

# Bitstream-based Perceptual Quality Assessment of Compressed 3D Point Clouds

Honglei Su, *Member, IEEE*, Qi Liu, Yuxin Liu, Hui Yuan, *Senior Member, IEEE*, Huan Yang, *Member, IEEE*, Zhenkuan Pan and Zhou Wang, *Fellow, IEEE*

**Abstract**—With the increasing demand of compressing and streaming 3D point clouds under constrained bandwidth, it has become ever more important to accurately and efficiently determine the quality of compressed point clouds, so as to assess and optimize the quality-of-experience (QoE) of end users. Here we make one of the first attempts developing a bitstream-based no-reference (NR) model for perceptual quality assessment of point clouds without resorting to full decoding of the compressed data stream. Specifically, we first establish a relationship between texture complexity and the bitrate and texture quantization parameters based on an empirical rate-distortion model. We then construct a texture distortion assessment model upon texture complexity and quantization parameters. By combining this texture distortion model with a geometric distortion model derived from Trisoup geometry encoding parameters, we obtain an overall bitstream-based NR point cloud quality model named streamPCQ. Experimental results show that the proposed stream-PCQ model demonstrates highly competitive performance when compared with existing classic full-reference (FR) and reduced-reference (RR) point cloud quality assessment methods with a fraction of computational cost.

**Index Terms**—Image quality assessment, Point cloud, Bitstream-based, No reference, G-PCC.

## I. INTRODUCTION

A 3D point cloud (PC) is a collection of points representing a 3D shape, object or environment. Each point has its own geometric coordinates and other associated attributes. With the fast development of immersive media communication (IMC), there has been a growing number of 3D PC applications, including immersive telephone, smart shopping, digital museum, distance education, and smart city, among many others. Nevertheless, the gigantic data volume of 3D PCs demands heavy compression to transmit PC data under strict

bandwidth constraints, which subsequently leads to major quality control challenges. Therefore, there has been an urgent need of effective and efficient PC quality assessment (PCQA) methods [1], which may not only be used to assess and compare the performance of PC processing algorithms, but also generate valuable feedback that greatly help the design, optimization and parameter tuning of novel PC processing systems.

Human beings are the ultimate receivers in immersive media applications, thus subjective quality assessment is thought as the most reliable method of PCQA. But it is time-consuming, cumbersome, and sometimes infeasible. It is therefore crucial to establish objective PCQA models. Like image/video quality assessment problems, objective PCQA approaches can be classified into full-reference (FR), reduced-reference (RR) and no-reference (NR) methods, according to whether a pristine reference PC is fully available, partially available, or not available. NR-PCQA metrics can be further divided into media-based (MB) and bitstream-based (BB) models – the former looks at points of fully decoded PCs, while the latter is based on compressed bitstreams without full decoding. BB models provide a low complexity solution and are highly desirable in time-critical applications.

A general representation of a typical IMC transmission system is depicted in Fig. 1, where a PCQA metric may be deployed at different points for quality monitoring along the data transmission workflow. At node  $\mathcal{A}$ , the original and compressed PCs are accessible before and after the encoder, and thus FR, RR, and NR methods are all applicable, and may be used to optimize PC compression (PCC). By contrast, at nodes  $\mathcal{B}$ ,  $\mathcal{C}$  and  $\mathcal{D}$ , since the reference PC is generally not accessible, an NR-PCQA metric is better suited. More specifically, MB NR-PCQA methods may be used at point  $\mathcal{D}$  where the PC has been fully decoded, while BB NR-PCQA would be a better choice at points  $\mathcal{B}$  and  $\mathcal{C}$  for real-time quality monitoring [2], for which the decoding complexity may be too costly. There are generally two sources of quality degradations: compression distortions and transmission errors. Compression distortion assessment is fundamental in PCQA, because it can be used to assess the PC quality in absence of transmission channel errors. When channel errors are present, compression distortion assessment still serves as a critical part for the overall PCQA.

In this work, we make one of the first attempts to develop a BB NR-PCQA method, namely streamPCQ, without fully decoding compressed PC bitstreams. The diagram of the streamPCQ model is shown in Fig. 2, where encoding param-

Honglei Su and Qi Liu contributed equally to this work. This work was supported in part by Natural Sciences and Engineering Research Council of Canada, in part by the National Natural Science Foundation of China under Grants (62222110, 62172259, 61772294), in part by the Shandong Provincial Natural Science Foundation, China, under Grants (ZR2022MF275, ZR2022ZD38, ZR2021MF025, ZR2022QF076), in part by the Taishan Scholar Project of Shandong Province (tsqn202103001) and the Major Scientific and Technological Innovation Project of Shandong Province under Grant 2020CXGC010109. (Corresponding author: Huan Yang.)

Honglei Su and Qi Liu are with the College of Electronic Information, Qingdao University, Qingdao, 266071, China (e-mail: suhonglei@qdu.edu.cn, sdqi.liu@gmail.com). Yuxin Liu, Huan Yang and Zhenkuan Pan are with the College of Computer Science and Technology, Qingdao University, Qingdao, 266071, China (e-mail: liuyuxin61699@gmail.com, cathy\_huanyang@hotmail.com, zkpan@qdu.edu.cn). Hui Yuan is with the School of Control Science and Engineering, Shandong University, Ji'nan, 250061, China (e-mail: yuanhui0325@gmail.com). Zhou Wang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada (e-mail: zhou.wang@uwaterloo.ca).

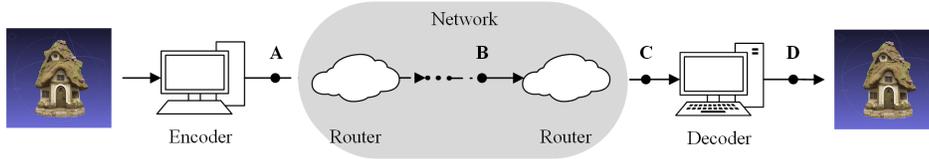


Fig. 1. Diagram of PC transmission system and potential quality monitoring points at nodes A, B, C and D.

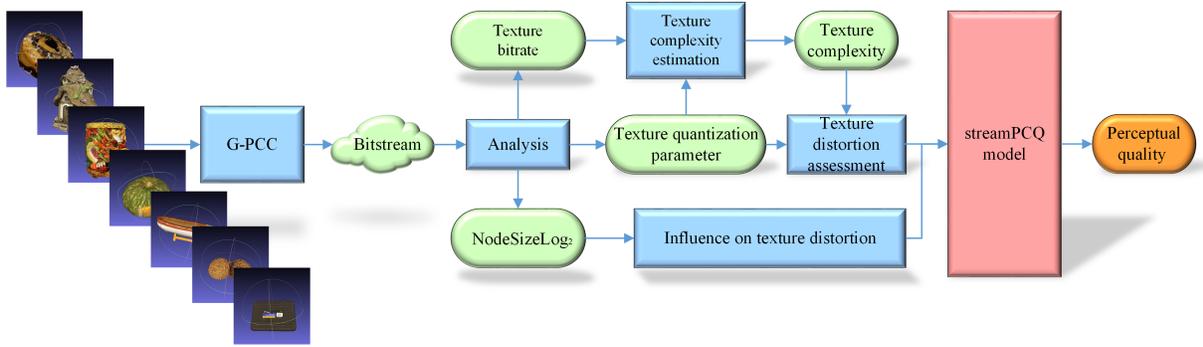


Fig. 2. Diagram of the proposed streamPCQ model.

eters, including Trisoup node size (TNS), texture quantization parameter ( $Q_p$ ), and texture bitrate ( $R_p$ ), are extracted by a bitstream analyzer, followed by texture complexity ( $\sigma_{TC}$ ) estimation, texture distortion assessment, geometric distortion modeling, before an overall streamPCQ quality prediction is reached. When tested on the Waterloo Point Cloud (WPC) database [3], the proposed streamPCQ model demonstrates state-of-the-art quality prediction performance with very low computational complexity.

The remainder of this paper is outlined as follows. In Section II, we discuss related work, including existing PCQA models, the relevant technical details of Trisoup and Region Adaptive Hierarchical Transform (RAHT) in Geometry-based PCC (G-PCC), and the standardization of bitstream-based visual quality assessment. In Section III, we establish the relationship between  $Q_p$  and the perceptual quality of geometric lossless PCs, derive a TC estimator by rate-distortion theory, and construct the streamPCQ model by incorporating the influence of texture and geometric distortions on perceptual quality. We report our experimental results in Section IV and draw conclusions in Section V.

## II. RELATED WORK

### A. PCQA Models

Objective PCQA aims to create accurate mathematical models to estimate PC quality. Just like image/video quality assessment, objective PCQA models can be categorized into FR, RR and NR PCQA models.

1) *FR-PCQA Models*: FR-PCQA models can be broadly classified into point-based, projection-based and feature-based models. More specifically, point-to-point metrics [4]–[7] assess geometric distortions by Euclidean distances, and point-to-plane methods [5] assess geometric distortions by projected errors along normal vector directions. The plane-to-plane metric predict the geometric distortions by calculating the

similarity of local surface approximations [8]. In [9], [10], point-to-distribution quality metrics are proposed by exploiting the correspondence between a point and a distribution of points from a small PC region.

Projection-based models [3], [11]–[19] use the projected image of PC information to evaluate the PC quality. Specifically, in [11], both reference and test PCs are projected onto six planes of their bounding boxes. Then, the average scores of the projected images are computed by state-of-the-art image quality metrics. In [12], projection-based objective quality assessment is extended by assigning weights to perspectives based on user interactivity data. Yang et al. [14] propose a projection-based method via perspective projection onto six planes and extracted global and local features of depth and color images obtained by projection. Wu et al. [16] propose two projection-based objective quality evaluation methods: a weighted view projection-based model and a patch projection-based model. He et al. [15] propose a PCQA method that combines colored texture and curvature projection. He et al. [17] also propose a PCQA method based on texture and geometry projection. Wang et al. [19] use the saliency maps to facilitate quality prediction.

Feature-based models extract features from both reference and distorted PCs to evaluate objective scores. Similarity measures [20] proven to be effective in general image quality assessment are extended to PCQA [21]–[28]. In these methods, geometry-based, normal-based, curvature-based color-based and graph-based features are extracted from both reference and distorted PCs, then these feature similarities are evaluated and combined to overall objective scores. Zhang et al. [29] proposed a metric based on graph signal and color features. Statistics of a variant of the Local Binary Pattern (LBP) [30], [31], perceptual color distance patterns (PCDP) [32] and local luminance patterns (LLP) [33] descriptors are introduced to PCQA. Besides, Diniz et al. [34] propose a PCQA metric, named BitDance, which uses color and geometry texture

descriptors. Alexiou et al. [35] propose a metric that compares local shape and appearance measurements between a reference and a distorted PC. Considering the visual masking effect of PC's geometric information and the color perception of human eyes, the CPC-GSCT metric [36] uses geometric segmentation and color transformation respectively to construct geometric and color features and then to predict the PC quality. In [37], color histograms and correlograms are employed to evaluate the impairment of a distorted PC with respect to its reference. Color-only and geometry-only approaches are then combined to obtain a rendering-independent objective PCQA metric. Yang et al. [38] proposed multi-scale potential energy discrepancy (MPED), a distortion quantification to measure point cloud geometry and color difference. The proposed MPED is able to capture both geometrical and color impairments by quantifying the total distortion between reference and distorted samples. It's interesting that Chetouani et al. [39], [40] and Tliba et al. [41] propose deep learning-based FR-PCQA methods that efficiently predict the quality of distorted PCs with reference.

Although these methods demonstrate promising results, they are FR models and require full decoding, which is time-consuming and is unsuitable for real-time quality monitoring at network nodes in practical PC transmission systems.

2) *RR-PCQA Models*: RR-PCQA models only require partial information of the reference PCs [42], [43]. In [42], an RR-PCQA metric is developed that extracts geometry-based, luminance-based and normal-based features from the reference PC. Such features are then transmitted alongside the content, and are employed at the receiver side (e.g. point  $\mathcal{D}$  in Fig. 1) and compared with the distorted PC. The best combination of the features is obtained through a linear optimization algorithm. In [43], two color features are proposed to estimate three content-dependent parameters, and then a RR-PCQA model is established. The reason is that the content has a masking effect on the coding distortion that is consistent with the characteristics of the human visual system. That is to say, the parameters are highly content dependent.

3) *NR-PCQA Models*: NR-PCQA models require no information of the reference PCs [40], [44]–[58]. Van et al. [57] present a cluster-based objective NR QoE assessment model for point cloud video. Tao et al. [45] propose a point cloud projection and multi-scale feature fusion network to assess the PC quality. The proposed method includes three modules, that is, joint color-geometric feature extractor, two-stage multi-scale feature fusion, and spatial pooling module. Considering the visual masking effect of PC's geometric information and the color perception of human eyes, the BQE-CVP metric [46] uses geometric feature, color feature and joint feature to develop a blind quality evaluator. Liu et al. [47] propose a deep learning-based no reference point cloud quality assessment method, namely PQA-Net. Specifically, the PQA-Net consists of a multi-view-based joint feature extraction and fusion (MVFEF) module, a distortion type identification (DTI) module, and a quality vector prediction (QVP) module. By using the distortion type labels, the DTI and the MVFEF modules are first pre-trained to initialize the network parameters, based on which the whole network is

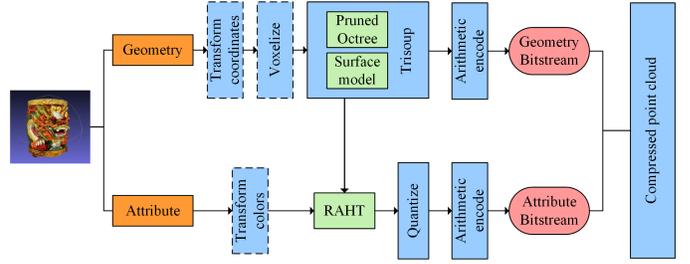


Fig. 3. Framework of G-PCC encoder in this work.

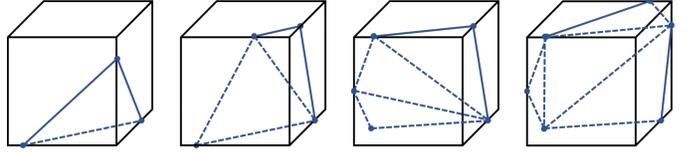


Fig. 4. Modeling geometry within each block: Trisoup.

then jointly trained to finally evaluate the point cloud quality. Zhang et al. [48] propose a metric for colored 3D models represented by both PC and mesh. Zhou et al. [49] propose an PC quality index with Structure Guided Resampling (SGR) to automatically evaluate the perceptually visual quality of 3D dense PCs. Yang et al. [50] present a metric, the image transferred point cloud quality assessment (IT-PCQA), for 3D PCs. Cao et al. [44] define PC quality as a function of the bitrate and observation distance. Nevertheless, bitrate alone cannot accurately estimate PC quality, and the observation distance, a parameter often used in the rendering algorithm, is often not available in practical systems. Liu et al. [58] propose a bitstream-layer model to evaluate V-PCC compressed PCs. Bitrate and quantization parameters are used in this model to evaluate PC quality without fully decoding compressed PCs. This gives us inspiration to develop the model in this paper.

## B. MPEG G-PCC

In 2017, MPEG initiated PCC standardization [59], [60]. Subsequently, two technologies were chosen as test models: G-PCC for static content and dynamically capturing, and Video-based PC Compression (V-PCC) for dynamic content, respectively. In G-PCC, there are two geometry encoding modes (Octree and Trisoup) and three attribute encoding modes (Predicting, Lifting and RAHT), respectively. Here we focus on static PCs and therefore we choose Trisoup with RAHT for G-PCC encoder, as shown in Fig. 3.

1) *Trisoup geometry encoding*: As illustrated in Fig. 4, Trisoup encoding is a geometry coding option that represents the object surface as a series of triangle mesh. TNS defines the size of the triangle nodes in unit of voxel. The octree encoding and decoding stop at leaf level  $L$ , in which case the leaf nodes of the octree represent cubes of width  $W = 2^{\max\_NodeSize \log_2 - L}$ , or blocks, where the octree is pruned.

If TNS is larger than 0, then the blocks are  $2 \times 2 \times 2$  or larger, and it is necessary to represent the collection of voxels within the block by some model. Geometry is represented within each

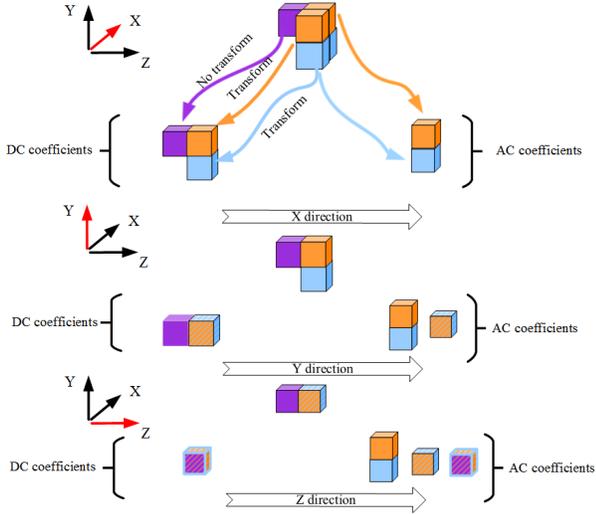


Fig. 5. Diagram of RAHT attribute encoding.

block as a surface that intersects each edge of the block at most once. Since there are 12 edges of a block, there can be at most 12 such intersections within a block. Each intersection creates a vertex. A vertex along an edge is detected if and only if there is at least one occupied voxel adjacent to the edge among all blocks that share the edge. The position of a detected vertex along an edge is the average position along the edge of all such voxels adjacent to the edge among all blocks that share the edge. This triangle mesh is encoded and used to obtain the positions of the reconstructed (or decoded) points. When it comes to decoding, the decoder generates PC from the mesh surface in the specified voxel granularity so that it assures the density of the reconstructed PC.

2) *RAHT*: RAHT is an attribute encoding option based on the geometric position of a known PC. The main idea behind RAHT is to use the attribute values in a lower octree level to predict the values in the upper level. After sorting nodes by Morton code [61], the lowest bits of Morton codes are deleted and then RAHT is applied to nodes with the same Morton code. It starts from the leaves of the octree (highest level) and proceeds backwards until it reaches its root (lowest level). At each decomposition, instead of grouping eight voxels at a time, the transform is applied to each node and is performed in three steps, one in each  $x$ ,  $y$ ,  $z$ , directions, as illustrated in Fig. 5. A brief description is as below and interested readers may refer to [62] for more details.

Attribute values of nodes are processed along  $x$  direction in the first step as shown in Fig. 5. Since the purple block has no adjacent nodes in the  $x$  direction, its attribute value, the DC coefficient, will be directly assigned to the next step. Both orange and blue nodes have adjacent nodes in the  $x$  direction, and at every grouping of two voxels with their weights, the following transform is applied:

$$\begin{bmatrix} C_{DC} \\ C_{AC} \end{bmatrix} = T_{w_1, w_2} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad (1)$$

where  $V_1, V_2$  are values of two adjacent voxels, and

$$T_{w_1, w_2} = \frac{1}{\sqrt{w_1 + w_2}} \begin{bmatrix} \sqrt{w_1} & \sqrt{w_2} \\ -\sqrt{w_2} & \sqrt{w_1} \end{bmatrix}, \quad (2)$$

where  $w_1, w_2$  are the weights of  $V_1$  and  $V_2$ , respectively. The transform matrix changes at all times, adapting to the weight values. As a result,  $C_{DC}$  and  $C_{AC}$  are assigned to two new orange nodes as the DC and AC coefficients, respectively. The same transform is also applied to two blue nodes. Then, the orange and blue nodes are adjacent along  $y$  direction, and their values are transformed into a new DC and AC coefficient, respectively. Finally, the last two DC coefficients are adjacent along  $z$  direction, thus they are sent to the same transform again. Consequently, the last AC coefficient and the final DC coefficient are obtained. Eventually, a series of AC coefficients and the final DC coefficient are obtained.

### C. Bitstream-based Visual Quality Assessment

Bitstream-based visual quality assessment is designed to use the information extracted from packet headers or payload for real-time and non-intrusive quality monitoring. It is most appropriate and sometimes the only choice for networked video services, when complete video decoding or reconstruction is not preferred. In fact, bitstream-based approaches have a special advantage over pure pixel-based schemes since additional information can be obtained from the bitstream, such as the bitrate, frame rate, frame type, quantization parameter, motion vectors, and detailed information about data corruption rising from packet loss [2].

For bitstream-based techniques, the performance significantly depends on the level of access to the bitstream. In general, the more information can be utilized, the better the performance will be. However, extracting more information from the bitstream is usually accompanied by increasing computational complexity. Three broad categories of models are identifiable in the state-of-the-art techniques and standards recommendations for bitstream-based quality assessment for networked video, depending on the level of access to the bitstream.

Specifically, a parametric model predicts perceptual video quality based on the general parameters obtained through statistical analysis of the bitstream, such as bitrate, frame rate, packet loss rate, and so on. Originally designed for service planning, the parametric planning model has been densely deployed in industry for quality monitoring of individual services. ITU-T G.1070 model [63], ITU-T G.1071 model [64] and Mode 0 of ITU-T P.1203 and P.1204 model belong to parametric models.

A packet layer model allows the use of packet header information to predict the video quality without resorting to any media-related payload information. Similar to the case of the parametric model, the packet layer model initially aims at estimating the average service quality. Along with its development, however, much expectation has been raised to monitor the QoE of individual video streams as the packet header information provides some insight into the content

characteristics. ITU-T P.1201 model [65], Mode 1 of ITU-T P.1203 [66] and P.1204 [67] model belong to packet layer models.

A bitstream layer model utilizes further information about the encoded bitstream (e.g., frame type, quantization parameter, motion vector, temporal complexity, spatial complexity, pixel values, the actual position of packet loss, the number of lost packet, etc.), in addition to the packet layer information. This model is dedicated to in-service non-intrusive monitoring, with more weight placed on the accuracy issue. The bitstream layer model can be further divided into two modes: parsing mode and full decoding mode. The former does not completely decode the payload. Any kind of analysis of the bitstream, without using the pixel information, can be applied. Mode 1 of ITU-T P.1202 model [68] and Mode 2 of ITU-T P.1203 and P.1204 model belong to the parsing mode. While the model in the latter mode can decode parts or all of the video sequence, and the pixel information can be used for quality estimation. Apparently, when full decoding is employed, this mode is similar to the hybrid model, which needs the input of the bitstream and the reconstructed video. Mode 2 of ITU-T P.1202 model and Mode 3 of ITU-T P.1203 and P.1204 model belong to the full decoding mode.

In a word, the bitstream-based model initially aims at quality assessment of networked video services. Along with its development, however, much expectation has been raised to monitor the quality of immersive media services (e.g., PC). It is clear that bitstream-based PCQA should follow the methodology of bitstream-based visual quality assessment, however, it presents new characteristics for the new media format. First, there is no geometric information in the video signal, however, geometric distortion has a great impact on the PC quality. Second, the PC is sparse in 3D space, while the video frame is regular in 2D space. Third, the subjective test of PC and video is very different, which will lead to the difference in objective algorithm design. The detailed design of proposed bitstream-based PCQA model is shown in Section III.

### III. PROPOSED STREAMPCQ MODEL

#### A. Perceptual Quality and Texture Quantization Parameter

We first establish a perceptual quality assessment model of geometric lossless PCs, called texture distortion assessment model. Similar to conventional image and video compression, blocking artifacts are often the main encoding distortion in reconstructed PCs due to RAHT. The blocking artifact here appears as virtual boundary discontinuities of virtual surfaces in a PC. It is caused by the independent quantization of RAHT coefficients across adjacent virtual surfaces. Texture quantization is the main source of texture distortion and therefore perceptual texture distortion is closely tied to the texture quantization step or its corresponding  $Q_p$ . To investigate the underlying relationship, 10 standardized PCs with both geometric and texture diversities (*Biscuits*, *House*, *Litchi*, *Mushroom*, *Pen\_container*, *Pineapple*, *Puer\_tea*, *Pumpkin*, *Ship*, *Toolbox*) from the WPC database were selected as the Training Set and encoded with constant  $Q_p$ s. The relationship between the mean opinion score (MOS) and  $Q_p$  is shown in Fig. 6 and can be

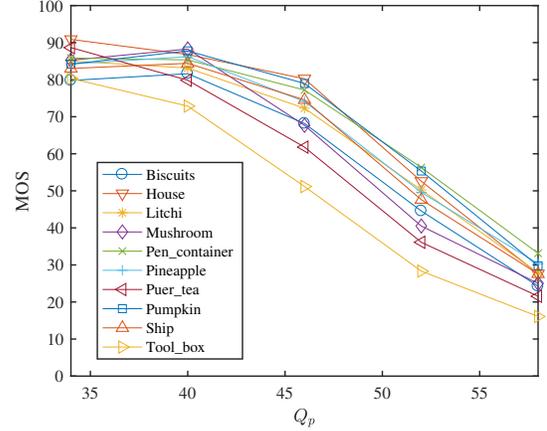


Fig. 6. Relationship between MOS and  $Q_p$ .



(a) *Pen\_container*, MOS = 56.30 (b) *Tool\_box*, MOS = 29.32

Fig. 7. Texture masking effects at  $Q_p = 52$ .

roughly depicted as a sigmoid curve for any specific PC, but the shapes vary for different PCs. Specifically, as shown in Fig. 7, the PCs richer in texture, such as *Pen\_container*, have a comparatively higher MOS at certain  $Q_p$  levels, while the PCs with less texture information such as *Tool\_box* have lower MOS. This can be explained by the texture masking effects of the Human Visual System (HVS), which suggests HVS characteristics and content-dependency, in particular  $\sigma_{TC}$ , play important roles in PCQA.  $\sigma_{TC}$  evaluation is the main focus of Section III-B.

#### B. Texture Complexity Estimation

When the attribute values of the original reference PC is available, a direct measurement of  $\sigma_{TC}$  is the averaged standard deviation of the attribute values in local blocks [69]. However, in NR-PCQA, the original attribute values are not available. To overcome the problem, we estimate  $\sigma_{TC}$  using information extracted from bitstreams of compressed PCs, such as the texture bitrate  $R_p$  and the  $Q_p$ . We derive a rate-distortion (RD) model to represent the relationship between  $\sigma_{TC}$ ,  $R_p$ , and  $Q_p$ . Assuming the PC signal has been decomposed into sub-bands, and the RAHT coefficients  $X$  within a given sub-band are independent and identically distributed and have a Laplacian distribution with parameter  $b$  [62]:

$$p(x) = \frac{1}{2b} e^{-\frac{|x|}{b}}, \quad (3)$$

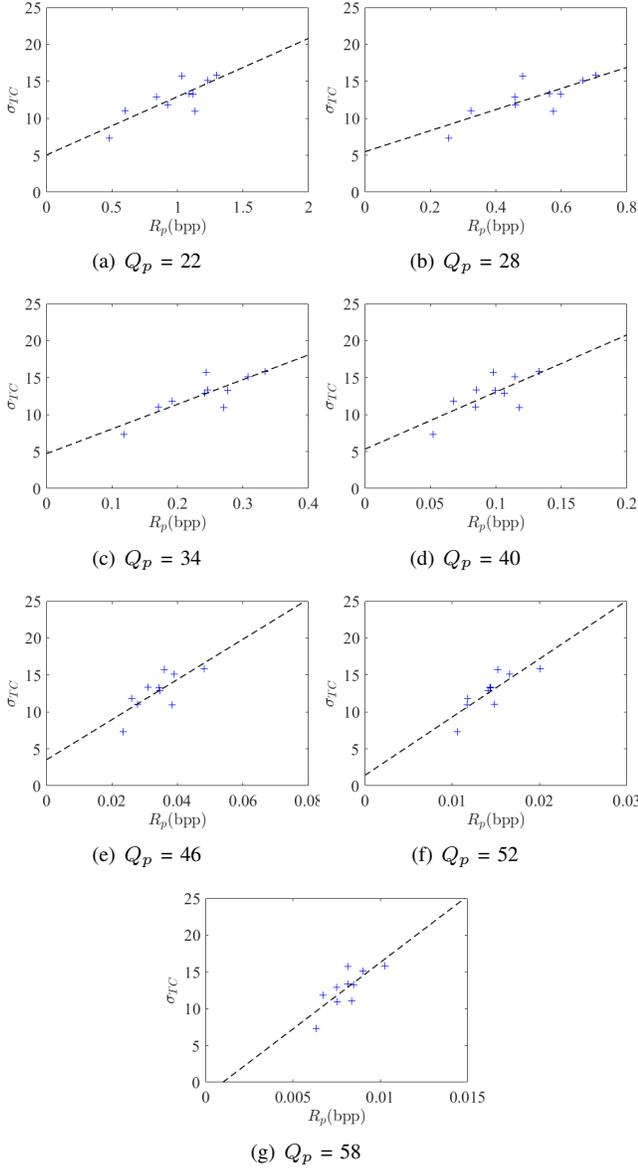


Fig. 8. Relation between  $\sigma_{TC}$  and  $R_p$ .

where the variance may be computed by  $\sigma^2 = 2b^2$ . The RD function of a Laplacian source in terms of the magnitude error criterion is given by [70]:

$$R(D) = \begin{cases} \ln(\frac{b}{D}) & 0 < D < b \\ 0 & D \geq b \end{cases}. \quad (4)$$

A Taylor series expansion of (4) leads to:

$$R(D) = \ln\lambda + \lambda^{-1}(\frac{b}{D} - \lambda) + R'(D) \approx (\ln\lambda - 1) + \frac{b}{\lambda}D^{-1}, \quad (5)$$

where  $\lambda > 1$  satisfies the condition for convergence. Since the magnitude error criterion is used as the distortion measure in (5),  $D$  can be expressed as [71]

$$D = Q_s/4, \quad (6)$$

where  $Q_s$  is the texture quantization step, and

$$Q_s = 2^{\frac{Q_p - 4}{6}}. \quad (7)$$

Substituting (6) into (5) and neglecting the higher-order terms, a rate-quantization model is formulated:

$$R_m = \alpha + \beta b_m Q_s^{-1} \quad \text{for } m = 1, 2, \dots, M, \quad (8)$$

where  $R_m$  is the average number of bits for the RAHT coefficients in the  $m$ th sub-band,  $b_m$  is the parameter of Laplace distribution in the  $m$ th sub-band,  $\alpha$  and  $\beta$  are the corresponding coefficients, and  $M$  is the number of sub-band. For a Laplacian distributed source, the standard deviation  $\sigma = \sqrt{2}b$ . Accordingly, (8) becomes

$$R_m = \gamma + \delta \sigma_m Q_s^{-1} \quad \text{for } m = 1, 2, \dots, M, \quad (9)$$

where  $\gamma$  and  $\delta$  are the corresponding coefficients and  $\sigma_m^2$  is the variance of the Laplacian distribution for the RAHT coefficient in the  $m$ th sub-band. In practical applications, given a RAHT coefficient, it is very difficult to determine the standard deviation. Nevertheless, previous research suggests that the distribution of the RAHT coefficients can also be modeled by the Laplacian distribution [62], and thus the variance of the RAHT coefficients can be estimated by:

$$\sigma_m^2 = \theta_m \sigma_{TC}^2, \quad (10)$$

where  $\theta_m$  is a parameter related to the  $m$ th sub-band and  $\sigma_{TC}$  is the standard deviation of the attribute values before the RAHT transform. Therefore, the texture bitrate (bits per point, bpp),  $R_p$ , in a transform block is given by

$$\begin{aligned} R_p &= \frac{1}{M} \sum_{m=1}^M R_m \\ &= \frac{1}{M} \sum_{m=1}^M (\gamma + \delta \sigma_m Q_s^{-1}) \\ &= \eta + \omega \sigma_{TC} Q_s^{-1}, \end{aligned} \quad (11)$$

where  $\eta$  and  $\omega$  are the corresponding parameters. Therefore given a constant  $Q_s$ ,  $R_p$  and  $\sigma_{TC}$  have approximately a linear relationship:

$$\sigma_{TC} = s(Q_s)R_p + i(Q_s), \quad (12)$$

where  $s(Q_s)$  and  $i(Q_s)$  are the slope and intercept values at  $Q_s$ , respectively. Seven examples of the relationship between  $\sigma_{TC}$  and  $R_p$  are shown in Fig. 8 for  $Q_p = 22, 28, 34, 40, 46, 52$  and  $58$  ( $Q_s = 8, 16, 32, 64, 128, 256, 512$ ), respectively. Markers in each sub-figure of Fig. 8 correspond to 10 original point clouds in the Training Set, and the abscissa and ordinate represent  $R_p$  and  $\sigma_{TC}$ , respectively. It is clear that a linear relationship exists between  $\sigma_{TC}$  and  $R_p$  for a given constant  $Q_p$ , which is consistent with the model presented in (12).

From Figs. 9 and 10, we observe that  $s(Q_s)$  is roughly a proportional to  $Q_s$ , and there is a V-curve relationship between  $i(Q_s)$  and  $Q_s$ , which has a minimum value when  $Q_s$  is around 32. These can be modeled by

$$s(Q_s) = c_1 Q_s \quad (13)$$

and

$$i(Q_s) = d_1 \ln(d_2 |Q_s - d_3|) + d_4, \quad (14)$$

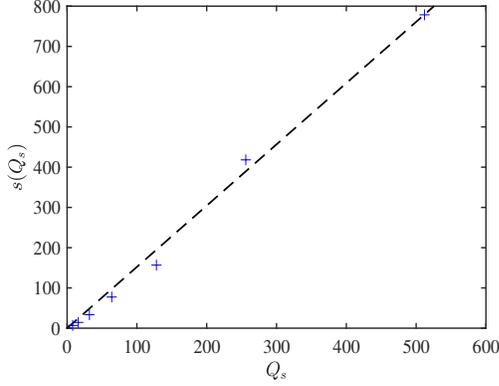


Fig. 9. Relationship between  $Q_s$  and  $s(Q_s)$ .

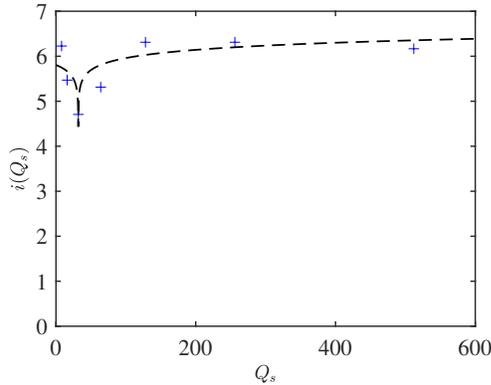


Fig. 10. Relationship between  $Q_s$  and  $i(Q_s)$ .

where  $c_1, d_1, d_2, d_3$  and  $d_4$  are parameters obtained by least square fitting. Submitting (13) and (14) into (12), we obtain our TC estimator of a distorted PC as

$$\sigma_{TC} = (c_1 Q_s) R_p + d_1 \ln(d_2 |Q_s - d_3|) + d_4. \quad (15)$$

### C. Texture Distortion Assessment

Fig. 6 exhibits a tripartite relationship between MOS and  $Q_p$ : The value of MOS decreases slowly at low- $Q_p$  or high quality range, followed by a sharp transition at mid- $Q_p$  range, and then saturates at the high- $Q_p$  or low quality region. This trend may be captured by a logistic function given by

$$MOS_T = \frac{f_1}{1 + e^{\frac{Q_p - f_2}{f_3}}} + f_4, \quad (16)$$

where  $f_1, f_2, f_3$  and  $f_4$  are empirical parameters, and  $MOS_T$  represents the MOS of PC of texture distortion only. The perceptual quality of a distorted PC and subsequently its parameters in (16) also depends on the content of the PC characterized by its  $\sigma_{TC}$ .

Table I presents the best fitting values of  $f_1, f_2, f_3$  and  $f_4$  calculated for each PC in the Training Set. There are several useful observations from Fig. 6 and Table I. In particular,  $f_1 + f_4$  for each PC corresponds to its maximum subjective score, and  $f_1$  stays relatively constant across different PCs. This is consistent with the subjective test results: when  $Q_p$

TABLE I  
OPTIMAL FITTING VALUES OF  $f_1, f_2, f_3$  AND  $f_4$  FOR EACH PC.

PC	$f_1$	$f_2$	$f_3$	$f_4$
<i>Biscuits</i>	64.66	50.85	3.436	17.22
<i>House</i>	71.36	51.70	3.218	18.74
<i>Litchi</i>	71.66	51.93	4.102	14.61
<i>Mushroom</i>	66.01	48.79	3.086	22.29
<i>Pen_container</i>	64.70	52.43	3.570	21.94
<i>Pineapple</i>	61.06	50.71	3.140	24.99
<i>Puer_tea</i>	77.80	48.00	4.487	13.84
<i>Pumpkin</i>	63.60	52.15	2.844	22.60
<i>Ship</i>	61.11	50.84	2.817	23.16
<i>Tool_box</i>	71.86	46.92	4.112	11.71

TABLE II  
 $\sigma_{TC}$  AND RETRAINED  $f_2$  AND  $f_4$  VALUES FOR EACH PC.

PC	$\sigma_{TC}$	$f_2$	$f_4$
<i>Biscuits</i>	10.99	51.16	14.79
<i>House</i>	15.76	51.24	22.55
<i>Litchi</i>	11.82	51.61	17.86
<i>Mushroom</i>	13.25	48.89	21.34
<i>Pen_container</i>	15.83	52.82	19.18
<i>Pineapple</i>	15.13	51.35	19.59
<i>Puer_tea</i>	12.92	47.64	20.32
<i>Pumpkin</i>	13.37	52.44	19.61
<i>Ship</i>	11.00	51.34	18.14
<i>Tool_box</i>	7.273	46.81	14.28

goes to its maximum value, MOS converges to  $f_4$ , the lower bound of MOS of the PC content. On the other extreme, if  $Q_p$  approaches its minimum value, MOS converges to  $f_1 + f_4$ . Although the best fitting of  $f_3$  change for different PCs, the resulting MOSs are insensitive to  $f_3$ . Therefore, we set both  $f_1$  and  $f_3$  constants in our model using the best fitting values for the full training set. Once  $f_1$  and  $f_3$  are fixed,  $f_2$  and  $f_4$  in (16) are retrained and the results are shown in Table II. Also shown in Table II are the  $\sigma_{TC}$  values of the corresponding PCs. It can be observed that a higher  $f_2$  or  $f_4$  corresponds generally to a PC of higher  $\sigma_{TC}$  (e.g., *Pen\_container*) and vice versa (e.g., *Tool\_box*). Therefore, we adapt  $f_2$  and  $f_4$  to content empirically:

$$f_2 = h_1 \sigma_{TC} + h_2 \quad (17)$$

and

$$f_4 = j_1 \sigma_{TC} + j_2, \quad (18)$$

where  $h_1, h_2, j_1$  and  $j_2$  are obtained through training using the least square fitting of the Training Set.

Finally, submitting (17) and (18) into (16), we have

$$MOS_T = \frac{f_1}{1 + e^{\frac{Q_p - h_1 \sigma_{TC} - h_2}{f_3}}} + j_1 \sigma_{TC} + j_2. \quad (19)$$

Such a model may be applied directly for quality prediction of PCs free of geometric distortion, and may also be used as a key component for quality assessment of PCs with both texture and geometric distortions.

### D. Geometric Distortion and Overall Quality Assessment

As described in Section II-B1, Trisoup encoding is a geometry coding mechanism that represents the object surface

as a series of triangle mesh, for which the TNS represented by the  $NodeSizeLog_2$  (NSL) parameter defines the size of the triangle nodes, and thus the accuracy of geometry representations. Fig. 11 shows the relationship between MOS and  $Q_p$  at different NSL levels for different PC content. Clearly, across all content, MOS generally reduces with  $Q_p$  for fixed NSL, and with NSL for fixed  $Q_p$ . The question is whether the impact of  $Q_p$  and NSL, which determines the texture and geometric distortions, respectively, are related to each other. To investigate the independence between the two factors, we define a normalized MOS (NMOS) as

$$NMOS(NSL, Q_p) = \frac{MOS(NSL, Q_p)}{MOS(NSL, Q_{p,min})}. \quad (20)$$

We then plot NMOS versus  $Q_p$  at different NSL levels in Fig. 12. Compared with Fig. 11, the curves are much more overlapped with each other, suggesting significant independence between  $Q_p$  and NSL. This also implies that the impact of texture and geometric distortions are roughly separable factors.

To model MOS against Trisoup block size, we use the following logistic function to predict the geometric distortion from the NSL parameter:

$$D_G(NSL) = \frac{l_1}{1 + e^{\frac{-NSL+l_2}{l_3}}} + l_4, \quad (21)$$

where  $l_1$ ,  $l_2$ ,  $l_3$  and  $l_4$  are empirical parameters, and  $D_G(NSL)$  represents the decay degree at NSL.

As described earlier, the impact of texture distortion and geometric distortion on MOS is roughly separable. Therefore, we propose the overall G-PCC compressed bitstream-based perceptual PC quality model, namely streamPCQ, as:

$$MOS_{est} = MOS_T(Q_p, R_p) \cdot D_G(NSL). \quad (22)$$

#### IV. EXPERIMENTAL RESULTS AND DISCUSSION

To validate the proposed streamPCQ model, we select 10 standardized PCs in the WPC database [3], [72] (*Bag, Banana, Cake, Cauliflower, Flowerpot, Glasses\_case, Honeydew\_melon, Statue, Ping-pong\_bat, Stone*) as the Validation Set. The PCs in the Validation Set cover a wide range of both geometric and textual complexities. They are employed for performance evaluation only, and has no overlap with the Training Set (which were used to determined the empirical parameters of the streamPCQ model, as described in Section III). For each source PC, four  $Q_p$  and three NSL values are used for encoding at  $\{40, 46, 52, 58\}$  and  $\{2, 4, 6\}$ , respectively. The remaining coding configuration parameters are set as default [13]. Finally, 120 distorted PCs are generated and employed for performance evaluation.

Operational parameters given in Table III are determined based on least square fitting with the Training Set following the descriptions in Section III-C. To be specific, the process of obtaining these parameters is divided into three steps. Firstly, the values of  $c_1$ , and  $d_1, d_2, d_3, d_4$  were obtained according to (13) and (14), respectively. Secondly, the values of  $f_1, f_3, h_1, h_2, j_1$ , and  $j_2$  were obtained according to (16), (17) and (18). Finally, the values of  $l_1, l_2, l_3$  and  $l_4$  were obtained

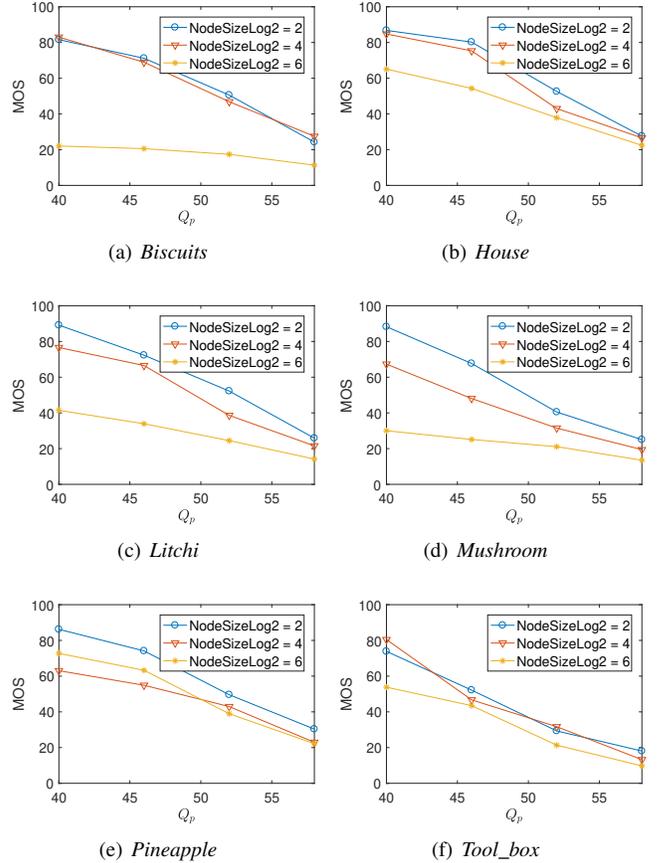


Fig. 11. MOS vs.  $Q_p$  at different  $NodeSizeLog_2$  (NSL).

TABLE III  
MODEL PARAMETERS OBTAINED USING THE TRAINING SET.

$c_1$	$d_1$	$d_2$	$d_3$	$d_4$	$f_1$	$f_3$	$h_1$
1.522	0.202	1.150	31.86	5.079	67.38	3.481	0.450
$h_2$	$j_1$	$j_2$	$l_1$	$l_2$	$l_3$	$l_4$	
44.80	0.811	8.438	0.461	4.491	-0.47	0.541	

according to (21). Once determined, these parameters are fixed for all the remaining experimental results reported in this paper.

To verify the effectiveness of (15) and (19), we carried out the separate validation on the Validation Set as shown in Table IV. The results suggest that both  $\sigma_{TC}$  and  $MOS_T$  can be correctly predicted.

To understand the contributions of the texture distortion and geometric distortion models in the proposed approach, an ablation test was performed, where the same training and validation procedures as in streamPCQ were used to create the texture distortion only  $MOS_T(Q_p, R_p)$  model and the geometric distortion only  $D_G(NSL)$  model. The test results are shown in Table V, which suggest that both models play important roles and make significant contributions to the overall streamPCQ model.

We compare the proposed streamPCQ model with  $PCM_{RR}$  [42], PointSSIM [21], PCQM [22], GraphSIM [26] and MPEG PSNR $_{\gamma}$  [73] models, as shown in Table VI. These

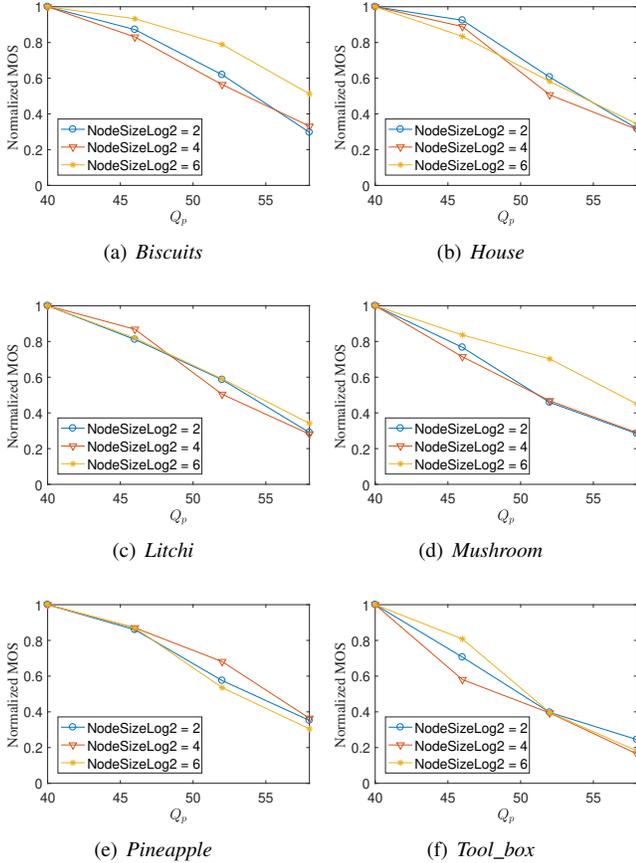


Fig. 12. Normalized MOS vs.  $Q_p$  at different  $NodeSizeLog_2$  (NSL).

TABLE IV  
SEPARATE VALIDATION RESULTS

Equation	PLCC	SRCC	RMSE
(15)	0.8344	0.7243	3.1395
(19)	0.9054	0.8692	11.00

TABLE V  
ABLATION TEST RESULTS

PCQA model	PLCC	SRCC	RMSE
$MOS_T(Q_p, R_p)$	0.6872	0.6879	18.26
$D_G(NSL)$	0.4346	0.4269	22.24
streamPCQ	<b>0.8062</b>	<b>0.8116</b>	<b>15.16</b>

models are chosen for three reasons. First, geometry distortion metrics, such as point-to-point, point-to-plane, angle similarity, PC-MSDM, etc. [4]–[9], [23], [73], are not suitable for assessing colored PCs. Second, projection-based metrics [3], [11]–[19] are governed by the selection of viewpoints (i.e., camera position and distance), and the rendering mechanism to consume the content [13], [37], [74], [75], where the rendering parameters, such as splat size and orientation, distance from the camera, and viewpoint, are unknown. Third, algorithms not publicly available are not included [24], [30], [31], [36], [37], [43]. Among the competing models, PointSSIM [21], PCQM [22], GraphSIM [26] and MPEG PSNR<sub>Y</sub> [73] are FR-

PCQA models, PCM<sub>RR</sub> [42] is a RR-PCQA model, and the proposed streamPCQ is the only NR-PCQA model. Results of the NR-PCQA model in [44] are not reported here as it is oversimplified and does not produce adequate quality prediction performance in our experiment.

Table VI reports the performance evaluation results in terms of Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SRCC), and Root Mean Squared Error (RMSE) on the Validation Set. Fig. 13 shows the scatter plots of objective score vs. MOS for all testing PCQA models on the WPC database, along with the best fitting logistic functions. From Table VI and Fig. 13, we can see that the proposed model delivers the best performance in most test cases of individual PC content, and is on par with the PCQM [22] model in the case of overall assessment. PCQM is used directly here without retraining, so its performance may be improved if we retrain it on the database in this paper. However, it needs to be emphasized that streamPCQ is the only NR model under comparison and PCQM [22] model is an FR model.

To test the generalizability of the streamPCQ model, a cross-database validation was performed on the M-PCCD database [13]. Because the M-PCCD database is rather small to train a model, we use it for testing purpose only. Specifically, without adapting the parameters of streamPCQ, we test it directly on the M-PCCD database, and the results are shown in Table VII, which also compares the performance of other PCQA models. Note that all other models are FR or RR methods, and streamPCQ is the only NR model and does not require full decoding of the compressed bitstreams. It is worth mentioning that many source PCs in the M-PCCD database are used for training the other models under comparison (except for PSNR<sub>Y</sub>), and such content is completely unforeseen to streamPCQ. It should also be noted that the performance drops of PCM<sub>RR</sub>, PointSSIM and GraphSIM from the M-PCCD to the WPC (Table VI) databases are quite significant (implying overfitting). In comparison, PCQM (an FR model) and streamPCQ produce relatively consistent performances across the two databases. Therefore, these results suggest good generalizability of the proposed streamPCQ method in comparison with existing methods in the literature.

To verify the robustness of the proposed streamPCQ model, we calculated the PLCC, SRCC and RMSE between MOS and objective score for 1000 random splits of training/validation sets. The results, shown in Fig. 14, confirmed the accuracy of the proposed model.

It is worth deep investigations of the outlier cases of the proposed streamPCQ model. In particular, in Fig. 13 (f), the outliers in the upper left side of the diagonal line often correspond to PCs with smooth surfaces, as exemplified by Fig. 15 (c), or with excessive brightness such as Fig. 15 (f). On the other hand, the outliers in the bottom right side are typically the PCs with a thinner side where the thickness is smaller than one TNS, as exemplified by Figs. 15 (a) and (b), or with holes, such as Figs. 15 (d) and (e). These observations suggest interesting directions for future improvement.

To ascertain that the improvement of the proposed model is statistically significant, we carried out a statistical significance

TABLE VI  
PERFORMANCE COMPARISON OF PCQA MODELS

Content	PCMRR [42]			PointSSIM [21]			PCQM [22]			GraphSIM [26]			PSNR <sub>Y</sub> [73]			streamPCQ		
	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC	PLCC	RMSE	SRCC
<i>Bag</i>	0.8248	16.83	0.7972	0.8482	<b>11.97</b>	0.8601	0.8299	13.77	0.8392	0.8045	13.21	0.7972	0.6579	25.15	<b>0.8811</b>	<b>0.8755</b>	14.07	0.8671
<i>Banana</i>	0.7683	19.06	0.9231	0.6971	22.82	0.6154	0.6982	14.82	0.7413	0.8714	17.90	0.8671	0.9475	<b>8.319</b>	<b>0.9582</b>	<b>0.9536</b>	17.27	0.9371
<i>Cake</i>	0.0382	24.25	0.1748	0.8848	<b>19.72</b>	0.8741	0.7823	22.76	0.8811	0.7657	24.60	0.7273	0.7522	20.59	0.8811	<b>0.8854</b>	20.64	<b>0.9231</b>
<i>Cauliflower</i>	0.2185	21.66	0.1888	0.8549	18.56	0.8252	0.9508	<b>10.22</b>	0.9231	0.8862	17.07	0.7832	0.9514	12.61	0.9513	<b>0.9813</b>	11.64	<b>0.9653</b>
<i>Flowerpot</i>	0.7589	32.54	0.7483	0.2659	30.13	0.1259	0.9598	<b>13.95</b>	0.8811	0.9551	17.20	<b>0.9441</b>	<b>0.9722</b>	21.63	0.9091	0.8147	17.59	0.6154
<i>Glasses_case</i>	0.6592	17.69	0.6993	<b>0.8573</b>	<b>12.27</b>	0.7692	0.7748	15.07	<b>0.8322</b>	0.6388	15.87	0.6503	0.5398	23.71	<b>0.8322</b>	0.6435	19.13	0.6224
<i>Honeydew_melon</i>	0.7539	15.78	0.8266	0.8139	16.63	0.8112	0.4198	24.03	0.3287	0.8678	15.83	0.9587	0.7892	19.46	0.9654	<b>0.9367</b>	<b>8.483</b>	<b>0.9791</b>
<i>Ping-pong_bat</i>	0.8362	17.32	0.9650	0.5652	21.56	0.6154	<b>0.9457</b>	<b>13.43</b>	<b>0.9790</b>	0.0029	34.75	0.1338	0.9119	19.99	0.9721	0.8352	16.58	0.8182
<i>Statue</i>	0.4332	27.66	0.2028	0.4627	29.30	0.3427	<b>0.9351</b>	<b>9.469</b>	<b>0.9301</b>	0.2531	28.56	0.0560	0.8035	17.88	0.8811	0.8847	12.30	0.8252
<i>Stone</i>	0.2665	20.44	0.2238	0.9006	8.608	0.8322	0.8442	12.13	0.9513	0.9040	10.11	0.9578	0.7052	21.31	0.9512	<b>0.9579</b>	<b>8.101</b>	<b>0.9720</b>
Overall	0.4448	21.93	0.2403	0.5575	20.33	0.5459	0.7885	<b>15.09</b>	0.7963	0.5291	20.78	0.4335	0.5956	19.67	0.5529	<b>0.8062</b>	15.16	<b>0.8116</b>

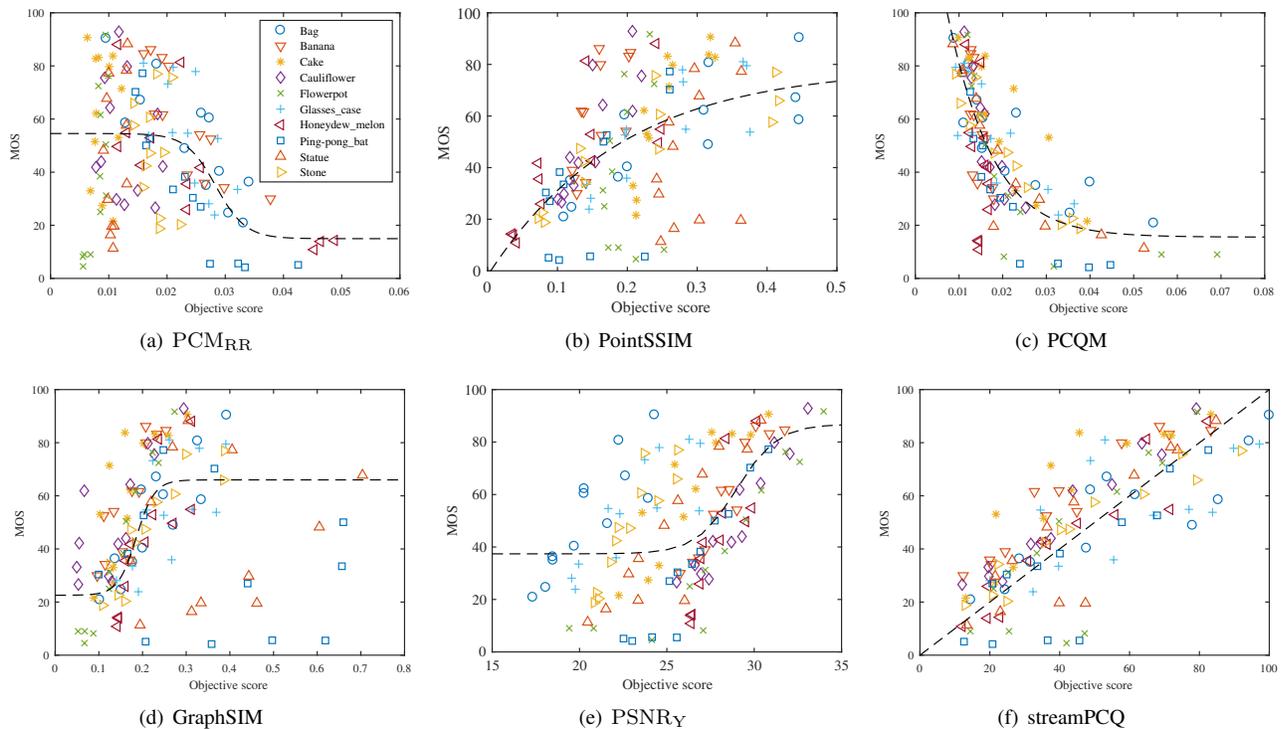


Fig. 13. Scatter plots of MOS vs. objective model prediction.

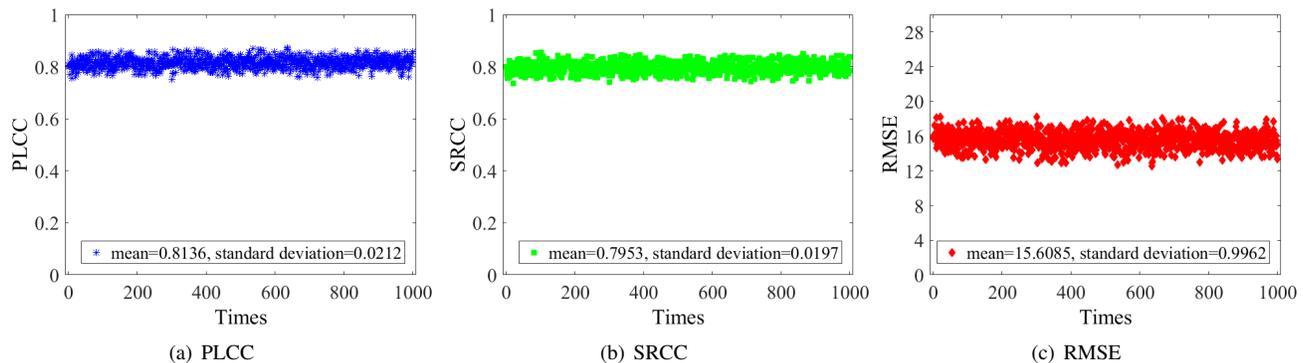


Fig. 14. Robustness of the proposed streamPCQ model. PLCC, SRCC and RMSE between MOS and objective score are computed for 1000 random splits of training/validation sets.

analysis by following the approach introduced in [76]. First, a nonlinear regression function is applied to map the objective quality scores to predict the subjective scores. We observe that the prediction residuals all have zero-mean, and thus the model with lower variance is generally considered better than

the one with higher variance. We conduct a hypothesis testing using F-statistics. Since the number of samples exceeds 50, the Gaussian assumption of the residuals approximately hold [77]. The test statistic is the ratio of variances. The null hypothesis is that the prediction residuals from one quality model come from

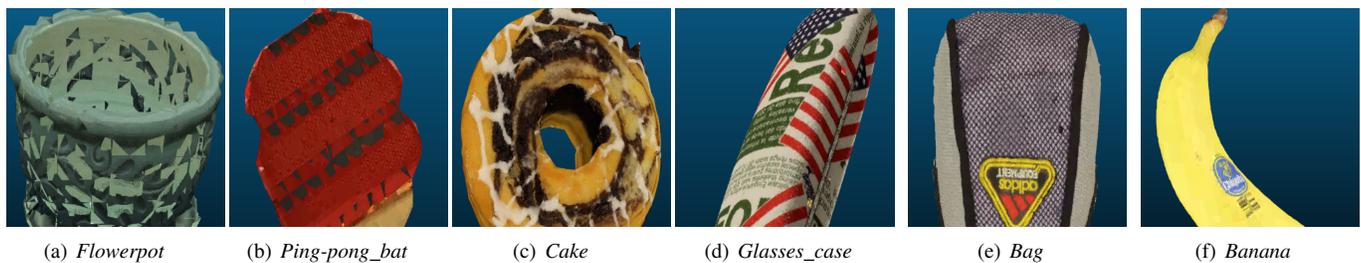


Fig. 15. Outlier PC content.

TABLE VII  
CROSS-DATABASE VALIDATION RESULTS ON M-PCCD

PCQA model	PLCC	SRCC	RMSE
PCM <sub>RR</sub>	0.7820	0.8755	1.3783
PointSSIM	0.9331	0.8915	0.4958
PCQM	0.7182	0.8950	1.3783
GraphSIM	<b>0.9376</b>	<b>0.9474</b>	<b>0.4791</b>
PSNR <sub>γ</sub>	0.6508	0.5786	1.0464
streamPCQ	0.6909	0.7760	0.9965

the same distribution and are statistically indistinguishable (with 95% confidence) from the residuals from another model. We compare every possible pairs of objective models. The results are summarized in Table VIII, where a symbol "1" means that the performance of the row model is statistically better than that of the column model, a symbol "0" means the opposite, and a symbol "-" means that the two models are statistically indistinguishable. It can be observed that most of the existing PCQA metrics are statistically indistinguishable from each other. Moreover, the streamPCQ and PCQM models are statistically indistinguishable but better than all other models, demonstrating the outstanding performance of streamPCQ as the only NR model under comparison.

Recall that one of the major motivations in developing BB NR-PCQA models is the potential low computational cost. To compare the computational complexity of streamPCQ against other PCQA models, all competing models are evaluated in terms of their execution time for three representative PCs at low, middle and high content complexity levels (measured as their total points in the PCs). The test is conducted on a Dell Precision 5820 Tower with a 3.6GHz Intel(R) Xeon(R) W-2123 processor, 32GB of RAM, Seagate BUP SCSI Disk Device, and Windows 10 Professional operating system. The execution times of all models are given in Table IX, where models were sorted in ascending order with respect to their execution time. For easy comparison, we also provided normalized execution time relative to the number of points in the PC, and relative to the streamPCQ model. It can be observed that the execution times of PCQM, PointSSIM, PSNR<sub>γ</sub> and PCM<sub>RR</sub> are roughly proportional to the number of points in the cloud, while that of GraphSIM expands at a much slower speed. By contrast, the execution time of streamPCQ does not grow with the complexity of the PCs, and is by far the lowest among all competing models. The time saving is the highest on PCs of high content complexity. The speed advantage, together with its NR nature, makes the proposed streamPCQ model a

better choice in practical especially time-critical applications.

## V. CONCLUSION

In this work, we tackle the problem of bitstream-based perceptual quality assessment of compressed 3D PCs. We make one of the first attempts to develop a BB NR-PCQA method, namely streamPCQ, without fully decoding compressed PC bitstreams. Encoding parameters, including Trisoup node size, texture quantization parameters, and texture bitrates, are extracted by a bitstream analyzer, followed by texture complexity estimation, texture distortion assessment, and geometric distortion modeling, before an overall streamPCQ quality prediction is reached. We compare the proposed NR streamPCQ model with state-of-the-art FR and RR PCQA models and find that streamPCQ delivers highly competitive performance with a fraction of computational cost.

It should be noted that the limitation of the streamPCQ model is that only Trisoup+RAHT distortion is measured. Future work will focus on developing bitstream-based PCQA models that consider other compression distortion types, such as Octree+RAHT, Octree+Lifting, Trisoup+Lifting, V-PCC and so on.

## REFERENCES

- [1] Q. Liu, H. Yuan, J. Hou, R. Hamzaoui, and H. Su, "Model-based joint bit allocation between geometry and color for video-based 3D point cloud compression," *IEEE Trans. Multimedia*, 2020.
- [2] F. Yang and S. Wan, "Bitstream-based quality assessment for networked video: a review," *IEEE Comm. Magazine*, vol. 50, no. 11, pp. 203–209, 2012.
- [3] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, "Perceptual quality assessment of colored 3D point clouds," *IEEE Trans. Visualization and Computer Graphics*, 2022.
- [4] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, April 2017.
- [5] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *Proc. IEEE Int. Conf. Image Processing*. IEEE, 2017, pp. 3460–3464.
- [6] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Improving PSNR-based quality metrics performance for point cloud geometry," in *Proc. IEEE Int. Conf. Image Processing*. IEEE, 2020, pp. 3438–3442.
- [7] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A generalized hausdorff distance based quality metric for point cloud geometry," in *Proc. IEEE Int. Conf. Quality of Multimedia Experience*. IEEE, 2020, pp. 1–6.
- [8] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *Proc. IEEE Int. Conf. Multimedia and Expo*. IEEE, 2018, pp. 1–6.
- [9] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Mahalanobis based point to distribution metric for point cloud geometry quality evaluation," *IEEE Signal Processing Letters*, vol. 27, pp. 1350–1354, 2020.

TABLE VIII  
STATISTICAL SIGNIFICANCE COMPARISON MATRIX BASED ON QUALITY PREDICTION RESIDUALS

	PCM <sub>RR</sub>	PointSSIM	PCQM	GraphSIM	PSNR <sub>Y</sub>	streamPCQ
PCM <sub>RR</sub>	-	-	0	-	-	0
PointSSIM	-	-	0	-	-	0
PCQM	1	1	-	1	1	-
GraphSIM	-	-	0	-	-	0
PSNR <sub>Y</sub>	-	-	0	-	-	0
streamPCQ	1	1	-	1	1	-

TABLE IX  
SPEED COMPARISON OF PCQA MODELS

Content (Point number)	PCQA model	Execution time (Seconds)	Execution time (10 <sup>-7</sup> second per point)	Execution time (Relative to streamPCQ)
Pingpong_bat (703879 points)	streamPCQ (NR)	<b>0.1649</b>	<b>2.343</b>	1
	PCQM (FR)	15.33	217.8	92.97
	PointSSIM (FR)	51.25	728.0	310.8
	PSNR <sub>Y</sub> (FR)	51.49	731.5	312.2
	GraphSIM (FR)	263.7	3747	1599
	PCM <sub>RR</sub> (RR)	360.9	5128	2189
Stone (1086453 points)	streamPCQ (NR)	<b>0.1584</b>	<b>1.458</b>	1
	PCQM (FR)	20.02	184.3	126.4
	PSNR <sub>Y</sub> (FR)	91.92	846.0	580.3
	PointSSIM (FR)	94.78	872.4	598.4
	PCM <sub>RR</sub> (RR)	661.6	6089	4177
	GraphSIM (FR)	824.4	7588	5205
Cake (2486566 points)	streamPCQ (NR)	<b>0.1846</b>	<b>0.7424</b>	1
	PCQM (FR)	57.73	232.2	312.7
	PSNR <sub>Y</sub> (FR)	192.5	774.3	1043
	PointSSIM (FR)	200.4	805.7	1086
	PCM <sub>RR</sub> (RR)	1474	5927	7985
	GraphSIM (FR)	3920	15760	21240

- [10] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A point-to-distribution joint geometry and color metric for point cloud quality assessment," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, IEEE, 2021, pp. 1–6.
- [11] E. Torlig, E. Alexiou, T. Fonseca, R. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Proc. SPIE Optical Engineering+Applications*, 2018, pp. 107 520I.1–107 520I.17.
- [12] E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," in *Proc. IEEE Int. Conf. Quality of Multimedia Experience*, IEEE, 2019, pp. 1–6.
- [13] E. Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Trans. on Signal and Information Processing*, vol. 8, 2019.
- [14] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3D-to-2D projection-based exploration," *IEEE Trans. Multimedia*, 2020.
- [15] Z. He, G. Jiang, Z. Jiang, and M. Yu, "Towards a colored point cloud quality assessment method using colored texture and curvature projection," in *Proc. IEEE Int. Conf. Image Processing*, IEEE, 2021, pp. 1444–1448.
- [16] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong, "Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4630–4644, 2021.
- [17] Z. He, G. Jiang, M. Yu, Z. Jiang, Z. Peng, and F. Chen, "TGP-PCQA: Texture and geometry projection based quality assessment for colored point clouds," *Journal of Visual Communication and Image Representation*, vol. 83, p. 103449, 2022.
- [18] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Joint geometry and color projection-based point cloud quality metric," *IEEE Access*, vol. 10, pp. 90 481–90 497, 2022.
- [19] Z. Wang, Y. Zhang, Q. Yang, Y. Xu, J. Sun, and S. Liu, "Point cloud quality assessment using 3D saliency maps," *arXiv preprint arXiv:2209.15475*, 2022.
- [20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [21] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, IEEE, 2020, pp. 1–6.
- [22] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: A full-reference quality metric for colored 3D point clouds," in *Proc. IEEE Int. Conf. Quality of Multimedia Experience*, 2020.
- [23] G. Meynet, J. Digne, and G. Lavoué, "PC-MSDM: A quality metric for 3D point clouds," in *Proc. IEEE Int. Conf. Quality of Multimedia Experience*, IEEE, 2019, pp. 1–3.
- [24] L. Hua, M. Yu, G. Jiang, Z. He, and Y. Lin, "VQA-CPC: a novel visual quality assessment metric of color point clouds," in *Optoelectronic Imaging and Multimedia Technology VII*, vol. 11550. International Society for Optics and Photonics, 2020, p. 1155012.
- [25] Y. Xu, Q. Yang, L. Yang, and J.-N. Hwang, "EPES: Point cloud quality modeling using elastic potential energy similarity," *IEEE Trans. Broadcasting*, 2021.
- [26] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2020.
- [27] Y. Zhang, Q. Yang, and Y. Xu, "MS-GraphSIM: Inferring point cloud quality via multiscale graph similarity," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 1230–1238.
- [28] Z. Lu, H. Huang, H. Zeng, J. Hou, and K.-K. Ma, "Point cloud quality assessment via 3D edge similarity measurement," *IEEE Signal Processing Letters*, vol. 29, pp. 1804–1808, 2022.
- [29] K.-x. Zhang, G.-y. Jiang, and M. Yu, "FQM-GC: Full-reference quality metric for colored point cloud based on graph signal features and color features," in *ACM Multimedia Asia*, 2021, pp. 1–5.
- [30] R. Diniz, P. G. Freitas, and M. C. Farias, "Towards a point cloud quality assessment model using local binary patterns," in *Proc. IEEE Int. Conf. Quality of Multimedia Experience*, IEEE, 2020, pp. 1–6.
- [31] R. Diniz, P. G. Freitas, and M. C. Farias, "Multi-distance point cloud quality assessment," in *Proc. IEEE Int. Conf. Image Processing*, IEEE, 2020, pp. 3443–3447.
- [32] R. Diniz, P. G. Freitas, and M. Farias, "A novel point cloud quality assessment metric based on perceptual color distance patterns," *Electronic Imaging*, vol. 2021, no. 9, pp. 256–1, 2021.
- [33] R. Diniz, P. G. Freitas, and M. C. Farias, "Local luminance patterns for point cloud quality assessment," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, IEEE, 2020, pp. 1–6.
- [34] R. Diniz, M. Q. Farias, and P. Garcia-Freitas, "Color and geometry texture descriptors for point-cloud quality assessment," *IEEE Signal Processing Letters*, 2021.
- [35] E. Alexiou, I. Viola, and P. Cesar, "PointPCA: Point cloud objective quality assessment using PCA-based descriptors," *arXiv preprint arXiv:2111.12663*, 2021.
- [36] L. Hua, M. Yu, Z. He, R. Tu, and G. Jiang, "CPC-GSCT: Visual quality assessment for coloured point cloud based on geometric segmentation and colour transformation," *IET Image Processing*, 2021.
- [37] I. Viola, S. Subramanyam, and P. César, "A color-based objective quality metric for point cloud contents," in *Proc. IEEE Int. Conf. Quality of Multimedia Experience*, 2020.
- [38] Q. Yang, S. Chen, Y. Xu, J. Sun, M. S. Asif, and Z. Ma, "Point cloud distortion quantification based on potential energy for human and machine perception," *arXiv preprint arXiv:2103.02850*, 2021.
- [39] A. Chetouani, M. Quach, G. Valenzise, and F. Dufaux, "Convolutional neural network for 3D point cloud quality assessment with reference," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, IEEE, 2021, pp. 1–6.
- [40] A. Chetouani, M. Quach, G. Valenzise, and F. Dufaux, "Deep learning-based quality assessment of 3D point clouds without reference," in *Proc. IEEE Int. Conf. Multimedia and Expo Workshops*, IEEE, 2021, pp. 1–6.
- [41] M. Tliba, A. Chetouani, G. Valenzise, and F. Dufaux, "Point cloud quality assessment using cross-correlation of deep features," in *Proceedings of the 2nd Workshop on Quality of Experience in Visual Multimedia Applications*, 2022, pp. 63–68.

- [42] I. Viola and P. Cesar, "A reduced reference metric for visual quality evaluation of point cloud contents," *IEEE Signal Processing Letters*, 2020.
- [43] Q. Liu, H. Yuan, R. Hamzaoui, H. Su, J. Hou, and H. Yang, "Reduced reference perceptual quality model with application to rate control for video-based point cloud compression," *IEEE Trans. Image Processing*, vol. 30, pp. 6623–6636, 2021.
- [44] K. Cao, Y. Xu, and P. Cosman, "Visual quality of compressed mesh and point cloud sequences," *IEEE Access*, vol. 8, pp. 171 203–171 217, 2020.
- [45] W.-x. Tao, G.-y. Jiang, Z.-d. Jiang, and M. Yu, "Point cloud projection and multi-scale feature fusion network based blind quality assessment for colored point clouds," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 5266–5272.
- [46] L. Hua, G. Jiang, M. Yu, and Z. He, "BQE-CVP: Blind quality evaluator for colored point cloud based on visual perception," in *Proc. IEEE Int. Sym. Broadband Multimedia Systems and Broadcasting*. IEEE, 2021, pp. 1–6.
- [47] Q. Liu, H. Yuan, H. Su, H. Liu, Y. Wang, H. Yang, and J. Hou, "PQA-Net: Deep no reference point cloud quality assessment via multi-view projection," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4645–4660, 2021.
- [48] Z. Zhang, W. Sun, X. Min, T. Wang, W. Lu, and G. Zhai, "No-reference quality assessment for 3D colored point cloud and mesh models," *IEEE Trans. Circuits and Systems for Video Technology*, 2022.
- [49] W. Zhou, Q. Yang, Q. Jiang, G. Zhai, and W. Lin, "Blind quality assessment of 3D dense point clouds with structure guided resampling," *arXiv preprint arXiv:2208.14603*, 2022.
- [50] Q. Yang, Y. Liu, S. Chen, Y. Xu, and J. Sun, "No-reference point cloud quality assessment via domain adaptation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2022, pp. 21 179–21 188.
- [51] Y. Liu, Q. Yang, Y. Xu, and L. Yang, "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM Trans. on Multimedia Computing, Communications, and Applications (TOMM)*, 2022.
- [52] Z. Zhang, W. Sun, X. Min, Q. Zhou, J. He, Q. Wang, and G. Zhai, "MM-PCQA: Multi-modal learning for no-reference point cloud quality assessment," *arXiv preprint arXiv:2209.00244*, 2022.
- [53] R. Tu, G. Jiang, M. Yu, T. Luo, Z. Peng, and F. Chen, "V-PCC projection based blind point cloud quality assessment for compression distortion," *IEEE Trans. Emerging Topics in Computational Intelligence*, 2022.
- [54] Z. Zhang, W. Sun, X. Min, Y. Fan, and G. Zhai, "Treating point cloud as moving camera videos: A no-reference quality assessment metric," *arXiv preprint arXiv:2208.14085*, 2022.
- [55] Y. Fan, Z. Zhang, W. Sun, X. Min, N. Liu, Q. Zhou, J. He, Q. Wang, and G. Zhai, "A no-reference quality assessment metric for point cloud based on captured video sequences," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*. IEEE, 2022, pp. 1–5.
- [56] S. Bourbia, A. Karine, A. Chetouani, M. El Hassouni, and M. Jridi, "No-reference point clouds quality assessment using transformer and visual saliency," in *Proceedings of the 2nd Workshop on Quality of Experience in Visual Multimedia Applications*, 2022, pp. 57–62.
- [57] S. Van Damme, M. T. Vega, J. van der Hooft, and F. De Turck, "Clustering-based psychometric no-reference quality model for point cloud video," in *Proc. IEEE Int. Conf. Image Processing*. IEEE, 2022, pp. 1866–1870.
- [58] Q. Liu, H. Su, T. Chen, H. Yuan, and R. Hamzaoui, "No-reference bitstream-layer model for perceptual quality assessment of V-PCC encoded point clouds," *IEEE Trans. Multimedia*, 2022.
- [59] 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.
- [60] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Trans. Signal and Information Processing*, vol. 9, 2020.
- [61] G. M. Morton, "A computer oriented geodetic data base and a new technique in file sequencing," 1966.
- [62] R. L. De Queiroz and P. A. Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Trans. Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016.
- [63] *Opinion Model for Video-telephony Applications*, ITU-T Recommendation G.1070, 2007.
- [64] *Opinion Model for Network Planning of Video and Audio Streaming Applications*, ITU-T Recommendation G.1071, 2015.
- [65] *Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality*, ITU-T Recommendation P.1201, 2012.
- [66] *Parametric Bitstream-based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services Over Reliable Transport*, ITU-T Recommendation P.1203, 2017.
- [67] *Video Quality Assessment of Streaming Services Over Reliable Transport for Resolutions up to 4K*, ITU-T Recommendation P.1204, 2020.
- [68] *Parametric Non-intrusive Bitstream Assessment of Video Media Streaming Quality*, ITU-T Recommendation P.1202, 2012.
- [69] R. J. Clarke, "Transform coding of images," 1985.
- [70] A. J. Viterbi and J. K. Omura, "Principles of digital communication and coding," 2013.
- [71] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Information Theory*, vol. 14, no. 5, pp. 676–683, 1968.
- [72] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual quality assessment of 3D point clouds," in *Proc. IEEE Int. Conf. Image Processing*. IEEE, 2019, pp. 3182–3186.
- [73] R. J. Mekuria, Z. Li, C. Tulvan, and P. Chou, "Evaluation criteria for PCC (point cloud compression)," *ISO/IEC JTC1/SC29/WG11 MPEG, N16332*, 2016.
- [74] G. Lavoué and R. Mantiuk, "Quality assessment in computer graphics," in *Visual Signal Quality Assessment*. Springer, 2015, pp. 243–286.
- [75] M. Corsini, M.-C. Larabi, G. Lavoué, O. Petřík, L. Váša, and K. Wang, "Perceptual metrics for static and dynamic triangle meshes," in *Computer Graphics Forum*, vol. 32, no. 1. Wiley Online Library, 2013, pp. 101–125.
- [76] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.
- [77] D. C. Montgomery and G. C. Runger, "Applied statistics and probability for engineers," 2014.



ing, immersive media processing, computer vision, etc.



Waterloo, ON, Canada. Her research interests include point cloud compression and point cloud quality assessment.

**Honglei Su** (M'19) received a B.A. degree from Shandong University of Science and Technology, Qingdao, China in 2008 and a Ph.D. degree from Xidian University, Xi'an, China in 2014. From 2014.09 to the present, he works as an Assistant Professor with the School of Electronic Information, Qingdao University, Qingdao, China. From 2018.03-2019.03, he also worked as a visiting scholar with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include perceptual image process-

**Qi Liu** received a B.S. degree from Shandong Technology and Business University, Yantai, China in 2011; an M.S. degree from Xidian University, Xi'an, China in 2014; and a Ph.D. degree from Shandong University, Qingdao, China in 2021. From 2021.12 to the present, she works as an Assistant Professor with the School of Electronic Information, Qingdao University, Qingdao, China. From 2018.09-2019.08, she also worked as an international visiting graduate student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Her research interests include point cloud compression and point cloud quality assessment.



**Yuxin Liu** was born in Zhoukou, Henan, HN, China in 1997. He is a postgraduate student in electronic engineering from Qingdao University, Qingdao in 2021. His research interests include immersive media quality assessment.



**Zhou Wang** (S'99-M'02-SM'12-F'14) received the Ph.D. degree from The University of Texas at Austin, TX, USA, in 2001. He is currently a Professor and a Canada Research Chair with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include image and video processing and coding; visual quality assessment and optimization; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 200 publications in these fields with over 70,000 citations (Google Scholar).

Dr. Wang serves as a member for the IEEE Multimedia Signal Processing Technical Committee from 2013 to 2015 and the IEEE Image, Video and Multidimensional Signal Processing Technical Committee from 2020 to 2022. He was elected as a Fellow of Canadian Academy of Engineering in 2016 and the Royal Society of Canada: Academy of Science in 2018. He was a recipient of the 2009 IEEE Signal Processing Society Best Paper Award, the 2013 *IEEE Signal Processing Magazine* Best Paper Award, the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2015 Primetime Engineering Emmy Award, and the 2016 IEEE Signal Processing Society Sustained Impact Paper Award. He serves as a Senior Editor for *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING* from 2022 to 2024, a Senior Area Editor for *IEEE TRANSACTIONS ON IMAGE PROCESSING* from 2015 to 2019, an Associate Editor for *IEEE SIGNAL PROCESSING LETTERS* from 2006 to 2010, *IEEE TRANSACTIONS ON IMAGE PROCESSING* from 2009 to 2014, and *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY* from 2016 to 2018, and a Guest Editor for *IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING* from 2013 to 2014 and from 2007 to 2009, among other journals.



**Hui Yuan** (S'08-M'12-SM'17) received a B.E. and Ph.D. degree in Telecommunication Engineering from Xidian University, Xi'an, China, in 2006 and 2011, respectively. From 2011.04 to the present, he works as a Lecturer, Associate Professor, and Full Professor at Shandong University (SDU), Jinan, China. From 2013.01-2014.12 and 2017.11-2018.02, he also worked as a postdoctoral fellow (Hong Kong Scholar) and a research fellow, respectively, with the Department of Computer Science, City University of Hong Kong (CityU). His current research interests

include video/image/immersive media compression, adaptive video streaming, computer vision, etc.



**Huan Yang** received the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2015, the M.S. degree in computer science from Shandong University, China, in 2010, and the B.S. degree in computer science from the Heilongjiang Institute of Technology, China, in 2007. Currently, she is working in the College of Computer Science and Technology, Qingdao University, Qingdao China. Her research interests include image/video processing and analysis, perception-based modeling and quality assessment, object detection/recognition, and machine learning.



**Zhenkuan Pan** received his Ph.D. in 1992 from the Shanghai Jiao Tong University of Science and Technology. Since 1996, he has been a full professor at Qingdao University. He is also a member of the virtual reality professional committee of China graphic image association. He is the author of more than 300 papers. His interests include computer vision, image processing, and pattern recognition. He is also working on the application of multibody system dynamics and control.