TOWARDS A POINT CLOUD STRUCTURAL SIMILARITY METRIC

Evangelos Alexiou and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
Emails: FirstName.LastName@epfl.ch

ABSTRACT

Point cloud is a 3D image representation that has recently emerged as a viable approach for advanced content modality in modern communication systems. In view of its wide adoption, quality evaluation metrics are essential. In this paper, we propose and assess a family of statistical dispersion measurements for the prediction of perceptual degradations. The employed features characterize local distributions of point cloud attributes reflecting topology and color. After associating local regions between a reference and a distorted model, the corresponding feature values are compared. The visual quality of a distorted model is then predicted by error pooling across individual quality scores obtained per region. The extracted features aim at capturing local changes, similarly to the well-known Structural Similarity Index. Benchmarking results using available datasets reveal best-performing attributes and features, under different neighborhood sizes. Finally, point cloud voxelization is examined as part of the process, improving the prediction accuracy under certain conditions.

Index Terms— Point cloud, objective quality metric, visual quality assessment.

1. INTRODUCTION

Objective quality evaluation is a research area that aims at investigating and proposing algorithms that are able to predict the visual quality of content representations, typically as perceived by human end-users. This research field is impactful on several tasks that are related to information and communication systems. For instance, having access to accurate predictions of quality for contents after encoding or transmission can greatly assist in improving user experience by updating corresponding configurations of the underlying systems. Moreover, the benchmarking of new solutions can be facilitated by using well-performing objective quality metrics instead of subjective evaluation experiments. The latter are considered to reveal ground truth scores of visual quality; yet, they are costly and cumbersome, as well as limited in terms of large scale realization and ad-hoc implementation.

The development of algorithms to accurately predict the level of distortion introduced in content representations under realistic types of degradations (e.g., compression, noise) has been at the center of attention for the research community for many years. The initial focus was naturally drawn by conventional images, where it was early understood that naive implementations of error quantification in a pixel-by-pixel basis, e.,g., Mean Square Error (MSE), did not correlate well with human judgements. As a consequence, efforts were concentrated on approaches that consider characteristics of the human visual system. These, in principle, can be categorized as *bottom-up*, and *top-down*. The former denote theoretical approaches that aim at measuring perceived errors in a content, whereas the latter signify engineering solutions that aim at capturing properties of human visual perception. Objective quality metrics can also be clustered based on the availability of the original version of the content at run-time as *full-reference*, *reduced-reference* and *no-reference* metrics.

In the field of three-dimensional imaging, *top-down*, *full-reference* approaches are the most common. Although largely explored in the case of polygonal meshes, objective quality evaluation for point clouds still remains a widely open problem. This type of media content has recently drawn a significant amount of interest due to the flexibility and efficiency in acquisition, processing and rendering. In fact, the MPEG standardization body is releasing its first point cloud coding standard, which will facilitate compatibility across devices, while the relevant efforts have triggered research on the development of more efficient encoding solutions. However, there is only a limited number of available objective quality assessment methods so far with reported weaknesses [1, 2], which urges the demand for better-performing algorithms.

In this study we aim to shed light on the effectiveness of a wide range of features that we construct from explicit and/or implicit information that is carried in a point cloud model. The adopted features are estimators of statistical dispersion in local neighborhoods, computed using several related formulas. As part of the study, the impact of the neighborhood size on the obtained quality scores is analysed, based on standardized benchmarking indexes. Moreover, the impact of a

voxelization step prior to the computation of features is investigated, obtaining promising results, which bring us to introduce the concept of "multi-scale" approaches in point clouds. The performance evaluation is conducted using two available datasets with diverse characteristics. This study can be regarded as an exploration of the applicability of the Structural Similarity (SSIM) index [3] in a higher dimensional, irregular space (volumetric content), incorporating not only color, but also topological coherence among local regions.

2. RELATED WORK

Polygonal mesh representation has been the prevailing 3D modality in computer graphics. Thus, a substantial amount of related work has preceded and is reported in the literature. In particular, existing objective quality metrics for meshes can be divided in two categories, namely, "image-based" and "model-based" [4]. The first essentially exploits the development of high-performing solutions from 2D imaging, which are applied on a set of representative views of the model. The second category relies on geometric errors, dihedral angles [5], curvature statistics [6], and roughness measurements [7], among others. The reader can refer to [8] for an excellent review study on this topic.

Point cloud objective quality metrics can also be clustered as "image-based" and "model-based" approaches, similarly to mesh modelling counterpart. The idea of converting point clouds to meshes prior to application of relevant algorithms was discarded quickly, as this additional processing step is commonly lossy. Similarly to mesh models case, "image-based" metrics rely on 2D imaging algorithms applied on model views [9, 10]. Although able to capture geometry, color and rendering distortions, this type of metrics are governed by the selection of viewpoints (i.e., camera position and distance), the rendering mechanism to consume the content [1], as well as environmental and lighting conditions set on the virtual scene, which affect the perception of colors.

The most common "model-based" approaches so far assess geometry and depend on Euclidean distances (point-to-point), or projected errors along normal vectors (point-to-plane) [11]. Algorithms based on similarity of local surface approximations (plane-to-plane) [12] and curvature statistics (PC-MSDM) [13] have been introduced, showing advances in predicting the perceived distortions, especially in colorless point cloud datasets. In [14], measurements of color errors based on MSE and PSNR, applied either on the RGB or the YCbCr color space, have been proposed. More recently, the first attempts to combine geometry and color features have been reported in [15, 16]. In [15], the evaluation of geometry distortions is based on curvature statistics, whereas several color features based on lightness, chroma and hue are examined, showing that lightness comparison and structure perform better. In [16], color statistics such as histograms are used to predict the texture distortion of point cloud contents,
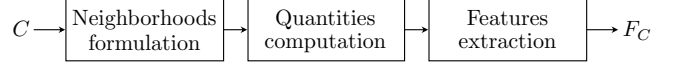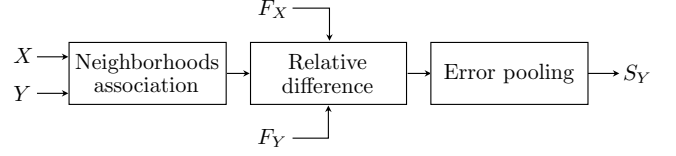

Fig. 1: Feature extraction.

Fig. 2: Structural similarity score computation.

and a linear combination of geometry and color measures is proposed to obtain a global indicator of distortion.

In this work, we extend previous efforts by generalizing feature extraction process, while also exploring a higher dimensional feature space to obtain visual quality scores; that is, we include information from additional point cloud attributes, such as geometry and normal vectors, and evaluate their efficiency. Moreover, we propose voxelization of models to improve prediction accuracy.

3. POINT CLOUD STRUCTURAL SIMILARITY

To measure structural similarity, we construct features that quantify statistical dispersion of quantities that characterize local topology and appearance of a point cloud. To capture local properties, neighborhoods are formed around every point of a model. Quantities to reflect local properties are computed from point cloud attributes, which are either present (e.g., geometry), or can be estimated in case of absence (e.g., normal vectors). In this study, four types of attributes are explored: (i) geometry, (ii) normal vectors, (iii), curvature values, and (iv) colors; yet, the same working principle can be easily extended to any other attribute.

The geometry-related quantities, in our case, are computed based on Euclidean distances between a point of focus and each point belonging to its neighborhood, and are employed to assess the coherence of the local geometric structure. The normal-related quantities are obtained by computing the angular similarity between the normal vector of a particular point and each neighbor, in order to examine the uniformity of the shape of the local surface. For the same purpose, the curvature values are used. Finally, the color-related quantities consist of luminance values that are employed to estimate the local contrast, similarly to SSIM [3].

The features are extracted per neighborhood after applying dispersion statistics on the aforementioned quantities. Statistical dispersion measurements are often utilized to estimate scale parameters; that is, population parameters that indicate the spread of a distribution. Several such estimators exist. Yet, in this study, the following are adopted: median (m_A), variance (σ_A^2), mean absolute deviation (μAD_A), median absolute deviation (mAD_A), coefficient of variation

2

(COV$_A$), and quartile coefficient of dispersion (QCD$_A$), using Equations 1-4 for the last four metrics, respectively

$$\mu\text{AD}_A = E(A - \mu_A) \tag{1}$$

$$m\text{AD}_A = E(A - m_A) \tag{2}$$

$$\text{COV}_A = \frac{\sigma_A}{\mu_A} \tag{3}$$

$$\text{QCD}_A = \frac{Q_A(3) - Q_A(1)}{Q_A(3) + Q_A(1)} \tag{4}$$

where μ_A indicates the mean, σ_A the standard deviation, and $Q_A(i)$ denotes the i-th quartile of a data set A.

A schematic diagram of the feature extraction process, as explained above, is illustrated in Figure 1, with C denoting an input point cloud and F_C indicating the output features. Note that each combination of dispersion estimator and attribute quantity leads to a different F_C. For simplicity reasons, the reference to features henceforth implies a particular combination of these parameters, without limiting the generality.

The perceptual quality prediction is based on the similarity of feature values that are extracted from a reference X and a point cloud under evaluation Y, as presented in Figure 2. For this purpose, each neighborhood of Y is associated with a neighborhood of X, by identifying for every point p of Y its nearest point q in X. Then, the similarity is measured as the relative difference between the corresponding feature values, using Equation 5

$$S_Y(p) = \frac{\mid F_X(q) - F_Y(p) \mid}{\max\{\mid F_X(q) \mid, \mid F_Y(p) \mid\} + \varepsilon} \tag{5}$$

with ε expressing an arbitrarily small number to avoid undefined operations; here, we set ε equal to the machine rounding error for floating point numbers. A total similarity score S_Y for the model under evaluation Y is estimated through error pooling across all points N_p, based on Equation 6

$$S_Y = \frac{1}{N_p} \sum_{p=1}^{N_p} S_Y(p)^k \tag{6}$$

with $k = \{1, 2\}$, denoting the mean and MSE, respectively.[1]

4. VALIDATION METHODOLOGY

4.1. Datasets

Two datasets of static voxelized point clouds under compression artifacts from state-of-the-art codecs are used [1, 2] for performance evaluation. Hereafter, we refer to them as MPCQA and IRPC, accordingly. The first dataset consists of 8 point clouds whose geometry and color were compressed
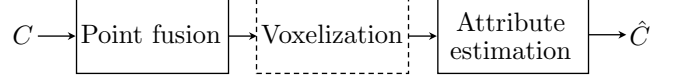
---

[1] A prototype MATLAB implementation is made available online at https://www.epfl.ch/labs/mmspg/pointssim/.

Fig. 3: Model pre-processing. With dashed lines the step introduced for measurements at different scales.

at 6 levels using five codecs, namely, V-PCC and the four G-PCC test model variations. The point clouds were evaluated in an interactive platform side-by-side, using point-based rendering with adaptive point size.

The second dataset consists of 6 point clouds whose geometry was compressed using three codecs, namely, V-PCC, G-PCC (TriSoup module) and PCL, at three degradation levels. The point clouds were evaluated in three different sessions, with one of them being relevant to this study; that is a fixed-size point-based rendering with color information obtained from the original models after a re-coloring step. The point clouds were evaluated in a passive scenario, using video sequences of both the reference and the distorted models, which were shown one after the other.

Both datasets consist of point clouds with diverse characteristics, resulting from the different nature of the represented models and the acquisition technologies that were employed. Moreover, the wide span of encoding schemes that were used lead to different types of artifacts, making them representative and suitable candidates for benchmarking purposes.

4.2. Structural similarity computation

Prior to feature extraction, a pre-processing methodology outlined in Figure 3 is proposed and followed in this study. In particular, point fusion is recommended in order to enable identification of duplicated coordinates in a point cloud. In our implementation, the redundant locations are discarded and the corresponding color values are blended. This step prevents enlisting the same location in neighborhood formulation more than once (Figure 1), and eliminates unnecessary correspondences in neighborhood association (Figure 2).

High-quality surface approximations are essential to benefit from curvature- and normal-based features. To estimate relevant attributes, we adopt the algorithm described in [15] for quadric surface fitting. In this context, the k-nearest neighbors of each point are initially identified ($k = 12$ in our case). A Principal Component Analysis (PCA) is then issued to provide an orthonormal basis and a linear approximation of the local surface, which passes from the centroid of the neighborhood. A least-squares error quadratic fitting function is computed across the normal of the plane, after transferring the origin of the orthonormal basis from the centroid to the transformed point of focus. The normal vector in this new coordinate system is obtained by simply computing the gradient of the locally fitted quadric surface at that point. Then, the inverse transform brings the estimated normal vector back to the original coordinate system. Moreover, the mean curvature
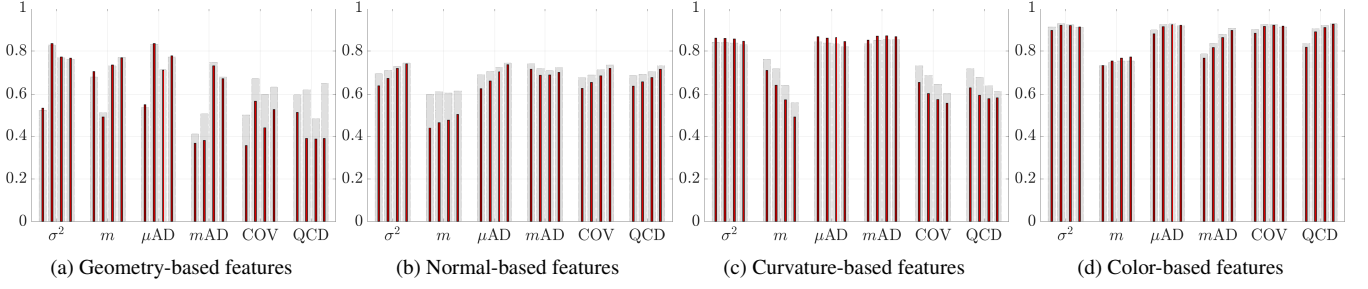
Fig. 4: MPCQA dataset. PCC (thick bars) and SROCC (thin bars) are grouped per metric. In every group, the neighborhood size is $6, 12, 24$ and 48, from left to right.
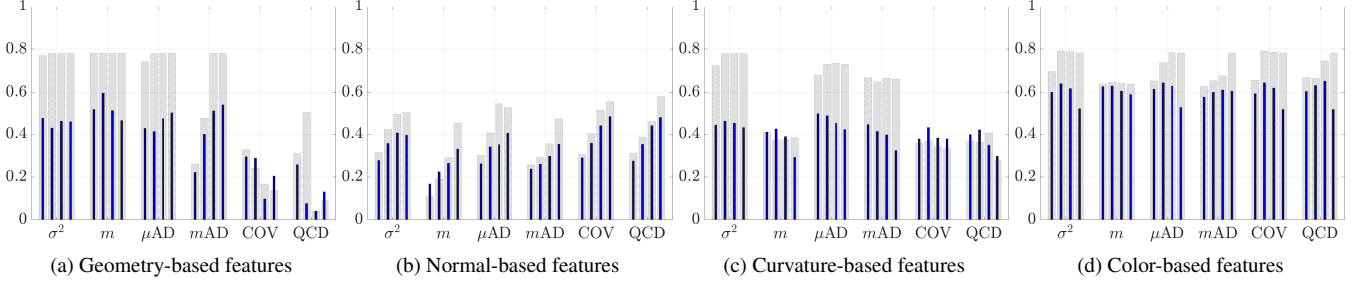

Fig. 5: IRPC dataset, *rcolor* session. PCC (thick bars) and SROCC (thin bars) are grouped per metric. In every group, the neighborhood size is $6, 12, 24$ and 48, from left to right.

value at the point of focus is computed from the coefficients of the fitted quadric surface, as described in [15].

The feature extraction process described in Section 3, is subsequently performed. In regard to the neighborhood formulation (Figure 1), two approaches are common, namely, the k-nearest neighbor, and the range search. In the first case, the set is extended until the specified number of points is reached, whereas in the second case, the set consists of points whose distance is smaller than the specified range. Thus, in the former case, the range of the set is adaptive in terms of size and the number of points is fixed, whereas in the latter case the range is fixed and the number of points can vary. In our implementation, we follow the k-nearest neighbor in order to fixate the population of quantities (distributions) over which the features are computed. To examine the impact of the neighborhood size in the prediction accuracy, in our analysis k takes values from the set $\{6, 12, 24, 48\}$.

A structural similarity score for a model under evaluation is computed based on relative differences of feature values using Equation 6, after following the methodology described in Section 3, and illustrated in Figure 2. In our analysis, the same procedure is repeated using both the original and the distorted models as a reference, resulting in quality scores that are referred to as *asymmetric-original* and *asymmetric-distorted*, respectively. The *symmetric* error is also computed as the minimum out of the two *asymmetric* scores, leading to a total of 144 quality predictors per neighborhood size.

Finally, we explore the prediction potentials of structural similarity measurements obtained from scaled point clouds.

For this purpose, a voxelization step is introduced in our pre-processing pipeline, as depicted in Figure 3. Voxelization is realized by quantizing the coordinates of a model and by color blending between points that fall in the same voxel. The resolution of the voxel grid is defined by a target voxel depth. In our implementation, no clipping is applied on coordinates lying outside of the grid, to avoid introducing extra loss.

4.3. Benchmarking

To evaluate how well an objective metric is able to estimate perceptual quality, Mean Opinion Scores (MOS) computed from ratings of subjects that participate in an experiment are required and serve as ground truth. The objective quality scores are typically benchmarked after application of a regression model. In our case, the logistic function is used following the Recommendation ITU-T P.1401 [17]. The Pearson linear correlation coefficient (PCC), the Spearman rank order correlation coefficient (SROCC), and the Root-Mean-Square Error (RMSE) are computed to conclude on the linearity, monotonicity, and accuracy of objective predictors, respectively.

5. EXPERIMENTAL RESULTS

In Figures 4 and 5, the benchmarking results are provided for both testing datasets. Each figure indicates all the dispersion estimators (i.e., metrics) that were employed, per attribute. In the provided plots, the thick bars correspond to the PCC index, while the thin bars indicate the SROCC. They

are grouped per metric, indicated on the x-axis, and in each group, the four selected neighborhoods are displayed in an increasing order. The reported results correspond to predictions based on *asymmetric-distortion* scores, which were found to perform slightly better with respect to the alternatives.

In Figure 4, the performance of the metrics is presented for the MPCQA dataset. It is evident that color-based features are over-performing, achieving high scores, with the best being a PCC of 0.928 and SROCC of 0.920, for σ^2 and $k = 12$. We observe that median is the worst-performing metric, although still achieving good results. In principle, the neighborhood size is not critical, albeit better performance is obtained for the majority of the metrics in mid-ranges (i.e., k equal to 12 or 24). The curvature-based features are the second best-performing solutions in this dataset. For the dispersion estimators that work better, namely, σ^2, μAD and mAD, the number of neighbors k is also not crucial. Regarding geometry-based features, they are rather unstable with respect to the local region size, while the majority of normal-based features tend to improve as the neighborhoods are expanding.

In Figure 5, similar plots are employed to present the performance of the metrics in the IRPC dataset. The color-based features are found again to be the most accurate predictors. However, their performance is notably deteriorated with respect to the MPCQA dataset, achieving a maximum of 0.792 PCC and 0.643 SROCC using COV with $k = 12$. This performance decrease can be explained by the fact that the color is not directly degraded in this dataset. Nonetheless, distortions are inherently added from point re-positioning and downsampling due to geometry encoding. The second best option is given by the geometry-based features, in regard to the PCC index. However, the low SROCC values indicate that the predictions are not very reliable. The majority of features that capture uniformity of surface shape perform very poorly, with the exception of some metrics, namely, σ^2, μAD and mAD, applied on curvature values. The general poor performance can be explained by the fact that (a) the dataset consists of several rather noisy point clouds, and (b) the original color values used for the decompressed models act as distractors. Our results are in accordance with what is seen in [15].

5.1. Towards multi-scale structural similarity

The performance of structural similarity measurements is finally investigated after point cloud scaling. The latter is implemented through voxelization, which enables color smoothing and regular down-sampling of geometry, only for models whose original geometric resolution is higher than the target voxel depth.[2] Moreover, color blur is introduced, reducing blocking artifacts, simulating visual inspection from far distances. Note that if the original voxel depth is smaller than the target, the color distribution remains unaltered, while the

---

[2] This essentially applies when the *effective* voxel depth of a model is larger than the target; the *effective* implies no prior up-scaling.

(a) MPCQA dataset with $k = 12$ (b) IRPC dataset with $k = 24$
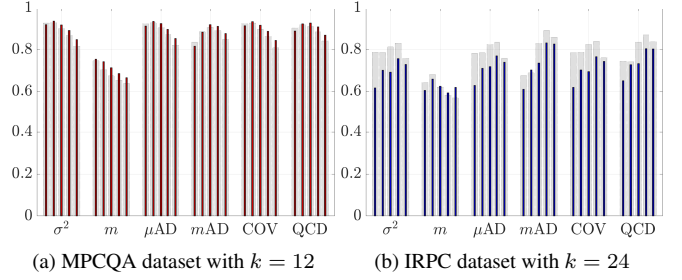
Fig. 6: Performance of color-based features after voxelization. PCC (thick bars) and SROCC (thin bars) are grouped per metric. In every group, the corner left bar corresponds to no voxelization, whereas the rest of the bars correspond to voxel depths obtained after decreasing the lowest reference resolution of the dataset by $0, 1, 2$ and 3, from left to right.

topology is up-scaled without impacting the number of points.

In Figure 6a, the performance of color-based features is presented in the MPCQA dataset, after voxelization using bit-depths equal and below the lowest resolution among original models (i.e., this dataset consists of 9-bit and 10-bit models). The best results are obtained with voxel depth equal to 9, with a maximum PCC of 0.929 and SROCC of 0.936 using σ^2 with $k = 12$. It can be observed that as the target voxel depth is further decreasing, the performance of the metrics is deteriorating. This can be explained by the increasing levels of blurring artifacts that appear at lower bit-depths, which are enhanced for models with severe color compression distortions. For this demonstration, k was chosen equal to 12, while very similar trends are obtained for other neighborhood sizes.

The benefits of model voxelization are more evident in our next paradigm. In Figure 6b, the performance of color-based features is demonstrated for the IRPC dataset, following the same rationale (i.e., this dataset consists of 10-bit and 12-bit models). Based on our results, a remarkable performance increase is observed, with a maximum of 0.893 for PCC and 0.832 for SROCC at 8-bit voxel depth using mAD with $k = 24$. Similar tendencies are noted for other k values. This outcome is explained by the fact that geometry degradations that are exhibited in blocks, as well as the resolution of the blocks, over which color blending is performed, affect the output color values. In other terms, through voxelization, geometry degradations are reflected on the output color values.

In general, voxelization can be seen as a way to reduce cross-content resolution, potentially providing a more suitable scale for objective quality predictors. However, the identification of appropriate voxel depths for point cloud models or datasets, is not explored in this study.

In Table 1 the best-performing metrics reported in the literature and attained in this study, are summarized. For the latter, results including and excluding model voxelization are handled separately, and are presented following the notation: (attribute, metric, neighborhood, voxel depth).

Table 1: Best performing metrics per dataset.

Dataset	Metric	PCC	SROCC
MPCQA	point-to-plane with MSE [1]	0.858	0.884
MPCQA	(color, σ^2, $k = 12$, original)	0.928	0.920
MPCQA	(color, σ^2, $k = 12$, 9-bit)	0.929	0.936
IRPC	PCQM [15]	0.873	0.807
IRPC	(color, COV, $k = 12$, original)	0.792	0.643
IRPC	(color, mAD, $k = 24$, 8-bit)	0.893	0.832

6. CONCLUSIONS

In this paper we propose and evaluate a wide set of features based on estimators of a distribution's dispersion. The features are extracted from computed quantities that depend on point cloud attributes, including geometry, normal vectors, curvature values, and colors. A pre-processing methodology for the models under evaluation is devised and applied to ensure high-quality measurements. Structural similarity scores are computed based on relative difference of feature values and reflect distortions, as expressed in the corresponding attributes' information. Our results show that color-based alternatives achieve the best performance in predicting perceptual quality, in both datasets that were examined. Moreover, we propose the use of voxelization prior to feature extraction, in order to mitigate objective quality scores that are achieved at high-resolution models. This extra processing step was found effective in case of colored point clouds whose geometry is only compressed, leading to more accurate visual quality predictions. This study aims to report the results of an extensive performance analysis of features that measure local topological and color consistencies. Future work will focus on the impact of the neighborhood formulation algorithm and the neighborhood sizes, on extending the work to non-voxelized point clouds, and to combine the attributes and scales presented above in a single metric.

7. REFERENCES

[1] 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 Transactions on Signal and Information Processing*, vol. 8, pp. e27, Nov 2019.

[2] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality," 2019.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[4] G. Lavoué, M. C. Larabi, and L. Váša, "On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 8, pp. 1987–1999, 2016.

[5] L. Váša and J. Rus, "Dihedral Angle Mesh Error: A Fast Perception Correlated Distortion Measure for Fixed Connectivity Triangle Meshes," *Comput. Graph. Forum*, vol. 31, no. 5, pp. 1715–1724, Aug 2012.

[6] G. Lavoué, "A Multiscale Metric for 3D Mesh Visual Quality Assessment," *Computer Graphics Forum*, vol. 30, pp. 1427–1437, Jul 2011.

[7] M. Corsini, E. D. Gelasca, T. Ebrahimi, and M. Barni, "Watermarked 3-D Mesh Quality Assessment," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 247–256, 2007.

[8] G. Lavoué and R. Mantiuk, *Quality Assessment in Computer Graphics*, pp. 243–286, Springer International Publishing, Cham, 2015.

[9] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*. 2018, vol. 10752, pp. 174 – 190, SPIE.

[10] E. Alexiou and T. Ebrahimi, "Exploiting user interactivity in quality assessment of point cloud imaging," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–6.

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

[12] E. Alexiou and T. Ebrahimi, "Point Cloud Quality Assessment Metric Based on Angular Similarity," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018, pp. 1–6.

[13] G. Meynet, J. Digne, and G. Lavoué, "PC-MSDM: A quality metric for 3D point clouds," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–3.

[14] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Updates and Integration of Evaluation Metric Software for PCC," ISO/IEC JTC1/SC29/WG11 Doc. MPEG2017/M40522, Apr 2017.

[15] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: A Full-Reference Quality Metric for Colored 3D Point Clouds," in *12th International Conference on Quality of Multimedia Experience (QoMEX 2020)*, 2020.

[16] I. Viola, S. Subramanyam, and P. César, "A color-based objective quality metric for point cloud contents," in *12th International Conference on Quality of Multimedia Experience (QoMEX 2020)*, 2020.

[17] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul 2012.