A QoE Model in Point Cloud Video Streaming

Jie Li†, Xiao Wang‡, Qiyue Li‡ and Zhi Liu‡

†School of Computer and Information, Hefei University of Technology, China
‡School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, China

Abstract—Point cloud video has been widely used by augmented reality (AR) and virtual reality (VR) applications as it allows users to have an immersive experience of six degrees of freedom (6DoFs). Yet there is still a lack of research on quality of experience (QoE) model of point cloud video streaming, which cannot provide optimization metric for streaming systems. Besides, position and color information contained in each pixel of point cloud video, and viewport distance effect caused by 6DoFs viewing procedure make the traditional objective quality evaluation metric cannot be directly used in point cloud video streaming system. In this paper we first analyze the subjective and objective factors related to QoE model. Then an experimental system to simulate point cloud video streaming is setup and detailed subjective quality evaluation experiments are carried out. Based on collected mean opinion score (MOS) data, we propose a QoE model for point cloud video streaming. We also verify the model by actual subjective scoring, and the results show that the proposed QoE model can accurately reflect users’ visual perception. We also make the experimental database public to promote the QoE research of point cloud video streaming.

Index Terms—Volumetric Video Streaming, Point Cloud, Quality of Experience, Subjective Quality Evaluation

I. INTRODUCTION

In recent years, volumetric video has become more and more popular, and there are many augmented reality (AR) and virtual reality (VR) applications in education, entertainment, medicine and other industries [1]. Compared with 360-degree video, volumetric video can provide six degrees of freedom (6DoFs) immersive experience to achieve extraordinary viewing procedure.

Currently, the most popular data format of volumetric video is the point cloud [2, 3] due to its flexibility and simplicity. Point cloud is composed of quantities of points in three-dimensional space, and each point contains space coordinates and RGB attributes. Due to having more multidimensional information than regular video (including 360-degree video), the bandwidth requirement for transmitting point cloud video is extremely large. Taking a commonly used 30 frames per second (fps) video as an example, when the number of points is approximately 760,000 per frame, the volume of data without any compression is approximately 12 MB, which means the point cloud video streaming bandwidth demand is up to 2.9 Gbps [4]. This requirement will be even higher when delivering high-quality video.

Adaptive streaming is one of the basic topics for many related applications. Similar to the Dynamic Adaptive Streaming over HTTP (DASH) used in traditional VR videos [5, 6, 7], point cloud video can be divided into 3D tiles and compressed into different quality representations, which are then uploaded to the video server [8]. The client selects which tiles to transmit and at which quality representation under the premise of satisfying current bandwidth condition, according to current network analysis, user’s FoV prediction and current buffer status of playback device. Finally, according to selected tile quality representation, the client can obtain the corresponding tiles in the user’s FoV from the server, integrate the tiles into a point cloud video, and render the point cloud video to enjoy.

Point cloud video can be compressed to further reduce bandwidth demand. For example, MPEG Point Cloud Compression (PCC) codec can achieve a fairly high compression ratio [9]. Thus the network bandwidth consumption can be greatly reduced during transmission. However, decompression of point cloud video will consume computational resources and relatively long decoding time. For example, a 511 MB volumetric video called longdress [10] containing 30 raw frames can be compressed into a binary file of only 21 KB using the MPEG-VPCC-TMC2 v12.0 codec, but it takes approximately 34 seconds to decode.

A point cloud video streaming system should be optimized to maximize users’ perceived quality of experience (QoE). Yet there is still a lack of research on the QoE model for point cloud video streaming. QoE models are generally divided into objective metric and subjective metric. The former is usually measured by objective quality metric such as Peak Signal to Noise Ratio (PSNR). The latter generally conducts user subjective experiments, then statistics the scores reflecting user’s direct viewing experience, and finally obtains a QoE model through mean opinion score (MOS).

The QoE model of point cloud video streaming is different from traditional system [11]. First of all, the ob-
jective PSNR of traditional video usually only represents color information of the video \[^{[12]}\] , and cannot represent position information of pixels in point cloud video. So the traditional PSNR cannot be directly used to measure the quality of point cloud video, especially the number and positions of points will change after compression. Secondly, subjective QoE model usually includes stall and quality switch due to network bandwidth variations in video streaming. Different from traditional video, the original frame of point cloud video can be compressed to reduce transmission time, but it takes a long time to decode. Thirdly, commonly used 3D tiling scheme in point cloud video streaming will cause the viewpoint distance effect. That is the user’s perception of quality switch caused by faraway tiles is weaker than that caused by nearby ones. Therefore, the distance between different tiles to the user’s viewpoint affect the quality switch and user’s quality of visual perception, while traditional video does not.

In this paper, to establish a more comprehensive QoE model for point cloud video streaming, we explore the influencing factors related to QoE metric, which mainly include objective quality of point cloud video, stall and quality switch during the video streaming. For objective quality, we extend PSNR to point cloud video, named Point Cloud PSNR (PCPSNR), by integrating position and color information of all pixels. For the playback stall, we consider the transmission time and decoding time changes caused by point cloud compression and decompression to measure the stall time more accurately. For the quality switch, we also consider the viewpoint distance effect to better reflect the effect of the distance from tile to user’s viewpoint on quality switch.

We set up an experimental system to simulate the point cloud video streaming procedure. Firstly, the point cloud video is divided into 3D tiles, compressed into different quality representations. Secondly, the client performs quality selection based on the field of view (FoV) prediction and network bandwidth, and obtains the quality representation of the tiles in the user’s FoV. After that, the point cloud video can be rendered to watch, while the stall time and quality switch can be calculated. Finally, the point cloud video is played with a self-developed player based on Unity, which is the most widely used 3D development platform \[^{[13]}\] , and user subjective experiments are conducted. We collect the subjective MOS data, analyze the relationship between MOS and related influencing factors, and establish a QoE model of point cloud video streaming.

As far as we know, this is the first work to study the QoE model for point cloud video streaming. The contributions of this article are summarized as follows:

1) We analyze the influencing factors related to QoE of PCC video streaming, which mainly include the objective quality of the point cloud video, stall and quality switch during streaming.

2) We propose a new metric, PCPSNR, to evaluate the objective quality of the 6DoFs point cloud video, which can reflect the position information and color information of points simultaneously.

3) We conduct detailed subjective experiments on the established experimental system. By analyzing the experimental data, an accurate QoE model for point cloud video streaming is obtained.

4) We made public the subjective experimental MOS data, which may stimulate more research on this area.

The rest of this article is organized as follows. Section II introduces the point cloud video streaming system and its quality evaluation metric. Section III introduces in detail the factors affecting the QoE metric of volumetric video streaming. Section IV describes our experimental system. The experiment procedure and results are discussed in Section V and Section VI respectively. Finally, we summarize the whole article in Section VII.

II. related work

In this section, we first introduce volumetric video compression and streaming. Then we discuss objective quality metrics and subjective quality evaluation of point clouds. Finally, we introduce the related work of QoE modeling.

A. Point Cloud Video Compression and Streaming

Video is usually first compressed before the transmission. Mekuria et al. \[^{[14]}\] developed an inter-prediction algorithm for real-time compression of 3D tele-immersive video, and proposed a general point cloud codec. Another common compression method is V-PCC recommended by MPEG \[^{[9]}\] , which first converts the dynamic point cloud into two independent volumetric video sequences, and then uses an encoder to compress the geometric and textural data of the point cloud video.

After compression, a volumetric video can be transmitted and the corresponding 6DoFs scene can be rendered for the user to view. Park et al. \[^{[15]}\] proposed a transmission system for volumetric video streaming based on 3D tiles, and introduced playback buffer to reduce latency. Han et al. \[^{[16]}\] designed a streaming system called ViVo to deliver high-quality volumetric content to commercial mobile devices. The authors proposed three visibility-aware optimizations in ViVo to effectively reduce mobile resource overhead, and studied point cloud encoding, decoding, segmentation and user’s FoV in detail. Li et al. \[^{[17]}\] proposed a volumetric video streaming framework for joint communication and computational resources allocation to maximize the defined QoE based on the user’s FoV and the point cloud video features. These tiles have different quality levels according to the user’s FoV, thus saving bandwidth by paying attention to what the user is watching.

B. Objective and Subjective Quality Evaluation for Point Cloud Video

The quality of point cloud is generally expressed by objective quality metric. Mekuria et al. \[^{[18]}\] proposed...
two kinds of quality evaluation metrics by comparing original and compressed point cloud, which are obtained by calculating point-to-point distance and PSNR per YUV component in the YUV color space. Tian et al. [19] proposed a point-to-plane metric used to calculate the quality of compressed point clouds. This metric is calculated by projecting the point-to-point distance onto the normal of the original point cloud plane where the corresponding point is located. Viola et al. [20] proposed a color-based metric to represent the objective quality of point cloud by extracting color features, and then combined geometry- and color-based methods to create a rendering-independent metric to better capture point cloud distortion. Due to lack of unified metric that considers position and color information, a new objective metric that can comprehensively represent the quality of point cloud is needed.

Subjective quality evaluation of point cloud is mainly divided into static point cloud and dynamic point cloud. Zerman et al. [21] conducted a subjective and objective evaluation on the TMC2 PCC software of volumetric video. After evaluation, volumetric video quality database and subjective data are publicly available. Hooft et al. [22] studied the QoE of the point cloud in the case of 6DoFs adaptive streaming and considered changes in different network conditions and configurations. In the above studies, the user’s QoE was finally compared with objective quality metrics, which ignored the effect of network changes on the user’s QoE in volumetric video streaming. Therefore, a more specific QoE model is needed to more accurately reflect the user’s QoE.

C. QoE Modeling

For QoE modeling, there have been many studies in traditional video and VR video. Liu et al. [23] proposed a new user experience model to measure the user experience of DASH video, combining three factors that affect the user experience: initial delay, stall and level change, to form an overall user experience model. Filho et al. [24] proposed PERCEPTION, which is used to evaluate the user’s QoE in adaptive VR video streaming. Through machine learning, the playback performance of VR video is predicted, and then the playback performance is used to model and predict the user QoE. Li et al. [25] studied the QoE modeling of VR video streaming based on wireless network. After considering the tiling and user FoV changes of VR video streaming, a QoE model of panoramic video wireless streaming was obtained through the analysis of experimental data. The above researches on QoE modeling mainly focus on traditional video and VR video, while there are few researches on volumetric video streaming.

III. QoE Model for Point Cloud Video Streaming

In this section, we discuss several factors related to QoE metric for point cloud video streaming.

A. Overview

In point cloud video streaming, objective quality has the greatest impact on QoE, as it can directly reflect the video quality received by users. We propose PCPSNR to better reflect the video quality of point cloud video, by integrating the position and color information of all points. In addition, due to the huge amount of point cloud data, current volumetric video streaming systems generally compress the video first before transmission, and then the clients can decode, render and play on their devices. However, the transmission time affected by different network bandwidths, as well as decoding time affected by different computational resources and decoder may cause playback to stall, which has a negative impact on QoE [17] . Finally, due to the adaptive mechanism of streaming system, users will see video contents of different quality representations, the quality switch of video content seen by users will also impair QoE [25] . Based on the above analysis, we conclude that the factors affecting the QoE metric of volumetric video streaming are the objective quality of point cloud video represented by PCPSNR, stall and quality switch caused by streaming system, as shown in Fig. [1]

![Fig. 1. Influencing factors related to QoE metric in adaptive point cloud video streaming.](image)

B. Objective Quality

To obtain PCPSNR, we first calculate the distance quality metric PSNRp based on the point-to-plane distance, then integrate original PSNR of each YUV components into the color quality metric PSNRc, and finally combine the distance metric and color metric.

1) Position information of points: For the calculation of distance quality metric representing the position information of points, we use the point-to-plane distance [19] . As shown in Fig. [2], for each point i in the received point cloud Z’, the nearest neighbouring point in the original reference point cloud Z can be found, and then the normal vector is estimated through the r-radius neighborhood, i.e., the normal vector of the plane where the nearest neighbor is located [26] . The final point-to-plane (P2Plane) distance is obtained by projecting the point-to-point distance in the direction of the normal
vector. In this paper, we take the root mean square distance (RMSD) to calculate the P2Plane distance of the whole point cloud. Specifically, the mean distance $d_{RMSD}$ is obtained by averaging the P2Plane distance between all the points in the received point cloud $Z'$ (the number of points is $N_{Z'}$) and the nearest neighbor in the reference point cloud $Z$:

$$d_{RMSD} = \frac{1}{N_{Z'}} \sum_{i=1}^{N_{Z'}} d_{Z,Z'}^z(i)$$

where $d_{Z,Z'}^z(i)$ is the P2Plane distance between point $i$ in the received point cloud $Z'$ and the nearest neighbor in the reference point cloud $Z$.

2) Color information of points: For the calculation of color quality metric representing the color information of points, we convert the RGB value contained in each point to YUV format according to ITU-R Rec.B.709 formula [27]. As shown in Fig. 2, the YUV mean square error (MSE) of point $i$ in the received point cloud video $Z'$ and its nearest neighbor in the reference point cloud $Z$ is calculated:

$$MSE_{Y,Y'} = \frac{1}{N_{Z'}} \sum_{i=1}^{N_{Z'}} (Y_i^r - Y_i^z)^2$$

$$MSE_{U,U'} = \frac{1}{N_{Z'}} \sum_{i=1}^{N_{Z'}} (U_i^r - U_i^z)^2$$

$$MSE_{V,V'} = \frac{1}{N_{Z'}} \sum_{i=1}^{N_{Z'}} (V_i^r - V_i^z)^2$$

where $Y_i^r$, $U_i^r$, $V_i^r$, and $Y_i^z$, $U_i^z$, $V_i^z$ are the Y, U, V value of point $i$ in $Z'$, respectively. $N_{Z'}$ represents the number of points in $Z'$.

Similarly, we calculate the MSE from $Z'$ to $Z$ and $Z$ to $Z'$, and then take their maximum as the symmetric MSE to calculate the $PSNR_Y$:

$$PSNR_Y = 10 \log_{10} \left( \frac{I_{MAX}^2}{MSE_{Y}} \right)$$

where $I_{MAX}$ represents the maximum value of the color information, which is equal to 255 in our case. $PSNR_U$ and $PSNR_V$ are calculated in the same way as $PSNR_Y$. According to [28], the color quality metric of point cloud can be expressed as:

$$PSNR_C = 6 \cdot PSNR_Y + PSNR_U + PSNR_V$$

3) PCPSNR Metric: Based on position and color information of points, we can define PCPSNR metric as follows:

$$PCPSNR = PSNR_P + PSNR_C$$

It is worth noting that the proposed PCPSNR metric is also applicable to traditional 2D video. In that scenario, $PSNR_P = 0$ and PCPSNR will degenerate to $PSNR_C$, complying with quality metric definition of traditional 2D video.

C. Stall Impairment

The stall during the user’s playback is a factor negatively correlated with QoE. In traditional video streaming systems, when network conditions degrade, the required transmission time may become longer and playback will stall if the playback buffer is empty. In point cloud video streaming system, due to the large amount of data in each frame, the server usually compress the video to improve the transmission efficiency in addition to producing videos with different quality representations. But this introduces extra decoding time.
in the client side, which is significantly longer than that required in traditional video decoding. Therefore, in the point cloud video streaming system, stall impairment is mainly caused by both downloading time and decoding time.

D. Quality Switch Impairment

Similar to the stall, quality switch is another factor negatively correlated with QoE in point cloud video streaming. During playback, the client requests 3D tiles of different quality representations to adapt to time-varying network bandwidth. Therefore, users may experience a quality switch during the viewing process, which will cause QoE impairment. In addition, point cloud video has the characteristic of 6DoFs, which will bring the viewport distance effect. As illustrated in Fig. 3, since the blue tile is closer to the viewpoint than the red tile, it is obvious that the quality switch caused by the blue tiles is more perceptible to the user. This effect is called the viewport distance effect, which is caused by the different viewpoint positions of the 3D tiles. The viewport distance effect should be introduced in the measurement of quality switch impairment.

To sum up, as shown in Fig. 1, we can model QoE metric with PCPSNR, stall and quality switch of point cloud video during streaming transmission. That is $QoE = f(PCPSNR, ST, QS)$, where $f$ is the metric mapping function, $ST$ represents the stall duration, $QS$ represents the quality switch times. In the following section, we will study the detailed QoE metric based on our adaptive point cloud streaming simulation system and subjective experiments.

IV. Experiment setup

A. Experimental System

In this section, we attempt to construct an adaptive point cloud video streaming experimental system, then conduct subjective experiments to analyze the relationship between QoE and the influencing factors, and finally establish a QoE model.

A widely used streaming system is shown in Fig. 4, which consists of a client and a server. In particular, the point cloud video is segmented into 3D tiles and each tile is encoded into compressed tiles with different quality representations. The client conducts network bandwidth prediction to determine the current network environment and receives the user’s FoV information. After FoV analysis, the system can transmit only the tiles in the user’s FoV to avoid unnecessary transmission, or the system can transmit a larger area to prevent prediction errors. The client also needs to monitor the current buffer state. After combining the above three information on the client side, the quality selection process is started to determine the optimal quality representation of tiles in the user’s FoV. The results of the tile quality selection are uploaded to the server through the HTTP interface. Finally, the server sends the selected tiles to the client through the HTTP interface. The client decodes the received tiles, merges the tiles, and then renders the video for the user to view.

As shown in Fig. 3, FoV analysis can be used to enable the transmission of only the point cloud video in the user’s field of view, which can reduce overhead of computational and communication resources. In actual system, FoV information can be obtained from user’s Head-Mounted Displays (HMD) in real-time, or it can be predicted based on the user’s most recent viewpoint information. Therefore, we simulate these two FoV analysis methods in the following sections.

For a point cloud video, it is usually divided into $K$ Group of Frames (GoFs) in time dimension and C 3D tiles of the same size in space dimension. For each tile, it can be assumed there are $R$ kinds of quality representations. Define a binary variable $x_{i,j}$, where $x_{i,j} = 1$ means tiles with quality level $l$ in GoF $k$ in FoV are selected to transmit, otherwise $x_{i,j} = 0$. Here we assume that for GoF $k$, all the tiles in FoV have the same quality level during transmission.

B. Calculations of Stall Duration and Quality Switch Times

1) Stall Duration and Buffer Dynamics: For GoF $k$, denote $T_{d,k}$ as download time, $T_{c,k}$ as decoding time, $T_{p,k}$...
as playback time. Then $T w_k$ can be expressed as:

$$T w_k = \sum_{i=1}^{R} \frac{(S_{k,i} \times x_{k,i})}{B_k}$$  \hspace{1cm} (8)

where $S_{k,i}$ is the data size of GoF $k$ in user’s FoV with quality level $i$ and $B_k$ represents the network bandwidth at time of GoF $k$.

Decoding time $T e_k$ can be expressed as:

$$T e_k = \sum_{i=1}^{R} (T e_{k,i} \times x_{k,i})$$  \hspace{1cm} (9)

where $T e_{k,i}$ is the decoding time of GoF $k$ in user’s FoV with quality level $i$.

Then the total download and decoding time $T r_k$ of GoF $k$ can be defined as:

$$T r_k = T w_k + T e_k$$

$$= \sum_{i=1}^{R} \frac{(S_{k,i} \times x_{k,i})}{B_k} + \sum_{i=1}^{R} (T e_{k,i} \times x_{k,i})$$  \hspace{1cm} (10)

The playback time of GoF $k$ can be defined as:

$$T p_k = \frac{f}{f p s}$$  \hspace{1cm} (11)

where $f$ is the number of frames of GoF $k$ in user’s FoV, and $f p s$ is the frame rate, i.e., how many frames are played per second.

Assuming that the playback device only caches one GoF’s data, then the stall duration $S T_k$ can be expressed as:

$$S T_k = \max (T r_k - T p_k, 0)$$

$$= \max \left[ \sum_{i=1}^{R} \frac{(S_{k,i} \times x_{k,i})}{B_k} + \sum_{i=1}^{R} (T e_{k,i} \times x_{k,i}) - \frac{f}{f p s} \right]$$  \hspace{1cm} (12)

2) Quality Switch: We define quality switch $Q S_k$ as the absolute value of the difference between quality level of two adjacent GoFs, which can be expressed as:

$$Q S_k = \sum_{c=1}^{C_{FoV}} \left( \left| \sum_{l=1}^{R} (l \times x_{k,l}) - \sum_{l=1}^{R} (l \times x_{k-1,l}) \right| \times \frac{B S_{k,c}}{D I_{k,c}} \right)$$  \hspace{1cm} (13)

where $C_{FoV}$ is the tiles set inside user’s FoV when GoF $k$ is played. $D I_{k,c}$ represents the distance between tile $c$ and the user’s viewpoint, and $B S_{k,c}$ is the diagonal length of Bounding-Box for GoF $k$ where the tile $c$ is located. As shown in Eq. 13 the quality switch is multiplied by a weight representing the viewport distance effect, which is the ratio of $B S_{k,c}$ to $D I_{k,c}$. Obviously, the farther the tile is from the viewpoint, the smaller the quality switch will be, that is, the smaller the quality switch affects the user’s QoE, and vice versa.

C. Point cloud video processing

In point cloud video streaming, we need to solve the following multi-objective optimization problem by trying to maximize the quality level while minimizing stall duration and quality switch.

$$P 1 : \begin{cases} \max (\sum_{i=1}^{R} (l \times x_{k,i})) \\ \min (S T_k) \\ \min (Q S_k) \end{cases}$$  \hspace{1cm} (14)

During the experiment setup, we select four volumetric videos of 300 frames \cite{10} longdress, loot, soldier with lower resolution, and Basketball player with higher resolution. Each volumetric video is divided into 30 GoFs, each containing 10 frames and being divided into 12 tiles. We use MPEG PCC TMCM\cite{3} software to generate up to 5 different quality levels representations. For longdress and loot, we set four different bandwidths (70, 120, 230, 340) Mb/s. While for soldier and Basketball player, we set another four bandwidths (130, 230, 350, 480) Mb/s. We also use two user’s FoV prediction schemes, perfect and most recent. Perfect means FoV data can be obtained in real-time from user’s HMD, and most recent is another prediction method with one GoF delay of perfect FoV data.

We select two groups of subjects to use HTC vive HMD to view the original complete point cloud video. During the viewing procedure of first group, we record their FoV data, which is used for viewing of second group.

Thus we can get 32 (4 videos × 4 bandwidths × 2 prediction schemes) different video configurations, and then solve the optimization problem $P 1$ with a simple brute-force search method. With the optimization variable $x$, we can select the compression level of all tiles in user’s FoV, and then merge them into a new GoF based on the FoV information. Finally all processed point cloud videos can be obtained, and the PCPSNR, stall duration and quality switch of all videos can be calculated. The detailed results are as follows.

1) Bandwidth versus PCPSNR and Stall: As shown in Fig. 5, with the increase of network bandwidth, the PCPSNR of four videos increases continuously, which indicates that the objective quality of the videos becomes higher. Fig. 7 shows that the stall duration of four videos decreases with the increase of bandwidth. In the case of most recent FoV analysis, due to that the streaming scheme selects higher quality at a larger bandwidth, the stall time of the loot point cloud video at 340 Mb/s is slightly increased compared with 230 Mb/s.

2) Bandwidth and Quality Switch: As shown in Fig. 6, the quality switch of four videos does not monotonously change with the increase of bandwidth, which is larger in the middle two bandwidths than the two sides. This is because we hope to conduct a more comprehensive subjective experiment, that is, each GoF of the videos is assigned the lowest quality by the system when the

\footnote{https://github.com/MPEGGroup/mpeg-pcc-tmcm2}
minimum bandwidth is set, and the highest quality when the maximum bandwidth is set. So there is almost no quality switch on two sides.

Fig. 6 shows the original and generated versions of one single frame in the four point cloud videos. We use HTC vive as the HMD to play the processed point cloud videos through the compiled Unity program, and conduct user’s subjective experiments.

V. SUBJECTIVE EXPERIMENT

This section describes our subjective experiment in detail.

Experimental apparatus and environment: We use ITU-R BT.500-14\footnote{https://www.itu.int/rec/R-REC-BT.500-14-201910-I/en} as the reference standards of our subjective experiments. We design a point cloud video player based on Unity, ask subjects wear the HTC vive HMD to enjoy the generated point cloud video, as shown in the Fig. 9.

Participants: We select 34 subjects from our school to participate in the subjective evaluation experiment, including 25 males and 9 females. In order to make the subjects more comprehensive, 6 of them are experts in point cloud video, and the remaining 28 participate in the point cloud video experiment for the first-time.

Training and evaluating: As shown in Fig. 9, the subjects train and evaluate the point cloud videos during the experiment. Fig. 9 shows the scenario of training and evaluating, and Fig. 10 shows the corresponding user interfaces (UI) of point cloud video player. The subjects participating in point cloud video experiment for the first time may not be familiar with point cloud video, so they are given some time to watch them. Specifically, we train the first-time subjects by watching point cloud videos of different contents and qualities, while point cloud video experts are not trained. After a short rest, all subjects evaluate the received point cloud videos using the double-stimulus impairment scale (DSIS\footnote{https://www.itu.int/rec/R-REC-BT.500-14-201910-I/en} method.)
32 pairs of processed point cloud videos and corresponding original point cloud videos are played randomly. Subjects first view the reference original point cloud video and then view the received point cloud video.

**Scoring:** After viewing the videos, the subjects score the processed point cloud videos. The score is set from 1 to 5, divided into five grades: excellent, good, average, poor and exceptionally poor. For each generated point cloud video, we calculate the mean of all user opinion scores. The mean opinion score (MOS) represents the overall user QoE of each processed point cloud video.

### VI. Experimental results and analysis

After the subjective experiment, we can analyze the relationship between the subjective score MOS and PCPSNR, stall time and quality switch to obtain the specific QoE model. Besides we make the subjective experiment MOS public.

**A. The relationship between MOS and PCPSNR, stall and quality switch**

Fig. 11 shows the relationship between bandwidth and MOS. The MOS of the four videos all increase with the increase of the bandwidth, as higher bandwidth enables the selection of higher quality point cloud video, and at the same time leads to lower stall time, which results in better user experience. This also reflects that the subjects are more sensitive to the changes of PCPSNR and stall compared with the quality switch, because the large change in the quality switch does not affect the trend of MOS increasing.

The MOS of longdress, loot, and soldier grow slowly in the middle two bandwidths compared to the two sides. This is because the quality switch is very high in the middle bandwidth setting, resulting in lower user MOS. However, loot and longdress have a big change in MOS when the bandwidth changes from 70 Mb/s to 120 Mb/s, which is caused by a large reduction in stall duration. From this analysis, we can see that subjects are more concerned about the negative effects of stall than quality switch. The MOS and stall of Basketballplayer are relatively correlated, which is because the PCPSNR and quality switch of Basketballplayer change slightly with the increase of bandwidth.

From Fig. 11 we can observe that PCPSNR, i.e., objective quality, is positively correlated with MOS, while stall and quality switch are negatively correlated with MOS. This result corresponds to our analysis of the influencing

---

3https://github.com/Siwwv/Volumetric-video-streaming-subjective-experiment-database
factors of QoE in Section III, proving the validity of our QoE modeling analysis. We also find out that subjects are more concerned about the effects of PCPSNR and stall than quality switch.

B. Function fitting and correlation verification of MOS model

In order to obtain specific QoE metric, we take a function fit to the experimental data to calculate a mathematical model between the MOS and the three influencing factors. The specific fitting procedure is divided into three parts, as follows

1) Sample scatter diagram: We have made a sample scatter plot of the dependent variable MOS and their respective variables, PCPSNR, stall and quality switch, as shown in Fig. 12. It can be seen that these points are roughly distributed beside a straight line, and linear regression can be adopted.

2) Multiple linear regression: In order to get the final QoE model, we use 60% of the MOS data for training to get the mathematical model and the rest of data is used for validating. During the training process, we used SPSS software for data analysis, and the relevant results are shown in Table I.

Thus the QoE metric can be obtained as follows:

$$QoE = -0.471 + 0.037 \times PCPSNR$$
$$- 0.313 \times ST - 0.007 \times QS$$ \hspace{2cm} (15)

The R-square value of the model is 0.931, indicating that PCPSNR, stall, and quality switch could explain 93.1% of the variation of QoE. It can be noticed that the model passed the $F$ test ($F = 66.948, p = 0.000 < 0.05$), indicating that at least one of PCPSNR, stall and quality switch would have an influence on QoE. In addition, according to the multi-collinearity test of the model, all VIF (variance inflation factor) values in the model are less than 5, indicating that there is no collinearity
problem. Moreover, the D-W value is near 2, which indicates that the model does not have auto-correlation and there is no correlation between the sample data used for training.

The final analysis shows that the regression coefficient of PCPSNR is 0.037\(t = 4.585, p = 0.000 < 0.01\), indicating that PCPSNR has a significant positive influence on MOS. The regression coefficient of stall is \(-0.313\(t = -6.830, p = 0.000 < 0.01\), indicating that stall would have a significant negative influence on MOS. The regression coefficient of quality switch is \(-0.007\(t = -5.667, p = 0.000 < 0.01\), indicating that quality switch has a significant negative influence on MOS.

3) Model validation: In order to further verify the MOS model, we use the remaining 40% of the data to calculate the predicted MOS through the fitted MOS model, and then compare the correlation between the subjective MOS and predicted MOS, as shown in Fig. 11. It can be seen that the Goodness of fit R-square value is 0.924, indicating the fitted MOS model is correct.

In general, the results inferred by the above two statistical methods are consistent, indicating that there is a significant linear correlation between the dependent variable MOS and the independent variable, and can well predict the QoE of users for point cloud video during volumetric video streaming.

VII. Conclusion

We have studied the QoE model of volumetric video streaming, which takes into account three factors closely related to the user’s QoE. These three factors are the objective quality related to the point cloud, stall and quality switch during video streaming. To properly calculate the point cloud video quality, we combine the position and color information of points to expand the objective quality metric of the point cloud, and take the viewpoint distance effect into account to model an accurate QoE metric. With the detailed double-stimulus subjective experiments, we obtain the final QoE metric through the analysis of subjective MOS data. We find out that objective quality is positively correlated with QoE, while stall and quality switch have negatively correlation with QoE. In addition, subjects are more concerned about the effects of PCPSNR and stall than quality switch. The experiments prove that our QoE model can well predict the subjective quality experience of user in volumetric video streaming, which has a good inspiration for the subsequent related QoE research and the optimization of volumetric video streaming.

References

[1] R. Pagés, K. Ampliantitis, D. Monaghan, J. Ondřej, and A. Smolčić, “Affordable content creation for free-viewpoint video and vr/ar applications,” Journal of Visual Communication and Image Representation, vol. 53, pp. 192–201, 2018.
[2] 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.
[3] Z. Liu, Q. Li, X. Chen, C. Wu, J. Li, Y. Ji et al., “Point cloud video streaming: Challenges and solutions,” IEEE Network, 2021.
[4] J. Li, C. Zhang, Z. Liu, W. Sun, W. Hu, and Q. Li, “Demo abstract: Narwhal: a dash-based point cloud video streaming system over wireless networks,” in IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2020, pp. 1326–1327.
[5] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, “Towards 6dof http adaptive streaming through point cloud compression,” in Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 2405–2413.
[6] M. Hosseini and C. Timmerer, “Dynamic adaptive point cloud streaming,” in Proceedings of the 23rd Packet Video Workshop, 2018, pp. 25–30.
[7] L. Wang, C. Li, W. Dai, J. Zou, and H. Xiong, “Qoe-driven and tile-based adaptive streaming for point clouds,” in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 1930–1934.
[8] J. Park, P. A. Chou, and J.-N. Hwang, “Volumetric media streaming for augmented reality,” in 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018, pp. 1–6.
[9] S. Schwarz, M. Preda, V. Baroclini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuca, S. Lasserre, Z. Li et al., “Emerging mpeg standards for point cloud compression,” in IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 133–148, 2018.
[10] E. d’Eon, B. Harrison, T. Myers, and A. Chou, Philip, 8i Voxeelfed Full Bodies-A Voxelized Point Cloud Dataset, ISO/IEC JTC1/SC29 Joint WG11/WGI (MPEG/JPEG) input document WG11M40059/WG11M74006, Geneva, January 2017.
[11] E. Duminic and L. A. da Silva Cruz, “Point cloud coding solutions, subjective assessment and objective measures: A case study,” Symmetry, vol. 12, no. 12, p. 1955, 2020.
[12] S. Winkler and P. Mohandas, “The evolution of video quality measurement: From psnr to hybrid metrics,” IEEE transactions on Broadcasting, vol. 54, no. 3, pp. 660–668, 2008.
[13] L. Gao, H. Bai, G. Lee, and M. Billinghurst, “An oriented point-cloud view for mr remote collaboration,” in SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications, ser. SA ’16. New York, NY, USA: Association for Computing Machinery, 2016.
Fig. 12. Scatter plots between the MOS and PCPSNR, stall and quality switch.

Fig. 13. Correlation between subjective MOS and predicted MOS calculated by the fitting MOS model.

[14] R. Mekuria, K. Blom, and P. Cesar, “Design, implementation, and evaluation of a point cloud codec for tele-immersive video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 4, pp. 828–842, 2016.

[15] 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.

[16] B. Han, Y. Liu, and F. Qian, “Vivo: Visibility-aware mobile volumetric video streaming,” in Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, ser. MobiCom ’20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3372224.3380888

[17] 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 ICC 2020 - 2020 IEEE International Conference on Communications (ICC), 2020, pp. 1–6.

[18] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, “Evaluation criteria for pcc (point cloud compression). output document n 16332, iso,” IEC JTC1/SC29/WG11 MPEG, Tech. Rep., 2016.

[19] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, “Geometric distortion metrics for point cloud compression,” in 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017, pp. 3460–3464.

[20] I. Viola, S. Subramanyam, and P. Cesar, “A color-based objective quality metric for point cloud contents,” in 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2020, pp. 1–6.

[21] E. Zerman, P. Gao, C. Ozcinar, and A. Smolic, “Subjective and objective quality assessment for volumetric video compression,” Electronic Imaging, vol. 2019, no. 10, pp. 323–1, 2019.

[22] J. van der Hooft, M. T. Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz, “Objective and subjective qoe evaluation for adaptive point cloud streaming,” in 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX). IEEE, 2020, pp. 1–6.

[23] Y. Liu, S. Dey, F. Uluapinar, M. Luby, and Y. Mao, “Deriving and validating user experience model for dash video streaming,” IEEE Transactions on Broadcasting, vol. 61, no. 4, pp. 651–665, 2015.

[24] R. I. T. da Costa Filho, M. C. Luizelli, M. T. Vega, J. van der Hooft, S. Petrangeli, T. Wauters, F. De Turck, and L. P. Gaspar, “Predicting the performance of virtual reality video streaming in mobile networks,” in Proceedings of the 9th ACM Multimedia Systems Conference, 2018, pp. 270–283.

[25] 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). IEEE, 2018, pp. 1–7.

[26] A. Javaferi, C. Brites, F. Pereira, and J. Ascenso, “Subjective and objective quality evaluation of 3d point cloud denoising algorithms,” in 2017 IEEE International Conference on Multimedia & Expo Worksop (ICMWM). IEEE, 2017, pp. 1–6.

[27] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, “Comparison of the coding efficiency of video coding standards—including high efficiency video coding (hevc),” IEEE Transactions on circuits and systems for video technology, vol. 22, no. 12, pp. 1669–1684, 2012.