INVITED PAPER

Scalable 360° Video Stream Delivery: Challenges, Solutions, and Opportunities

This article introduces a novel networking scenario, namely 360° video streaming, discussing its challenges, existing approaches, and the research opportunities which it enabled.

By Michael Zink, *Senior Member IEEE*, Ramesh Sitaraman, *Fellow IEEE*, and Klara Nahrstedt, *Fellow IEEE*

ABSTRACT | In recent years, virtual reality and augmented reality applications have seen a significant increase in popularity. This is due to multiple technology trends. First, the availability of new tethered and wireless head-mounted displays allows viewers to consume new types of content. Second, 360° omnidirectional cameras, in combination with production software, make it easier to produce personalized 360° videos. Third, beyond these new developments for creating and consuming such content, video sharing websites and social media platforms enable users to publish and view 360° video content. In this paper, we present challenges of 360° video streaming systems, give an overview of existing approaches for 360° video streaming, and outline research opportunities enabled by 360° video. We focus on the data model for 360° video and the different challenges and approaches of creating, distributing, and presenting 360° video content, including 360° video recording, storage, distribution, edge delivery, and quality-of-experience evaluation. In addition, we identify major research opportunities with respect to efficient storage, timely distribution, and cybersickness-free personalized viewing of 360° videos.

KEYWORDS | 360° videos; content delivery; virtual reality (VR)

Manuscript received July 31, 2018; revised November 7, 2018; accepted January 11, 2019. Date of publication February 18, 2019; date of current version March 25, 2019. This work was supported by the National Science Foundation under Grant CNS-1413998 and Grant CNS-1763617. *(Corresponding author: Michael Zink.)*

M. Zink is with the Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA 01003 USA (e-mail: zink@ecs.umass.edu).

R. Sitaraman is with the College of Information and Computer Science, University of Massachusetts Amherst, Amherst, MA 01003 USA (e-mail: ramesh@cs.umass.edu).

K. Nahrstedt is with the Department of Computer Science, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA (e-mail: klara@illinois.edu).

Digital Object Identifier 10.1109/JPROC.2019.2894817

I. INTRODUCTION

In recent years, virtual reality (VR) and augmented reality (AR) contents and applications have seen a significant increase in popularity. This is mainly due to two recent technology trends. The first trend is the availability of new tethered and wireless head-mounted displays (HMDs), such as Oculus Rift or HTC Vive, that allow viewers to consume new types of content. According to Cisco's Visual Network Index report [1], VR headsets are expected to grow fivefold, from 18 million in 2016 to nearly 100 million by 2021. The report also predicts that more than half of the VR headsets will be connected to smartphones; most of the remaining will be connected to PCs and consoles, while a few will be standalone. The second trend is the availability of omnidirectional cameras that make it easy to produce personalized 360° videos, such as GoPro OmniAll and Insta360 One. Beyond the end-user technologies that enable the creation and consumption of 360° content, video sharing websites (e.g., YouTube) and social media platforms, such as Facebook, allow users to publish and disseminate such content. Since it has become easier to create, distribute, and consume personalized 360° videos, streaming of such videos has turned into a popular VR application.

Even though the technology to create and consume 360° videos has become widely available, the delivery of high-quality 360° videos over the Internet to a globally distributed set of users poses significant challenges that we outline in the following. We believe that these challenges have to be addressed by the research communities and industries to support VR streaming at a global scale and under the premise of the projected increase in popularity of such content.

A. Ultrahigh Bandwidth Requirements

The bandwidth required to stream a 360° video is an order of magnitude larger than that required for a traditional (2-D) video. The data rate of a 360° video that delivers a 4k stream to each eye and allows a full 360° viewing range requires about 400 Mb/s, compared to about 25 Mb/s for a traditional 4k video.

B. Ultralarge Storage Requirements

The ultrahigh bandwidth of 360° videos also translates to ultralarge storage requirements. Specifically:

1) multiple views of each scene of the 360° video must be stored;
2) since there is a large variety of client devices, such as VR goggles, smart phones, or desktops, the 360° video content needs to be stored in many different formats (projection method, temporal/spatial resolution, and so on);
3) to account for fluctuations in available bandwidth between 360° video source and client, adaptive formats are required to support adaptive bitrate (ABR) streaming.

For example, if we assume that the bitrate of a 360° video is 400 Mb/s, a total of 15 GB would be required to store a 5-min video and 270 GB would be required in the case of a 90-min video. The required storage space would increase if multiple alternative quality versions would have to be stored to support the ABR streaming.

C. Ultralow Motion-to-Photon Delay

To prevent cybersickness, ultralow motion-to-photon delay is required, i.e., when the user turns her head, the delay in rendering the new view should take no more than a few tens of milliseconds. It is known that this delay should be in the order of tens of milliseconds to avoid motion sickness [2], [3]. Traditional 2-D Video-on-Demand (VoD) applications do not have such stringent delay requirements, and a live broadcast 2-D video that is delayed by multiple seconds is often acceptable. Furthermore, the consequences of high delay in traditional videos are just annoyance, rather than causing motion sickness that is much more serious!

D. Complex View Adaptation

The 360° video adds the additional complexity of adapting the video delivery to the viewport of the user, as he/she moves her head, in addition to the traditional bitrate adaptation that is also required for the 2-D videos. There exist two general approaches to tackle this issue. In the first one, the complete 360° panorama is transmitted to the client, and only the area covered by the viewport is rendered in the display. This results in wastage of consumed bandwidth since a significant part of the content streamed to the client is not consumed by the viewer. An alternative approach to perform viewport adaptation is to split the 360° video into several regions that are specified as tiles [see Fig. 1(b)]. With this approach, a 360° video is composed of many streams, where each stream represents a specific direction (tile) in the 360° panorama and each of these streams may be available in many qualities (spatial and temporal resolution) to support the ABR streaming. The main challenge in view adaptation is predicting the movement of the user and using the adaptive control principles to manage both the server–client bandwidth and the client-side buffer that stores tiles prior to being played out. The complexity of performing seamless view/bitrate adaptation requires innovative algorithms and architectures that make intelligent decisions to store, cache, and prefetch content, taking into account the cinematographic rules. For example, to guarantee ultralow delay to prevent cybersickness, the tile that the user will see in the near future should already be prefetched into a proximal edge server or even to the client's buffer so that it can be rendered quickly. In addition, buffering methods that are traditionally used in clients in the case of bitrate adaptation for 2-D videos cannot be directly applied to the 360° video streaming.

E. Complex Rules and Metadata for Viewing the Videos

Traditional videos are watched from a single viewpoint that is predetermined by the director of the video. In contrast, for 360° videos, the freedom of a user to view a 360° panorama means that he/she can watch the video in complex ways that are not predetermined or predictable at the production time. However, to optimize the viewing experience, the director may still want to restrict the user by allowing her to watch the video only from specific viewpoints that conform with cinematographic rules. For instance, the director of a 360° movie may want to only allow views that conform with the 180° rule that keeps the subjects in the same relative order on the screen. Encoding the director's rules in metadata and enforcing them in video delivery are the examples of additional complexity which are the key parts of 360° video delivery with no counterparts in traditional single-view video delivery.

F. Video QoE

Traditional video Quality of Experience (QoE) has resulted in extensive research over the years [4]–[8], including some of our work [9]–[12]. However, what contributes to the 360° video QoE is much less understood and requires conceiving of new metrics.

The above-mentioned challenges make the current state-of-the-art video delivery architectures unsuitable for 360° video delivery, except for the low-quality 360° videos that we see today. In this paper, we survey several approaches in the areas of content creation (see Section III), 360° video distribution (see Section IV), and QoE (see Section V) for scalable 360° video streaming systems. In addition, we identify important future research challenges for the 360° video delivery.

II. OVERVIEW

Before we start describing the existing approaches and challenges, we provide an overview on how 360° videos are created, stored, and delivered.

A. Data Model

We present a data model that describes how 360° video are captured, encoded, and stored.

Omnidirectional videos are spherical videos (in some cases, mapped into a 3-D geometry), where the viewer can change the viewport [8]—the section of the sphere that is currently in the user's Field of View (FoV)—during playback using either an HMD or a regular screen (e.g., TV or computer monitor). In the case of an HMD, head movements determine the change in viewport, while a mouse or a remote has to be used for control in the case of regular screens. Such videos are captured by a centric camera system where multiple cameras are centrally mounted and facing outward, so as to cover the whole sphere with the camera system at the center [see Fig. 1(a)]. The sphere can be projected and subdivided into tiles, as shown in Fig. 1(b). The viewport selected by the user can be reconstructed at the user's viewing device by downloading and stitching together the relevant tiles.

The majority of the 360° video content offered today is of this centric type. Relatively low-cost camera systems and the easier setup of such systems enable even nonprofessionals to produce 360° videos. Since such videos allow viewers to choose a viewport from a complete sphere, they are called omnidirectional [13]. A series of encoding techniques for such omnidirectional videos have been presented in the past [14]–[16]. Recently, layered encoding schemes have been proposed [17], [18], which have the goal of reducing stalling and adapting quicker to viewport changes. While many approaches require special players, browsers can be used to watch 360° videos if they support the WebVR standard [19].

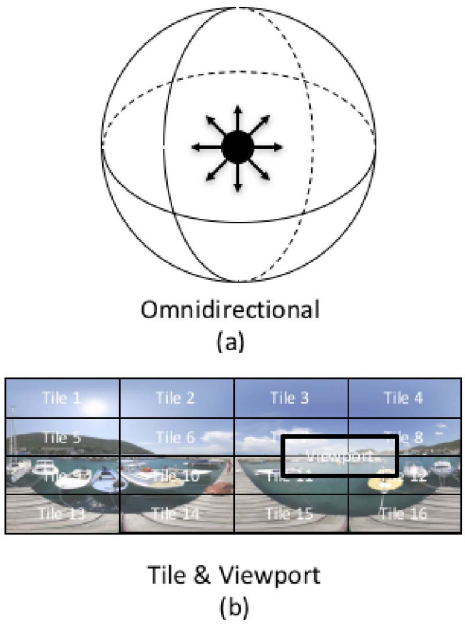

Fig. 1. *(a) Omnidirectional 360° video creation system where the camera or camera system is physically located in a central spot. (b) ERP of a 360° image from an omnidirectional camera and its subdivision into tiles and a user selected viewport.*

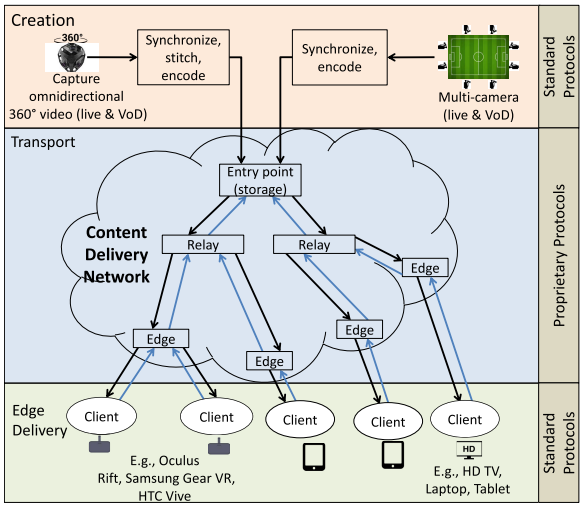

Fig. 2. *Architecture of a 360° video stream delivery system.*

B. Video Delivery Architecture

We envision that the delivery architecture for scalable, high-quality 360° video will build on existing architectures used for the video content delivery networks (CDNs) [20], [21], albeit with significant novel enhancements to support the additional requirements. Our assumption is based on the fact that most of the traditional videos are delivered by the CDNs today, and they provide a natural architectural platform that can be extended to the 360° video streaming. An overview of a potential 360° video delivery architecture is shown in Fig. 2. The architecture is comprised of three major components: creation, transport, and edge delivery. The creation component is concerned with the encoding, storage, and preparation for the streaming of 360° videos. The transport component focuses on the delivery of such videos. This includes the creation of distribution trees and the analysis of different dimensions of coherence that can be used for efficient distribution. Finally, the edge delivery component focuses on how the stringent delay requirements for 360° video streaming can be met by establishing a tight interaction between the clients and CDN edge servers. All three components will be discussed in more detail in Sections III–V. Note that the resulting QoE of a viewer watching a 360° video is impacted by the individual components of this architecture. For example, having the relevant tiles of

两种编码

a 360° video cached at the edge server closest to the requesting client might allow a smooth viewport adaptation (based on the viewer's head movement). However, if the relevant tiles are not present at the edge servers, the longer round-trip to obtain them could result in a perceptible lag and a lower QoE.

In Section III, we focus on the content creation component that is more complex than that for the traditional videos. In Section IV, we give an overview of the transport component that is responsible for transporting videos across the WAN from their points of creation to the edge servers that are proximal to the clients. Due to the strict timing constraints of 360° video streaming (e.g., response to head movement), increased attention has to be given to video distribution, which we will also discuss in Section IV. Similar to how traditional videos are delivered today, a CDN could use proprietary protocols for transporting videos between their own nodes to gain a competitive advantage, while standard protocols are likely to be used elsewhere. This approach allows a large diversity of content creation and client systems that use standard protocol interfaces to seamlessly enter the 360° video ecosystem.

III. CONTENT CREATION

As with all streaming systems, the creation of content is one of its major components. While there has been a plethora of work on the content creation, we focus on the approaches for 360° video streaming. Multicamera video content creation for 360° video is taking on many different forms, and we present the different forms and representations that are important for streaming, distribution, and viewing of the video content. For multicamera systems, streams are recorded and/or distributed with different representations. One potential representation is MPEG Dynamic Adaptive Streaming over HTTP (DASH), which represents the content stream of each camera, hence generating k independent MPEG-DASH video streams for k distributed 2-D camera environments. Other systems may take the 2-D multicamera content, e.g., multiple cameras collocated on one stick, and stitch the content, generating a 360° video, to generate a 3-D teleimmersive video. In the following, we will discuss the characteristics of each of these video representations.

A. MPEG-DASH Video Representation

We start with a description of how traditional 2-D videos are represented with MPEG-DASH and encoded with an MPEG video standard, such as H.264/AVC [22] or High Efficiency Video Coding (HEVC) [23], and transmitted via the DASH [24]. It is informative to understand MPEG-DASH as there has been a recent push to extend this framework to omnidirectional 360° videos with the MPEG-1 Omnidirectional Media Access Format (OMAF) standard [25]–[27]. It should be mentioned that in addition to DASH, there exist proprietary ABR implementations, such as Microsoft's Smooth Streaming [28], Apple's HTTP Live Streaming (HLS) [29], and Adobe's HTTP Dynamic Streaming (HDS) [30]. With the introduction of the Common Media Application Format [31], MPEG has recently started an initiative to create a single-standard segment format that is supported by DASH and HLS (and potentially others).

The MPEG-DASH stream is divided into segments that can be encoded in different bitrates or spatial resolutions. The resulting segments are stored on a web server and requested from a video client via standard HTTP, as shown in Fig. 3. In the example shown in Fig. 3, MPEG-DASH video is represented in three qualities—best, medium, and low—and divided into chunks of equal time length. Clients, then, adapt according to their available bitrate. To describe the spatial and temporal relation between segments, MPEG-DASH introduces the so-called media presentation description (MPD). The MPD is an XML file which represents the different qualities of the media content and the individual segments of each quality with the HTTP uniform resource locator (URL). This structure provides the binding of the segments to the bitrate (resolution

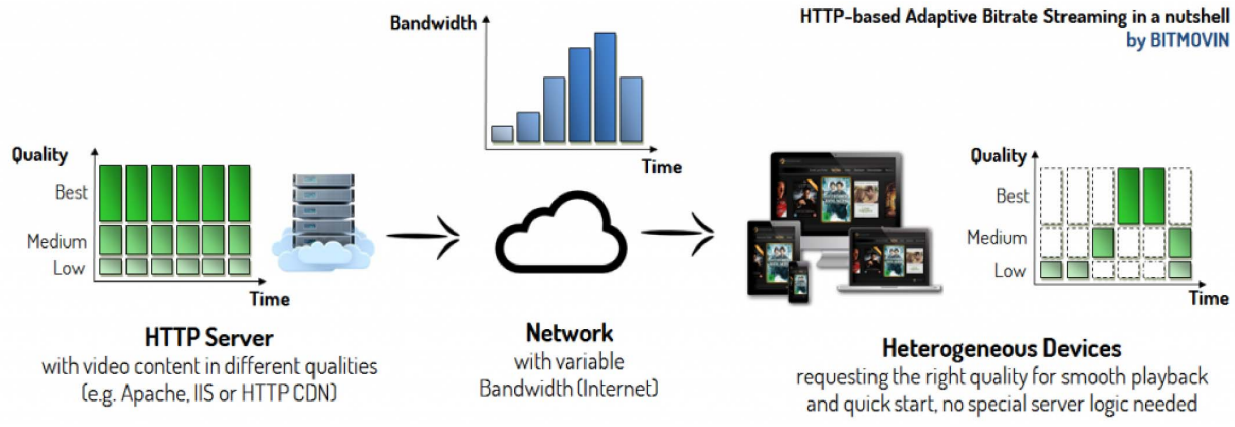

Fig. 3. ***MPEG-DASH video framework [32].***

and so on) and temporal information (e.g., start time and duration of segments). As a consequence, each client will first request the MPD that contains the temporal and spatial information for the media content, and based on that information, it will request the individual segments that fit best for its requirements.

The MPD is a hierarchical data model that can contain one or more periods. Each period contains video components with different view angles or codecs and audio components with different languages, subtitles, or captions. Those components have certain characteristics, such as bitrate, frame rate, and audio channels, which do not change within a period. Typically, media components are organized into adaptation sets. Each period contains one or more adaptation sets, in which different multimedia components that logically belong together are grouped. For example, components with the same codec, language, resolution, and audio channel format (e.g., 5.1 stereo) could belong to the same adaptation set. Note that the adaptation sets consist of a set of representations containing interchangeable versions of the respective content. This allows the client to choose a representation for playback, adapt the media stream to current network conditions, and achieve a smooth playback without stalling and with better QoE.

Representations are in turn chopped up into segments to enable the switching between individual representations during playback. Those segments are described by a URL and, in certain cases, by an additional byte range if those segments are stored in a bigger, continuous file. The segments in a representation usually have the same duration (usually between 2 and 10 s although MPEG-DASH does not restrict the segment length or give advice on the optimal length) and are arranged according to the media presentation timeline, which represents the timeline for synchronization, enabling a smooth switching of representations during playback. The length of segments presents a tradeoff. Longer segments provide efficient compression as the Group of Pictures could be longer or reduce network overhead since bigger chunks will be transmitted. In contrast, shorter segments are used for live scenarios as well as for constrained scenarios, such as mobile networks, as they enable faster and flexible switching of individual bitrates.

Recently, in response to the increased popularity of 360° video, MPEG issued a draft for an omnidirectional media application format [33]. This format specifies a set of projection mappings for the conversion of 360° video into the 2-D plane (see Fig. 4). In addition, it specifies a storage format for omnidirectional video in the ISO Base Media File Format (ISOBMFF), and the encapsulation, signaling, and streaming of omnidirectional video over DASH are specified. Finally, it declares which video and audio codecs as media configurations can be used for the compression of 360° video.

As in the case of streaming traditional, non-VR video, 360° video requires metadata to describe spatial and temporal relations between the segments for transport and viewing. The MPEG-DASH Spatial Relationship Descriptor standard [34] extends the MPEG-DASH MPD to allow the streaming of spatial subparts (in our case tiles) of a video by describing the spatial relationships between the associated pieces of a video content. This additional information is provided in addition to the metadata already offered in the standard DASH MPD which provides information, such as segment length and quality levels of each available segment.

B. Projection and Preparation for Streaming

Omnidirectional 360° video is known as a spherical video, as it involves a 360° view of the scene captured from a single point. The captured video maps to the internal surface of a sphere, as shown in Fig. 1(a). An HMD views only a limited portion of the video as seen from the center of the sphere. This view is called the user's viewport. The area covered by the viewport is limited by the HMD's FoV, and its coordinates are based on the orientation of the user's head. Presenting complete high-quality video within the user's viewport requires a complete high-quality 360° frame. For example, the Oculus Rift's viewing resolution is 1080×1200 per eye with an FoV of 110°, meaning that the complete 360° frame should have 6k resolution to exploit the highest quality that the device can offer. This requires 400 Mb/s of bandwidth to stream the 360° video content to the client.

1) Projection: Since encoding of spherical videos is not supported by the existing video coding standards directly, we need to first project video to a 2-D plane using a projection method [18], [35]. Examples of the projection methods are equirectangular [36], [37], cubemap [13], [38], or pyramid projection [39], as shown in Fig. 4. The most common projection is the equirectangular projection (ERP) that maps a sphere into a rectangle. This projection introduces severe stretching at the north and south poles of the sphere, which reduces the efficiency of encoding. The cubemap projection (CMP) maps 90° FoVs to the sides of a cube. It has less quality degradation across the video than ERP and lower bandwidth overhead. Furthermore, cubemap requires less rendering processing power [40] and

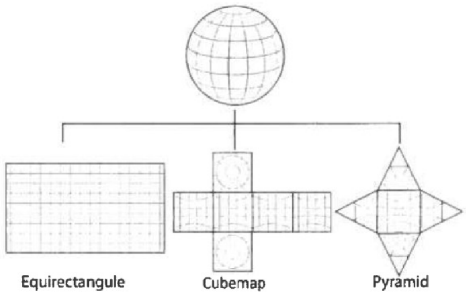

Fig. 4. ***Projection methods [40].***

reduces the file size by 25% compared to ERP. Applying the pyramid projection approach results in a file size reduction of up to 80% [39] but suffers from quality degradation at the sides of the pyramid and it requires high GPU processing.

2) Preparation of 360° Video for Streaming: To enable 360° video to stream from a server, it needs to be prepared for streaming in the spatial and time domains. In the time domain, the video sequence is divided into segments of fixed duration. This is very similar to the DASH video format described earlier. In the spatial domain, one can divide each 360° video frame into tiles (see Fig. 1), and each tile is made available in a set of different quality levels. The tiles are then encoded into multibitrate segments. The viewer receives segments that incorporate tiles that correspond to the user's viewport. With the recent standard for the Omnidirectional MediA Format (OMAF) [33], [41], MPEG is in the process of specifying storage and delivery formats of the 360° videos. OMAF is building on top of the MPEG-DASH and the ISOBMFF [42]. Currently, equirectangular and CMPs are the only ones considered by the standard, while others are under consideration. The integration of DASH into OMAF is realized by the DASH's extension tools, which allow the definition of new property descriptors to expose information related to 360° content in the MPD. One example is the Projection Format Descriptor that specifies the projection method (e.g., 0 for ERP).

The client adapts the video quality depending on the user's viewport in the spatial domain. For instance, the client can prefetch higher quality tiles for the area covered by the user's viewport as far as the available bandwidth allows [43], [44]. In tile-based HTTP adaptive streaming, the client has to perform two adaptations: rate adaptation to adapt time-varying bandwidth and viewport adaptation to cope with the user's head movement. The selection of quality of tiles is one of the major issues with tile representation. One can formulate this issue as a bandwidth-optimal problem with selecting the quality of tiles that maximizes the quality of viewport depending on the available bandwidth [45]. Another approach could be to predict the user's future viewport and select the optimal sequence of tiles, which minimizes the bandwidth consumption. Some of the related work's results [44] show that one can predict the viewport quite accurately for the next second. But the major difficulty with the tile-based methods is that it requires multiple decoders at the client side. Some solutions have explored the usage of motion-constrained tile sets feature of HEVC [37] and used a single hardware decoder to decode a set of selected tiles from the user's viewport. Other challenges of the tile-based representation and streaming are as follows.

1) Tile prefetching errors since the motion-to-photon latency requirement for VR is less than 20 ms [46], which is less than the Internet request–reply delay. Therefore, we must prefetch tiles by viewport prediction. However, it is difficult to do the long-term prediction (>3 s) [44].
2) Rebuffering and stalls under small playback buffer, which happens because the time-based methods keep small playback buffers to help with the short-term constraints of viewport prediction.
3) Border effects of mixed bitrate tiles, which happens due to the spatial partition of video. When video tiles are encoded with different bitrates, these mixed-bitrate tiles can result in visible border and quality inconsistency in combined tiles rendering [47].

A different alternative is to create the viewport adaptive representations for each video chunk independently. This requires that one has to define a set of overlapping viewports that are of interest to users. Then, for each predefined viewport, one takes the entire spherical frame, including the viewport, and provides higher quality within the predefined viewport and lower quality in the rest of the frame. Facebook and Pixvana [48] employ this representation and method for adaptive 360° video streaming [49]. Facebook transforms a 360° video into a viewport-dependent multiresolution panorama, which decreases the overall resolution without decreasing the quality of the viewport. In this case, 30 different viewports covering the 360° video are created with a viewport separation of 30°. The main benefit of this method is that the video can be decoded by a single decoder. In comparison to the tile-based approach, data for the entire spherical frame are transmitted. This may lead to increased storage and bandwidth requirements for this approach.

C. Cinematographic Rules

There are various cinematographic rules that the film industry employs as movies are created, edited, and displayed to the users to guarantee a satisfying viewing experience [50]. An example of a cinematographic rule is the 180° rule, where during a conversation, the camera view should stay on a 180° line that keeps two or more characters in the same position on the screen. Since personalized control of 360° video content is a new feature that is not offered in a traditional video, it will be important that the 360° video delivery systems also utilize these rules to maximize the viewing quality. The need for the incorporation of such rules with respect to the 360° video has also recently been identified by MPEG. For example, the OMAF standard [25] allows the content's author to specify an initial viewport in the form of metadata. With this metadata, a viewport trace can be expressed, which can be seen as a "director's cut" for VR storytelling [27].

Several approaches provide automated selection of views (including pan, tilt, and zoom of a camera), which is especially the case in event broadcasting. The standard is that a content provider (e.g., NBC) edits views and delivers the selected views via the broadcasting mechanism to a broad audience. This approach leaves no room for the personalization of views by the user. Hence, there has

been much effort to understand users and how to enable personalized viewing of videos. Gaddam *et al*. [51] present a study that compares the user's perceived quality when a panoramic video system for soccer is operated by a human being (director) versus different algorithms. Similar work has been presented by Ariki *et al*. [52] (also for soccer) and Carr *et al*. [53] for the capturing of basketball matches. Chen and Vleeschouwer [54] propose a personalized production system for basketball matches with the goal of preventing production and storytelling artifacts.

Dambra *et al.* [55] are amongst the first to present an approach that considers film editing to limit and control the viewer's head movement with the goal of reducing the bandwidth consumption in the case of 360° video streaming. The approach is based on the idea that a new scene (after a scene change) starts with a viewport the viewer would most likely select. This helps a video player decide which tiles of a 360° video should be buffered in a high quality and which not.

Obviously, limiting the viewers' flexibility in watching 360° videos based on the cinematographic rules can create conflict. A viewer might actually decide to choose a viewport that violates these rules. Studies to determine the impact of QoE will have to be conducted to better understand the impact of the cinematographic rules in the case of 360° videos.

D. Video Recording and Storage

The 360° video recording and storage systems must consider two major issues: 1) the generation of control metadata to describe media stored in the video segments and 2) the continuous capture and storage of video segments in a storage cloud. In the area of video metadata, recent work in MPEG has focused on the specification of the 360° video information. The MPEG DASH Spatial Relationship Descriptor standard [34] extends the MPEG DASH MPD to allow the streaming of spatial subparts (i.e., tiles in the case of 360° video) of a video by describing the spatial relationships between the associated pieces of a video content. This spatial information is provided in addition to the metadata already offered in the standard DASH MPD that provides information, such as segment length and quality levels each segment is available in. Extensions for distributed 360° video have been presented in the literature [56] but are currently not considered as a part of an extension for DASH MPD. In addition, the MPEG OMAF standard (see Section III-B) builds on top of the MPEG DASH and supports metadata for projection formats (currently, equirectangular and cubemap).

Furthermore, a large amount of research on multimedia storage systems has been published [57]–[59] and extended toward storage of correlated 3-D streams at single servers [59]–[61]. In addition, large-scale map-reduce storage systems, including the Hadoop Distributed File System [62] and the Google File System [63], are available, but they have not been fully optimized for the storage of 360° videos. In the case of the 360° video content, we need to consider a large number of cameras, views, and tiles for near-term and long-term storage with very large viewership. Hence, the available storage systems will need to be revisited and new architectures to be considered.

IV. VIDEO DELIVERY

The increased popularity of 360° videos has led to research that focuses on increasing the efficiency of 360° video delivery. The goal of the proposed approaches is to reduce the amount of data that has to be transported from the source to the edge servers and to the clients while keeping the perceived quality at the client to be high and the motion-to-photon latency to be low. However, due to the lack of a standard QoE evaluation approach for 360° video (see Section V), much of the literature uses only image quality metrics to assess the viewers' perceived QoE.

The 360° videos are much more personalized and interactive than the traditional 2-D videos for applications, such as Skype, PPlive [64], CoolStreaming [65], LiveSky [66], and YouTube. Scalable 360° video delivery systems have to be able to combine the interactive real-time content generation with broadcast distribution and deliver 360° video content to viewers who: 1) interactively watch the activities of the content producers and 2) select a viewport of the activities at run time. Even though the current IPTV solutions [64]–[67] provide efficient frameworks for large-scale dissemination, they do not consider multiparty multiview contents with viewport change dynamics to serve 360° video content.

Nahrstedt *et al.* [68] envisioned an Internet Interactive Television, where viewers can select multiple contents they want to watch together in a smart room and the contents are generated from different heterogeneous sources (e.g., mobile phones and TV cameras) by different entities distributed all over the world. Nevertheless, the content was not semantically correlated with each other. In the 360° video content delivery systems, video content is correlated by locality, such as an omnidirectional camera that takes many views around the camera's location, and by event, such as many views of a single event such as soccer.

A. Exploiting Coherence for Efficient 360° Video Delivery

The delivery of 360° videos can be performed by forming distribution trees that disseminate the video content from the source to the edge servers that request that video. The edge servers in turn send the video segments to their downstream clients. As shown in Fig. 2, the nodes of the tree are CDN servers that can store and send the video segments. The nodes of a distribution tree for on-demand videos have the ability to cache and forward the video segments [69]. The nodes of the distribution tree for live videos act as relays and forward segments down the tree. Such architecture is sometimes referred to as the application-level multicast.

Constructing distribution trees for traditional 2-D videos is a well-studied problem. In our work [70], we study constructing the application-level multicast trees for live 2-D videos by solving an optimization problem that yielded low cost and low packet loss paths for disseminating the video. Our work was subsequently applied to the production live streaming network at Akamai [71]. In addition to the application-level multicast, there have also been several key proposals for forming scalable IP-level multicast trees for traditional 2-D videos, such as the core-based trees [72] and the Mbone [73].

Constructing the distribution trees for 360° videos is much more complicated, as it takes prohibitive amounts of bandwidth to transport an entire high-quality 360° video from its source to the edge servers and to the clients. The worst case situation for 360° video transport is if every user is watching a different and unpredictable viewport at a different and unpredictable quality level, but that worst case is not likely to occur in practice. In fact, viewers likely watch 360° video in spatially, temporally, locationally, and behaviorally coherent ways as follows.

1) Spatial Coherence: It dictates that the viewport of a client is made up of tiles that are spatially adjacent to each other. This means that the edge server can proactively prefetch and cache tiles that are spatially adjacent to the downstream client's current viewport. In the case of a viewport change, the required tiles for creating the new view may already be in cache at a proximal edge server, drastically reducing the delay.

2) Temporal Coherence: It dictates that a client's viewport is likely to evolve in a predictable way over time. For example, if we know the trajectory of a client's viewport, one can predict the motion into the future, prefetch the tiles that are needed next, and store it at a proximal edge server, reducing the motion-to-photon delay.

3) Locational Coherence: It dictates that two clients in close proximity, e.g., adjacent seats in a stadium, likely have similar views. Clients in the same location may also have similar connectivity and require tiles of the same video quality, e.g., two clients in a home who share the same connection to the Internet. Such locational coherence can be exploited to reduce the amount of traffic on the network.

4) Behavioral Coherence: It dictates that although clients can in principle view 360° videos in arbitrary and uncorrelated ways, their viewports will likely be very correlated [74], [75]. In many cases, there is a point of interest that most clients will focus on. For example, in the case of a soccer match, the eyes of most viewers will likely follow the ball as it moves across the field. The use of the cinematographic rules to constrain the client's viewing behavior will likely increase such coherence. If it can be predicted, such coherence can also be exploited to send less traffic through the network.

Exploiting such viewing coherences avoids transporting all tiles at all qualities to the edge servers and drastically reduces the traffic and delay by prefetching only what is needed before it is needed. The major challenge in the 360° video transport is how (and whether) these coherences can be exploited in a scalable, accurate, and efficient manner, a topic that requires much future research.

B. Recent Work

Recent work has started to explore viewing coherences in 360° videos. Graf *et al.* [36] propose a set of strategies, in which only tiles of a viewport and their immediate neighbors are streamed to the client, which can lead to bitrate savings up to 65%. Zhou *et al.* [76] reengineer the Facebook's approach to stream 360° videos to an Oculus HMD. Their evaluation reveals that abrupt changes in viewport direction result in 20% waste of the total download bandwidth due to the retrieval of unwatched segments. Bao *et al.* propose a motion-based approach for the omnidirectional 360° video [74], [77], where they demonstrate that the users' selection of viewport are correlated. This knowledge can be used to send a multicast stream to the geographically colocated viewers.

Looking at the next-generation HMDs, where gaze detection is possible, exploiting spatial coherence becomes even more challenging though more rewarding. With gaze detection, the HDM can provide accurate spatial information that can pinpoint the focus of the client. The human eye can see a rich image only for a small central viewing angle of about 20° [78] beyond which the vision acuity and color perception drastically decrease till it reaches the full FoV of about 100°. With gaze detection, gaze information will flow as control information from the client to the upstream edge and nodes of the distribution tree. An approach to drastically reduce the amount of traffic is to encode the video in multiple resolutions and tile sizes, similar to the approaches proposed for a single-view video [79]. Small-size high-resolution tiles can be used for the central vision for the client, while the large-size low-resolution tiles can be used for the peripheral vision.

V. EXPERIENCE

QoE [4], [80] has been used for many years to assess the subjective experience of a viewer when viewing 2-D streaming content. It has become clear that in systems that present audio and visual contents to a user, pure objective metrics, such as the ones often used to determine the Quality of Service (QoS), are not sufficient since they do not capture a user's preferences and subjective assessment of the quality of the presented content. For example, while the average bitrate could be quite high, it could fluctuate significantly on a short time scale, leading to many changes in quality level during a streaming session, which can be quite annoying for the viewer.

The increased popularity of ABR streaming has led to a variety of proposals for QoE metrics [5], [8], [10]. Recently, an ITU recommendation [81] has been published, which specifies the quality assessment for HTTP adaptive streaming. Most of these proposals and the ITU

recommendation are limited to the traditional 2-D, non-interactive video. The results of these investigations show that stalling (i.e., the pausing of the play out of content due to lack of available data at the client) either at the start of a video [7], [82] or during play out has the most significant impact on QoE. Other significant factors are the average quality, in which the video is streamed to the client and how often the quality changes during play out [6], [7]. Furthermore, there have been efforts to map QoS to QoE for the ABR streaming [83], [84]. While there is a large body of work on QoE for the traditional 2-D video streaming, only little is known in the area of 360° video delivery. Ghosh *et al*. [85] present a rate adaptation algorithm for the tile-based 360° video, in which they define QoE metrics. While these metrics present an initial approach, their validity has not been demonstrated through subjective assessment. Schatz *et al*. [86] present results from a QoE evaluation of 360° video, but this approach only focuses on the impact of stalling. The impact of other factors, such as the motion-to-photon delay, caused by delay or jitter and its potential to cause cybersickness are not considered.

Singla *et al*. [87] present results on QoE and cybersickness for the case of omnidirectional 360° video, but their evaluation focuses on the effects caused by using different HMDs, and the influence of video creation and delivery is not considered. A test bed for the subjective evaluation of omnidirectional content is proposed by Upenik *et al* [88]. Currently, this test bed only allows the evaluation of omnidirectional still images.

While several data sets for 360° video already exist [75], [89], [89], [90], they have not been used for the purpose of QoE evaluation but rather for the modeling of user behaviors.

We believe that more structured research in the area of QoE for 360° video streaming is required to better evaluate how viewers perceive the viewing of such content. This will require a series of subjective assessments, similar to the ones that have been performed for the traditional 2-D video. A standard that describes how such assessments should be performed, similar to the one for television pictures [91], would assure that assessment results generated by different entities are comparable. Such research would also benefit from a test bed for such assessments.

There are several aspects of QoE for 360° video streaming, which should be assessed. We believe that the assessment of factors that could impact cybersickness should be investigated with high priority since this is an effect that does not occur if one watches traditional 2-D videos on a regular screen (in opposite to VR goggles and glasses). Our hypothesis is that cybersickness caused by viewing omnidirectional 360° videos via HMDs and abrupt scene changes when viewing 360° video are the two significant contributors to QoE.

One major focus with respect to future research on QoE should be the investigation of the impact of cybersickness on the users' viewing experience. In addition, the influence of network characteristics, such as delay, jitter, available bandwidth, view/tile degradation ratio, and streaming characteristics, such as stalling, quality changes, and viewport changes on cybersickness, should be studied. It might also be important to study if physiological symptoms, such as heart rate and respiratory rate [92], [93], can be linked to the subjective experience of viewers when viewing 360° videos. It should also be evaluated to what extent the provision of viewing guidance according to cinematographic rules will impact QoE. As mentioned in Section III, certain kinds of scene changes which do not adhere to cinematographic rules may cause a viewing discomfort, such as disruption in continuous viewing and violation of synchronized playback. While studies of the camera switching and its impact on QoE have been performed in the past [94], those have been very rudimentary and further studies are required to better understand the relationship between the two. Finally, the impact of interaction between real world and cyber world on QoE should be studied. For example, in the case of untethered HMDs, users will be able to walk freely. In addition, body parts of the viewer (e.g., hands, legs, and feet) are not visible, which might be another contributing factor to the diminished QoE and cybersickness. This will require mechanisms that provide feedback about physical objects close to the viewer. In the case of AR and teleimmersion, the impact of synchronization (or the lack of) between real world and cyber world on QoE should also be studied.

VI. CONCLUSION

Omnidirectional and distributed 360° videos are increasing in popularity. This increase can be attributed to new content creation systems, such as the omnidirectional cameras and end devices such as the VR headsets. With this increase in popularity comes a new set of challenges for the video streaming systems. In this paper, we focus on three major facets of the 360° videos: content creation, video distribution, and QoE. For each of these facets, we identify major challenges and identify the existing approaches that have been proposed.

While a significant amount of research has been performed in the area of omnidirectional and distributed 360° videos in the recent past, many of the existing challenges have not been solved and need further attention. First, standardized QoE metrics that focus on the characteristics of 360° video do not yet exist. Second, new mechanisms that can meet the stringent latency requirements of 360° video, building on top of the existing CDN architectures, have not yet been devised. Finally, while content creation for 360° video has matured significantly in recent years, the integration of cinematographic rules into content creation techniques is required to guide the viewer while watching 360° content. In the long term, the increase of popularity in 3-D teleimmersion applications driven by advances in holographic displays will require new research and streaming standards. ■

REFERENCES

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021 white paper," White paper, 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] A. Berthoz and G. Weiss. The brain's sense of movement. Harvard Univ. Press. (2000). [Online]. Available: http://www.jstor.org/stable/j.ctt1cc2mz2

[3] P. Fuchs, *Virtual Reality Headsets—A Theoretical and Pragmatic Approach*. Boca Raton, FL, USA: CRC Press, 2017. [Online]. Available: https://books.google.com/books?id=P640DgAAQBAJ

[4] R. Jain, "Quality of experience," *IEEE MultiMedia*, vol. 11, no. 1, pp. 95–96, Jan./Mar. 2004.

[5] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, 1st Quart., 2015.

[6] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing effect sizes of influence factors towards a QoE model for HTTP adaptive streaming," in *Proc. 6th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Sep. 2014, pp. 111–116

[7] T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, "Internet video delivery in YouTube: From traffic measurements to quality of experience," in *Data Traffic Monitoring and Analysis*. Berlin, Germany: Springer, 2013, pp. 264–301, doi: 10.1007/978-3-642-36784-7_11.

[8] M.-N. Garcia *et al.*, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. 6th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Sep. 2014, pp. 141–146

[9] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 2001–2014, Dec. 2013.

[10] M. Zink, J. Schmitt, and R. Steinmetz, "Layer-encoded video in scalable adaptive streaming," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 75–84, Feb. 2005.

[11] C. Wang, D. Bhat, A. Rizk, and M. Zink, "Design and analysis of QoE-aware quality adaptation for dash: A spectrum-based approach," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 3s, pp. 45:1–45:24, Jul. 2017. [Online]. Available: http://doi.acm.org/10.1145/3092839

[12] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang, "Quality of experience in distributed interactive multimedia environments: Toward a theoretical framework," in *Proc. 17th ACM Int. Conf. Multimedia*, New York, NY, USA, 2009, pp. 481–490. [Online]. Available: http://doi.acm.org/10.1145/1631272.1631338

[13] 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.

[14] M. Budagavi, J. Furton, G. Jin, A. Saxena, J. Wilkinson, and A. Dickerson, "360 degrees video coding using region adaptive smoothing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 750–754.

[15] K.-T. Ng, S.-C. Chan, and H.-Y. Shum, "Data compression and transmission aspects of panoramic videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 82–95, Jan. 2005.

[16] I. Bauermann, M. Mielke, and E. Steinbach, "H.264 based coding of omnidirectional video," in *Computer Vision and Graphics*. Dordrecht, The Netherlands: Springer, 2006, pp. 209–215, doi: 10.1007/1-4020-4179-9_30.

[17] K. K. Sreedhar, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "Standard-compliant multiview video coding and streaming for virtual reality applications," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 295–300.

[18] A. TaghaviNasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using layered video coding," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2017, pp. 347–348.

[19] V. Vukicevic, B. Jones, K. Gilbert, and C. V. Wiemeersch. (2017). *Webvr*. [Online]. Available: https://w3c.github.io/webvr/spec/1.1/

[20] J. Dilley, B. Maggs, J. Parikh, H. Prokop, R. Sitaraman, and B. Weihl, "Globally distributed content delivery," *IEEE Internet Comput.*, vol. 6, no. 5, pp. 50–58, Sep. 2002.

[21] E. Nygren, R. K. Sitaraman, and J. Sun, "The Akamai network: A platform for high-performance Internet applications," *ACM SIGOPS Oper. Syst. Rev.*, vol. 44, no. 3, pp. 2–19, 2010.

[22] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, 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.

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

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

[25] *Information Technology—Coded Representation of Immersive Media—Part 2: Omnidirectional Media Format*, Int. Org. Standardization, Geneva, Switzerland, Standard ISO/IEC 23090-2:2019, Jul. 2017.

[26] A. Zare, A. Aminlou, and M. M. Hannuksela, "6K effective resolution with 4K HEVC decoding capability for OMAF-compliant 360 video streaming," in *Proc. 23rd Packet Video Workshop (PV)*, New York, NY, USA, 2018, pp. 72–77. [Online]. Available: http://doi.acm.org/10.1145/3210424.3210425

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

[28] Microsoft. (2016). *Microsoft Smooth Streaming*. [Online]. Available: http://www.iis.net/downloads/microsoft/smooth-streaming

[29] Apple Inc. (2016). *Apple HTTP Live Streaming*. [Online]. Available: https://developer.apple.com/resources/http-streaming

[30] A. S. Incorporated. (2016). *Adobe HTTP Dynamic Streaming*. [Online]. Available: http://www.adobe.com/products/hds-dynamic-streaming.html

[31] *Common Media Application Format*, Standard ISO/IEC 23000-19, Int. Org. Standardization, Geneva, Switzerland, Jun. 2016.

[32] *MPEG-DASH (Dynamic Adaptive Streaming over HTTP)*, Bitmovin, Standard ISO/IEC 23009-1, 2015. [Online]. Available: https://bitmovin.com/dynamic-adaptive-streaming-http-mpeg-dash/ and https://bitmovin.com/dynamic-adaptive-streaming-http-mpeg-dash/

[33] *Omnidirectional Media Format*, International Organization for Standardization, Standard ISO/IEC 23000-20, Geneva, Switzerland, Jul. 2017.

[34] O. A. Niamut, E. Thomas, L. D'Acunto, C. Concolatom, F. Denoual, and S. Y. Lim, "MPEG DASH SRD: Spatial relationship description," in *Proc. 7th Int. Conf. Multimedia Syst. (MMSys)*, New York, NY, USA, 2016, pp. 5:1–5:8. [Online]. Available: http://doi.acm.org/10.1145/2910017.2910606

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

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

[37] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," in *Proc. ACM Multimedia Conf.*, New York, NY, USA, 2016, pp. 601–605. [Online]. Available: http://doi.acm.org/10.1145/2964284.2967292

[38] D. Salomon, *Transformations and Projections in Computer Graphics*. Secaucus, NJ, USA: Springer-Verlag, 2006.

[39] Facebook. (2017). *Next-Generation Video Encoding Techniques for 360 Video and VR*. [Online]. Available: https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding-techniques-for-360-video-and-vr/

[40] S. Afzal, J. Chen, and K. K. Ramakrishnan, "Characterization of 360-degree videos," in *Proc. Workshop Virtual Reality Augmented Reality Netw.*, New York, NY, USA, 2017, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/3097895.3097896

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

[42] *Information Technology–Coding of Audio-Visual Objects—Part 15: Carriage of Network Abstraction Layer (NAL) Unit Structured Video in the ISO Base Media File Format*, International Organization for Standardization, Geneva, Switzerland, Standard ISO/IEC 14496-15, Feb. 2017.

[43] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 107–110. [Online]. Available: http://arxiv.org/abs/1609.08729

[44] 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 (ATC)*, New York, NY, USA, 2016, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/2980055.2980056

[45] P. R. Alface, J. Macq, and N. Verzijp, "Interactive omnidirectional video delivery: A bandwidth-effective approach," *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 135–147, Mar. 2012.

[46] R. Yao, T. Heath, A. Davies, T. Forsyth, N. Mitchell, and P. Hoberman. (2014). *Oculus VR Best Practices Guide*. [Online]. Available: https://static.oculus.com/documentation/pdfs/intro-vr/latest/bp.pdf

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

[48] Pixvana. (2017). *2017 VR Award Winners*. [Online]. Available: http://blog.pixvana.com/2017-vr-award-winners

[49] E. Kuzyakov and D. Pio. (2015). *Under the Hood: Building 360 Video*. [Online]. Available: https://code.facebook.com/posts/1638767863078802/under-the-hood-building-360-video/

[50] D. Arijon, *Grammar of the Film Language*. West Hollywood, CA, USA: Silman-James Press, 1991. [Online]. Available: https://books.google.com/books?id=6bQlAQAAIAAJ

[51] V. R. Gaddam, R. Langseth, H. K. Stensland, P. Gurdjos, V. Charvillat, and C. Griwodz, "Be your own cameraman: Real-time support for zooming and panning into stored and live panoramic video," in *Proc. 5th ACM Multimedia Syst. Conf. (MMSys)*, New York, NY, USA, 2014, pp. 168–171. [Online]. Available: http://doi.acm.org/10.1145/2557642.2579370

[52] Y. Ariki, S. Kubota, and M. Kumano, "Automatic production system of soccer sports video by digital camera work based on situation recognition," in *Proc. 8th IEEE Int. Symp. Multimedia (ISM)*, Dec. 2006, pp. 851–860.

[53] P. Carr, M. Mistry, and I. Matthews, "Hybrid robotic/virtual pan-tilt-zom cameras for

autonomous event recording," in *Proc. 21st ACM Int. Conf. Multimedia*, New York, NY, USA, 2013, pp. 193–202. [Online]. Available: http://doi.acm.org/10.1145/2502081.2502086

[54] F. Chen and C. De Vleeschouwer, "Personalized production of basketball videos from multi-sensored data under limited display resolution," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 667–680, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314210000287

[55] S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio-Pardo, and A.-M. Pinna-Déry, "Film editing: New levers to improve VR streaming," in *Proc. 9th ACM Multimedia Syst. Conf. (MMSys)*, New York, NY, USA, 2018, pp. 27–39. [Online]. Available: http://doi.acm.org/10.1145/3204949.3204962

[56] Z. Gao, S. Chen, and K. Nahrstedt, "Omniviewer: Multi-modal monoscopic 3D DASH," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2015, pp. 449–452.

[57] D. B. Terry and D. C. Swinehart, "Managing stored voice in the etherphone system," *ACM Trans. Comput. Syst.*, vol. 6, no. 1, pp. 3–27, Feb. 1988. [Online]. Available: http://doi.acm.org/10.1145/35037.35038

[58] P. V. Rangan and H. M. Vin, "Designing file systems for digital video and audio," in *Proc. 13th ACM Symp. Oper. Syst. Princ. (SOSP)*, New York, NY, USA, 1991, pp. 81–94. [Online]. Available: http://doi.acm.org/10.1145/121132.121149

[59] P. Agarwal, R. R. Toledano, W. Wu, K. Nahrstedt, and A. Arefin, "Bundle of streams: Concept and evaluation in distributed interactive multimedia environments," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2010, pp. 25–32.

[60] A. Arefin, Z. Huang, K. Nahrstedt, and P. Agarwal, "4D TeleCast: Towards large scale multi-site and multi-view dissemination of 3DTI contents," in *Proc. IEEE 32nd Int. Conf. Distrib. Comput. Syst.*, Jun. 2012, pp. 82–91.

[61] S. Chen, K. Nahrstedt, and I. Gupta, "3DTI amphitheater: A manageable 3DTI environment with hierarchical stream prioritization," in *Proc. 5th ACM Multimedia Syst. Conf. (MMSys)*, New York, NY, USA, 2014, pp. 70–80. [Online]. Available: http://doi.acm.org/10.1145/2557642.2557654

[62] K. Shvachko *et al.*, "The Hadoop distributed file system," in *Proc. IEEE 26th Symp. MSST*, May 2010, pp. 1–10.

[63] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," in *Proc. 19th ACM Symp. Oper. Syst. Princ. (SOSP)*, New York, NY, USA, 2003, pp. 29–43. [Online]. Available: http://doi.acm.org/10.1145/945445.945450

[64] PPLive. (2017). *Pplive*. [Online]. Available: http://www.pptv.com/

[65] X. Zhang, J. Liu, B. Li, and Y.-S. P. Yum, "Coolstreaming/DONet: A data-driven overlay network for peer-to-peer live media streaming," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, vol. 3, Mar. 2005, pp. 2102–2111.

[66] H. Yin *et al.*, "Design and deployment of a hybrid CDN-P2P system for live video streaming: Experiences with livesky," in *Proc. 17th ACM Int. Conf. Multimedia*, New York, NY, USA, 2009, pp. 25–34. [Online]. Available: http://doi.acm.org/10.1145/1631272.1631279

[67] Akamai. (2017). *Iptvserver*. [Online]. Available: https://www.akamai.com/us/en/resources/iptv-server.jsp

[68] K. Nahrstedt, B. Yu, J. Liang, and Y. Cui, "Hourglass multimedia content and service composition framework for smart room environments," *Pervasive Mobile Comput.*, vol. 1, no. 1, pp. 43–75, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1574119205000076

[69] R. K. Sitaraman, M. Kasbekar, W. Lichtenstein, and M. Jain, "Overlay networks: An Akamai perspective," in *Advanced Content Delivery, Streaming, and Cloud Services*. 2014, pp. 305–328.

[70] K. Andreev, B. M. Maggs, A. Meyerson, and R. K. Sitaraman, "Designing overlay multicast networks for streaming," in *Proc. 15th Annu. ACM Symp. Parallel Algorithms Archit.*, 2003, pp. 149–158.

[71] L. Kontothanassis *et al.*, "A transport layer for live streaming in a content delivery network," *Proc. IEEE*, vol. 92, no. 9, pp. 1408–1419, Sep. 2004.

[72] T. Ballardie, P. Francis, and J. Crowcroft, "Core based trees (CBT)," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 23, no. 4, pp. 85–95, 1993.

[73] H. Eriksson, "Mbone: The multicast backbone," *Commun. ACM*, vol. 37, no. 8, pp. 54–61, 1994.

[74] 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.

[75] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in *Proc. 8th ACM Multimedia Syst. Conf. (MMSys)*, New York, NY, USA, 2017, pp. 199–204. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083215

[76] C. Zhou, Z. Li, and Y. Liu, "A measurement study of oculus 360 degree video streaming," in *Proc. 8th ACM Multimedia Syst. Conf. (MMSys)*, New York, NY, USA, 2017, pp. 27–37. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083190

[77] Y. Bao, H. Wu, A. A. Ramli, B. Wang, and X. Liu, "Viewing 360 degree videos: Motion prediction and bandwidth optimization," in *Proc. IEEE 24th Int. Conf. Netw. Protocols (ICNP)*, Nov. 2016, pp. 1–2.

[78] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral vision and pattern recognition: A review," *J. Vis.*, vol. 11, no. 5, p. 13, 2011, doi: 10.1167/11.5.13.

[79] D. Agrafiotis, S. J. C. Davies, N. Canagarajah, and D. R. Bull, "Towards efficient context-specific video coding based on gaze-tracking analysis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 3, no. 4, pp. 4:1–4:15, Dec. 2007. [Online]. Available: http://doi.acm.org/10.1145/1314303.1314307

[80] K. Brunnström *et al.*, "Qualinet white paper on definitions of quality of experience," Novi Sad, Tech. Rep., Mar. 2013. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00977812

[81] *Parametric Bitstream-Based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services Over Reliable Transport*, Int. Telecommun. Union, document ITU-T P.1203, Oct. 2017.

[82] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 2001–2014, Dec. 2013.

[83] T. Mäki, M. Varela, and D. Ammar, "A layered model for quality estimation of HTTP video from QoS measurements," in *Proc. 11th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Bangkok, Thailand, Nov. 2015, pp. 591–598, doi: 10.1109/SITIS.2015.41.

[84] D. Ammar and M. Varela, "QoE-aware routing for video streaming over wired networks," in *Proc. IEEE 23rd Int. Symp. Qual. Service (IWQoS)*, Jun. 2015, pp. 71–72.

[85] A. Ghosh, V. Aggarwal, and F. Qian. (2017). "A rate adaptation algorithm for tile-based 360-degree video streaming." [Online]. Available: https://arxiv.org/abs/1704.08215

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

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

[88] E. Upenik, M. Řeřábek, and T. Ebrahimi, "Testbed for subjective evaluation of omnidirectional visual content," in *Proc. 32nd Picture Coding Symp. (PCS)*, Dec. 2016, pp. 1–5.

[89] W.-C. Lo, C.-L. Fan, J. Lee, C.-Y. Huang, K.-T. Chen, and C.-H. Hsu, "360ř video viewing dataset in head-mounted virtual reality," in *Proc. 8th ACM Multimedia Syst. Conf. (MMSys)*, New York, NY, USA, 2017, pp. 211–216. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083219

[90] 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. (MMSys)*, New York, NY, USA, 2017, pp. 193–198. [Online]. Available: http://doi.acm.org/10.1145/3083187.3083210

[91] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Int. Telecommun. Union, Geneva, Switzerland, Standard ITU-R BT.500-13, Jan. 2012.

[92] A. M. Gavgani, K. V. Nesbitt, K. L. Blackmore, and E. Nalivaiko, "Profiling subjective symptoms and autonomic changes associated with cybersickness," *Autonomic Neurosci.*, vol. 203, pp. 41–50, Mar. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1566070216301096

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

[94] N. Staelens *et al.*, "On the impact of video stalling and video quality in the case of camera switching during adaptive streaming of sports content," in *Proc. 7th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, May 2015, pp. 1–6.

ABOUT THE AUTHORS

Michael Zink (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany.

He is currently an Associate Professor with the Electrical and Computer Engineering Department, University of Massachusetts Amherst, Amherst, MA, USA. He is also the Co-Director of the National Science Foundation (NSF) Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere, University of Massachusetts Amherst, where his research focuses on the closed-loop sensing system for severe weather detection and warning. As a Principal investigator for several Global Environment for Network Innovations (GENI) projects, he contributed to the creation of one of the largest at-scale Future Internet test bed. His current research interests include cyber–physical systems, multimedia distribution, and Future Internet Architectures. In the area of multimedia streaming, his work has focused on the available bitrate streaming and the quality of experience.

Dr. Zink was a recipient of the NSF CAREER Award. He received the DASH-IF Excellence in DASH Award for his work on quality adaptation for Dynamic Adaptive Streaming over HTTP (DASH).

Ramesh Sitaraman (Fellow, IEEE) received the B.Tech. degree from IIT Madras, Chennai, India, and the Ph.D. degree in computer science from Princeton University, Princeton, NJ, USA.

As a Principal Architect, he helped create the Akamai Content Delivery Network (CDN), the world's first major CDN that currently delivers a significant fraction of the Internet traffic. He retains a part-time role as the Akamai's Chief Consulting Scientist. He is currently a Professor with the College of Information and Computer Sciences, University of Massachusetts at Amherst, Amherst, MA, USA. His current research interests include the Internet-scale distributed systems, including algorithms, architectures, performance, energy efficiency, security, and economics.

Dr. Sitaraman is a member of the Association for Computing Machinery (ACM) and the American Association for the Advancement of Science (AAAS). He was a recipient of the inaugural ACM SIGCOMM Networking Systems Award for his work on the Akamai CDN, the DASH-IF Excellence in DASH Award for his work on support adaptive bitrate algorithms, the National Science Foundation CAREER Award, the College of Natural Sciences Outstanding Teacher Award, and the Lilly Fellowship.

Klara Nahrstedt (Fellow, IEEE) received the Diploma degree in mathematics and numerical analysis from the Humboldt University of Berlin, Berlin, Germany, in 1985, and the Ph.D. degree from the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA, in 1995.

She was a Research Scientist with the Institut für Informatik, Berlin, until 1989. She is currently a Ralph and Catherine Fisher Professor with the Computer Science Department and the Director of the Coordinated Science Laboratory, the University of Illinois at Urbana–Champaign, Urbana, IL, USA. Her current research interests include multimedia systems, teleimmersive systems, video 360° systems, trusted cyber–physical systems, quality-of-service management in wired and wireless networks, and distributed and pervasive mobile systems.

Dr. Nahrstedt is a Fellow of the Association for Computing Machinery (ACM). She is also a member of the German Academy of Sciences (Leopoldina Society). She was a recipient of the Humboldt Research Award, the IEEE Computer Society Technical Achievement Award, and the ACM SIGMM Technical Achievement Award.