# OPTIMIZING PROJECTION-BASED POINT CLOUD QUALITY ASSESSMENT WITH HUMAN PREFERRED VIEWPOINTS SELECTION

*Zicheng Zhang*[1], *Yu Fan*[1], *Wei Sun*[1], *Xiongkuo Min*[1], *Xiaohong Liu*[1], *Chunyi Li*[1],
Haoning Wu[2], Weisi Lin[2], Ning Liu[1†], and Guangtao Zhai[1]

[1]Shanghai Jiaotong University, [2]S-Lab, Nanyang Technological University

## ABSTRACT

Viewpoint selection plays a pivotal role in projection-based point cloud quality assessment (PCQA). Generally speaking, sole reliance on a single projection fails to capture adequate quality information, leading to the prevalent use of multi-projection approaches. It is important to recognize that viewpoint selection is significantly influenced by human preferences and viewpoints that align with human predilections exert a greater impact on PCQA. Therefore, we introduce the first viewpoint selection database for PCQA, which comprises 405 distorted point clouds, accompanied by preferred viewpoints collected from humans. Then we propose a novel human preference index, devised from the Visible-Points Ratio and Visible-Color-Entropy Ratio, to guide the selection of viewpoints. Our experimental findings confirm that this human preference index correlates more closely with human preferences than traditional viewpoint selection settings. Moreover, the proposed PCQA method optimized with the human preference index demonstrates competitive performance as well.

***Index Terms***— Viewpoint selection, Projection-based, Point cloud quality assessment

## 1. INTRODUCTION

Point cloud quality assessment (PCQA) is targeted at predicting the visual quality levels of the point clouds. During the last decade, many PCQA models have been developed, which can be divided into model-based and projection-based PCQA methods. The model-based PCQA methods [1, 2, 3, 4, 5, 6, 7] extract quality-aware features directly from the point cloud. Benefiting from the advanced 2D quality assessment techniques [8, 9, 10, 11, 12, 13, 14, 15], the projection-based PCQA models [16, 17] infer the point cloud quality via the rendered projections, where viewpoint selection is very significant.

This work was supported in part by NSFC (No.62225112, No.61831015, No. 62301316), the Fundamental Research Funds for the Central Universities, NKPs No. 2021YFE0206700, Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102), STCSM 22DZ2229005, CPSF No. 2023TQ0212 and 2023M742298, and the Postdoctoral Fellowship Program of CPSF No. GZC20231618.

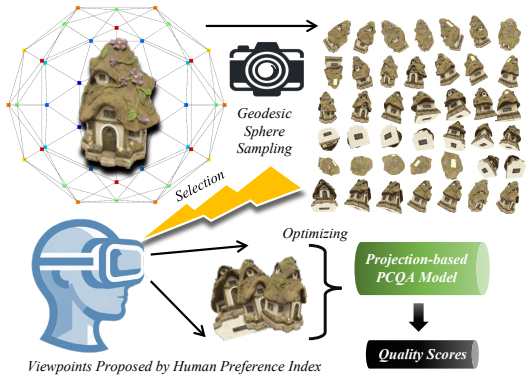


**Fig. 1**. The framework of the proposed approach. 42 projections are generated from the geodesic sphere sampling process. Then the viewpoints are proposed with the Human Preferred Index and forwarded as the input of the projection-based PCQA model.

However, few efforts have been put into investigating the viewpoint selection for the PCQA task currently. Therefore, in this paper, we propose the first-of-a-kind viewpoint selection database for the PCQA task. Specifically, we collect 15 high-quality source point clouds and derive 405 distorted point clouds for the subjective rating. During the subjective experiment, the human subjects can freely rotate the point cloud (with fixed viewing distance) for observation and are required to record their preferred viewpoints (at least *THREE*) based on the principle that these viewpoints are suitable for quality evaluation from their perspective. Considering the viewpoints are infinite, we project the recorded viewpoints to the closest geodesic sphere sampled (GSS) viewpoints as the ground truth. Then we can obtain the most-preferred GSS-viewpoints list ranked by the votes from the human subjects. Afterward, we design a novel viewpoint preference index to predict the preferred viewpoints of the point cloud based on the Visible-Points Ratio and Visible-Color-Entropy Ratio. The experimental results reveal that the proposed viewpoint preference index has a high correlation with human preference. More importantly, the proposed PCQA method optimized with the human preference index achieves the best performance among the PCQA competitors, which further confirms the effectiveness of the human preference index.


**Fig. 2**. Overview of the selected 15 reference point clouds, which contain 8 humans and 7 objects.

## 2. RELATED WORKS

During the early stages of PCQA development, MPEG introduces several point-based full-reference (FR) PCQA techniques, notably p2point [1] and p2plane [2]. A point-based PSNR-yuv is later introduced to handle colored point clouds [3]. However, these methods face challenges in accurately depicting complex distortions at the point level, leading to the development of advanced FR-PCQA metrics such as PCQM [4], GraphSIM [5], and PointSSIM [6], which incorporate structural features and demonstrate significant improvements in performance.

Various NR-PCQA methods are also developed to cater to a broader range of applications. Chetouani et al. [18] apply classical CNN models for quality regression after extracting patch-wise hand-crafted features. PQA-net [16] uses multi-view projection for feature extraction. Zhang et al. [19] estimate quality-aware parameters from the distributions of geometry and color attributes using several statistical distributions. Liu et al. [7] utilize an end-to-end sparse CNN for quality prediction, while Zhou et al. [20] adopt structure-guided resampling for extracting quality-related features. Some research [21, 22] even converts point clouds into videos and leverages video quality assessment (VQA) techniques for evaluating perceptual quality. Yang et al. [23] additionally transfer quality information from natural images to comprehend the quality of point cloud rendering images using domain adaptation. However, these methods focus on single-modal information and do not effectively integrate multi-modal quality information.

**Table 1**. Details for distortion generation. The distortion values are adjusted manually to cover the majority quality range.

| Type | Distortion Parameter Description | Value |
|---|---|---|
| Color Noise | Color noise standard deviation | 20, 30, 40, 50 |
| Downsampling | Points Removal Proportion | 25%, 50%, 75%, 90% |
| Geometry Noise | Geometry noise standard deviation | 5, 25, 50, 70 |
| G-PCC (octree) | Octree-based G-PCC compression parameter | 28, 34, 40, 46, 51 |
| G-PCC (trisoup) | Trisoup-based G-PCC compression parameter | 22, 28, 34, 40, 46 |
| V-PCC | V-PCC compression parameter | 20, 24, 28, 32, 37 |

## 3. DATABASE CONSTRUCTION

### 3.1. Point Clouds Generation

To ensure the diversity of the reference point clouds, we select both high-quality human and object point clouds. The overview of the selected point clouds is exhibited in Fig.2, from which we can see that the collected point clouds are rich in color and texture. To investigate the impact of different distortions on the human-preferred viewpoint selection, we further degrade the point clouds with 6 common distortions, which include color noise, downsampling, geometry noise, G-PCC (octree) compression [24], G-PCC (trisoup) compression [24], and V-PCC compression [24]. The detailed distortion description and parameters are listed in Table 1. It's worth mentioning that there are five levels for the compression distortions and four levels for the others. Then $15\times27$=405 distorted point clouds are obtained in the end. Additionally, the examples of the distorted *Biplane* point clouds are illustrated in Fig. 3.

### 3.2. Human-Preferred Viewpoints Collection

A total of 20 human subjects are invited to participate in the subjective experiment. We design a 2D interface on the screen which allows the viewers to freely rotate the point clouds with the default viewing distance for observation. During the observation, the viewers are required to select at least *THREE* preferred (based on the principle that the viewpoints are suitable for PCQA) viewpoints by pressing the space key, and the preferred viewpoints' camera parameters are recorded for the collection. Thus there are more than $15\times3\times405$=18,225 annotations collected at last.

### 3.3. Geodesic Sphere Sampling

A point cloud is visible from infinite viewpoints on the viewing sphere. The subjective experiment offers continuous viewpoints on this sphere. However, the viewpoint selection methods work with a finite set of viewpoints, necessitating the sampling of the viewing sphere. Consequently, we employ a geodesic sphere sampling (GSS) approach [25] to achieve uniform sampling of viewpoints from the viewing sphere. The detailed explanation for the geodesic sphere is shown in Fig. 4. We then project the human-preferred perspectives

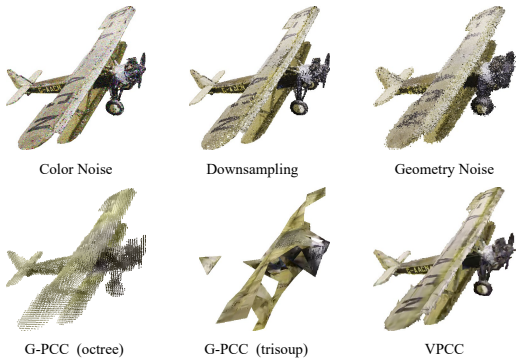


**Fig. 3**. Examples of the distorted *Biplane* point clouds.

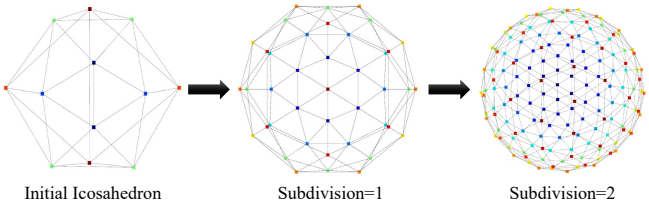


**Fig. 4**. The evolving process of the geodesic sphere. A geodesic sphere is initiated from a *regular icosahedron*. The triangular faces of the icosahedron are repeatedly *subdivided* into smaller triangles to increase the sphere's resolution. Considering the computation complexity and rendering resources consumption, we choose the *'Subdivision=1' geodesic sphere* as the default GSS setting, which includes *42 uniform viewpoints*.

onto the closest GSS-viewpoints as the ground truth. Therefore, for each viewer, we can obtain at least *THREE* human-preferred GSS-viewpoints for a single point cloud. Then we can derive the most-preferred GSS-viewpoints list ranked by the votes of the viewpoints for each distorted point cloud (the viewpoints that are voted for only once are ignored).

### 3.4. Subjective Data Analysis

With the collected human-preferred GSS-viewpoints, we exhibit the heatmaps of the viewpoint distributions for each group of point clouds in Fig. 5. From the heatmaps, we can make several interesting observations: a) The preferred GSS-viewpoints for human point clouds tend to be more consistent and centered compared to those for object point clouds. Upon analysis, it is evident that viewers generally favor the frontal view when it comes to human point clouds, a preference that does not extend to object point clouds. b) The symmetric point clouds tend to have more scattered viewpoints such as *Biplane* and *Statue*. This is because the scattered viewpoints might have similar viewing content due to the symmetry.

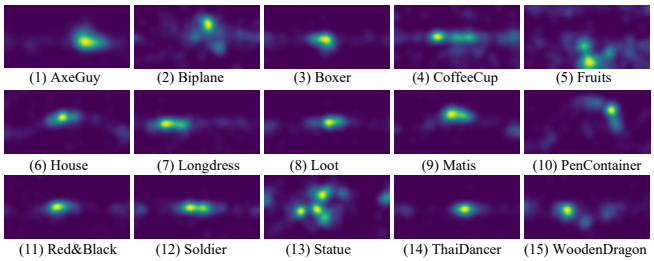


**Fig. 5**. The heatmaps for the human-preferred viewpoints distribution. It is apparent that the preferred GSS-viewpoints of object point clouds are more scattered than the human point clouds.

## 4. VIEWPOINT SELECTION & QUALITY ASSESSMENT

In this section, we propose a viewpoint selection method based on the visible points and show how to advance the NR-PCQA methods with the selected viewpoints.

### 4.1. Viewpoint Selection Module

#### *4.1.1. Visible-Points Ratio*

As proven in [25], humans prefer viewpoints with more viewing area since these viewpoints can provide more information. Therefore, we utilize the visible-points ratio to represent the viewing area:

$$\begin{aligned} \mathcal{R}_P^{v_i} &= \frac{\hat{N_p^{v_i}}}{N_p}, \\ \mathcal{V} &= \{v_i | i = 0 \sim N_v - 1\}, \end{aligned} \tag{1}$$

where $\mathcal{V}$ indicates the set of GSS-viewpoints, $N_v$ is the number of GSS-viewpoints, $\hat{N_p^{v_i}}$ represents the visible points number of the $i$-th GSS-viewpoint, $N_p$ is the number of all the points in the point cloud, and $\mathcal{R}_P^{v_i}$ stands for the visible-points ratio (VPR) of the $i$-th GSS-viewpoint.

#### *4.1.2. Visible-Color-Entropy Ratio*

Furthermore, to incorporate color information into the viewpoint selection, we propose a visible-color-entropy ratio to address this issue:

$$\mathcal{R}_{CE}^{v_i} = \frac{\hat{E_r^{v_i}}}{E_r} + \frac{\hat{E_g^{v_i}}}{E_g} + \frac{\hat{E_b^{v_i}}}{E_b}, \tag{2}$$

where $\hat{E_r^{v_i}}$, $\hat{E_g^{v_i}}$, and $\hat{E_b^{v_i}}$ denote the *RGB* entropy values of the $i$-th GSS-viewpoint, $E_r$, $E_g$, and $E_b$ stand for the *RGB* entropy values of all points in the point cloud, and $\mathcal{R}_{CE}^{v_i}$ represents the visible-color-entropy ratio (VCER).

#### 4.1.3. Viewpoint Preference Index

Afterward, the final viewpoint preference index can be obtained as the weighted sum of VPR and VCER:

$$\mathcal{VPI}^{v_i} = \mathcal{W}_1 \mathcal{R}_P^{v_i} + \mathcal{W}_2 \mathcal{R}_{CE}^{v_i}, \tag{3}$$

where $\mathcal{VPI}^{v_i}$ is the preference index for the $i$-th GSS-viewpoint, and $\mathcal{W}_1$ and $\mathcal{W}_2$ are the weight parameters. Finally, we can rank the GSS-viewpoint set according to the preference index:

$$\tilde{\mathcal{V}} = sort\left(\mathcal{V}, key = \mathcal{VPI}^{v_i}\right), \tag{4}$$

where $\tilde{\mathcal{V}}$ is the preference index sorted GSS-viewpoints list.

### 4.2. Quality Assessment Module

#### 4.2.1. Projection Rendering

To investigate the impact of the viewpoint selection module, we simply employ a multi-projection NR-PCQA approach as the quality assessment module. Given the top-$k$ preferred viewpoints from the sorted GSS-viewpoints list $\tilde{\mathcal{V}}$, we first obtain the rendered projections of the point cloud:

$$\mathcal{P} = \{p_j | p_j = \phi(v_j), j = 0 \sim k-1\}, \tag{5}$$

where $\phi(\cdot)$ indicates the rendering process, $\mathcal{P}$ is the set of the projections from the top-$k$ preferred viewpoints. The rendering process is assisted with the *Python opend* package, where we keep the default rendering setup. Additionally, the white background of the projections is cropped out.

#### 4.2.2. Feature Extraction

We then incorporate the rendered 2D projections into quality-aware space with a 2D image encoder:

$$\begin{aligned} F_j &= \alpha(p_j), \\ \widetilde{F} &= \frac{1}{k}\sum_{t=0}^{k-1} F_j, \end{aligned} \tag{6}$$

where $F_j \in \mathbb{R}^{1\times C}$ stands for the quality-aware embedding for the $j$-th rendered projection $p_j$, $C$ denotes the number of output channels of the 2D image encoder $\alpha(\cdot)$, and $\widetilde{F} \in \mathbb{R}^{1\times C}$ is the pooled results after average fusion.

#### 4.2.3. Quality Regression

For simplification, we employ a two-stage fully-connected layer to map the averaged quality-aware embedding $\widetilde{F}$ to the estimated quality scores. Our evaluation not only considers the prediction precision but also emphasizes their relative rankings. Consequently, the loss function comprises both

**Table 2**. Performance results of *Precision@k* (k=1,2,3) for random viewpoints, cube-like viewpoints, and proposed viewpoints selection. *Precision@k* indicates the possibility of involving the human-preferred viewpoints in the corresponding top-$k$ viewpoints selection. For instance, *Precision@6*=36.34% represents that there are about 2.18=39.91%×6 human-preferred viewpoints in the top-*6* viewpoints selection on average. Best in **BOLD**.

| *Precision@k* | k=1 | k=3 | k=6 |
|---|---|---|---|
| Random Viewpoints Selection | 12.83% | 12.51% | 13.37% |
| Cube-like Viewpoints Selection | 9.62% | 21.89 % | 18.40% |
| Proposed Viewpoints Selection | **62.96%** | **50.04%** | **36.34%** |

Mean Squared Error (MSE) for accuracy and a rank error for ordering. The MSE can be expressed as:

$$L_{\text{MSE}} = \frac{1}{n}\sum_{\eta=1}^{n}(q_\eta - q'_\eta)^2, \tag{7}$$

where $q_\eta$ represents the estimated quality scores, $q'_\eta$ denotes the actual quality labels for the point cloud, and $n$ is the batch size. To better differentiate quality distinctions between point clouds with similar quality labels, we introduce a rank loss. This is approximated using the differentiable ranking function from [26], which is formulated as:

$$\begin{aligned} L_{rank}^{ij} &= \max\left(0, |q_i - q_j| - e\left(q_i, q_j\right)\cdot\left(q'_i - q'_j\right)\right), \\ e\left(q_i, q_j\right) &= \begin{cases} 1, q_i \geq q_j, \\ -1, q_i < q_j, \end{cases} \end{aligned} \tag{8}$$

where $i$ and $j$ indicate two distinct point clouds within a batch. The rank loss is then computed as:

$$L_{\text{rank}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} L_{\text{rank}}^{ij}, \tag{9}$$

The overall loss function is a weighted sum of the MSE and rank losses:

$$\text{Loss} = \lambda_1 L_{\text{MSE}} + \lambda_2 L_{\text{rank}}, \tag{10}$$

where $\lambda_1$ and $\lambda_2$ balance the contribution of MSE and rank losses respectively.

## 5. EXPERIMENT

### 5.1. Towards Viewpoint Selection Accuracy

We first assess the efficacy of the viewpoint selection module. For each distorted point cloud, we can get a preference index sorted GSS-viewpoints list and the human-preferred GSS-viewpoints list (ranked by the votes of the human subjects). Therefore, we employ the *Precision* metric to reflect the accuracy for including the human-preferred GSS-viewpoints in the proposed 6 viewpoints. The outcomes, detailed in Table 2,

reveal that our viewpoint preference index substantially outperforms both random and conventional cube-based selection methods in accuracy. These findings further corroborate the robustness of our viewpoint selection approach.

### 5.2. Towards Quality Assessment Accuracy

#### *5.2.1. Quality Assessment Competitors*

We choose 14 popular PCQA methods for comparison, which include 8 FR-PCQA methods and 6 NR-PCQA methods. The FR-PCQA methods consist of MSE-p2point (MSE-p2po) [1], Hausdorff-p2point (HD-p2po) [1], MSE-p2plane (MSE-p2pl) [2], Hausdorff-p2plane (HD-p2pl) [2], PSNR-yuv [3], PCQM [4], GraphSIM [5], and PointSSIM [6]. The NR-PCQA methods consist of BRISQUE [8], NIQE [9], IL-NIQE [10], ResSCNN [7], PQA-net [16], and 3D-NSS [19].

It's worth noting that BRISQUE, NIQE, IL-NIQE are image-based quality assessment metrics and are validated on the projections. Three mainstream criteria are utilized for evaluation, which include Spearman Rank Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Squared Error (RMSE).

#### *5.2.2. Implementation Details*

The subjective point cloud assessment database (SJTU-PCQA) [27] and the Waterloo point cloud assessment database (WPC) [28] are used for validation, which are applied with 9-fold and 5-fold cross-validation split strategy as stated in [29] respectively. There is no content overlap between the training and testing sets.

The Adam optimizer is utilized with a weight decay of $10^{-4}$. The initial learning rate is set to $5 \times 10^{-5}$, and the batch size is set to 8. The model undergoes training for a total of 50 epochs. The default number of the proposed viewpoints is established at 4. The projections are rendered with the assistance of Open3d with a resolution of 1080P, which are then randomly cropped into image patches with dimensions of 224×224×3 as input. The image encoder is implemented using the Swin Transformer tiny model [30], initialized with pre-trained weights from the ImageNet database [31]. The weighting coefficients $\lambda_1$ and $\lambda_2$ for the loss functions $L_{MSE}$ and $L_{rank}$ are both set to 1.

#### *5.2.3. Performance Discussion*

The experimental results are presented in Table 3, which clearly demonstrates that the proposed method outperforms existing approaches on the two PCQA databases under consideration. Furthermore, it is noteworthy that all evaluated methods exhibit a marked decline in performance when transitioning from the SJTU-PCQA to the WPC database. This is attributed to the more intricate and nuanced distortion levels featured in the WPC database, posing greater challenges for performance maintenance.

**Table 3**. Performance comparison with state-of-the-art approaches on the SJTU-PCQA and WPC databases. Best in **BOLD**.

| Type | Methods | SJTU-PCQA | | | WPC | | |
|---|---|---|---|---|---|---|---|
| | | SRCC↑ | PLCC↑ | RMSE↓ | SRCC↑ | PLCC↑ | RMSE↓ |
| FR | MSE-p2po | 0.7294 | 0.8123 | 1.3613 | 0.4558 | 0.4852 | 19.8943 |
| | HD-p2po | 0.7157 | 0.7753 | 1.4475 | 0.2786 | 0.3972 | 20.8990 |
| | MSE-p2pl | 0.6277 | 0.5940 | 2.2815 | 0.3281 | 0.2695 | 22.8226 |
| | HD-p2pl | 0.6441 | 0.6874 | 2.1255 | 0.2827 | 0.2753 | 21.9893 |
| | PSNR-yuv | 0.7950 | 0.8170 | 1.3151 | 0.4493 | 0.5304 | 19.3119 |
| | PCQM | 0.8644 | 0.8853 | 1.0862 | 0.7434 | 0.7499 | 15.1639 |
| | GraphSIM | 0.8783 | 0.8449 | 1.0321 | 0.5831 | 0.6163 | 17.1939 |
| | PointSSIM | 0.6867 | 0.7136 | 1.7001 | 0.4542 | 0.4667 | 20.2733 |
| NR | BRISQUE | 0.3975 | 0.4214 | 2.0937 | 0.2614 | 0.3155 | 21.1736 |
| | NIQE | 0.1379 | 0.2420 | 2.2622 | 0.1136 | 0.2225 | 23.1415 |
| | IL-NIQE | 0.0837 | 0.1603 | 2.3378 | 0.0913 | 0.1422 | 24.0133 |
| | ResSCNN | 0.8600 | 0.8100 | - | - | - | - |
| | PQA-net | 0.8372 | 0.8586 | 1.0719 | 0.7026 | 0.7122 | 15.0812 |
| | 3D-NSS | 0.7144 | 0.7382 | 1.7686 | 0.6479 | 0.6514 | 16.5716 |
| | Proposed | **0.9041** | **0.9155** | **0.9263** | **0.8329** | **0.8137** | **12.9895** |

**Table 4**. Results of ablation study on the SJTU-PCQA and WPC databases, where the proposed preference viewpoints selection is replaced with the random and cube-like viewpoints selection. Best in **BOLD**.

| Models | SJTU-PCQA | | | WPC | | |
|---|---|---|---|---|---|---|
| | SRCC↑ | PLCC↑ | RMSE↓ | SRCC↑ | PLCC↑ | RMSE↓ |
| *with* Random | 0.8517 | 0.8649 | 1.3273 | 0.7800 | 0.7815 | 14.6358 |
| *with* Cube-like | 0.8760 | 0.8689 | 1.3089 | 0.7990 | 0.8023 | 14.0019 |
| *with* Proposed | **0.9041** | **0.9155** | **0.9263** | **0.8329** | **0.8137** | **12.9895** |

Moreover, an ablation study is carried out to evaluate the efficacy of the proposed viewpoint selection module. The findings are delineated in Table 4. The results indicate that substituting the proposed viewpoint selection algorithm with random or cube-like projections results in a decrease in performance. This observation underscores the significance and effectiveness of the proposed viewpoint selection approach.

## 6. CONCLUSION

In conclusion, our research underscores the vital role of viewpoint selection for projection-based PCQA models. We have introduced the first comprehensive viewpoint selection database tailored for PCQA. Our novel human preference index, developed using the Visible-Points Ratio and Visible-Color-Entropy Ratio, emerges as a pivotal tool in guiding viewpoint selection, ensuring it aligns more closely with human preferences. The experimental results from our study not only validate the correlation of our index with human preferences but also demonstrate the enhanced performance of our proposed PCQA method.

## 7. REFERENCES


[1] R Mekuria, Z Li, C Tulvan, and P Chou, "Evaluation criteria for point cloud compression," *ISO/IEC MPEG*, 2016.

[2] Dong Tian, Hideaki Ochimizu, Chen Feng, Robert Cohen, and Anthony Vetro, "Geometric distortion metrics for point cloud compression," in *IEEE ICIP*, 2017.

[3] Eric M Torlig et al., "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*, 2018.

[4] Gabriel Meynet, Yana Nehmé, Julie Digne, and Guillaume Lavoué, "Pcqm: A full-reference quality metric for colored 3d point clouds," in *QoMEX*, 2020.

[5] Qi Yang et al., "Inferring point cloud quality via graph similarity," *IEEE TPAMI*, 2020.

[6] Evangelos Alexiou and Touradj Ebrahimi, "Towards a point cloud structural similarity metric," in *ICMEW*, 2020.

[7] Yipeng Liu et al., "Point cloud quality assessment: Dataset construction and learning-based no-reference metric," *ACM TOMM*, 2022.

[8] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE TIP*, 2012.

[9] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE SPL*, 2012.

[10] Lin Zhang, Lei Zhang, and Alan C Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE TIP*, 2015.

[11] Haoning Wu, Chaofeng Chen, Jingwen Hou, Liang Liao, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin, "Fast-vqa: Efficient end-to-end video quality assessment with fragment sampling," in *ECCV*, 2022.

[12] Haoning Wu, Chaofeng Chen, Liang Liao, Jingwen Hou, Wenxiu Sun, Qiong Yan, Jinwei Gu, and Weisi Lin, "Neighbourhood representative sampling for efficient end-to-end video quality assessment," 2023.

[13] Zicheng Zhang, Wei Sun, Xiongkuo Min, Wenhan Zhu, Tao Wang, Wei Lu, and Guangtao Zhai, "A no-reference evaluation metric for low-light image enhancement," in *IEEE ICME*, 2021.

[14] Zicheng Zhang, Yingjie Zhou, Long Teng, Wei Sun, Chunyi Li, Xiongkuo Min, Xiao-Ping Zhang, and Guangtao Zhai, "Quality-of-experience evaluation for digital twins in 6g network environments," *IEEE TBC*, 2024.

[15] Haoning Wu et al., "Q-bench: A benchmark for general-purpose foundation models on low-level vision," *ICLR*, 2023.

[16] Qi Liu et al., "Pqa-net: Deep no reference point cloud quality assessment via multi-view projection," *IEEE TCSVT*, 2021.

[17] Zicheng Zhang, Wei Sun, Haoning Wu, Yingjie Zhou, Chunyi Li, Zijian Chen, Xiongkuo Min, Guangtao Zhai, and Weisi Lin, "Gms-3dqa: Projection-based grid mini-patch sampling for 3d model quality assessment," *ACM TOMM*, 2023.

[18] Aladine Chetouani, Maurice Quach, Giuseppe Valenzise, and Frédéric Dufaux, "Deep learning-based quality assessment of 3d point clouds without reference," in *ICMEW*, 2021.

[19] Zicheng Zhang, Wei Sun, Xiongkuo Min, Tao Wang, Wei Lu, and Guangtao Zhai, "No-reference quality assessment for 3d colored point cloud and mesh models," *IEEE TCSVT*, 2022.

[20] Wei Zhou et al., "Blind quality assessment of 3d dense point clouds with structure guided resampling," *arXiv preprint arXiv:2208.14603*, 2022.

[21] Yu Fan, Zicheng Zhang, Wei Sun, Xiongkuo Min, Ning Liu, Quan Zhou, Jun He, Qiyuan Wang, and Guangtao Zhai, "A no-reference quality assessment metric for point cloud based on captured video sequences," in *IEEE MMSP*. IEEE, 2022.

[22] Zicheng Zhang, Wei Sun, Yucheng Zhu, Xiongkuo Min, Wei Wu, Ying Chen, and Guangtao Zhai, "Evaluating point cloud from moving camera videos: A no-reference metric," *IEEE TMM*, 2023.

[23] Qi Yang et al., "No-reference point cloud quality assessment via domain adaptation," in *IEEE/CVF CVPR*, 2022.

[24] Sebastian Schwarz et al., "Emerging mpeg standards for point cloud compression," *IEEE J. Emerg. Sel.*, 2018.

[25] Helin Dutagaci et al., "A benchmark for best view selection of 3d objects," in *ACM 3DOR Worshop*, 2010.

[26] Wei Sun, Xiongkuo Min, Wei Lu, and Guangtao Zhai, "A deep learning based no-reference quality assessment model for ugc videos," in *ACM MM*, 2022, pp. 856–865.

[27] Qi Yang et al., "Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration," *IEEE TMM*, 2020.

[28] Qi Liu, Honglei Su, Zhengfang Duanmu, Wentao Liu, and Zhou Wang, "Perceptual quality assessment of colored 3d point clouds," *IEEE TVCG*, 2022.

[29] Zicheng Zhang, Wei Sun, Xiongkuo Min, Quan Zhou, Jun He, Qiyuan Wang, and Guangtao Zhai, "Mm-pcqa: Multi-modal learning for no-reference point cloud quality assessment," *IJCAI*, 2023.

[30] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *IEEE/CVF ICCV*, 2021.

[31] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE/CVF CVPR*, 2009.