IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 7 June 2023, accepted 25 June 2023, date of publication 7 July 2023, date of current version 12 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3292335

SURVEY

A Survey on Optimizing Mobile Delivery of 360° Videos: Edge Caching and Multicasting

MOATASIM MAHMOUD[1], STAMATIA RIZOU[1], ANDREAS S. PANAYIDES[2], (Senior Member, IEEE), NIKOLAOS V. KANTARTZIS[3], (Senior Member, IEEE), GEORGE K. KARAGIANNIDIS[3,4], (Fellow, IEEE), PAVLOS I. LAZARIDIS[5], (Senior Member, IEEE), AND ZAHARIAS D. ZAHARIS[3], (Senior Member, IEEE)

[1]Research and Development Department, Singular Logic, 14564 Athens, Greece
[2]VIDEOMICS Group, CYENS Centre of Excellence, 1016 Nicosia, Cyprus
[3]School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[4]Cyber Security Systems and Applied AI Research Center, Lebanese American University, Beirut 1102-2801, Lebanon
[5]School of Computing and Engineering, University of Huddersfield, HD1 3DH Huddersfield, U.K.

Corresponding author: Moatasim Mahmoud (mmoatasim@singularlogic.eu)

This work was supported by the European Union through the Horizon 2020 Marie Skłodowska-Curie Innovative Training Networks Program "Mobility and Training for Beyond 5G Ecosystems (MOTOR5G)" under Grant 861219.

ABSTRACT Recently, there has been an ever-growing demand for virtual reality (VR) and 360° video applications. Different from conventional 2D videos, 360° videos take users into an immersive experience by providing them with a navigable panoramic view. However, achieving adequate quality of experience (QoE) levels poses significant network challenges, especially in mobile delivery setups. Despite the tremendous improvements offered by 5G and beyond mobile networks, streaming 360° videos in a similar fashion to 2D videos is suboptimal, while scaling at high numbers questions the feasibility of the endeavor. This paper explores the utilization of caching and multicasting solutions for the mobile delivery of VR and 360° videos. First, an overview of immersive technologies and their distinctive characteristics is provided. Then, we discuss the network challenges associated with 360° videos and the role of implementing robust caching and multicasting schemes that exploit the unique features of 360° videos and capitalize on the correlations among end-users' viewports. Having established the foundations and challenges of 360° video streaming, we continue with a comparison of the state-of-the-art literature, while focusing on video streaming optimization aspects. We conclude our work by discussing the status and future research directions.

INDEX TERMS Edge caching, multi-access edge computing (MEC), multicast, optimization, video streaming, VR, 360° videos.

I. INTRODUCTION

In recent years, immersive technologies have been utilized for a variety of applications including education, healthcare, sports, and cultural events [1], [2], [3], [4]. Several augmented and virtual reality (AR/VR) equipment options exist in the market nowadays, which contributes to accelerating the adoption of AR/VR technologies, in the broader extended/mixed reality (XR/MR) era. In fact, over 10 million AR/VR headsets have been shipped in 2022 [5] with over 171 million VR users worldwide [6]. Meta Quest, formerly Oculus Quest, [7], HTC Vive VR headsets [8], and PSVR [9] are among the most reputable options for VR headsets. Microsoft HoloLens [10] and Lenovo's ThinkReality A3 [11] smart glasses are examples of commercial AR headsets. Available head-mounted displays (HMDs) differ in terms of resolution, field of view (FoV), frame rate, and other technical features giving users the option to choose according to their needs and preferences.

The associate editor coordinating the review of this manuscript and approving it for publication was Nurul I. Sarkar.

The rapid development of multimedia technologies and the increasing demand for new forms of immersive content make it difficult for mobile networks to meet the specifications required for such applications [12]. 360° videos tend to have significantly higher bitrate demands than 2D videos [13]. The spherical nature of VR 360° videos contributes to a large volume of data that needs to be stored and transmitted. The close distance between the display screen and the viewer's eyes (typically a few centimeters) necessitates the use of high spatial resolutions. For example, the task of delivering 8K resolution videos of 90Hz temporal resolution, suggested to ensure comfortable use, creates additional challenges for mobile network operators in terms of high data rates that translate to a huge burden on core networks. As a result, mobile network operators invest and resort to transformative solutions in their future designs that would allow them to address the projected massive network traffic.

In addition to the high data rates, VR and 360° videos are subjected to strict latency requirements. Quality of experience (QoE) in VR applications entails additional distinctive aspects to those required by traditional videos and other applications [14], [15]. Presence is one such key feature needed to provide a satisfying virtual experience. Presence implies a sensation of being physically present or located in the virtual world. On the other hand, immersion refers to the illusion of being enveloped by the virtual environment [16]. Realizing a present and immersive experience relies heavily on displaying convincing and realistic content that instantly responds to changes in users' viewing direction. A mismatch between users' movement and the movement of the VR environment causes a feeling of discomfort, dizziness, or nausea among other symptoms [17]. The term cybersickness or VR sickness is used to describe this phenomenon that is experienced by VR users during or after VR sessions. High motion-to-photon (MTP) latency, which is the time needed to display movements in the VR environment to the user, is a primary cause of cybersickness [18]. The specific QoE characteristics and their corresponding quality of service (QoS) requirements demanded by VR videos impose significant challenges in mobile communication networks. For example, stringent delay requirements, needed to provide a smooth and immersive experience, involve a recommended MTP latency of less than 20ms [19], [20].

Contemporary 5G and future 6G networks need to address the upcoming challenges by means of innovative network designs and smart content delivery algorithms. With the development of VR applications and their associated network demands, providing a reliable service that can keep up with the high QoS requirements for AR/VR content has triggered the need for drastic network improvements [21], [22], [23]. 5G networks provide gigabit speeds and millisecond latencies with new network functionalities that can help in handling immersive applications. However, the huge data traffic and ultra-low latency requirements necessitate not only new network architecture designs but also smart network management algorithms which consider the specific features of immersive videos.

Caching popular videos at a close distance from mobile users reduces the experienced latency and relieves core networks from handling unnecessary traffic. When multiple users are watching a common video, multicast transmission allows for the efficient use of the spectrum resources. In the case of 2D video streaming, many approaches have been proposed for the optimization of conventional video delivery [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. Nonetheless, more sophisticated algorithms are needed for the specific case of 360° videos.

Implementing new caching and multicasting schemes is essential for bringing VR 360° videos over wireless networks [34]. 360° videos are normally projected and divided into 2D rectangular tiles. At any specific moment, a VR user views only part of the overall tiles. Specifically, those comprising their current FoV. Moreover, analyzing user head movements while watching 360° videos reveals interesting patterns and correlations among users' viewports as humans tend to focus on specific engaging parts of their given scene. Consequently, designing caching and multicasting methods for 360° videos should take advantage of these criteria for optimal network utilization. Streaming tiles related to a user's FoV, for instance, reduces the required bandwidth significantly. Moreover, correlations between multiple users' FoVs can be leveraged for optimizing the available network resources. Such observations have been examined for designing efficient 360° video delivery algorithms in mobile networks, including mobile multicasting and caching.

This paper is dedicated to surveying the different approaches in optimizing the mobile delivery of VR/360° videos. In section II, we present some of the existing surveys and identify the unique contributions of our work. We then introduce spherical video projection methods, 360° video viewing behaviors, and network-related aspects in section III. In section IV, we present state-of-the-art 360° caching and multicasting solutions along with a detailed comparison. Finally, we discuss some key challenges and future research directions in section V.

II. EXISTING SURVEYS

Our work is motivated by the proliferation of VR-enabled applications that rely on the delivery of VR/360° video content in mobile networks. Optimized streaming techniques of conventional videos have been the focus of earlier surveys, as depicted in Table 1. In the case of immersive video applications, however, existing surveys focus on general aspects and challenges of 360° video streaming, as well as on network-level solutions, but fail to provide a comprehensive overview of the specific techniques adopted for optimizing the resources of mobile networks. In this context, we dedicated our work in this article to reviewing state-of-the-art VR and 360° video streaming studies with emphasis

TABLE 1. Summary of existing surveys.

Ref.	VR/360°	Caching	Multicasting	Optimization Techniques
[35]	◐	●	×	◐
[36]	◐	●	×	●
[37]	×	●	×	●
[38]	×	●	×	●
[39]	×	●	×	●
[40]	×	×	●	×
[41]	×	×	●	●
[42]	◐	◐	×	×
[43]	●	◐	×	×
[44]	●	◐	◐	×
[46]	◐	●	×	◐
Our survey	●	●	●	●

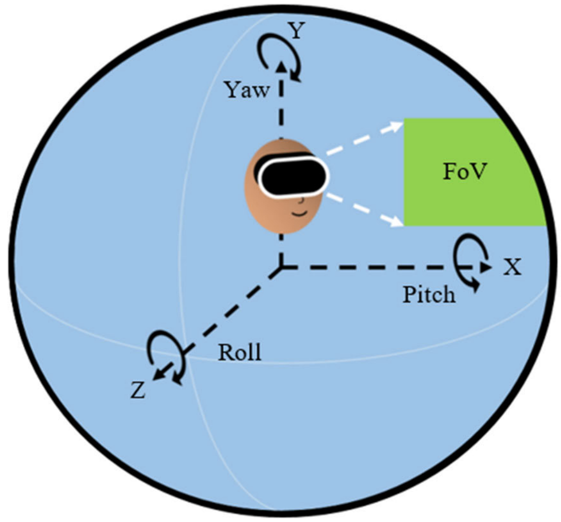

FIGURE 1. Watching 360° videos.

on the optimization techniques utilized in mobile 360° video caching and multicasting.

References [35] and [36] discuss the role of multi-access edge computing (MEC) in video streaming and survey state-of-the-art developments of edge video caching and processing. Although [35] and [36] bring some attention to the challenges presented by AR/VR and 360° video streaming, they do not explicitly focus on existing streaming solutions in this regard.

The articles [37], [38], [39], [40], [41] overview existing works in caching and multicasting of conventional videos. In [37], the authors survey popularity-based video caching methods in wireless networks. They analyze the presented works based on their adopted prediction algorithms, features, and performance evaluation metrics. [38] reviews machine learning (ML)-based edge caching strategies and makes an extensive comparison based on ML aspects, caching policies, and network scenarios. Similarly, [39] investigates the use of deep learning (DL) techniques for data caching in edge networks. Multicasting in mobile networks has also been a major focus of some survey papers. Reference [40] discusses the arising challenges for multicast applications in 5G networks, whereas [41] provides a comprehensive survey of multicast/broadcast schemes in mobile networks and makes an intricate classification of the presented works.

References [42] and [43] discuss edge caching and computing for AR/VR applications. Both papers, however, are brief and serve as an introduction to the topic in question. Reference [44] presents a survey on adaptive 360° video streaming. The survey focuses on end-to-end adaptive streaming based on MPEG-DASH [45]. Although they consider the roles of MEC, they are mostly focused on the network architectures of the presented works and do not delve into the employed optimization techniques. The authors in [46] overview network architectures and enabling solutions for video delivery, caching, and analytics at the wireless edge. Although the authors address some of the challenges and streaming solutions of 360° videos, an extensive review that is more dedicated to VR/360° videos is still lacking. Furthermore, the rapidly evolving merit of the topic necessitates a more up-to-date survey that covers the latest developments.

Therefore, this article is distinguished from earlier surveys by reviewing state-of-the-art 360° video MEC-based caching and multicasting techniques, while focusing on the optimization methods and goals adopted in the reviewed works. Table 1 summarizes the discussed surveys in this section and illustrates the contribution of our survey.

III. BACKGROUND

The objective of the present study is to capture the current landscape and identify the challenges concerning the mobile delivery of immersive videos. Thus, it is important to discuss the fundamental principles of 360° videos and their streaming properties, in addition to providing an adequate introduction to relevant mobile networking paradigms. The following subsections aim to introduce the necessary basic concepts.

A. VIEWING 360° VIDEOS

360° videos, also known as panoramic videos, are typically captured by omnidirectional cameras, or generated by stitching multiple views recorded simultaneously by a group of cameras. Users have the option of watching 360° videos on their mobile devices, personal computers, or on customized HMDs. They can navigate through the spherical view either by mouse clicks or by moving their heads when an HMD equipped with a head-tracking mechanism is used. In this case, users move around the spherical scene by simply moving their heads toward the desired direction as depicted in Figure 1.

VR provides users with a fabricated environment imitating real-life visual and audio scenes. An HMD is typically required to view VR content and to position the user in an immersive world. At any given time, the user's viewing direction, along with the HMD specifications, define the FoV to be rendered and shown to the user. The axes or directions over which users are allowed to navigate the videos are referred to as the degrees of freedom (DoF). In a 3DoF environment, three rotational movements (pitch, yaw, and roll) define the three axes which can be navigated by the user through their head movements while standing still or sitting on a chair. In a 6DoF scenario, the user can freely move around in physical space which yields three transitional movements in addition to the rotational head movements [47], [48].

B. 360° VIDEO CODING AND PROJECTION

As the size of panoramic videos highly exceeds that of traditional videos, developing new means for reducing the required streaming bandwidth is essential for providing 360° video content at a high scale. In practice, spherical scenes are typically transformed by mapping methods into other geometrical shapes or formats to benefit from advanced 2D video codecs such as Advanced Video Coding (AVC) / H.264 [49], High-Efficiency Video Coding (HEVC) / H.265 [50], and AV1 [51]. Versatile Video Coding (VVC) / H.266 [52] was launched in 2020 to claim the best compression efficiency to date and to accommodate a diverse set of applications, including 360° video streaming. VVC provides specific coding tools for facilitating 360° video streaming.

One of the simplest and most adopted mapping methods is the equirectangular projection (ERP). ERP has been used for map projection where Earth's longitude and latitude lines are flattened, along with the spherical shape of Earth, to form the vertical and horizontal lines in a 2D rectangular map [53]. When applied to 360° videos, the ERP mapping scheme results in suboptimal pixel distribution across the scene, as more information is given to the upper and lower points of the sphere (i.e., the poles) compared to other parts.

Cubemap projection tries to overcome this problem by reforming the spherical view into a six-sided cube where each side of the cube represents part of the original sphere. Cubemap reduces video size by 25% compared to ERP and reduces deformations that occur in the transformation process. However, Cubemap projection still causes a notable imbalance in pixel density. Hence, the equi-angular cubemap (EAC) was proposed to address this problem by adjusting sampling to be at uniform distances and assigning equal pixel densities regardless of the position on the sphere [54].

Pyramid projection is an example of a viewport-dependent projection scheme in which the mapping process relies on the requested FoV. Like Cubemap, pyramid projection stretches the entirety of the spherical scene and projects it inside of a pyramid-shaped view [55]. However, in pyramid format only currently watched FoV is rendered at full resolution (i.e., the bottom side of the pyramid). Other sides of the pyramid are still rendered to account for changes in FoV but at a gradually decreasing quality. Pyramid encoding is reported to reduce video size by 80%. Figure 2 illustrates the discussed projection methods in this subsection.

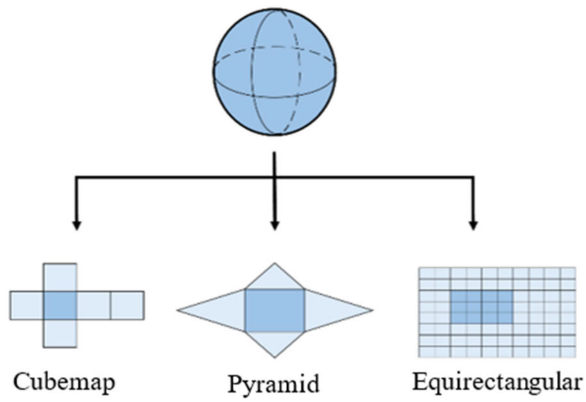

FIGURE 2. Common projection methods.

Modern video codecs divide videos into independent rectangular regions (i.e., tiles) which can be compressed in a parallelized manner. Tiling provides additional advantages when applied in panoramic video streaming. Streaming tiles outside users' FoVs results in a significant waste of bandwidth. Therefore, many recent works suggest segmenting the equirectangular projected scene into equal tiles and transmitting only tiles that are relevant to the user to avoid wasting resources on unwatched tiles [56]. In some scenarios, it is advisable to anticipate users' head movements by additionally streaming tiles that are adjacent to the current FoV to serve as a safe margin. Another approach is to transmit the whole panoramic scene at a lower resolution, in addition to the higher resolution FoV tiles. These approaches mitigate cases where users end up viewing black/empty regions and achieve higher QoE at the price of a slight increase in bandwidth usage.

C. USERS VIEWING BEHAVIOR AND PERCEPTION

Observing human behavior while watching 360° and VR videos is integral in designing new optimized streaming techniques specifically for this type of application. When analyzing people's viewing habits, we can identify patterns and correlations among their head movements, eye gaze positions, and/or FoVs. These patterns can be leveraged to optimize video transmission by anticipating users' interests and prioritizing certain video parts, and by designing the video compression and streaming strategies accordingly.

A dataset for viewing trajectories is provided in [57]. The dataset includes navigation trajectories, given as yaw and pitch values, of people viewing twelve 360° videos belonging to different categories. The authors provide an analysis of these traces and show that, depending on the video category, most people tend to view the same areas of the videos as their attention is directed towards specific objects in each scene. The analysis also shows that horizontal exploration of the videos (yaw) is more dominant than vertical exploration (pitch) and that the vertical direction is concentrated in central regions.

HMDs are capable of tracking head movements and viewers' eye gaze movements. Many works use previously collected viewing data, oftentimes assisted by information about the video content, to predict the future viewing point [58], [59], [60], [61], [62]. Having accurate predictions of viewers' future FoVs can help in taking proactive and efficient network decisions. Other works even consider directing users' attention in omnidirectional videos to guide them through the videos and make them focus on the most important parts for a more engaging experience. This can be achieved either by changing the user's viewing direction forcibly or by giving guidance to the user through visual or auditory hints [63], [64], [65]. These studies show that although automatic control results in faster redirection, it may cause discomfort and less feeling of presence for the users.

D. 360 VIDEO QUALITY ASSESSMENT

The endmost objective of a content delivery service is to provide the user with the best quality of experience possible. Oftentimes, streamed and downloaded videos reach users with a level of distortion caused during compression, transmission, and processing. This in turn results in the degradation of the perceived visual quality of the videos. Video quality assessment (VQA) methods provide objective metrics as an indication of the actual perceived video quality. Previously, full reference VQA primarily relied on the application of image quality assessment (IQA) metrics on a video frame-by-frame basis [66]. Nowadays, predicting perceptual quality using objective VQA metrics that integrate vital temporal (i.e., motion) aspects in quality assessment is an area of active research [67], [68], [69]. No-reference (i.e., blind) methods that utilize regularized natural scene statistics descriptors, such as normalized pixel intensities distribution, or employ machine learning to learn from a set of distortions, are increasingly used and likewise are emerging DL models [70].

The performance of VQA metrics is measured in terms of the correlation between the objective VQA scores and the perceptual ratings provided by human subjects as differential or mean opinion scores (DMOS/ MOS) [71], [66]. Similarly in medical VQA, the medical experts assess the diagnostic capacity of a compressed video, based on clinically established protocols [72].

In terms of 360° VQA, early methods such as spherical PSNR (S-PSNR) [73], Craster parabolic projection PSNR (CPP-PSNR) [74], and weighted-to-spherically-uniform PSNR (WS-PSNR) [75] focused on extending 2D IQA methods for 360° video quality assessment. Recent methods attempt to integrate content and visual attention in the assessment process [76], [77], [78], also leveraging DL methods [79], while emerging ones integrate the temporal dimension for a realistic and objective video evaluation [80], [81], [82], [83].

Understanding user perception of visual content plays a vital role in designing effective streaming algorithms. For instance, [84] provides a subjective study of human perception of omnidirectional images. As human eyes are most sensitive to quality in central regions, they fix the resolution of central regions and gradually reduce the resolution of peripheral parts of users' FoVs until they notice a drop in quality. At a tile level, the authors in [85] examine the effect of mixing tiles encoded at different quality levels. They show that considerable bandwidth savings can be achieved without affecting the perceived quality or at the cost of an acceptable drop in quality. These observations can be integrated for more flexible delivery solutions.

E. 360° VIDEOS DATASETS AND TOOLS

Datasets of 360° videos and user-collected data are required to test and verify the various techniques related to omnidirectional video delivery. Moreover, specialized tools are needed for processing and evaluating 360° videos. Thus, the development of 360° video delivery and the advancement of research in this field is subject to the availability of such software tools and datasets. In this subsection, we present the most prominent and useful open-source tools and datasets related to omnidirectional videos and 360° video transmission.

Real-life viewing data of 360° videos is essential for implementing and testing 360° video delivery methods in practical settings. Reference [86] provides a set of 28 omnidirectional videos and records the viewport traces of 60 subjects (17 female and 43 male) watching these videos using HMDs. Reference [87] similarly presents a dataset of head tracking information for 48 users watching 18 omnidirectional videos. In [88], both head motion and eye gaze information have been collected. The dataset contains 27 videos from different categories and the viewing data has been captured from one hundred participants at 120Hz sampling frequency. AVtrack360 [89] presents the head movement data collected from 48 users watching 20 videos, in addition to a software tool for recording the rotational angles when a user is watching a 360° video using an HMD.

Other datasets have been provided for the study of the quality assessment of 360° videos. Reference [90] presents a dataset of 60 reference videos in 4K to 8K resolutions. The dataset also contains 540 impaired videos resulting from mapping and compressing the reference videos. Three projection schemes and three compression levels have been used to produce the impaired videos. Moreover, 221 users participated in the collection of eye movement (EM) and head movement (HM) data, MOS, and DMOS scores. Reference [78] first provides a dataset of 48 videos and the recorded viewing directions of 40 users. The authors analyze the dataset to show that users tend to view content close to the equator more than other regions with high consistency of viewing behavior among different users. Based on their findings, they develop subjective and objective VQA methods that incorporate human viewing direction.

There are several available tools for processing and evaluating 360° videos. FFMPEG [91] offers a set of libraries and tools for encoding, manipulating, and streaming video content which can be applied to 360° videos. The dash-360 [92] and video2DASH [93] tools can be used for tiling and preparing 360° videos for DASH-based adaptive streaming. 360tools [94] provides tools for 360° video projection and quality evaluation. It allows for the calculation of 360° VQA metrics and the conversion between nine supported projection formats, which makes it a powerful tool for compression and VQA experiments. Likewise, Omnieval [73] offers conversion and quality assessment functionalities for omnidirectional videos, whereas Transform360 [95] can be used for transforming ERP to cubemap format with high efficiency in terms of memory usage and processing speed.

F. RELEVANT NETWORK CONCEPTS

Recent advancements in mobile communications paved the way toward the realization of reliable mobile XR and mobile 360° video streaming. 5G and beyond systems offer exceptional data speeds and ultra-low latency. Several technologies and network paradigms have contributed to this, including device-to-device (D2D) communications, small cells, and massive multiple-input multiple-output (MIMO) transmission [96], [97]. Moreover, the adoption of millimeter-wave (mmWave) communications utilizes the huge bandwidth resources available at the mmWave spectrum, allowing for the beamforming of signals [98].

MEC systems bring computational and storage capabilities closer to the mobile user, i.e., to the network edge. Rather than traveling all the way to a far centralized cloud, data can be fetched from a nearby MEC server, reducing the end-to-end latency and offloading data traffic from core networks [99], [100]. The deployment of MEC servers can be performed at base stations (BSs) or access points (APs) to achieve high benefits for mobile users and network operators. MEC is considered a key enabler technology in 5G networks for delivering new services, including panoramic videos and various AR/VR applications.

Conventional video streaming methods employ edge caching by placing popular videos in the proximity of end users. By anticipating users' requests for popular videos, optimized caching schemes result in reducing backhaul network traffic and delivery delay [24], [25], [26], [27], [28]. Moreover, video multicasting allows for the simultaneous delivery of content to a group of users watching a common video [101]. Multicasting avoids redundant transmission of the same videos by enabling point-to-multipoint transmission. Many algorithms have been proposed for optimizing the available resources using efficient multicast/unicast decisions and multicast grouping [29], [30], [31], [32].

IV. DELIVERING 360° VIDEOS IN MOBILE NETWORKS

Several approaches have been investigated to address the challenges of delivering high-quality next-generation 360° videos on mobile networks. These efforts have focused on lessening the burden on the backhaul network and on meeting the low latency requirements of these applications [102]. Segmenting 360° videos into tiles was suggested to support efficient VR transmission in cellular networks [103]. With the help of head movement prediction, transmitting relevant tiles, rather than whole videos, greatly reduces bandwidth consumption. Moreover, sending slightly larger viewports than the ones required by the users to mitigate imperfections in head movement prediction was found to increase QoE. The authors in [104] investigate the potential benefits of MEC-enabled FOV rendering. The authors address the task of delivering VR videos in mobile networks as a trade-off between latency, throughput, and computation. They show through experiments in a 4G long-term evolution (LTE) lab setup how rendering VR content at the network edge to only transmit the user's FOV could result in ∼80% bandwidth savings compared to transmitting the full 360° content to be rendered at the user's side. In the presence of multiple VR users, additional optimization opportunities arise from correlations among users' requests. Caching and multicasting schemes can utilize such correlations by storing popular tiles at nearby MEC servers or by grouping mobile users with common tile requests to be served simultaneously. Thus, achieving higher scalability while maintaining lower latency and optimized resource utilization.

This section presents state-of-the-art efforts in mobile caching and multicasting of 360° videos. We provide a discussion on the scenarios and optimization techniques used in each of the presented works and make a comprehensive comparison between them.

A. CACHING

Caching popular content in the proximity of mobile users (e.g., at base stations) allows for shorter delays and mitigates the need for sending the same content repeatedly over the backhaul network [105]. Many caching policies have been suggested for traditional video content to meet users' requests in MEC systems [24], [25], [26], [27], [28]. Although these caching schemes can be directly applied to 360° videos, the performance would be suboptimal. Therefore, it was necessary to develop caching algorithms that target the uniqueness of omnidirectional videos.

1) TILE-BASED CACHING

In the context of VR and 360° videos, information about users' FoVs can be leveraged for the development of smarter caching policies where only certain parts (e.g., subsets of the overall tiles) of popular videos need to be cached.

The authors in [106] propose a caching strategy that exploits the correlation between users' FoVs for determining which tiles need to be cached at high quality. In their work, a base station with caching capabilities serves users equipped with VR headsets. A user chooses a 360° video to watch and request tiles within their FoV in high or low resolution depending on the underlying network condition (i.e., available bandwidth). They also request tiles outside their FoV but only at the lower available resolution. The proposed caching policy learns from historical requests of the users to predict whether a certain tile in a video will lie within users' FoVs. The caching policy estimates the tile probability of potentially being requested, which translates to the need of being cached at the higher available resolution. The acquired information is thus utilized for cache replacement decisions upon the arrival of new requests. The provided simulation results show a significant improvement in cache hit ratio (CHR) and bandwidth savings, as well as in the number and duration of rebuffering events, compared to the legacy baseline approaches: least recently used (LRU) and least frequently used (LFU) [107].

An additional degree of caching flexibility can be introduced by considering the difference between requested and delivered qualities. A satisfactory tile resolution can be subjected to its position within the user's FoV. A drop in the quality of a tile at the FoV edge may not affect the user's QoE. Whereas the importance of center tiles or tiles which lie directly toward the viewer's gaze is much higher. The authors in [108] consider this viewing characteristic and aim at optimizing tile caching where users are expected to have different resolution needs across their viewports. They introduce a metric that portrays the difference between requested and delivered tile resolutions to be used in their optimization. A layered video scenario and a multiple-resolution video scenario are both considered separately and are formulated as a multiple-choice knapsack problem and as a k-medoids problem, respectively. The authors evaluate their quality-aware model against a popularity-based approach and demonstrate the improvements introduced by their method in terms of higher CHR and lower delivery time. Nonetheless, the overall delivered quality must also be assessed to study the trade-off introduced by the proposed method.

In [109], users' requests can be found in a common MEC cache either in 3D or 2D forms. In the first case, the 3D tiles are directly transmitted whereas in the latter case, the MEC server converts the requested 2D tiles into 3D before transmission. In case the requested tiles are not found on the MEC server, they are sent from a remote server to the MEC server. 2D and 3D caching decisions consider both the downloading delay and the MEC energy consumption resulting from 3D projections. The authors utilize the combinatorial multi-armed bandit (CMAB) theory for co-optimizing delay and energy consumption in a sequential decision-making process. They specifically develop an improved combinatorial UCB (CUCB) algorithm which considers the nature of VR content distribution. They compare the results of their improved algorithm to LRU, LFU, and another UCB-based algorithm. The presented simulation results show considerable improvements delivered by the proposed CUCB algorithm in the form of reduced delay and energy consumption.

Most 360° video caching schemes rely on dividing the video into multiple tiles. However, they usually assume fixed tiling in their implementation. Rather than dividing video frames into fixed tiles, dynamic tiling, with a varying tile size and location, proves to be a more flexible and bandwidth-efficient solution. To investigate the gains of applying variable tiling in MEC-based caching, OpCACH [110] has been proposed as an optimized caching strategy with dynamic tiling. In OpCACH, the overlapping between requested and cached tiles is found to decide which tiles should be retrieved from the content server. The authors formulate an integer linear programming (ILP) optimization problem aiming at finding the optimal cache tile configuration that minimizes the need for retrieving data from the content server and minimizes cached pixel redundancy. The simulation results show that OpCASH provides better cache utilization and reduces the overall delay compared to traditional methods.

2) COLLABORATIVE CACHING

The caching policies introduced by [106], [108], [109], [110] address the question of which tiles should be cached and at what qualities. In the presence of multiple cooperating cache entities, however, the question of "where" to cache those tiles also arises.

Reference [111] presents a collaborative approach between small-cell base stations (SBSs) to optimize caching, computing, and streaming of AR/VR content in cellular networks. In the adopted setting, the requested viewpoint content is rendered and served by the local base station or by a neighboring base station, or it can be retrieved from the internet. The authors consider the limited caching and processing capabilities of the SBSs in their optimization and formulate reward values to be maximized. The assigned rewards depend on the quality of the delivered videos and the place where the rendering/serving occurs (highest reward for local rendering/serving). Dynamic programming is used to provide an approximate solution for the formulated NP-hard problem. Simulation results confirm the benefits of the proposed algorithm as a significant reduction in energy consumption compared to LRU and non-cooperative caching. In multi-cell MEC 5G networks, collaboration among MEC servers for optimized caching can result in further latency reduction and efficient use of resources.

In [112], the authors propose a collaborative caching scheme for VR videos and couple it with the optimization of the rendering operation. They first perform tile popularity prediction based on video popularity and estimated saliency maps. Then they formulate the rendering-aware caching problem as a multi-knapsack problem under storage and computational constraints. The authors also propose a routing algorithm to deliver the requested tiles based on their cache location within the network to achieve maximized latency saving. Simulation results show significant latency reductions compared to non-cooperative caching and to solutions with decoupled rendering/caching.

Within the scope of cooperative tile-based caching, [113] discusses a scenario of multiple SBSs, each equipped with caching capabilities and connected to a content delivery network (CDN) through a common macrocell base station (MBS). In this work, each of the available videos is encoded at a base quality layer and at several enhancement layers which are used to construct higher resolutions of the requested tiles. Long-short-term memory (LSTM) is used for popularity forecasting of videos at the base layer and of each tile at higher resolutions. Each SBS performs popularity prediction and uses the acquired predictions to cache popular videos at the base layer and popular tiles at higher quality in place of the least popular in the cache. Three baseline methods are used for assessing this approach: LFU, LRU, and the first-in-first-out (FIFO) scheme. The evaluation demonstrates that the proposed caching strategy outperforms the simulated baseline methods in terms of delivered video quality and CHR while maintaining the least backhaul usage.

Another cooperative MEC caching and transmission strategy is proposed in [114]. The authors of this work consider a multi-MEC 5G architecture. They formulate the collaborative transmission and caching problem as a k-shortest paths problem and suggest a low-complexity optimized solution to the problem. The suggested solution results in low-latency transmission while maintaining high scalability.

3) PROACTIVE CACHING AND VIEWPORT PREDICTION

Proactive tile-based caching anticipates users' head movements and prefetches tiles with high viewing probabilities. Successful proactivity requires accurate long-term viewport predictions to achieve tangible latency savings. To this end, many 360° video caching frameworks employ DL models for predicting future viewpoints and tile popularities. These models are trained using previous viewing directions and sometimes are assisted by the actual video content.

A DL approach that makes use of the 360° video content represented by the saliency maps of the videos can be found in [115]. First, the authors justify their choice of using saliency maps by providing an analysis of users' behavior and viewing patterns while watching 360° videos. They suggest the use of two neural networks within their MEC caching scheme. An LSTM network is employed for video popularity prediction based on previous video requests. Saliency maps of popular videos are fed to a convolutional neural network (CNN). The CNN performs binary classification to decide which tiles within each video should be cached at the BS. Without assuming the popularity distributions, the proposed solution yields an improvement over the baseline methods in terms of CHR and backhaul usage and proves to be the closest to the optimal solution.

The authors in [116] propose another proactive caching scheme for VR videos, in which they leverage saliency maps and historical viewing orientations. In this work, saliency maps are obtained beforehand using a CNN+LSTM+ Gaussian mixture model (GMM). An LSTM prediction model at the MEC server uses previous viewing directions and video saliency maps to produce tile viewing probabilities. Tiles with the highest probabilities are hence proactively cached at the MEC server. An analysis of the proposed system performance is provided by applying diffusion approximation and is validated by means of computer simulation.

A proactive and collaborative VR caching strategy is suggested in [117]. In this work, interactive content generation and background delivery are performed independently. Interactive VR content contains dynamic data that changes depending on human interaction with the environment. Oftentimes, however, the scene background is stable and can be separated from the changing interactive content, and thus transmitted independently. For the low latency transmission of background content, the authors propose a proactive and collaborative caching algorithm that utilizes graph neural networks (GNN). The authors examine the effect of the proactivity period on users' request prediction accuracy. They show that their GNN model is more accurate than LSTM and motion-based prediction methods. Moreover, the evaluation of the proposed caching scheme shows significant improvement over LFU-based caching in terms of CHR and QoE.

4) RL-BASED DYNAMIC CACHING

Implementing optimized caching policies becomes more challenging as the system complexity increases. With the increased network heterogeneity associated with emerging network designs, researchers started investigating the utilization of reinforcement learning (RL) for making 360° video caching decisions. RL makes efficient decisions in a variety of dynamic environments and can be used to implement intelligent caching policies that adapt to the system intricacies. In [118], deep reinforcement learning (DRL) is utilized for implementing a transcoding-enabled 360° VR video caching and wireless transmission framework. The authors consider a collaborative case of a heterogeneous network, consisting of a centralized MBS and multiple SBSs. All BSs are assisted by edge caching and computing capabilities and employ non-orthogonal multiple access (NOMA)-based multicasting. To minimize the total computational and communication latency, the authors first employ multi-agent DRL to address the delivery preparation phase which involves transcoding-assisted caching decisions. After that, they address the multicasting subproblem in the delivery execution phase using a two-tier matching algorithm. A thorough performance evaluation of the suggested framework shows its effectiveness in terms of latency and CHR.

Another DRL-based strategy for 360° video caching and transcoding is proposed in [119]. A system with collaborating transcoding-enabled BSs is considered, where requested tiles can be delivered directly from a nearby home base station, other collaborating neighboring base stations, or from the CDN server. A deep deterministic policy gradient (DDPG) agent is trained to make collaborative edge caching and transcoding decisions with the goal of reducing delay, transmission cost, and quality mismatch levels. The proposed DRL

TABLE 2. Summary of works on VR/360° video caching techniques.

Ref.	Goals	Optimization Method	Collaborative /Multiple BSs	Multiple Videos	Multiple Resolutions
Tile-based caching (basic methods)					
[106]	Improve CHR	Probabilistic-model-based solution	✗	✓	✓
[108]	Minimize a quality distance metric	K-Medoids, Multiple-choice Knapsack	✗	✗	✓
[109]	Co-optimize latency & energy consumption	CMAB / CUCB	✗	✓	✓
[110]	Optimize pixel redundancy and compress cached tile size	ILP	✗	✓	✓
Collaborative caching					
[111]	Maximize an aggregate reward combining delivered quality and location of rendering/retrieval	Dynamic programming	✓	✓	✗
[112]	Optimize end-to-end latency	Multi-knapsack	✓	✓	✗
[113]	Maximize video quality and reduce backhaul usage	LSTM	✓	✓	✓
[114]	Reduce transmission latency	K-shortest paths	✓	✓	✗
Proactive caching &viewport prediction					
[115]	Maximize CHR	LSTM & CNN	✗	✓	✗
[116]	Minimize end-to-end latency	GMM & LSTM	✗	✗	✗
[117]	Maximize cache and minimize cache update cost	GNN	✓	✓	✗
RL-based dynamic caching					
[118]	Average latency minimization	Multi-agent DRL (Actor-critic algorithm)	✓	✓	✓
[119]	Reduce quality mismatch level, delay, and transmission cost	DRL (DDPG)	✓	✓	✓
[120]	Maximize users' reliability (meeting delay requirements)	Echo-liquid state DRL	✓	✓	✗
[121]	Optimize CHR, video quality, and rebuffering time	FDRL	✓	✓	✓
[122]	Minimize transmission rate requirements	DDPG	✓	✓	✗

CHR: cache hit ratio, LSTM: long-short-term memory, CNN: convolutional neural network, GMM: Gaussian mixture model, DRL: deep reinforcement learning, DDPG: deep deterministic policy gradient, GNN: graph neural network, ILP: integer linear programming, CMAB: combinatorial multi-armed bandit, CUCB: combinatorial upper confidence bound.

method surpasses LRU and a transcoding-enabled version of LRU as shown through the simulation results.

Other variations of RL have also been considered for different mobile communication settings. In [120], the authors study caching and transmission in cellular networks where unmanned aerial vehicles (UAVs) are the source of VR videos. Cache-enabled SBSs receive VR content from UAVs to serve mobile users' requests. An SBS can receive either full 360° videos or the visible contents of specific users. The authors develop transmission and caching algorithms that consider the processing time at the SBSs and at the UAVs to provide a reliable VR experience that meets the application delay requirements. They use a novel DRL approach that uses a liquid model and an echo state network (ESN) model, rather than feedforward neural networks (FNNs).

The authors in [121] employ federated deep reinforcement learning (FDRL) to provide a stable and effective solution. The work on jointly optimizing caching and bitrate adaptation of VR videos in hierarchical clustered MEC networks. An agent is trained to optimize a reward function that incorporates CHR, video quality, quality changes, and rebuffering time, in addition to bandwidth and transcoding costs. Performance analysis shows that FDRL results in improving CHR and QoE over other DRL-based algorithms.

In [122], the VR device and a set of serving APs are all equipped with caching and computing capabilities. Thus, a viewpoint-based cooperative computational offloading and caching strategy is proposed to minimize transmission rate requirements. The optimization problem is formulated as a Markov decision process which is then tackled using a deep DDPG algorithm. The provided simulation results show the benefits of the proposed method in terms of a significant reduction in the required transmission rates. In this work, even though video popularity is considered, the viewpoint popularity is assumed to be uniform across each video. User viewing data and statistical analysis can be applied for practical viewpoint popularity modeling that would help at improving the obtained results.

Table 2 summarizes the literature discussed in this subsection. It draws a comparison between 360° video caching techniques in terms of optimization methods and objectives, in addition to the adopted scenario in each work.

B. MULTICASTING

When multiple mobile users belonging to a common BS or AP are expected to be requesting the same content, multicasting of the data to these users can result in lower latency and noteworthy bandwidth savings while allowing for higher scalability.

1) TILE-BASED MULTICASTING

360° video tiles lying within the FoVs of multiple users can be sent via multicasting for more efficient transmission. [123], [124] take advantage of this fact and aim at optimizing the multicasting transmission of VR videos. In both works, the authors assume a single BS serving multiple users with random viewing directions that follow a Zipf distribution. In [123], they assume a time division multiple access (TDMA) system, whereas in [124], they adopt an orthogonal frequency multiple access (OFDMA) system in which the BS transmits the 360° video tiles to the users through multicasting on a set of available subcarriers. With the assumption of knowing the system channel state and users' FoVs, the authors formulate two non-convex optimization problems for each multiple-access scenario and provide their solutions. The first problem is to minimize the required energy for specified encoding rates. In the second problem, they assume a given energy budget and maximize the video quality. The benefits of applying tile-based multicast over the unicast approach are illustrated for both systems through simulation as the proposed optimized multicasting schemes result in lower power consumption and higher delivery rates. Although multiple quality levels are considered in these works, all users receive the tiles at the same quality levels. In practice, users experience different channel conditions and might have different quality preferences, which necessitates more flexible solutions that adapt to all users' conditions and needs.

The authors in [125] take users' heterogeneity into account and explore the transmission gains associated with mixing tiles of different resolutions. Mixing tiles of different, but close, quality levels within a viewer's FoV oftentimes does not degrade their perceived QoE. Given this, the authors consider two cases for tile quality variation: one where all the tiles within an FoV must be of the same quality level, i.e., absolute smoothness, and one where tiles can vary in quality within a specified range, i.e., relative smoothness. They also consider two playback modes where transcoding at the user side is and is not enabled. In the four resulting cases, the authors aim to utilize the available multicast chances associated with each case to achieve optimal wireless streaming of multi-quality 360° VR videos. In wireless TDMA systems where multiple users are connected to a single BS (or AP), they provide mathematical modeling to minimize energy consumption for the four cases. They consider transmission energy in the two cases where user transcoding is disabled whereas in the two user-transcoding-enabled cases they minimize a weighted sum of transmission and transcoding energy consumption. In the case of absolute smoothness without user transcoding, they transform the resulting non-convex problem into an equivalent convex problem and solve it using standard convex optimization techniques. In the other three cases, they follow the convex-concave procedure for obtaining the optimized solutions. The provided simulation shows that introducing more multicasting opportunities, which emerge from enabling relative smoothness and user transcoding, results in reduced energy consumption in wireless VR streaming.

In [126], a multi-user VR delivery system is introduced where the authors suggest a bit-assignment algorithm for tiles based on the available bandwidth. They propose a unicast/multicast tile delivery scheme that takes the decisions depending on the estimated viewing tiles and the similarities among users' views. Simulation of the suggested method shows an improvement in delivered video quality while avoiding redundant transmissions. In particular, the resulting viewport-normalized PSNR and the associated traffic rates were evaluated to demonstrate the benefits of the proposed solution compared to the uniform bit assignment scheme.

2) PROACTIVE MULTICASTING AND VIEWPORT PREDICTION

The previously presented works assume that the users' FoVs, or their estimated ones, are known to the system. However, the low MTP delay requirements for immersive applications motivated many researchers to incorporate viewport prediction models within their transmission frameworks [127], [128], [129].

In [127], motion patterns among users are exploited and temporal motion prediction is performed to assist the wireless multicasting of 360° videos. A neural network is trained to forecast motion over the three rotational axes based on current and previous viewing positions. Based on the acquired predictions, the wireless AP decides on which parts of the video should be sent through multicast or unicast transmission. The proposed multicasting scheme is designed to optimize the use of available bandwidth in wireless networks and achieves 50% bandwidth saving compared to the simple full-frame multicast.

Reference [128] suggests another proactive wireless multicasting framework where users' FoVs are predicted using a deep recurrent neural network (DRNN) based on gated recurrent units (GRUs). In the presence of multiple SBSs, the authors develop a multicasting scheme that clusters the users depending on their FoV overlaps and relative locations in a VR theater. SBSs mounted across the theater, operating in the mmWave band, transmit a set of VR 360° videos to the users. Each group of users is assigned to a specific SBS and a beam of the multi-beam transmission of that SBS. Scheduling the users for their multicast beams is formulated as a matching theory game to maximize viewing quality under low latency constraints. Simulation of the proposed solution shows considerable improvements in delivery rates compared to the unicast transmission while keeping lower delay values than the real-time multicast approach.

In [129], the authors use linear regression (LR) for producing tile weights estimations and assess the transmission based on the resulting viewport quality. Tiles of 360° videos are encoded using Scalable Video Coding (SVC) where each tile is encoded as a base layer and several enhancement layers. Based on the obtained tile weights estimations, resource blocks are assigned to multicast and unicast streams with the goal of maximizing the overall QoE, which is described as a weighted summation of the delivered tiles. A formulation of the optimization problem is provided using binary integer programming (BIP), and an algorithm is suggested to prioritize tiles with higher utility over cost ratio by transmitting them at higher layers. Simulation results demonstrate the benefits of the proposed method in terms of higher delivered viewport quality compared to the reference methods.

3) MULTICASTING IN INNOVATIVE 5G NETWORKS

As discussed in section III-F, the recent developments in 5G and beyond networks paved the way for a more reliable VR mobile transmission. Therefore, recent multicasting techniques consider emerging network paradigms in their designs. Although such paradigms create new multicasting opportunities, the increased system complexity makes the resulting multicast optimization problems more difficult to solve. In [130] for instance, a transmission mode selection algorithm is suggested for D2D-assisted 5G heterogeneous networks (5G-HetNets). In the adopted scenario, the mobile VR user receives VR content through a macro-cell broadcast, a small-cell mmWave unicast transmission, or through D2D multicast. To address the complexity of the problem, selection between the three transmission modes is performed using online reinforcement learning with the goal of optimizing the total system throughput. Furthermore, an unsupervised learning strategy is presented to construct the D2D clusters depending on users' locations. Evaluation of the proposed multicasting scheme shows a throughput increase over traditional broadcasting strategies while keeping adequate resource utilization levels.

The authors in [131] also study VR video delivery in 5G-HetNets and propose a multi-user crowd-assisted framework. Their adopted scenario consists of one MBS and multiple SBSs with edge computing capabilities and mobile users with enabled D2D communication among them. Contrary to many existing approaches that focus on transmission delay, the authors introduce a multicast method to optimize the users' average buffer level, which is more tied to viewers' QoE. Considering video transcoding at the base stations which is managed by the MBS, they propose a multicast-aware transcoding offloading algorithm that jointly optimizes the multicasting and transcoding tasks. Furthermore, they introduce a crowd-assisted delivery algorithm on the users' side to complement the discussed algorithm in the case of segment loss. When poor channel conditions exist between a user and a BS, the user requests the missing tiles from other users in their multicast cluster through D2D communication and drops its viewing resolution for future requests. The simulation of the proposed framework shows improvement in terms of throughput and latency compared to other existing solutions and demonstrates the highest QoE.

Among the wireless technologies adopted in 5G communications, beamforming techniques and the use of multiple antennas enhance system performance and capacity. In [132], the authors investigate the optimality of 360° VR wireless streaming in MIMO-OFDMA systems. Specifically, a multi-antenna BS that streams a multi-quality tiled 360° VR video to single-antenna mobile users. They consider two separate cases: with and without user transcoding. For the first case,

TABLE 3. Summary of works on VR/360° video multicast techniques.

Ref.	Goal	Optimization Method	Multiple BSs	Multiple Videos	Multiple Resolutions
Tile-based multicasting (basic methods)					
[123]	Minimize energy, maximize video quality	Non-convex optimization	×	×	✓
[124]	Minimize energy, maximize video quality	Non-convex optimization	×	×	✓
[125]	Minimize energy consumption	Convex optimization & convex-concave procedure	×	×	✓
[126]	Reduce redundant transmission and enhance video quality	Heuristic	×	×	✓
Proactive multicasting & viewport prediction					
[127]	Optimize bandwidth consumption	NN for motion prediction. A heuristic approach for unicast/multicast decisions	×	×	×
[128]	Maximize quality under latency constraints	DRNN for FoV prediction. Matching theory for scheduling	✓	✓	×
[129]	Overall QoE	LR for viewport prediction. BIP for resource allocation	×	×	✓
Multicasting in innovative 5G networks					
[130]	Optimize system throughput	Online RL	✓	×	×
[131]	Buffer-nadir maximization	Stochastic optimization	✓	✓	✓
[132]	Minimize total power	Mixed discrete-continuous optimization	×	×	✓
[116]	Average latency minimization	Matching theory	✓	✓	✓

QoE: quality of experience, NN: neural network, DRNN: deep recurrent neural network, LR: linear regression, RL: reinforcement learning.

they formulate the problem of minimizing the total transmission power as a mixed discrete-continuous optimization problem. Similarly, they formulate a two-timescale mixed optimization problem for minimizing the weighted sum of transmission and transcoding power in the case where user transcoding is enabled. In both cases, they provide general suboptimal solutions and globally optimal solutions under special circumstances. Simulation results demonstrate the benefits of the proposed algorithms in reducing power consumption and show that new multicast opportunities arising from enabling user transcoding result in reducing the overall power consumption.

Table 3 summarizes the literature discussed in this subsection. It makes a comparison between the multicasting schemes used with 360° videos in terms of optimization methods and objectives and delivery scenarios.

C. LESSONS AND RESEARCH DIRECTIONS

In this subsection, we provide potential research topics based on the discussion provided in the last two subsections and the limitations of the presented works. We identify some gaps in the literature that need to be addressed in the upcoming future to keep up with the challenges associated with 360° video delivery. We divide our research directions into five main points as follows:

1) QoE-BASED TRANSMISSION

By analyzing the caching and multicasting techniques in this section, we observe that most existing optimization methods target network-level metrics rather than optimizing the actual perceived quality. Instead of maximizing the delivered video rate, for instance, a more ambitious approach would be to maximize the corresponding objective 360° VQA metrics by leveraging the video content and viewport information in the optimization process. Although higher video rates typically result in better QoE, the mapping between the two is not straightforward. Hence, benefiting from the advancements in 360° VQA methods, more reliable solutions can be implemented. Quality assessment metrics give a better indication of end users' experience and can result in a better overall QoE. This will also be dependent on the development and verification of accurate objective VQA metrics for omnidirectional videos that reflect users' subjective quality.

2) PROJECTION SCHEMES

Although there are several projection methods available, most delivery methods assume ERP projection with fixed-size tiles. This is due to the simplicity of ERP compared to other projection schemes. However, serious research efforts must be undertaken towards investigating variable size tiling and the various existing projection methods for more flexible 360° video caching and multicasting schemes.

3) VIEWPORT PREDICTION

The importance of leveraging viewport information is undeniably crucial for implementing effective 360° video delivery frameworks. Analyzing human behavior while watching omnidirectional videos reveals interesting patterns that are used, with the help of viewport prediction, for making intelligent network decisions. Viewport prediction has become an integral part of 360° video streaming and the development of more accurate models that extend the prediction time would allow for higher proactivity. Moreover, implemented transmission methods must be able to identify cases of viewport prediction uncertainties or failures in order to address these cases and achieve an uninterrupted watching experience.

4) EMERGING VARIATIONS OF 360° VIDEOS

A variety of immersive applications that need researchers' attention have been appearing recently. While 3DoF 360° videos are intended to be watched from a fixed position, 6DoF VR provides users with additional navigation axes that correspond to the user's position in the physical environment. Volumetric videos provide a similar experience by representing the environment as 3D scenes and objects. Streaming of 6DoF and volumetric videos is even more challenging than 3DoF 360° videos as they require higher bandwidth and computational power [133], [134], [135]. Such videos are also more challenging in prediction and have a lower potential for exploiting common FoVs. In view of this, new caching and multicasting methods must be developed to handle the particular challenges of these applications.

5) 6G ERA

Several caching and multicasting approaches have been proposed in the literature for delivering omnidirectional videos in mobile networks. On the network architecture level, many of the existing works utilize the new developments in 5G networks, such as mmWave and D2D communications, which allow for implementing different innovative solutions. Recent technologies that have been proposed for advancing towards 6G communications, including THz communications and reconfigurable intelligent surfaces (RIS) [136], create new opportunities for VR and 360° video delivery. Therefore, upcoming delivery techniques of omnidirectional videos must aim at benefiting from these network changes for more efficient transmission.

V. CONCLUSION

The availability of advanced commercial HMDs enabled the widespread usage of VR and 360° videos in a variety of applications. At the same time, the high data rate demands and ultra-low latency requirements of omnidirectional videos pose significant challenges for mobile communications. In the case of conventional 2D video delivery, mobile edge caching, and multicasting are two effective networking paradigms, which hold great potential also for 360° video delivery. Consequently, the unique criteria of 360° videos inspired researchers to explore innovative caching and multicasting solutions that are specific to 360° video delivery.

In this work, we surveyed 360° video caching and multicasting techniques in mobile communication systems. First, we introduced the viewing mechanisms and criteria of 360° videos and presented prominent encoding and projection methods. A discussion was then provided on human perception and behavior while watching omnidirectional videos. Next, we analyzed state-of-the-art caching and multicast schemes with an emphasis on optimization methods and objectives, while also discussing the network architectures and scenarios in the surveyed works. Based on our analysis, we highlighted the limitations of current 360° video caching and multicasting methods and suggested some critical research areas that need further investigation.

Mobile networks need to accommodate the upcoming demand for omnidirectional videos and be ready to provide a satisfying experience for mobile users at a high scale. To achieve this, delivery methods must exploit the characteristics of 360° videos. Moreover, new transmission schemes must consider emerging mobile communication technologies and architectures, while leveraging the advancements in optimization and DL techniques.

REFERENCES

[1] C. Coates. (Jul. 31, 2021). Virtual reality is a big trend in museums, but what are the best examples of museums using VR? MuseumNext. [Online]. Available: https://www.museumnext.com/article/how-museums-are-using-virtual-reality/

[2] M. C. Hsieh and J. J. Lee, "Preliminary study of VR and AR applications in medical and healthcare education," *J. Nursing Health Stud.*, vol. 3, no. 1, p. 1, 2018.

[3] J. Pirker and A. Dengel, "The potential of 360° virtual reality videos and real VR for education—A literature review," *IEEE Comput. Graph. Appl.*, vol. 41, no. 4, pp. 76–89, Jul. 2021.

[4] (May 12, 2022). *NBA Lane Comes to Life through Virtual Reality*. [Online]. Available: https://www.nba.com/news/nba-lane-horizon-worlds-meta

[5] IDC. (Sep. 20, 2022). *AR & VR Headsets Market Share*. [Online]. Available: https://www.idc.com/promo/arvr

[6] C. Petrov. (Nov. 26, 2022). 45 virtual reality statistics that rock the market in 2022. Techjury. [Online]. Available: https://techjury.net/blog/virtual-reality-statistics/.

[7] *Meta Quest VR Headsets, Accessories & Equipment: Meta Quest (No Date) Meta*. Accessed: May 22, 2023. [Online]. Available: https://www.meta.com/quest/

[8] *VR Headsets, Games, and Metaverse Life: United States (No Date) VIVE*. Accessed: May 22, 2023. [Online]. Available: https://www.vive.com/us/

[9] *PlayStation VR: Live the Game With the PS VR Headset (No Date) PlayStation*. Accessed: May 22, 2023. [Online]. Available: https://www.playstation.com/en-us/ps-vr/

[10] *Microsoft Hololens: Mixed Reality Technology for Business (No Date) Microsoft HoloLens | Mixed Reality Technology for Business*. Accessed: May 22, 2023. [Online]. Available: https://www.microsoft.com/en-us/hololens

[11] *ThinkReality A3 Smart Glasses (No Date) Lenovo Official U.S. Site*. Accessed: May 22, 2023. [Online]. Available: https://www.lenovo.com/us/en/thinkrealitya3/

[12] C. Westphal, "Challenges in networking to support augmented reality and virtual reality," in *Proc. ICNC*, Jan. 2017, pp. 1–5.

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

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

[15] C.-L. Fan, T.-H. Hung, and C.-H. Hsu, "Modeling the user experience of watching 360° videos with head-mounted displays," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 18, no. 1, pp. 1–23, Jan. 2022.

[16] M. J. Schuemie, P. van der Straaten, M. Krijn, and C. A. P. G. van der Mast, "Research on presence in virtual reality: A survey," *CyberPsychol. Behav.*, vol. 4, no. 2, pp. 183–201, 2001.

[17] P. Caserman, A. Garcia-Agundez, A. G. Zerban, and S. Göbel, "Cybersickness in current-generation virtual reality head-mounted displays: Systematic review and outlook," *Virtual Reality*, vol. 25, no. 4, pp. 1153–1170, Dec. 2021.

[18] J.-P. Stauffert, F. Niebling, and M. E. Latoschik, "Latency and cybersickness: Impact, causes, and measures. A review," *Frontiers Virtual Reality*, vol. 1, Nov. 2020, Art. no. 582204.

[19] S. M. LaValle, A. Yershova, M. Katsev, and M. Antonov, "Head tracking for the Oculus Rift," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 187–194.

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

[21] F. Hu, Y. Deng, W. Saad, M. Bennis, and A. H. Aghvami, "Cellular-connected wireless virtual reality: Requirements, challenges, and solutions," *IEEE Commun. Mag.*, vol. 58, no. 5, pp. 105–111, May 2020.

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

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

[24] H. Pang, J. Liu, X. Fan, and L. Sun, "Toward smart and cooperative edge caching for 5G networks: A deep learning based approach," in *Proc. IEEE/ACM 26th Int. Symp. Quality Service (IWQoS)*, Jun. 2018, pp. 1–6.

[25] T. X. Tran, P. Pandey, A. Hajisami, and D. Pompili, "Collaborative multi-bitrate video caching and processing in mobile-edge computing networks," in *Proc. 13th Annu. Conf. Wireless On-demand Netw. Syst. Services (WONS)*, Feb. 2017, pp. 165–172.

[26] S. Yang, Y. Tseng, C. Huang, and W. Lin, "Multi-access edge computing enhanced video streaming: Proof-of-concept implementation and prediction/QoE models," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1888–1902, Feb. 2019.

[27] Z. Wan and Y. Li, "Deep reinforcement learning-based collaborative video caching and transcoding in clustered and intelligent edge B5G networks," *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1–16, Dec. 2020.

[28] X. Huang, L. He, L. Wang, and F. Li, "Towards 5G: Joint optimization of video segment caching, transcoding and resource allocation for adaptive video streaming in a multi-access edge computing network," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10909–10924, Oct. 2021.

[29] S. Sharangi, R. Krishnamurti, and M. Hefeeda, "Energy-efficient multicasting of scalable video streams over WiMAX networks," *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 102–115, Feb. 2011.

[30] J. Montalban, P. Scopelliti, M. Fadda, E. Iradier, C. Desogus, P. Angueira, M. Murroni, and G. Araniti, "Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 91–95, Mar. 2018.

[31] D. Jiménez-Soria, F. J. Martín-Vega, and M. C. Aguayo-Torres, "Coordinated multicast/unicast transmission on 5G: A novel approach for linear broadcasting," *Wireless Pers. Commun.*, vol. 121, no. 2, pp. 1273–1287, Nov. 2021.

[32] N. Chukhno, O. Chukhno, S. Pizzi, A. Molinaro, A. Iera, and G. Araniti, "Efficient management of multicast traffic in directional mmWave networks," *IEEE Trans. Broadcast.*, vol. 67, no. 3, pp. 593–605, Sep. 2021.

[33] G. Esakki, A. S. Panayides, V. Jalta, and M. S. Pattichis, "Adaptive video encoding for different video codecs," *IEEE Access*, vol. 9, pp. 68720–68736, 2021.

[34] E. Bastug, M. Bennis, M. Medard, and M. Debbah, "Toward interconnected virtual reality: Opportunities, challenges, and enablers," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 110–117, Jun. 2017.

[35] M. A. Khan, E. Baccour, Z. Chkirbene, A. Erbad, R. Hamila, M. Hamdi, and M. Gabbouj, "A survey on mobile edge computing for video streaming: Opportunities and challenges," *IEEE Access*, vol. 10, pp. 120514–120550, 2022.

[36] X. Jiang, F. R. Yu, T. Song, and V. C. M. Leung, "A survey on multi-access edge computing applied to video streaming: Some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 871–903, 2nd Quart., 2021.

[37] H. S. Goian, O. Y. Al-Jarrah, S. Muhaidat, Y. Al-Hammadi, P. Yoo, and M. Dianati, "Popularity-based video caching techniques for cache-enabled networks: A survey," *IEEE Access*, vol. 7, pp. 27699–27719, 2019.

[38] J. Shuja, K. Bilal, W. Alasmary, H. Sinky, and E. Alanazi, "Applying machine learning techniques for caching in next-generation edge networks: A comprehensive survey," *J. Netw. Comput. Appl.*, vol. 181, May 2021, Art. no. 103005.

[39] Y. Wang and V. Friderikos, "A survey of deep learning for data caching in edge network," *Informatics*, vol. 7, no. 4, p. 43, Oct. 2020.

[40] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over emerging 5G networks: Challenges and perspectives," *IEEE Netw.*, vol. 31, no. 2, pp. 80–89, Mar. 2017.

[41] D. Striccoli, G. Piro, and G. Boggia, "Multicast and broadcast services over mobile networks: A survey on standardized approaches and scientific outcomes," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1020–1063, 2nd Quart., 2019.

[42] M. Erol-Kantarci and S. Sukhmani, "Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G," in *Proc. 9th Int. Conf. Ad Hoc Networks (AdHocNets)*. Niagara Falls, ON, Canada: Springer, Sep. 2017, pp. 169–177.

[43] S. Sukhmani, M. Sadeghi, M. Erol-Kantarci, and A. El Saddik, "Edge caching and computing in 5G for mobile AR/VR and tactile internet," *IEEE Multimedia Mag.*, vol. 26, no. 1, pp. 21–30, Jan. 2019.

[44] A. Yaqoob, T. Bi, and G. Muntean, "A survey on adaptive 360° video streaming: Solutions, challenges and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2801–2838, 4th Quart., 2020.

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

[46] B. Jedari, G. Premsankar, G. Illahi, M. D. Francesco, A. Mehrabi, and A. Ylä-Jääski, "Video caching, analytics, and delivery at the wireless edge: A survey and future directions," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 431–471, 1st Quart., 2021.

[47] *5G; Extended Reality (XR) in 5G*, Standard ETSI TR 126 928, Version 16.0.0, ETSI, Nov. 2020.

[48] H. Baker. (May 18, 2022). Beginner's guide to VR: FAQ and everything you need to know. UploadVR. [Online]. Available: https://uploadvr.com/beginners-guide-vr-faq-everything-you-need-to-know/

[49] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

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

[51] P. de Rivaz and J. Haughton. *AV1 Bitstream & Decoding Process Specification*. Accessed: Jul. 7, 2023. [Online]. Available: https://aomediacodec.github. io/av1-spec/av1-spec.pdf

[52] B. Bross, Y. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J. Ohm, "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021.

[53] J. P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections*. Chicago, IL, USA: Univ. of Chicago Press, 1997.

[54] C. Brown. (Mar. 14, 2017). Bringing pixels front and center in VR video. Google. [Online]. Available: https://blog.google/products/google-ar-vr/bringing-pixels-front-and-center-vr-video/

[55] E. Kuzyakov and D. Pio. (Jan. 21, 2016). Next-generation video encoding techniques for 360 video and VR. Engineering at Meta. [Online]. Available: https://engineering.fb.com/2016/01/21/virtual-reality/next-generation-video-encoding-techniques-for-360-video-and-vr/

[56] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 261–271.

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

[58] P. Dong, R. Shen, X. Xie, Y. Li, Y. Zuo, and L. Zhang, "Predicting long-term field of view in 360-degree video streaming," *IEEE Netw.*, early access, Oct. 31, 2022, doi: 10.1109/MNET.106.2100449.

[59] H. Nguyen, T. N. Dao, N. S. Pham, T. L. Dang, T. D. Nguyen, and T. H. Truong, "An accurate viewport estimation method for 360 video streaming using deep learning," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 9, no. 4, p. e2, Sep. 2022.

[60] A. Nguyen, Z. Yan, and K. Nahrstedt, "Your attention is unique: Detecting 360-degree video saliency in head-mounted display for head movement prediction," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 1190–1198.

[61] C. Wu, "A spherical convolution approach for learning long term viewport prediction in 360 immersive video," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 14003–14040.

[62] J. Li, L. Han, C. Zhang, Q. Li, and Z. Liu, "Spherical convolution empowered viewport prediction in 360 video multicast with limited FoV feedback," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 1, pp. 1–23, Jan. 2023.

[63] Y.-C. Lin, Y.-J. Chang, H.-N. Hu, H.-T. Cheng, C.-W. Huang, and M. Sun, "Tell me where to look: Investigating ways for assisting focus in 360° video," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2017, pp. 2535–2545.

[64] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. M. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin, "Missing the point: An exploration of how to guide users' attention during cinematic virtual reality," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, Nov. 2016, pp. 229–232.

[65] B. Masia, J. Camon, D. Gutierrez, and A. Serrano, "Influence of directional sound cues on users' exploration across 360° movie cuts," *IEEE Comput. Graph. Appl.*, vol. 41, no. 4, pp. 64–75, Jul. 2021.

[66] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[67] Z. Sinno and A. C. Bovik, "Large-scale study of perceptual video quality," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 612–627, Feb. 2019.

[68] P. C. Madhusudana, N. Birkbeck, Y. Wang, B. Adsumilli, and A. C. Bovik, "Capturing video frame rate variations via entropic differencing," *IEEE Signal Process. Lett.*, vol. 27, pp. 1809–1813, 2020.

[69] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of temporal effects on subjective video quality of experience," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5217–5231, Nov. 2017.

[70] Z. Tu, Y. Wang, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "UGC-VQA: Benchmarking blind video quality assessment for user generated content," *IEEE Trans. Image Process.*, vol. 30, pp. 4449–4464, 2021.

[71] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J. Ohm, and G. J. Sullivan, "Video quality evaluation methodology and verification testing of HEVC compression performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 76–90, Jan. 2016.

[72] A. Panayides, M. S. Pattichis, C. S. Pattichis, C. P. Loizou, M. Pantziaris, and A. Pitsillides, "Atherosclerotic plaque ultrasound video encoding, wireless transmission, and quality assessment using H.264," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 387–397, May 2011.

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

[74] V. Zakharchenko, K. P. Choi, and J. H. Park, "Quality metric for spherical panoramic video," *Proc. SPIE*, vol. 9970, Sep. 2016, Art. no. 99700C.

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

[76] X. Xiu, Y. He, Y. Ye, and B. Vishwanath, "An evaluation framework for 360-degree video compression," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[77] S. Yang, J. Zhao, T. Jiang, J. Wang, T. Rahim, B. Zhang, Z. Xu, and Z. Fei, "An objective assessment method based on multi-level factors for panoramic videos," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

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

[79] F.-Y. Chao, L. Zhang, W. Hamidouche, and O. Déforges, "A multi-FoV viewport-based visual saliency model using adaptive weighting losses for 360° images," *IEEE Trans. Multimedia*, vol. 23, pp. 1811–1826, 2021.

[80] M. Orduna, C. Díaz, L. Muñoz, P. Pérez, I. Benito, and N. García, "Video multimethod assessment fusion (VMAF) on 360 VR contents," *IEEE Trans. Consum. Electron.*, vol. 66, no. 1, pp. 22–31, Feb. 2020.

[81] P. Gao, P. Zhang, and A. Smolic, "Quality assessment for omnidirectional video: A spatio-temporal distortion modeling approach," *IEEE Trans. Multimedia*, vol. 24, pp. 1–16, 2022.

[82] J. Guo, L. Huang, and W.-C. Chien, "Multi-viewport based 3D convolutional neural network for 360-degree video quality assessment," *Multimedia Tools Appl.*, vol. 81, no. 12, pp. 16813–16831, May 2022.

[83] T. An, S. Sun, and R. Liu, "Panoramic video quality assessment based on spatial–temporal convolutional neural networks," in *Signal and Information Processing, Networking and Computers*. Singapore: Springer, 2022.
[84] P. Guo, Q. Shen, M. Huang, R. Zhou, X. Cao, and Z. Ma, "Modeling peripheral vision impact on perceptual quality of immersive images," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.
[85] H. Wang, V.-T. Nguyen, W. T. Ooi, and M. C. Chan, "Mixing tile resolutions in tiled video: A perceptual quality assessment," in *Proc. Netw. Operating Syst. Support Digit. Audio Video Workshop*, Mar. 2014, pp. 25–30.
[86] A. T. Nasrabadi, A. Samiei, A. Mahzari, R. P. McMahan, R. Prakash, M. C. Q. Farias, and M. M. Carvalho, "A taxonomy and dataset for 360° videos," in *Proc. 10th ACM Multimedia Syst. Conf.*, Jun. 2019, pp. 273–278.
[87] C. Wu, Z. Tan, Z. Wang, and S. Yang, "A dataset for exploring user behaviors in VR spherical video streaming," in *Proc. 8th ACM Multimedia Syst. Conf.*, Jun. 2017, pp. 193–198.
[88] Y. Jin, J. Liu, F. Wang, and S. Cui, "Where are you looking? A large-scale dataset of head and gaze behavior for 360-degree videos and a pilot study," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 1025–1034.
[89] S. Fremerey, A. Singla, K. Meseberg, and A. Raake, "AVtrack360: An open dataset and software recording people's head rotations watching 360° videos on an HMD," in *Proc. 9th ACM Multimedia Syst. Conf.*, Jun. 2018.
[90] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the gap between VQA and human behavior on omnidirectional video: A large-scale dataset and a deep learning model," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018.
[91] *FFmpeg*. Accessed: Jul. 7, 2023. [Online]. Available: https://ffmpeg.org/
[92] Y. Liu. (2020). *Dashify Tile-Based Adaptive 360° Videos*. [Online]. Available: https://github.com/Yang-Jace-Liu/dash-360
[93] N. Kan. (2018). *360 video2DASH Segment*. [Online]. Available: https://github.com/confiwent/video2DASH
[94] Samsung Electronics. (2016). *360 Tool*. [Online]. Available: https://github.com/Samsung/360tools
[95] Facebook. (2015). *Transform360*. [Online]. Available: https://github.com/facebook/transform360
[96] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
[97] A. Dogra, R. K. Jha, and S. Jain, "A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies," *IEEE Access*, vol. 9, pp. 67512–67547, 2021.
[98] W. Hong, Z. H. Jiang, C. Yu, D. Hou, H. Wang, C. Guo, Y. Hu, L. Kuai, Y. Yu, Z. Jiang, Z. Chen, J. Chen, Z. Yu, J. Zhai, N. Zhang, L. Tian, F. Wu, G. Yang, Z. Hao, and J. Y. Zhou, "The role of millimeter-wave technologies in 5G/6G wireless communications," *IEEE J. Microw.*, vol. 1, no. 1, pp. 101–122, Jan. 2021.
[99] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," *ETSI White Paper*, vol. 11, pp. 1–16, Dec. 2015.
[100] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.
[101] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: Overview and Rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, Nov. 2012.
[102] E. S. Wong, N. H. A. Wahab, F. Saeed, and N. Alharbi, "360-degree video bandwidth reduction: Technique and approaches comprehensive review," *Appl. Sci.*, vol. 12, no. 15, p. 7581, Jul. 2022.
[103] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proc. 5th Workshop All Things Cellular, Oper., Appl. Challenges*, Oct. 2016, pp. 1–6.
[104] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran, and M. D. Silva, "VR is on the edge: How to deliver 360° videos in mobile networks," in *Proc. Workshop Virtual Reality Augmented Reality Netw.*, Aug. 2017, pp. 30–35.
[105] D. Liu, B. Chen, C. Yang, and A. F. Molisch, "Caching at the wireless edge: Design aspects, challenges, and future directions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 22–28, Sep. 2016.
[106] A. Mahzari, A. T. Nasrabadi, A. Samiei, and R. Prakash, "FoV-aware edge caching for adaptive 360° video streaming," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018.
[107] H. Chen and Y. Xiao, "Cache access and replacement for future wireless internet," *IEEE Commun. Mag.*, vol. 44, no. 5, pp. 113–123, May 2006.
[108] G. Papaioannou and I. Koutsopoulos, "Tile-based caching optimization for 360° videos," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2019.
[109] Z. Yu, J. Liu, S. Liu, and Q. Yang, "Co-optimizing latency and energy with learning based 360° video edge caching policy," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2022, pp. 2262–2267.
[110] C. Madarasingha, K. Thilakarathna, and A. Zomaya, "OpCASH: Optimized utilization of MEC cache for 360-degree video streaming with dynamic tiling," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2022, pp. 34–43.
[111] J. Chakareski, "VR/AR immersive communication: Caching, edge computing, and transmission trade-offs," in *Proc. Workshop Virtual Reality Augmented Reality Netw.*, Aug. 2017.
[112] Y. Liu, J. Liu, A. Argyriou, L. Wang, and Z. Xu, "Rendering-aware VR video caching over multi-cell MEC networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2728–2742, Mar. 2021.
[113] P. Maniotis and N. Thomos, "Tile-based edge caching for 360° live video streaming," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4938–4950, Dec. 2021.
[114] J. Xia, L. Chen, Y. Tang, T. Yang, and W. Wang, "Multi-MEC cooperation based VR video transmission and cache using K-shortest paths optimization," 2023, *arXiv:2303.04626*.
[115] S. Kumar, L. Bhagat, A. A. Franklin, and J. Jin, "Multi-neural network based tiled 360° video caching with mobile edge computing," *J. Netw. Comput. Appl.*, vol. 201, May 2022, Art. no. 103342.
[116] Q. Cheng, H. Shan, W. Zhuang, L. Yu, Z. Zhang, and T. Q. S. Quek, "Design and analysis of MEC- and proactive caching-based 360° mobile VR video streaming," *IEEE Trans. Multimedia*, vol. 24, pp. 1529–1544, 2022.
[117] J. Yang, Z. Guo, J. Luo, Y. Shen, and K. Yu, "Cloud-edge-end collaborative caching based on graph learning for cyber-physical virtual reality," *IEEE Syst. J.*, early access, Apr. 12, 2023, doi: 10.1109/JSYST.2023.3262255.
[118] H. Xiao, C. Xu, Z. Feng, R. Ding, S. Yang, L. Zhong, J. Liang, and G. Muntean, "A transcoding-enabled 360° VR video caching and delivery framework for edge-enhanced next-generation wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1615–1631, May 2022.
[119] T. Yang, Z. Tan, Y. Xu, and S. Cai, "Collaborative edge caching and transcoding for 360° video streaming based on deep reinforcement learning," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 25551–25564, Dec. 2022.
[120] M. Chen, W. Saad, and C. Yin, "Echo-liquid state deep learning for 360° content transmission and caching in wireless VR networks with cellular-connected UAVs," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6386–6400, Sep. 2019.
[121] Y. Li, "Federated deep reinforcement learning-based caching and bitrate adaptation for VR panoramic video in clustered MEC networks," *Electronics*, vol. 11, no. 23, p. 3968, Nov. 2022.
[122] Y. Feng, D. Wang, and Y. Hou, "Communication, computing, and caching trade-off in VR networks," *Electronics*, vol. 12, no. 7, p. 1577, Mar. 2023.
[123] C. Guo, Y. Cui, and Z. Liu, "Optimal multicast of tiled 360 VR video," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 145–148, Feb. 2019.
[124] C. Guo, Y. Cui, and Z. Liu, "Optimal multicast of tiled 360 VR video in OFDMA systems," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2563–2566, Dec. 2018.
[125] K. Long, Y. Cui, C. Ye, and Z. Liu, "Optimal wireless streaming of multi-quality 360 VR video by exploiting natural, relative smoothness-enabled, and transcoding-enabled multicast opportunities," *IEEE Trans. Multimedia*, vol. 23, pp. 3670–3683, 2021.
[126] T. Okamoto, T. Ishioka, R. Shiina, T. Fukui, H. Ono, T. Fujiwara, T. Fujihashi, S. Saruwatari, and T. Watanabe, "Edge-assisted multi-user 360-degree video delivery," in *Proc. IEEE 20th Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2023, pp. 194–199.
[127] Y. Bao, T. Zhang, A. Pande, H. Wu, and X. Liu, "Motion-prediction-based multicast for 360-degree video transmissions," in *Proc. 14th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2017, pp. 1–9.
[128] C. Perfecto, M. S. Elbamby, J. D. Ser, and M. Bennis, "Taming the latency in multi-user VR 360°: A QoE-aware deep learning-aided multicast framework," *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2491–2508, Apr. 2020.

[129] D. Nguyen, N. V. Hung, N. T. Phong, T. T. Huong, and T. C. Thang, "Scalable multicast for live 360-degree video streaming over mobile networks," *IEEE Access*, vol. 10, pp. 38802–38812, 2022.
[130] L. Feng, Z. Yang, Y. Yang, X. Que, and K. Zhang, "Smart mode selection using online reinforcement learning for VR broadband broadcasting in D2D assisted 5G HetNets," *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 600–611, Jun. 2020.
[131] L. Zhong, X. Chen, C. Xu, Y. Ma, M. Wang, Y. Zhao, and G. Muntean, "A multi-user cost-efficient crowd-assisted VR content delivery solution in 5G-and-beyond heterogeneous networks," *IEEE Trans. Mobile Comput.*, vol. 22, no. 8, pp. 4405–4421, Aug. 2023.
[132] C. Guo, L. Zhao, Y. Cui, Z. Liu, and D. W. K. Ng, "Power-efficient wireless streaming of multi-quality tiled 360 VR video in MIMO-OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5408–5422, Aug. 2021.
[133] J.-B. Jeong, S. Lee, I.-W. Ryu, T. T. Le, and E.-S. Ryu, "Towards viewport-dependent 6DoF 360 video tiled streaming for virtual reality systems," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 3687–3695.
[134] J. Chakareski, M. Khan, T. Ropitault, and S. Blandino, "Millimeter wave and free-space-optics for future dual-connectivity 6DOF mobile multiuser VR streaming," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 2, pp. 1–25, May 2023.
[135] I. Viola and P. Cesar, "Volumetric video streaming: Current approaches and implementations," 2022, *arXiv:2209.01982*.
[136] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

MOATASIM MAHMOUD received the B.Sc. degree in electrical engineering from the University of Jordan, in 2018, the M.Sc. degree in informatics and telecommunications from the National and Kapoditrian University of Athens, in 2020, and the M.Sc. degree in optical networks and photonic systems from Institut Polytechnique de Paris, in 2021. He is currently pursuing the Ph.D. degree in video distribution. He is also an Early-Stage Researcher (ESR) with the European Union's Horizon 2020 MOTOR5G Project.

STAMATIA RIZOU received the Diploma degree from the Electrical and Computer Engineering Department, National Technical University of Athens, in 2007, and the Ph.D. degree in distributed systems from the University of Stuttgart, in 2013. Since 2012, she works as a Researcher/Project Manager at the Research and Development Department, Singular Logic. Her research interests include data management, cloud computing, and the Internet of Things and its applications. She has been a Project Coordinator, Technical Coordinator, and/or an Exploitation Manager of several EU projects. She has also been an Industrial Supervisor of more than four Ph.D. students in the context of projects funded under Marie Skłodowska-Curie Actions and she has authored or coauthored more than 30 research publications. Finally, she has been acting as an Expert Evaluator at the H2020 and Horizon EU Programs.

ANDREAS S. PANAYIDES (Senior Member, IEEE) received the B.Sc. degree in computer systems and applications from the National and Kapodistrian University of Athens, Athens, Greece, in 2004, the M.Sc. degree in computing and internet systems from Kings College London, London, U.K., in 2005, and the Ph.D. degree in computer science from the University of Cyprus, Nicosia, Cyprus, in 2011. From 2012 to 2014, he was a Marie Skłodowska Curie Fellow with the Communications and Signal Processing Group, Imperial College London, and from 2015 to 2017, he was a Visiting Research Assistant Professor with the Image and Video Processing and Communications Laboratory, The University of New Mexico, USA. He has been a Senior Research Fellow with the Electronic Health Laboratory, Computer Science Department, University of Cyprus, from 2015 to 2022. Since 2022, he has been the VIDEOMICS Group Leader at the CYENS Centre of Excellence. His research interests include adaptive video processing and delivery for real-time applications, computer vision for healthcare applications, and mHealth and eHealth systems and services. He has published more than 80 peer-reviewed journals and conference papers, and book chapters, in these areas, including one patent. He is an Engineer with the Medicine and Biology Society (EMBS) and the Signal Processing Society (SPS) Joint Cyprus Chapter Chair. He was a Co-Guest of the *IET Healthcare Technology Letters* Special Issue on "mHealth-Emerging Mobile Health Systems and Services," in 2016, *Frontiers in Digital Health* Special Issue on "Connected Health: Status and Trends," in 2020, and IEEE Journal of Biomedical and Health Informatics Special Issue on "Large-Scale Medical Image and Video Analytics for Clinical Decision Support," in 2023. He is an Associate Editor of IEEE Journal of Biomedical and Health Informatics.

NIKOLAOS V. KANTARTZIS (Senior Member, IEEE) received the Diploma and Ph.D. degrees in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1994 and 1999, respectively. In 1999, he joined the Applied and Computational Electromagnetics Laboratory, Aristotle University of Thessaloniki, where he is currently a Professor. He has coauthored *Higher-Order FDTD Schemes for Waveguide and Antenna Structures* (Morgan & Claypool Publishers: San Rafael, CA, USA, 2006) and *Modern EMC Analysis Techniques I & II* (Morgan & Claypool Publishers: San Rafael, CA, USA, 2008). Moreover, he has authored or coauthored more than 170 peer-reviewed journal papers and more than 250 publications in international conference proceedings. His main research interests include computational electromagnetics, EMC modeling, metamaterials/metasurfaces, graphene, and microwave structures. He was a recipient of several scholarships, distinctions, and awards, such as the 1999 URSI General Assembly Young Scientist Award, the 2002 COMPEL Most Outstanding Paper Award of Excellence, the EMC Europe 2004 Best Paper Award, the 2013 Metamaterials Third Prize Best Paper Award, the 2014 Computation in Electromagnetics Conference Best Paper Award, and the 2014 SPIE Photonics Europe Best Paper Award (Metamaterials Section).

GEORGE K. KARAGIANNIDIS (Fellow, IEEE) is currently a Professor with the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece, and the Head of the Wireless Communications and Information Processing (WCIP) Group. He is also Faculty Fellow with the Cyber Security Systems and Applied AI Research Center, Lebanese American University. His research interests include wireless communications systems and networks, signal processing, optical wireless communications, wireless power transfer and applications and communications, and signal processing for biomedical engineering. Recently, he received three prestigious awards such as the 2021 IEEE ComSoc RCC Technical Recognition Award, the 2018 IEEE ComSoc SPCE Technical Recognition Award, and the 2022 Humboldt Research Award from Alexander von Humboldt Foundation. He is one of the highly cited authors across all areas of electrical engineering, recognized by Clarivate Analytics as a Web-of-Science Highly Cited Researcher in the eight consecutive years 2015–2022. He was a past Editor in several IEEE journals and from 2012 to 2015, he was the Editor-in-Chief of IEEE COMMUNICATIONS LETTERS. From September 2018 to June 2022, he served as an Associate Editor-in-Chief of IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. Currently, he is in the Steering Committee of IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING.

PAVLOS I. LAZARIDIS (Senior Member, IEEE) received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1990, the M.Sc. degree in electronics from Université Pierre and Marie Curie (Paris 6), Paris, France, in 1992, and the Ph.D. degree from the École Nationale Supérieure des Télécommunications (ENST) Paris and Université Paris 6, in 1996. From 1991 to 1996, he was involved in research at France Télécom and teaching at ENST Paris. In 1997, he became the Head of the Antennas and Propagation Laboratory, Télédiffusion de France/the France Télécom Research Center (TDF–C2R Metz). From 1998 to 2002, he was a Senior Examiner with the European Patent Office (EPO), The Hague, The Netherlands. From 2002 to 2014, he was involved in teaching and research with the ATEI of Thessaloniki, Thessaloniki, and Brunel University, London, U.K. He is currently a Professor of electronics and telecommunications with the University of Huddersfield, U.K. He has been involved in several international research projects, such as EU Horizon 2020 MOTOR5G and RECOMBINE, and NATO-SfP ORCA. He has published over 150 research articles and several national and European patents. He is a member of IET (MIET), a Senior Member of URSI, and a fellow of the Higher Education Academy (FHEA). He is serving as an Associate Editor for IEEE ACCESS.

ZAHARIAS D. ZAHARIS (Senior Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronics, the Ph.D. degree in antennas and propagation modeling for mobile communications, and the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1987, 1994, 2000, and 2011, respectively. From 2002 to 2013, he was with the Administration of the Telecommunications Network, Aristotle University of Thessaloniki. Since 2013, he has been with the School of Electrical and Computer Engineering, Aristotle University of Thessaloniki. He has been involved in several international research projects, such as EU Horizon 2020 MOTOR5G and RECOMBINE. He is the author of 81 scientific journal articles, 77 international conference papers, one national patent, five book chapters, and one textbook. His current research interests include the design and optimization of antennas and microwave circuits, signal processing on smart antennas, development of evolutionary optimization algorithms, and neural networks. He is a member of the Technical Chamber of Greece. Recently, he was elected as the Chair of the Electron Devices/Microwave Theory and Techniques/Antennas and Propagation Joint Chapter of the IEEE Greece Section. He is currently serving as an Associate Editor for IEEE ACCESS.