Review

360-Degree Video Bandwidth Reduction: Technique and Approaches Comprehensive Review

En Sing Wong [1], Nur Haliza Abdul Wahab [1,*], Faisal Saeed [2] and Nouf Alharbi [3]

1 School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor 81310, Malaysia; ensing@graduate.utm.my
2 DAAI Research Group, Department of Computing and Data Science, School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK; faisal.saeed@bcu.ac.uk
3 College of Computer Science and Engineering, Taibah University, Madina 42353, Saudi Arabia; nmoharbi@taibahu.edu.sa
* Correspondence: nur.haliza@utm.my

Abstract: Recently, the usage of 360-degree videos has prevailed in various sectors such as education, real estate, medical, entertainment and more. The development of the Virtual World "Metaverse" demanded a Virtual Reality (VR) environment with high immersion and a smooth user experience. However, various challenges are faced to provide real-time streaming due to the nature of high-resolution 360-degree videos such as high bandwidth requirement, high computing power and low delay tolerance. To overcome these challenges, streaming methods such as Dynamic Adaptive Streaming over HTTP (DASH), Tiling, Viewport-Adaptive and Machine Learning (ML) are discussed. Moreover, the superiorities of the development of 5G and 6G networks, Mobile Edge Computing (MEC) and Caching and the Information-Centric Network (ICN) approaches to optimize the 360-degree video streaming are elaborated. All of these methods strike to improve the Quality of Experience (QoE) and Quality of Service (QoS) of VR services. Next, the challenges faced in QoE modeling and the existing objective and subjective QoE assessment methods of 360-degree video are presented. Lastly, potential future research that utilizes and further improves the existing methods substantially is discussed. With the efforts of various research studies and industries and the gradual development of the network in recent years, a deep fake virtual world, "Metaverse" with high immersion and conducive for daily life working, learning and socializing are around the corner.

Keywords: Virtual Reality (VR); 360-degree video; bandwidth reduction; metaverse; DASH; tiling; viewport-adaptive; machine learning; network optimization; Quality of Experience (QoE)

1. Introduction

A 360-degree video is a video filmed in all directions by an omnidirectional camera or numerous cameras simultaneously, encompassing a whole 360-degree 3D sphere view, hence creating a Virtual Reality (VR) environment. When played back on a 2D flat screen (mobile or computer), viewers may alter the viewing direction and view the film from whichever angle they like, similar to a panorama. It can also be played on a display like a head-mounted display or projectors organized in the shape of a sphere or a portion of a sphere. The potential of 360-degree video and VR is enormous. The development of VR, AR and 360-degree video could be seen in education, real estate, medical, economics and more. Due to the COVID-19 pandemic situation where close contact with physical connections is forbidden, the work from home and study from home culture has further encouraged the development of these technologies. For instance, hologram conference meetings [1], home shopping with AR [2], etc. Based on the study in [3], the growth of the 360-degree video is demonstrated in Figure 1.
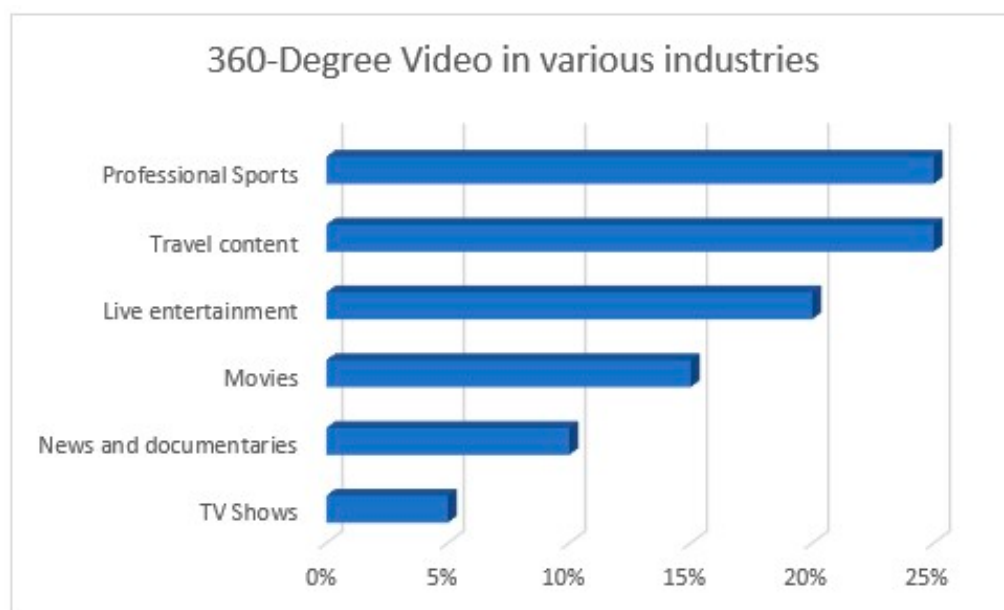
Figure 1. Growth of the 360-degree video in various industries.

The exponential expansion in online material, more inexpensive and cost-effective technologies and the remarkable advances offered to mobile technology have sparked attentiveness in the employment of immersive technologies and 360° video, especially in sectors such as education [4,5]. The 360-degree videos are a step forward from traditional two-dimensional videos in that they provide the user with the sensation of being immersed in the middle of a scene that can be examined freely and interactively by just rotating the head to a specific Region of Interest (RoI).

360-degree video streaming is deemed to be the essential instrument to develop technical creativity and implement research and project activities among students. In addition, the 360-degree video may improve the understanding of more complicated subjects, ideas, or concepts by providing a more active rather than the traditional passive education atmosphere that is conducive to embedding learning and elevating education to new heights. Thus, it helps in increasing student motivation on exploring more on the subject that they are learning limitlessly.

The development of these technologies is further driven by the big corporations, Facebook officially announced that it would be renaming its company "Meta" in October 2021 [6]. Mark Zuckerberg, the Chief Executive Officer (CEO) of Meta has propagated the concept of bringing social connection, entertainment, gaming, health, work, education, and commerce, all of which are daily activities, into the virtual universe known as the "Metaverse". Along with the other technologies Blockchain, Non-Fungible Token (NFT) and Cryptocurrency, people could even run their businesses and own assets such as houses inside the Metaverse. The business potential and opportunities are impossible to overlook. For instance, JP Morgan, the central bank and global financial institution in the United States (US) sees a $1 Trillion market opportunity in Metaverse and launched a virtual lounge on Blockchain-based VR platform "Decentraland" on 15 February 2022 [7].

Afterwards, Second Life (SL), which is known as a 3-D virtual environment offers a user something different, where users could pose to be anybody or anything they desire, and this SL receives massive attention. They can go to social events such as a concert, press conferences and classes, buy land, clothes and gadgets or visit their friends [8]. Moreover, in February 2022, one of the Asia entertainment companies SM Culture Universe (SMCU) forged a metaverse collaboration with a company called "The Sandbox", which is the utmost metaverse platform in the world. The goal of this collaboration is to host events in the metaverse, such as concerts and fan gatherings, as well as to release various games and

NFT products [9]. Furthermore, completing tasks in the Metaverse may occur considerably faster than we expect.

Another statement about bringing VR/AR to step as top of technologies; Bill Gates, the co-founder of Microsoft and a millionaire, stated in December 2021 that Metaverse work sessions will take place in 2 to 3 years [10]. It is possible to see more and more influential companies from different sectors expand their industry into the Metaverse in the coming years.

As for this, the superiorities of 360-degree video can be concluded as: (a) Boost interest and creativity in education; (b) Generate various business and job opportunities in Metaverse; (c) Providing a virtual communication platform highly similar to face-to-face interaction; (d) Enabling a supreme experience in entertainment: games, concerts, etc.

Although lots of benefits can be listed on 360-degree video, there are a few problems such as lack of tools and network barriers. Due to the extremely high bandwidth demands, providing a great Quality of Experience (QoE) to viewers while streaming 360 videos over the Internet is particularly difficult. Both academics and businesses are currently looking for more effective ways to bridge the gap between the user experience of VR apps and the VR networking issues such as high bandwidth requirements. Currently, many VR devices can only deliver a primitive and limited user experience with a generally unsatisfying result.

In this paper, we try to tackle several issues that occur in 360-degree video, not only in the foundation understanding of what 360-degree video is, until the solutions that can be utilized to accomplish the evolvement of this technology. Thus, the main contributions of this paper are:

(a)   Discuss challenges faced by 360-degree video streaming;
(b)   Discuss existing approaches and techniques available for bandwidth reduction for 360-degree video;
(c)   Discuss existing network approaches in optimizing the streaming of 360-degree video;
(d)   Discuss existing Quality of Experience (QoE) measurement metrics of 360-degree video;
(e)   Discuss the future works to improve existing approaches.

2. Challenges Faced by 360-Degree Video Streaming

One of the biggest obstructions in realizing the real-time interaction and communication inside the VR is the high bandwidth requirement. The immersive experience will be broken if a delay occurs or the resolution drops. With the same resolution, 360-degree video requires 4 to 6 times the bandwidth of a regular video [11]. In order to ensure a good immersion experience, the displayed viewport usually has a high pixel resolution, typically 4K or in size of (3840 × 2160) and the resolution of the 360-degree movie has to be not less than 12K or in size of (11,520 × 6480). [12]. Moreover, an immersive 360-degree video can require a video frame rate equivalent to the Head-Mounted Device (HMD) refresh rate, normally in a range of a hundred (100) frames per second (fps), with the very top frame rate (up to 100 fps) and very huge in resolution (which can reach to maximum 12K). With respect to immersion, 360-degree videos can require a video frame rate equivalent to the HMD refresh rate. For example, a High-Efficiency Video Coding (HEVC) encoded 8K video at 60 fps has a bitrate of roughly 100 Mbps [12]. Table 1 summarizes the network requirement of VR, the bandwidth required for each VR type and the estimated network latency.

Table 1. Network requirement of VR.

| VR	Resolution	Equivalent TV Res.	Bandwidth	Latency (ms)
Early stage VR	1K × 1K@visual field 2D_30fps_8bit_4K	240P	25 Mbps	40
Entry level VR	2K × 2K@visual field 2D_30fps_8bit_8K	SD	100 Mbps	30
Advanced VR	4K × 4K@visual field 2D_60fps_10bit_12K	HD	400 Mbps	20
Extreme VR	8K × 8K@visual field 3D_120fps_12bit_24K	4K	2.35 Gbps	10

Next, intensive computing power is needed by VR devices, especially in VR information processing. Multiple activities, including scene depth estimation, picture semantic interpretation, three-dimensional reconstruction, and higher realism rendering, for example, need a large amount of computational power and have to be completed in real-time to give a natural and seamless user engagement [13]. The processing latency of VR is affected by the computational power of the computing equipment and the computational capacity of the jobs [14].

In addition, in order to achieve a high level of immersion and interactivity, high accuracy and high definition (HD) are also owing to the restricted space in the 360-degree film delay tolerance of human eyes. Experiment studies [15,16] clearly show that users can perceive delays greater than 10 ms as annoying, although users can tolerate the higher latencies [17]. Moreover, the human eye sees precise and smooth motion with a motion-to-photon (MTP) delay of fewer than 20 milliseconds (ms) [18–20]. The vestibular-ocular reflex (VOR) gets contradictory signals when MTP levels are high, which can cause dizziness and motion sickness. Thus, real-time communication with limited delay tolerance demanded accurate, low-latency communication services.

The problems can be concluded as: (a) High bandwidth requirement; (b) Intensive computing power requirement; (c) Stringent latency tolerance.

3. Available Techniques to Reduce the Bandwidth of the 360-Degree Video

Four categories of solutions proposed by various research are Dynamic adaptive HTTP streaming (DASH), tiling, viewport-adaptive and Machine learning (ML) as illustrated in Figure 2.
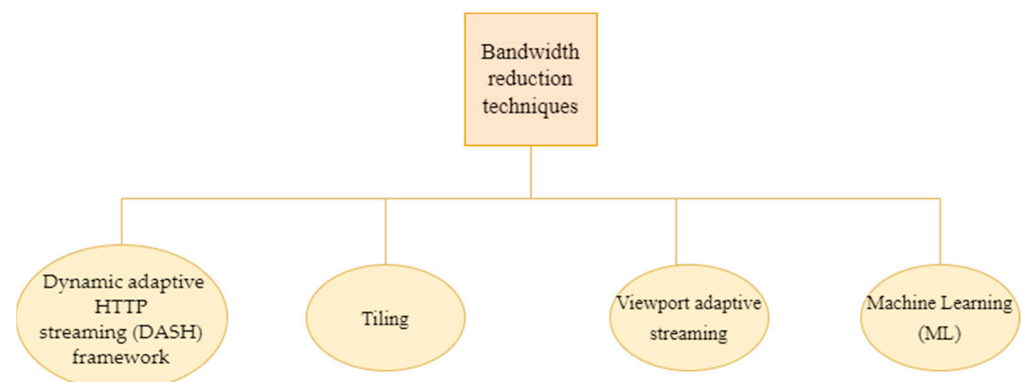
Figure 2. Bandwidth reduction techniques.

3.1. Dynamic Adaptive HTTP Streaming (DASH) Framework

Dynamic adaptive HTTP streaming (DASH) is an MPEG standard that provides a multimedia style and specification for sending material over HTTP using an adjustable

bitrate method [21]. DASH is extremely compatible with the existing internet infrastructure due to its minimal processing burden and transparency to middleboxes, and the ability to apply alternative adaption methods makes it adaptable to diverse network conditions standard is generally extensively utilized for two-dimensional video streaming over the world wide web recently. DASH streaming works by splitting videos into short segments, each segment on the DASH server maintains a number of video streams with varying bitrates [19]. By requesting the proper HTTP resource, based on the view on the streaming client, the main viewpoint segment stream with higher resolution and the other viewpoint segment stream with lower resolution. A video player can switch from one quality level to another in the middle of the video playback without interruption.

Table 2 demonstrates the major steps in the DASH streaming process:

Table 2. Major steps in the DASH streaming process.

Step	Process
Stitching	Stitch videos collected by many cameras/an omnidirectional camera onto diverse planar models such as cubic and affine transformation models match up the various camera images, merging and distorting the views to a sphere's surface [3]. For successful coding and transmission, the 360-degree sphere is projected to a 2D planar format such as Cubic Mapping Projection (CMP) and Equirectangular Projection (ERP).
Encoding and segmentation	The video file is segmented into smaller parts of a few seconds in length by the origin server. Each section is encoded in numerous bitrate or quality level variants.
Delivery	The encoded video segments are sent out to client devices over a content delivery network (CDN).
Decoding, rendering and play	Decodes the streamed data. With adaptive bitrate streaming, it plays the video and automatically adjusts the quality of the picture according to the network condition/user's views at the client device.

Another extension of DASH or other streaming systems is the Omnidirectional Media Format (OMAF) standard specifying the spatial information of video segments [22]. For the DASH OMAF scheme, storage space is sacrificed to increase the bandwidth of the VR video streaming [23]. Figure 3 shows the technical framework of the DASH OMAF architecture network. Furthermore, OMAF specifies several requirements for users, bringing the standard specification for omnidirectional streaming one step closer to completion. Players based on OMAF have already been implemented and demonstrated [24].

OMAF also defines tile-based streaming and Viewport-Based Streaming approaches where the Field of View (FoV) is downloaded at the highest quality possible, along with the lower quality of the other viewable region which will be discussed in detail in the next section. This enables the client to download a collection of tiles with varying encoding qualities or resolutions, with the visible region prioritized to improve the quality of experience (QoE) while consuming less bandwidth.

Next, OMAF also specifies video profiles based on the High-Efficiency Video Coding (HEVC) coding standard, as well as HEVC-based or older Advanced Video Coding (AVC), AVC-based viewport-dependent profiles that support Equirectangular Projection (ERP), Cubic Mapping Projection (CMP), and tile-based streaming [25]. The comparison of ERP and CMP is shown in Figure 4.
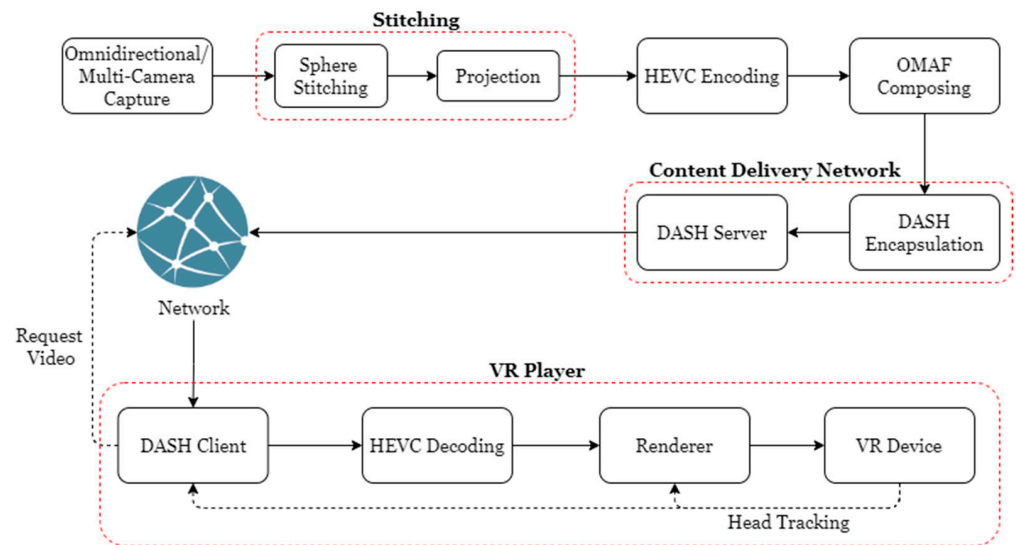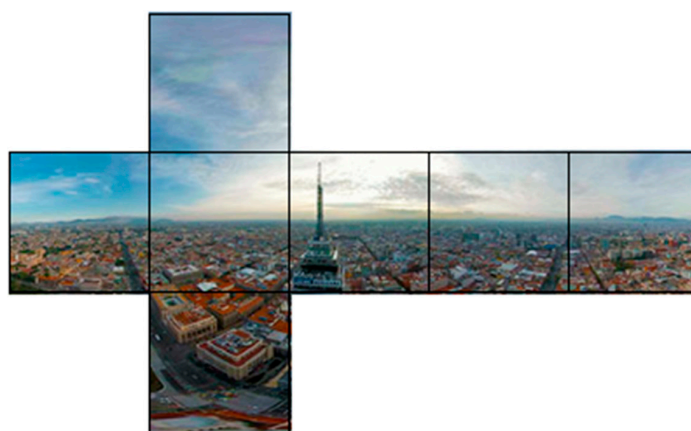
Figure 3. DASH-OMAF architecture network.

Figure 4. Equirectangular projection (EMP) and cube map projection (CMP) comparison.

Clients can stream omnidirectional video from a DASH SRD or OMAF compliant server. The server will deliver segments with different viewport-dependent projections or independent tiles based on the choices of the client, which will be further discussed in the next sections. The client then downloads the appropriate segments, potentially discarding low viewing probability segments or downloads with lower quality to save bandwidth. Next, the features of HEVC of fast Field of View (FoV) switching allow the client to request the segments based on users' head movements in high quality [26], users can even zoom into the region of interest within the 360-degree video [27], providing a smooth user experience with minimal server-side changes.

In recent years, some researchers have enhanced the Quality of Experience (QoE) of 360 videos streaming with the DASH architecture [28]. At any one point in a VR 360-degree movie, the user can at most see a portion of the 360-degree film. As a result, sending the entire picture wastes bandwidth and processing power. With the DASH-based viewpoint of adaptive transmission, these problems may be resolved. The client must pre-download the video material to ensure seamless playing, which needs the client predicting the user's future viewpoint.

Based on HTTP 2.0, a real-time video streaming technology with low latency has been developed by Huang, Ding [29]. The MPEG DASH prototype implements HTTP 2.0 server push functionality to actively deliver live video from the server to the client with low latency whereas Nguyen, Tran [30] suggested an efficient adaptive VR video stream approach based on the DASH transport architecture via HTTP/2 that implements stream prioritization and stream termination.

3.2. Tiling

Tiling is one of the typical solutions proposed by various researchers in order to overcome the bandwidth issues of 360-degree videos by projecting and splitting video frames into numerous sections known as tiles. In general, this technique divides a frame into several sections known as tiles, focusing on the quality of the Region of interest (RoI)/Quality Emphasis Region (QER)/Field of View (FoV) while reducing the others to overcome the bandwidth issue. Most of the solutions are based on the DASH framework as discussed in Section 3.1.

Figure 5 illustrates the small region of FoV in an equirectangular mapped 2K picture. Following that, the most popular HMDs have a small FoV. For example, Google Cardboard [31] and Samsung Gear VR [32] have an FoV of 100 degrees whereas Oculus Rift and HTC Vive [33] have wider 110 degrees of FoV as demonstrated in Figure 6.
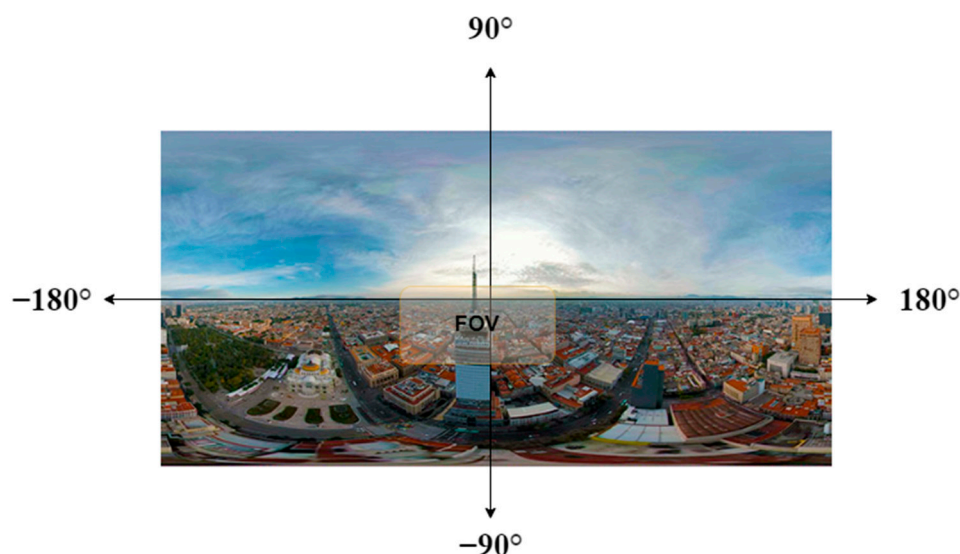

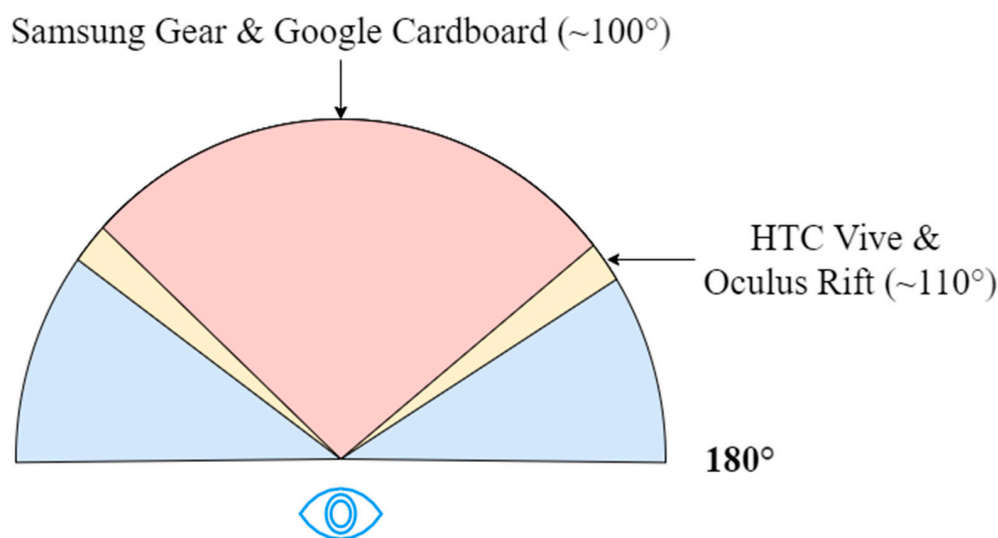**Figure 5.** FoV in a full 360-degree video frame.

Figure 6. FoV associated with the human eye.

Figure 7 shows the methods using the tiling technique whereas Table 3 summarizes and compared the characteristics of each tiling scheme.
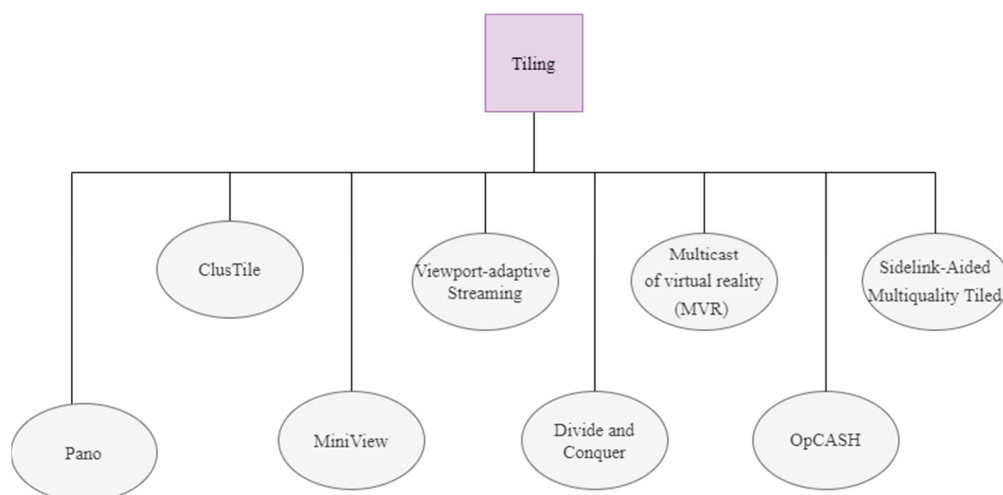
Figure 7. Methods using the tiling technique.

3.2.1. ClusTile

Research as Zhou, Xiao [34] proposed ClusTile, a tiling approach that schemes each tile represents a DASH segment covering a portion of the 360-degree view with typically fixed time intervals, formulated by solving the set of integer linear programs (ILPs). Although this work mentions a decrease of such a high percentage in bandwidth reduction (76%), it does not allow varying the solution of representations but only their bitrate. The increasing number of tiles in the process is not sufficient for the segments downloaded and uploaded.

3.2.2. PANO

Guan, Zheng [35] propose a quality model named Pano for 360° videos that capture the factors that affect the QoE of 360° video including difference in depth-of-field (DoF), relative viewpoint-moving speed and change in scene luminance. The proposed tiling scheme with variable-sized tiles aims to find the tradeoff between the video quality and efficiency of video encoding. Pano achieves 41–46% less bandwidth consumption than Zhou, Xiao [34] with the same Peak Signal-to-Perceptible-Noise Ratio (PSPNR) [35].

3.2.3. MiniView Layout

To reduce the bandwidth requirement of 360-degree video streaming, Xiao, Wang [36] proposed the MiniView Layout which has saved up to 16% of the encoded video without downgrading the visual qualities. In this method, the video was projected into equalized tiles with each MiniView independently encoded into segments. It increases the number of segments and higher in the number of requests parallelly to the streaming client. Plus, Ref. [36] showed improvements in projection efficiency as it created a set of views with the rectilinear projection referred to as "miniview", which has smaller FOVs than cube faces, hence able to save encoded 360-degree videos' storage size without quality loss. Each miniview has its parameters which include FOV, orientation and pixel density [36].

3.2.4. Viewport Adaptive Streaming

In [12], The adaption algorithm initially chooses the video's Quality Emphasized Region (QER) based on the viewport center and the Quality Emphasis Center (QEC) of the available QERs. Each QER-based video is composed of a pre-processed collection of tile representations that are then encoded at various quality levels. This allows for faster server maintenance (fewer files, resulting in a smaller media presentation description (MPD) file), a simpler selection procedure for the client (through a distance computation), and no need to reconstruct the video prior to viewport extraction. However, improved adaption algorithms are required to predict head movement, as well as a new video encoding approach to do quality-differentiated encoding for high-resolution videos.

3.2.5. Divide and Conquer

Research by Hosseini and Swaminathan [37] proposed a divide and conquer approach to increase the bandwidth efficiency of the 360 VR video streaming system. The hierarchical resolution degrading enables a seamless video quality-switching process hence providing a better user experience. Compared to the other method which uses equirectangular projection [37], implements hexaface sphere projection as illustrated in ([37] Figure 4), and significantly saved 72% bandwidth compared to other tiling approaches without viewport awareness. To improve the performance of this approach, an adaptive rate allocation method for tile streaming based on available bandwidth is needed.

3.2.6. Multicast Virtual Reality (MVR)

In [38], the Multicast Virtual Reality (MVR) streaming technique, which is a basic rate adaptation mechanism, serves all members in a multicast group with the same data rate to ensure that all members can receive the video. The data rate is selected based on the member with the poorest network conditions. However, a better tile weighting technique with data-driven probabilistic and an improved rate adaption algorithm is required to improve the user experience.

3.2.7. Sidelink-Aided Multiquality Tiled

Dai, Yue [39] adapt sidelink is a modification of the basic LTE standard that enables device-to-device (D2D) communication in 360-degree streaming without the use of a base station. Allocate tile weight based on long-term weight (how often the tile was visited) and short-term weight (tile distance from the FOV). To find suboptimal solutions with minimal computational cost, a two-stage optimization technique is used to pick sidelink Receivers and Senders in stage 1 and allocate bandwidth and select tile quality level in stage 2.

3.2.8. OpCASH

In [40], a tiling scheme with variable-sized tiles is proposed. To deliver optimal cached tile coverage to user viewports (VP), Mobile Edge Computing (MEC) cache usage is used. Next, an ILP-based technique is used to determine the best cache tile configuration to decrease the redundancy of stored variable tiles at a MEC server while limiting queries to faraway servers, lowering delivery delay, and increasing cache utilization. OpCASH

successfully reduces data fetched from content servers by 85% and overall content delivery time by 74% with MEC.

Table 3. Comparison of existing tiling approaches.

Source	Technique	Result	Limitation
[34]	- Dynamic adaptive HTTP streaming (DASH). - Integer linear programs (ILPs). - Artificial Neural Networks (ANN).	- Saved 76% bandwidth in comparison to the non-tiling scheme. - Saved 52% downloaded volume in comparison to fixed tiling schemes.	A fixed tiling scheme requires tile selection algorithms.
[12]	- Dynamic adaptive HTTP streaming (DASH). - An adaptation algorithm first chooses the Quality Emphasized Region (QER) of the viewport's video-based left and the Quality Emphasis Center (QEC) of the available QERs.	Enable High-quality service with high interactivity with HMD with low management required from VR providers.	Improved adaption algorithms are required to predict head movement, as well as a new video encoding approach to do quality-differentiated encoding for high-resolution videos.
[38]	- Dynamic adaptive HTTP streaming (DASH). - Heuristic algorithm: Multicast Virtual Reality streaming algorithm (MVR).	Increased video bitrates ($\leq$46%) for video tiles in users' viewports.	Require a better tile weighting approach with data-driven probabilistic as well as an improved rate adaption algorithm.
[37]	- MPEG-DASH SRD. - hierarchical resolution degrading - hexaface sphere projection.	72% bandwidth savings.	Improve performance with an adaptive rate allocation method for tile streaming based on available bandwidth.
[35]	- Variable-sized tiling scheme. - A new quality model for 360° videos captures the factors that affect the QoE of 360° video including Difference in depth-of-field (DoF), Relative viewpoint-moving speed and Change in scene luminance.	The same PSPNR was obtained with 41–46 percent reduced bandwidth consumption than [34].	The 360JND model is based on the results of a survey in which the values of 360° video-specific characteristics were varied individually.
[36]	- Higher sphere-to-2D projection efficiency. - The ffmpeg360 program transcodes 360-degree videos and assesses the quality of 360-degree videos based on user head movement patterns. - Created collection of views with the rectilinear projection referred to as "miniview", which have smaller FOVs than cube faces, hence able to save encoded 360-degree videos' storage size while maintaining the quality.	Saved up to 16% encoded video size without much quality loss.	Fixed tiles, each miniview might well be encoded into segments individually, and the streaming client could request these segments as needed.
[39]	- Adapt sidelink. - Weighted tile allocation. - Two-stage optimization technique.	Dai, Yue [39] formulated optimization problems based on the interaction between tile quality level selection, sidelink sender selection, and bandwidth allocation to optimize the overall utility of all users.	When the number of groups is increased from 10 to 50, the tile quality degrades because less bandwidth can be provided to each group as the number of groups grows.
[40]	- Variable-sized tiling scheme. - Adapt MEC cache usage. - ILP-based technique for determining the best cache tile configuration on the MEC server.	OpCASH obtained more than 95 percent VP coverage from cache after only 24 views of the video. When compared to a baseline that illustrates standard tile-based caching, OpCASH reduces data fetched from content servers by 85% and overall content delivery time by 74%.	Improve real-time tile encoding features on content servers by including tile quality selection in the ILP formulation and increasing the variable quality level tiles streaming in. Next, in a lab scenario, interact with many edge nodes using real-world user testing to achieve the biggest benefit at the edge layer.

3.3. Viewport-Based Streaming

In the case of 360-degree video, it would be a waste of network resources to transmit the entire panoramic content as the users typically only see the scenes in the viewport. The bandwidth requirement can be decreased, and transmission efficiency could be improved by identifying and transmitting the current viewport content and the predicted viewport corresponding to the head movement of users. Similar to the tiling technique in the previous section, the server contains a number of video representations that range not just in bitrate but also in the quality of various scene areas. Then, the region of the viewport is dynamically selected and streams in the best quality while the other regions are in lower quality or not being delivered at all to reduce the bandwidth transmission. In other words, the highest bitrate is assigned to tiles in users' viewports, while some other tiles possess bitrates that are proportionate to the likelihood that users may switch viewports, which is also similar to DASH. However, the number of adaption variants of the same content increases dramatically to smooth the viewport-switching due to the sudden head movements. As a result, storage is sacrificed, and the transmission rate increases.

Ribezzo, De Cicco [41] proposed a DASH 360° Immersive Video Streaming Control System which consists of control logic with two cooperating components: quality selection algorithm (QSA) and view selection algorithm (VSA) to dynamically select the demanded video segment. The QSA functions similarly to traditional DASH adaptive video streaming algorithms whereas VSA aims to identify the proper view representation based on the current head position of the users. Ref. [41] reduced segments bitrate around 20% with improved visual quality. In [12], the adaptation algorithm first selects the Quality Emphasized Region (QER) of the video based on the viewport center and the Quality Emphasis Center (QEC) of the available QERs, hence providing high interactive service to head-mounted device (HMD) users with low management. However, improved adaption algorithms are required to predict head movement, as well as a new video encoding approach to do quality-differentiated encoding for high-resolution videos.

High responsiveness and processing power are required to adapt to rapid changes in viewports and viewport prediction to ensure smooth viewport switching with accurate prediction. Many viewport prediction approaches have been developed to cover the demands, such as historical data-driven probabilistic, popularity-based, deep content analysis, and so on as summarized in Table 4.

Table 4. Viewport prediction scheme of the viewport adaptive streaming approach.

Source	Viewport Prediction Scheme	Descriptions
[42]	Historical viewport movement	Prediction with Linear Regression (LR) and Ridge Regression (RR) using viewing data collected from 130 users.
[43]	Cross-user similarity	Cross-Users Behaviors (named CUB360) based on k-NN and LR take into account both the user's specific information and cross-user behavior information to forecast future viewports.
[44]	Popularity-based model	Predict based on the popularity of the tiles where they are visited with a higher frequency at a certain time, might be due to the nature of the video like interesting content along with the evaluation of the rate-distortion curve for each tile.
[45]	Popularity-based model	Similar to [44] and provide the popularity of each shown viewport (heatmap) and rate-distortion function for each tile-representation for the interested segments periodically to clients during each downloading.
[46]	Content Analysis + Popularity	Sensor- and content-based predictive mechanisms, similar to [47] with linear regression (LR). When a transition due to insufficient bandwidth occurs, the tile popularity is solely used to determine the tile quality levels.

Table 4. *Cont.*

Source	Viewport Prediction Scheme	Descriptions
[48]	k-Nearest Neighbors (k-NN)	Improve the accuracy of traditional linear regression (LR) with cross-users watching behaviors that take advantage of prior users' data by identifying common scan paths and allocating a higher chance to future FoVs from those users.
[47]	Deep content analysis	Concurrently leverage sensor characteristics (HMD orientations) and content-related information (image saliency maps and motion maps) with LSTM to predict the viewer fixation in the future. The estimated viewing probability for each equirectangular tile may then be used in the quality optimization based on probability.
[49]	3D-CNN (convolutional neural networks)	3D-CNN to extract the Spatio-temporal features (saliency, motion, and FoV info) from the videos, has better performance than [47].
[50]	Content Analysis + Cross-user similarity	PARIMA, which is a hybrid of Passive Aggressive (PA) Regression and Auto-Regressive Integrated Moving Average (ARIMA) times series models to predict viewports based on users' behavior and the YOLOv3 algorithm on the stitched image to recognize the objects and retrieve their bounding box coordinates in each frame.
[51]	Content Analysis + Cross-user similarity	2 dynamic viewport selection (DVS) which changes the streaming areas depending on content complexity and user head movements to assure viewport accessibility and non-delay visual views for virtual reality users. To achieve higher accuracy, DVS1 focuses on the adjusted prediction distance between two prediction mechanisms whereas DVS2 selects the tiles for the following segment based on the modified prediction difference between actual and predicted perspectives based on content complexity variations.

3.4. Machine Learning (ML)

Machine learning (ML) is used to predict bandwidth and views as well as increase video streaming bitrate to improve the Quality of Experience (QoE) [14]. Table 5 summarizes the many papers that use machine learning to increase QoE in video streaming applications. The proposed scheme in [11] significantly reduces bandwidth consumption by 45% with less than a 0.1% failure ratio while minimizing performance degradation with Naïve linear regression (LR) and neural networks (NN). Next, Dasari, Bhattacharya [52] developed a system called PARSEC (PAnoRamicStrEaming with neural Coding) to reduce bandwidth requirements while improving video quality based on super-resolution, where the video is significantly compressed at the server and the client runs a deep learning model to enhance the video quality. As for this, although Dasari, Bhattacharya [52] successfully reduce the bandwidth requirement and enhance the quality of the video, deep learning is large in models. It also results in the slowest inference rate. Furthermore, Yu, Tillo [53] present a method for adapting to changing video streams with the combination of the Markov Decision Process and Deep Learning (MDP-DL). In Filho, Luizelli [54], a strategy for adapting to fluctuating video streams (the Reinforcement Learning (RL) model) is researched. Next, a Recurrent Neural Network-Long Short-Term Memory(RNN-LSTM) and Logistic Regression-Ridge Regression(LR-RR)) to predict bandwidth and viewpoint is researched by Qian, Han [42] and Zhang, Guan [55]. To increase QoE, Vega, Mocanu [56] suggested a Q-learning technique for adaptive streaming systems. In [57], the deep reinforcement learning (DRL) model uses eye and head movement data to assess the quality of 360-degree videos.

Table 5. Machine learning (ML)-based approaches.

Source	Technique	Scope
[11]	• Naïve linear regression (LR). • Neural networks (NN).	Motion detection and prediction.
[52]	• Deep neural networks (DNN). • Neural-aware adaptive bitrate (ABR) algorithm (§IV).	Reduce bandwidth requirement and Improve video quality.
[53]	• Markov Decision Process-Deep Learning (MDP-DL).	Improve Variable bitrate (VBR).
[54]	• Reinforcement Learning (RL) model.	Improve Adaptive VR Streaming.
[42]	• Logistic Regression-Ridge Regression (LR-RR).	Viewpoint prediction and Bandwidth prediction.
[55]	• Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM).	Viewpoint prediction and Bandwidth prediction.
[56]	• Q-Learning Reinforcement Learning (RL).	Improve constant bitrate (CBR).
[58]	• Markov decision process (MDP). • Deep reinforcement learning (DRL)-based algorithm.	Viewpoint prediction and Optimal bitrate allocation.
[59]	• Encoder-Decoder Based LSTM Model. • Model Predictive Control (MPC)-based rate adaptation.	Viewpoint prediction and Rate adaptation.
[60]	• Markov Decision Process (MDP). • Deep-Q-Network (DQN).	Reactive caching and Viewport prediction.

Kan, Zou [58] deploys RAPT360, a reinforcement learning-based Rate Adaptation with adaptable prediction and tiling for 360-degree video streaming, addresses the needs for precise viewport prediction and efficient bitrate allocation for tiles. Younus, Shafi [59] presents an Encoder-Decoder based Long-Short Term Memory (LSTM) model that transforms data instead of receiving direct input to more correctly capture the non-linear relationship between past and future viewport locations to predict future user movement. To ensure that the 360 films sent to end-users are of the highest possible quality, Maniotis and Thomos [60] propose a reactive caching scheme that uses the Markov Decision Process (MDP to determine the content placement of 360○ videos in edge cache networks and then using the Deep Q-Network (DQN) algorithm, a variant of Q-learning to determine the optimal caching placement and cache the most popular 360○ videos at base quality along with a virtual viewport in high quality.

3.5. Comparison between Techniques

Firstly, the DASH framework, tiling and viewport-adaptive techniques are correlated to each other as most of the tiling and viewport-adaptive techniques are using the DASH framework. Some of the tiling techniques [12,34,37,38] and the viewport-adaptive approach [45,48,49] are all using DASH to stream the areas covered by users' FOV in high quality while some other tiles are streamed in lower quality. The differences between these techniques are the mapping projection, encoding, tiling scheme and tile selection algorithm.

However, there are several limitations to the tiling and viewport-adaptive method. Firstly, more bandwidth is required to stream a screen-size movie at viewport devices as compared to a typical 2D laptop screen at the same quality. As illustrated in Figure 5, streaming a viewport region with a width of 110 degrees is still significantly wider than a normal laptop screen with a width of 48 degrees roughly [3,14]. Furthermore, most tiling solutions employ the viewport-driven technique, in which only the viewport that is the viewed area of the viewer is streamed in high resolution, yet it may also suffer from a significant delay due to the switching of the viewport, which might be due to the video content from the other viewports are not being delivered at the moment. So, when the user

abruptly switches his/her viewport during the display time of the current video segment, a delay occurs. Next, as human eyes have a low delay and error tolerance, any viewport prediction errors can cause rebuffering or quality degradation and result in a break of immersion and poor user Quality of Experience (QoE). Furthermore, to accommodate users' random head movements, causing the need to increase the number of tiles of the video has and thus the video size increases significantly. Therefore, the implications of smooth viewport switching, minimized delays, with lessened video size and bandwidth should be addressed during 360-degree video delivery.

DASH, tiling and viewport-adaptive are focused on improving the streaming efficiency of the 360-degree video with lower bandwidth by streaming the demanded region of the 360-degree video with higher quality. On the other hand, Machine Learning (ML) techniques not only focus on lowering the bandwidth but also focusing on the improvement of QoE of the streaming. The proposed scheme ML also improves video quality, improves bitrate and predicts viewpoint in real-time which as is also effectively reduces bandwidth consumption while minimizing performance degradation. Some of the tiling and viewport-adaptive methods also use some algorithms such as Artificial Neural Network [34], Heuristic algorithm [38] and adaptive algorithm [12] to optimize tile selection and predict the users' viewpoint.

4. Network Approaches to Optimize 360-Degree Video Streaming

Five types of network solutions are proposed by various research to optimize the streaming of 360-degree video to increase the network availability, decrease transmission latency and ease the intensive computing power needed by 360-degree video as illustrated in Figure 8.
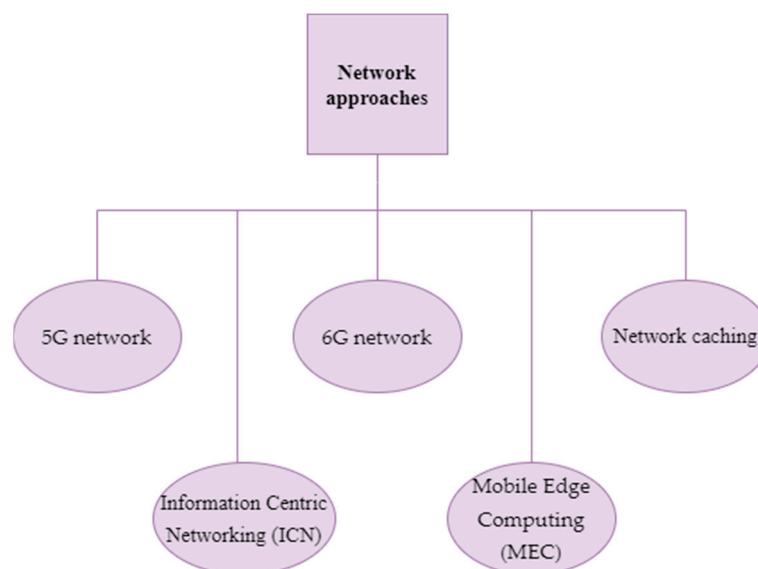

Figure 8. Network approaches to optimize 360-degree video streaming.

4.1. 5G Network

As long-term evolution (LTE), also known as fourth-generation (4G) networks, give way to fifth-generation (5G) networks, 5G components such as edge computing and edge caching are bringing content and computing resources closer to users as compared to cloud computing [61] offers big leaps in the transmission 360-degree virtual reality videos via networks. 5G networks not only boost network capacity and efficiency but also includes computing resources directly into the communication network, which solves the limited computing power of VR devices, resulting in significant gains in user satisfaction [14]. To improve overall QoS, Ref. [62] have proposed a 5G enabled tactile internet-based 360 VR video streaming architecture with a new multicasting mechanism to transmit 360 TI-P2P (TI-Peer-to-peer) live streaming traffic over a MEC-enabled software-defined next-generation

Ethernet passive optical network (NG-EPON) architecture. The 5G network with ultra-low latency and high bandwidth is significant in enabling the highly delay-sensitive 360-degree video streaming with high bandwidth requirements.

4.2. 6G Network

6G has been introduced by Peltonen, Bennis [63] to further improve the 5G network. A proposed AI-powered 6G service application is mobile extended reality (XR), consisting of virtual reality (VR), augmented reality (AR), and mixed reality (MR) [63]. Visuo-haptic XR, which pushes massive real-time data at the network edge, enables remote communication with actual and virtual components in real-time. As a result, it allows computationally complex and data-intensive applications with low delay jitter. In addition, Edge intelligence has shown a lot of potential for allowing XR services especially on devices with limited resource and battery. Intelligent task segmentation, computing offloading, and learning model sharing will all play important roles in providing consumers with high immersion by addressing device battery consumption, computation power, and network latency constraints.

4.3. Network Caching

One of the fundamental elements that allow VR applications to run is the cache. Caching presented at the mobile network's edge is designed to optimize the bandwidth and latency required by VR 360-degree video streaming. Mangiante, Klas [64] demonstrated a mobile network edge solution for optimizing the bandwidth and latency required for VR 360-degree video streaming. Matsuzono, Asaeda [65] offer L4C2, in-network caching with low latency, low loss streaming technique for low delay-tolerance streaming with improved quality in real-time. Chakareski [66] created an optimization framework that enables base stations to choose cooperative caching/rendering/streaming techniques that optimize the accumulated reward they receive while serving customers. In order to maximize the overall performance of 360-degree videos provided to users, Maniotis and Thomos [60] proposed a 360 Video Caching approach using Deep Reinforcement Learning, which is a reactive caching strategy that utilizes the Markov Decision Process (MDP) to determine the location of 360 video content in edge cache networks and then uses the Deep Q-Network (DQN) algorithm, a variant of Q-learning, to determine the optimal caching placement and cache the most popular 360 videos at base quality along with a virtual viewport in high quality.

4.4. Information-Centric Networking (ICN)

Westphal [67] demonstrated how Information-Centric Networking (ICN) may alleviate the network latency problems related to 360-degree video streaming. ICN is a network architecture that transitions from the traditional host-oriented communication model to a content-centric model, relying on location-independent naming schemes, in-network pervasive caching, and content-based routing to enable effective content distribution across the network, allowing content retrieval via any available network interfaces [68].

4.5. Mobile Edge Computing (MEC)

Mobile Edge Compute (MEC) is one of the important elements of 5G [69] specifically beneficial for VR streaming. VR applications require high computing power and data processing power, the CPU and storage capacity of mobile nodes is insufficient for the rendering and computational tasks in VR. In this context, the MEC server may assist by calculating the necessary blocks as target tasks and then delivering the complete task to the mobile VR device [14]. Next, to reduce data related to virtual reality content directly sent to mobile VR devices, the cloud server will first pre-render the VR material, followed by secondary rendering on the mobile VR device.

MEC architectures are also successful in improving network responsiveness and latency, as well as reducing communication resources by leveraging the caching and computational capacity of VR devices for mobile use [70–73]. Liu, Chen [74] claimed that

the MEC architecture can solve insufficient computing power issues that existed on most mobile VR devices. However, the rate of growth of mobile VR data outpaces the rate of growth of wireless network capacity, resulting in a large communication load when sending VR content with the present MEC architecture. Although the impact is minimal, the combination of edge computing with mmWave in mobile VR has been investigated [75,76]. Perfecto, Elbamby [75] researched a user clustering method to increase user field-of-view frame requests, whereas Elbamby, Perfecto [76] investigated an adaptive computation and caching method on the interactive VR video frames to reduce VR game traffic. However, these approaches are primitive, and they only consider 360-degree video lower quality and lower resolution (4K), which affected the delivered experience significantly.

A few recent works display the fusion of MEC and tiling approaches that maximize the advantages of each other. Kumar, Bhagat [77] proposed a tiled 360° caching solution based on Mobile Edge Computing (MEC) with Long–Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), where the LSTM model predicts the future viewport based on popularity and the CNN model identifies the most engaging tiles based on the video's saliency map. This effectively improves the cache hit rate by at least 10% and decreases backhaul utilization by at least 35% while reducing end-to-end latency by at least 35%. Yu, Liu [78] store all versions of 2D and 3D tiles at the MEC server and process the projection of 2D into 3D at the MEC side. In [78], the weighted-sum technique is used to solve a 360° video caching optimization issue, which is a sequential decision-making problem with the combinatorial multi-armed bandit (CMAB) theory and an upgraded combinatorial UCB (ICUCB). Zhang and Chakareski [79] suggested an Unmanned aerial vehicle (UAV)-assisted MEC network and formulated a combined UAV deployment, MEC and radio resource allocation, and 360-degree video content layer assignment (UAV-MV) to enhance QoE for all mobile VR users.

4.6. Comparison between Network Approaches

Table 6 summarizes the scope of each network approach. All of the approaches are able to improve the QoS and QoE of 360-degree video streaming. 5G network, Network caching and ICN mainly improve the network availability for 360-degree video streaming and decrease the network delay. 6G network and MEC mainly optimize the streaming on the client-side by reducing the intensive computing demand and battery consumption on the client device.

Table 6. Comparison of network approaches.

Network Approach	Scope
5G network	Edge computing and Edge caching brings content and resources nearer to the client.
6G network	AI-powered 6G service applications (AR, VR, XR, MR) reduce device battery consumption, computation capacity, and end-to-end latency.
Network caching	Cache the VR content to optimize the bandwidth and latency.
Information Centric Networking (ICN)	Content-centric, location-independent models enable retrieval of content over any network interfaces available.
Mobile Edge Computing (MEC)	Reduce the intensive computing burden on VR devices. The MEC server assists the mobile VR device by processing some computational and rendering tasks and then delivering the task to the mobile device.

5. Quality of Experience (QoE) Assessment

According to the International Telecommunication Union (ITU), Quality of Experience (QoE) could be defined as the overall acceptability of an application or service perceived subjectively by the end-user [80]. QoE assessment is the practice of measuring or estimating

the QoE for a group of users of an application or a service using a defined approach and taking into account the affecting variables [80].

The nature of 360-degree videos which can cause different distortions provides distinct challenges for modeling QoE measurement. To begin, 360-degree videos are captured using many dioptric cameras from different angles, which are then stitched together using a mosaicking algorithm [3]. Azevedo, Birkbeck [81] found that distortions such as blurring, apparent seams, ghosting, broken edges, missing information, and geometrical distortions might arise during the stitching process or capturing process because of the camera inconsistency. For example, the lighting may alter depending on the camera angle which causes the stitched 360-degree video with different illumination affected the user's viewing experience.

Next, 360-degree video streaming required high bandwidth requirements as mentioned in Section 2 of this article. As a result, the video content must be compressed to lower quality, resulting in distortions such as blurring, blocking, ringing, and the staircase effect, all of which can impact the user experience heavily [81]. Due to network restrictions, delays, rebuffering, and quality variations can produce disorientation and cybersickness by disturbing the flow of movement, which also has a detrimental impact on the watching experience.

Using an HMD to view VR 360-degree videos provides an immersive experience. A realistic environment with a higher immersion level and user presence could influence the QoE [82,83]. Thus, more factors than just video quality need to be considered, such as visual realism, acoustic realism and proprioceptive matching between the video content and user's movements influence the sense of presence strongly [84]. Furthermore, Ref. [85] discovered that more diegetic VR interfaces, or situations in which all of the things perceived by a user belong to the virtual world, lead to better user experiences. In comparison to traditional 2D displays, the HMD is significantly closer to the eye, which may cause distortions to be perceived clearly and cause more eye strain and tiredness [86]. Space between pixels may also be noticeable due to the proximity of the screen to the eyes. These effects may raise perceptual and cognitive load (PCL), resulting in stress [87]. Compared to 2D video, it provides more interactivity with the viewing angle but may also cause cybersickness easily.

Currently, there are two main approaches in the QoE assessment of 360-degree video, which are Objective QoE assessment and Subjective QoE assessment.

5.1. Objective QoE Assessment

Objective QoE assessments are video-centric models which analyze the performance of QoE directly bases on the video quality by comparing the presented video's distortions to the original version. This method is known as video quality evaluation (VQA) [88]. The VQA metrics of 360 videos are developed on the basis of the 2D video metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). Various research [89–92] apply Viewport PSNR (V-PSNR) and Spherical Peak Signal-to-Noise Ratio (S-PSNR) [93] based on PSNR modified by sphere-to-plane mappings for 360° video streaming. Sun et al. [94] presented weighted-to-spherically-uniform PSNR (WS-PSNR), which computes PSNR on each pixel of the projected picture before multiplying each pixel with a weight that represents the sphere–plane relationship. Next, adaptations of SSIM, Spherical SSIM (S-SSIM) [95] and Weighted to Spherically Uniform SSIM (WS-SSIM) [96] adjust the structural similarity to compensate for the geometrical distortion using a weighting function. Content Preference PSNR (CP-PSNR) and Content Preference SSIM (CP-SSIM) in [97] showed adaptation to viewport direction and content saliency with a predictive model to predict future viewing direction.

In the above works, the measurement of VQA metrics needs to refer to the reference videos which are the original 360-degree videos without distortion. To overcome this restriction, QoE evaluation models without reference videos have been created [98–100] by assessing metrics derived from the properties of degraded videos or network statistics such as bandwidth, packet loss, and latency.

Croci, Ozcinar [101] divided each 360° video into several patches with equally spaced pixels on the sphere and measure the QoE with extended objective metrics of 2D video using Voronoi patches. VI-PSNR, VI-SSIM, and VI-VMAF are the quality measures that arise from this process. VI-VMAF has the strongest association with the MOSs of total QoE, according to their user survey. Croci, Ozcinar [102] further develop weighted variations of their measures named Visual Attention (VA) maps [31]. VI-VA-metrics, such as VI-VA-PSNR and VI-VA-VMAF are the resultant metrics. According to the findings, VI-VA-VMAF and VI-VA-MS-SSIM are the two quality indicators that have the strongest impact on the overall QoE.

However, in most real-world video streaming settings, objective QoE measurement is problematic because it ignores the perception of the users such as immersiveness and cybersickness. Moreover, these widely used 360° video quality indicators have been demonstrated to have a low impact on user satisfaction [103]. Thus, the assessment of QoE in 360-degree material using 2D video theory and methodology requires additional specialized 360-video QoE study.

5.2. Subjective QoE Assessment

Visual attention-enhanced models have been introduced to overcome the limitation of objective QoE assessment. Users in an immersive virtual reality environment can only be perceived from their FOV, and usually only focus on certain objects, i.e., buildings which grab their attention. Thus, distortions in various parts of the 360-degree video have distinct effects on QoE. In estimating the PSNR based on the viewer's visual attention distribution, Xu, Li [104] employed the usual PSNR metric to apply weights to the pixel-wise distortion. VQA-OV [57] employs a similar strategy. To induce visual attention, a VR headset's inertial sensors and eye tracker monitor the viewer's head and eye movements. In [105], they created a subject's field of view (FoV) and saliency map to aid VQA evaluation.

Although these models incorporate visual information, it is mostly influenced by the user's visual attention. Lately, several studies have recently begun to include human factors in 360-degree video QoE evaluations. A link between eye-based signals and human perceptions has been acknowledged by [106,107]. For instance, Ref. [80] employs physiological features of viewers' ocular motions such as eye gazing, fixations, saccades, pupillometry, and different types of eye-opening and closing events to infer how contented they are with 360-degree videos.

However, the results in [86] found that adding eye-tracking metrics in the model did not further explain individual variation in subjective assessment. The proportion of the total viewing area examined had no impact on the quality of experience. Participants become less sensitive to quality distortions as they look at more moving things, and the effect of freezing becomes rather more adverse. The effects discovered are minor, and the variation on a participant-by-participant basis remains substantial. Visual attention implementation based on these findings is most likely insufficient to enhance objective QoE measures.

In VR applications, biosensors may also be utilized to estimate subjective quality. When people watch 360° videos, for example, Egan, Brennan [108] record their heart rate and electrodermal activity. The relationship between these two measurements and subjective scores is examined and discussed. Their findings show that HMDs provide higher subjective quality than 2D displays.

Singla, Göring [109] perform a user study, particularly on tiled 360° video streaming. They consider the effects of many aspects including bandwidth, latency, and resolution. They also assess perceived quality and cybersickness in response to various forms of latency, such as tile switching and network delays. Their findings suggest that a network delay of 47 milliseconds is acceptable and does not impact quality ratings. Fan, Hung [110] calculate the QoE of tiled-based streaming with Mean Opinion Score (MOS) and Individual Score (IS) under uniform-bitrate templates. MOS stands for the average quality scores of a group of subjects assessing an experience of watching tiled 360° films with HMDs,

whereas IS stands for the quality ratings provided by each subject on his or her personal experience. The findings show that the content variables such as video complexity in Temporal Information (TI) and Spatial Information (SI), as well as video quality, dominate the factor categories for overall QoE and most QoE features.

van Kasteren, Brunnström [86] investigate the impacts of quality degradations, freezing, and content on QoE, as well as evaluate visual attention as a factor in QoE. Quality degradation did not affect QoE until the threshold. In every circumstance, freezing occurrences lower the QoE below an acceptable threshold. Additional compression is favored over freezing when network resources are limited. Furthermore, the findings of this paper reveal that perceptual and cognitive load (PCL) and cybersickness do not have a significant impact on QoE and are unaffected by manipulations. Freezing affects visual attention but switching between different levels of degradation does not affect visual attention to the experience.

QoE assessment metrics are still in development and subjective methods are yet to be standardized. A comprehensive QoE assessment model is required to evaluate the performance of different approaches such as tiling, machine learning and viewport adaptive approach that strive to improve the QoE of 360-degree video.

6. Discussion and Future Works

First, a more efficient 360-degree video technique with the combination of the techniques in Section 3 is plausible. With the combination of the strengths of DASH, tiling and ML, a better streaming approach with low bandwidth, high video quality and high QoE could be produced. However, the comparison of the particular techniques still needs further discussion which includes the comparison of the tiling schemes, ML algorithms, DASH encoded schemes to filter the best among them. Moreover, the other factors such as areas of projection, encoding and bitrate adaption need to be considered as well. This is very challenging, as for now, a unified, standardized assessment has not existed to measure the QoE and QoS 360-degree video streaming. For instance, Ref. [35] proposed the survey-based 360JND model whereas [36] used the ffmpeg360 tool to evaluate the quality of the videos. Thus, a standardized assessment is required to have a better and more accurate evaluation of the QoE and QoS of 360-degree video services. Quantitative and qualitative measurements are required to perform the comparison between the QoE assessment model.

On the other hand, the improvement of networks such as 5G and 6G is optimistic for the development of the 360-degree video. 5G and 6G networks will play a main role to overcome the delay and compensate for the high bandwidth requirements of 360-degree video by increasing the network availability along with the other optimized network streaming approaches such as edge computing, network caching and MEC which bring the content nearer to the client and solve the intensive computing requirement on the client-side. With the fast evolution because of the fast evolution of the network in recent years, a deep fake virtual world, "Metaverse" with high immersion and conducive for daily life working, learning and socializing are feasible. Moreover, there are more emerging research studies that implement the combination of the bandwidth reduction techniques in Section 3 and network approaches in Section 4 to deliver 360-degree video streaming more efficiently. For instance, Refs. [77,78] combine the strength of MEC, ML and tiling to improve the efficiency in 360-degree video streaming in terms of tile selection and reduce the computational requirement with the MEC server.

All the research studies mentioned in this paper strive to improve the QoE and QoS of 360-degree video, AR and VR. For a good user experience, immersion of users is the key. With the promotion of big corporates such as Microsoft and Meta, a day where people highly depend on these technologies to carry out daily activities in the virtual world seems to be around the corner.

7. Conclusions

Many scholars are intrigued by the development of 360-degree video technology. It is widely used in various sectors such as education, entertainment and economics. Many good attempts have been made over the years to improve 360-degree video streaming. However, due to their high video quality ($\geq$6K), such videos have always required a higher bandwidth than typical videos.

This paper described the streaming architecture of 360-degree video, MPEG-DASH and DASH-OMAF compliant. Modern streaming technologies, such as tile-based streaming, viewport-based streaming and machine learning, are given and discussed in an effort to reduce bandwidth-delay needs and increase the quality of high-resolution content. Following that, this article explores network techniques such as network caching, 5G and 6G networks, and MEC to optimize 360-degree video streaming. Next, the challenges faced when modeling QoE assessment of 360-degree video and the existing QoE assessment methods which include various objective and subjective assessment methods have been discussed to evaluate the QoE of 360-degree video.

Despite the topic's popularity and extensive research efforts, there are significant research challenges, most notably in the fields of bitrate optimization, encoding schemes, tiling weighting schemes, projection formats, viewport prediction approach, and so on. In terms of delivering relevant knowledge for 360-degree video streaming, standardized projects are already showing a lot of potential. Such problems should be addressed before deployment to enable the utmost user experience.

References

1. Brewer, J. Cisco Launches Webex Hologram, An AR Meeting Solution. 2021. Available online: https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=2202545 (accessed on 28 December 2021).
2. Indigo9 Digital Inc. 10 of the Best Augmented Reality (AR) Shopping Apps to Try Today. 2021. Available online: https://www.indigo9digital.com/blog/how-six-leading-retailers-use-augmented-reality-apps-to-disrupt-the-shopping-experience (accessed on 5 December 2021).
3. Shafi, R.; Shuai, W.; Younus, M.U. 360-Degree Video Streaming: A Survey of the State of the Art. *Symmetry* **2020**, *12*, 1491. [CrossRef]
4. Reyna, J. The Potential of 360-Degree Videos for Teaching, Learning and Research. *INTED Proc.* **2018**, 1448–1454.
5. Lampropoulos, G.; Barkoukis, V.; Burden, K.; Anastasiadis, T. 360-degree video in education: An overview and a comparative social media data analysis of the last decade. *Smart Learn. Environ.* **2021**, *8*, 20. [CrossRef]
6. Meta. Introducing Meta: A Social Technology Company. 2021. Available online: https://about.fb.com/news/2021/10/facebook-company-is-now-meta/ (accessed on 28 December 2021).

7.   Sriram, S. JPMorgan Opens a Lounge in Decentraland, Sees $1 Trillion Metaverse Opportunity. 2022. Available online: https://www.benzinga.com/markets/cryptocurrency/22/02/25655613/jpmorgan-opens-a-lounge-in-decentraland-sees-1-trillion-metaverse-opportunity (accessed on 18 February 2022).

8.   Strickland, J.; Pollette, C. How Second Life Works. 2021. Available online: https://computer.howstuffworks.com/internet/social-networking/networks/second-life.htm (accessed on 14 December 2021).

9.   SM Entertainment. SM Brand Marketing Signs a Metaverse Partnership with the World's Largest Metaverse Platform, The Sandbox. 2022. Available online: https://smentertainment.com/PressCenter/Details/7936 (accessed on 27 February 2022).

10.   Molina, B. Bill Gates Predicts Our Work Meetings Will Move to Metaverse in 2–3 years. 2021. Available online: https://www.usatoday.com/story/tech/2021/12/10/bill-gates-metaverse-work-meetings-predictions/6459911001/ (accessed on 15 December 2021).

11.   Bao, Y.; Wu, H.; Zhang, T.; Ramli, A.A.; Liu, X. Shooting a moving target: Motion-prediction-based transmission for 360-degree videos. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; pp. 1161–1170.

12.   Corbillon, X.; Simon, G.; Devlic, A.; Chakareski, J. Viewport-adaptive navigable 360-degree video delivery. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017.

13.   Zhou, Y.; Sun, B.; Qi, Y.; Peng, Y.; Liu, L.; Zhang, Z.; Liu, Y.; Liu, D.; Li, Z.; Tian, L. Mobile AR/VR in 5G based on convergence of communication and computing. *Telecommun. Sci.* **2018**, *34*, 19–33.

14.   Ruan, J.; Xie, D. Networked VR: State of the Art, Solutions, and Challenges. *Electronics* **2021**, *10*, 166. [CrossRef]

15.   Grzelka, A.; Dziembowski, A.; Mieloch, D.; Stankiewicz, O.; Stankowski, J.; Domanski, M. Impact of video streaming delay on user experience with head-mounted displays. In Proceedings of the 2019 Picture Coding Symposium (PCS), Ningbo, China, 12–15 November 2019; IEEE: New York, NY, USA, 2019.

16.   Mania, K.; Adelstein, B.D.; Ellis, S.R.; Hill, M.I. Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization, Los Angeles, CA, USA, 7–8 August 2004.

17.   Albert, R.; Patney, A.; Luebke, D.; Kim, J. Latency Requirements for Foveated Rendering in Virtual Reality. *ACM Trans. Appl. Percept.* **2017**, *14*, 25. [CrossRef]

18.   Chen, M.; Saad, W.; Yin, C. Virtual Reality Over Wireless Networks: Quality-of-Service Model and Learning-Based Resource Management. *IEEE Trans. Commun.* **2018**, *66*, 5621–5635. [CrossRef]

19.   Doppler, K.; Torkildson, E.; Bouwen, J. On wireless networks for the era of mixed reality. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017; IEEE: New York, NY, USA, 2017.

20.   Ju, R.; He, J.; Sun, F.; Li, J.; Li, F.; Zhu, J.; Han, L. Ultra wide view based panoramic VR streaming. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, Los Angeles, CA, USA, 25 August 2017.

21.   Gohar, A.; Lee, S. Multipath Dynamic Adaptive Streaming over HTTP Using Scalable Video Coding in Software Defined Networking. *Appl. Sci.* **2020**, *10*, 7691. [CrossRef]

22.   Hannuksela, M.M.; Wang, Y.-K.; Hourunrant, A. An overview of the OMAF standard for 360 video. In Proceedings of the 2019 Data Compression Conference (DCC), Snowbird, UT, USA, 26–29 March 2019; IEEE: New York, NY, USA, 2019.

23.   Monnier, R.; van Brandenburg, R.; Koenen, R. Streaming UHD-Quality VR at realistic bitrates: Mission impossible? In Proceedings of the 2017 NAB Broadcast Engineering and Information Technology Conference (BEITC), Las Vegas, NV, USA, 22–27 April 2017.

24.   Skupin, R.; Sanchez, Y.; Podborski, D.; Hellge, C.; Schierl, T. Viewport-dependent 360 degree video streaming based on the emerging Omnidirectional Media Format (OMAF) standard. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; IEEE: New York, NY, USA, 2017.

25.   Chiariotti, F. A survey on 360-degree video: Coding, quality of experience and streaming. *Comput. Commun.* **2021**, *177*, 133–155. [CrossRef]

26.   Song, J.; Yang, F.; Zhang, W.; Zou, W.; Fan, Y.; Di, P. A fast fov-switching dash system based on tiling mechanism for practical omnidirectional video services. *IEEE Trans. Multimed.* **2019**, *22*, 2366–2381. [CrossRef]

27.   D'Acunto, L.; Van den Berg, J.; Thomas, E.; Niamut, O. Using MPEG DASH SRD for zoomable and navigable video. In Proceedings of the 7th International Conference on Multimedia Systems, Klagenfurt, Austria, 10–13 May 2016.

28.   Xie, L.; Xu, Z.; Ban, Y.; Zhang, X.; Guo, Z. 360 probdash: Improving qoe of 360 video streaming using tile-based http adaptive streaming. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017.

29.   Huang, W.; Ding, L.; Wei, H.Y.; Hwang, J.N.; Xu, Y.; Zhang, W. Qoe-oriented resource allocation for 360-degree video transmission over heterogeneous networks. *arXiv* **2018**, arXiv:1803.07789.

30.   Nguyen, D.; Tran, H.T.; Thang, T.C. A client-based adaptation framework for 360-degree video streaming. *J. Vis. Commun. Image Represent.* **2019**, *59*, 231–243. [CrossRef]

31.   Google Cardboard. 2021. Available online: https://arvr.google.com/cardboard/ (accessed on 30 December 2021).

32.   Samsung Gear VR. 2021. Available online: https://www.samsung.com/global/galaxy/gear-vr/ (accessed on 30 December 2021).

33.   HTC Vive VR. 2021. Available online: https://www.vive.com/ (accessed on 30 December 2021).

34. Zhou, C.; Xiao, M.; Liu, Y. Clustile: Toward minimizing bandwidth in 360-degree video streaming. In Proceedings of the IEEE INFOCOM 2018-IEEE Conference on Computer Communications, Honolulu, HI, USA, 16–19 April 2018; IEEE: New York, NY, USA, 2018.

35. Guan, Y.; Zheng, C.; Zhang, X.; Guo, Z.; Jiang, J. Pano: Optimizing 360 video streaming with a better understanding of quality perception. In Proceedings of the ACM Special Interest Group on Data Communication, Beijing China, 19–23 August 2019; pp. 394–407.

36. Xiao, M.; Wang, S.; Zhou, C.; Liu, L.; Li, Z.; Liu, Y.; Chen, S. Miniview layout for bandwidth-efficient 360-degree video. In Proceedings of the 26th ACM international Conference on Multimedia, Seoul, Korea, 22–26 October 2018.

37. Hosseini, M.; Swaminathan, V. Adaptive 360 VR video streaming: Divide and conquer. In Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016; IEEE: New York, NY, USA, 2016.

38. Ahmadi, H.; Eltobgy, O.; Hefeeda, M. Adaptive Multicast Streaming of Virtual Reality Content to Mobile Users. In Proceedings of the on Thematic Workshops of ACM Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 170–178.

39. Dai, J.; Yue, G.; Mao, S.; Liu, D. Sidelink-Aided Multiquality Tiled 360° Virtual Reality Video Multicast. *IEEE Internet Things J.* **2022**, *9*, 4584–4597. [CrossRef]

40. Madarasingha, C.; Thilakarathna, K.; Zomaya, A. OpCASH: Optimized Utilization of MEC Cache for 360-Degree Video Streaming with Dynamic Tiling. In Proceedings of the 2022 IEEE International Conference on Pervasive Computing and Communications (PerCom), Pisa, Italy, 22–25 March 2022.

41. Ribezzo, G.; De Cicco, L.; Palmisano, V.; Mascolo, S. A DASH 360 ° immersive video streaming control system. *Internet Technol. Lett.* **2020**, *3*, e175. [CrossRef]

42. Qian, F.; Han, B.; Xiao, Q.; Gopalakrishnan, V. Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, New Delhi, India, 29 October–2 November 2018.

43. Ban, Y.; Xie, L.; Xu, Z.; Zhang, X.; Guo, Z.; Wang, Y. CUB360: Exploiting Cross-Users Behaviors for Viewport Prediction in 360 Video Adaptive Streaming. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.

44. Chakareski, J.; Aksu, R.; Corbillon, X.; Simon, G.; Swaminathan, V. Viewport-Driven Rate-Distortion Optimized 360° Video Streaming. In Proceedings of the EEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018; pp. 1–7.

45. Rossi, S.; Toni, L. Navigation-Aware Adaptive Streaming Strategies for Omnidirectional Video. In Proceedings of the IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, 16–18 October 2017; pp. 1–6.

46. Koch, C.; Rak, A.-T.; Zink, M.; Steinmetz, R.; Rizk, A. Transitions of viewport quality adaptation mechanisms in 360° video streaming. In Proceedings of the 29th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video, Amherst, MA, USA, 21 June 2019; pp. 14–19.

47. Fan, C.L.; Lee, J.; Lo, W.C.; Huang, C.Y.; Chen, K.T.; Hsu, C.H. Fixation Prediction for 360° Video Streaming in Head-Mounted Virtual Reality. In Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video, Taipei, Taiwan, 20–23 June 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 67–72.

48. Xu, Z.; Ban, Y.; Zhang, K.; Xie, L.; Zhang, X.; Guo, Z.; Meng, S.; Wang, Y. Tile-Based Qoe-Driven Http/2 Streaming System For 360 Video. In Proceedings of the 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), San Diego, CA, USA, 23–27 July 2018; pp. 1–4.

49. Park, S.; Bhattacharya, A.; Yang, Z.; Dasari, M.; Das, S.R.; Samaras, D. Advancing User Quality of Experience in 360-degree Video Streaming. In Proceedings of the 2019 IFIP Networking Conference (IFIP Networking), Warsaw, Poland, 20–22 May 2019; pp. 1–9.

50. Chopra, L.; Chakraborty, S.; Mondal, A.; Chakraborty, S. PARIMA: Viewport Adaptive 360-Degree Video Streaming. In Proceedings of the Web Conference 2021, Ljubljana Slovenia, 19–23 April 2021.

51. Yaqoob, A.; Togou, M.A.; Muntean, G.-M. Dynamic Viewport Selection-Based Prioritized Bitrate Adaptation for Tile-Based 360° Video Streaming. *IEEE Access* **2022**, *10*, 29377–29392. [CrossRef]

52. Dasari, M.; Bhattacharya, A.; Vargas, S.; Sahu, P.; Balasubramanian, A.; Das, S.R. Streaming 360-Degree Videos Using Super-Resolution. In Proceedings of the IEEE INFOCOM 2020—IEEE Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020.

53. Yu, L.; Tillo, T.; Xiao, J. QoE-driven dynamic adaptive video streaming strategy with future information. *IEEE Trans. Broadcasting* **2017**, *63*, 523–534. [CrossRef]

54. Filho, R.I.T.D.C.; Luizelli, M.C.; Petrangeli, S.; Vega, M.T.; Van der Hooft, J.; Wauters, T.; De Turck, F.; Gaspary, L.P. Dissecting the Performance of VR Video Streaming through the VR-EXP Experimentation Platform. *ACM Trans. Multimedia Comput. Commun. Appl.* **2019**, *15*, 1–23. [CrossRef]

55. Zhang, Y.; Guan, Y.; Bian, K.; Liu, Y.; Tuo, H.; Song, L.; Li, X. EPASS360: QoE-aware 360-degree video streaming over mobile devices. *IEEE Trans. Mob. Comput.* **2020**, *20*, 2338–2353. [CrossRef]

56. Vega, M.T.; Mocanu, D.C.; Barresi, R.; Fortino, G.; Liotta, A. Cognitive streaming on android devices. In Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), Ottawa, ON, Canada, 11–15 May 2015; IEEE: New York, NY, USA, 2015.

57. Li, C.; Xu, M.; Du, X.; Wang, Z. Bridge the gap between VQA and human behavior on omnidirectional video: A large-scale dataset and a deep learning model. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018.

58. Kan, N.; Zou, J.; Li, C.; Dai, W.; Xiong, H. RAPT360: Reinforcement Learning-Based Rate Adaptation for 360-Degree Video Streaming with Adaptive Prediction and Tiling. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 1607–1623. [CrossRef]

59. Younus, M.U.; Shafi, R.; Rafiq, A.; Anjum, M.R.; Afridi, S.; Jamali, A.A.; Arain, Z.A. Encoder-Decoder Based LSTM Model to Advance User QoE in 360-Degree Video. *Comput. Mater. Contin.* **2022**, *71*, 2617–2631.

60. Maniotis, P.; Thomos, N. Viewport-Aware Deep Reinforcement Learning Approach for 360° Video Caching. *IEEE Trans. Multimed.* **2022**, *24*, 386–399. [CrossRef]

61. Partners, S. How 5G and Edge Computing Will Transform AR & VR Use Cases. Available online: https://stlpartners.com/articles/edge-computing/how-5g-and-edge-computing-will-transform-ar-vr-use-cases/ (accessed on 18 February 2022).

62. Ganesan, E.; Liem, A.T.; Hwang, I.-S. QoS-Aware Multicast for Crowdsourced 360° Live Streaming in SDN Aided NG-EPON. *IEEE Access* **2022**, *10*, 9935–9949. [CrossRef]

63. Peltonen, E.; Bennis, M.; Capobianco, M.; Debbah, M.; Ding, A.; Gil-Castiñeira, F.; Jurmu, M.; Karvonen, T.; Kelanti, M.; Kliks, A.; et al. 6G White Paper on Edge Intelligence. *arXiv* **2020**, arXiv:2004.14850.

64. Mangiante, S.; Klas, G.; Navon, A.; GuanHua, Z.; Ran, J.; Silva, M.D. VR is on the Edge: How to Deliver 360° Videos in Mobile Networks. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, Los Angeles, CA, USA, 25 August 2017; pp. 30–35.

65. Matsuzono, K.; Asaeda, H.; Turletti, T. Low latency low loss streaming using in-network coding and caching. In Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; IEEE: New York, NY, USA, 2017.

66. Chakareski, J. VR/AR immersive communication: Caching, edge computing, and transmission trade-offs. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, Los Angeles, CA, USA, 25 August 2017.

67. Westphal, C. Challenges in Networking to Support Augmented Reality and Virtual Reality. In Proceedings of the IEEE ICNC 2017, Silicon Valley, CA, USA, 26–29 January 2017.

68. Westphal, C. *Adaptive Video Streaming in Information-Centric Networking (ICN)*; IRTF RFC7933, ICN Research Group. (August 2016); IETF: Fremont, CA, USA, 2016.

69. Lav Gupta, R.J.; Chan, H.A. Mobile Edge Computing—An Important Ingredient of 5G Networks. 2016. Available online: https://sdn.ieee.org/newsletter/march-2016/mobile-edge-computing-an-important-ingredient-of-5g-networks (accessed on 18 February 2022).

70. Dai, J.; Liu, D. An mec-enabled wireless vr transmission system with view synthesis-based caching. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), Marrakech, Morocco, 15–18 April 2019; IEEE: New York, NY, USA, 2019.

71. Dai, J.; Zhang, Z.; Mao, S.; Liu, D. A View Synthesis-Based 360° VR Caching System Over MEC-Enabled C-RAN. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 3843–3855. [CrossRef]

72. Liu, Y.; Liu, J.; Argyriou, A.; Ci, S. MEC-Assisted Panoramic VR Video Streaming Over Millimeter Wave Mobile Networks. *IEEE Trans. Multimed.* **2018**, *21*, 1302–1316. [CrossRef]

73. Yang, X.; Chen, Z.; Li, K.; Sun, Y.; Liu, N.; Xie, W.; Zhao, Y. Communication-Constrained Mobile Edge Computing Systems for Wireless Virtual Reality: Scheduling and Tradeoff. *IEEE Access* **2018**, *6*, 16665–16677. [CrossRef]

74. Liu, H.; Chen, Z.; Qian, L. The three primary colors of mobile systems. *IEEE Commun. Mag.* **2016**, *54*, 15–21. [CrossRef]

75. Perfecto, C.; Elbamby, M.S.; Del Ser, J.; Bennis, M. Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework. *IEEE Trans. Commun.* **2020**, *68*, 2491–2508. [CrossRef]

76. Elbamby, M.S.; Perfecto, C.; Bennis, M.; Doppler, K. Edge computing meets millimeter-wave enabled VR: Paving the way to cutting the cord. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018; IEEE: New York, NY, USA, 2018.

77. Kumar, S.; Bhagat, L.A.; Franklin, A.A.; Jin, J. Multi-neural network based tiled 360° video caching with Mobile Edge Computing. *J. Netw. Comput. Appl.* **2022**, *201*, 103342. [CrossRef]

78. Yu, Z.; Liu, J.; Liu, S.; Yang, Q. Co-Optimizing Latency and Energy with Learning Based 360° Video Edge Caching Policy. In Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022.

79. Zhang, L.; Chakareski, J. UAV-Assisted Edge Computing and Streaming for Wireless Virtual Reality: Analysis, Algorithm Design, and Performance Guarantees. *IEEE Trans. Veh. Technol.* **2022**, *71*, 3267–3275. [CrossRef]

80. ITU-T. Vocabulary for Performance, Quality of Service and Quality of Experience. 2017. Available online: https://www.itu.int/rec/T-REC-P.10-201711-I/en (accessed on 1 June 2022).

81. Azevedo, R.G.D.A.; Birkbeck, N.; De Simone, F.; Janatra, I.; Adsumilli, B.; Frossard, P. Visual Distortions in 360° Videos. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2524–2537. [CrossRef]

82. Tran, H.T.; Ngoc, N.P.; Pham, C.T.; Jung, Y.J.; Thang, T.C. A subjective study on QoE of 360 video for VR communication. In Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Luton, UK, 16–18 October 2017; IEEE: New York, NY, USA, 2017.

83. Cummings, J.J.; Bailenson, J.N. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychol.* **2014**, *19*, 272–309. [CrossRef]

84. Zou, W.; Yang, F.; Zhang, W.; Li, Y.; Yu, H. A Framework for Assessing Spatial Presence of Omnidirectional Video on Virtual Reality Device. *IEEE Access* **2018**, *6*, 44676–44684. [CrossRef]

85. Salomoni, P.; Prandi, C.; Roccetti, M.; Casanova, L.; Marchetti, L.; Marfia, G. Diegetic user interfaces for virtual environments with HMDs: A user experience study with oculus rift. *J. Multimodal User Interfaces* **2017**, *11*, 173–184. [CrossRef]

86. Van Kasteren, A.; Brunnström, K.; Hedlund, J.; Snijders, C. Quality of experience of 360 video—Subjective and eye-tracking assessment of encoding and freezing distortions. *Multimed. Tools Appl.* **2022**, *81*, 9771–9802. [CrossRef]

87. Sweller, J. Cognitive load theory. In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2011; pp. 37–76.

88. Zhu, H.; Li, T.; Wang, C.; Jin, W.; Murali, S.; Xiao, M.; Ye, D.; Li, M. EyeQoE: A Novel QoE Assessment Model for 360-degree Videos Using Ocular Behaviors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 39. [CrossRef]

89. Wang, Y.; Liu, D.; Ma, S.; Wu, F.; Gao, W. Spherical Coordinates Transform-Based Motion Model for Panoramic Video Coding. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 98–109. [CrossRef]

90. Yu, M.; Lakshman, H.; Girod, B. A framework to evaluate omnidirectional video coding schemes. In Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality, Fukuoka, Japan, 29 September–3 October 2015; IEEE: New York, NY, USA, 2015.

91. Alshina, E.; Boyce, J.; Abbas, A.; Ye, Y. JVET common test conditions and evaluation procedures for 360° video. JVET document, JVET-G1030. In Proceedings of the Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 7th Meeting, Torino, Italy, 13–21 July 2017.

92. Lo, W.-C.; Fan, C.-L.; Yen, S.-C.; Hsu, C.-H. Performance measurements of 360 video streaming to head-mounted displays over live 4G cellular networks. In Proceedings of the 2017 19th Asia-Pacific Network Operations and Management Symposium (APNOMS), Seoul, Korea, 27–29 September 2017; IEEE: New York, NY, USA, 2017.

93. Yu, M.; Lakshman, H.; Girod, B. Content adaptive representations of omnidirectional videos for cinematic virtual reality. In Proceedings of the 3rd International Workshop on Immersive Media Experiences, Brisbane, Australia, 26–30 October 2015.

94. Sun, Y.; Lu, A.; Yu, L. Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video. *IEEE Signal Process. Lett.* **2017**, *24*, 1408–1412. [CrossRef]

95. Chen, S.; Zhang, Y.; Li, Y.; Chen, Z.; Wang, Z. Spherical structural similarity index for objective omnidirectional video quality assessment. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; IEEE: New York, NY, USA, 2018.

96. Zhou, Y.; Yu, M.; Ma, H.; Shao, H.; Jiang, G. Weighted-to-spherically-uniform SSIM objective quality evaluation for panoramic video. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018; IEEE: New York, NY, USA, 2018.

97. Rai, Y.; le Callet, P.; Guillotel, P. Which saliency weighting for omni directional image quality assessment? In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; IEEE: New York, NY, USA, 2017.

98. Sun, W.; Min, X.; Zhai, G.; Gu, K.; Duan, H.; Ma, S. MC360IQA: A Multi-channel CNN for Blind 360-Degree Image Quality Assessment. *IEEE J. Sel. Top. Signal Process.* **2019**, *14*, 64–77. [CrossRef]

99. Singla, A.; Fremerey, S.; Robitza, W.; Lebreton, P.; Raake, A. Comparison of subjective quality evaluation for HEVC encoded omnidirectional videos at different bit-rates for UHD and FHD resolution. In Proceedings of the Thematic Workshops of ACM Multimedia 2017, Mountain View, CA, USA, 23–27 October 2017.

100. Fei, Z.; Wang, F.; Wang, J.; Xie, X. QoE Evaluation Methods for 360-Degree VR Video Transmission. *IEEE J. Sel. Top. Signal Process.* **2019**, *14*, 78–88. [CrossRef]

101. Croci, S.; Ozcinar, C.; Zerman, E.; Cabrera, J.; Smolic, A. Voronoi-based Objective Quality Metrics for Omnidirectional Video. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2019; pp. 1–6.

102. Croci, S.; Ozcinar, C.; Zerman, E.; Knorr, S.; Cabrera, J.; Smolic, A. Visual attention-aware quality estimation framework for omnidirectional video using spherical Voronoi diagram. *Qual. User Exp.* **2020**, *5*, 4. [CrossRef]

103. Upenik, E.; Rerabek, M.; Ebrahimi, T. On the performance of objective metrics for omnidirectional visual content. In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; IEEE: New York, NY, USA, 2017.

104. Xu, M.; Li, C.; Chen, Z.; Wang, Z.; Guan, Z. Assessing Visual Quality of Omnidirectional Videos. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 3516–3530. [CrossRef]

105. Li, C.; Xu, M.; Jiang, L.; Zhang, S.; Tao, X. Viewport Proposal CNN for 360° Video Quality Assessment. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 10169–10178.

106. Marandi, R.Z.; Madeleine, P.; Omland, Ø.; Vuillerme, N.; Samani, A. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Sci. Rep.* **2018**, *8*, 13148. [CrossRef] [PubMed]

107. O'Dwyer, J.; Murray, N.; Flynn, R. Eye-based Continuous Affect Prediction. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), Cambridge, UK, 3–6 September 2019; IEEE: New York, NY, USA, 2019.

108. Egan, D.; Brennan, S.; Barrett, J.; Qiao, Y.; Timmerer, C.; Murray, N. An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. In Proceedings of the 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016; IEEE: New York, NY, USA, 2016.

109. Singla, A.; Göring, S.; Raake, A.; Meixner, B.; Koenen, R.; Buchholz, T. Subjective quality evaluation of tile-based streaming for omnidirectional videos. In Proceedings of the 10th ACM Multimedia Systems Conference, Amherst, MA, USA, 18–21 June 2019.

110. Fan, C.-L.; Hung, T.-H.; Hsu, C.-H. Modeling the User Experience of Watching 360° Videos with Head-Mounted Displays. *ACM Trans. Multimedia Comput. Commun. Appl.* **2022**, *18*, 1–23. [CrossRef]