Point Cloud Video Streaming in 5G Systems and Beyond: Challenges and Solutions

Zhi Liu [1], Qiyue Li [2], Xianfu Chen [2], Celimuge Wu [2], susumu ishihara [2], Jie Li [2], and Yusheng Ji [2]

[1]The University of Electro-Communications
[2]Affiliation not available

October 30, 2023

Abstract

Volumetric media (or hologram video), the medium for representing natural content in VR/AR/MR, is presumably the next generation of video technology and a typical use case for 5G and beyond wireless communications. To realize volumetric media applications, efficient volumetric media streaming methods are in critical demand. To this end, this article studies the challenges for and solutions to wireless transmission systems of point cloud video, which is the most popular and favored way to represent volumetric media and significantly differs from the other types of videos. In particular, we first discuss point cloud video technology and its applications and then the challenges of and solutions to point cloud video streaming in 5G systems and beyond, including encoding, tiling, viewing angle prediction, decoding, quality assessment and transmission optimization. A prototype of DASH-based point cloud video streaming system is introduced as a preliminary study, along with more simulation results to verify the performance of the proposed scheme. Moreover, future research directions are identified for providing high-quality point cloud video streaming.

Point Cloud Video Streaming in 5G Systems and Beyond: Challenges and Solutions

Zhi Liu, *Senior Member, IEEE,* Qiyue Li, *Member, IEEE,* Xianfu Chen, *Member, IEEE,* Celimuge Wu, *Senior Member, IEEE,* Susumu Ishihara, *Member, IEEE,* Jie Li, *Senior Member, IEEE,* Yusheng Ji, *Senior Member, IEEE,*

***Abstract*—Volumetric media (or hologram video), the medium for representing natural content in VR/AR/MR, is presumably the next generation of video technology and a typical use case for 5G and beyond wireless communications. To realize volumetric media applications, efficient volumetric media streaming methods are in critical demand. To this end, this article studies the challenges for and solutions to wireless transmission systems of point cloud video, which is the most popular and favored way to represent volumetric media and significantly differs from the other types of videos. In particular, we first discuss point cloud video technology and its applications and then the challenges of and solutions to point cloud video streaming in 5G systems and beyond, including encoding, tiling, viewing angle prediction, decoding, quality assessment and transmission optimization. A prototype of DASH-based point cloud video streaming system is introduced as a preliminary study, along with more simulation results to verify the performance of the proposed scheme. Moreover, future research directions are identified for providing high-quality point cloud video streaming.**

***Index Terms*—Volumetric media, hologram video, point cloud, viewing angle prediction, machine learning, beyond 5G**

I. Introduction

With the capability of providing immersive viewing experience and low-latency, high-bandwidth network requirements, virtual reality (VR) video [1] has become the quintessential application of 5G wireless communication and has been widely studied in both academia and industry. However, VR users are not allowed to move forwards or backwards, which degrades the immersive viewing experience.

Volumetric media (or hologram video) can provide users with even more immersive viewing experiences than VR with six degrees of freedom (DoF), i.e. forward/backward (surging), up/down (heaving), and left/right (swaying). Thus, users can freely select any preferred viewing angle of the 3D scene, which is not applicable in VR video, as it only provides 3 DoF. With the advantages of providing extraordinary viewing experiences and enabling viewing from different angles, volumetric media can be widely used in many areas, including education, healthcare and industry, as shown in Fig. 1, and thus has drawn great attention from both academia and industry. It is expected to be the next generation of video technology and is regarded as a typical use case of beyond 5G wireless communications [2], which is also mentioned in many industry white papers such as Samsung and NTT DoCoMo 6G white paper, to name but a few. Currently, 6 DoF games and other applications are also available. It is predicted that the global holography market for industrial applications alone will reach 22.5 billion USD by 2024 [3]. Note that the pandemic of COVID-19 draws further research interests on new holographic technique and applocations, and the market is expected to have an even bigger expansion. To further enhance volumetric media applications, networked volumetric media is in critical demand, which allows users to untether themselves.

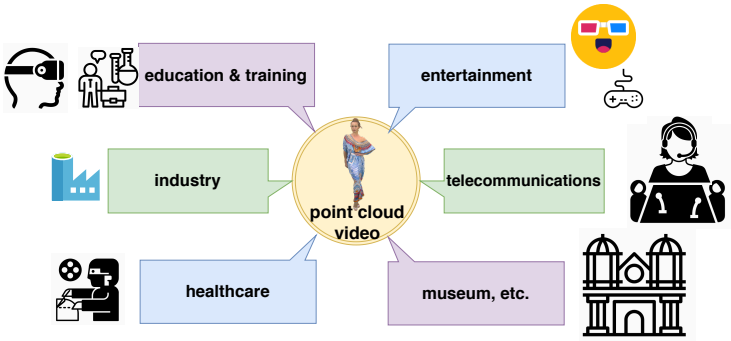

Fig. 1: Illustration of typical point cloud video applications.

However, transmitting volumetric media over the current wireless network is nontrivial. The inherent technical challenges mainly lie in the lack of transmission-friendly encoding, efficient and adaptive transmission schemes to handle the high-bandwidth and low-latency requirements, as well as effective quality metrics and accurate viewing angle prediction methods. In general, volumetric media can be represented in an image-based manner (e.g., light fields) or point cloud videos according to the adopted capturing and processing methods. Point clouds, as the most popular and favored way to represent volumetric media, are composed of points represented in 3D space, and each point is associated with multiple attributes such as coordinate and color. The focus of this article is to discuss the challenges of and solutions to the inherent components in point cloud video streaming systems and provide a prototype of a DASH-based point cloud video streaming system as a preliminary study.

A typical point cloud video streaming system, as shown in Fig. 2, mainly consists of cameras (capturing the scene of

Zhi Liu and Celimuge Wu are with The University of Electro-Communications, Japan. E-mail: liu@ieee.org
Qiyue Li is Hefei University of Technology, China
Xianfu Chen is with VTT Technical Research Centre of Finland, Finland
Susumu Ishihara is with Shizuoka University, Japan
Jie Li is with Shanghai Jiao Tong University, China
Yusheng Ji is with National Institute of Informatics, Japan

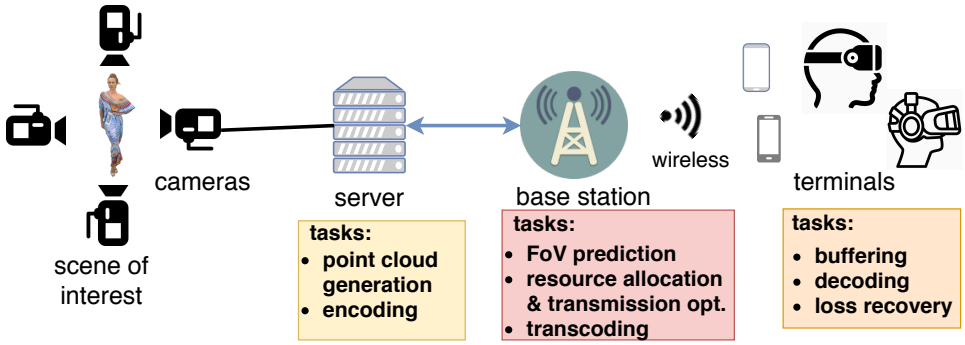

Fig. 2: Illustration of a typical point cloud video streaming system and the inherent tasks. In particular, the scene is captured by multiple RGB-Depth or RGB cameras from different angles simultaneously. These cameras are connected with a server that generates the point cloud video from the captured video. The point cloud video is then encoded/compressed for storage. The server is connected with a base station (or access point) using a cable with sufficient bandwidth. The base station predicts the viewing angle of the user and correspondingly manages the network resources to optimize the transmission performance (with/without the support of transcoding, which could generate the video at a lower rate). The wireless communications can be based on cellular (e.g., 5G) or WiFi networks, with limited bandwidth and dynamic channel conditions. Each client first buffers the video, then decodes it for playback and recovers the missing packet if required.

interest from different angles simultaneously), servers (generating the point cloud video from these captured videos and encoding/compressing the generated point cloud video), base stations or access points (allocating the available network resources to efficiently deliver the video content to the users), and users (buffering, decoding and watching the video content). The significant difference between the point cloud video data structure and traditional video frame data structure makes it distinguishable from the components in the wireless transmission of the other types of videos.

The research in this area mainly focuses on point cloud compression [4, 5] with the goal of optimizing the coding efficiency, resulting in a low source rate but a high complexity. Academia and industry (such as Fraunhofer Institute for Telecommunications and Futurewei Technologies) also pay tremendous efforts on the streaming solutions. However, point cloud video transmission has not been fully studied until now, and most of the recent point cloud video streaming schemes [6–9] are based on VR video streaming methods [1]. In other words, these schemes do not fully exploit the point cloud video features, which degrades the transmission performance. In general, research on point cloud video streaming systems is still in its infancy and requires much effort to investigate the unknowns.

To this end, this article studies the challenges for and solutions to wireless transmission systems of point cloud video. In particular, we first discuss point cloud video technology and its applications and then the challenges of and solutions to point cloud video streaming in 5G systems and beyond. A prototype DASH-based point cloud video streaming system is introduced as a preliminary study, along with more simulation results to verify the performance of the proposed scheme. Moreover, future research directions are identified for providing high-quality point cloud video streaming.

II. Characteristics, Challenges and Solutions

This section discusses the characteristics and challenges of point cloud video streaming systems, followed by alternatives to solve inherent technical problems.

A. Characteristics

The point cloud video is synthesized using the videos captured by multiple RGB-Depth cameras or RGB cameras located at different angles; the cases with RGB cameras require extra processing to estimate the depth maps. The dense camera setup results in high point cloud video quality. As a typical example, the point cloud video *Longdress*, which is shown in Fig. 1 and Fig. 2, is captured by 42 RGB cameras configured into 14 clusters (with each cluster acting as a logical RGB-Depth camera) at 30 frames per second. One spatial resolution is used for each sequence: a cube of 1024x1024x1024 voxels, which is also known as depth 10. In each cube, only voxels near the surface of the subjects are occupied. The attributes of an occupied voxel are the red, green, and blue components of the surface color. For each sequence, the cube is scaled so that it is the smallest bounding cube that contains the entire sequence [10].

Point cloud video enables 6 DoF and is very different from the other types of videos in many aspects, including the data volume, encoding, decoding, delay tolerance and inherent key research issues. As shown in Table I, the point cloud video streaming system requires an extremely high transmission rate (at the Gbps level [10]), which sometimes might surpass the capacity of a 5G wireless communication network. Moreover, the encoding tools (e.g., VPCC-TMC2) for point cloud video are very different from those for other types of videos. These tools encounter very high encoding and decoding complexity, which is one of the critical issues in implementation. In addition, the 6 DoF immersive viewing experience and high degree of user interaction require even lower network latency than a VR video streaming system. These characteristics bring

new and unique challenges for point cloud video streaming systems.

B. Technical challenges and state-of-the-art alternatives

These aforementioned characteristics lead to unique research issues. Next, we discuss these research issues and state-of-the-art alternatives for each research issue.

1) Point cloud video encoding: Point clouds are composed of points represented in 3D space, and each point is associated with multiple attributes, such as coordinates and color. The size of raw point cloud video frame is large; thus, the transmission bandwidth for a point cloud video with 30 frames per second can be as high as 6 Gbps [10], leading to a large amount of pressure on the current transmission and storage technologies. Efficient point cloud video encoding or compression hence becomes very important for high-quality point cloud video applications, which draws great attention from both academia and industry. ISO/IEC MPEG is also working on the point cloud video encoding for its international standard.

There are two major classes of point cloud encoding methods according to the different distributions of point cloud data. For point cloud data with a relatively uniform distribution in 3D space, we can leverage well-known 2D video technologies by projecting the points into 2D frames. For point cloud data with a sparse distribution, we can decompose the 3D space into a hierarchical structure of cubes and encode each point as an index of the cube to which it belongs.

Note that the existing point cloud video encoding methods require higher computation complexity than the methods for encoding traditional video [7, 8]. On the one hand, this issue poses a new computation constraint on the point cloud video streaming system design and might even require mobile edge computing or cloud computing to provide extra computation capability for encoding. On the other hand, this issue affects the point cloud video encoding design, and designing a computationally friendly encoding method thus becomes an important and urgent research issue. Another alternative is to introduce a parallel rendering and streaming mechanism to reduce the add-on streaming latency by pipelining the rendering, encoding, transmission and decoding procedures.

To further cope with adaptive and scalable point cloud video transmissions, efficient scalable encoding and multiple description coding for point cloud video are essential. As a representative work, Huang et al. [11] provide a generic point cloud codec to encode attributes, including the position and color of points sampled from 3D objects with arbitrary topology; this scheme enables scalable point cloud video encoding.

2) Tiling: Similar to VR video users, a point cloud video user may only watch a part of the whole video each time, and this viewing area is called the field of view (FoV). Instead of sending the whole video content, we may only transmit the FoV each time to avoid wasting precious communication resources. Note that unlike a VR video system, where the tiles play the same role, tiles in point cloud video streaming may have different impacts on the received video quality due to the different distances between the user and the scene.

To enable efficient streaming, the point cloud video needs to be partitioned into smaller segments by sacrificing the coding efficiency, and each segment can satisfy the FoV or partial FoV requirements and is called a *tile*. The most popular way to partition a point cloud video is through equal 3D tiling [6, 8], which evenly divides the whole point cloud video into cuboids, and each cuboid is one tile. However, uniform tiling may sacrifice the coding efficiency and ignore the video content itself and its associate semantics, where different parts are associated with different impacts on the user experience.

3) Viewing angle prediction: Predicting user behavior can more effectively serve users with limited computation and communication resources by avoiding transmitting unnecessary video tiles, providing more efficient video predownloading/buffering, etc. Viewing angle prediction has been well studied in VR and 360 video streaming systems, especially machine learning-driven viewing angle prediction. However, user behaviors in point cloud video streaming systems become more complicated and difficult to predict due to the enriched 6 DoF; i.e., point cloud video streaming system users change not only the viewing angles but also the distances from the scene.

In general, the prediction can be classified into model-based schemes (e.g., linear prediction) and machine learning-based schemes (e.g., long short-term memory (LSTM) or gated recurrent units (GRUs)). Both require viewing logs, but unfortunately, the public dataset of user viewing logs for point cloud video streaming systems is still currently unavailable. Note that collecting the viewing logs is labor intensive, and a generative adversarial network (GAN) may be useful for effectively increasing the dataset, which has been successfully used in increasing the quantity of the samples in other scenarios, such as indoor localization.

Hou et al. provide pioneering work for the viewing angle prediction of such 6 DoF systems [12]. This work considers head movements and body movements separately; an LSTM model is used for body motion prediction, and a multilayer perceptron model is used for head motion prediction with high precision. The authors generate viewing logs and achieve a better prediction performance than the baseline schemes. However, whether it is good to separate head motions from body motions is unclear, although this separation simplifies the prediction. Moreover, video information such as the saliency has not yet been considered in viewing angle prediction, which may further improve the prediction performance.

4) High decoding complexity: The high decoding complexity [7, 8] makes point cloud video streaming systems very different from the other video streaming systems. The most straightforward solution is to reduce the decoding complexity, which requires revisiting point cloud video encoding and decoding. Mobile edge computing is popular in 5G networks and beyond, which can provide augmented computation capability and may become another option to satisfy the computation requirements.

Li et al. [8] provide another alternative to the high decoding complexity, which allows the transmission of uncompressed tiles at different quality levels in addition to compressed tiles. The uncompressed tiles do not require decoding, and thus, the

TABLE I: Overview of different types of videos [1, 4, 5, 7, 8, 10]. Note that the low or high data volume is compared with a 2D video, which has a higher volume of data than normal data. The delay tolerance is also compared with the other types of videos. There is no efficient codec for VR video, which is usually projected into 2D video and then encoded using a codec for 2D video.

	2D video	**multiview video**	**360 video**	**VR video**	**point cloud video**
user freedom	NA	2 DoF with limited angles	2 DoF	3 DoF	6 DoF
data volume (whole video)	low (∼1 Mbps)	medium (∼10 Mbps)	high (∼100 Mbps)	high (∼100 Mbps)	very high (∼1 Gbps)
encoding	HEVC, VCC	MV-HEVC, HEVC, VCC	HEVC, VCC	NA	mpeg-pcc
decoding	trivial	trivial	trivial	trivial	high complexity
delay tolerance	medium	medium	high	high	very high
key unique technical issues	adaptive streaming	representation, view switching, view selection, synthetic-based optimization	viewing angle prediction, tiling, resource allocation	encoding, projection, transmission	encoding, decoding complexity, viewing angle prediction, tiling

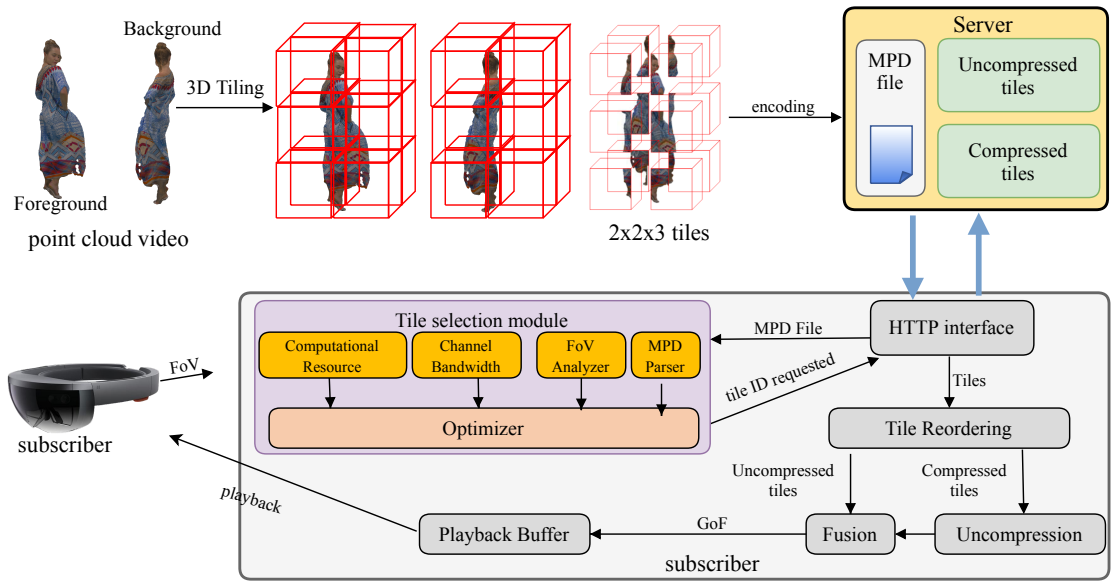

Fig. 3: System architecture of the case study.

decoding complexity is reduced. However, these uncompressed tiles consume more channel bandwidth, which requires the tradeoff between the user device's computation capability and channel bandwidth.

5) Quality assessment: Quality assessment studies how to effectively measure video quality, and it directly determines how encoding and transmission should be conducted. Considering the different distances between the user and the scene and the different visual effects between the foreground and background, the existing quality assessment methods cannot be directly adopted. The current quality assessment tool for point cloud video streaming systems is basically a variant or extension of counterparts from conventional approaches. In particular, the point cloud video PSNR can be calculated from the point-to-point distortion (if no corresponding point at the exact location can be found, the point at the nearest position is used instead) [13] or the angular similarity to measure the objective quality. This calculation is also supported by the MPEG PCC quality metric software [14], which can compare an original point cloud video with an adapted model and provide numerical values for the point cloud video PSNR. Moreover, Park et al. [6] use a more advanced model by using

Fig. 4: Overview of the prototype and the transmission performance.

the underlying quality of the representation, the level of details to the user's viewpoint and the device resolution. We can find that these schemes basically inherit the state-of-the-art quality metrics for traditional videos and do not fully consider the features of point cloud videos, especially the visual difference caused by the different distances between the scene and the viewer.

Another domain is the quality of experience (QoE), which directly reflects how users perceive the video and is affected by many factors, including the video quality, the amount of quality level switches and stalls during playback. Different from VR videos, in which each viewer only observes a part of the video each time, the point cloud video streaming system is even more complicated due to the 6 DoF. Viewers may enjoy a part of or the entire point cloud video according to the different relative distances. To the best of our knowledge, how to measure the QoE of point cloud video streaming systems is still in its infancy. Effective and efficient objective and subjective quality assessment for point cloud video streaming systems, which takes the various features of the point cloud video into consideration, remains an open issue.

6) Efficient point cloud video transmission: Most of the existing point cloud video transmission schemes [6–9] are extensions of the VR video streaming schemes [1]; i.e., the point cloud video is divided into smaller tiles, and only tiles inside the user's FoV are transmitted. Optimization can be conducted following this method to optimize the defined objective function [6]. [8] considers the high computational complexity of point cloud video decoding and solves the inherent transmission optimization problem. Note that these are model-based schemes and are not adaptive to scenarios with dynamic network conditions.

With the development of artificial intelligence, reinforcement learning has become increasingly popular in network resource allocation for video streaming systems with satisfactory performance [15]. Reinforcement learning along with other similar approaches is capable of dynamically adapting its behaviors by interacting with the environment to optimize the received video quality. The difficulties of reinforcement learning-based schemes mainly include how to properly define the reward and how to efficiently train the reinforcement learning model to achieve good performance.

Moreover, point cloud videos have three modes of distributions: live transmission, on-demand transmission and telecommunication. Different distributions have different details (e.g., buffer management), and how these should be considered in transmission optimization is an important yet challenging research issue.

III. Performance Evaluation

MPEG-DASH is an adaptive HTTP-based streaming solution that is an international standard. This section first overviews the DASH-based case study and explains the prototype with its experimental results. Then more simulation results are introduced to verify the performance of the proposed scheme.

A. Case study: a DASH-based solution

1) System overview: The DASH-based video streaming system of interest considers a server and a subscriber, as shown in Fig. 4. The server evenly partitions the point cloud video into 3D tiles and encodes each tile into different quality levels with different source rates. The compressed tiles are decoded at the server so that the uncompressed tiles at different quality levels are also available for transmission. The uncompressed tiles have larger source rates than the compressed tiles at the same quality level but do not require decoding at the user side, which releases the computation burden and serves as an alternative to solve the high decoding complexity issue. A media presentation description (MPD) file is then recorded for DASH-based point cloud video streaming. The MPD file includes the information on the file size, computation resource requirement for decoding, the number of points within a frame and the URL of each file.

The subscriber has an HTTP interface, a tile-reordering module, a decoding module, a fusion module, a tile selection module and a buffer. The tile selection module is the 'brain' of the system. It calculates the tiles residing in the subscriber's FoV and selects the proper quality level for each tile to maximize the viewing experience. In this case study, the selection is formulated as an optimization problem to optimally allocate the available computation and communication resources according to the status of the subscriber's FoV, available bandwidth and computation resources. The key point is that we

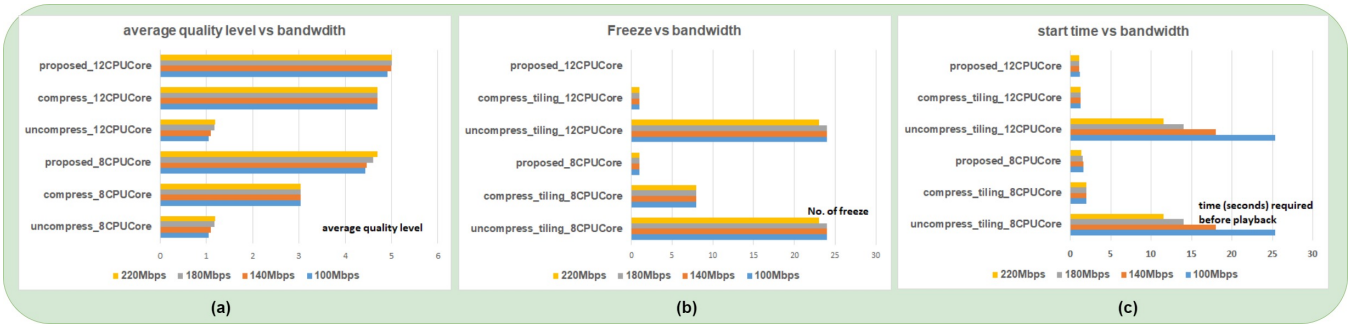

Fig. 5: Simulation results at different bandwidth conditions and CPU cores: (a) shows the simulation results in terms of the average quality levels obtained. (b) shows the simulation results in terms of the times of freeze during the playback, and (c) shows the required time before user can start watching the video.

take into consideration the user device's limited computation capability, which is insufficient for decoding, and thus includes transmitting the uncompressed tiles to reduce the computation requirement for decoding at the cost of higher bandwidth consumption. Due to the page and equation limitations, we omit the detailed explanations of the optimization used in this case study. Please refer to [8] for an example of such an optimization scheme.

2) Prototype and results: We prototype this DASH-based point cloud video streaming system with one laptop (equipped with an i5-7500 CPU, an RTX2060 GPU, 16 GB RAM and a 1 TB SSD) as the server and one PC as the client in the laboratory, as shown in Fig 4. To better demonstrate the performance, one PC is used to simulate the playback device, and the same visual range parameters ($43° \times 29°$) with HoloLens2 are used. A Xiaomi network router is used for wireless transmission with a bandwidth of up to 1200 Mbps.

We use *VS2017+PCL1.9.1+ Qt5.12* to build a point cloud video player on the client, where *PCL* is a point cloud library and *Qt* is a C++ graphical user interface application development framework. VPCC-TMC2-v7, officially recommended by MPEG, is employed as the encoding and decoding tool in the server and player, respectively. We build an HTTP web server *nginx1.16* at the server side to implement the file transmission.

The point cloud video sequence *Longdress* is used as the test sequence, which has a total of 300 frames, and each group of frames (GoF) contains 10 frames. There are approximately 780K points in each frame, and the point cloud video is divided into $2 \times 2 \times 3$ tiles. The player buffers up to 5 GoFs before it starts playing to avoid stalls.

Fig. 4(a)-(e) shows the visual experimental results. In particular, Fig. 4(a) shows one view watched by the user, and Fig. 4(c) shows the corresponding original frame. Both figures demonstrate the video from the user's viewing perspective, and we can observe that their quality levels are almost the same. With the aforementioned setup, the PSNR achieved by this system reaches up to 71 dB in terms of the point-to-point PSNR. To further verify its effectiveness, we also exhibit the viewing angle in Fig. 4(b) and only the tiles that intersect the cone (i.e., the tiles inside the FoV) are delivered to the user. As a comparison reference, the original video is presented in Fig. 4(d).

B. Simulation setup and results

To further verify the performance of the core algorithm used in the aforementioned prototype in Section III-A1, the following simulations are conducted with different network bandwidth conditions and numbers of CPU cores. The same *Longdress* with 300 frames is used, and each GOF contains 10 frames. The video is partitioned into $3 \times 3 \times 4$ tiles, and each tile is encoded into 5 quality levels. A higher quality level is associated with a higher video quality, but requires more bandwidth resources. Both compressed versions and uncompressed versions are provided for each tile. The player buffers up to 5 GoFs before the user starts playing to avoid stalls. This paper considers the baseline schemes, which only consider compressed version or uncompressed version. These schemes are optimized to select the tiles with the most proper quality levels. The performance results are from the average of 20 trails.

Fig. 5 shows the simulation results. In particular, Fig. 5 (a) shows that the proposed scheme provides the highest video quality (highest average quality level) given that it can optimally utilize the merits of both compressed and uncompressed tiles. With the increase of CPU cores, there will be more available computing resources, which results in better overall performance by enabling the user to decode more compressed tiles. With more bandwidth, our scheme achieves better performance by allowing transmitting more high quality level tiles. Fig. 5 (b) shows the number of freeze during the playback. The proposed scheme optimally uses the bandwidth and computing resources and thus has the fewest freeze. With more CPU resources available, the schemes considering compressed tiles have fewer freeze given the higher capability to decode the compressed tiles. With more bandwidth resources available, the schemes considering uncompressed tiles have fewer freeze given the higher capability to transmit the video sources at higher quality levels. Fig. 5 (c) shows the waiting time before the user can playback the video, where the proposed scheme provides the smallest waiting time. With more computing or network bandwidth resources, the waiting time decreases.

IV. Future Research Directions

We envision that the point cloud video streaming system will play a vital role in future society by offering an enriched, 6

DoF immersive viewing experience. This area opens up many exciting and critical future research directions.

A. Error-resilient point cloud video encoding and streaming

Point cloud video streaming is extremely sensitive to delays, and wireless communication is lossy in nature due to shadowing, channel fading and intersymbol interference. How to deal with the transmission error while maintaining smooth playback is vital. The possible solutions include error-resilient point cloud video encoding, source/channel coding during transmission or even error recovery at the user end, which are still open yet very important research issues for error-resilient point cloud video streaming systems.

B. AI-empowered resource allocation

Resource allocation mainly focuses on effectively utilizing network resources for a high-quality viewing experience. In particular, physical-layer resource allocation studies the modulation and coding scheme selection, transmission power and time allocation, subcarrier allocation in OFDMA systems, etc., to maximize a defined objective function such as optimizing the received video quality in terms of the PSNR or minimizing the power consumption using mathematical tools. The point cloud video streaming system brings extra constraints (such as computation) and variables.

In addition, most of the existing resource allocation schemes are mode based. How to use AI techniques to design dynamic resource allocation for point cloud video streaming systems, accounting for the computation constraint and other point cloud video features, is an interesting topic for future studies.

C. Performance boosting with edge computing

Point cloud videos are dense in terms of the number of points and require more computation resources and time for encoding, transcoding and decoding than traditional videos, which may exceed the computation capability of user devices. Low-latency computation capability hence becomes important in a high-quality point cloud video transmission system. As edge computing has already greatly accelerated the development of conventional video streaming, it can serve as an alternative to provide the required computation capability with low latency. The inherent research issue is how to explicitly use edge computing to boost the point cloud video streaming, including where to place the edge server and when and how to offload the tasks.

D. Point cloud video streaming in various types of networks and scenarios

Point cloud video is expected to be used in various scenarios, e.g., in vehicular ad hoc networks (VANETs) or information-centric networking (ICN). How to design point cloud video transmission to cope with the features of these scenarios to promote transmission performance is still open.

As an example in VANETs, the driving routing, relative distances between neighboring vehicles and network infrastructures, and user behaviors can be jointly investigated in the system to further improve the transmission performance. Note that the bandwidth conditions and future video content requests can be estimated using the information of user behaviors.

Moreover, social media platforms such as TikTok and YouTube have become increasingly popular and have new requirements and opportunities. These live-streaming systems have more stringent delay, real-time encoding, decoding and transmission scheduling requirements. How to efficiently serve point cloud video users in these services remains unsolved. Moreover, these systems have strong connections with social networks, exploring which could provide performance enhancement.

V. Conclusion

Point cloud videos enable 6 DoF viewing experiences and are expected to be the next-generation video technology. Point cloud video streaming is one fundamental research topic to facilitate point cloud video applications. In this article, we discussed the challenges of and solutions to point cloud video streaming systems, followed by a proposed DASH-based point cloud video streaming system. We prototyped this system with off-the-shelf devices and achieved satisfactory performance compared with the baseline scheme. Future research directions were also introduced to help further understand and study this topic.

References

[1] M. Zink, R. Sitaraman, and K. Nahrstedt, "Scalable 360° video stream delivery: Challenges, solutions, and opportunities," *Proceedings of the IEEE*, 2019.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *arXiv preprint arXiv:1902.10265*, 2019.

[3] I. Global Industry Analysts, "Growing Significance of Product Testing and Quality Control to drive Growth of Holography in Industrial Applications," https://www.strategyr.com/MarketResearch/ViewInfoGraphNew.asp?code=MCP-1160, 2019, [Online; accessed 2-Dec-2019].

[4] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018.

[5] C. Cao, M. Preda, and T. Zaharia, "3d point cloud compression: A survey," in *The 24th International Conference on 3D Web Technology*. ACM, 2019, pp. 1–9.

[6] J. Park, P. A. Chou, and J.-N. Hwang, "Rate-utility optimized streaming of volumetric media for augmented reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 149–162, 2019.

[7] F. Qian, B. Han, J. Pair, and V. Gopalakrishnan, "Toward practical volumetric video streaming on commodity smartphones," in *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*. ACM, 2019, pp. 135–140.

[8] J. Li, C. Zhang, Z. Liu, W. Sun, and Q. Li, "Joint communication and computational resource allocation for qoe-driven point cloud video streaming," in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[9] A. Clemm, M. T. Vega, H. K. Ravuri, T. Wauters, and F. De Turck, "Toward truly immersive holographic-type communication: Challenges and solutions," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 93–99, 2020.

[10] E. d'Eon, B. Harrison, T. Myers, and P. Chou, "8i voxelized full bodies-a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva*, 2017.

[11] Y. Huang, J. Peng, C.-C. J. Kuo, and M. Gopi, "A generic scheme for progressive point cloud coding," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 440–453, 2008.

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

[13] E. Dumic, C. R. Duarte, and L. A. da Silva Cruz, "Subjective evaluation and objective measures for point clouds—state of the art," in *2018 First International Colloquium on Smart Grid Metrology (SmaGriMet)*. IEEE, 2018, pp. 1–5.

[14] R. Mekuria, Z. Li, C. Tulvan, and P. A. Chou, "Evaluation criteria for pcc (point cloud compression)," 2016.

[15] H. Mao, R. Netravali, and M. Alizadeh, "Neural adaptive video streaming with pensieve," in *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*, 2017, pp. 197–210.