

# Joint Communication and Computational Resource Allocation for QoE-driven Point Cloud Video Streaming

Jie Li<sup>†</sup>, Cong Zhang<sup>†</sup>, Zhi Liu<sup>‡</sup>, Wei Sun<sup>‡</sup>, Qiyue Li<sup>‡</sup>

<sup>†</sup>School of Computer and Information, Hefei University of Technology, China

<sup>‡</sup>Department of Mathematical and Systems Engineering, Shizuoka University, Japan

<sup>‡</sup> School of Electrical Engineering and Automation, Hefei University of Technology Hefei, China

Email: {lijie@hfut.edu.cn, zhangcong@mail.hfut.edu.cn, liu@shizuoka.ac.jp

wsun@hfut.edu.cn, liqiyue@mail.ustc.edu.cn}

**Abstract**—Point cloud video is the most popular representation of hologram, which is the medium to precedent natural content in VR/AR/MR and is expected to be the next generation video. Point cloud video system provides users immersive viewing experience with six degrees of freedom and has wide applications in many fields such as online education, entertainment. To further enhance these applications, point cloud video streaming is in critical demand. The inherent challenges lie in the large size by the necessity of recording the three-dimensional coordinates besides color information, and the associated high computation complexity of encoding. To this end, this paper proposes a communication and computation resource allocation scheme for QoE-driven point cloud video streaming. In particular, we maximize system resource utilization by selecting different quantities, transmission forms and quality level tiles to maximize the quality of experience. Extensive simulations are conducted and the simulation results show the superior performance over the existing schemes.

**Index Terms**—point cloud video, hologram video, QoE, video streaming, immersive video, resource allocation

## I. INTRODUCTION

With the rise of immersive video, technologies such as Virtual Reality (VR) and Augmented Reality (AR) are increasingly favored by users. Currently, 360 degree video combined with VR technology are the most widely used because 360 degree video use mature common video compression technology in terms of transmission. However, 360 degree video is limited by the principle and can only achieve two degrees of freedom (DOF), that is, the user can only change elevation and horizontal angle of the field of view (FOV) through the left and right and up and down movements of the head, and can not change FOV by moving the spatial position of the user, which has a great impact on user interaction. Thus, hologram video which can support six DOF, combining with wireless transmission, can give users a real immersive experience [1].

However, higher degrees of freedom also means a much larger amount of data, thus the wireless streaming of hologram videos brings more challenges. Nowadays, light field and point cloud are two commonly used technologies in hologram video. Constrained by huge data size, light field technology is rarely used in practice [2], and

dynamic point cloud is the best choice for hologram video application. The point cloud is composed of quantities of points in three-dimensional space, and each point contains RGB attributes. However the point cloud needs to store three dimensions of information, The amount of data is very large compared to ordinary video [3]. For example, when the number of 3D points is around 2800,000, the amount of traffic without any compression is approximately 78 Mbits for a single frame, which means that the bandwidth requirement of hologram video streaming is 2300Mbps when frame rate is just 30, and that would be even much larger when processing high quality point cloud video. Therefore, higher bandwidth requirement is placed on the transmission of hologram video.

At present, there are limited researches related to hologram video streaming, and it mainly focuses on the compression coding of point clouds. Hu proposes a novel point cloud compression method for attributes, based on geometric clustering and Normal Weighted Graph Fourier Transform (NWGFT) [4]. Mohammad proposes spatially sub-sample point clouds video in the 3D space method to reduce data amount and combines the adaptive scheme in DASH protocol to make the point cloud video transmission adaptive [5]. To reduce the huge data amount of hologram video, efficient video compression is an important technology. At present, some mature compression schemes have shown excellent performance. Such as Google's Draco [6] project and point cloud compression (PCC) [7] in MPEG-I. However, in our experiment tests, the running time of compressing and uncompressing are also very long. For example, the time for a typical computer with i7 processor to decode 30 frames of point cloud video may be at the minute level. Drawing on the idea of VR video streaming, Jounsup proposes a volumetric media transmission scheme that cut the point cloud video into tiles spatially, while different tiles have various quality levels, culling them or reducing their level of quality depending on their relation to the users view frustum and distance to the user [8]. However, the long uncompressing and decoding time before playback is avoided in their research, just taking the transmission time into account. Arunkumar studies the relationship between decoding

time and transmission time in VR video [9], and proposes to balance the relationship between the two by partially transferring raw tiles of VR video.

Due to the huge data traffic requirement, it is impossible to directly transmit the raw point cloud video. We can use the idea of VR video processing method [10] by cutting the point cloud video into 3D tiles and only transmitting them within the user's viewport. Besides, the current mainstream point cloud video compression and decompression technology has high efficiency, and supports multi-core, multi-thread operation, which can compress one second of point cloud video of 500 Mbits to 4 Mbits, with 23 seconds uncompressing time. However the PSNR will be dropped from 75dB to 66dB. In this manuscript, we try to jointly allocate the computational resources of the user's playback device as well as the communication resources, so that the users can adaptively select the video quality according to the processing capability, network bandwidth, and user viewport to maximize the QoE. To the best of our knowledge, this is the first research jointly considers communication resource and computational resource for hologram video streaming.

Our contributions are summarized as follows:

- (1) we propose the joint communication and computational resources allocation framework for point cloud video streaming, in which computational resources are represented as the performance of decoding the compressed 3D tiles on user's playback device.
- (2) We propose a QoE evaluation metric for dynamic point cloud video based on user's perspective and point cloud characteristics, combining distance and quality weights of different 3D tiles.
- (3) We establish a transmission system that balances the utilization of communication resources and computational resources to maximize the QoE value of dynamic point clouds video.

The remainder of this paper is organized as follows. Section II introduce the proposed transmission system model and introduce how our tiling selection strategy can improve system resource utilization. Section III introduces the joint mathematical allocation problem to maximize user's QoE value. In section IV, experiments are carried out to verify the feasibility of our transmission scheme and its performance. And we finally conclude this paper in Section V.

## II. QOE DRIVEN HOLOGRAM VIDEO STREAMING SYSTEM

In this section, we will introduce our hologram video streaming system and the basic optimization idea of how to maximize user's QoE when jointly considering allocation of communication and computational resources.

### A. System Architecture

As mentioned above, due to huge amount of data in point cloud video, it cannot be directly transmitted via the wireless network. Similar to VR video streaming, hologram video is also sliced into multiple 3D tiles and only part of tiles are transmitted according to user's FoV

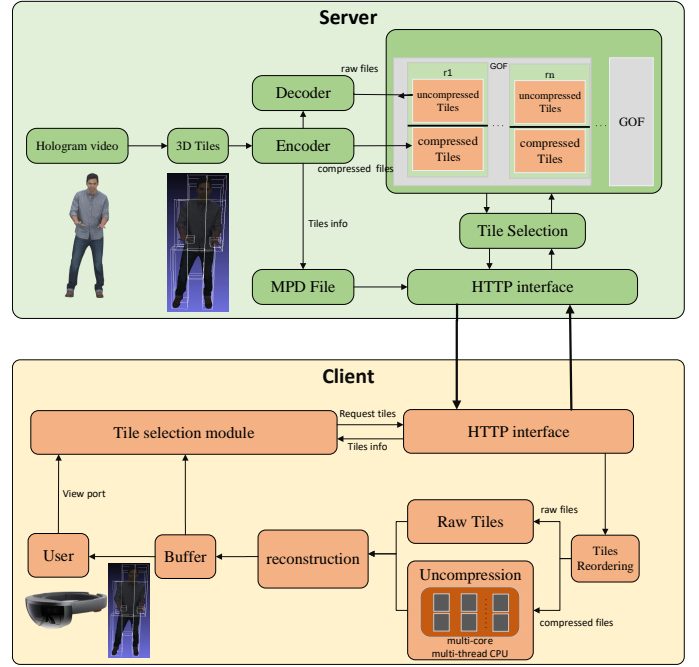


Fig. 1. Dynamic point cloud adaptive transmission system

to reduce bandwidth requirement [8]. Besides, the video compression technology such as H.265 can be applied to relatively keep video quality while dramatically reducing data size. However, due to the particularity of 3D video, decompression will consume a lot of computational resources and running time. Although the multi-core and multi-threading operation are supported, with different available computational resources and time-varying network bandwidth, how to optimize the transmission scheme to achieve a better user's QoE still remains a non-trivial issue. In this paper, we propose a dynamic point cloud adaptive streaming system to optimize the transmission time and uncompressing time by adjusting the balance between computational resources and communication resources, as shown in Fig. 1.

The whole system can be divided into two parts, the server side and the client side. The server performs pre-processing on point cloud video, cutting it into 3D tiles, and compressing each tile into different quality representations with Group of Frame (GOF) as the minimum unit. The information of all the point cloud video tiles is stored in a Media Presentation Description (MPD) file, similar to MPEG-DASH [11], and the server will send corresponding tiles after receiving streaming requests from the client.

After all the tiles are received and uncompressed, they will be reconstructed, as show in Fig. 2, which can be seen that some tiles quality changes in the reconstructed frame. Then it will be sent to the buffer, waiting for rendering and playback. Obviously to maintain a continuous playback, the buffer cannot be drained.

The core of the client side is tile selection module. It calculates user's FoV, selects the tiles residing inside FoV with appropriate quality representation to maximize user's QoE, according to wireless bandwidth status, buffer depth, and the available computational resources.

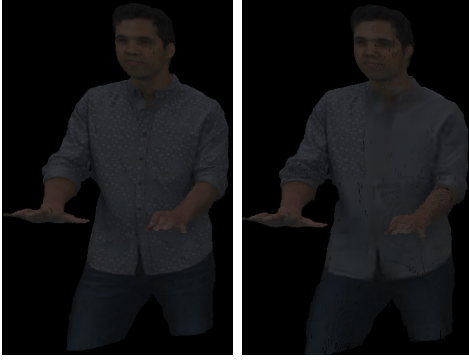


Fig. 2. Left:raw frame. Right:reconstruct frame

### B. Tile Selection Module

Although the compression efficiency of dynamic point clouds is already very high, most of the current test data sets just have simple and low quality scenarios with less than one million points. With the development of point cloud acquisition technology, the future point cloud video will be the reappearance of various realistic and complex scenes [12]. For example, a street scene, there will be hundreds of pedestrians and vehicles and other complex things, the number of points will be more than one billion, and the size of compressed video will still be extremely large and hard to transmit. Thus We use the idea of VR video processing method by cutting the point cloud video into 3D tiles and only transmitting them within the users viewport. The 3D tile cutting method is shown Fig.3.

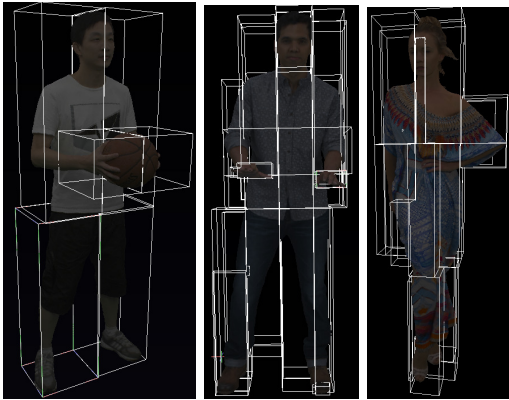


Fig. 3. 3D Tiles samples

With the help of tiling scheme, we can reduce the transmission of redundant tiles and select different quality level for each tile to optimize the transmission. The former can reduce the bandwidth requirement by detecting the user's viewport, and the latter can be implemented by transmitting the compressed version of some tiles, while the other raw tiles can be requested to improve reconstructed quality. Although the transmission time is increased, the computational resources required for decoding can be reduced to realize real-time uncompressing and playback. It is most suitable for the scenario where communication resource utilization is not high and computational resources are insufficient.

## III. PROBLEM FORMULATION

### A. Computational Resources and Decoding Time

Through the available data set [13], it can be found that the decoding time and the required computational resource of point cloud video is related to it's total number of points and quality level, as shown in Table I.

TABLE I  
COMPUTATIONAL RESOURCE AND BIT RATE LEVEL, NUMBER OF POINTS

Sequence	Rep	No.point	Encoder time(s)	Decoder time(s)
Basketball	r1	84258322	8366	112.9
Basketball	r3	91814344	8895	118.5
Basketball	r5	92079625	9818	120.6
Queen	r1	30274532	6114	41.38
Queen	r3	32735750	6414	43.46
Queen	r5	31426902	6764	45.02

Assume the minimum required computational resource of all tiles with lowest quality is  $C$ , we can find a parameter  $\partial_{g,k}$  so that the required computational resource of tile  $k$  with quality level  $r$  in GOF  $g$  can be described as  $C_{g,k,r} = \partial_{g,k,r} \times C$ , and it can be quantized into clock cycles to describe the required time to decode. All the consumed computational resource  $C_{g,k,r}$  can be regarded as known parameters stored in the MPD file. We assume  $CU_1$  is the computational resources that one single processor core of user's playback device can provide in one GOF time, and  $NC$  is the number of cores, then  $CU_{NC} = NC \times \tau_{NC} \times CU_1$ , where  $CU_{NC}$  is the total computational resource available, and  $\tau_{NC}$  means the conversion efficiency when the uncompression program is running in multi-core and multi-thread mode. Obviously, the consumed computational resource cannot exceed the amount that the user's device can provide.

Then we can obtain the decoding time  $Td_g$  for GOF  $g$ , which can be expressed as:

$$Td_g = \frac{f}{fps} \times \frac{\sum_{k=1}^{M_g} e_{g,k} \times \sum_{r=1}^R C_{g,k,r} \times x_{g,k,r}}{NC \times \tau_{NC} \times CU_1} \quad (1)$$

where  $f$  is the number of frames in one GOF and  $fps$  is the number of frames of the point cloud video.  $e_{g,k}$  indicates the transmission form of the tiles, and when  $e_{g,k} = 1$  indicates that the compressed version is selected, while  $e_{g,k} = 0$  means that the raw tile is selected.  $M_g$  means the number of tiles in user's FOV, and  $R$  is the total quality levels available.  $x_{g,k,r}$  is a binary variable and  $x_{g,k,r} = 1$  means compressed version of tile  $k$  with quality level  $r$  in GOF  $g$  is transmitted, otherwise  $x_{g,k,r} = 0$ .

In our system, all the retrieved tiles are inserted to a buffer before playback. Let the  $Tb_g$  is the current depth of the buffer, which means the frame length (measured in time) of point cloud video stored in buffer. Obviously to maintain a continuous playback, we have  $Tb_g > 0, \forall g \in [1, G]$ .

Then the dynamics of the playback buffer can be expressed by the following difference equations:

$$Tb_g = Tb_{g-1} - Tu_g + Ti_g \quad (2)$$

where  $Ti_g$  is the increased playback time when GOF  $g$  enters the buffer, which is a constant value and equals to  $\frac{f}{fps}$ .  $Tu_g$  indicates the time consumed by the transmitting all the tiles resides in FOV and decoding the compressed tiles in GOF  $g$ . Thus  $Tu_g$  can be expressed as:

$$Tu_g = Ts_g + Td_g \quad (3)$$

$$Ts_g = \frac{\text{raw tiles.size} + \text{compressed tiles.size}}{Bw_g} \quad (4)$$

$$\begin{aligned} &= \frac{\sum_{k=1}^{M_g} e_{g,k} \times \sum_{r=1}^R \text{bin}S_{g,k,r} \times x_{g,k,r}}{Bw_g} \\ &+ \frac{\sum_{k=1}^{M_g} (1 - e_{g,k}) \times \sum_{r=1}^R \text{ply}S_{g,k,r} \times x_{g,k,r}}{Bw_g} \end{aligned} \quad (5)$$

where  $\text{bin}S_{g,k,r}$  indicates the data size of the compressed tile  $k$  with quality level  $r$  in GOF  $g$ , while  $\text{ply}S_{g,k,r}$  represents the data size of raw tile  $k$  with quality level  $r$  in GOF  $g$ , and the parameters are all stored in the MPD file.  $Bw_g$  is the predicated wireless bandwidth at the time of GOF  $g$ .

### B. The Quality of Experience Model

Unlike to VR video, the hologram video can support 6DOF, and the 3D tiled viewing process is shown in Fig. 4. During the playback, it can be found that different tiles have different distances from the viewer's position, and they have different contributions to the total QoE. We assume that the closer the tile is to the user's viewpoint position, the greater its contribution. Besides, for each 3D tile, there are up to  $R$  compressed quality levels available, and when different quality level is requested, different QoE will be yielded.

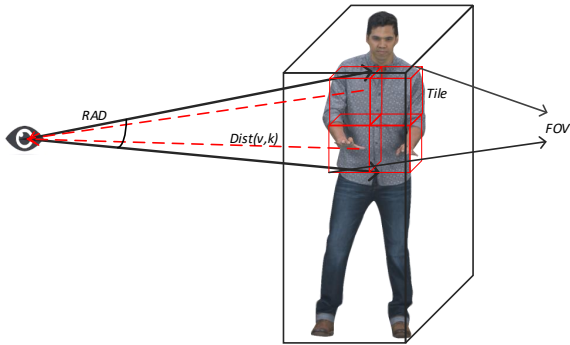


Fig. 4. Filed of view

Then for a single 3D tile  $k$  in GOF  $g$ , we can define its QoE contribution as:

$$Q_{g,k,r} = \sum_{r=1}^R p_{g,k} \times r \times QT_{g,k} \times x_{g,k,r} \quad (6)$$

where  $p_{g,k}$  and  $QT_{g,k}$  are distance weight and quality weight respectively for 3D tile  $k$  in GOF  $g$ .

In this paper we define  $p_{g,k}$  as:

$$p_{g,k} = \frac{1}{\text{dist}_{g,k}(v_g, \text{pos}_{g,k})} \quad (7)$$

where  $v_g$  is the user viewpoint position when watching GOF  $g$ , and  $\text{pos}_{g,k}$  is the position of tile  $k$  in GOF  $g$ .  $\text{dist}_{g,k}$  is a function of the tangential distance from the viewpoint.

Regarding the quality weight  $QT_{g,k}$ , due to unevenness of the point cloud video, the number of points contained in each tile is different. Thus we can define the quality weight for a 3D tile  $k$  as the ratio of number of points in that tile to the total number of points in whole FoV, which can be expressed as:

$$QT_{g,k} = \frac{N_{g,k,R}}{\sum_{k=1}^{M_g} N_{g,k,R}} \quad (8)$$

The quality of the point cloud image is generally evaluated by level of density (LOD), which represents the number of points in the unit volume, and the user views each point in the FOV area. Then for the whole point cloud video, we can define the QoE as [14]:

$$\text{QoE} = \log \left( \frac{\sum_{g=1}^G \sum_{k=1}^{M_g} Q_{g,k,r}}{\sum_{g=1}^G \sum_{k=1}^{M_g} Q_{g,k,R}} \right) \quad (9)$$

### C. Total Problem Formulation

Overall, the QoE driven communication and computational resource allocation point cloud video streaming can be formulated as:

$$\max_{x_{g,k,r}, e_{g,k}} \text{QoE} = \log \left( \frac{\sum_{g=1}^G \sum_{k=1}^{M_g} Q_{g,k,r}}{\sum_{g=1}^G \sum_{k=1}^{M_g} Q_{g,k,R}} \right) \quad (10)$$

$$= \log \left( \frac{\sum_{g=1}^G \sum_{k=1}^{M_g} \sum_{r=1}^R \left( \frac{1}{\text{dist}_{g,k}} \times r \times QT_{g,k} \times x_{g,k,r} \right)}{\sum_{g=1}^G \sum_{k=1}^{M_g} Q_{g,k,R}} \right) \quad (11)$$

s.t.

$$\sum_{r=1}^R x_{g,k,r} = 1, \forall g \in [1, G], k \in [1, M_g] \quad (12)$$

$$Tb_g > 0, \forall g \in [1, G] \quad (13)$$

This is a constrained nonlinear 0-1 programming problem. Among them Equ. 13 is as follows:

$$\begin{aligned} Tb_g &= Tb_{g-1} - Tu_g + Ti_g \\ &= Tb_{g-1} - \left( \frac{\sum_{k=1}^{M_g} [e_{g,k} \times \sum_{r=1}^R (\text{bin}S_{g,k,r} \times x_{g,k,r})]}{Bw_g} \right. \\ &\quad \left. + \frac{\sum_{k=1}^{M_g} [(1 - e_{g,k}) \times \sum_{r=1}^R (\text{ply}S_{g,k,r} \times x_{g,k,r})]}{Bw_g} \right) \\ &\quad + \frac{f}{fps} \times \left( \frac{\sum_{k=1}^{M_g} [e_{g,k} \times \sum_{r=1}^R (C_{g,k,r} \times x_{g,k,r})]}{CU_{NC}} \right) + \frac{f}{fps} \end{aligned} \quad (14)$$



It can be simplified to

$$Tb_g - Tb_{g-1} = a_g = f(x_{g,k,r}) \quad (15)$$

$$Tb(0) = b > 0 \quad (16)$$

The constraint is a first-order non-homogeneous linear difference equation,  $x$ ,  $e$  is a binary vector, and  $b$  is the initial buffer size. Then we have

$$Tb(g) = \sum_{i=1}^g (a_i) + b = \sum_{i=1}^g [f(x_{i,k,r})] + b > 0 \quad (17)$$

The problem is a nonlinear integer programming (NIP) problem. If the solution variable of the problem is relaxed to a continuous variable, its objective function and constraint are convex functions and convex sets that are easy to solve. Therefore, the paper firstly relaxes the problem, converts the NIP problem into the NP problem, and uses the KKT-condition to find the solution of the nonlinear programming problem. Then the branch-and-bound method is used to solve the 0-1 variable solution of the original problem.

#### IV. PERFORMANCE EVALUATION

We built a simulation platform to verify the feasibility of our proposed transmission scheme, calculate the different results under different computational resources  $CU_{NC}^i$  and different communication resources  $Bw^i$ , and then compare with the traditional tiling scheme, that is, the scheme of only transmitting the compressed tiles, in terms of system resource utilization and QoE values.

Limited to the effectiveness of bandwidth prediction, we set the whole dynamic point cloud time to 2s, including 3 GOFs per second. The tiling scheme  $M = 2 \times 2 \times 6$ , total quality levels  $r = 5$ , initial buffer length  $b = 2s$ . The required computational resource with the smallest size among all the tiles is set to  $C = 1$ , and the computational resources required for each tile are obtained according to the size. Other simulation parameters are listed in Table II:

TABLE II  
SIMULATION PARAMETERS SETTING

Test Group ID	NC (number of cores)	Bw(Gbps)
1	2	0.5
2	2	0.8
3	2	1.0
4	4	0.5
5	4	0.8
6	4	1.0
7	6	0.5
8	6	0.8
9	6	1.0

Firstly, we study the performance of our proposed system. Fig.5 is the number of transmitted raw tiles and compressed tiles under various conditions. It can be seen that when the computational resources are relatively large, the scheme preferentially transmits the compressed tile to improve the resource utilization. When the bandwidth is relatively large, the scheme preferentially transmits the

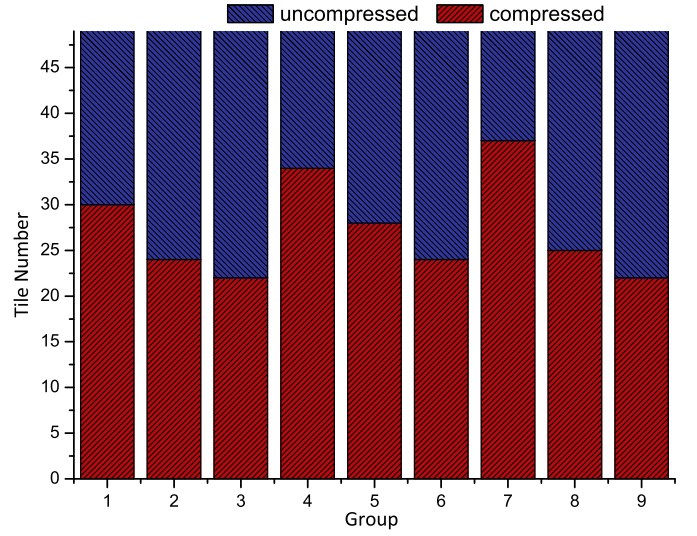


Fig. 5. Number of tile transmission forms under different schemes

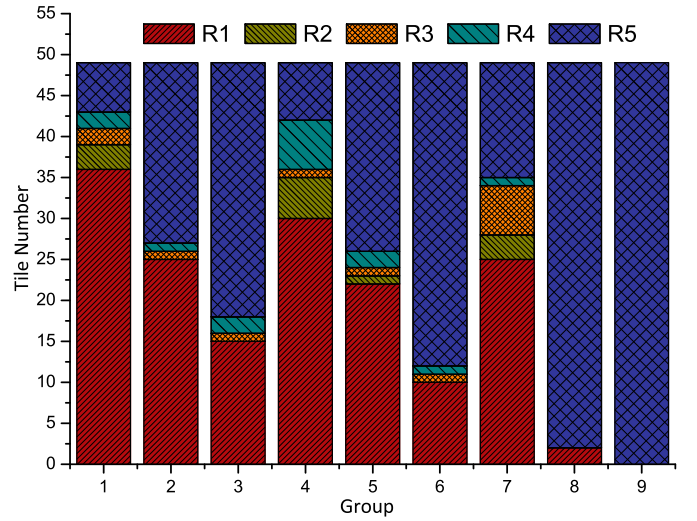


Fig. 6. Comparison of the number of tile in different quality under different condition

uncompressed tile to make full use of the communication resources. In order to achieve the best system resource utilization.

Fig.6 illustrates the difference in terms of tile numbers in different quality level transmitted under different conditions. When the computational resources and communication resources are low, the system will choose to transmit tiles at a lower bit rate and quality level to reduce resource requirements, and when the system resources are sufficient, high different quality tile will be transmitted to achieve the best quality experience.

Then we compare with traditional hologram transmission system, which delivers only compressed tiles. Fig.7 illustrates system resource utilization expressed by  $p = \frac{1}{2} \times (\frac{tiles.C}{system.C} + \frac{tile.B}{system.B})$ , where  $B$  means communication resource and  $C$  means computational resource. It can be seen that compared to the traditional hologram transmission scheme, our system has a better system resource utilization.

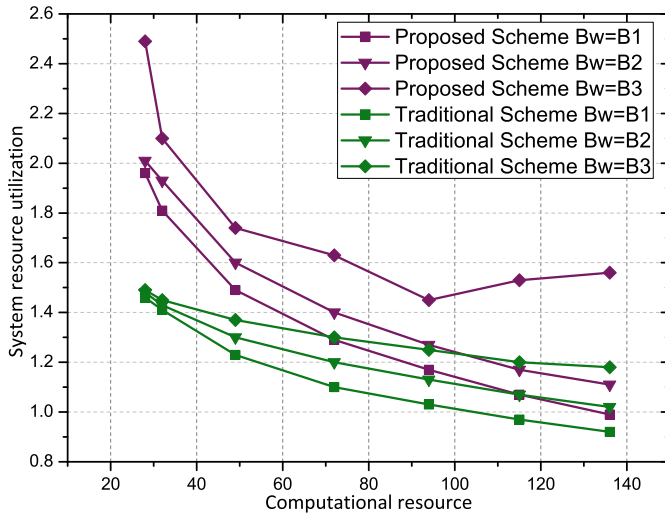


Fig. 7. System utilization comparison

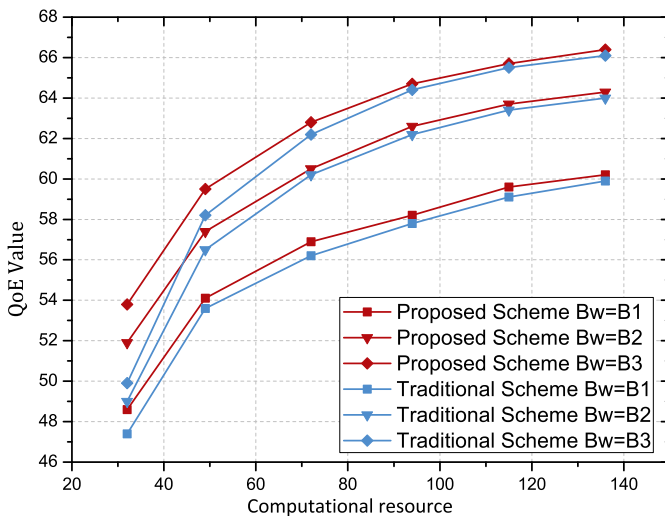


Fig. 8. Maximum QoE comparison

The Fig.8 is a comparison of QoE values between our scheme and the traditional scheme. It can be seen that since our transmission scheme makes full use of communication resources and computational resources, it can maximize the quality level of the tiles, and can obtain a greater QoE value. Besides it can be seen that When the computing resources are low, which is typical scenario for mobile user playback equipment, the QoE value is significantly higher than the traditional solution.

## V. CONCLUSION

The streaming of hologram video is a major problem at present. Many studies focus on how to improve the efficiency of compression, that is, to maximize the utilization of computational resources, while neglecting the utilization of communication resources in the streaming process. End up with the solution that the transmission time is reduced, but the decoding and uncompressing time is greatly improved, thus it is still hard to implement continuous playback.

In this paper, we propose an idea of jointly allocating the communication and computational resources for hologram video streaming. And a corresponding point cloud adaptive transmission system is established. Then the allocation scheme is formulated into a 0-1 integer programming problem, by selecting different quantities, transmission forms and quality level tiles to maximize the quality of experience. Finally, we conduct extensive simulations to show that our system can outperform over the traditional point cloud transmission scheme in terms of system resource utilization and QoE.

## ACKNOWLEDGMENT

This research is supported in part by the Fundamental Research Funds for the Central Universities, Grant No. JZ2019HGTB0089, JZ2018HGTB0253 and PA2019GDQT0006, National Natural Science Foundation of China, Grant No. 51877060.

## REFERENCES

- [1] M. J. Richardson and J. D. Wiltshire, *What is a Hologram?* IEEE, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8068913>
- [2] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, Oct 2017.
- [3] L. Cui, H. Xu, and E. S. Jang, "Hybrid color attribute compression for point cloud data," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, July 2017, pp. 1273–1278.
- [4] Y. Xu, W. Hu, S. Wang, X. Zhang, S. Wang, S. Ma, and W. Gao, "Cluster-based point cloud coding with normal weighted graph fourier transform," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 1753–1757.
- [5] M. Hosseini and C. Timmerer, "Dynamic adaptive point cloud streaming," in *Proceedings of the 23rd Packet Video Workshop*, ser. PV '18. New York, NY, USA: ACM, 2018, pp. 25–30. [Online]. Available: <http://doi.acm.org/10.1145/3210424.3210429>
- [6] "Google.2018.draco: 3d data compression.retrieved march 3,2018 from," <http://github.com/google/draco>.
- [7] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokua, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "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, March 2019.
- [8] J. Park, P. A. Chou, and J. 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, March 2019.
- [9] A. Ravichandran, I. K. Jain, R. Hegazy, T. Wei, and D. Bharadia, "Facilitating low latency and reliable vr over heterogeneous wireless networks," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 723–725. [Online]. Available: <http://doi.acm.org/10.1145/3241539.3267781>
- [10] J. Li, R. Feng, Z. Liu, W. Sun, and Q. Li, "Modeling qoe of virtual reality video transmission over wireless networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–7.
- [11] "Information technology-dynamic adaptive streaming over http (dash)-part 1: Media presentation description and segment formats.iso/iec 23009-1:2014/pdam 3.2015-02-20," <https://mpeg.chiariglione.org/standards/mpeg-dash/media-presentation-description-and-segment-formats>.
- [12] O. Schreer, I. Feldmann, S. Renault, M. Zepp, M. Worchel, P. Eisert, and P. Kauff, "Capture and 3d video processing of volumetric video," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 4310–4314.
- [13] "mpeg point cloud compression common test condition reporting template n18175," <https://www.interdigital.com/download/5d2072018934bf9bb4000968>.

- [14] W. Huang, L. Ding, H. Wei, J. Hwang, Y. Xu, and W. Zhang, "Qoe-oriented resource allocation for 360-degree video transmission over heterogeneous networks," *CoRR*, vol. abs/1803.07789, 2018. [Online]. Available: <http://arxiv.org/abs/1803.07789>