.
[301] M. S. Feurstein, "Towards an integration of 360-degree video in higher education. Workflow, challenges and scenarios," in *Proc. DeLFI Workshops 16th e-Learn. Conf. German Comput. Soc.*, 2018, pp. 1–12.
[302] E. J. David, J. Gutiérrez, A. Coutrot, M. P. Da Silva, and P. L. Callet, "A dataset of head and eye movements for 360 videos," in *Proc. 9th ACM Multimedia Syst. Conf.*, 2018, pp. 432–437.
[303] W.-C. Lo, C.-L. Fan, J. Lee, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "360 video viewing dataset in head-mounted virtual reality," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 211–216.
[304] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "A public database of immersive VR videos with corresponding ratings of arousal, valence, and correlations between head movements and self report measures," *Front. Psychol.*, vol. 8, p. 2116, Dec. 2017.
[305] T. Xu, B. Han, and F. Qian, "Analyzing viewport prediction under different VR interactions," in *Proc. 15th Int. Conf. Emerg. Netw. Exp. Technol.*, 2019, pp. 165–171.
[306] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5G technologies: Potential solutions, recent advancements and future directions," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 196–248, 1st Quart., 2020.
[307] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for Internet video," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 339–350, 2013.
[308] X. Feng, V. Swaminathan, and S. Wei, "Viewport prediction for live 360-degree mobile video streaming using user-content hybrid motion tracking," *ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–22, 2019.

Abid Yaqoob (Graduate Student Member, IEEE) received the B.Sc. degree in computer systems engineering from the Islamia University of Bahawalpur, Pakistan, in 2014, and the M.Sc. degree in network and information security from Northwestern Polytechnical University, Xi'an, China, in 2018. He is currently pursuing the Ph.D. degree with the Performance Engineering Laboratory and the Insight Centre for Data Analytics, School of Electronic Engineering, Dublin City University, Ireland. His research interests include mobile wireless communication, priority-aware multiview video streaming, and immersive multimedia processing and delivery solutions.

Ting Bi (Member, IEEE) received the B.Eng. degree in software engineering from Wuhan University, China, in 2010, and the M.Eng. and Ph.D. degrees in telecommunications from Dublin City University, Ireland, in 2011 and 2017, respectively, where he is a Postdoctoral Researcher with the Performance Engineering Laboratory and the Insight Centre for Data Analytics, School of Electronic Engineering. He has published in prestigious international conferences and journals. His research interests include mobile and wireless communications, multimedia and multisensory media streaming over wireless access networks, user quality of experience, handover and network selection strategies, and energy saving for mobile devices. He is a member of ACM, RDA, and IEEE Communications and IEEE Broadcast Technology Societies.

Gabriel-Miro Muntean (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees in software engineering from the "Politehnica" University of Timisoara, Romania, in 1996 and 1997, respectively, and the Ph.D. degree for research on adaptive multimedia delivery from Dublin City University, Ireland, in 2004, where he is an Associate Professor with the School of Electronic Engineering and the Co-Director of the Performance Engineering Laboratory. He has published over 350 papers in top-level international journals and conferences, authored four books and 19 book chapters, and edited six additional books. His research interests include quality, performance, and energy saving issues related to multimedia and multiple sensorial media delivery, technology-enhanced learning, and other data communications over heterogeneous networks. He is an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING, the Multimedia Communications Area Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and chair and reviewer for important international journals, conferences, and funding agencies. He is senior member of IEEE Broadcast Technology Society.