

# Point Cloud Video Streaming: Challenges and Solutions

Zhi Liu, Qiyue Li, Xianfu Chen, Celimuge Wu, Susumu Ishihara, Jie Li, Yusheng Ji

**Abstract**—Volumetric video (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 video applications, efficient volumetric video streaming is in critical demand. This article responds to the challenges of and propose 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 introduce point cloud video technology and its applications, and then discuss the challenges of and solutions to point cloud video streaming, including encoding, tiling, viewing angle prediction, decoding, quality assessment and transmission optimization. Furthermore, we explain a prototype of MPEG DASH-based point cloud video streaming system as a preliminary study, along with more simulation results to verify its performance. Finally, we identify future research directions for providing high-quality point cloud video streaming.

**Index Terms**—volumetric media, hologram video, point cloud, video streaming, 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 video users are unable to move forwards or backwards freely, which degrades the immersive viewing experience. Volumetric video (or hologram video) can conquer this limitation by providing users with six degrees of freedom (6DoF) immersive viewing experience, i.e. users can freely move forward/backward (surging), up/down (heaving), or left/right (swaying) to select the favorite viewing angle of the 3D scene, and hence enjoy three more degrees of freedom in comparison with 3DoF VR video users.

By enabling viewing from different angles, volumetric video can be widely used in many areas, including education, healthcare and industry, as shown in Fig. 1, and thus becomes a research hotspot. It is expected to be the next generation of video technology and is considered a typical application of 5G and beyond wireless communications. The global holography market for industrial applications alone is expected to reach 22.5 billion USD by 2024 [2]. Note that the pandemic of COVID-19 draws further research interests on new holographic technique and its applications, and the market is expected to further expand. To further enhance volumetric video applications, networked volumetric video is in critical demand, which allows users to untether themselves.

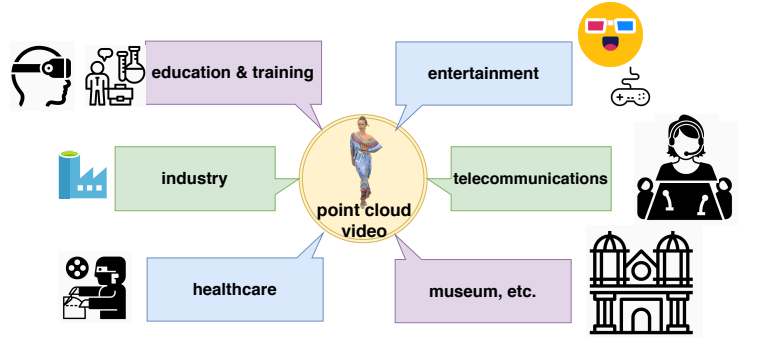


Fig. 1: Examples of typical point cloud video applications.

However, transmitting volumetric video over the current wireless network is challenging due to the lack of transmission-friendly encoding, adaptive transmission methods, as well as effective quality metrics and accurate viewing angle prediction. 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. As the most popular and favored representation of the volumetric media, point clouds consist of 3D points, each with multiple attributes, such as coordinates and color. The focus of this article is to discuss the challenges of and solutions to the point cloud video streaming and to provide a prototype of a MPEG Dynamic Adaptive Streaming over HTTP (DASH)-based point cloud video streaming system as a preliminary study. As shown in Fig. 2, a typical point cloud video streaming system mainly consists of cameras (capturing the scene of 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 (receiving, decoding and watching the video content).

The significant difference between the point cloud video data structure and traditional video data structure distinguishes the point cloud video streaming from the usual video streaming. Research on point cloud video systems have focused on point cloud compression [3, 4], with the goal of optimizing the coding efficiency so as to achieve low source rate, which, however, results in high complexity. Academia and industry (such as Fraunhofer Institute for Telecommunications) have also spent great efforts on point cloud video streaming solutions. These recent point cloud video streaming schemes [5–9] are based on the VR video streaming approach [1]. In

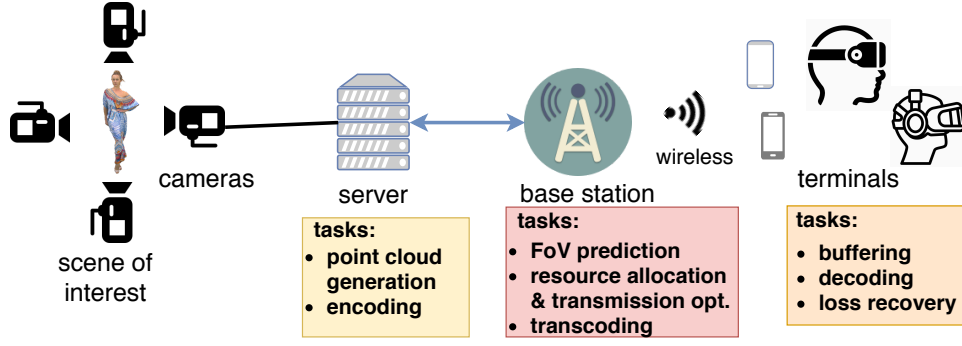


Fig. 2: Illustration of a typical point cloud video streaming system and the inherent tasks. In particular, the scene is captured by multiple RGB-depth or RGB cameras from different angles simultaneously. These cameras are connected with a server that generates the point cloud video from the captured videos. 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.

other words, these schemes do not fully exploit the point cloud video features, which degrades the system performance. For example, the high decoding complexity is not fully considered in these schemes. This brings huge computation burden to the user devices. Overall, research on point cloud video streaming systems is still in its infancy and requires much effort to investigate the unknowns.

This article responds to the challenges of and propose solutions to point cloud video streaming. In particular, we first introduce point cloud video and its applications, and then discuss the challenges of and solutions to point cloud video streaming from the aspects of encoding, tiling, viewing angle prediction, decoding, quality assessment and transmission optimization. Furthermore, we propose a DASH-based point cloud video streaming solution and prototype it as a preliminary study, along with more simulation results to verify its performance. Finally, we identify future research directions 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 these inherent technical challenges.

### A. Characteristics

Point cloud video is generated by synthesizing the videos captured simultaneously 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. A denser camera setup results in a higher point cloud video quality. As a typical example, to capture the point cloud video sequence *Longdress*, shown in Fig. 1 and Fig. 2, 14 clusters of 42 RGB cameras are used, each as a logical RGB-depth camera at 30 frames per second. This sequence uses a spatial resolution with  $1024 \times 1024 \times 1024$  voxels as a cube, within which only the surface voxels are taken. These voxels are

characterized with three colors, i.e., red, green and blue. Each cube is adjusted as small as possible to contain the whole video according to its size [10].

Point cloud video enables 6DoF immersive viewing experience 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, point cloud video streaming system requires an extremely high transmission rate (at the Gbps level [10]), which may surpass the capacity of a 5G wireless communication network sometimes. 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 6DoF 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 designing 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 alternatives to solve each of them.

1) *Point cloud video encoding*: Point clouds consist of 3D points, each with multiple attributes, such as coordinates and color. As the current transmission and storage technologies face great pressure caused by the large raw point cloud video frames, efficient encoding and compression of point cloud videos thus are becoming vital [10] and draw great attention from both academia and industry. ISO/IEC MPEG is also working on the point cloud video encoding for its international standard. To the best of our knowledge, two groups of point cloud encoding methods are usually adopted. Based on the different distributions of point cloud data, we tend to use well-established 2D video technologies to project the points into 2D

TABLE I: Features of different types of videos [1, 3, 4, 6, 7, 10]. Here the low or high data volume is compared with a 2D video, which has a higher data volume than normal data. The delay tolerance is also compared with the other types of videos. Note that VR video is usually projected into 2D video and then encoded using state-of-the-art 2D video codec such as HEVC.

	2D video	multiview video	360 video	VR video	point cloud video
user freedom	NA	2DoF with limited angles	2DoF	3DoF	6DoF
data volume (whole video)	low ( $\sim 25$ Mbps (4K))	medium ( $\sim \#view \times 25$ Mbps (4K))	high ( $\sim 400$ Mbps (4K) [8])	high ( $\sim 400$ Mbps (4K) [8])	very high ( $\sim 1$ Gbps (1080p)) [8]
encoding	HEVC, VCC	MV-HEVC, HEVC, VCC	HEVC, VCC	NA	mpeg-pcc
decoding	trivial	trivial	trivial	trivial	high complexity
delay tolerance	medium	medium	low	low	very low
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

frames for those with uniform distribution in 3D space, while we allocate those with a sparse distribution into hierarchically structured cubes and encode each point as an index of the cube [9].

Note that the existing point cloud video encoding methods require higher computation complexity than the methods for traditional video encoding [6, 7]. 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. Thus, designing a computationally friendly encoding method becomes an important and urgent research issue. Another option is to parallelly render, encode, transport and decode the point cloud video to reduce the overall streaming latency. To further cope with adaptive and scalable point cloud video transmissions, efficient scalable encoding and multiple description coding for point cloud video are essential.

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 [5, 7], 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 serve users more effectively with limited computation and communication resources by avoiding transmitting unnecessary video tiles, providing more efficient video predownload-ing/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 6DoF; 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 based prediction) and machine learning-based schemes (e.g., long short-term memory (LSTM) or gated recurrent units (GRUs) based prediction). Both require viewing logs, but unfortunately, the public dataset of user viewing logs for point cloud video streaming systems is still not available. Note that collecting the viewing logs is labor intensive, and 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 a pioneering study for the viewing angle prediction of such 6DoF systems [11]. 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 been used in VR viewing angle prediction and has boosted the prediction accuracy. How this can help the viewing angle prediction in point cloud video system requires further study.

4) *High decoding complexity*: The high decoding complexity [6, 7] makes point cloud video streaming systems

differ greatly from the other video streaming systems, the decoding in which is trivial. The most straightforward solution to handling the high decoding complexity issue is to reduce the decoding complexity, which requires revisiting point cloud video encoding and decoding. Mobile edge computing is popular in 5G and beyond wireless systems, which can provide augmented computation capability and may become another option to satisfy the computation requirement by the decoding.

Li et al. [7] provide another alternative to solve the high decoding complexity, which transmits the uncompressed tiles at different quality levels in addition to compressed tiles. The uncompressed tiles do not require decoding, and thus, can reduce the computation requirement by the decoding. However, these uncompressed tiles consume more channel bandwidth, and tradeoff between the user device's computation capability and channel bandwidth should be carefully conducted.

5) *Quality assessment*: Quality assessment studies how to effectively measure video quality, and it directly affects how video streaming is 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 in point cloud video streaming system. Current quality assessment tools for point cloud video streaming systems are basically variants or extensions 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) [12] or the angular similarity to measure the objective quality. This calculation is also supported by MPEG PCC quality metric software [13], which can compare an original point cloud video with an adapted model and provide numerical values for the point cloud video PSNR. We 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 of quality assessment is 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 video systems, in which each viewer only observes a part of the video each time, the point cloud video streaming system is more complicated due to the 6DoF. Viewers may enjoy a part of or the entire point cloud video according to the viewing angle and relative distances between the scene and the user. 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 or subjective quality assessment scheme for point cloud video streaming systems, which takes the various features of the point cloud video into consideration, remains an open issue.

6) *Optimizing point cloud video streaming*: Most of the existing point cloud video transmission schemes [5–8] are extensions of the VR video streaming schemes [1]; i.e., the point cloud video is encoded, stored and divided into smaller tiles at the server, and only tiles inside the user's FoV are

transmitted under the motion to photon latency requirement. Optimization can be conducted following this method to optimize the defined objective function [5], jointly considering the point cloud video features (such as the high data rate and decoding complexity).

These optimization problems can be solved mathematically or using AI techniques. For example, [7] considers the high computational complexity of the point cloud video decoding and solves the inherent transmission optimization problem mathematically. Note that these are model-based schemes and are not adaptive to scenarios with dynamic network conditions, which are hard to be modeled. With the development of artificial intelligence, reinforcement learning has become increasingly popular in resource allocation for video streaming systems with satisfactory performance [14]. Reinforcement learning along with other similar approaches are capable of dynamically adapting its behaviors by interacting with the environment to optimize the received video quality. The difficulties of reinforcement learning-based point cloud video streaming optimization 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. PROTOTYPE AND PERFORMANCE EVALUATION

MPEG-DASH is an adaptive HTTP-based streaming solution that is an international standard. This section first overviews the DASH-based solution 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'

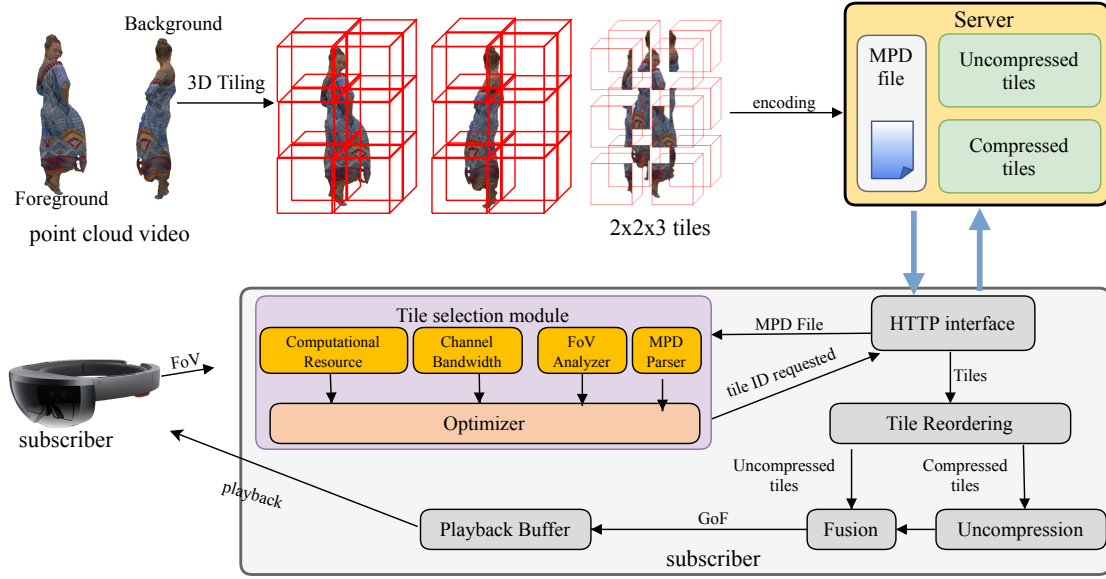


Fig. 3: System architecture of the case study.

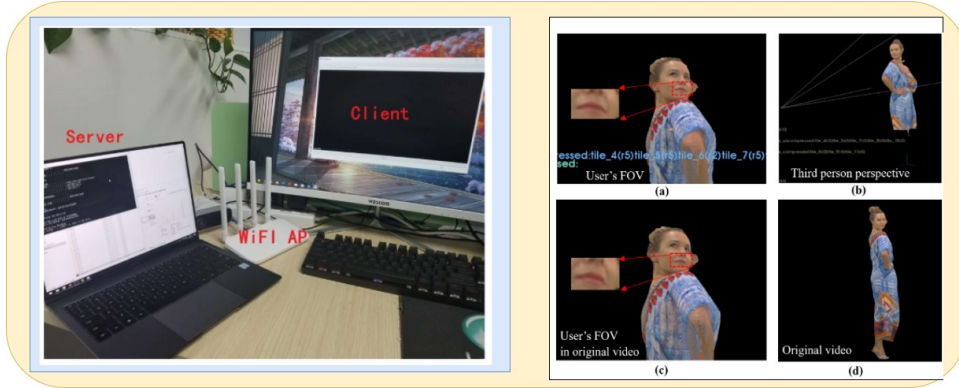


Fig. 4: Overview of the prototype and the transmission performance.

of the DASH-based point cloud video 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. This optimization problem is a non-linear integer programming problem. After the continuous relaxation of discrete variables, the new relaxed problem can be solved easily based on KKT-conditions, and branch and bound method. The key point is that we 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. Please refer to [7] for details.

2) *Prototype and results*: The prototype of this DASH-based point cloud video streaming solution uses a laptop as a server and a PC as a client in the laboratory, as shown in Fig 4. Another PC is used as the playback device and the same visual range parameters ( $43^\circ \times 29^\circ$ ) with HoloLens2 are used. An off-

the-shelf network router is used for wireless transmission with a bandwidth of up to 1200 Mbps.

VS2017, PCL1.9.1 and Qt5.12 are used to build a point cloud video player on the client. The MPEG officially recommended encoder VPCC-TMC2-v7 is used as the encoding and decoding tool for the server and player respectively. An HTTP web server *nginx1.16* is built at the server in order to realize the video transfer. The point cloud video sequence used is selected from *Longdress*, which consists 300 frames and each frame has about 780K points. Each 10 consecutive frames form a group of frames (GoF), which is divided into  $2 \times 2 \times 3$  tiles. The player buffers up to 5 GoFs before starting playback in order to avoid stalls,

Fig. 4(a)-(e) show 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 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. We also exhibit the viewing angle in Fig. 4(b) and we can observe that only the tiles that intersect the cone (i.e., the tiles inside the FoV)



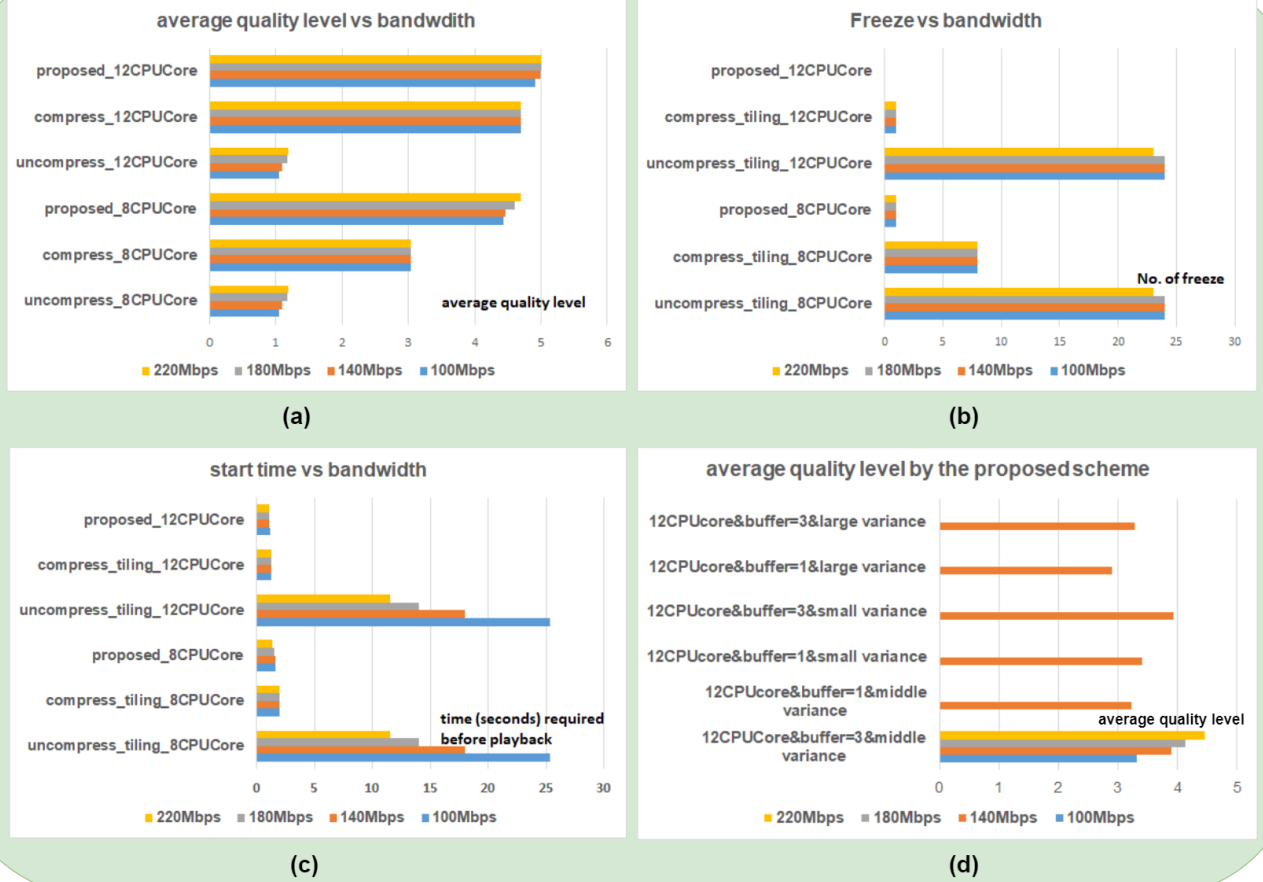


Fig. 5: Simulation results at different bandwidth conditions and CPU cores: (a) shows the simulation results in terms of the average quality levels obtained, (b) shows the simulation results in terms of the times of freeze during the playback, (c) shows the required time before user can start watching the video, and (d) shows the average quality levels obtained with different network traces and buffer lengths. In (d), buffer =  $x$  means the buffer length limit is  $x$  GoFs.

are streamed to the client. These results show the effectiveness of the proposed solution. As a reference for comparison, the original video is shown 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 in order 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 trials.

Fig. 5 (a)-(c) show 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 result in better overall performance by enabling the user to decode more compressed tiles. With larger 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.

We also run more simulations using 4G/LTE bandwidth logs

[15] with different buffer lengths, and the results are shown in Fig. 5 (d). Bandwidth logs at three scenarios (car, bus and tram) with different variations are selected, and scaled to have the same average bandwidth with Fig. 5 (a) for comparison. Although the performance drops slightly compared with the case without bandwidth variation (as shown in 5 (a)), the proposed scheme can still provide satisfactory performance. While different buffer lengths result in different performances, a longer buffer provides better performance as shown in 5 (d), especially when the channel bandwidth variation is large.

#### 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, 6DoF 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 for 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

Studies of resource allocation mainly solve how to effectively utilize the 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. Point cloud video streaming system brings extra constraints (such as computation) and variables, and makes the resource allocation more complicated.

In addition, most of the existing resource allocation schemes are mode based. How to use AI techniques to design dynamic resource allocation methods 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. 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, for point cloud video streaming in VANETs, the driving routing, relative distances between neighboring vehicles and network infrastructures, and user behaviors can be jointly investigated to improve the transmission performance.

Moreover, social media platforms such as TikTok and YouTube have become increasingly popular and bring 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 6DoF 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] 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].
- [3] 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.
- [4] 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.
- [5] 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.

- [6] 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.
- [7] 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.
- [8] 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.
- [9] Z. Liu, J. Li, X. Chen, C. Wu, S. Ishihara, and Y. Ji, "Fuzzy logic-based adaptive point cloud video streaming," *IEEE Open Journal of the Computer Society*, vol. 1, pp. 121–130, 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/WG11M74006*, Geneva, 2017.
- [11] 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.
- [12] E. Dumić, 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.
- [13] R. Mekuria, Z. Li, C. Tulvan, and P. A. Chou, "Evaluation criteria for pcc (point cloud compression)," 2016.
- [14] 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.
- [15] J. van der Hooft, S. Petrangeli, T. Wauters, R. Huysegems, P. R. Alfaced, T. Bostoen, and F. De Turck, "HTTP/2-Based Adaptive Streaming of HEVC Video Over 4G/LTE Networks," *IEEE Communications Letters*, vol. 20, no. 11, pp. 2177–2180, 2016.