Review

Networked VR: State of the Art, Solutions, and Challenges

Jinjia Ruan [ID] and Dongliang Xie *

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China; jinjia_ruan@bupt.edu.cn
* Correspondence: xiedl@bupt.edu.cn

Abstract: The networking of virtual reality applications will play an important role in the emerging global Internet of Things (IoT) framework and it is expected to provide the foundation of the expected 5G tactile Internet ecosystem. However, considerable challenges are ahead in terms of technological constraints and infrastructure costs. The raw data rate (5 Gbps–60 Gbps) required achieving an online immersive experience that is indistinguishable from real life vastly exceeds the capabilities of future broadband networks. Therefore, simply providing high bandwidth is insufficient in compensating for this difference, because the demands for scale and supply vary widely. This requires exploring holistic solutions that exceed the traditional network domain, and integrating virtual reality (VR) data capture, encoding, network, and user navigation. Emerging services are extremely inefficient in terms of mass use and data management, which significantly reduces the user experience, due to their heuristic design choices. Other key aspects must be considered, such as wireless operation, ultra-low latency, client/network access, system deployment, edge computing/cache, and end-to-end reliability. A vast number of high-quality works have been published in this area and they will be highlighted in this survey. In addition to a thorough summary of recent progress, we also present an outlook of future developments in the quality of immersive experience networks and unified data set measurement in VR video transmission, focusing on the expansion of VR applications, security issues, and business issues, which have not yet been addressed, and the technical challenges that have not yet been completely solved. We hope that this paper will help researchers and developers to gain a better understanding of the state of research and development in VR.

Keywords: virtual reality; 360 videos; video streaming; networked

1. Introduction

With the continuous development of augmented reality (AR) and virtual reality (VR) technologies, cyber space VR plays an important role in social development. VR is revolutionizing how people know the world. Industry and academia have highly valued the development of VR. Through virtual reality, we perceive how remote objects and people exist in the environment around us, similar to virtual people and objects, in order to achieve long-range transmission. Currently, most VR applications are used for game and entertainment composite content. Because most of the current VR applications are wired for operation, the interactivity is weakest, considerably limiting the mobility and interactivity of VR systems in remote communication scenarios. Networked VR will play a vital role in the future development of the network communication field. The global Internet of Things (IoT) framework is expected to become the 5G tactile Internet ecosystem and even provide an interactive mechanism in order to maintain perceptual illusions in 6G white paper [1]. However, significant challenges are expected due to technological and infrastructure constraints. Overcoming the current networked VR challenges and technical limitations will help to guide society into the envisioned VR future, which requires a departure from traditional network solutions. With the advances and improvement in VR technology, the performance gap between the requirements of networked VR and existing and upcoming network technologies is only expected to increase [2,3]. Therefore,

networked solutions need to provide a new generation of network technologies with faster data rates or lower transmission delays in order to cope with emerging applications. The overall solution should be studied beyond the traditional network domain to tightly integrate the basic user navigation of VR data capture, encoding, network, and next-generation task interaction. In addition, attention should be paid to the development of measurement standards and unification, the mapping of quality of service (QoS) to quality of experience (QoE) (QoE–QoS) evaluation settings, and testing platforms for publicly available benchmark data sets to help accelerate such research while also promoting repeatability and standardization. Several technological advances have started to enter the VR landscape. First, the developing 5G networks [4] provide new opportunities for VR networked. The peak data rate of 5G network will reach 10 Gbps and the service delay will be less than 5 ms, which is a leap forward in transmission rate and delay when compared to 4G [5]. Second, the introduction of some new characteristics, such as edge computing, device-to-device (D2D) communication, and mmWave, provides an adaptive and scalable communication mechanism for the deployment and promotion of VR. Therefore, the 5G networks and rapid performance improvement of VR devices have laid a solid foundation for the practical deployment and application of VR on a large scale.

The main objective of this survey was to present the details and challenges of integrating networks with VR applications in the VR networking process, and the state-of-the-art technology. To this end, this survey starts with the requirements of VR networking and the various associated challenges. As a part of the survey, we examined several papers that focus on the foundations of VR and its applicability. One of these surveys analyzed the foundations of VR image processing, including a review of the three core aspects of 360-degree video/image processing, including perception, evaluation, and compression [6]. In addition, we attempted to capture the unique advantages of various relevant spherical features and visual attention models in the context of VR image processing. Subsequently, several survey papers were examined that focus on enumerating the four main use cases of cellular-connected wireless VR and identifying their unique research challenges [7]. A case study is presented in order to demonstrate the effectiveness of a quality of service (QoS) solution that defines wireless VR and the unique QoS performance requirements for VR transmissions when compared to traditional video services in cellular networks.

Increasing numbers of studies are focusing on multiple aspects of 360-degree video streaming, including acquisition, transmission, and display [8]. We analyzed several survey papers as part of an effort to review these recent investigations in the literature. The advent of 5G networks will improve network performance, but it is unclear whether it will be sufficient to provide new applications for delivering augmented and virtual reality services [9]. We then focused on the multiple research challenges that are related to important typical transmission components of the networking process at the basic representation level of VR; we concluded with an examination of three main state-of-the-art optimizations that have been implemented in order to overcome some of these challenges. Throughout the literature survey, the key focus was examining the various methodologies that are associated with networked breakthroughs. Each of these methodologies was analyzed by focusing on their applicability to the networking implementation process. The main contributions can be summarized, as follows:

- This paper discusses the architecture of VR video streaming. The VR content preprocessing stages, such as content acquisition, projection, and encoding, are organized and discussed. Subsequently, the transmission and consumption of 360-degree video is described in detail.
- The proven streaming technologies for 360-degree video are presented and discussed in detail, including viewport-based, tile-based, and viewport-tracking delivery solutions. We describe how high-resolution content can be delivered to single or multiple users. Different technical- and design-related challenges and implications are presented for the interactive, immersive, and engaging experience of VR video.

- We describe the state of the art in some recent research optimizing VR transmission by leveraging wireless communication, computational, and caching resources at the network edge in order to significantly improve the performance of VR networking.
- We outline some open research questions in the field of VR and some interesting research directions in order to stimulate future research activities in related areas.

The rest of this paper is organized, as follows: Section 2 presents the representation principles of VR and three typical VR transmission mechanisms, as well as the challenges and enabling technologies for networking VR. Section 3 summarizes the different VR networked optimization approaches that are based on edge-computing design for user-centric, node-related assisted associations, and the QoE push VR implementation. Section 4 discusses the open challenges in different ways. We conclude this paper in Section 5.

2. Background: VR Representation Principles and Typical Transmission Mechanism

In this section, we provide an overview of the representation of VR and summarize the three typical VR transmission mechanisms, followed by the challenges that VR will face when applied to real cases. Finally, we detail some enabling technologies that are necessary or recommended for the implementation of VR.

2.1. Capture and Representation of VR

The core problem of VR services is how to transmit and store panoramic VR video from the camera capture side to the final display side.The technical architecture of panoramic VR media mainly consists of video stitching and mapping, video encoding and decoding, an transmission technologies. Currently, several companies have proposed video coding schemes, and various model schemes are available for the projection methods.

2.1.1. Projection Conversion

For 360-degree video, since each image is captured by the camera at different angles, they are not on the same projection plane, so, if the overlapping images are directly and seamlessly stitched together, the visual consistency of the actual scenery will be destroyed. Therefore, the images need to be transformed by projection first, and then stitched together.

Before the video encoding process is performed by the 360-degree video source, the video that is captured by the different viewing angles must be replaced on the 2D plane. The Joint Video Exploration Group (JVET) has proposed projection solutions, including Hybrid Cubemap Projection [10], Octahedral Projection (OHP) [11], Truncated Square Pyramid Projection (TSP) [12], Icosahedral Projection (ISP) [13], and Segmented Ball Projection (SSP) [14]. In 2016, Facebook proposed the famous cube map [15] and pyramid [16] projection methods and coding schemes, specifically for 360-degree video streams, with better compression improvements, respectively.

2.1.2. Video Encoding

In VR application systems, a media file (in the case of live video, a stream including chunks of audio-visual data) is encoded or transcoded into multiple representations. HEVC/H.265 is currently the most widely used video coding format. This video coding standard was introduced by the Moving Picture Experts Group (MPEG) in collaboration with The ITU-T Video Coding Experts Group (VCEG). In 2018, the MPEG developed standardization work (MPEG-I) for immersive media, and panoramic video is the video part of immersive media [17]. The Joint Video Exploration Team (JVET) has also embarked on a video capture standard, High Efficiency Video Coding (HEVC) for panoramic video [18]. The MPEG organization plans for specific technical work in the next five years that is based on future video application trends and industry needs, as shown in Figure 1 [19].
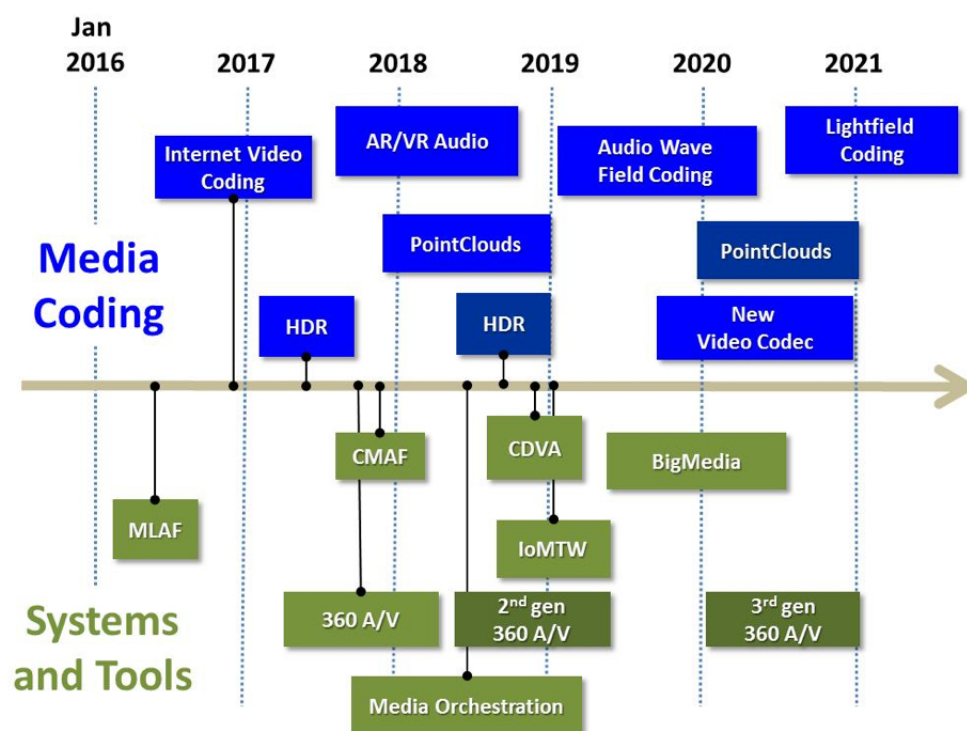
**Figure 1.** Moving Picture Experts Group (MPEG) standardization roadmap.

MPEG is currently working on ISO/IEC 23090 MPEG-I in order to support 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.

2.2. Typical VR Transmission Mechanisms

The high resolution of VR video means that a huge amount of data must be transmitted, creating a challenge for the bandwidth and real-time capabilities of the network. We consider adaptive streaming of omnidirectional/360-degree video content of virtual reality (VR) to be a challenging task. The research indicates that VR video transmission requires intelligent coding and streaming technologies to meet today's and tomorrow's application and service needs. We explored various options that enable rich and efficient omnidirectional video adaptive streaming. Currently, the main transmission mechanisms are based on the dynamic adaptive HTTP streaming (DASH) scheme and the VR videos transmission scheme that is based on tile and view switching.

2.2.1. VR Video Transmission Based on DASH

For the DASH scheme for OMAF, to improve the bandwidth used for the transmission of VR videos storage space is sacrificed [20]. The VR video transmission is mainly achieved by dynamic adaptive streaming technology with code rate and perspective. Each view stores multiple video streams of different bitrates on the DASH server. According to the view information on the client, the main perspective slice stream of the higher code rate and the other perspective slice stream of the lower code rate are transmitted. Figure 2 [21] shows its technical framework.
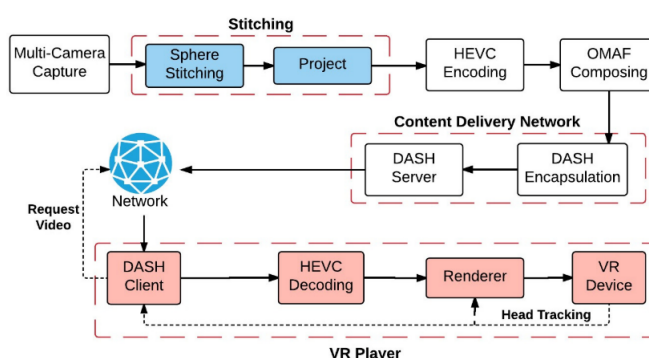
Figure 2. Dynamic adaptive HTTP streaming (DASH) Omnidirectional Media Format (OMAF) architecture network.

In recent years, some of the studies have improved QoE of 360 video streaming to a certain extent based on DASH framework [22]. In VR 360-degree video transmission, the user only sees part of the 360-degree video at each moment. Therefore, transmitting all of the content of the panorama wastes bandwidth and computing resources. These problems can be avoided by using DASH-based viewpoint adaptive transmission. In order to ensure smooth playback, the client needs to pre-download the video content, which requires the client to predict the future viewpoint of the user. Huang et al. [23] developed low-latency real-time video streaming technology based on HTTP 2.0. When encountering the available VR videos clips, the new server push feature of HTTP 2.0 is used to actively stream live video from the web server to the client, and the low-latency mechanism based on server pushes is implemented in the MPEG dynamic adaptive HTTP streaming (DASH) prototype. Nguyen et al. [21] introduced an efficient adaptive VR video stream method over HTTP/2 that is based on the DASH transport architecture, which uses stream prioritization and stream termination. In order to ensure adaptability, the 360-degree VR video is divided into multiple faces, with each face divided into time segments. VR video is also stored on the server at different quality levels.

2.2.2. Transmission Scheme Based on Tile and View Switching

The main viewpoint code stream is usually dynamically switched according to the user's perspective, which can remove the single perspective redundancy and reduce the bandwidth demands. With the design of VR video transmission schemes based on tile and view switching, the codec scheme is usually closely related [24]. In tile-based streaming, the panoramic image is divided into multiple tiles at the encoding end, and each tile has a different bitrate, which is then encoded into a different stream. This allows for tiles to cover the user's viewport (e.g., what is displayed on the device) while maintaining high quality. This ensures high quality, while the tile covers the user's viewport, while other tiles are of lower quality. One implementation is a tile-based streaming framework [25–28].

Zare et al. [29] divided the VR panoramic image of the video encoding end. Multiple tiles are encoded as streams of different qualities. The media streams of different resolutions and code streams are dynamically switched in network transmission according to user view information. At the video decoding end, a high quality mixed image of the main view and low quality background is combined. Petrangeli et al. [30], for tile-based streaming, divided the panoramic image into multiple tiles at the encoding end, and each tile again has a different bitrate, which is then encoded into a different stream. This allows for tiles to cover the user's viewport while maintaining high quality. Only tiles belonging to the viewport (the video area viewed by the user) are streamed at the highest quality; the other tiles are streamed at a lower mass. The authors also proposed an algorithm for predicting future viewport locations and minimizing quality transitions to viewport changes. Hosseini et al. [31] spatially partitioned the underlying three-dimensional (3D) mesh into multiple 3D sub-grids and constructed an efficient 3D geometric mesh, called

hexaface sphere, to best represent the tiled 360-degree VR video in 3D space. The 360-degree encoding was spatially divided into multiple tiles during encoding and packaging, and tiles in the field of view (FoV) were prioritized for view-aware adaptation. Xavier et al. [32] defined the concept of tile and tiled partitions in order to extend their models to tiled versions. A tile is a set of contiguous regions, and a tile is a set of non-overlapping tiles that are overlaid. In the tiling scheme, the service provider can generate a version of each tile without providing the entire video. In this case, unlike the case where the service provider decides which video version to generate, the client needs to select each tile version individually in order to use the tile version to generate a model of the visual content that was represented by the tile. Kashyap et al. [21] stated that, when viewing content using head-mounted display (HMD), a subset of the entire 360-degree video is displayed at a single point in time. Viewport-based encoding is required in order to improve the resolution and image quality of the displayed content. They proposed multi-resolution versions with equal resolution and cubemap projections by studying various viewport-related projection schemes.

2.2.3. The Progress of Viewport-Tracking Optimization

In the case of VR video, since users can usually only view the scenes in the viewport, most of the current types of transmission solutions are therefore designed to reduce bandwidth waste and improve transmission efficiency by transmitting the current and predicted viewport corresponding screens, instead of transmitting the complete panoramic content. There is currently an increasing interest in viewport-driven transmission optimization methods [25–27,29,33,34]. The viewport-driven approach combines transmission with video encoding in such a way that the viewports that are of interest to the user will be transmitted in a high quality manner, while other areas will be encoded in a low quality manner or not transmitted at all. Some work has also been done in order to accommodate slight viewport movement by taking the nearest viewport and rescaling it to a large region [29,34], but, if the viewport presence moves too much, it can still miss the live viewport [35]. In order to address this problem, many viewport prediction schemes [22,36–38] have been developed to infer a user's viewport from historical viewport movement [35], cross-user similarity [39], or deep content analysis [40].

In addition to predicting viewport location, studies [41] have predicted new quality determinants (viewport movement speed, luminance, and the degree of freedom (DoF)) by borrowing ideas from previous viewport prediction algorithms (e.g., history-based prediction).

2.3. The Main Challenges Facing VR Networking

In order to overcome the challenges facing VR networking, both academia and industry are now seeking more efficient approaches to compensate for the gap between the user experience with VR applications and limited network capacity. At present, various VR terminals can only provide a simple and limited experience, and the overall effect is unsatisfactory. We can classify the VR networked optimization approaches into three types, depending on the challenges when VR is networked, as described below.

2.3.1. VR Network Computing Power Challenges

The network applications of VR devices are placing unprecedented demands in computing power, especially the VR information processing process given the system's intensive computing power requirements. Studies [42] have shown that VR information processing requires computationally intensive tasks, such as scene depth estimation, image semantic understanding, 3D scene reconstruction, and high realism rendering, to be completed in real time in order to guarantee that users have a natural and smooth experience. The processing latency of VR is determined by the computing power of the computing nodes and the computational volume of the task. Although some of the remote cloud servers can relieve some of the computational pressure, it cannot guarantee latency perfor-

mance. Research [43] showed that, despite the limited computational power, significant challenges are facing the endpoints. Future mobile networks will integrate mobile edge computing (MEC), VR processing, and transmission problem analysis at different levels of the base station in order to accommodate intensive computation.

As mentioned above, future mobile networks will integrate MEC nodes to provide computing and storage functions, which can increase mobile VR content and provide multiple aspects of mobile VR services that are close to the user, and it can effectively address the challenges in terms of the multi-level computing capabilities of mobile terminals and future 5G mobile network VR computing needs.

2.3.2. The Challenge of VR Network Communication Efficiency

Network requirements are another crucial problem facing VR. When considering the limitations of the computing and rendering functions of VR mobile devices, to provide a higher quality user experience, computing-intensive tasks are usually delivered to the cloud/network edge server to improve performance. The MEC paradigm can further lower the communication delay for VR applications. Because of the high cost of deploying edge computing systems, the infrastructure has not yet been popularized in current 3G/4G mobile networks. Some studies [44–46] have shown that MEC and D2D technologies will be based on adaptive and scalable computing and communication paradigms to more flexibly promote service provision on mobile VR applications.

### 2.3.3. The Challenge of VR Network Service Latency

In the previous section, we established that edge computing will play a prominent part in end-to-end design in order to help address prospective long end-to-end delays. Both computational (image processing and frame rendering) and communication (queuing and over-the-air transmission) latency are major bottlenecks for VR systems. The human eye experiences accurate and smooth motion at low (less 20 ms) motion-to-photon (MTP) delays [47–49]. High MTP values send conflicting signals to the vestibular-ocular reflex (VOR), which may lead to dizziness or motion sickness. Currently, it will be critical to provide deterministic, low-latency communication services due to the stringent requirements for real-time communication and the low tolerance for delay jitter, especially at the wireless edge. Efficient VR transmissions while using radio communications at the edge of the network, including computing and caching resources, have been developed. Edge caching and edge computing are considered to be key technologies, and they significantly improve performance in 5G networks. Recently, a study [49] used radio communication, computational, and cache resources at the edge of the network for efficient transmission of VR. The quality of the web immersion experience and its dependence on various system/network/client aspects are also critical. This will require the consideration of user navigation patterns, as opposed to traditional quality measures that only consider the fidelity of the reconstructed data. In this context, interactivity (latency) poses an even greater challenge than providing a large amount of data to the user.

2.4. Enabling Technologies for VR

Some advanced technologies are emerging in order to meet the basic requirements of VR and provide performance improvement approaches.

2.4.1. Future Network System Architecture

Some studies [50,51] have indicated that future network architecture will provide some useful benefits for the development of VR. Especially for VR applications, the characteristics of information-centric networking (ICN) can support local multicast and multipoint-to-multipoint communication semantics, and they can combine more common blocks in order to form the required perspective. Zhang et al. [52] proposed a VR video conferencing system that was built on top of named data networking (NDN).They propose a framework that is shown in Figure 3.
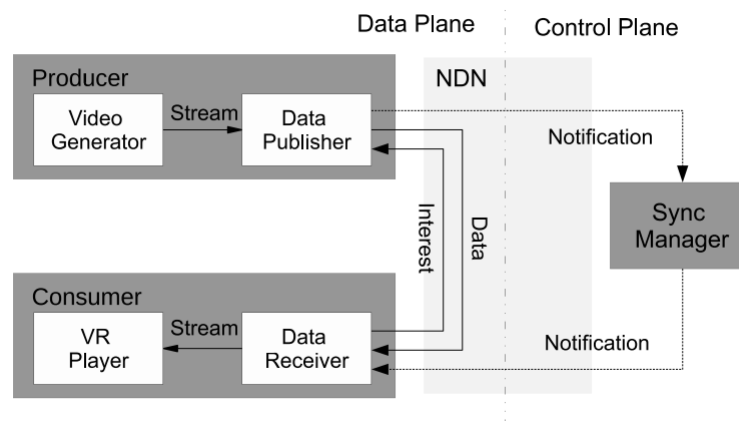
Figure 3. Framework of the named data networking (NDN) virtual reality (VR) conferencing system.

2.4.2. VR System Design

Some advanced technologies have emerged in order to meet the basic requirements of VR and provide performance improvement methods. Several practical implementations of VR video streaming systems have been proposed. Recently developed VR systems include FlashBack [53], Furion [54], and LTE-VR [55], and Flare [37], to name a few.

1. FlashBack [53]: Boos et al. proposed FlashBack in order to solve the problem faced by products, such as Google Cardboard and Samsung Gear VR, in providing VR with limited GPU power, which cannot produce acceptable frame rates and delays. FlashBack proactively pre-computes and caches all possible images that VR users may encounter. Record rendering works in offline steps to build a cache full of panoramic images. FlashBack constructs and maintains a tiered storage cache index at runtime in order to quickly find images that a user should view. For cache misses, a fast approximation of the correct image is used, while more closely matched entries are fetched from the cache for future requests. In addition, FlashBack is not only suitable for static scenes, but also for dynamic scenes of moving and animated objects.

2. Furion [54]: to enable high quality VR applications on unrestricted mobile devices such as smartphones, Lai et al. introduced Furion, a framework that enables high-quality, immersive mobile VR on today's mobile devices and wireless networks. Furion leverages key insights into VR workloads, namely the predictability of foreground interaction and background environments as compared to rendering workloads, and uses a split renderer architecture that runs on phones and servers. This is complemented by video compression, the use of panoramic frames, and the parallel decoding of multiple cores on the phone.

3. LTE-VR [55]: Tan et al. designed LTE-VR, a device-side solution for mobile VR that requires no changes to device hardware or LTE infrastructure. LTE-VR adapts the signaling operations that are involved in delay-friendliness. LTE-VR can passively use two innovative designs: (1) it adopts a cross-layer design in order to ensure rapid loss detection and (2) it only has rich side-channel information available on the device to reduce VR perception delays.

4. Flare [37]: Qian et al. designed Flare, a practical VR videos streaming system for commodity mobile devices. Flare uses a viewport adaptive method: instead of downloading the entire panoramic scene, it predicts the future viewport of the user and only obtains the part that the audience will consume. When compared with the prior methods, Flare reduces bandwidth usage or improves the quality of acquiring VR content of the same bandwidth. In addition, Flare is a universal 360-degree video streaming framework that does not rely on specific video encoding technologies.

3. Different VR Networking Optimization Approaches

In order to break the challenges when VR is networked, both academia and industry are now seeking more efficient melioration approaches to compensate for the gap between the user experience with VR applications and limited network capacity. At present, various VR terminals can only provide a simple limited experience, and the overall effect is unsatisfactory. We can classify the VR networking optimization approaches into three types, depending on the challenges that are experienced when VR is networked, as follows.

3.1. User-Centric Design for Edge Computing

VR is a computing and data-intensive application. Computing and rendering tasks in VR require an efficient runtime environment due to the limited computing and storage capabilities of mobile devices. The MEC server calculates all of the corresponding blocks as target tasks and then delivers the entire task to the mobile VR device. Some studies have developed communication-constrained MEC frameworks for wireless VR to minimize communication resource consumption, while considering a trade-off between communication, computation, and caching (3C) task scheduling strategies. In order to avoid an excessive amount of directly transmitted VR content data, to ensure the realtime transmission of mobile VR, the cloud server can usually perform the preliminary rendering of the VR content, and then the mobile VR device can perform the secondary rendering. The ability to use VR mobile devices is also one of the mainstream research directions for future wireless VR transmission systems. The Juniper [56] argued that the demand for VR video produce more data than the demand for 4K video; therefore, faster data transfer speeds are needed in order to efficiently transmit VR video content. Liu et al. [57] argued that the MEC architecture can help with solving the problem of inadequate computing power of mobile VR devices, but the growth rate of mobile VR content data far exceeds the growth rate of wireless network capacity, and transmitting VR video while using the current MEC architecture will result in a huge communication load. Numerous studies [44,45,58,59] have shown that MEC architectures can be used to improve network responsiveness and reduce latency, and we can try to save communication resources by using the computational and caching resources on mobile VR devices. Other studies [60,61] considered the integration of edge computing and mmWave in mobile VR, although the contribution is limited. Perfecto et al. [60] investigated a user clustering strategy in order to maximize user field-of-view frame requests. Elbamby et al. [61] studied active computation and caching of interactive VR video frames with the constraint of minimizing the traffic of VR games. However, these approaches are heuristic and they only consider low quality/low resolution (4K) 360-degree content, and these shortcomings significantly impact the quality of the delivered experience.

3.2. Optimization of Node-Related Associations

With the long-term evolution (LTE) network being gradually replaced by fifth-generation (5G) networks, edge caching and mobile edge computing are bringing content and computing resources closer to users. The 5G networks need not merely continue to increase the capacity and efficiency of network functions: it is necessary to directly integrate computing resources into the communication network. The key technology edge caching capabilities and edge computing capabilities have been implemented in 5G networks with significant performance gains.

3.2.1. Optimization Based on Caching

Caching will significantly impact VR performance. One study [62] considered optimizing the parameters of a single base station buffer, and others [63,64] studied hierarchical buffering in cellular backhaul networks. In [65,66], the information theory of hierarchical caching was studied, which simultaneously runs on client devices (personalized view caching), edge, and cloud, and it may require novel multi-level caching architectures. In particular, when caching is pushed to the edge, the traditional understanding of mass

data caching methods may no longer be applicable. In addition, instead of traditional caching methods, personalization- and viewport-driven strategies should be investigated in order to capture the spatial and temporal localization caused by user navigation of VR data. Likewise, we must understand how the interaction of virtual and physical functions in such applications affects caching, which is another new source of expected data location that can be exploited. A number of problems related to caching in VR systems has been studied [61,67–69]. In some studies, existing caching techniques were used to leverage various lateral information, such as user location, personalization characteristics, mobility patterns, and social relationship attributes, in order to determine what content to cache and where to cache it, improving the efficiency of accessing content servers on request.

3.2.2. Optimization Based on Access Control (AC) Scheduling

The streaming of 360-degree video collaboration from node-related AC scheduling to wireless VR clients is a novel topic. Closely related areas include multi-camera wireless sensing for multi-view systems [70], immersive remote collaboration [71,72], multi-view video encoding/communication [73–75], and individual 360-degree video Internet streams [25,35]. Existing work on wireless base station caching includes [76], which considered the problem of estimating base station content popularity and minimizing total content retrieval latency, referring to the latter as the backpack problem [77]. Shanmugam et al. [78] considered the problem of using caching in wireless assistant nodes, which are small cell base stations with high storage capacity and low coverage, in order to reduce the latency of content delivery and distinguish between available assistants that are based on their proximity to the service nodes.

3.2.3. Optimization Based on Content Awareness

Most content-based prediction algorithms use significant detection and neural networks to understand the region of interest (ROI) of the VR content. When compared to traditional 360-degree video, the ROI for predicting 360-degree video is inherently different and more challenging, because 360-degree video is omnidirectional. It cannot meet the requirements of real-time video streaming. Accordingly, we delved into viewing behavior across users to understand video content. There are currently two main solutions: (1) the use of the strong correlation between the user browsing the contents to determine the future perspective area. Borji et al. [79] studied the prediction of content-related features and significant target detection for still images. (2) Another type of method is to make predictions that are based on the salient features of the video. Advanced machine learning techniques are often used and a variety of supervised learning methods are employed, including neural networks, in order to better perform feature extraction and prediction accuracy in gaze detection [80–82]. We think that it is intuitive to measure the user's head movements (i.e., viewports) and prefetch the tiles that the user will use. However, many challenges remain in designing such a system, the first of which is high responsibility. We should be highly responsive to fast-paced viewport changes and viewport prediction (VP) updates. Secondly, processing power must be reasonable. We need to design where the prediction is performed, and we may need to define the processing power of the device. Finally, time-varying matching is required. The time window of the viewport prediction accuracy limits the total time budget for the entire process flow.

3.3. VR Implementation Driven by QoE

For the transmission process of VR streaming, we think that a suitable mechanism would improve the VR transmission. Most of the VR studies have optimized 360-degree videos transmissions while using the QoE model. QoE research has provided important insights into the design and optimization of video streaming services. An appropriate QoE model can help video providers to determine how to partition and encode 360-degree video and provide a benchmark for network operators to design QoE-aware scheduling algorithms. The literature [23,83,84] provides in-depth studies of QoE-driven cross-layer

design schemes for scalable video DASH services, and proposes a cross-layer designs framework for joint optimization of application, Medium Access Control(MAC), and physical layer parameters. The framework provides efficient wireless resource allocation between different services, thereby maximizing the network resource use and user QoE.

Machine learning (ML) is used to predict bandwidth views and video streaming bitrate improvements, which can bridge the gap between streaming approaches in terms of objective and subjective QoE assessments. Table 1 provides a summary of the different works that define ML in order to improve QoE in video streaming applications. In [37,85], model (Recurrent Neural Network-Long Short-Term Memory(RNN-LSTM) & Logistic Regression-Ridge Regression(LR-RR)) optimization for bandwidth and viewpoint predictions to support QoE are investigated. In [86], a method (Reinforcement Learning (RL) model) of adapting to variable video streams is investigated and a two-stage model for QoE evaluation is proposed. In [87], a method (Markov Decision Process-Deep Learning (MDP-DL)) of adapting to variable video streams is investigated. The study in [88] aimed to address the quality variation that affects QoE. The deep reinforcement learning (DRL) model [89] considers eye and head movement data in the quality assessment of 360-degree videos. Other authors [90] proposed a Q-learning algorithm for adaptive streaming services in order to improve the QoE in variable environments.

Table 1. Machine learning (ML)-based approaches to improve the quality of experience (QoE).

Reference	Method	Sheme
[85]	RNN-LSTM	Predicted Viewpoint/Predicted Bandwidth
[37]	LR-RR	Predicted Viewpoint/Predicted Bandwidth
[86]	RL model	Improved Adaptive VR Streaming
[87]	MDP-RL	Improved Variable bitrate (VBR)
[88]	Post-decision state	Improved constant bitrate (CBR)
[89]	DRL model	Improved video quality
[90]	Q-Learning RL	Improved CBR

4. Open Research Challenges

The emergence of networked VR has helped to promote VR applications on a large scale. However, various obstacles are waiting for the proper technologies to be available and affordable. The popularity of networked VR has been inspiring. In this section, we detail these insights and provide some further discussions.

4.1. Construction of Mapping the Relationship between QoE and QoS

Typically, most existing studies only singularly improved QoS on the subscriber side, and networks and service providers need to understand the relationship between network conditions and VR service performance. QoE applications effectively and accurately map to the respective QoS network/communication system, which can ensure overall end-to-end operations. Therefore, a number of newly completed studies [91–94] on VR QoE evaluation methods have focused progressively on reflecting the functionality of the network. The performance of VR services can be considered to be an indicator for detecting and evaluating the network environment and for planning the network behavior in order to satisfy VR functions.

VR applications are sensitive to latency and throughput, and different types of VR applications (e.g., video on demand, live streaming, interactive VR) have different requirements (interactive VR applications are the most demanding). In order to help address potentially long end-to-end delays, edge computing support is important and end-to-end designs should emerge. Because of the stringent requirements for real-time communication and low tolerance for latency jitter, providing deterministic, low-latency communication services will be critical, especially at the wireless edge. In order to ensure a quality user experience, we need more measurements and analytics to quantify the QoE of different VR applications, and we need to map the QoE requirements to the QoS requirements.

To ensure the QoE provided to the user, more measurements and analysis are needed for characterizing the QoE that is associated with different VR applications. Similarly, it is important to examine the mapping of valid and accurate QoE application requirements to the respective network/communication system QoS requirements to ensure overall end-to-end performance.

4.2. Unified Data Set

Recently, the content and data sets of most 360-degree videos that promote repeated research have been published. In order to facilitate the fair comparison between different VR solutions, the existing data set mainly provides three aspects of information: audience demographic information, viewing behavior information, and video feature information.

- Viewer information: sex, age, experience with the device, and visual health status.
- Video features: capturing projection models, encoding bit rate, resolution, etc.
- Viewing behavior: content type, visual track, experience rating, etc.

Wu et al. [95] collected a HMD sensor tracking dataset that is composed of 48 users (24 women and 24 men) watching 18 sphere videos from five categories. The video categories include performances, sports, movies, documentaries, and talk shows. Lo et al. [96] watched 10 video clips with the Oculus Rift DK2 headset. The data set includes 10 video clips (one minute) from YouTube and the HMD sensor tracks collected data on 50 topics. The video that is used in the NTHU [96] dataset is different from the video in the THU [95] dataset. Corbillon et al. [97] used three Orah Live Spherical VR Camera 4i 360-degree cameras to collect data from four users (all women) between the ages of 27 and 34 to watch MVP 360-degree video, while using the Google Daydream controller to switch viewpoint data. All of the users had already watched single-viewpoint 360-degree videos. The collected data included recording the observer's viewpoint direction, as well as shifting the viewpoint and switching decision time. Upenik et al. [98] described a test platform that demonstrates subjective quality assessments for omnidirectional visual content. The experimental data that were obtained by the test bench included subjective ratings, stimulation time, and viewing direction trajectories. The software allows for the user's viewing direction and other data to be captured at a selected sampling rate. De Abreu et al. [99] described navigation patterns that were collected during 360-degree image viewing with a HMD. They collected viewport center trajectories (VCTs) of 32 participants for 21 omnidirectional images (ODIs) and propose a method for transforming the gathered data into saliency maps. The developed database and testbed are publicly available with this paper.

Problems exist with the available datasets: they lack a relatively uniform standard; different sources, different resolutions, and different contents make it difficult to fairly compare the results in the study. For example, when the user fixes the viewport on the static content, they will have a better viewing experience, but, when the user moves their head frequently, the same content may produce a worse experience. Because the data sets and experimental evaluation methods are interrelated, the lack of a common standard design increases the challenge to consistent and fair evaluation of other system designs. The current general slowdown in VR development makes publicly available benchmark datasets, source code, evaluation sets, and testbeds more attractive. We think that this is a good way to improve the repeatability and standardization.

4.3. The Evolution of 6-DoF VR Applications

When compared to traditional video, the era of VR video media has provided a large amount of information. There is considerable potential for future applications of six degrees of freedom, refocusing based on the viewer's line of sight and depth map, bringing the immersive environment closer to the real world. For 6-DoF, predicting the spatial location of the user in the virtual environment will allow the user to participate in the VR environment. The 6-DoF use case will allow the user to intuitively interact with a high quality immersive virtual world while moving freely within the virtual environment. This use case will need to meet the unique challenges that are associated with supporting high

data rates (400–600 Mb/s) and low latency (5–20 ms), and accurately locate the VR user. More importantly, new human perception requirements are introduced to VR applications, where traditional video services and 3-DoF VR video (i.e., 360-degree video) content can tolerate unstable QoS through jitter buffering, whereas rendering 6-DoF content requires real-time delivery with low interaction latency. The hardware and software with 6-DoF tracking are much more complex than with 3-DoF tracking. Current HMD studies on 6-DoF [40] on HMDs [100–103] have focused on the more realistic and immersive experience that they provide to the user. Some of these HMDs are equipped with eye-tracking and ultrasonic positioning sensors, and enable new applications, such as concave rendering, gaze movement, and refocusing.

The challenges facing acquiring, presenting, storing, and transmitting VR technology implementations are enormous. Today, an increasing number of organizations and companies are involved in the development of standards for immersive media events in order to facilitate innovation and progress in this work. International standards organizations will continue to play an important role in compression coding. Standards, such as OMAF [104,105], have largely enabled 3-DoF video, whereas the 6-DoF video still requires further development.

5. Conclusions

This paper presented a survey of the state-of-the-art research in the area of networked VR. We highlighted several challenges that are associated with VR representation principles, typical transmission mechanisms, and enabling technologies. We discussed existing networked VR optimization approaches and highlighted their advantages and shortcomings. This review also outlined open research directions for QoE-QoS modeling, dataset measurement, and the evolution of 6-DoF VR applications.

Author Contributions: This work was mainly performed by J.R. (conceptualization, investigation, methodology, data curation, formal analysis, project administration, resources, software, visualization, and original draft preparation) and was completed with the key contribution of D.X. (conceptualization, supervision, validation, manuscript review & editing and funding acquisition). All authors have read and agreed to the published version of the manuscript.

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

Abbreviations

The following abbreviations are used in this manuscript:

2D	2-dimensional
3D	3-dimensional
3C	Communication computation and caching
3G	The 3rd generation of mobile phone mobile communication technology standards
4G	The 4th generation of mobile phone mobile communication technology standards
4K	4K resolution
5G	5th generation mobile networks or 5th generation wireless systems
6G	6th generation mobile networks or 6th generation wireless systems
AC	Access control
AP	Access point
AR	Augmented reality
CBR	Constant bit rate
DASH	Dynamic adaptive streaming over HTTP
D2D	Device-to-device
DRL	Deep reinforcementlLearning
DoF	Degree of Freedom
GPU	Graphics Processing Unit
HEVC	High-Efficiency Video Coding
HMD	Head-Mounted Display

HTTP	HyperText Transfer Protocol
JVET	Joint Video Experts Team
ICN	Information-Centric Networking
ISP	Internet Service Provider
LR	Logistic Regression
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
MAC	Medium Access Control
MEC	Mobile Edge Computing
ML	Machine Learning
MPEG	Moving Picture Experts Group
MTP	Motion-To-Photons
NDN	Named Data Networking
QoE	Quality of Experience
QoS	Quality of Service
QVGA	Quarter VGA
OHP	OctaHedral mapping Projection
OMAF	Omnidirectional Media Application Format
RNN	Recurrent Neural Network
ROI	Region of Interest
RR	Ridge Regression
SSP	Segmented Sphere Projection,
TSP	Truncated Square Pyramid projection
VBR	Variable Bit Rate
VOR	Vestibulo-ocular reflex
VP	Viewport Prediction
VR	Virtual Reality
VVC	Versatile Video Coding

References

1. Peltonen, E.; Bennis, M.; Capobianco, M.; Debbah, M.; Ding, A.; Gil-Castineira, F.; Jurmu, M.; Karvonen, T.; Kelanti, M.; Kliks, A.; et al. 6g white paper on edge intelligence. *arXiv* **2020**, arXiv:2004.14850.
2. Knightly, E. Scaling WI-FI for next generation transformative applications. In Proceedings of the Keynote Presentation, IEEE INFOCOM, Atlanta, GA, USA, 1–4 May 2017.
3. Begole, B. Why the Internet Pipes Will Burst When Virtual Reality Takes off. 2016. Available online: https://www.forbes.com/sites/valleyvoices/2016/02/09/why-the-internet-pipes-will-burst-if-virtual-reality-takes-off/?sh=5ae310d63858 (accessed on 12 January 2021).
4. Sharma, S.K.; Woungang, I.; Anpalagan, A.; Chatzinotas, S. Toward tactile internet in beyond 5g era: Recent advances, current issues, and future directions. *IEEE Access* **2020**, *8*, 56948–56991. [CrossRef]
5. 3GPP. *System Architecture for the 5G System*; version 1.2.0:TS23.501[S]; Sophia Antipolis: Valbonne, France, 2017.
6. 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. [CrossRef]
7. Hu, F.; Deng, Y.; Saad, W.; Bennis, M.; Aghvami, A.H. Cellular-connected wireless virtual reality: Requirements, challenges, and solutions. *IEEE Commun. Mag.* **2020**, *58*, 105–111. [CrossRef]
8. Fan, C.; Lo, W.; Pai, Y.; Hsu, C. A survey on 360 video streaming: Acquisition, transmission, and display. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–36. [CrossRef]
9. He, D.; Westphal, C.; Garcia-Luna-Aceves, J.J. Network support for ar/vr and immersive video application: A survey. *ICETE* **2018**, *1*, 525–535.
10. Duanmu, F.; He, Y.; Xiu, X.; Hanhart, P.; Ye, Y.; Wang, Y. Hybrid cubemap projection format for 360-degree video coding. In Proceedings of the 2018 Data Compression Conference, Snowbird, UT, USA, 27–30 March 2018; p. 404.
11. Lin, H.C.; Li, C.Y.; Lin, J.L.; Chang, S.K.; Ju, C.C. An Efficient Compact Layout for Octahedron Format. 2016. Available online: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=2943 (accessed on 12 January 2021).
12. der Auwera, G.V.; Coban, H.M.; Karczewicz, M. Truncated Square Pyramid Projection (tsp) for 360 Video. 2016. Available online: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=2767 (accessed on 12 January 2021).
13. Zhou, M. A Study on Compression Efficiency of Icosahedral Projection. 2016. Available online: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=2719 (accessed on 12 January 2021).
14. Zhang, C.; Lu, Y.; Li, J.; Wen, Z. Segmented Sphere Projection (ssp) for 360-Degree Video Content. 2016. Available online: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=2726 (accessed on 12 January 2021).
15. Kuzyakov, E.; Pio, D. *Under the Hood: Building 360 Video*; Facebook Engineering: Menlo Park, CA, USA, 2015.
16. Kuzyakov, E.; Pio, D. Next-Generation Video Encoding Techniques for 360 Video and vr. *Facebook*, 2016. Available online: https://engineering.fb.com/2016/01/21/virtual-reality/next-generation-video-encoding-techniques-for-360-video-and-vr/ (accessed on 12 January 2021).

17. Domanski, M.; Stankiewicz, O.; Wegner, K.; Grajek, T. Immersive visual media—mpeg-i: 360 video, virtual navigation and beyond. In Proceedings of the 2017 International Conference on Systems, Signals and Image Processing (IWSSIP), Poznan, Poland, 22–24 May 2017; pp. 1–9.
18. Sullivan, G.; Ohm, J. Meeting report of the 13th meeting of the joint collaborative team on video coding (jct-vc). In *ITU-T/ISO/IEC Joint Collaborative Team on Video Coding*; ITU: Incheon, Korea, 2013.
19. MPEG. MPEG Strategic Standardisation Roadmap. In Proceedings of the 119th MPEG Meeting,Turin, Italy, 17–21 July 2017.
20. Monnier, R.; van Brandenburg, R.; Koenen, R. Streaming uhd-quality vr at realistic bitrates: Mission impossible. In Proceedings of the 2017 NAB Broadcast Engineering and Information Technology Conference (BEITC), Las Vegas, NV, USA, 22–27 April 2017.
21. Nguyen, D.V.; Tran, H.T.T.; Thang, T.C. A client-based adaptation framework for 360-degree video streaming. *J. Vis. Commun. Image Represent.* **2019**, *59*, 231–243. [CrossRef]
22. 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.
23. Huang, W.; Ding, L.; Wei, H.; Hwang, J.; Xu, Y.; Zhang, W. Qoe-oriented resource allocation for 360-degree video transmission over heterogeneous networks. *arXiv* **2018**, arXiv:1803.07789.
24. 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.
25. Corbillon, X.; Simon, G.; Devlic, A.; Chakareski, J. Viewport-adaptive navigable 360-degree video delivery. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017; pp. 1–7.
26. Gaddam, V.R.; Riegler, M.; Eg, R.; Griwodz, C.; Halvorsen, P. Tiling in interactive panoramic video: Approaches and evaluation. *IEEE Trans. Multimed.* **2016**, *18*, 1819–1831. [CrossRef]
27. 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.
28. Brandenburg, R.; van Koenen, R.; Sztykman, D. CDN Optimization for vr Streaming. 2017. Available online: https://www.ibc.org/cdn-optimisation-for-vr-streaming-/2457.article (accessed on 12 January 2021).
29. 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, 23–27 October 2016; pp. 601–605.
30. Petrangeli, S.; Swaminathan, V.; Hosseini, M.; Turck, F.D. An http/2-based adaptive streaming framework for 360 virtual reality videos. In Proceedings of the 25th ACM international conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 306–314.
31. 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.
32. Corbillon, X.; Devlic, A.; Simon, G.; Chakareski, J. Optimal set of 360-degree videos for viewport-adaptive streaming. In Proceedings of the 25th ACM international conference on Multimedia, New York, NY, USA, 23–27 October 2017; pp. 943–951.
33. Duanmu, F.; Kurdoglu, E.; Hosseini, S.A.; Liu, Y.; Wang, Y. Prioritized buffer control in two-tier 360 video streaming. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, Los Angeles, CA, USA, 25 August 2017; pp. 13–18.
34. Xie, X.; Zhang, X. Poi360: Panoramic mobile video telephony over lte cellular networks. In Proceedings of the 13th International Conference on emerging Networking Experiments and Technologies, Incheon, Korea, 12–15 December 2017; pp. 336–349.
35. 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, 3–7 October 2016; pp. 1–6.
36. Nasrabadi, A.T.; Mahzari, A.; Beshay, J.D.; Prakash, R. Adaptive 360-degree video streaming using scalable video coding. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1689–1697.
37. Qian, F.; Han, B.; Xiao, Q.; Gopalakrishnan, V. Flare: Practical viewport-adaptive 360-degree video streaming for mobile devices. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, New Delhi, India, 29 October–2 November 2018; pp. 99–114.
38. Xie, L.; Zhang, X.; Guo, Z. Cls: A cross-user learning based system for improving qoe in 360-degree video adaptive streaming. In Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 564–572.
39. Ban, Y.; Xie, L.; Xu, Z.; Zhang, X.; Guo, Z.; Wang, Y. Cub360: Exploiting cross-users behaviors for viewport prediction in 360 video adaptive streaming. In Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.
40. Fan, C.-L.; Lee, J.; Lo, W.-C.; Huang, C.-Y.; Chen, K.-T.; Hsu, C.-H. Fixation prediction for 360 video streaming in head-mounted virtual reality. In Proceedings of the 27th Workshop on Network and Operating Systems Support for Digital Audio and Video, Taipei, Taiwan, 20–23 June 2017; pp. 67–72.

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

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

43. Yanli, Q.; Yiqing, Z.; Ling, L.; Lin, T.; Jinglin, S. Mec coordinated future 5g mobile wireless networks. *J. Comput. Res. Dev.* **2018**, *55*, 478.

44. Dai, J.; Liu, D. An mec-enabled wireless vr transmission system with view synthesis-based caching. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), Marrakech, Morocco, 15–18 April 2019; pp. 1–7.

45. Dai, J.; Zhang, Z.; Mao, S.; Liu, D. A view synthesis-based 360° vr caching system over mec-enabled c-ran. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *10*, 3843–3855. [CrossRef]

46. Dang, T.; Peng, M. Joint radio communication, caching, and computing design for mobile virtual reality delivery in fog radio access networks. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1594–1607. [CrossRef]

47. Chen, M.; Saad, W.; Yin, C. Virtual reality over wireless networks: Quality-of-service model and learning-based resource management. *IEEE Trans. Commun.* **2018**, *66*, 5621–5635. [CrossRef]

48. Doppler, K.; Torkildson, E.; Bouwen, J. On wireless networks for the era of mixed reality. In Proceedings of the 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12–15 June 2017; pp. 1–5.

49. Ju, R.; He, J.; Sun, F.; Li, J.; Li, F.; Zhu, J.; Han, L. Ultra wide view based panoramic vr streaming. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network,Los Angeles, CA, USA, 25 August 2017; pp. 19–23.

50. Xylomenos, G.; Ververidis, C.N.; Siris, V.A.; Fotiou, N.; Tsilopoulos, C.; Vasilakos, X.; Katsaros, K.V.; Polyzos, G.C. A survey of information-centric networking research. *IEEE Commun. Surv. Tutor.* **2013**, *16*, 1024–1049. [CrossRef]

51. Zhang, L.; Afanasyev, A.; Burke, J.; Jacobson, V.; Claffy, K.C.; Crowley, P.; Papadopoulos, C.; Wang, L.; Zhang, B. Named data networking. *ACM SIGCOMM Comput. Commun. Rev.* **2014**, *44*, 66–73. [CrossRef]

52. Zhang, L.; Amin, S.O.; Westphal, C. Vr video conferencing over named data networks. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, Los Angeles, CA, USA, 25 August 2017; pp. 7–12.

53. Boos, K.; Chu, D.; Cuervo, E. Flashback: Immersive virtual reality on mobile devices via rendering memoization. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, Singapore, 26–30 June 2016; pp. 291–304.

54. Lai, Z.; Hu, Y.C.; Cui, Y.; Sun, L.; Dai, N.; Lee, H.S. Furion: Engineering high-quality immersive virtual reality on today's mobile devices. *IEEE Trans. Mob. Comput.* **2020**, *7*, 1586–1602. [CrossRef]

55. Tan, Z.; Li, Y.; Li, Q.; Zhang, Z.; Li, Z.; Lu, S. Supporting mobile vr in lte networks: How close are we? *Proc. ACM Meas. Anal. Comput. Syst.* **2018**, *2*, 1–31.

56. Juniper. Virtual Reality Markets: Hardware, Content & Accessories 2017–2022.[J/OL]. 2017. Available online: https://www.juniperresearch.com/researchstore/enabling-technologies/virtual-reality/hardware-content-accessories (accessed on 12 January 2021).

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

58. Liu, Y.; Liu, J.; Argyriou, A.; Ci, S. Mec-assisted panoramic vr video streaming over millimeter wave mobile networks. *IEEE Trans. Multimed.* **2018**, *21*, 1302–1316. [CrossRef]

59. Yang, X.; Chen, Z.; Li, K.; Sun, Y.; Liu, N.; Xie, W.; Zhao, Y. Communication-constrained mobile edge computing systems for wireless virtual reality: Scheduling and tradeoff. *IEEE Access* **2018**, *6*, 16665–16677. [CrossRef]

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

61. Elbamby, M.S.; Perfecto, C.; Bennis, M.; Doppler, K. Edge computing meets millimeter-wave enabled vr: Paving the way to cutting the cord. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018; pp. 1–6.

62. Bastug, E.; Bennis, M.; Kountouris, M.; Debbah, M. Cache-enabled small cell networks: Modeling and tradeoffs. *EURASIP J. Wirel. Commun. Netw.* **2015**, *1*, 1–11.

63. Erman, J.; Gerber, A.; Hajiaghayi, M.; Pei, D.; Sen, S.; Spatscheck, O. To cache or not to cache: The 3g case. *IEEE Internet Comput.* **2011**, *15*, 27–34. [CrossRef]

64. Ahlehagh, H.; Dey, S. Video caching in radio access network: Impact on delay and capacity. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Paris, France, 1–4 April 2012; pp. 2276–2281.

65. Karamchandani, N.; Niesen, U.; Maddah-Ali, M.A.; Diggavi, S.N. Hierarchical coded caching. *IEEE Trans. Inf. Theory* **2016**, *62*, 3212–3229. [CrossRef]

66. Maddah-Ali, M.A.; Niesen, U. Fundamental limits of caching. *IEEE Trans. Inf. Theory* **2014**, *60*, 2856–2867. [CrossRef]

67. Chakareski, J. Vr/ar immersive communication: Caching, edge computing, and transmission trade-offs. In Proceedings of the Workshop on Virtual Reality and Augmented Reality Network, Los Angeles, CA, USA, 25 August 2017; pp. 36–41.

68. Chen, M.; Saad, W.; Yin, C. Echo-liquid state deep learning for 360 content transmission and caching in wireless vr networks with cellular-connected uavs. *IEEE Trans. Commun.* **2019**, *67*, 6386–6400. [CrossRef]

69. Sukhmani, S.; Sadeghi, M.; Erol-Kantarci, M.; Saddik, A.E. Edge caching and computing in 5g for mobile ar/vr and tactile internet. *IEEE Multimed.* **2018**, *26*, 21–30. [CrossRef]

70. Chakareski, J. Uplink scheduling of visual sensors: When view popularity matters. *IEEE Trans. Commun.* **2014**, *63*, 510–519. [CrossRef]

71. Vasudevan, R.; Zhou, Z.; Kurillo, G.; Lobaton, E.; Bajcsy, R.; Nahrstedt, K. Real-time stereo-vision system for 3d teleimmersive collaboration. In Proceedings of the 2010 IEEE International Conference on Multimedia and Expo, Singapore, 19–23 July 2010; pp. 1208–1213.

72. Hosseini, M.; Kurillo, G. Coordinated bandwidth adaptations for distributed 3d tele-immersive systems. In Proceedings of the 7th ACM International Workshop on Massively Multiuser Virtual Environments, Portland, OR, USA, 18–20 March 2015; pp. 13–18.

73. Cheung, G.; Ortega, A.; Cheung, N. Interactive streaming of stored multiview video using redundant frame structures. *IEEE Trans. Image Process.* **2010**, *20*, 744–761. [CrossRef]

74. Chakareski, J.; Velisavljevic, V.; Stankovic, V. User-action-driven view and rate scalable multiview video coding. *IEEE Trans. Image Process.* **2013**, *22*, 3473–3484. [CrossRef]

75. Chakareski, J. Wireless streaming of interactive multi-view video via network compression and path diversity. *IEEE Trans. Commun.* **2014**, *62*, 1350–1357. [CrossRef]

76. Blasco, P.; Gunduz, D. Learning-based optimization of cache content in a small cell base station. In Proceedings of the 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, 10–14 June 2014; pp. 1897–1903.

77. Martello, S. Knapsack problems: Algorithms and computer implementations. *Wiley-Intersci. Ser. Discret. Math. Optim. Tion* **1990**, *6*, 513.

78. Shanmugam, K.; Golrezaei, N.; Dimakis, A.G.; Molisch, A.F.; Caire, G. Femtocaching: Wireless video content delivery through distributed caching helpers. *arXiv* **2011**, arXiv:1109.4179.

79. Borji, A.; Cheng, M.; Hou, Q.; Jiang, H.; Li, J. Salient object detection: A survey. *Comput. Vis. Media* **2019**, *2*, 117–150. [CrossRef]

80. Alshawi, T.; Long, Z.; AlRegib, G. Understanding spatial correlation in eye-fixation maps for visual attention in videos. In Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA, USA, 11–15 July 2016; pp. 1–6.

81. Chaabouni, S.; Benois-Pineau, J.; Amar, C.B. Transfer learning with deep networks for saliency prediction in natural video. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1604–1608.

82. Nguyen, T.V.; Xu, M.; Gao, G.; Kankanhalli, M.; Tian, Q.; Yan, S. Static saliency vs. dynamic saliency: A comparative study. In Proceedings of the 21st ACM international conference on Multimedia, Barcelona, Spain, 21–25 October 2013; pp. 987–996.

83. Hashim, F.; Nasimi, M. Qoe-oriented cross-layer downlink scheduling for heterogeneous traffics in lte networks. In Proceedings of the IEEE Malaysia International Conference on Communications, Kuala Lumpur, Malaysia, 26–28 November 2013.

84. Rubin, I.; Colonnese, S.; Cuomo, F.; Calanca, F.; Melodia, T. Mobile http-Based Streaming Using Flexible Lte Base Station Control. In Proceedings of the 16th IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks, Boston, MA, USA, 14–17 June 2015.

85. Zhang, Y.; Guan, Y.; Bian, K.; Liu, Y.; Tuo, H.; Song, L.; Li, X. Epass360: Qoe-aware 360-degree video streaming over mobile devices. *IEEE Trans. Mob. Comput. (Early Access)* **2020**, *1*, 1. [CrossRef]

86. Filho, R.I.T.D.C.; Luizelli, M.C.; Petrangeli, S.; Vega, M.T.; der Hooft, J.V.; 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]

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

88. 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.

89. 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.

90. 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.

91. Wang, Y.; Xu, J.; Jiang, L. Challenges of system-level simulations and performance evaluation for 5g wireless networks. *IEEE Access* **2014**, *2*, 1553–1561. [CrossRef]

92. Fei, Z.; Xing, C.; Li, N. Qoe-driven resource allocation for mobile ip services in wireless network. *Sci. China Inf. Sci.* **2015**, *58*, 1–10. [CrossRef]

93. Agiwal, M.; Roy, A.; Saxena, N. Next generation 5g wireless networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1617–1655. [CrossRef]

94. Wang, F.; Fei, Z.; Wang, J.; Liu, Y.; Wu, Z. Has qoe prediction based on dynamic video features with data mining in lte network. *Sci. China Inf. Sci.* **2017**, *60*, 042404. [CrossRef]

95. Wu, C.; Tan, Z.; Wang, Z.; Yang, S. A dataset for exploring user behaviors in vr spherical video streaming. In Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, 20–23 June 2017; pp. 193–198.

96. Lo, W.C.; Fan, C.L.; Lee, J.; Huang, C.Y.; Chen, K.T.; Hsu, C.H. 360 video viewing dataset in head-mounted virtual reality. In Proceedings of the 8th ACM on Multimedia Systems Conference, Taipei, Taiwan, 20–23 June 2017; pp. 211–216.

97. Corbillon, X.; Simone, F.D.; Simon, G.; Frossard, P. Dynamic adaptive streaming for multi-viewpoint omnidirectional videos. In Proceedings of the 9th ACM Multimedia Systems Conference, Amsterdam, The Netherlands, 12–15 June 2018; pp. 237–249.

98. 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.

99. Abreu, A.D.; Ozcinar, C.; Smolic, A. Look around you: Saliency maps for omnidirectional images in vr applications. In Proceedings of the 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), Erfurt, Germany, 31 May–2 June 2017; pp. 1–6.

100. Oculus Rift VR. Available online: https://www.oculus.com/rift/ (accessed on 12 January 2021).

101. Daydream, G. Available online: https://vr.google.com/ (accessed on 12 January 2021).

102. Samsung Gear VR. Available online: http://www.samsung.com/global/galaxy/gear-vr (accessed on 12 January 2021).

103. HTC Vive VR. Available online: https://www.vive.com/ (accessed on 12 January 2021).

104. Jeong, J.; Jang, D.; Son, J.; Ryu, E. 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 Island, Korea, 17–19 October 2018; pp. 581–586.

105. Jeong, J.; Jang, D.; Son, J.; Ryu, E. 3dof+ 360 video location-based asymmetric down-sampling for view synthesis to immersive vr video streaming. *Sensors* **2018**, *18*, 3148. [CrossRef] [PubMed]