

Review

360-Degree Video Streaming: A Survey of the State of the Art

Rabia Shafi ^{1,†}, Wan Shuai ^{1,†} and Muhammad Usman Younus ^{2,*,†}

¹ School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China; rabibabi@mail.nwpu.edu.cn (R.S.); swan@nwpu.edu.cn (W.S.)

² Ecole Doctorale Mathématiques, Informatique, Télécommunications de Toulouse, Université Paul Sabatie, 31062 Toulouse, France

* Correspondence: usman1644@gmail.com; Tel.: +33-753413108

† These authors contributed equally to this work.

Received: 24 July 2020; Accepted: 3 September 2020; Published: 10 September 2020



Abstract: 360-degree video streaming is expected to grow as the next disruptive innovation due to the ultra-high network bandwidth (60–100 Mbps for 6k streaming), ultra-high storage capacity, and ultra-high computation requirements. Video consumers are more interested in the immersive experience instead of conventional broadband televisions. The visible area (known as user's viewport) of the video is displayed through Head-Mounted Display (HMD) with a very high frame rate and high resolution. Delivering the whole 360-degree frames in ultra-high-resolution to the end-user significantly adds pressure to the service providers' overall intention. This paper surveys 360-degree video streaming by focusing on different paradigms from capturing to display. It overviews different projections, compression, and streaming techniques that either incorporate the visual features or spherical characteristics of 360-degree video. Next, the latest ongoing standardization efforts for enhanced degree-of-freedom immersive experience are presented. Furthermore, several 360-degree audio technologies and a wide range of immersive applications are consequently deliberated. Finally, some significant research challenges and implications in the immersive multimedia environment are presented and explained in detail.

Keywords: Virtual Reality (VR); 360-degree video; Tile-based streaming; Audio techniques; Head-Mounted Display (HMD)

1. Introduction

Currently, Virtual Reality (VR) has achieved great significance due to the advancements of computing and display technologies. Filmmakers have already started to think creatively about VR technologies because it is not just a gaming trend that is going to get wider. The healthcare industry, immersive telepresence, telehealth, sports, education, etc. are being rapidly commercialized to meet the demands of the market and consumer expectations, etc. The VR market expects revenue of 108 billion USD until 2021 [1].

As one of the essential VR applications, 360-degree video facilitates the user with an interactive experience that was never thought before. Many commercial broadcasters and video-sharing platforms are showing considerable interest in this domain. Microsoft has released its Windows Mixed [2] Another platform called ARTE360 VR by ARTE [3] enables the sharing and accessing of various omnidirectional videos through mobile applications.

Different 360-degree contents, including Natural Image (NI) and Computer Generated (CG) videos, are well suited to be visualized using the new Head-Mounted Displays (HMDs), like Oculus Rift [4], HTC Vive [5], Samsung Gear VR [6], Google Cardboard [7], among others. These HMDs

equipped with multiple sensors are much more commonly used than conventional display devices to view 360-degree videos. 360-degree video lets the user control viewport via head movements within a spherical video coverage of 360×180 degree [8]. 360-degree videos can also be experienced within the HTML5 environment. In this context, WebVR [9] is a JavaScript API that uses the WebGL API to facilitate the web-based support to obtain a more relaxed VR experience. Within this framework, 360-degree videos need more optimization regarding camera settings, encryption, delivery, and rendering of immersive media content. The web-based 360-degree video rendering on the high-resolution desktop monitors provide no or very little sense of immersion. The latest HMDs, on the other hand, is the most demanding, fully immersive VR systems that offer a compelling experience of realism. Figure 1 shows an equirectangular mapped 2K image where the yaw angle (-180 to 180 degrees) and pitch angle (-90 to $+90$ degrees) are mapped to the x-axis (0 to 1920 pixels) and y-axis (0 to 1080 pixels), respectively. Problems associated with 360-degree video streaming include huge storage requirements, limited Filed of View (FoV) related to the human visual system and display devices, interactivity, smooth user navigation, resource-intensive coding, etc. are considering high-resolution representations [10]. A high-resolution 360-degree video is usually four to five times larger than regular videos [11]. The popular HMDs have a limited FoV, i.e., 100 degrees for Google Cardboard and Samsung Gear VR and 110 degrees for Oculus Rift and HTC Vive, as shown in Figure 2.

The transmission of 360-degree video is rather challenging, especially over the current generation cellular networks because of the limited capacity and dynamic nature. Various 360-degree streaming solutions exist, while one common solution is to project and split an equirectangular frame into several rectangular regions known as tiles, to overcome the bandwidth issue [12,13]. Among such solutions, some only stream a subset of tiles that cover the user's current viewing region of the user [14,15]. Such schemes restrict the user to visualize only a limited part of the video in possible high quality. On the other hand, one may transmit all the tiles of a 360-degree frame in variable qualities [16,17] to compensate for the viewport prediction errors. The streaming of 360-degree video requires higher network bandwidth, as pixels are transmitted to users from every direction. Whether the user may unicast or multicast, the views depend on what type of application will be used. For VR/AR applications, the system uses the user information into an enhanced view (e.g., virtual reality classrooms [18]).

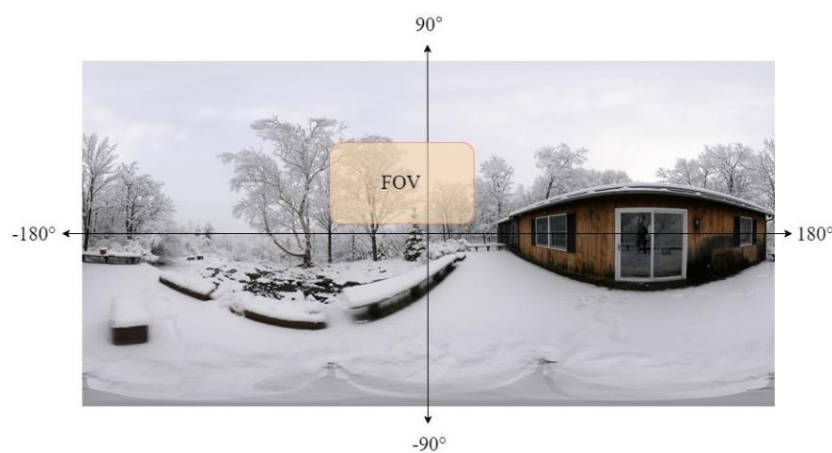


Figure 1. FoV in full 360-degree video frame.

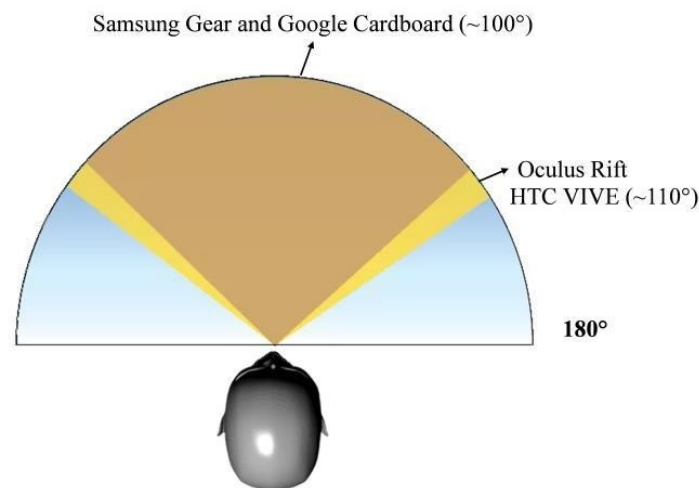


Figure 2. Field of view associated with human eye [19].

Multicast has a unique potential for reducing bandwidth consumption by 360-degree videos. In contrast, the unicast streaming of immersive content uses the network resources and cannot meet each user's demand to wear their HMD for watching the same content. On the other hand, multicast is considered a highly feasible solution because it introduces multiple challenges such as interactivity, ensuring fairness, ensuring smooth quality, etc. Multicast of the 360-degree video has gained importance in the literature so far. A multicast of virtual reality (MVR) [20] has been proposed in LTE networks by considering the adaptive streaming of VR content. This algorithm divides the users by weight of tiles and finds the bitrate for each tile. Similarly, VRCast in [21] was designed for cellular networks by supporting the multicast (e.g., LTE). It solves the complex live streaming issue, divides the 360-degree video into small tiles, maximizes the viewport's quality, and ensures fairness between users. As current works on streaming always transmit panorama pictures in a unicast manner. As a result, viewers only watch a small portion of the video, wasting the extra bits being transmitted. The partial 360-degree video frames are transmitted to a single viewer in [22]. Although an approach was proposed in [23] for the optimization of network bandwidth through multicast that transmits the 360-degree video efficiently. The feasibility of partial multicast frames was presented by reducing the prediction errors that ensure the user quality of experience (QoE) [24]. It is also essential for a seamless experience for end-users. 360-degree videos are complex, requiring fast decoding instances and sophisticated projection schemes that may aid in high overhead. This paper presents and discusses key technologies related to support 360-degree video streaming to enable interactive and immersive experiences. Specifically, the general video streaming system, different streaming approaches, the immersive standardization/project efforts, the latest tools, open software, and the possible challenges and implications are discussed. The main contributions can be gathered into the following:

1. This paper addresses the architecture of 360-degree video streaming. The purpose is to stay as close as possible on 360-degree video principles by considering both low and high-level perspectives. The content pre-processing stages, e.g., content acquisition, stitching, projection, and encoding, are cogitated. Then the transmission and consumption of the 360-degree video are over-viewed.
2. The sophisticated streaming technologies for 360-degree video, including viewport-based, tile-based, and QoE enabled solutions, are presented and discussed in detail. It also describes how high-resolution content is transmitted to single or multiple users.
3. The audio-related technologies that support immersive experience are illustrated.
4. The technological efforts to enable the technologies for an extended degree of freedom in immersive multimedia consumption are explained.

5. Different technical and design-related challenges and implications are presented for the sake of an interactive, immersive, and engaging experience with 360-degree video.
6. The potential of 360-degree video and guidelines for readers approaching research on 360-degree video streaming are presented.

The paper’s structure is as follows: Section 2 provides an overview of a 360-degree video streaming system. Section 3 outlines the major streaming approaches for 360-degree video. Section 4 briefs some technical issues of spatial audio for 360 media. Section 5 explains various technological efforts that aim to bring the virtual world close to the real world. Section 6 signifies the potential growth of 360-degree video based on applications. Section 7 provides some technical challenges and implications to create an immersive, interactive, and engaging experience with 360-degree video. Lastly, Section 8 presents the discussion and conclusion. The schematic map of the paper is shown in Figure 3.

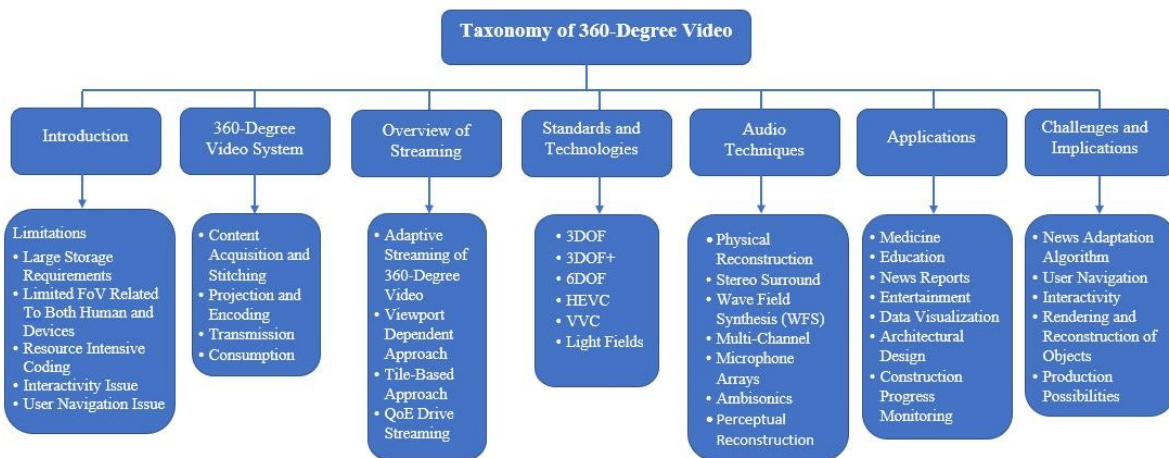


Figure 3. Structure of the paper.

2. 360-Degree Video Streaming System

The concept of streaming media has gained significant attention because of its advancements in video compression technologies. The industry and academia are trying to come up with multimedia streaming solutions. However, supporting 360-degree video streaming in real life remains challenging. Such real-time demands are the key differentiators between multimedia and other data network traffic that need special attention. Figure 4 describes an ecosystem for 360-degree video streaming principles. Each step from acquisition to consumption by the end-user is briefly described here.

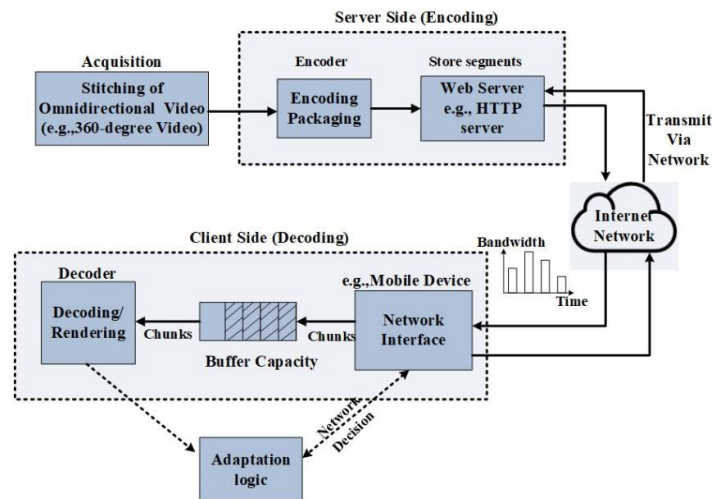


Figure 4. 360-degree video streaming framework.

2.1. Content Acquisition and Stitching

Several omnidirectional cameras [25], such as the Gear 360, the Ricoh Theta, and the Orah cameras, etc., are equipped with multiple sensors to capture a full 360-degree scene. Recently, some stereoscopic omnidirectional camera systems such as Jump [26] and Facebook surrounds 360 capture the stereo views in all directions. However, stereoscopic omnidirectional cameras' capturing of dynamic scenes is very challenging to build a professional capturing system because of self-occlusion problems among the cameras. As in the automatic image stitching process, the different types of planar models (e.g., affine, cubic transformation models) are used to align the different views from the camera, thus blending and distorting the views to a surface of a sphere [27]. However, in video stitching, video stabilization, and video synchronization are essential for moving cameras and individual sensors, respectively [28].

The stitching of video follows a seamless 360-degree representation (e.g., planar representation) either in real time or offline before mapping and encoding [29]. The reuse of the existing image and video content distribution is allowed for a planar representation and involves encoding, packaging, and transmission steps. The acquisition may add some serious visual distortions. As the lack of synchronization between the cameras can become a cause of motion discontinuities and automatically impacts the overall 360-degree video streaming framework. Moreover, the capturing of omnidirectional stereoscopic 3D content added up to its challenges [30,31]. However, if the camera shares the same projection center, multiple views can be synchronized together using the planar transformation models. Moreover, the keystone (that is the result of a converging position of two cameras to slightly different planes) and cardboard effect (unnatural flattening of objects) can occur.

2.2. Projection and Encoding

After the content acquisition and stitching through advanced tools, the 360-degree sphere is projected to a 2D planar format for effective coding and transmission over bandwidth constraints networks. 360-degree video compression can take advantage of different projection approaches to determine the better compression and coding processes. A straightforward solution is provided by using equirectangular projection (ERP). Several 360-degree video streaming services use the ERP format, such as YouTube and Jaunt VR. The most common example of ERP is the world map. It can be defined as flattening a sphere on to a 2D surface around the viewer [32]. Nevertheless, ERP is not considered the most efficient representation of a sphere. One of its main drawbacks is that the significant network bandwidth is wasted due to the expensive encoding of less interesting regions. Alternatively, the other planar representations (e.g., cubemap (CMP), octahedron, etc.) are proposed to address the problems of ERP [33]. Among these, CMP is the most common and well-known in graphics frameworks (e.g., Open GL). In this projection, a combination of the cube's six faces is used to map the pixels on the sphere to corresponding pixels on a cube. VR gaming applications widely use it. This technique saves space and reduces the video file size by 25% compared to similar user-perceived quality in ERP format [34]. A significant disadvantage of the cubemap technique is that the rendering of a limited user's FoV is smaller than the encoded 360-degree image.

Based on CMP, a Modified Cubemap Projection (MCP), also known as Hybrid Equirectangular projection (HEC) [35], was proposed to achieve the improvements in coding efficiency. It offers a highly efficient representation of 360-degree video by combining the mapping functions of the outerra spherical cubemap (OSC) [36] and equi-angular cubemap (EAC) [36]. Other different projection formats such as Hybrid Cubemap Projection (HCP) [37] and Hybrid angular cubemap projection (HAC) [38] were proposed to improve the coding performance.

A pyramid projection projects a sphere onto a pyramid based on the user's current viewing area. It is proposed by Facebook to support variable quality mappings [39]. It mainly faces two issues, including; (1) the users rotate their heads by 120 degrees. Therefore, when they turn their heads to the back of the pyramid, the quality drops by the same amount, (2) it is not supported on GPUs and is not as effective as the cubemap for rendering. Offset cubemap projection is a regular cubemap, which provides a variable mapping while solving the pyramid projection problems.

It facilitates a smoother degradation of quality than pyramid projection. The main disadvantage of the offset cubemap projection is the expensive storage requirements. Figure 5 depicts different projection approaches for 360-degree video, and Table 1 represents a summary of different projection, coding, and streaming schemes for 360-degree video streaming. Coding efficiency is a significant factor in assessing video compression. There is still a need for more effective video compression techniques to stream panorama, ultra HD (UHD), and 360-degree video content. High-Efficiency Video Coding (HEVC) is currently the latest standard implemented worldwide, standardized by the Joint Collaborative Team on Video Coding (JCT-VC). ISO/IEC Moving Picture Expert Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) have developed HEVC to explore future video coding (FVC) [40] technologies.

The Joint Exploration Test Model integrates the most promising future technologies. MPEG has issued the needs for the functionalities of FVC in 2016. Moreover, many companies in parallel have developed their video coding formats (e.g., VP8 [41], VP9 [42], and VP10 [43]). The novel Versatile Video Coding (VVC) [44] standard is defined in MPEG-I Part3. It supports three types of videos called High Dynamic Range (HDR), 360-degree VR-oriented video, and Standard Dynamic Range (SDR). The target compression performance of VVC is 30-50 % better than HEVC at the same subjective video quality. Table 2 provides a comparative analysis of the HEVC and VVC video codecs. Although Alliance for Open Media [45] was developed for next-generation media formats, the first version of the AOM video format, namely AV1, provides a substantial coding efficiency over the most advanced video codecs available today. Researchers have employed different content preparation schemes to achieve higher coding efficiency for 360-degree video. In [46], a region-based adaptive smoothing scheme is proposed to improve the perceived quality with different encoding settings. It is applied to equirectangular mapping because we get a high number of pixels at the top and bottom in such 2D projected images. Another attempt [47] to encode high-quality region-of-interest (ROI) is to use Scalable Video Coding (SVC). This approach mainly optimizes the user experience by using a layered-based streaming framework to minimize transmission delay, enhancing ROI quality, and evade rebuffering. The authors in [48] present an in-network bitrate adaptation strategy for SVC video streaming over Long-term Evolution (LTE).

The spherical-to-plane projection [49] may result in geometrical distortions and discontinuous regions and may also involve interpolations and resampling, leading to the aliasing and blurring distortions in the signal. Since the conventional 2D video compression schemes are used for the planar representation. It involves the same compressions issues as discussed in Table 3, because of the quantization errors [50]. Moreover, the planar representation of omnidirectional video content means that vectors cannot be predicted because the motion is no longer planar at some parts. Hence, the intra-prediction and motion model [51] is not optimal in regions (e.g., ERP poles and CMP discontinuities), which can lead to bitrates and compression problems.

Table 1. Summary of Projection, Coding, and Streaming Techniques.

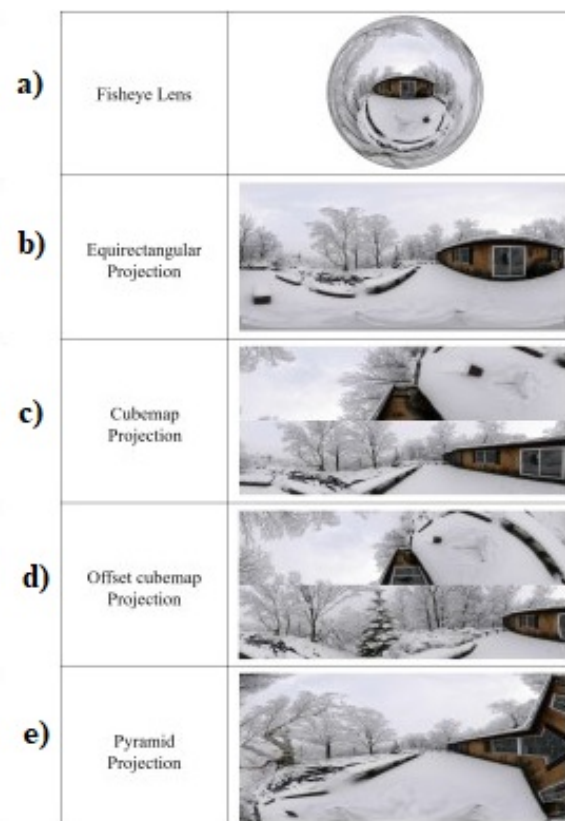
Reference	Year	Projection	Description	Scheme
[52]	2017	CMP	It handles the irregular motion for the cubic map projection of 360-degree video by projecting the pixels in both reference and current picture from unblocking cube back to the sphere.	Projection-based advanced motion model
[53]	2018	ERP	It compresses the motion information of omnidirectional content efficiently by applying a scaling scheme based on the location of the video to facilitate a uniform motion behavior.	Geometry-based MV scaling method
[54]	2019	EPR	It considers the motion of each coding block as 3D translation in the spherical domain with 2D MV to improve coding efficiency.	Spherical Coordinates Transform-based motion model (SCTMM)
[55]	2017	CMP	It performs spherical rotation of the input video prior to HEVC/JEM encoding that improves the coding efficiency.	SR-HEVC/JEM encoding
[56]	2018	CMP	It predicts the sphere rotation to yield the maximal compression rate. A convolutional neural network learns the association between the compressibility at a different rotation of CMP and its visual content.	Learning-based approach
[35]	2019	HEC	This scheme achieves more information sampling and reduces the boundary artifacts by presenting a hybrid equi-angular cubemap projection. It offers better improvements in coding efficiency.	Hybrid Equi-angular CMP scheme
[38]	2018	HAC	The proposed coding scheme improves the coding efficiency by keeping the sampling continuity at borders. It achieves the significantly higher coding efficiency of 33.9% and 13.5% BD-rate than JEM and HM.	Hybrid angular CMP projection
[57]	2016	Multiple	This streaming approach transmits the front viewport with high resolution while other parts are transmitted with low resolution. It also offers an HM-tracking algorithm to compress the real-time 360-degree video.	Viewport-adaptive streaming system
[58]	2019	ERP	It maximizes the rendered video quality while keeping the streaming continuity against the network bandwidth.	Two-tier system streaming system
[59]	2017	Offset-cubic	It performs a comparison between the quality level adaptation and view orientation adaptation for 360-degree video. It saves the bitrate from 5.6% to 16.4%.	Oculus HMD-based viewport adaptive streaming
[1]	2017	Multiple	This type of streaming has several advantages such as; reduces the bitrate, highest quality display in case user do not move, still better quality if the user move.	Viewport-adaptive streaming

Table 2. Comparison of HEVC and VVC.

Category	High-Efficiency Video Coding (HEVC)	Versatile Video Coding (VVC)
Deployment	2014	2020+
Partition	8×8 to 64×64 (coding unit)	Maximum size of 28×128 , Chroma Separate Tree (CST)
Licensing	Complex	Not Defined
Live Deployment Environment	Appliance SW cloud	Cloud
Reference Model	HM	JEM
Inter-Prediction	Weighted Prediction	Affine Motion Compensation (AFF), MV Prediction
Intra-Prediction	35 Predictions	65 Directional Predictions
Quantization	Fixed Quantization	Dependent Quantization (DQ)
Filters	De-blocking Filters (DF), Sample Adaptive Offset (SAO)	Adaptive Loop Filters (ALF), SAO, DF
Transform Coding	Square IDCT from 32×32 to 4×4 +DST luma Intra 4×4	Square and multiple Transform (up to 64×64 size), Multiple Transform Set (MTS)
Supportive Media	Full-HD mobile, UHD Broadcasting	VR 360, Point Cloud, Light Field, High Dynamic Range

Table 3. Compression characteristics and artifacts for the traditional video.

Artifact	Properties
Blocking	It occurs by the coarse quantization of low-detail regions.
Blurring	It occurs because of the loss of spatial details when high-frequency components are quantized to zero.
Ringing	It seems as “halos” (ripple structure) around strong edges.
Pattern	The incapability of basic functions (horizontal/vertical) such as building blocking of DCT for the representations of diagonal edges.
Flickering	It leads to frequent variations in luminance or chrominance with temporal dimensions.
Floating	It appears because of illusion motion in certain regions with the background.

**Figure 5.** Video mapping approaches. (a) Fisheye Lens (b) Equirectangular (c) Cubemap (d) Offset-cubemap (e) Pyramid.

2.3. Transmission

Many technical challenges on content distribution are presented by 360-degree video because of low latency and high video rate requirements of omnidirectional signals [60]. The same traditional video frame transmission protocols are used for 360-degree delivery. In the packaging step, data is packaged using a state-of-the-art streaming framework such as Dynamic Adaptive Streaming over HTTP (DASH) to facilitate over-the-top (OTT) services. MPEG-DASH-based content delivery solution is the most prominent once for 360-degree videos [61–64], because it exploits the existing delivery architecture without many extensions. Table 4 describes the resolution requirements of commercially deployed 360-degree video services. The major streaming technologies, such as viewport-based streaming [57] and tile-based streaming [65], use the DASH framework to manage the viewing region’s quality according to the network conditions available. The adaptive streaming of omnidirectional content may suffer from transmission losses and can degrade the user experience. The typical DASH distortions (e.g., quality loss on fast head movements, spatial quality variance, buffering, delay, etc.)

can strongly lower the viewing experience in viewport-based streaming compared to viewport-agnostic streaming. They can impact the QoE of immersive applications and are still mainly overlooked when the compressed content is viewed by an HMD [57].

2.4. Rendering and Displaying

The inverse steps (such as decoding, unpacking, converting to display geometry, and rendering) are performed before a user can interact with 360-degree content. The inverse mapping from a plane to sphere visualizes the rendered content typically implemented on HMD, monitor, tablet, or smartphone. The consumption of 360-degree content on the latest HMDs with advanced display features can still introduce new distortions (e.g., aliasing, blurring, etc.) that are needed to be resolved. Finally, the distortions related to stereoscopic displays are still present in HMD and can produce several issues such as misperception of the display, speed of objects, etc. Table 5 defines the different types of capturing to displays issues and distortions in 360-degree video.

Table 4. Commercially installed 360-degree video services.

Function	Resolution	Assessment
Capturing, Stitching, Encoding	4K × 4K	Easy implementation
Storage and Delivery	4K × 2K	High bitrate, not sensitive to network delay
Display	1K × 1K	Low video quality

Table 5. Different types of issues and distortions in 360-degree video.

Content	Issues and Distortions
Capturing and Stitching	Omnidirectional recording systems consist of multiple cameras that subject to common optical distortions (e.g., chromatic aberrations, noise, motion blur, etc.). Some issues may also occur due to the inconsistencies between cameras. The stitching process also creates some spatial distortions (e.g., blurred circle, and missing object parts), temporal distortions (e.g., blurred motion, geometrical distortions, etc.), and stereoscopy (e.g., keystone) distortions.
Projection and Encoding	Projecting a sphere to a plane is a common problem in map projections and it adds some discontinuities and geometrical distortions that may result in aliasing, blurring, and ringing. Compression also becomes a cause of blocking, blurring, the spatial pattern changes, temporal changes (that includes floating, jerkiness, and flickering, etc.), and cardboard effect in stereoscopy.
Transmission	Transmission of rich media content over highly dynamic network channels can badly affect the user experience levels due to the channel distortions (such as tiling artifacts and spatial quality distortions) and temporal discontinuities (e.g., viewport deviation, stalling, and temporal quality variance, etc.).
Display	Traditional displays related artifacts (e.g., aliasing, motion blur, etc.) may also affect HMD displays. Some display limitations (such as crosstalk as inter-perspective aliasing and motion-to-photon delay) in stereoscopic and HMD are also included.

3. Overview of Video Streaming

Academic and industrial research growth has made a great deal of effort by concentrating on coming up with solutions to stream the multimedia. Real-time requirements are important changes, which need special attention, between multimedia and other data network traffic. A lot of standardization organizations and protocols were obtained to enable multimedia streaming, and Quality-of-Service (QoS)-based streaming. Some early protocols built on top of the Internet Protocol (IP) were Integrated Services (IntServ) [66] and Differentiated Services (DiffServ) [67] that will ensure QoS-aware streaming and multimedia streaming. Resources are specifically stored by executing a signaling protocol [68] called Resource Reservation Protocol (RSVP) to satisfy the application specifications. This protocol is used by IntServ that defines the QoS requirements for an application's traffic. Real-time Transport Protocol (RTP) [69] enables end-to-end multimedia streaming in real time by proposing a standardized packet format. Application-layer framing and integrated layer processing are two main concepts that are used to design RTP. A client-server-based connection is established by Real-time Streaming Protocol (RTSP) [70] by controlling the multimedia servers before downloading the required video content. Some researchers in [71] have shown that Transmission

Control Protocol (TCP) is beneficial for transmitting the delay-tolerant-based videos. In return for its reliability, it permits efficient data transmission but must suffer from unpredictable delays. HTTP's design over TCP ensures progressive downloading to download a video file of constant quality as quickly as TCP enables it. A major downside is that clients receive the same video content under different network conditions, which can lead to unnecessary stalls or rebuffering. This situation has led researchers to turn to the development of HTTP adaptive streams. 360-degree video streaming is different from traditional video streaming. This section presents a detailed overview of existing 360-degree video streaming solutions, followed by a summary of existing solutions.

3.1. Adaptive 360-Degree Video Streaming

360-degree video streaming has gained vast importance in the multimedia world over the years. Implementing adaptive streaming in a VR environment for 360-degree video content is difficult because it needs smart streaming and encoding techniques to deal with present and future services as well as applications. The video compression standard exploits information theory that provides source coding and characteristics of the human visual system to minimize spatial-temporal redundancies. Three essential aspects of video coding, visual perception, and quality assessment have focused on the research of perceptual compression of 360-degree video. Furthermore, a user can randomly switch to neighboring views during 360-degree video playback. The actual challenge is to facilitate a smooth viewport switching by providing a certain level of resilience to errors to eliminate error propagation due to different encoding frames. Thus, mostly the viewport-based streaming strategies save resources while transmitting video streams.

3.2. Viewport-Based Streaming

Viewport-based adaptive streaming has gained attention in both industrial and academic communities. The end-users' corresponding viewport can be identified in viewport-dependent streaming based on the user's head movement. Therefore, such solutions are adaptive during the streaming of 360-degree videos, as they dynamically select regions and adjust the quality to minimize the transmitted bitrate. It provides several adaptation sets at the server-side because it is a viable option to smooth the viewport during abrupt head movements. Each adaptation set contains the associated video area with a given viewing orientation. In [57], the authors proposed a differentiated quality approach where the front viewport is transmitted with relatively high resolution compared to the other parts. They compared ERP and CMP multi-resolution variants with the current pyramid projection variants. Similarly, a viewport-adaptive 360-degree video streaming in [1] is suggested to reduce the bandwidth. The concept of quality-focused regions (QERs) was introduced, making a particular region of higher quality video than the rest of the video. However, streaming approaches do not involve quality adjustment based on head movement prediction errors. The authors in [72] evaluated the impacts of response delay based on viewport-based adaptive streaming. The system provides a server-based quality adjustment and view transmission to reduce the response delay by estimating the throughput and viewport signals. Based on the client's response, network throughput is estimated by the proposed framework for the future viewing position. On the server-side, the necessary tiles are then streamed to satisfy the delay constraints. The viewport-dependent adaptive streaming is based primarily on small adaptation and buffering delays indicated in real-world experiments. The initial results illustrate this type of streaming is effective in case of short response delay. The n th interval of estimated viewport $V^e(n)$ and the n th adaptation interval of estimated throughput $T^e(n)$ is given below:

$$V^e(n) = V_{fb} \quad (1)$$

$$T^e(n) = T_{fb} \quad (2)$$

where the V_{fb} and T_{fb} are the last reported viewport position and throughput, respectively. The bitrate computation R_{bits} based on $T^e(n)$ is given as:

$$R_{bits} = (1 - \beta)T^e(n) \quad (3)$$

where β is the safety margin. In [73], a joint adaptation was observed based on network and buffer delay. The proposed framework dynamically adjusts the viewport area to visualize the high likelihood of the scene at the time of rendering. It has shown that the proposed design provides flexible adaptation support to consume the available bandwidth efficiently. A navigation-aware optimization problem is studied in [74] to reduce both view switching delay and navigation distortions. An optimal solution is provided to polynomial time complexity through a dynamic algorithm. The k th frame of quality objective $VQ(k)$ is computed, such as:

$$VQ(k) = \sum_{t=1}^{T_n} w_k(k) * D_t^n(v_t, k) \quad (4)$$

The weight w_k in Equation (4) shows that at k th frame how much the tile t overlaps the viewport. The weight w_k is computed as:

$$w_k = \frac{A(t, k)}{A_{vp}} \quad (5)$$

where $A(t, k)$ is the overlapped area of title t and A_{vp} is the total area of the viewport. The quality objective VQ is computed as follows:

$$VQ = \frac{1}{L} \sum_{k=1}^L VQ(k) \quad (6)$$

Oculus' viewport-based streaming implementation is performed in [59], indicating that this implementation is inefficient: 20% of the bandwidth is lost downloading video segments that have never been used. The asymmetric viewport-based technique streams 360-degree content with different spatial resolutions to save bandwidth. During video playback, the client requests a version of the video based on the user's orientation. This approach's advantage is that even if the client incorrectly anticipates the user's orientation, the low-quality content can still be made in user viewport. However, such a scheme involves huge storage and processing overheads in most cases. Viewport-independent streaming is the straightforward way that streams the content of 360-degree video since the entire frame is transmitted in the same quality as conventional videos. Simplicity implementation has been an appealing gateway to viewport-independent streaming. Though, the coding efficiency is 30% lower than viewport-dependent streaming [75]. Additionally, invisible areas require a lot of bandwidth and decoding resources. This form of streaming [76,77] mainly applies to content streaming.

3.3. Tile-Based Streaming

In tile-based streaming, a video is divided temporally into segments as in traditional HTTP-based adaptive streaming. Moreover, these video segments are spatially divided into tiles, so that several spatial tiles compose each temporal segment. Since the client needs to store some amount of video to ensure continuous playback so it pre-fetches video segments based on viewport prediction. As an earlier work, [78] performed tile-based coding that tries to adjust the resolution based on the user's viewport. The video tiles are encoded with two different levels of resolution. The frame reconstruction process integrates high and low-resolution tiles within and outside of the viewport, respectively. A study [79] explore the various tiling features by investigating 360-degree video steaming where each tile can be projected for quality adaptation based on different viewing regions. Moreover, the full delivery of basic streaming can save about 65% compared to full and partial delivery strategies.

An equirectangular video in [80] is partitioned into many tiles where a sampling weight is assigned to each horizontal tiles based on its content. The bitrate allocation is optimized based on sampling weight and bandwidth budget. An overlapping margin with two neighboring tiles is added to overcome the probability of viewport missing by applying an alpha blending on overlapping tile margins. In recent approaches [81–83], each tile has multiple types of hierarchical representation to choose from, based on the user’s viewport. As a result, smoother quality degradation can be obtained. By using SVC, they surmount the randomness of both the network channels and head movements. The authors in [81] use the visual attention metric that calculates the tiling patterns by introducing an adaptive-based streaming framework. Based on this metric, tiles are generated in different sizes to retain the advantages of larger tiles and smaller tiles with high coding efficiency and streaming decisions, respectively. The bitrate allocation strategy is assigned to the tiles belonging to different regions for optimal streaming for each selected pattern. The authors in [84] presented an optimization framework that tries to minimize the pre-fetched tiles error. It also ensures continuous playback within a small buffer and builds a probabilistic model that predicts the viewport. The SRD extension of DASH achieves a higher bitrate and thus we can stream the videos to users with the highest quality. In addition, the motion-constrained HEVC tiles in [85] minimize the complexity and synchronization problems between tiles such that a single decoder can be used. The three types of heuristic strategies are also presented for 360-degree video streaming. The experimental results indicate that the better coding efficiency has achieved by streaming the viewport tiles at the original captured resolution. The authors in [65] designed an end-to-end VR video streaming to transmit 8K 360-degree videos. The proposed methodology assigns higher bitrates to the viewport tiles and gradually lower quality to the tiles that are outside of the viewport. The bitrate assigned to the k th tile in the viewport is given as:

$$R_{VP_k} = (\gamma BW_{curr})w_k, \text{ where } k \in V, \&k \notin S^{out} \quad (7)$$

where V and S_{out} represent the viewport and a set of tiles outside of the viewport, respectively. γ is a constant that is defined by the client. BW_{curr} is the currently available bandwidth and w_k is the weight of the k th tile. The bitrate estimation for the k th tile in a set of tiles outside of the viewport $R_{S_k^{out}}$ is calculated as:

$$R_{S_k^{out}} = K_k((1 - \gamma)BW_{curr}), \text{ where } k \in S^{out}, \&k \notin V \quad (8)$$

Finally, for each tile representation, the client requests a bitrate which is represented as:

$$m \Leftarrow \min|r_m - R_k|, \text{ where } k \in (S_{in} \cup S_{out}) \quad (9)$$

where S_{in} is a set of the tiles inside the viewport and m is the DASH representation ID. The researchers in [58] considered fetching unviewed part of the video at the lowest quality based on user head movement prediction as well as to decide the video playback quality adaptively for the viewed part based on bandwidth prediction. A two-tier system for 360-degree video streaming has proposed in [86], where the entire video content is delivered by base tier at a lower quality with a long buffer. In contrast, the enhancement tier facilitates the predicted viewport with a short buffer at a higher quality. Consequently, a tradeoff between reliability and efficiency is achieved for 360-degree video streaming.

In [86], the authors predicted the head movement (HM) of the viewer from his/her previous HM data, considering both angular velocity and angular acceleration. According to the predicted HM, a different quantization parameter (QP) is allocated to each tile. The experimental evaluation showed that angular velocity and angular acceleration-based HM prediction significantly reduces the prediction error and introduced low delay and the associated loss in visual fidelity compared to baseline approaches. A very similar solution but with HTTP/2 feature is presented in [87] to overcome the bandwidth and request overheads. HTTP/2’s priority features enable priority transmission and stream termination features to enhance the user experience. Unlike the prosperity of above-mentioned viewport-based coding, the saliency-aware compression is still a challenge because

the existing 2D saliency approaches are difficult to employ for 360-degree video. A work in [88] proposes saliency-based sampling for a 360-degree video system, where low-resolution CMP is combined with unsampled salient regions. Spatial Relationship Description (SRD) feature extends the Media Presentation Description (MPD) that enables the DASH client to retrieve only certain user-relevant video streams at high resolution. The authors in [83,89] employed the MPEG-DASH SRD [90] extensions to support tiled streaming and described a video as an exclusive collection of synchronized video. They also present several SRD use cases (e.g., zoomable) where the users are provided with a seamless experience. SRD facilitates the spatial positions of content, and thus DASH clients can determine which tiles have to request. The users always download low-resolution tiles to avoid rebuffering while the current view region is presented to support a high-quality zooming feature. Table 6 represents a summary of different streaming schemes for 360-degree video streaming.

Table 6. Comparison of Different Approaches of 360-Degree Video Streaming.

Technique	Projection	Dataset	Duration	Metric	Distortion
Tile-based streaming	It streams the 360-degree video rectangular tiles in same or different qualities.	Medium	<20 ms	ERP, CMP, TSP, Offset Cubemap	Medium
Viewport-Dependent Streaming	It performs adaptation based on the network characteristics and the viewing orientation of the user.	Medium	<20 ms	Pyramid, TSP, Offset Cubemap	High
Viewport-Independent Streaming	It streams the 360-degree video content in equal quality.	High	High latency acceptable	ERP, CMP	Low

3.4. Quality of Experience Enabled Streaming

Multimedia streaming has gained considerable popularity among users everywhere, as there are many performance problems while delivering multimedia over different loaded networks. Even more so as the processing and transmission of 360-degree format bring along new challenges (i.e., bandwidth, distortions, etc.). To lower the bandwidth requirements, video material must be compressed to lower qualities, causing compression artifacts that may negatively affect the user's quality of experience (QoE). QoE refers to the measure of customer satisfaction and experience from a service such as TV broadcasts, phone calls, and web browsing. As with traditional 2D videos, quality assessment of 360-degree videos can be done through both subjective and objective tests, which have their advantages and disadvantages.

3.4.1. Subjective Quality Assessment

Many subjective video quality assessment methods have been found for 2D videos from the past two decades. Many subjective methodologies have been proposed by the international telecommunication union (ITU). Two metrics are widely used in subjective assessment quality, such as one metric is MOS [91], and another metric is DMOS [92]. Currently, different types of subjective assessment methods are identified for omnidirectional videos. The authors in [20] presented a testbed on omnidirectional video and image by suggesting an HMD as the displaying device. But, unfortunately, this study does not consider how to measure the subjective quality of 360-degree videos. Based on the testbed proposed in [20], a dual stimulus method in [93] has been used to measure the quality of High Dynamic Range (HDR) omnidirectional images. In contrast, authors in [94] present a single stimulus ACR-based study for omnidirectional images. It has been found that the ideal viewing duration for 360-degree images is 20 seconds to explore the content entirely by the user. Moreover, different people might explore the content differently, looking at other parts, resulting in different experiences. Therefore, visual attention and salience are important aspects to consider in the subjective assessment of 360-degree video content [95]. The authors in [96,97] elaborated on the subjective study by considering several parameters (i.e., resolution, bitrate, quantization parameter QP, content characteristics) and their effect on perceptual quality 360-degree video. A study by [98] was also conducted towards QoE of 360-degree video streaming that mainly focuses on the impact

of stalling. They performed subjective research in their lab, where they compared different stalling frequencies and duration and additionally compared results for the 360-degree video to traditional TV. Another study [99] on the QoE of streaming 360-degree videos found that delay, quality variations, and interruptions could support the evaluation of the QoE these factors into their model, indicating these factors do influence the quality perception. There is still a lack of standardized methodologies for subjective studies and metrics for 360-degree video. The debate on how to develop these is ongoing; consensus has not yet been reached within the research community. Nevertheless, some studies on the subjective experience of omnidirectional content have been performed adapting methodologies from classical video quality assessment. However, this adaptation is not trivial as viewing through an HMD is substantially different from a regular display that presents different experiences. The viewer is more immersed in the content, and challenges regarding strain and cybersickness arise. Cybersickness is a potential barrier to achieve higher QoE levels and can cause discomfort. In [100], two subjective experiments have conducted to evaluate the video perception level and cybersickness in viewport adaptive 360-degree video streaming with limited bandwidth and resolution. Also, a modified absolute category rating (M-ACR) method was proposed by using different devices [101,102] to analyze the cybersickness of 360-degree videos at varying conditions of bitrates and resolutions. Table 7 depicts the comparison of different subjective quality assessment approaches.

Table 7. Comparison of different subjective quality assessment approaches.

Reference	Dataset	Method/ Recommendation	Device	QoE Aspect	Duration	Published Year
[103]	6 images	ACR-HR	HMD mount (MergeVR)	Perception	30 s	2016
[94]	12 images	ACR Single Stimulus	HMD, HTC Vive	Spatial resolution and distortion	20 s	2018
[98]	2 videos	ITU-R BT.500-13	Static/Move VR	Perception, Presence	60 s	2017
[104]	48 users	ACR-HR	Samsung Galaxy S8, Mirage Solo	Usability, HMD influence, Fatigue	40 min	2020
[96]	60 videos	-	Samsung Gear VR, Samsung Galaxy S6	Cyber sickness Presence	30 s	2017
[102]	6 videos	Modified ACR	HMD Oculus Rift	Motion Sickness	10 s	2017
[101]	12 videos	ACR	Oculus Rift	Cyber sickness	60–65 s	2017

3.4.2. Objective Quality Assessment

Currently, objective quality in 360-degree video is measured in the planar projection through structure similarity (SSIM) and standard peak-signal-to noise ratio (PSNR). However, they give similar importance to all parts of the spherical image, even though different parts have different viewing probabilities and thus different importance. Additionally, they still do not give a good representation of subjective quality. Viewport-based PSNR or SSIM metrics could be a solution closer to what the users perceive. However, all objective metrics still fail to consider perceptual artifacts such as, for example, visible seams [95]. In a study by [105], three metrics especially designed for omnidirectional content, were compared to conventional 2D metrics. They evaluated the spherical PSNR (S-PSNR), weighted spherical PSNR (WS-PSNR), and crater parabolic projection PSNR (CPP-PSNR). The results showed only moderate correlations with the subjective scores. Compared to traditional methods, the metrics developed for 360-degree video content did not work better. This was confirmed once more by studies on various quality metrics by [106]. They considered 10 quality metrics in their study. Their results show a better correlation to the subjective metrics. Moreover, they showed as well that traditional PSNR outperformed the other metrics due to its simplicity. The data from subjective methods are used as ground truth, and the goal is to predict the quality scores (MOS) as close as possible through objective data about the video [107]. Several objective quality video assessment approaches advance the metric of PSNR. Hence, PSNR cannot represent subjective visual quality since human experience is not taken into account. For example, in region-of-interest (RoI), subjective quality is more likely to be affected by PSNR. The study [108] considered the multi-level quality features and fusion model where the quality features are compared with RoI maps. These multiple quality features are then

combined by a fusion model to obtain the overall quality score. Another study [109] introduced S-PSNR, where PSNR is calculated based on uniformly sampled points. S-PSNR can generate the objective quality assessment of 360-degree videos by applying interpolation algorithms under various projections. In contrast, a weighted PSNR (W-PSNR) [110] is proposed by using gamma-corrected pixel values. A Craster parabolic projection PSNR (CPP-PSNR) compares the different projection approaches by remapping pixels to CPP projection. SSIM is another quality evaluation metric to define multi-factor image distortion. The author in [111] analyzed the SSIM results and introduced a spherical-SSIM (S-SSIM) metric to compare the similarity of impaired and original 360-degree videos. The overview of different approaches to quality evaluation is given in the following Table 8.

Table 8. Summary of various approaches of objective quality assessment.

Reference	Projection	Dataset	Duration	Metric	Distortion
[108]	ERP	16 videos	10 s	Pixel-wise quality metric	6 bitrates
[112]	ERP, CMP, Dyadic	10 videos	10s	S-PSNR	4 QP
[111]	ERP	8 videos	10 s	S-SSIM	QP (22,27,32,37,42)
[96]	ERP	3 videos	30 s	S-PSNR, WS-PSNR, CPP-PSNR	QP (22,28,32,36,40)
[105]	CMP	4 images	-	S-PSNR, WS-PSNR, CPP-PSNR	Bitrates (0.25, 0.50, 0.75, 1.00)

Machine Learning (ML) can bridge the gap between streaming approaches through objective and subjective QoE assessments. Reinforcement learning (RL) methods are used for video streaming bitrates to improve the QoE. Table 9 provides a summary of different works in video streaming applications to define RL to improve QoE. In [113], a method was investigated to adapt the variable video streaming. A two-stage model [109] was proposed for QoE assessment. The research in [114] aims to address the issue of quality variation that affects the QoE. A DRL model [115] considered both eye and head movements data for the quality assessment of 360-degree video. The author in [116] proposed a Q-learning algorithm for adaptive streaming services to improve the QoE in variable environments.

Table 9. RL-based approaches to improve the QoE.

Reference	Method	Aspect	Video Streams	Experiment Setup
[113]	MDP-RL	QoE model	Variable bitrate (VBR)	Simulation (MATLAB)
[114]	Post-decision state	SSIM	Constant bitrate (CBR)	Simulation (MATLAB)
[116]	Q-Learning RL	bitrate	CBR	Lab test
[112]	Two-stage RL model	QoE model	Adaptive VR Streaming	Linux Ubuntu 14.04 operating system
[115]	DRL model	QoE model	Improved video quality	HTC Vive

In summary, QoE research is important for the development of video streaming technology to most efficiently handle the tradeoff between providing good quality and limiting network burden. Additionally, the 360-degree video offers a substantially different experience compared to regular 2D. Therefore, it would be prudent to do more research on the QoE in the 360-degree video specifically.

4. Audio Technologies for 360-Degree Video

360-degree and panoramic videos can be break or make by an audio effect. Spatial audio [117,118] is known as a full-sphere surround sound approach that employs multiple audio channels to mimic the audio representation that we have in real life. The 360-degree video becomes more reliable due to the spatial audio because of the channeling properties of sound that enable it to pass through time and space. The Google VR Software Development Kit (SDK) [119] optimizes the audio rendering engine for the mobile VR. The significance of the 360-degree video display system cannot be overstated in producing the spatial audio soundtrack. The Facebook spatial Workstation [120] has the templates

and numerous plugins that are used to support the synchronized audio playback for 360-degree video with the help of HMDs (e.g., albeit solely for OSX, Oculus Rift, etc.). The other audio production environment will be integrated with such type of video monitoring.

Two categories have been described for the reproduction techniques of spatial audio named physical reconstruction and perceptual reconstruction. The physical reconstruction technique is used to synthesize the whole sound field as close as possible to the desired signal. In contrast, the psychoacoustic techniques are used in perceptual reconstruction to produce a perception for the spatial sound characteristics [121]. The stereo configuration uses the two speakers in the most popular methods of sound reproduction to facilitate the more spatial information (that includes distance, direction sense, ambiance, and sound stage ensemble). While Multi-channel reproduction methods [122] are used in the acoustic environment and become popular in consumer devices.

A study in [123] provides multi-channel reproduction techniques. The same acoustical pressure field is also produced with the other physical reconstruction techniques, as called Ambisonics and Wave Field Synthesis (WFS), as existing in the surroundings. An array of a microphone is needed to capture the more spatial sound field. Consequently, the microphone recordings demand the post-processing because they cannot be used directly without processing for the analysis of the sound field characteristics. Microphone arrays are used in speech enhancement, source separation, echo cancellation, and sound reproduction.

Ambisonics [124], also known as 3D audio, is used to record, mix, and playing the 360-degree audio around a center point. Recently, it has been adopted in the VR industry and 360-degree applications but was investigated in the 1970s and never used before. Ambisonics audio is not like traditional surround technologies. The principle behind the two-channel stereo and traditional sound technologies is the same because all are used to create an audio by sending an audio signal to specific speakers. This is the reason Ambisonics becomes standard in the VR industry and 360-degree video. Ambisonics is not pre-limited to any specific speaker as it creates a smooth sound sphere even when the sound field rotates. Still, traditional surround formats provide excellent imaging only in case of audio scene static. Moreover, Ambisonics also delivers a full sphere to spread the sound evenly throughout the sphere.

There are six Ambisonics formats names as A, B, C, D, E, and G formats. The first-order Ambisonics or B-format microphones are used in the representation of linear VR by using a tetrahedral array. Furthermore, these are processed in four channels, such as "W" that provides a non-directional pressure level. At the same time, "X, Y, and Z" facilitate the front-to-back, side-to-side, and up-to-down directional information, respectively. The first-order Ambisonics is only useful for a comparatively smaller sweet spot because of its limited spatial fidelity that can affect the sound localization. For this, Higher-Order Ambisonics boosts the performance efficiency of first-order Ambisonics by adding more microphones. These are provided in linear VR and required more loudspeakers. The perceptual reconstruction techniques replicate the natural listening experience for spatial audio to represent the physical sound. Binaural recording [125] that is an extended form of stereo recording, provides a 3D sound experience. Binaural recordings replicate the human ears as closely as possible by using the two 360-degree microphones the same as regular stereo recordings that capture the sound with directional microphones. 360-degree microphones to the dummy head [125] are used to serve as proxies for the human ears because it provides the precise geometry of ears. The dummy head also produces the sound waves that interact with the human head contours. A spatially stereo image is captured more precisely as compared to any other recording method with the help of 360-degree microphones.

Head-Related Transfer Functions (HRTFs) [126] are used in real-time techniques of binaural audio to reproduce the complex cues that help us to localize the sounds by filtering an audio signal. The multiple factors such as ears, head, and listening environment) can affect the cues because, in reality, we reorient ourselves to localize the sounds. Hence, it is essential for soundscape researchers to choose the proper sound recording/reproducing technique to enable the playback sounds the same

as the natural listening scenarios. Table 10 provides comprehensive detail of audio techniques that are mentioned above.

Table 10. Summary of Audio techniques.

Reconstruction Techniques		Description	Drawbacks	Examples	Reproduction Techniques	Recording Techniques
Physical	Stereo	It provides more information on the sound field and direct playback over loudspeakers. It is widely used adopted in industry and has direct playback over loudspeakers.	It just covers 1D or 2D and not supportive for head movements.	Optimized Cardioid Triangle (OCT), IRT sound	✓	✓
	Surround	It is widely used adopted in industry and has direct playback over loudspeakers.	It has a poor spatial effect.	Hamasaki Square, IRT	✗	✓
	Wave Field Synthesis (WFS)	It creates the same acoustical pressure field as exist in surrounding.	It requires a large number of speakers.	Artificial wavefronts	✓	✗
	Multichannel	This technique is better than stereo setups.	It does not facilitate	Consumer devices,	✓	✗
	Multiple Microphone Arrays	It supports the head movements and has the ability to focus on certain sounds. So, it can record a more complete sound field.	A sophisticated signal processing is required to get the desired sound because it needs a great amount of microphones to show good performance.	Nokia OZO,	✗	✓
	Ambisonics	It can be used with any speaker arrangement can be used with any speaker arrangement because it provides efficient rendering for interactive applications by facilitating the 3D sound fields. It is also known as “evocative” that means a complete 360 representation of audio.	Such type is not good for non-diegetic sound (e.g., music) because it demands high-order Ambisonics. It uses expensive types of equipment.	Sound Field SPS200 Software Controlled Microphone, Core Sound Tetra Mic	✓	✓
Perceptual	Binaural	It is most commonly used due to its simplicity. Also, it provides direct playback over headphones.	It does not provide support for head movements. It has good spatial quality but limited interaction study of the soundscape.	Bruel and Kjaer 4101 Binaural Microphone, Free Space Binaural Microphone	✓	✓

5. Standards

Currently, immersive media has gained enormous significance in exploring its technological and scientific challenges. Significant activities are being undertaken by academics and research institutions to facilitate the immersive media standardization, and a multi-phase scheme is being pursued to complete this set of standards. MPEG is currently working on ISO/IEC 23090 MPEG-I to support the immersive media coding. MPEG-I consists of the following parts: (1) Technical Report on Immersive Media, (2) Omnidirectional Media Format (OMAF), (3) Versatile Video Coding (VVC), (4) Immersive Audio Coding, (5) Point Cloud Compression, (6) Metrics, (7) Metadata, (8) Network-Based Media Processing, (9) Geometry-based Point Cloud Compression, (10) Carriage of Point Cloud Data, (11) Implementation Guidelines for Network-based Media Processing, and (12) Immersive Video.

OMAF standard defines the storage and delivery formats for omnidirectional media applications, concentrating on images, audio, and synchronized text of 360 degrees videos. Its first edition [44] ensures the storage based on ISO Base Media File Format (ISOBMFF) and MPEG MediaTransport (MMT). OMAF includes several additions such as interactivity, temporal navigation, and natural viewing experience by supporting head motion parallax. MPEG has divided the standardization associated with VR into the following categories: monoscopic 360-degree video, binocular 3D 360-degree video, stereoscopic 360-degree video, and free-viewpoint video (FVV) [127]. A set of 4 to 6 cameras take 360-degree video shots and then stitches those cameras' images into a single spherical view. In the monoscopic 360-degree video, the data is represented as 2D images but with pixels coordinates interpreted as values. While viewing a 360-degree video on HMD, the movement of users can be explained with three directions (i.e., yaw, pitch, and roll). Therefore, 360-video is also called 3DoF (degree of freedom) because both the user's eyes see the same panorama, and there is no depth impression. At the end of 2017, the part 1a of the first phase of OMAF enabled the streaming of 3DoF 360-degree video with existing comparison technologies. In 3DoF, the user is static, but the head can change orientation to look around the 360-degree video.

Part 1b of the first phase of the OMAF aims to enhance the 360-degree video with depth information named enhanced 3DOF or 3DOF+ because 3DoF cannot represent the scene behind the objects. 3DoF+ ensures the accurate parallax for a limited range of motion and leverages much of the existing 360-degree video infrastructure. The additional sensor data is used by 3DoF+ to produce a depth map to allow a player to re-project the video frames that depict the virtual movement in space. However, 3DoF+ has some disadvantages as follows: (1) visible artifacts will be minimized if not eliminated by machine learning (ML) methods, (2) no current standards for depth layer representations,

and (3) user's movement is only in a limited range. The second phase of the OMAF aims to develop the full support for 6DoF [128,129] by including point cloud coding, natural 6DoF representation (i.e., light fields), and rendering centric interactive 6DoF.

In March 2019, a Call for Proposals (CFPs) on 3DoF+ videos was announced by MPEG to establish a coding solution based on standardization of the HEVC and 3DoF+ metadata. MPEG-I TM2 for immersive video common test conditions (CTC) is desirable to conduct coding experiments in a well-defined environment. In this context, the Test Model of Immersive Video TMIV [130] specifies the standard test conditions, i.e., coding efficiency, subjective quality, pixel rate, user experience, and assessment of immersive video applications. The technical approach follows these steps: (1) compressing test content, (2) synthesizing intermediate views from decoded views and metadata (when available), (3) rendering viewports of real/virtual pose traces with a limited or a wider movement, and (4) evaluating coding efficiency and parallax effect considering both decoded and synthesized views. The bit-stream should be viewer independent, meaning that neither the position nor the orientation of the same scene from a range of locations promise the incredibly realistic immersive imagery with correct specular effects. MPEG has carried out explorations on technologies that enable 6DoF to allow the user not only to change the viewer should be considered when compressing the test content. The range of supported possible viewer position is constrained and known. Three different anchors are used, the first one includes MIV (Metadata for Immersive Video) anchor based on HEVC+TMIV. The second one includes the MIV view anchor is also HEVC + TMIV-based but directly encodes a subset of the source views. The third anchor, the MV-HEVC anchor is based on MV-HEVC and VVS. Stereoscopic 360-degree video is a 3D extension of a 360-degree video, where two panoramas of a scene are used and represented with a circular projection. In each time frame, each panorama gives an image that is captured through a rotating camera with narrow horizontal FoV. Presenting different views for the left and right eyes produce the depth sensation in a scene. However, in such type of visualization, the user has limited movements because it can produce the unnatural 3D impression with fast head movements [131].

6. Applications of 360-Degree Video

The possibilities for new immersive experiences are endless with 360-degree video. Technology's adoption by consumers is still in its early stages but proves very popular in the gaming industry. The applications of 360-degree videos are just not confined to gaming. There are many more 360-degree video uses, which range from academic research to engineering, design, business, arts, and entertainment [132,133]. The user will be able to virtually attend live sports with a favorite seat, listen to a live singer, or watch movies. Several VR simulators have been designed for training and education purposes in different fields, e.g., power plants, submarines, cranes, surgery, planes operation, and air traffic control, etc. [134]. Figure 6 signifies the growth potential of the 360-degree video market based on applications such as professional sports, travel, live events, movies, news, and TV shows. Next, the applicability of 360-degree video to various fields is briefly described.

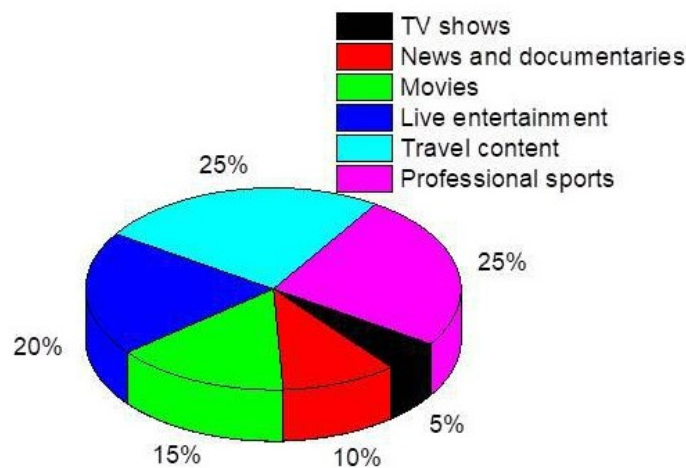


Figure 6. 360-degree video growth in different sectors.

6.1. Architectural Design

The architecture industry has achieved immense growth due to the increased immersive media technology. 360-degree video can present a model to millions of viewers just in few minutes with no or minimum loss of information. 360-degree video can preserve lifetime descriptions of engineering drawing or static components in the form of 3D models. This applicability enables researchers to demonstrate the components to be gathered, synthesized, tested and examined with possibly low time and cost consumption [134,135].

6.2. Construction Progress Monitoring

Presently, techniques of image-based visualization enable the reporting of the construction progress [136]. 360-degree interactive and immersive media can ensure the success of a construction project. Alternatively, it may be used to do exact measurement and performing advance control along other suitable procedures to be fulfilled in a specific time [137]. Researchers argue that this application is used as an e-learning tool and that [138] must be interoperable, robust, and reusable.

6.3. Medicine

The apparent and most practical applicability of 360-degree video extends to the medical area. It is proving popular in molecular modeling, ultrasound echography, computational neuroscience, and treating phobias, etc. These advancements have significantly saved time and practical costs at the training and education level. Another 360-degree video medicinal area targets to develop surgical skills without harming human beings or animals [139].

6.4. Data Visualization

It is used for graphical representation of information for making several characteristics or values more apparent. This type of application is implemented for a 3D data set resulting from Computational Fluid Dynamics (CFD) [140]. The data is visualized using the mapping of geometric objects, i.e., particle clouds or arrows to data values. For instance, arrows are implemented to data values to visualize the airflow where the width can show the volumetric flow rate, direction indicates airflow and color represent temperature.

6.5. News Broadcasting

News is always exciting and informative for a viewer. Different news broadcasters have set up 360-degree sections on their web portals, as shown in Figure 7.

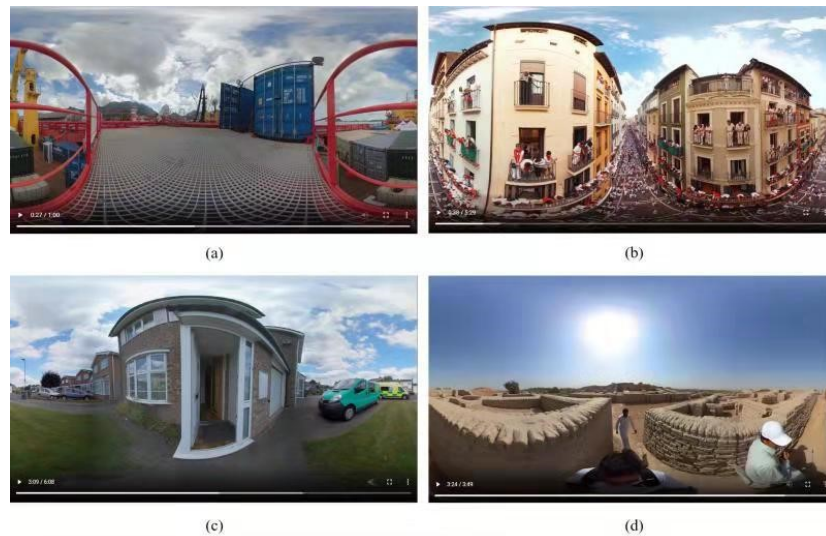


Figure 7. News reporting from different spots (a) Deep Ocean Live—Ocean Zephyr 360 by Sky News (b) Run with the bulls in Pamplona—360-degree Video by CNN (c) Ambulance: VR—360 | BBC One by BBC (d) Moenjodaro 360—A walkthrough of the ancient civilization by Al Jazeera.

6.6. Sports and Entertainment

360-degree video is found applicable for sports, for example, a round of golf can be played through a large projection screen. Presently, TV cartoons are also making use of 360-degree video applications (e.g., the BBC’s Ratz, during live broadcast, the cat is animated in real time using a tracking system on puppeteer.) [132,139]. Similarly, “Trump World” is also involved in tackling new things from technical perspectives. It was the first-ever effort to develop a system that can deliver 360-degree videos being synchronized with a television broadcast.

The media industry has deployed several technologies that enable synchronizing transmission with video delivery, including specific software and hybrid cast-capable televisions. The live 360-degree video technology possibly brings liveliness delay compared to recorded programs for which chunk files are prepared in advance. Moreover, the rapid creation of chunk files enables fast replays from 360-degree video perspectives even during a live broadcast of an event. The speculations for 2020 necessitate the development of this technology further and making more investments to deliver live 360-degree videos.

6.7. Education

360-degree video is used in education, showing complex scenes that are difficult to explain in the conventional video, images, and even words. In biological sciences, 360-degree video cameras are used to record field trips and the crime scenes in forensic science to help the students to examine it. 360-degree video recording can be a more authentic way to record classrooms as it is a powerful tool for pre-service teachers to explain all the activities performed by students. The advantages and disadvantages of every system are based on characteristics of the application environment. Some applications are highly beneficial if these are implemented using a fully immersive environment and not useful if these are implemented using a non-immersive environment. Table 11 depicts the applications with all possible systems and explains the type of system and whether it is good or not related to the application used.

Table 11. Use of 360-degree Video system based on applications.

References	Area	Fully Immersive	Semi-Immersive	Non-Immersive
[134]	Engineering and Architecture	See the specific building path to stimulate it	Visualize the 3D objects	Check the correctness of conception after designing the building
[136–139]	Construction progress monitoring	Effective by monitoring the progress of data	Image-based techniques are allowed by using visualization approaches (e.g., monitor the traffics inroads)	Enables the performance as close as possible to the wanted outcomes
[134]	Medicine	Very expensive to use fully immersive in medicine	Not expensive for every autopsy's experiment, use it in making surgery	Not useful in medicine
[141]	Education	Appropriate for training (e.g., submarines, ships, cranes, etc.)	Most efficient for education (e.g., engineering, cooking, drawing)	Good to depict the animated images by converting it into video
[132]	Entertainment and Sports	Effective for specific entertainment (e.g., some games for tourists to see the new places by using CAVE)	For play station	Good for entertainment such as cartoons and TV
[139]	Data Visualization	Having no benefits for direction of rains	Not useful for interacting with the data visualization	Appropriate for graphical representations to make specific values more apparent
[142,143]	Designing	Useful for the designing of a driving simulator and simulation of buildings	It affects the designing of 3D objects	Show 3D models

7. Challenges and Implications

360-degree videos provide an immersive experience that is difficult to find in traditional 2D videos. A significant number of production possibilities emerge because different events have been captured as 360-degree video. The rapid production in various fields has introduced the 360-degree video to wider audiences through social media platforms. A traditional virtual environment allows the user to navigate in complex theory geometries that reconstruct real areas attempting to stimulate and create real spaces. The 360-degree video introduces several challenges that need to be explored for a viable implementation of the streaming system. The major challenge experienced by the user in the virtual environment is a sense of presence. Such an understanding can be enhanced via the creation of close to the real environment while avoiding the visual cues. Many technical and design challenges and implications are explained here for the sake of an interactive, immersive, and engaging experience within 360-degree videos.

1. 360-degree video introduce several distortion from acquisition to display. To overcome the distortion issue in 360-degree video streaming, there should be a focus on adding new stitching, projection, and packaging formats that may introduce less noise.
2. The 3D objects are being included in the environment besides capturing and use of 360-degree video to represent the real world and actual interacting content. The incorporation of 3D objects is challenging for realistic view.
3. Since the user head movement is highly variable throughout the streaming session, using a fixed tiling scheme as in existing studies might lead to non-optimal viewport quality. When the viewport prediction accuracy is good, many tiles can be used as it can reduce the number of redundant pixels, which are the pixels not in the viewport. Meanwhile, redundant pixels in case of a small number of tiles can help to deal with high prediction errors. Therefore, the number of tiles in the streaming framework should be dynamically selected to improve the streaming quality.
4. Adaptation mechanisms should be smart enough to accurately adapt according to the environmental factors. In this context, deep reinforcement learning (DRL)-based strategies

should be developed to allocate suitable bitrates to the tiles in different regions of the 360-degree video frame.

5. The navigation in the 360-degree video is operated while using a backward or forward option for moving between frames or supported by camera movements [144,145]. Such an application enables designers to perform the naturally realistic task to provide non-real-world functionality and using analogous for commands at the time of need. One key challenge faced by the researchers is to support normal visual angle orientation while navigating through 360-degree video. The free navigation of the user through a 360-degree video can easily make him/her feel anxious about missing something important [146]. The rich environments should be equipped with novel orientation mechanisms for supporting full 360-degree video while reducing the cognitive load to overcome this problem.
6. The true navigation depends on viewport prediction mechanisms. The modern prediction approaches should use the spatial and temporal image features as well as the positional information of the user with suitable encoder-decoder convolutional LSTM architecture to mitigate long-term prediction errors.
7. As the immersive media technology aims at endowing the user with an unprecedented sense of full immersion in the real world. It dynamically varies with the user interaction and possible by projecting the user at the center of the scene. This interactivity for the immersive user experience is driven by HMDs or by remote control in free-viewpoint television. With the increasing use of 360-degree VR applications video in recent years, immersive media demands new ways of interactivity. Despite their immersive nature, these videos cannot directly interact. The novel challenges due to the user's interaction with the scene are created through the coding and transmission perspective [140,147]. Therefore, it is crucial to predict the user's behavior for the efficient coding and streaming of interactive content. Therefore, authors in [148] presented the interaction in the form of a hotspot. In [149,150], different interaction methods to control the 360-degree video playback system have been discussed. A similar technique in [151] was suggested to stream the interactive omnidirectional video. In addition, different technologies have been implemented for 360-degree video playbacks such as CAVEs [152], gesture-based over interface [153], large screens [154], and effect of immersiveness and future VR expectations [155]. A study in [144] defined a new technique for the interaction of 360-degree video using an immersive VR system. However, researchers have already investigated the different interaction aspects. However, more efforts for interactive 360-degree experience are highly needed.
8. There is a need for a concentrated effort towards designing quality assessment methods and metrics for 360-degree video. This is a complex and challenging problem because of the unknown network fluctuations and traditional video QoE models that do not consider the 360-degree content.
9. Special sound effects used in the 360-degree video require strong research intention before using it in the context of attracting attention.

8. Discussion and Conclusions

The emerging 360-degree video has attracted the attention of many researchers. It has been all-time popular in multimedia applications such as gaming, education, entertainment, tourism, and sports, among others. Through years a vast number of better works have been focused on improving 360-degree video streaming. However, it has always been challenging because such types of videos need a higher bitrate than traditional videos because of high-resolution (6K and beyond).

This paper explained the streaming architecture of 360-degree video that is compatible with MPEG-DASH and traditional CDNs. Several distortions associated with capturing, stitching, projection, encoding, transmission, and displaying are presented. Projection approaches play a critical role in deciding the overall quality of the frames. The cubemap projection is more efficient compared to the

equiangular version based on the current 4k encoding techniques. CMP transmits more information to the user's as compared to the un-oriented projections [36].

The modern streaming approaches such as viewport-based and tile-based streaming which aim to reduce the bandwidth and latency requirements of high-resolution content are presented and explained. Viewport-based streaming considers differential quality streaming and needs to prepare several adaptation sets at the server-side. Such types of adaptation involve huge storage and processing overheads. Tile-based streaming has low storage overhead and provides efficient caching and computation support [15,16]. The bitrate allocation decisions for both streaming technologies should try to balance several environmental factors such as viewport prediction errors, rebuffering, response delay, viewport quality, resource use. The audio and video related technologies and standardization efforts are explained in detail to enable a higher degree-of-freedom immersive environment. This paper described the salient features and technical challenges and implications for the viable implementation of 360-degree video.

Despite the popularity of the topic and abundant research efforts, several research challenges (mainly concerning projection, encoding, tiling selection, bitrate adaptation, viewport prediction, etc.) still exist. The standardization efforts are already showing much interest to provide important insights for 360-degree video streaming. Such issues should be addressed before real implementation to ensure the user's best experience.

Author Contributions: R.S. writing—original draft preparation. W.S. provided the guidance, refined the manuscript and provided valuable comments. M.U.Y. helps in writing manuscript and proofread the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Northwestern Polytechnical University Xian, China.

Acknowledgments: We would like to thank Kaifang Yang from Shaanxi Normal University for his insightful comments to improve the quality of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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), Klagenfurt, Austria, 10–13 May 2017; pp. 1–7.
2. Windows Mixed Reality Platform. Available online: <https://www.microsoft.com/en-us/windows/windows-mixed-reality> (accessed on 15 February 2020).
3. ARTE. Available online: <https://www.arte.tv/en/> (accessed on 15 February 2020).
4. Oculus Rift. Available online: <https://www.oculus.com/> (accessed on 15 February 2020).
5. Vive. Available online: <https://www.vive.com/> (accessed on 15 March 2020).
6. Samsung Gear VR. Available online: <https://www.samsung.com/global/galaxy/gear-vr/> (accessed on 19 March 2020).
7. Google Cardboard. Available online: <https://arvr.google.com/cardboard/> (accessed on 19 March 2020).
8. Xiao, M.; Zhou, C.; Liu, Y.; Chen, S. Optile: Toward optimal tiling in 360-degree video streaming. In Proceedings of the 25th ACM international conference on Multimedia, Mountain View, CA, USA, 17 October 2017; pp. 708–716.
9. WebV. Available online: <https://webvr.info/samples/> (accessed on 19 March 2020).
10. Xu, M.; Li, C.; Zhang, S.; Callet, P.L. State-of-the-Art in 360° Video/Image Processing: Perception, Assessment and Compression. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 5–26, doi:10.1109/JSTSP.2020.2966864. [CrossRef]
11. Lee, J.; Lee, J.; Lim, J.; Kim, M. Bandwidth-Efficient Live Virtual Reality Streaming Scheme for Reducing View Adaptation Delay. *TIIS* **2019**, *13*, 291–304.
12. Ghaznavi-Youvalari, R.; Zare, A.; Aminlou, A.; Hannuksela, M.M.; Gabbouj, M. Shared Coded Picture Technique for Tile-Based Viewport-Adaptive Streaming of Omnidirectional Video. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 3106–3120, doi:10.1109/TCSVT.2018.2874179. [CrossRef]

13. Ma, K.J.; Bartoš, R.; Bhatia, S. A survey of schemes for Internet-based video delivery. *J. Netw. Comput. Appl.* **2011**, *34*, 1572–1586. [[CrossRef](#)]
14. Qian, F.; Ji, L.; Han, B.; Gopalakrishnan, V. Optimizing 360 video delivery over cellular networks. In Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges, New York, NY, USA, 30 October 2016; pp. 1–6.
15. Hosseini, M. View-aware tile-based adaptations in 360 virtual reality video streaming. In Proceedings of the 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 18–22 March 2017; pp. 423–424.
16. van der Hooft, J.; Vega, M.T.; Petrangeli, S.; Wauters, T.; De Turck, F. Optimizing adaptive tile-based virtual reality video streaming. In Proceedings of the 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Washington, DC, USA, 8–12 April 2019; pp. 381–387.
17. 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.
18. Hou, X.; Lu, Y.; Dey, S. A novel hyper-cast approach to enable cloud-based virtual classroom applications. In Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016; pp. 533–536.
19. Wei, X.; Thibos, L. Design and validation of a scanning Shack Hartmann aberrometer for measurements of the eye over a wide field of view. *Opt. Express.* **2010**, *18*, 1134–1143. [[CrossRef](#)] [[PubMed](#)]
20. 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.
21. Eltobgy, O. VRCast: Mobile Streaming of Live 360-Degree Videos. Ph.D. Thesis, School of Computing Science, Vancouver, BC, Canada, 10 December 2018.
22. 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.
23. Bao, Y.; Zhang, T.; Pande, A.; Wu, H.; Liu, X. Motion-prediction-based multicast for 360-degree video transmissions. In Proceedings of the 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), San Diego, CA, USA, 12–14 June 2017; pp. 1–9.
24. Kumar, S.; Sarkar, A.; Sur, A. A resource allocation framework for adaptive video streaming over LTE. *J. Netw. Comput. Appl.* **2017**, *97*, 126–139. [[CrossRef](#)]
25. Gurrieri, L.E.; Dubois, E. Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: A review. *J. Electron. Imaging* **2013**, *22*, 030902. [[CrossRef](#)]
26. Anderson, R.; Gallup, D.; Barron, J.T.; Kontkanen, J.; Snavely, N.; Hernández, C.; Agarwal, S.; Seitz, S.M. Jump: Virtual reality video. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–13. [[CrossRef](#)]
27. Knorr, S.; Croci, S.; Smolic, A. A modular scheme for artifact detection in stereoscopic omni-directional images. In Proceedings of the Irish Machine Vision and Image Processing Conference, Maynooth University, Maynooth, Ireland, 30 August–1 September 2017.
28. Jiang, W.; Gu, J. Video stitching with spatial-temporal content-preserving warping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 42–48.
29. Akhshabi, S.; Anantakrishnan, L.; Dovrolis, C.; Begen, A.C. Server-based traffic shaping for stabilizing oscillating adaptive streaming players. In Proceedings of the 23rd ACM Workshop on Network and Operating Systems Support for Digital Audio and Video, Oslo, Norway, 26 February–1 March 2013; pp. 19–24.
30. Tan, J.; Cheung, G.; Ma, R. 360-degree virtual-reality cameras for the masses. *IEEE Multimed.* **2018**, *25*, 87–94. [[CrossRef](#)]
31. Younus, M.U.; Nadeem, M.A.; Abbas, W.; Yong, L.; Shafi, R. Design and implementation of wireless sensor networks integrated with RFID tags for navigation purpose. In Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 2240–2245.
32. Corbillon, X.; De Simone, F.; Simon, G. 360-degree video head movement dataset. In Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, 20–23 June 2017; pp. 199–204.

33. Chen, Z.; Li, Y.; Zhang, Y. Recent advances in omnidirectional video coding for virtual reality: Projection and evaluation. *Signal Process.* **2018**, *146*, 66–78. [[CrossRef](#)]
34. Jiang, J.; Sekar, V.; Zhang, H. Improving fairness, efficiency, and stability in http-based adaptive video streaming with festive. In Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies, Nice, France, 10–13 December 2012; pp. 97–108.
35. Lin, J.L.; Lee, Y.H.; Shih, C.H.; Lin, S.Y.; Lin, H.C.; Chang, S.K.; Wang, P.; Liu, L.; Ju, C.C. Efficient projection and coding tools for 360 video. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 84–97. [[CrossRef](#)]
36. Dimitrijević, A.M.; Lambers, M.; Rančić, D. Comparison of spherical cube map projections used in planet-sized terrain rendering. *Facta Univ. Ser. Math. Inform.* **2016**, *31*, 259–297.
37. He, Y.; Xiu, X.; Hanhart, P.; Ye, Y.; Duanmu, F.; Wang, Y. Content-adaptive 360-degree video coding using hybrid cubemap projection. In Proceedings of the 2018 Picture Coding Symposium (PCS), San Francisco, CA, USA, 24–27 June 2018; pp. 313–317.
38. Hanhart, P.; Xiu, X.; He, Y.; Ye, Y. 360 video coding based on projection format adaptation and spherical neighboring relationship. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2018**, *9*, 71–83. [[CrossRef](#)]
39. Kuznyakov, E.; Pio, D. Next-Generation Video Encoding Techniques for 360 Video and VR. 2020. Available online: <https://engineering.fb.com/virtual-reality/next-generation-video-encoding-techniques-for-360-video-and-vr> (accessed on 5 August 2020)
40. Grois, D.; Nguyen, T.; Marpe, D. Performance comparison of AV1, JEM, VP9, and HEVC encoders. In Proceedings of the Applications of Digital Image Processing XL. International Society for Optics and Photonics, San Diego, CA, USA, 8 February 2018; Volume 10396, p. 103960L.
41. Bankoski, J.; Wilkins, P.; Xu, Y. Technical overview of VP8, an open source video codec for the web. In Proceeding of the 2011 IEEE International Conference on Multimedia and Expo, Barcelona, Spain, 11–15 July 2011; pp. 1–6.
42. Mukherjee, D.; Bankoski, J.; Grange, A.; Han, J.; Koleszar, J.; Wilkins, P.; Xu, Y.; Bultje, R. The latest open-source video codec VP9—an overview and preliminary results. In Proceedings of the 2013 Picture Coding Symposium (PCS), San Jose, CA, USA, 8–11 December 2013; pp. 390–393.
43. Mukherjee, D.; Su, H.; Bankoski, J.; Converse, A.; Han, J.; Liu, Z.; Xu, Y. An overview of new video coding tools under consideration for VP10: the successor to VP9. In Proceedings of the Applications of Digital Image Processing XXXVIII. International Society for Optics and Photonics, San Diego, CA, USA, 10–13 August 2015; Volume 9599, p. 95991E.
44. Wien, M.; Boyce, J.M.; Stockhammer, T.; Peng, W.H. Standardization status of immersive video coding. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 5–17. [[CrossRef](#)]
45. Alliance for Open Media. Available online: <https://aomedia.org/> (accessed on 24 February 2020).
46. Budagavi, M.; Furton, J.; Jin, G.; Saxena, A.; Wilkinson, J.; Dickerson, A. 360 degrees video coding using region adaptive smoothing. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 750–754.
47. Batalla, J.M.; Krawiec, P.; Beben, A.; Wisniewski, P.; Chydzinski, A. Adaptive video streaming: Rate and buffer on the track of minimum rebuffering. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 2154–2167. [[CrossRef](#)]
48. Radhakrishnan, R.; Nayak, A. Cross layer design for efficient video streaming over LTE using scalable video coding. In Proceedings of the 2012 IEEE International Conference on Communications (ICC), Ottawa, ON, Canada, 10–15 June 2012; pp. 6509–6513.
49. Frederick Pearson, I. *Map Projections Theory and Applications*; Routledge: Milton Park, Abingdon, UK, 2018.
50. Unterweger, A. Compression artifacts in modern video coding and state-of-the-art means of compensation. In *Multimedia Networking and Coding*; IGI Global: Hershey, PA, USA, 2013; pp. 28–49.
51. De Simone, F.; Frossard, P.; Birkbeck, N.; Adsumilli, B. Deformable block-based motion estimation in omnidirectional image sequences. In Proceedings of the 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSp), Luton, UK, 16–18 October 2017; pp. 1–6.
52. Li, L.; Li, Z.; Budagavi, M.; Li, H. Projection based advanced motion model for cubic mapping for 360-degree video. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1427–1431, doi:10.1109/ICIP.2017.8296517. [[CrossRef](#)]
53. Ghaznavi-Youvalari, R.; Aminlou, A. Geometry-Based Motion Vector Scaling for Omnidirectional Video Coding. In Proceedings of the 2018 IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 10–12 December 2018; pp. 127–130, doi:10.1109/ISM.2018.00030. [[CrossRef](#)]

54. 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](#)]
55. Boyce, J.; Xu, Q. Spherical rotation orientation indication for HEVC and JEM coding of 360 degree video. In Proceedings of the Applications of Digital Image Processing XL, San Diego, CA, USA, 17 September 2017; Volume 10396, pp. 61–67.
56. Su, Y.C.; Grauman, K. Learning Compressible 360° Video Isomers. In Proceedings of the The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
57. Sreedhar, K.K.; Aminlou, A.; Hannuksela, M.M.; Gabbouj, M. Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications. In Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016; pp. 583–586.
58. Sun, L.; Duanmu, F.; Liu, Y.; Wang, Y.; Ye, Y.; Shi, H.; Dai, D. A two-tier system for on-demand streaming of 360 degree video over dynamic networks. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 43–57. [[CrossRef](#)]
59. Zhou, C.; Li, Z.; Liu, Y. A measurement study of oculus 360 degree video streaming. In Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, 20–23 June 2017; pp. 27–37.
60. Li, Y.; Markopoulou, A.; Apostolopoulos, J.; Bambos, N. Content-aware playout and packet scheduling for video streaming over wireless links. *IEEE Trans. Multimed.* **2008**, *10*, 885–895.
61. Liu, K.; Liu, Y.; Liu, J.; Argyriou, A.; Yang, X. Joint source encoding and networking optimization for panoramic video streaming over lte-a downlink. In Proceedings of the GLOBECOM 2017—2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–7.
62. Concolato, C.; Le Feuvre, J.; Denoual, F.; Mazé, F.; Nasseur, E.; Ouedraogo, N.; Taquet, J. Adaptive streaming of hevc tiled videos using mpeg-dash. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 1981–1992. [[CrossRef](#)]
63. Younus, M.U.; Bukhari, S.S.H.; Abbas, W.; Shafi, R. Optimized indoor lighting system to save energy through window blinds management. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 2817–2821.
64. Zhou, C.; Li, Z.; Osgood, J.; Liu, Y. On the effectiveness of offset projections for 360-degree video streaming. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2018**, *14*, 1–24. [[CrossRef](#)]
65. Ozcinar, C.; De Abreu, A.; Smolic, A. Viewport-aware adaptive 360 video streaming using tiles for virtual reality. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2174–2178.
66. Bernet, Y.; Ford, P.; Yavatkar, R.; Baker, F.; Zhang, L.; Speer, M.; Braden, R.; Davie, B.; Wroclawski, J.; Felstaine, E. A Framework for Integrated Services Operation Over Diffserv Networks; Technical Report; RFC 2998; 2000. Available online: <https://tools.ietf.org/html/rfc2998> (accessed on 5 August 2020).
67. Blake, S.; Black, D.; Carlson, M.; Davies, E.; Wang, Z.; Weiss, W. An Architecture for Differentiated Services. 1998. Available online: <https://www.semanticscholar.org/paper/An-Architecture-for-Differentiated-Services-Blake-Black/13c3630c03bec0c3e443ad5a3dc6d1951db74c20> (accessed on 5 August 2020).
68. Braden, R.; Zhang, L.; Berson, S.; Herzog, S.; Jamin, S. Resource ReSerVation Protocol: (RSVP); Version 1 Functional Specification; December 1997. Available online: <https://tools.ietf.org/html/rfc2205> (accessed on 5 August 2020).
69. Schulzrinne, H.; Casner, S.; Frederick, R.; Jacobson, V. RTP: A Transport Protocol for Real-Time Applications. Network Working Group; Technical Report; 2003. Available online: <https://www.rfc-editor.org/rfc/rfc3550.pdf> (accessed on 5 August 2020).
70. Schulzrinne, H.; Rao, A.; Lanphier, R. Real Time Streaming Protocol (RTSP). April 1998. Available online: <https://www.rfc-editor.org/info/rfc2326> (accessed on 5 August 2020).
71. Postel, J. Rfc0793: Transmission Control Protocol. Internet Engineering Task Force (IETF); Technical Report; September 1981. Available online: <https://tools.ietf.org/html/rfc793> (accessed on 5 August 2020).
72. Nguyen, D.V.; Tran, H.T.; Thang, T.C. Impact of delays on 360-degree video communications. In Proceedings of the 2017 TRON Symposium (TRONSHOW), Tokyo, Japan, 13–14 December 2017; pp. 1–6.
73. He, D.; Westphal, C.; Garcia-Luna-Aceves, J. Joint rate and fov adaptation in immersive video streaming. In Proceedings of the 2018 Morning Workshop on Virtual Reality and Augmented Reality Network, Budapest, Hungary, 24 August 2018; pp. 27–32.
74. Zhang, X.; Toni, L.; Frossard, P.; Zhao, Y.; Lin, C. Adaptive streaming in interactive multiview video systems. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 1130–1144. [[CrossRef](#)]

75. 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; p. 4592.
76. Roche, L.; Gal-Petitfaux, N. Using 360 video in physical education teacher education. In Proceedings of the Society for Information Technology & Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE), Austin, TX, USA, 5 March 2017; pp. 3420–3425.
77. Kelling, C.; Väättäjä, H.; Kauhanen, O. Impact of device, context of use, and content on viewing experience of 360-degree tourism video. In Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, Stuttgart, Germany, 26–29 November 2017; pp. 211–222.
78. Skupin, R.; Sanchez, Y.; Hellge, C.; Schierl, T. Tile based HEVC video for head mounted displays. In Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016; pp. 399–400.
79. Graf, M.; Timmerer, C.; Mueller, C. Towards bandwidth efficient adaptive streaming of omnidirectional video over http: Design, implementation, and evaluation. In Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, 20–23 June 2017; pp. 261–271.
80. 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, 30 October 2015; pp. 1–6.
81. Ozcinar, C.; Cabrera, J.; Smolic, A. Visual attention-aware omnidirectional video streaming using optimal tiles for virtual reality. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 217–230. [[CrossRef](#)]
82. Nguyen, D.V.; Tran, H.T.; Pham, A.T.; Thang, T.C. An optimal tile-based approach for viewport-adaptive 360-degree video streaming. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 29–42. [[CrossRef](#)]
83. 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; pp. 107–110.
84. Xie, L.; Xu, Z.; Ban, Y.; Zhang, X.; Guo, Z. 360probdash: 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; pp. 315–323.
85. Zare, A.; Aminlou, A.; Hannuksela, M.M.; Gabbouj, M. HEVC-compliant tile-based streaming of panoramic video for virtual reality applications. In Proceedings of the 24th ACM international conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 601–605.
86. de la Fuente, Y.S.; Bhullar, G.S.; Skupin, R.; Hellge, C.; Schierl, T. Delay impact on MPEG OMAF's tile-based viewport-dependent 360° video streaming. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2019**, *9*, 18–28. [[CrossRef](#)]
87. Nguyen, M.; Nguyen, D.H.; Pham, C.T.; Ngoc, N.P.; Nguyen, D.V.; Thang, T.C. An adaptive streaming method of 360 videos over HTTP/2 protocol. In Proceedings of the 2017 4th NAFOSTED Conference on Information and Computer Science, Hanoi, Vietnam, 24–25 November 2017; pp. 302–307.
88. Sitzmann, V.; Serrano, A.; Pavel, A.; Agrawala, M.; Gutierrez, D.; Masia, B.; Wetzstein, G. How do People Explore Virtual Environments? *arXiv* **2016**, arXiv:1612.04335.
89. 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; pp. 1–4.
90. Niamut, O.A.; Thomas, E.; D'Acunto, L.; Concolato, C.; Denoual, F.; Lim, S.Y. MPEG DASH SRD: Spatial relationship description. In Proceedings of the 7th International Conference on Multimedia Systems, Klagenfurt, Austria, 10–13 May 2016; pp. 1–8.
91. Tan, T.K.; Weerakkody, R.; Mrak, M.; Ramzan, N.; Baroncini, V.; Ohm, J.R.; Sullivan, G.J. Video quality evaluation methodology and verification testing of HEVC compression performance. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *26*, 76–90. [[CrossRef](#)]
92. Seshadrinathan, K.; Soundararajan, R.; Bovik, A.C.; Cormack, L.K. Study of subjective and objective quality assessment of video. *IEEE Trans. Image Process.* **2010**, *19*, 1427–1441. [[CrossRef](#)]
93. Perrin, A.F.; Bist, C.; Cozot, R.; Ebrahimi, T. Measuring quality of omnidirectional high dynamic range content. In Proceedings of the Applications of Digital Image Processing XL. International Society for Optics and Photonics, San Diego, CA, USA, 19 September 2017; Volume 10396, p. 1039613.

94. Huang, M.; Shen, Q.; Ma, Z.; Bovik, A.C.; Gupta, P.; Zhou, R.; Cao, X. Modeling the perceptual quality of immersive images rendered on head mounted displays: Resolution and compression. *IEEE Trans. Image Process.* **2018**, *27*, 6039–6050. [[CrossRef](#)]
95. Azevedo, R.G.A.; Birkbeck, N.; De Simone, F.; Janatra, I.; Adsumilli, B.; Frossard, P. Visual distortions in 360-degree videos. *arXiv* **2019**, arXiv:1901.01848.
96. 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; pp. 1–6.
97. Younus, M.U. Analysis of the impact of different parameter settings on wireless sensor network lifetime. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 16–21.
98. Schatz, R.; Sackl, A.; Timmerer, C.; Gardlo, B. Towards subjective quality of experience assessment for omnidirectional video streaming. In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; pp. 1–6.
99. Tran, H.T.; Ngoc, N.P.; Pham, A.T.; Thang, T.C. A multi-factor QoE model for adaptive streaming over mobile networks. In Proceedings of the 2016 IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–6 December 2016; pp. 1–6.
100. 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; pp. 232–242.
101. Singla, A.; Fremerey, S.; Robitza, W.; Raake, A. Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays. In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; pp. 1–6.
102. 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 on Thematic Workshops of ACM Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 511–519.
103. Upenik, E.; Řeřábek, M.; Ebrahimi, T. Testbed for subjective evaluation of omnidirectional visual content. In Proceedings of the 2016 Picture Coding Symposium (PCS), Nuremberg, Germany, 4–7 December 2016; pp. 1–5.
104. Orduna, M.; Pérez, P.; Díaz, C.; García, N. Evaluating the Influence of the HMD, Usability, and Fatigue in 360VR Video Quality Assessments. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; pp. 683–684.
105. 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; pp. 1–6.
106. Tran, H.T.; Ngoc, N.P.; Bui, C.M.; Pham, M.H.; Thang, T.C. An evaluation of quality metrics for 360 videos. In Proceedings of the 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, Italy, 4–7 July 2017; pp. 7–11.
107. Möller, S.; Raake, A. *Quality of Experience: Advanced Concepts, Applications And Methods*; Springer: Berlin/Heidelberg, Germany, 2014.
108. Yang, S.; Zhao, J.; Jiang, T.; Rahim, J.W.T.; Zhang, B.; Xu, Z.; Fei, Z. An objective assessment method based on multi-level factors for panoramic videos. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
109. da Costa Filho, R.I.T.; Luizelli, M.C.; Vega, M.T.; van der Hooft, J.; Petrangeli, S.; Wauters, T.; De Turck, F.; Gaspary, L.P. Predicting the performance of virtual reality video streaming in mobile networks. In Proceedings of the 9th ACM Multimedia Systems Conference, Amsterdam, The Netherlands, 12–15 June 2018; pp. 270–283.
110. Zakharchenko, V.; Choi, K.P.; Park, J.H. Quality metric for spherical panoramic video. In Proceedings of the Optics and Photonics for Information Processing X. International Society for Optics and Photonics, San Diego, CA, USA, 14 September 2016; Volume 9970, p. 99700C.
111. 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; pp. 1–6.

112. Filho, R.I.T.D.C.; Luizelli, M.C.; Petrangeli, S.; Vega, M.T.; Hooft, J.V.d.; Wauters, T.; Turck, F.D.; Gaspary, L.P. Dissecting the Performance of VR Video Streaming through the VR-EXP Experimentation Platform. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2019**, *15*, 1–23. [CrossRef]
113. Yu, L.; Tillo, T.; Xiao, J. QoE-driven dynamic adaptive video streaming strategy with future information. *IEEE Trans. Broadcast.* **2017**, *63*, 523–534. [CrossRef]
114. Chiariotti, F.; D’Aronco, S.; Toni, L.; Frossard, P. Online learning adaptation strategy for DASH clients. In Proceedings of the 7th International Conference on Multimedia Systems, Klagenfurt, Austria, 10–13 May 2016; pp. 1–12.
115. 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; pp. 932–940.
116. 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; pp. 1316–1321.
117. Bates, E.; Boland, F. Spatial Music, Virtual Reality, and 360 Media. In Proceedings of the Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality; Audio Engineering Society, Los Angeles, CA, USA, 1 October 2016.
118. Huang, H.; Solah, M.; Li, D.; Yu, L.F. Audible panorama: Automatic spatial audio generation for panorama imagery. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–11.
119. Google VR Software Development Kit. Available online: <https://developers.google.com/vr> (accessed on 23 May 2020).
120. Facebook spatial Workstation. Available online: <https://facebook360.fb.com/spatial-workstation/> (accessed on 23 May 2020).
121. He, J. *Spatial Audio Reproduction With Primary Ambient Extraction*; Springer: Berlin, Germany, 2017.
122. Riaz, H.; Stiles, M.; Armstrong, C.; Chadwick, A.; Lee, H.; Kearney, G. Multichannel Microphone Array Recording for Popular Music Production in Virtual Reality. In Proceedings of the 143rd Audio Engineering Society Convention - Jacob K. Javits Convention Center, New York, NY, USA, 18–20 October 2017.
123. Guastavino, C.; Katz, B.F.; Polack, J.D.; Levitin, D.J.; Dubois, D. Ecological validity of soundscape reproduction. *Acta Acust. United Acust.* **2005**, *91*, 333–341.
124. Frank, M.; Zotter, F.; Sontacchi, A. Producing 3D audio in ambisonics. In Proceedings of the 57th International Conference: The Future of Audio Entertainment Technology—Cinema, Television and the Internet, Hollywood, CA, USA, 6–8 March 2015.
125. Zieliński, S.K.; Lee, H. Automatic Spatial Audio Scene Classification in Binaural Recordings of Music. *Appl. Sci.* **2019**, *9*, 1724. [CrossRef]
126. Xie, B. *Head-Related Transfer Function and Virtual Auditory Display*; J. Ross Publishing: Richmond, VA, USA, 2013.
127. Jeong, J.; Jang, D.; Son, J.; Ryu, E.S. 3DoF+ 360 video location-based asymmetric down-sampling for view synthesis to immersive VR video streaming. *Sensors* **2018**, *18*, 3148. [CrossRef]
128. Jeong, J.; Jang, D.; Son, J.W.; Ryu, E.S. Bitrate efficient 3DoF+ 360 video view synthesis for immersive VR video streaming. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 17–19 October 2018; pp. 581–586.
129. Younus, M.U.; Kim, S.W. Proposition and Real-Time Implementation of an Energy-Aware Routing Protocol for a Software Defined Wireless Sensor Network. *Sensors* **2019**, *19*, 2739. [CrossRef]
130. Test Model of Immersive Video. Available online: <https://sites.google.com/site/dragoam/mpeg-i> (accessed on 12 June 2020).
131. Lu, A.; Sun, Y.; Wang, B.; Yu, L. Analysis on circular projection of 360 degree 3D video. In Proceedings of the 117th MPEG Meeting Concluded, Geneva, Switzerland, 20 January 2017.
132. Hui-Zhen, R.; Zong-Fa, L. Application and prospect of the virtual reality technology in college ideological education. In Proceedings of the 2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications, Zhangjiajie, China, 6–7 November 2013; pp. 125–128.

133. Neng, L.A.; Chambel, T. Get around 360 hypervideo. In Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 6–8 October 2010; pp. 119–122.
134. Alqahtani, A.S.; Daghestani, L.F.; Ibrahim, L.F. Environments and system types of virtual reality technology in STEM: A survey. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*. **2017**, *8*, 6.
135. Younus, M.U.; ul Islam, S.; Ali, I.; Khan, S.; Khan, M.K. A survey on software defined networking enabled smart buildings: Architecture, challenges and use cases. *J. Netw. Comput. Appl.* **2019**, *137*, 62–77. [[CrossRef](#)]
136. Sampaio, A.Z. Virtual reality technology applied in teaching and research in civil engineering education. *J. Inf. Technol. Appl. Educ.* **2012**, *1*, 152–163.
137. Son, H.; Kim, C. 3D structural component recognition and modeling method using color and 3D data for construction progress monitoring. *Autom. Constr.* **2010**, *19*, 844–854. [[CrossRef](#)]
138. Birzina, R.; Fernate, A.; Luka, I.; Maslo, I.; Surikova, S. E-learning as a challenge for widening of opportunities for improvement of students' generic competences. *Learn. Digit. Media* **2012**, *9*, 130–142. [[CrossRef](#)]
139. Cox, C. The Use of Computer Graphics and Virtual Reality for Visual Impact Assessments. Ph.D. Thesis, University of Nottingham, UK, 2003.
140. El-Ganainy, T.; Hefeeda, M. Streaming virtual reality content. *arXiv* **2016**, arXiv:1612.08350.
141. Wang, J. Virtual Reality Technology in the Design of the Space Environment Research. In Proceedings of the 2011 International Conference on Control, Automation and Systems Engineering (CASE), Singapore, 30–31 July 2011; pp. 1–4. doi:10.1109/ICCSE.2011.5997892. [[CrossRef](#)]
142. Kaufmann, H.; Schmalstieg, D. Designing Immersive Virtual Reality for Geometry Education. In Proceedings of the IEEE Virtual Reality Conference (VR 2006), Alexandria, VA, USA, 25–29 March 2006; pp. 51–58. doi:10.1109/VR.2006.48. [[CrossRef](#)]
143. Lorenzo, G.; Lledó, A.; Pomares, J.; Roig, R. Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Comput. Educ.* **2016**, *98*, 192–205, doi:10.1016/j.compedu.2016.03.018. [[CrossRef](#)]
144. Petry, B.; Huber, J. Towards effective interaction with omnidirectional videos using immersive virtual reality headsets. In Proceedings of the 6th Augmented Human International Conference, Singapore, 9–11 March 2015; pp. 217–218.
145. Younus, M.U.; Yong, L.; Shahbaz, M.; Shafi, R.; Hongkun, H. Robust security system for intruder detection and its weight estimation in controlled environment using Wi-Fi. In Proceedings of the 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 985–990.
146. Zoric, G.; Barkhuus, L.; Engström, A.; Önnvall, E. Panoramic video: Design challenges and implications for content interaction. In Proceedings of the 11th European Conference on Interactive TV And Video, Como, Italy, 24–26 June 2013; pp. 153–162.
147. 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; pp. 31–36.
148. Kallioniemi, P.; Keskinen, T.; Mäkelä, V.; Karhu, J.; Ronkainen, K.; Nevalainen, A.; Hakulinen, J.; Turunen, M. Hotspot Interaction in Omnidirectional Videos Using Head-Mounted Displays. In Proceedings of the 22nd International Academic Mindtrek Conference, Tampere, Finland, 10–11 October 2018; pp. 126–134.
149. Pakkanen, T.; Hakulinen, J.; Jokela, T.; Rakkolainen, I.; Kangas, J.; Piippo, P.; Raisamo, R.; Salmimaa, M. Interaction with WebVR 360 video player: Comparing three interaction paradigms. In Proceedings of the 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 18–22 March 2017; pp. 279–280.
150. Berning, M.; Yonezawa, T.; Riedel, T.; Nakazawa, J.; Beigl, M.; Tokuda, H. pARnorama: 360 degree interactive video for augmented reality prototyping. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, Zurich, Switzerland, 12–14 September 2013; pp. 1471–1474.
151. Quax, P.; Liesenborgs, J.; Issaris, P.; Lamotte, W.; Claes, J. A practical and scalable method for streaming omni-directional video to web users. In Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences, Barcelona, Spain, 22 October 2013; pp. 57–60.
152. Rovelo Ruiz, G.A.; Vanacken, D.; Luyten, K.; Abad, F.; Camahort, E. Multi-viewer gesture-based interaction for omni-directional video. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 4077–4086.

153. Atienza, R.; Blonna, R.; Saldares, M.I.; Casimiro, J.; Fuentes, V. Interaction techniques using head gaze for virtual reality. In Proceedings of the 2016 IEEE Region 10 Symposium (TENSymp), Bali, Indonesia, 9–11 May 2016; pp. 110–114.
154. Mousas, C. Performance-driven dance motion control of a virtual partner character. In Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, Germany, 18–22 March 2018; pp. 57–64.
155. Rupp, M.A.; Kozachuk, J.; Michaelis, J.R.; Odette, K.L.; Smither, J.A.; McConnell, D.S. The effects of immersiveness and future VR expectations on subjective-experiences during an educational 360 video. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, USA, 24 October 2016; Volume 60, pp. 2108–2112.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).