

PointPCA: Point Cloud Objective Quality Assessment Using PCA-Based Descriptors

Evangelos Alexiou, Irene Viola, *Member, IEEE* and Pablo Cesar, *Senior Member, IEEE*

Abstract—With the increasing popularity of extended reality technology and the adoption of depth-enhanced visual data in information exchange and telecommunication systems, point clouds have emerged as a promising 3D imaging modality. Similarly to other types of content representations, visual quality predictors for point cloud data are vital for a wide range of applications, enabling perceptually optimized solutions from acquisition to rendering. Recent standardization activities on point cloud compression have urged the need for objective quality evaluation methods, driving the research community to the development of relevant algorithms. In this work, we complement existing approaches by proposing a new quality metric that compares local shape and appearance measurements between a reference and a distorted point cloud. To this aim, a large set of geometric and textural descriptors is defined, and the prediction accuracy of corresponding statistical features is evaluated in the context of quality assessment. Different combination strategies are examined, providing insights regarding the effectiveness of different metric designs. The performance of the proposed method is validated against subjectively-annotated datasets, showing better performance against state-of-the-art solutions in the majority of cases. A software implementation of the metric is made available here: <https://github.com/cwi-dis/pointpca>.

Index Terms—point cloud, quality metric, objective quality assessment, PCA

I. INTRODUCTION

POINT clouds, nowadays, denote a prominent solution for the representation of 3-D photo-realistic content in immersive applications. This possibility was raised by advancements in depth sensing and graphics processing technologies the past years, and more recently fuelled by relevant activities of the JPEG [1] and MPEG [2] standardization bodies. As a result of these efforts, MPEG has crafted two standards, namely, Geometry-based Point Cloud Compression (G-PCC) [3], and Video-based Point Cloud Compression (V-PCC) [4] which are tailored for static and dynamic point cloud contents respectively, while JPEG has recently issued a Call for Evidence within the JPEG Pleno framework [5]. Released standards are expected to establish interoperability and facilitate the integration of point cloud technology in daily use-cases, unravelling new possibilities for remote communication and extended reality (XR) experiences.

Point clouds offer advantages in terms of ease of acquisition, accurate modelling, and photo-realistic rendering. However, for faithful representation of 3D visual information, a vast amount of data is required. Thus, lossy compression schemes are commonly employed, which achieve larger data size reductions in comparison to lossless counterparts, at the expense of

perceptual distortions. Point clouds might also undergo signal deformations during processing, transmission, and/or rendering, which have a negative impact on the fidelity of the original content. Therefore, there is need for mechanisms to quantify the induced visual impairments. To this aim, subjective and objective quality assessment methodologies are essential. Subjective quality evaluations rely on human opinions, and provide ground truth ratings of visual quality for distorted stimuli. Although accurate, these approaches are time and cost expensive. Objective quality metrics refer to algorithms that computationally predict the visual quality of distorted stimuli. They are easily operated, yet, their performance depends on their ability to assess distortions in perceptual terms. The prediction accuracy of objective quality metrics is validated through benchmarking and, specifically, after comparison with subjective ground truth.

Objective quality metrics can be classified based on their requirement for the original content (i.e., reference) at execution time as full-reference, reduced-reference, and no-reference. In the first class, the presence of the reference content is necessary; in the second class some related data are needed as input, whereas, in the third class no reference information is required. Point cloud metrics can additionally be categorized as projection-based and point-based. The former refer to 2D solutions that are applied on projected views, capturing geometric and textural point cloud distortions as reflected upon rendering on planar arrangements. These predictors commonly adopt or extend techniques that were devised for images in the past; however, they are view- and rendering-dependent [6]. Conversely, point-based quality metrics operate in the 3-D point cloud domain and are rendering-agnostic. Initial attempts built on simple distances between individual points, whereas more recent submissions utilize richer features that capture local patterns of geometric or textural information. The effectiveness of the latest point-based metrics heavily relies on the accuracy of textural predictors [7]–[9], which compute statistics over spatial neighborhoods, thus, carrying contributions from neighbors and capturing geometric distortions in an implicit manner. Yet, there is still need for explicit evaluations of geometric degradations, in order to improve the performance and robustness of quality metrics. Current point-based solutions rely on a very small set of geometrical features, often capturing a specific property of the underlying surfaces, as reflected by the statistical distribution of point locations [10], normal vectors [11]–[13], or curvature values [7], [13].

In this paper, we propose the use of geometric descriptors based on Principal Component Analysis (PCA) to estimate

E. Alexiou and I. Viola are with CWI.
Pablo Cesar is with Delft University and CWI.

structural distortions in point cloud contents. Such descriptors have been used with lidar data for urban classification [14], semantic interpretation [15], semantic segmentation [16], and contour detection [17], while more recently, a subset with well-behaving distributions was employed for no-reference objective quality assessment [18]. In order to better capture local variations in the distribution of the descriptors, we adopt statistical features that can estimate average trends and dispersion in a neighborhood. We complement our geometrical features by employing textural descriptors, based on perceptually-relevant luminance variations. Corresponding quality scores obtained in both domains are linearly combined to a global indication of visual degradation. Our results show the high performance of our proposed features in predicting the perceptual quality of point cloud contents under various impairments, with improvement with respect to state-of-the-art solutions.

II. RELATED WORK

The point-to-point and point-to-plane [11] denote the earliest attempts for the establishment of point-based objective quality metrics. The former measures the Euclidean distance between point coordinates, while the latter relies on the projected error of distorted points across reference normal vectors. In both metrics, the Mean Square Error (MSE) or the Hausdorff distance (HSD) is applied over the individual, per point error values, in order to deliver a global degradation score. In [19], the generalized Hausdorff distance is proposed to mitigate the sensitivity of the Hausdorff distance in outlying points, by excluding a percentage of the largest individual errors. The geometric Peak-Signal-to-Noise-Ratio (PSNR), defined in [20] for both metrics to account for differently scaled contents, was also revised in [21]. Specifically, the voxel grid's diagonal originally used as the peak value, was then replaced by the average over distances between neighbors in 3D space or after projection onto local planes, to represent the content's intrinsic or rendering resolution, respectively. The plane-to-plane metric is described in [12] and estimates the angular similarity of tangent planes, as expressed through unoriented normals. The point-to-distribution metric, introduced in [22], computes the Mahalanobis distance between a distorted point and a reference neighborhood. The PC-MSDM [23] evaluates the similarity of local curvature statistics, extracted after quadric fitting in support regions.

The PC-MSDM was extended to PCQM [7], by incorporating local statistical measurements from luminance, chrominance and hue components to evaluate textural impairments. The PointSSIM [13] relies on statistical dispersion of location, normal, curvature and luminance data. An optional pre-processing step of voxelization is proposed to enable different scaling effects and reduce intrinsic geometric resolution differences across contents. The VQA-CPC [24] computes statistics upon Euclidean distances between every sample and the arithmetic mean of the point cloud, using geometric coordinates and color values. An extension is presented in [25], involving a point cloud partition stage, before extraction of features per region. The geometric features consist of statistical moments

applied on Euclidean distances, angular distortions and local densities, which are weighted according to the roughness of a region. The textural features rely on the same statistics, after conversion to the HSV color space. The above algorithms follow a similar operating principle to the SSIM [26], in terms of identifying structural features that reflect local changes in either geometric, or textural information.

In [10], the point-to-distribution metric was extended to capture color degradations, by applying the same formula on the luminance, and averaging the obtained error values. A graph signal processing-based approach is described in [9], evaluating statistical moments of color gradients on keypoints of the reference content, which are identified after high-pass filtering on its topology. In [27], local binary patterns (LBP) on the luminance channel are applied in local neighborhoods. This work is extended in [28] considering the point-to-plane distance between point clouds, and the point-to-point distance between feature maps. A variant descriptor called local luminance patterns (LLP) is proposed in [29], introducing a voxelization stage. In [30], a texture descriptor to compare neighboring color values using the CIEDE2000 distance, is proposed. The color differences are coded as bit-based labels, which denote frequency values of pre-defined intervals. An extension is presented in [31], namely, BitDance, which incorporates bit-based labels from a geometric descriptor that relies on comparison of neighbouring normal vectors.

The point-to-point metric has been adapted to measure the MSE or the PSNR for color-only degradations [32]. This effectively simulates corresponding 2D algorithms that have been extensively used with color images. More sophisticated paradigms of texture-only metrics have been proposed in [33], which compute histograms or correlograms, of luminance and chrominance, in order to characterize color distributions.

The first reduced-reference objective quality metric is reported in [34], and relies on global features that are extracted from location, color and normal data. More recently, a reduced reference metric for point clouds encoded with V-PCC is presented in [35]. It is based on a linear model of geometry and color quantization parameters, with the model's parameters determined by a local and a global color fluctuation feature.

Regarding projection-based approaches, the prediction accuracy of 2D quality metrics was initially examined in [36], over images obtained after projecting point clouds on the six faces of a surrounding cube. The influence of the number of viewpoints in denser camera arrangements and the exclusion of background pixels is explored in [37], with a proposed weighting scheme based on user interactivity. In [38], a weighted combination of global and local features extracted from texture and depth images, is defined. The Jensen-Shannon divergence on the luminance channel serves as the global feature, whereas a depth-edge map, a texture similarity map, and an estimated content complexity factor, account for the local features. In [39], color and curvature values are projected on planar surfaces. Color impairments are evaluated using probabilities of local intensity differences, together with statistics of their residual intensities, and similarity values between chromatic components. Geometric distortions are assessed based on statistics of curvature residuals.

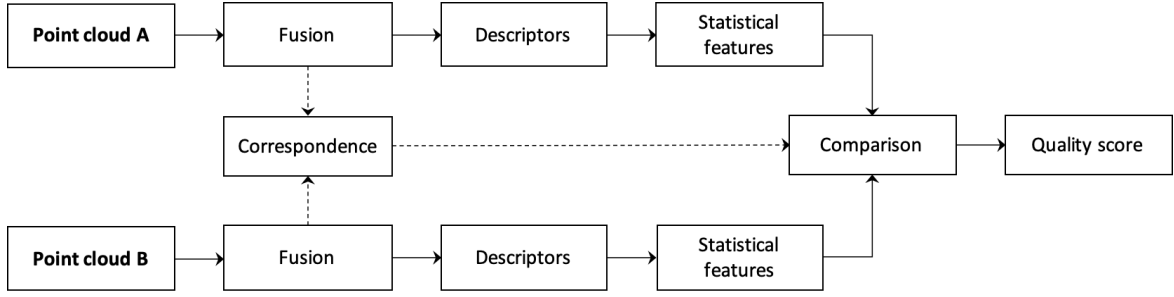


Fig. 1. PointPCA architecture. Both the original and the distorted point clouds are passing from every processing stage to compute a quality score.

A hybrid approach that makes use of both projection- and point-based algorithms, is proposed in [40]. The point clouds are divided into non-overlapping partitions, called layers, with a planarization process taking place at each layer, before applying the IW-SSIM [41] to assess geometric distortions. Color impairments are evaluated using RGB-based variants of similarity measurements defined in [7].

III. DESCRIPTION OF POINTPCA

The architecture of the proposed metric can be decomposed in five stages, namely, (a) fusion, (b) correspondence, (c) descriptors, (d) statistical features, and (e) comparison. An illustrative system diagram is provided in Figure 1. The metric requires the reference content during execution in order to provide a quality prediction for a distorted stimulus. Thus, a correspondence between the reference and distorted point cloud is computed, after a fusion step which removes duplicated points. Then, geometric and textural descriptors are computed per point, for both stimuli. For every descriptor, local relations are captured through statistics, which are estimated over support regions defined around points. This process is repeated for every descriptor, leading to a set of so-called statistical features, which are obtained per point. The statistical features extracted from both the reference and the distorted point clouds are eventually compared, given the correspondence function, and error maps are obtained. For each map, the error samples are pooled together and a quality prediction is obtained, per statistical feature. Finally, a total quality score for a distorted stimulus is computed as a linear combination of individual quality predictions. Below, we detail each of the aforementioned stages separately.

A. Fusion

In this step, duplicated coordinates in a point cloud are removed and corresponding color values are averaged. Thus, points with unique locations are employed during neighborhood formulation, while redundant correspondences between the reference and the model under evaluation, are avoided.

B. Correspondence

Identifying matches between two sets of points, is generally an ill-posed problem. To favor lower complexity, we make use of the nearest neighbor algorithm for identification of correspondences between two point clouds, similarly to the

TABLE I
DEFINITION OF GEOMETRIC DESCRIPTORS.

Descriptor	Definition
Eigenvalues	$d_v^g = \lambda_v$, with $v \in \{1, 2, 3\}$
Sum of eigenvalues	$d_4^g = \sum_v \lambda_v$
Linearity	$d_5^g = (\lambda_1 - \lambda_2)/\lambda_1$
Planarity	$d_6^g = (\lambda_2 - \lambda_3)/\lambda_1$
Sphericity	$d_7^g = \lambda_3/\lambda_1$
Anisotropy	$d_8^g = (\lambda_1 - \lambda_3)/\lambda_1$
Omnivariance	$d_9^g = \sqrt[3]{\lambda_1 \cdot \lambda_2 \cdot \lambda_3}$
Eigenentropy	$d_{10}^g = -\sum_v \lambda_v \cdot \ln(\lambda_v)$
Surface variation	$d_{11}^g = \lambda_3 / \sum_v \lambda_v$
Roughness	$d_{12}^g = (\mathbf{p}_i - \bar{\mathbf{p}}_i) \cdot \mathbf{e}_3 $
Parallelity _x	$d_{13}^g = 1 - \mathbf{u}_x \cdot \mathbf{e}_3 $
Parallelity _y	$d_{14}^g = 1 - \mathbf{u}_y \cdot \mathbf{e}_3 $
Parallelity _z	$d_{15}^g = 1 - \mathbf{u}_z \cdot \mathbf{e}_3 $

majority of existing solutions. For this purpose, one model is selected as the reference and the other is set under evaluation. In particular, for every point \mathbf{b}_i that belongs to a point cloud under evaluation \mathcal{B} (i.e., $\mathbf{b}_i \in \mathcal{B}$), a matching point $\mathbf{a}_i \in \mathcal{A}$ is identified as its nearest neighbor in terms of Euclidean distance, and is registered as its correspondence. Formally, for a point cloud under evaluation, a correspondence function is defined as $c^{\mathcal{B}, \mathcal{A}} : \mathcal{B} \rightarrow \mathcal{A}$ with $c^{\mathcal{B}, \mathcal{A}}(\mathbf{b}_i) = \mathbf{a}_i$. Note that, different sets of matching points are obtained when iterating over the points of \mathcal{B} to identify nearest neighbors in \mathcal{A} , with respect to starting from \mathcal{A} to find matches in \mathcal{B} ; that is, when setting \mathcal{A} , or \mathcal{B} as reference, respectively. In our case, we set both the pristine and the impaired models as reference, and we use a max operation in order to obtain a final prediction independent of the reference selection. This is commonly referred in the literature as symmetric error [12], [19].

C. Descriptors

A set of geometric and textural descriptors are defined per point, in order to reflect local properties of a point cloud with respect to its topology and appearance, respectively.

Geometric descriptors: We propose the use of low-level shape descriptors that are extracted from quantites obtained after PCA. Provided a query point $\mathbf{p}_i = (x_i, y_i, z_i)^T$, we identify a surrounding support region that belongs to the same

point cloud, forming a set of points $\mathbf{p}_n \in \mathbf{P}_i$. The covariance matrix Σ_i of this set is computed, as shown in Equation 1,

$$\Sigma_i = \frac{1}{|\mathbf{P}_i|} \sum_{n=1}^{|\mathbf{P}_i|} (\mathbf{p}_n - \bar{\mathbf{p}}_i)(\mathbf{p}_n - \bar{\mathbf{p}}_i)^T \quad (1)$$

with $|\mathbf{P}_i|$ indicating the cardinality, and $\bar{\mathbf{p}}_i$ the centroid of \mathbf{P}_i , which is given in Equation 2.

$$\bar{\mathbf{p}}_i = \frac{1}{|\mathbf{P}_i|} \sum_{n=1}^{|\mathbf{P}_i|} \mathbf{p}_n \quad (2)$$

Since the covariance matrix, also known as the 3D structure tensor, represents a symmetric positive-definite matrix, its eigenvalues exist, are non-negative and correspond to an orthogonal system of eigenvectors. The latter, denote directions across which the data are mostly dispersed, whereas the former reflect the variance of the transformed data across the principal axis. Eigenvalues and eigenvectors are employed in measurements that estimate interpretable local 3D shape properties.

Let us assume that \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 denote the eigenvectors corresponding to the eigenvalues λ_1 , λ_2 and λ_3 , with $\lambda_1 > \lambda_2 > \lambda_3$. Moreover, let us define $\mathbf{u}_x = (1, 0, 0)^T$, $\mathbf{u}_y = (0, 1, 0)^T$ and $\mathbf{u}_z = (0, 0, 1)^T$ to depict unit vectors across the x , y and z axis, respectively. In Table I, we present the definitions of the geometric descriptors, $\mathbf{d}^g \in \mathbb{R}^{1 \times 15}$. Intuitively, $d_1^g - d_4^g$ denote the individual (i.e., $d_1^g - d_3^g$) and the aggregated sum (i.e., d_4^g) of dispersion magnitudes for the points distribution across principal axis. Descriptors $d_5^g - d_7^g$ reveal behaviors of a neighborhood's point arrangement, capturing the dimensionality of the local surface. Another measurement indicating the shape of the local point distribution, is given by d_8^g and focuses on the data variation across the 1st and the 3rd principal directions. Descriptors d_9^g , d_{10}^g and d_{11}^g provide an estimate of spread, uncertainty and variation of the underlying surface, respectively, considering all principal axis. Descriptor d_{12}^g depends on the third eigenvector, \mathbf{e}_3 , and quantifies the projected error of the queried point from the centroid of the set, across the linearly approximated local surface. Finally, descriptors $d_{13}^g - d_{15}^g$ measure the projected error of the estimated normal vector (i.e., \mathbf{e}_3) across unit vectors parallel to the coordinate system axes where a point cloud is laying. Note that $d_5^g - d_{10}^g$ have been proposed in [42], d_{11}^g in [43], and d_{14}^g in [44] named as verticality, in order to measure local geometric properties.

Textural descriptor: We propose luminance-based measurements to capture distortions in the appearance of a point cloud, inspired by their high correlation with human perception proven by previous efforts [7], [8], [10]. Considering that RGB is the most common representation of color attributes, the ITU-R Recommendation BT.709 [45] is employed for conversion to the YCbCr color space, as shown in Equation 3.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.1146 & -0.3854 & 0.5000 \\ 0.5000 & -0.4542 & -0.0468 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (3)$$

After transformation, the obtained luminance of every point serves as our descriptor, with $d_1^t = Y$, and $\mathbf{d}^t \in \mathbb{R}^{1 \times 1}$.

Support regions: A support region around a point sample is required for the computation of geometric descriptors. There are two alternatives widely employed to specify point cloud neighborhoods; that is, the k nearest neighbor and the range search algorithms, hereafter, noted as k -nn and r -search, respectively. The former leads to neighborhoods of arbitrary extent and a fixed population, whereas the latter results in regions that span in the same volume, with a varying number of samples.

The r -search is selected to estimate descriptors. This choice is justified by our requirement to represent properties of the same length of surface areas in both the reference and the distorted stimuli. This behavior is granted by the r -search variant, as opposed to the k -nn algorithm, which is susceptible to different point densities. For example, in the presence of down-sampling, there is no difference between the size of regions identified in the pristine and the impaired point clouds using the r -search. However, when using the k -nn alternative, larger regions are considered in the impaired point cloud. Thus, descriptor values would reflect properties of different size of underlying surfaces and, in turn, such a comparison would introduce unreliable measurements.

D. Statistical features

Statistical functions are applied on geometric and textural descriptor values, which are computed per point, in order to capture inter-point relations that lie in the same neighborhood. In particular, the mean is computed, in order to reflect a smoother estimate of the surface property, accounting for a broader region. The standard deviation is also obtained, in order to quantify the level of variation of a surface property in the surrounding area. Considering a query point \mathbf{p}_i and a set of neighbors $\mathbf{p}_{\hat{n}} \in \hat{\mathbf{P}}_i$, the first statistical feature is computed per Equation 4,

$$\boldsymbol{\mu}_i = \frac{1}{|\hat{\mathbf{P}}_i|} \sum_{\hat{n}=1}^{|\hat{\mathbf{P}}_i|} \mathbf{d}_{\hat{n}} \quad (4)$$

where $\mathbf{d}_{\hat{n}} = [\mathbf{d}_{\hat{n}}^g, \mathbf{d}_{\hat{n}}^t] \in \mathbb{R}^{1 \times 16}$ is a row vector of concatenated geometric and textural descriptors that corresponds to $\mathbf{p}_{\hat{n}}$. The second statistical feature, is then obtained from Equation 5.

$$\boldsymbol{\sigma}_i = \sqrt{\frac{1}{|\hat{\mathbf{P}}_i|} \sum_{\hat{n}=1}^{|\hat{\mathbf{P}}_i|} (\mathbf{d}_{\hat{n}} - \boldsymbol{\mu}_i)^2} \quad (5)$$

A complete statistical features vector is composed of both terms, and is annotated by $\boldsymbol{\phi}_i = [\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i] \in \mathbb{R}^{1 \times 32}$.

Statistical features are able to better capture dependencies within local neighborhoods, and provide measurements that are more perceptually-relevant with respect to single points. Specifically, they are well-aligned with primary characteristics of human visual system, such as low-pass filtering and sensitivity to high-pass frequencies. Applying the mean in local regions mimics the former, whereas the standard deviation provides an estimate of the latter. Moreover, statistical features computed on a point contain contributions from its surroundings, alleviating the negative effects of an erroneous correspondence, or outlying descriptor values. That is, considering that distorted stimuli are characterized by point removal

or displacement with respect to their reference positions, errors might be introduced by the matching algorithm, or individual descriptor values might be poorly estimated. Thus, for instance, comparing the means instead of the individual descriptor values, mitigates the error.

Support regions: The k -nn is selected in order to compute statistical features. We argue that, in this case, the operating principle of this approach is beneficial towards revealing topological deformations. In particular, by appending neighboring samples until we reach k , we effectively extend to larger areas in case of lower point density, and recruit erroneous points in case of re-positioning. Thus, larger differences will be observed when compared to corresponding measurements taken from the pristine content. In simpler terms, using k -nn enables us to penalize point sparsity and displacement.

E. Comparison

Given a correspondence function $c^{\mathcal{B},\mathcal{A}}(\mathbf{b}_i) = \mathbf{a}_i$, the j -th statistical feature of point $\mathbf{b}_i \in \mathcal{B}$, namely $\phi_{i,j}^{\mathcal{B}}$, is compared the j -th statistical feature of point $\mathbf{a}_i \in \mathcal{A}$, namely $\phi_{i,j}^{\mathcal{A}}$, as shown in Equation 6,

$$r_{i,j}^{\mathcal{B},\mathcal{A}} = \frac{|\phi_{i,j}^{\mathcal{A}} - \phi_{i,j}^{\mathcal{B}}|}{\max(|\phi_{i,j}^{\mathcal{A}}|, |\phi_{i,j}^{\mathcal{B}}|) + \varepsilon} \quad (6)$$

where $r_{i,j}^{\mathcal{B},\mathcal{A}}$ indicates the relative difference that corresponds to \mathbf{b}_i , $1 \leq i \leq |\mathcal{B}|$, $1 \leq j \leq 32$, while ε represents a small constant to avoid undefined operations; in this case, we use the machine rounding error for floating point numbers. For every statistical feature j , a corresponding predictor is exported after pooling across all points of \mathcal{B} , as shown in Equation 7.

$$s_j^{\mathcal{B},\mathcal{A}} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} r_{i,j}^{\mathcal{B},\mathcal{A}} \quad (7)$$

The same computations are repeated after setting the point cloud \mathcal{B} as the reference and computing a correspondence function $c^{\mathcal{A},\mathcal{B}}(\mathbf{a}_k) = \mathbf{b}_k$. In an analogous way, for every statistical feature j , a corresponding predictor $s_j^{\mathcal{A},\mathcal{B}}$ is derived, and a symmetric quantity is obtained after a max operation, as shown in Equation 8.

$$s_j = \max(s_j^{\mathcal{B},\mathcal{A}}, s_j^{\mathcal{A},\mathcal{B}}) \quad (8)$$

F. Quality score

Each predictor s_j essentially provides a quality rating based on the j -th statistical feature. A total quality score can be obtained as a linear combination of domain-specific, or across-domain predictors, as per Equation 9

$$q^\alpha = \frac{\sum_{m=1}^{|\mathcal{S}^\alpha|} w_m \cdot s_m^\alpha}{\sum_{m=1}^{|\mathcal{S}^\alpha|} w_m}, \quad (9)$$

in which $\alpha = \{g, t, g \cup t\}$ denotes geometry, texture, and geometry-plus-texture attribute domains respectively, $s_m^\alpha \in \mathcal{S}^\alpha$ indicates a subset of predictors, with $\mathcal{S}^\alpha \subseteq \mathcal{S}$, $\mathcal{S} = \{s_j\}$, and $1 \leq j \leq 32$, while $m \in \{1, \dots, N^\alpha\}$, with N^α the number of selected predictors, and w_m the corresponding weight. Such weights can be either manually set to specific values, e.g.,

giving equal weight to each predictor, or can be learned. In our case, the latter is performed through an optimization problem minimizing the distance between the predicted and the ground truth Mean Opinion Score (MOS):

$$\begin{aligned} \underset{\mathbf{w}}{\operatorname{argmin}} \quad & \left| \text{MOS} - f\left(\sum_{m=1}^{|\mathcal{S}^\alpha|} w_m \cdot s_m^\alpha\right) \right|_2^2 \\ \text{s.t.} \quad & w_m \in [0, 1], \\ & \sum_{m=1}^{|\mathcal{S}^\alpha|} w_m = 1 \end{aligned} \quad (10)$$

with $f(\cdot)$ representing the selected fitting function to map the objective score to the MOS (see Section IV-B).

Additionally, in case of domain-specific quality scores, a final quality score encompassing both domains can be computed using the following linear combination:

$$\tilde{q} = \omega \cdot q^g + (1 - \omega) \cdot q^t \quad (11)$$

with ω a regularization term to drive the contributions of geometry and texture.

IV. BENCHMARKING SETUP

A. Datasets

A total of three subjectively annotated data sets is recruited in order to evaluate the performance of quality metrics under consideration, namely, ICIP2020 (D1) [46], M-PCCD (D2) [6] and SJTU (D3) [38]. D1 contains six colored static point clouds that represent human figures, whose geometry and color is encoded using V-PCC and two G-PCC variants (i.e., Octree-plus-Lifting and TriSoup-plus-Lifting) at 5 degradation levels, for a total of 96 stimuli. D2 consists of eight colored static point clouds illustrating both human figures and inanimate objects, whose geometry and color is encoded using V-PCC and four G-PCC variants (i.e., Octree-plus-Lifting, Octree-plus-RHAT, TriSoup-plus-Lifting and TriSoup-plus-RAHT), resulting in 240 stimuli. Finally, D3 is comprised of nine colored point clouds depicting both human figures and inanimate objects, that are subject to octree-based compression, color noise, geometry Gaussian noise, downscaling, and all super-imposed combinations of the different types of degradation excluding compression, for a sum of 387 stimuli.

B. Performance indexes

To evaluate how well an objective metric is able to estimate perceptual quality, the Mean Opinion Scores (MOS) computed from ratings of subjects that participate in an experiment are required and serve as ground truth. The metrics are typically benchmarked after applying a regression model in order to map the objective scores to the subjective quality range, while also to account for biases, non-linearities and saturations that might appear in subjective testing. In particular, let us define the result of execution of a particular objective metric indicates a Predicted Quality Score (PQS). A predicted MOS, denoted as $P(\text{MOS})$, is estimated by applying a fitting function on the $[\text{PQS}, \text{MOS}]$ data-set. In our analysis, the Recommendation ITU-T J.149 [47] is followed, using the logistic fitting function type II. The PLCC, the SROCC, and the RMSE are computed between the PQS and MOS to conclude on the linearity, monotonicity, and accuracy of the objective quality predictors, respectively.

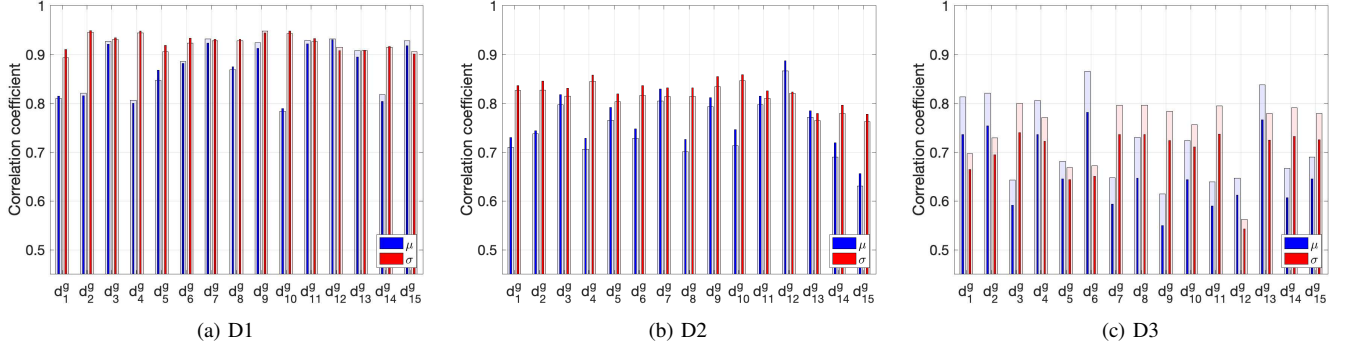


Fig. 2. Benchmarking of statistical features from geometric descriptors by means of PLCC (thin opaque bars) and SROCC (thick transparent bars). Each bar corresponds to a statistical feature, with blue and red color indicating mean and standard deviation, respectively, applied on corresponding descriptors.

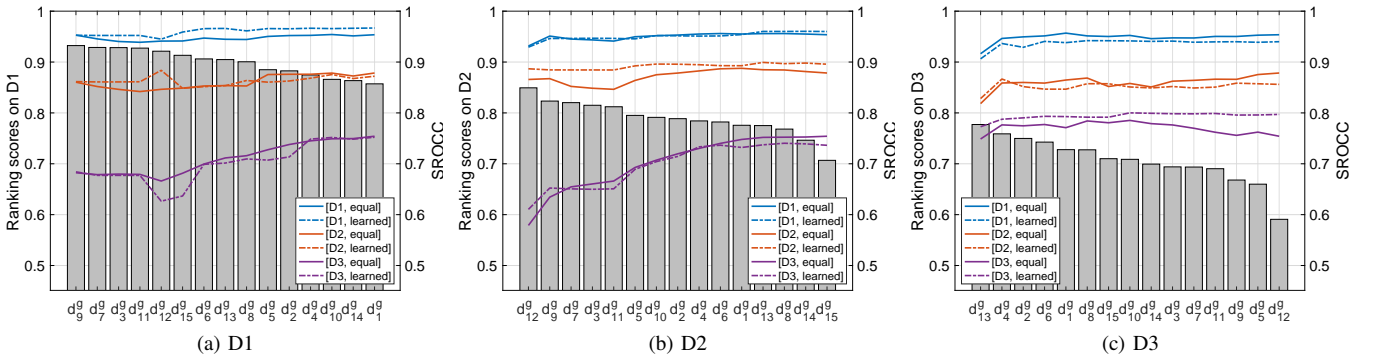


Fig. 3. Ranking scores of geometric descriptors (grey bars) computed as the average PLCC and SROCC over corresponding statistical features. The lines show the performance of progressively cumulating statistical features per dataset, using equal (solid) or learned (dashed) weights.

C. Execution of objective quality metrics

State-of-the-art objective quality metrics are employed in our performance evaluation analysis for comparison purposes. In particular, we use the point-to-point, point-to-plane [11] and plane-to-plane [12] metrics, which are focused on geometry-only distortions. The color PSNR on Y channel, as well as the histogram-based metric on luminance [33] (DistY) are recruited, assessing textural-only degradations. Finally, the joint point-to-distribution metric [10] using logarithmic values, the PointSSIM [13] applied on normal, curvature and luminance separately, the PCQM [7], as well as the reduced reference PCM_RR [34] are also benchmarked.

To compute the point-to-point and point-to-plane, the software version 0.13.5 [32] is used. For the latter, the required normal vectors are computed using quadric fitting with r -search and $r = 0.01 \times B_R$, where B_R indicates the maximum length of the bounding box of the reference point cloud. For plane-to-plane, the normal vectors are computed based on quadric fitting and r -search with $r = 0.02 \times B_R$, in a variant of the settings employed in [46]. In the point-to-distribution metric, neighborhoods consisting of $k = 31$ point samples are considered. For PointSSIM, the default parameters settings are employed, with the variance as the selected estimator of statistical dispersion, and $k = 12$. Quality predictions based on luminance, curvature and normal vectors, are computed, with the latter two obtained after quadric fitting with r -search ($r = 0.01 \times B_R$). In PCQM, the default configurations are

used, and the proposed weighted combination of the computed features is employed. For the histogram metric and the reduced reference point cloud metric, the script implementations with default settings are executed. Regarding the proposed metric, namely, PointPCA, to estimate the geometric descriptors, we use the r -search with r equal to $0.01 \times B_R$. For the computation of statistical features, we use the k -nn algorithm with $k = 25$.

V. RESULTS

A. Geometric descriptors

In Figure 2, the PLCC and SROCC of every statistical feature is illustrated in the form of bars, grouped per geometric descriptor, against subjectively annotated datasets. It can be noticed that the prediction accuracy of the proposed features is generally high, reaching a different performance plateau per dataset. We observe that, for a given descriptor, one statistical function might perform better than the other in one dataset, whereas the other might be superior in the rest. For instance, for d_1^g , the standard deviation exhibits higher accuracy in D1 and D2 with respect to the mean, with the opposite being true for D3. However, the differences remain limited. It is also remarked that there is no particular statistical feature performing poorly, or excelling consistently across all datasets.

As the next step, we examine different combination approaches for the proposed geometric descriptors. In Figure 3,

TABLE II
PERFORMANCE OF GEOMETRIC DESCRIPTORS.

Set	Testing dataset								
	D1			D2			D3		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
\mathbf{q}^g	0.955	0.954	0.338	0.856	0.878	0.705	0.818	0.754	1.411
\mathbf{q}_{D1}^g	0.972	0.967	0.269	0.853	0.871	0.713	0.812	0.751	1.430
\mathbf{q}_{D2}^g	0.962	0.960	0.311	0.878	0.896	0.652	0.803	0.736	1.450
\mathbf{q}_{D3}^g	0.947	0.940	0.368	0.842	0.859	0.736	0.870	0.796	1.204

the ranking scores of every descriptor, computed as the average PLCC and SROCC among corresponding statistical features, are shown with gray bars in descending order. The overlaying plots show how performance in terms of SROCC varies as new statistical features in ranking order are combined together, either with equal weights (solid line), or with optimized weights (dashed line). The latter is carried out by training a linear model using the leave- p -out method over each dataset, separately. In particular, a dataset is split into two equal groups, ensuring that all versions of the same content are included only in one group. Optimal weights are extracted from every partition under the problem formulation given in Equation 10, with the final weights obtained as the average across all partitions. Each plot additionally reports cross-dataset validation results, i.e., performance obtained by applying the weights learned on one dataset to the rest.

Focusing on the ranking scores, we observe resemblances in the ranking patterns between D1 and D2, due to the similar types of contents and distortions both datasets exhibit; yet, for D3, the trends differ. For instance, the top five descriptors in D1 coincides with D2, although at a different order, while they are ranked low in D3. Similarly, the top three descriptors in D3 are ranked below middle in D1 and D2. These results demonstrate that the visual quality of some types of distortions and contents are more accurately captured by certain descriptors.

Regarding the effect of different combinations and weighting strategies, the performance of statistical features varies depending on the dataset. In particular, considering all tested combinations and weightings, the SROCC is ranging for D1 in the interval [0.906, 0.967], for D2 in [0.818, 0.899], and for D3 in [0.579, 0.800]. Evidently, geometry-only predictors are excelling in D1, showing high resilience under any selected weighting approach. In D2, the performance decreases while maintaining good robustness, whereas in D3, the worst performance with the largest variability is found.

Considering a learning approach, we remark that, when weights are obtained from the same dataset used for testing, high performance is maintained across all examined combinations, with marginal differences (i.e., dashed blue, red and purple lines in Figures 3a, 3b and 3c, respectively). This indicates that optimal weights that are learned even on small sets of highly-performing descriptors on a dataset, lead to good results. When the weights are learned on other datasets, performance reductions naturally occur; this depends not only on the combination of statistical features in use, but also on the intrinsic characteristics of the testing dataset. For instance, under any set of weights, small deviations are observed when

TABLE III
PERFORMANCE OF TEXTURAL DESCRIPTOR.

Set	Testing dataset								
	D1			D2			D3		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
s_1	0.679	0.649	0.855	0.769	0.789	0.872	0.797	0.787	1.466
s_2	0.872	0.844	0.559	0.919	0.917	0.537	0.665	0.649	1.831
\mathbf{q}^t	0.810	0.785	0.669	0.924	0.929	0.522	0.690	0.681	1.764
\mathbf{q}_{D1}^t	0.857	0.831	0.589	0.922	0.923	0.529	0.672	0.659	1.811
\mathbf{q}_{D2}^t	0.801	0.776	0.685	0.924	0.929	0.523	0.690	0.687	1.762
\mathbf{q}_{D3}^t	0.695	0.666	0.835	0.813	0.832	0.794	0.805	0.797	1.443

testing on D1 and D2. On the contrary, evident reductions are noted in D3 under highly-ranked descriptors in D1 and D2 (which are less perceptually relevant for D3), implying the larger impact of a poorly tuned weighting scheme in more challenging datasets. Recruiting more statistical features, though, seems to be beneficial, as the performance is progressively improving. The same trend is observed when applying equal weights. Note that using the latter weighting scheme leads to performance reductions with respect to learning from the same testing dataset; yet, better results are obtained when compared to learning weights on other datasets.

Summarizing our analysis, high performance is achieved under a suitable subset of properly weighted geometric statistical features, per dataset. By incorporating a larger number of descriptors, lower performance variability across datasets is observed using either equal or learned weights, indicating higher generalization capabilities. Thus, we opt for the recruitment of all geometric descriptors. Corresponding performance results are presented in Table II, for equal weights (\mathbf{q}^g), and weights learned (\mathbf{q}_D^g) using every dataset $D = \{D1, D2, D3\}$.

B. Textural descriptor

The performance of every statistical feature is reported individually in Table III, with s_1 indicating the mean and s_2 the standard deviation. Moreover, we report performance indexes obtained after combining the above using either equal weights (\mathbf{q}^t), or weights learned (\mathbf{q}_D^t) in every dataset separately, similarly to the geometric descriptors analysis. We observe that the standard deviation performs better in D1 and D2, whereas the mean is superior in D3. Combining the two statistical functions with equal weights leads to performance gains in D2 with respect to using either of them individually. Conversely, the prediction accuracy drops for the remaining datasets, reaching values closer to the highest- and to the lowest-performing statistic for D1 and D3, respectively. By using learned weights, we observe improvements when the learning and testing are carried out on the same dataset, with declines occurring when learning on the rest. The only exception is observed when training with D2, where very similar results are obtained compared to the equal weights; indeed, the learned weights assign 0.55 to s_1 and 0.45 to s_2 .

Independently of the weights in use, the textural descriptor performs overall better in D2, and worse in D1 and D3. Note that the prediction accuracy of any combination is bounded by

TABLE IV
PERFORMANCE OF GEOMETRIC AND TEXTURAL
DESCRIPTORS COMBINATIONS.

Set	Testing dataset								
	D1			D2			D3		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
$\mathbf{q}^{g \cup t}$	0,954	0,952	0,342	0,871	0,894	0,670	0,827	0,768	1,384
$\mathbf{q}_{D1}^{g \cup t}$	0,965	0,958	0,300	0,907	0,922	0,573	0,864	0,794	1,228
$\mathbf{q}_{D2}^{g \cup t}$	0,931	0,920	0,417	0,935	0,946	0,483	0,880	0,853	1,154
$\mathbf{q}_{D3}^{g \cup t}$	0,910	0,897	0,475	0,890	0,906	0,621	0,912	0,887	0,998
$\tilde{\mathbf{q}}$	0,935	0,922	0,404	0,932	0,944	0,495	0,868	0,848	1,206

the performance of the individual statistical features, shown in Table III. We remark larger variability in performance for D1 with respect to D2 and D3, which indicates higher sensitivity for the former on the weights assigned to the statistical functions. It is noteworthy that learning weights on D1 leads to worse performance compared to using only the standard deviation. We recall that optimal weights are obtained using leave- p -out and by averaging across partitions that involve different contents for training and testing. Based on our results, learning in any subset of the 8i contents [48] leads to a dominant contribution of the mean, which steers the final weight allocation accordingly. This statistical function under-performs in D1, which explains this performance drop.

Summarizing, different statistical functions are proven more effective in capturing visual impairments across datasets. Sub-optimal weight allocation leads to substantial performance drops, as witnessed for example with D1. This can be explained by considering that only one descriptor is adopted; that is, only a single aspect of visual appearance is examined, implying lower robustness and higher sensitivity in weight assignment. Finally, comparing the results presented in Tables II and III, we note that the textural descriptor outperforms the geometric in D2, while the performance is notably lower in D1 and D3. Higher variability in prediction accuracy is observed across different combinations of textural predictors, for the same reason mentioned above. To this result, we should additionally account that the benchmarking is conducted in datasets that involve both textural and geometric distortions.

C. Geometric and textural descriptors

In Table IV, different combinations of geometry and texture descriptors are employed and corresponding performance results are reported. In particular, geometric and textural features are merged to a new set, with equal weights assigned ($\mathbf{q}^{g \cup t}$); that is, $w = 1/32$. Moreover, using the same set, weights are learned per dataset ($\mathbf{q}_D^{g \cup t}$). Finally, geometry and texture quality scores with equal weights (\mathbf{q}^d and \mathbf{q}^t , per Table II and III), are combined under Equation 11 and $\omega = 0.5$ ($\tilde{\mathbf{q}}$).

We observe that using equal weights across all statistical features results in similar or better performance with respect to the same weighting on geometry-only features. In comparison to textural-only features, a performance decrease in D2 is observed; however, remarkable improvements are noted in D1 and D3. Our results suggest that, by using equal weighting, geometric descriptors govern the outcome with the integration

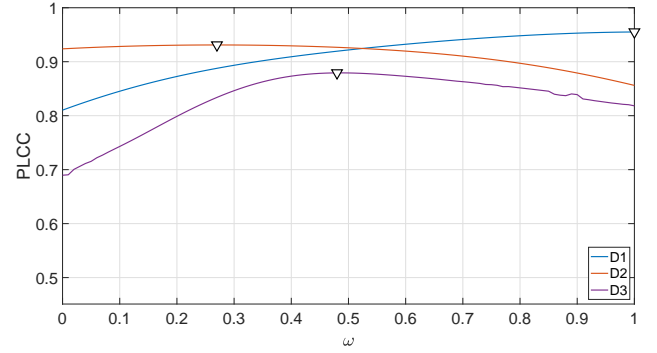


Fig. 4. PLCC achieved against each dataset, as a function of the regularizer term ω that defines the contributions between the geometric and textural quality scores. The markers indicate the maximum PLCC, per dataset.

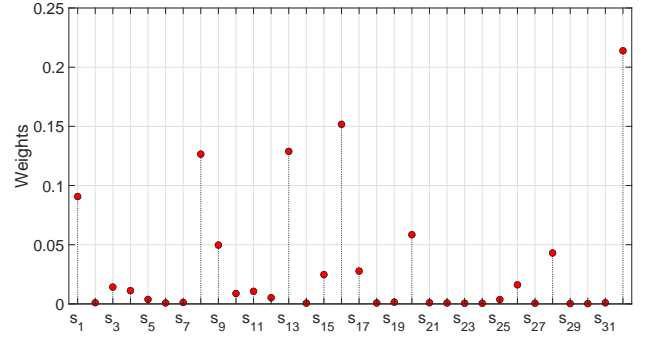


Fig. 5. Selected weights to compute a quality prediction with PointPCA.

of textural predictors being beneficial. The former is justified by the substantially larger number of geometric features employed, compared to the textural ones (i.e., 30 against 2).

Combining attribute-specific quality scores with same contributions leads to performance improvements in D2 and D3, while a slight decrease is observed in D1 with respect to $\mathbf{q}^{g \cup t}$. To understand how the optimal contribution of geometric and textural quality scores varies per dataset, we perform a grid search with values of $\omega \in [0, 1]$. For $\omega = 0$, only textural quality scores (\mathbf{q}^t) are considered, whereas for $\omega = 1$, only geometric quality scores are employed (\mathbf{q}^g). Results are shown for all datasets in Figure 4. The trends indicate that for D1, the integration of texture-based predictors doesn't improve the results, thus, using geometry-only descriptors is optimal. For D2, texture-only descriptors perform better than geometry-only, while their combination can lead to enhancements. For D3, geometry-only descriptors are found better than texture-only, with notable improvements observed when considering both. Evidently, a balanced contribution achieved with $\omega = 0.5$, provides a good compromise across all datasets.

Finally, by learning weights, it can be seen that the overall performance is improved when considering both geometric and textural descriptors, with respect to using descriptors from only one domain. Performance reductions when weights are learned on another dataset with respect to weights obtained from learning in the same dataset are observed, as expected. However, combining geometry and texture descriptors marks an improvement in both cases, excluding D1 where we note decreases; in this dataset, any weight allocation that drives contributions towards texture is sub-optimal, as noted earlier.

TABLE V
PERFORMANCE EVALUATION OF STATE-OF-THE-ART QUALITY METRICS.

Metric	Testing dataset								
	D1			D2			D3		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Point-to-point_MSE	0,945	0,935	0,375	0,844	0,867	0,731	0,237	0,463	2,361
Point-to-plane_MSE	0,952	0,945	0,349	0,866	0,885	0,681	0,220	0,483	2,370
PSNR point-to-point_MSE*	0,867	0,855	0,545	0,719	0,759	0,885	-	-	-
PSNR point-to-plane_MSE*	0,911	0,909	0,449	0,767	0,804	0,818	-	-	-
Plane-to-plane	0,882	0,876	0,537	0,802	0,810	0,815	0,796	0,736	1,470
PSNR Y	0,888	0,893	0,526	0,670	0,662	1,012	0,687	0,689	1,766
DistY	0,695	0,652	0,829	0,861	0,887	0,694	0,041	-0,112	2,428
LOG point-to-distribution-joint	0,922	0,912	0,441	0,901	0,914	0,591	0,606	0,594	1,934
PointSSIM (norm)	0,901	0,905	0,497	0,753	0,765	0,897	0,812	0,752	1,434
PointSSIM (curv)	0,937	0,939	0,398	0,829	0,844	0,763	0,828	0,762	1,383
PointSSIM (luma)	0,892	0,866	0,516	0,926	0,917	0,515	0,701	0,687	1,736
PCQM	0,967	0,970	0,291	0,899	0,915	0,597	0,864	0,854	1,225
PCM_RR	0,895	0,893	0,510	0,864	0,886	0,686	0,590	0,553	1,974
PointPCA	0,931	0,920	0,417	0,935	0,946	0,483	0,880	0,853	1,154

The above observations indicate the benefits in accounting for both types of attributes, which result in improving the accuracy and the resilience of the predictions. The best performance is attained by learning on the entire set of descriptors using D2. Thus, these weights, shown in Figure 5, are selected as the final weights in our metric to provide a quality score. The close-second performance achieved by \tilde{q} is noteworthy, considering that there is no learning involved in its computation. Thus, it provides a viable alternative that avoids concerns regarding over-fitting or optimistic biases, validating the efficiency of the proposed features.

D. Comparison with state-of-the-art metrics

In Table V, we report performance evaluation results for existing point cloud quality metrics, for comparison purposes. It can be seen that the proposed metric achieves competitive, if not better performance on all tested datasets. In the case of D1, it is worth noting that geometry-only approaches achieve very good performance, supporting our findings that the geometric descriptors exhibit stronger performance, as illustrated in Figure 4. In this dataset, PCQM achieves the best performance in terms of PLCC, SROCC, and RMSE. For D2, we observe that our proposed metric achieves the best performance for all indicators. PointSSIM using luminance, MMD-joint and PCQM also attain high correlation scores, with the former confirming our observations that luminance-based predictors are accurate in capturing visual distortions in this dataset. Finally, for D3 we remark that our proposed metric achieves the highest performance in terms of PLCC and RMSE, while reaching a close second SROCC. In this dataset, PCQM is a direct competitor, with the normal-based and curvature based variants of PointSSIM following, and the plane-to-plane denoting the last member of the top five metrics. Note that the results of point-to-point and point-to-plane with PSNR are not provided, considering that correlation computations were not possible for 54 out of 378 stimuli.

VI. CONCLUSION

In this paper, we propose a point cloud objective quality metric that relies on PCA-based shape descriptors and luminance-based predictors to evaluate distortions in the geometry and color domain, respectively. Statistical functions are applied on the descriptor values in order to capture local relationships between point samples, which are compared between the pristine and the distorted stimuli in order to provide a quality prediction for the latter. The proposed features are evaluated individually, showing good overall performance, although not consistently ranked across datasets. To alleviate this behaviour, several weighting combination strategies were examined, both within the same attribute domain of descriptors, as well as by combining them. Well-tuned configurations were shown to lead to substantial performance improvements and higher generalization capabilities. Learning weights over the entire set of geometric and textural descriptors was found to be the highest-performing approach. A balanced contribution between geometry and texture quality scores computed using equal weights, denoted a close alternative, confirming the effectiveness of the proposed features. Our results show that the proposed metric achieves state-of-the-art performance, outperforming existing solutions in the majority of cases across all tested datasets. Considering that certain descriptors are more efficient against particular types of contents and degradations, future work will focus on identification and adoption of most perceptually-relevant predictors, per case.

REFERENCES

- [1] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, 2016.
- [2] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging MPEG standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2019.

- [3] 3DG, “Text of ISO/IEC CD 23090-9 Geometry-based Point Cloud Compression,” ISO/IEC JTC1/SC29/WG11 Doc. N18478, Jul. 2019.
- [4] —, “Text of ISO/IEC DIS 23090-5 Video-based Point Cloud Compression,” ISO/IEC JTC1/SC29/WG11 Doc. N18670, Oct. 2019.
- [5] WG1, “Final Call for Evidence on JPEG Pleno Point Cloud Coding,” ISO/IEC JTC1/SC29/WG1 Doc. N88014, Jul. 2020.
- [6] 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, p. e27, 2019.
- [7] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, “PCQM: A full-reference quality metric for colored 3D point clouds,” in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.
- [8] E. Alexiou, N. Yang, and T. Ebrahimi, “PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality,” in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.
- [9] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, “Inferring point cloud quality via graph similarity,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.
- [10] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “A point-to-distribution joint geometry and color metric for point cloud quality assessment,” *arXiv preprint arXiv:2108.00054*, 2021.
- [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] E. Alexiou and T. Ebrahimi, “Towards a point cloud structural similarity metric,” in *2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 2020, pp. 1–6.
- [14] N. Chehata, L. Guo, and C. Mallet, “Airborne lidar feature selection for urban classification using random forests,” in *Laserscanning*, 2009.
- [15] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, “Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 286–304, 2015.
- [16] T. Hackel, J. D. Wegner, and K. Schindler, “Fast semantic segmentation of 3D point clouds with strongly varying density,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 177–184, 2016.
- [17] —, “Contour detection in unstructured 3D point clouds,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1610–1618.
- [18] Z. Zhang, W. Sun, T. Wang, W. Lu, X. Min, and G. Zhai, “No-reference quality assessment for 3D colored point cloud and mesh models,” *arXiv preprint arXiv:2107.02041*, 2021.
- [19] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “A generalized Hausdorff distance based quality metric for point cloud geometry,” in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.
- [20] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, “Evaluation metrics for point cloud compression,” ISO/IEC JTC1/SC29/WG11 Doc. M39966, Geneva, Switzerland, Jan. 2017.
- [21] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “Improving psnr-based quality metrics performance for point cloud geometry,” in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3438–3442.
- [22] 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.
- [23] 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.
- [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*, Q. Dai, T. Shimura, and Z. Zheng, Eds., vol. 11550, International Society for Optics and Photonics. SPIE, 2020, pp. 244 – 252.
- [25] 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.
- [26] 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.
- [27] R. Diniz, P. G. Freitas, and M. C. Q. Farias, “Towards a point cloud quality assessment model using local binary patterns,” in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.
- [28] R. Diniz, P. G. Freitas, and M. C. Farias, “Multi-distance point cloud quality assessment,” in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3443–3447.
- [29] —, “Local luminance patterns for point cloud quality assessment,” in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, 2020, pp. 1–6.
- [30] 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. 256–1–256–11, 2021.
- [31] R. Diniz, P. G. Freitas, and M. C. Q. Farias, “Color and geometry texture descriptors for point-cloud quality assessment,” *IEEE Signal Processing Letters*, vol. 28, pp. 1150–1154, 2021.
- [32] 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, Hobart, Australia, Apr. 2017.
- [33] I. Viola, S. Subramanyam, and P. Cesar, “A color-based objective quality metric for point cloud contents,” in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.
- [34] I. Viola and P. Cesar, “A reduced reference metric for visual quality evaluation of point cloud contents,” *IEEE Signal Processing Letters*, vol. 27, pp. 1660–1664, 2020.
- [35] 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 Transactions on Image Processing*, vol. 30, pp. 6623–6636, 2021.
- [36] 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*, A. G. Tescher, Ed., vol. 10752, International Society for Optics and Photonics. SPIE, 2018, pp. 174 – 190.
- [37] 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.
- [38] 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 Transactions on Multimedia*, pp. 1–1, 2020.
- [39] Z. He, G. Jiang, Z. Jiang, and M. Yu, “Towards a colored point cloud quality assessment method using colored texture and curvature projection,” in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 1444–1448.
- [40] T. Chen, C. Long, H. Su, L. Chen, J. Chi, Z. Pan, H. Yang, and Y. Liu, “Layered projection-based quality assessment of 3D point clouds,” *IEEE Access*, vol. 9, pp. 88 108–88 120, 2021.
- [41] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [42] K. F. West, B. N. Webb, J. R. Lersch, S. Pothier, J. M. Triscari, and A. E. Iverson, “Context-driven automated target detection in 3D data,” in *Automatic Target Recognition XIV*, F. A. Sadjadi, Ed., vol. 5426, International Society for Optics and Photonics. SPIE, 2004, pp. 133 – 143.
- [43] M. Pauly, M. Gross, and L. Kobbelt, “Efficient simplification of point-sampled surfaces,” in *IEEE Visualization, 2002. VIS 2002.*, 2002, pp. 163–170.
- [44] J. Demantké, B. Vallet, and N. Paparoditis, “Streamed vertical rectangle detection in terrestrial laser scans for facade database production,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 99–104, 2012.
- [45] ITU-R BT.709-6, “Parameter values for the HDTV standards for production and international programme exchange,” International Telecommunication Union, Jun. 2015.
- [46] S. Perry *et al.*, “Quality evaluation of static point clouds encoded using MPEG codecs,” in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3428–3432.
- [47] ITU-T J.149, “Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM),” International Telecommunication Union, Mar. 2004.
- [48] E. Eon, B. Harrison, T. Myers, and P. A. Chou, “8i voxelized full bodies, version 2 — a voxelized point cloud dataset,” ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) Doc. m40059/M74006, Jun. 2017.

- [49] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1338–1350, 2001.
- [50] G. Lavoué, E. D. Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3D distance metrics with application to watermarking," in *Applications of Digital Image Processing XXIX*, vol. 6312. SPIE, 2006, pp. 150 – 161.
- [51] G. Lavoué, "A multiscale metric for 3D mesh visual quality assessment," in *Computer Graphics Forum*, vol. 30, no. 5. Wiley Online Library, 2011, pp. 1427–1437.
- [52] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [53] C. Loop, Q. Cai, S. Escolano, and P. Chou, "Microsoft voxelized upper bodies - a voxelized point cloud dataset," ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) Doc. m38673/M72012, May 2016.

APPENDIX A

PARAMETER SELECTION

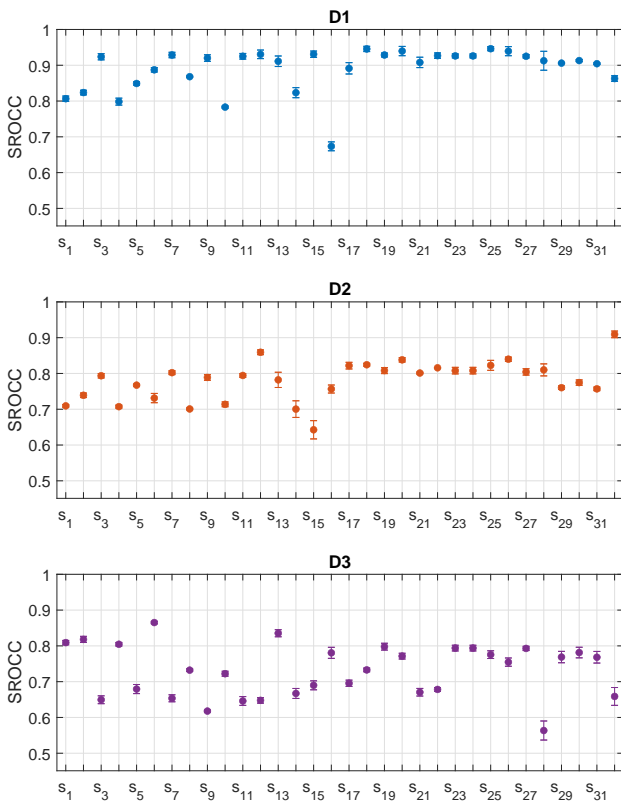


Fig. 6. Error bars indicating the mean and standard deviation of the SROCC index, for every predictor s_j , across neighborhood sizes using the k -nn algorithm with $k = \{9, 25, 49, 81\}$.

A. Support regions for statistical features

In Figure 6, the mean and standard deviation of SROCC values achieved by every predictor s_j with $1 \leq j \leq 32$ are illustrated, under neighborhoods formed using k -nn, with $k = \{9, 25, 49, 81\}$. Recall that s_1 - s_{16} indicates the use of mean, while from s_{16} - s_{32} the standard deviation is employed. For every statistic, the first 15 predictors refer to the geometry and the last to the texture domain. Regarding the neighborhood sizes, we choose the aforementioned k values, considering that the point clouds included in the datasets under exam are voxelized, dense, and represent large models. Thus, we

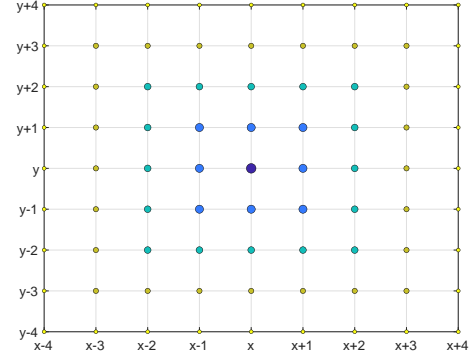


Fig. 7. Point arrangement on a planar grid (x, y) . The central is a queried point. The neighborhoods identified with the selected k 's are color- and size-coded, assuming full occupancy and planar underlying surface.

may assume that small point neighborhoods represent local regions, which in turn can be approximated by planar surfaces. The selected k 's represent the number of vertices in fully-occupied planars of length size equal to 2, 4, 6 and 8 times the distance between two voxels, as shown in Figure 7. Our results indicate that all predictors show high robustness against the tested neighborhood sizes.

It is worth noting that there is an inter-dependency between support regions used for the estimation of geometric descriptors, and the computation of corresponding statistical features. For example, increasing the support region to estimate a descriptor leads to lower deviations between neighboring descriptor values; concurrently, decreasing the population to compute statistical features has a similar effect. In our case, the relationship between the two is examined by fixing the neighborhood size over which the descriptors are estimated (i.e., r -search with $r = 0.01 \times B_R$) and altering the population size (i.e., k values) to compute statistical features. The obtained results confirm that the geometric descriptors' configuration leads to high prediction accuracy, with stable performance across all tested k 's. In the proposed settings of our metric, we set $k = 25$.

B. Color spaces

In Table VI, we present the performance indexes obtained from textural measurements computed identically, considering different color spaces. We aim at validating the selection of the luminance-only component, by examining the performance achieved with the proposed statistical features over alternative color spaces that are popular in the literature. In particular, we employ the YCbCr color space as a straight-forward baseline, the default RGB as an anchor, and we additionally recruit the GCM [49], which is reported to correlate well with human perception, and CIELAB recommended by the International Commission on Illumination in 1976 and designed for perceptual uniformity. In this analysis, we employ only textural information. As usual, we learn optimal weights per dataset and test their prediction accuracy on all. Our results show that luminance-only leads to higher robustness across datasets, with very similar, or substantially better performance than the alternatives. Specifically, when training with D1 or D2,

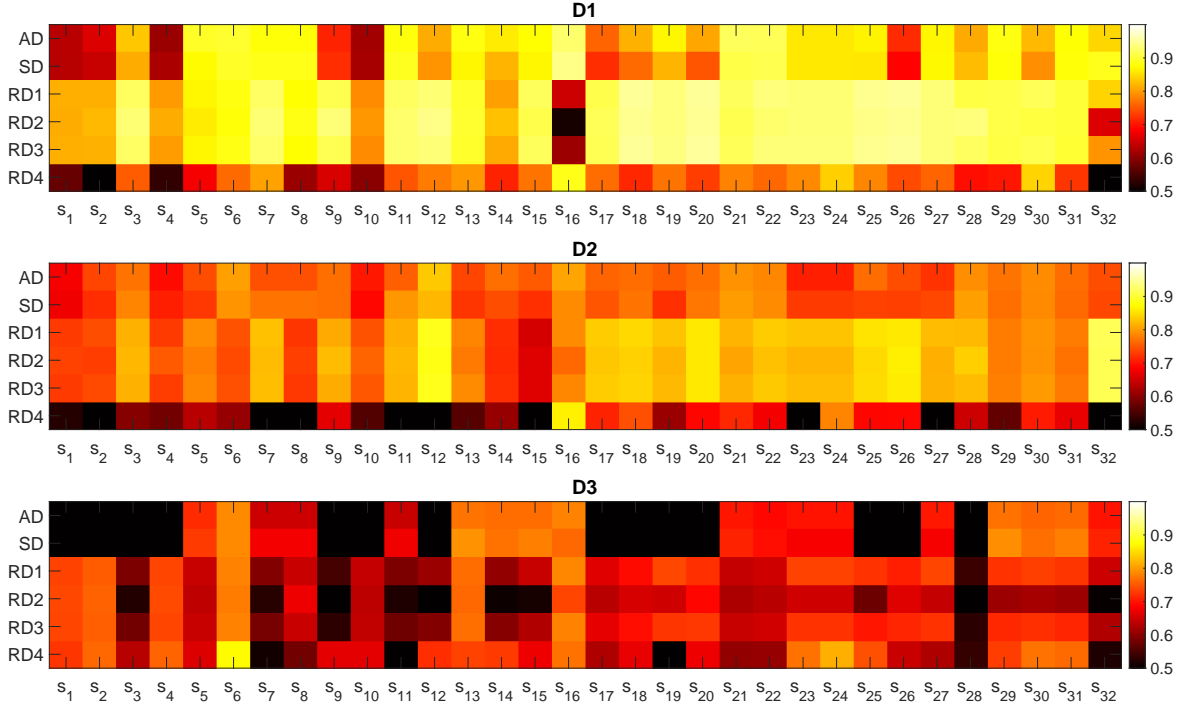


Fig. 8. SROCC achieved by every predictor s_j per dataset, using different methods to compare statistical features from reference and distorted contents.

TABLE VI
PERFORMANCE OF TEXTURAL DESCRIPTORS IN
DIFFERENT COLOR SPACES.

Space	D	Testing dataset								
		D1			D2			D3		
		PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
Y	D1	0.857	0.831	0.589	0.922	0.923	0.529	0.672	0.659	1.811
	D2	0.801	0.776	0.685	0.924	0.929	0.523	0.690	0.687	1.762
	D3	0.695	0.666	0.835	0.813	0.832	0.794	0.805	0.797	1.443
YCbCr	D1	0.837	0.829	0.626	0.901	0.908	0.591	0.679	0.689	1.785
	D2	0.800	0.779	0.685	0.921	0.928	0.533	0.650	0.654	1.849
	D3	0.701	0.672	0.827	0.814	0.832	0.792	0.804	0.798	1.445
RGB	D1	0.837	0.815	0.625	0.925	0.925	0.517	0.628	0.621	1.906
	D2	0.767	0.746	0.733	0.930	0.934	0.502	0.648	0.647	1.856
	D3	0.648	0.559	0.939	0.781	0.797	0.852	0.808	0.802	1.433
GCM	D1	0.851	0.838	0.600	0.928	0.933	0.508	0.639	0.637	1.875
	D2	0.776	0.756	0.720	0.927	0.935	0.510	0.598	0.603	1.950
	D3	0.690	0.662	0.840	0.812	0.828	0.796	0.794	0.788	1.476
CIELAB	D1	0.877	0.860	0.549	0.925	0.931	0.518	0.639	0.637	1.878
	D2	0.819	0.802	0.656	0.926	0.932	0.514	0.640	0.638	1.871
	D3	0.694	0.667	0.838	0.794	0.806	0.829	0.784	0.780	1.509

there are color spaces that perform slightly better in either of them; for instance, CIELAB performs better at both, while RGB and GCM are better at D2, with marginal differences in all cases. However, luminance-only out-performs the alternatives by substantial margins when testing on D3. By learning weights in D3, Y and YCbCr are found to attain almost equivalent performance. Our analysis indicates that the proposed predictors show similar performance under all tested color spaces; yet, the higher resilience observed using only the luminance component, justifies our selection.

C. Comparison methods

We analyze the impact of selecting a different comparison method between statistical features extracted from the reference and the distorted stimulus. Considering conventional 2D imaging formulas, the relative difference between pixel values has been proven more effective, as it follows the Weber's law and, hence, accounts for the higher sensitivity of human visual perception in lower luminance regions. Initially, mesh [50], [51] and, more recently, point cloud quality metrics [7], [13], [23]–[25] have been adopting relative difference formulas for both color and geometric measurements, without evidence though regarding how they compare to simpler alternatives. We shed light into the matter by illustrating performance differences between the descriptors proposed in the context of PointPCA. Specifically, we employ the comparison methods provided in Equations 12-17, with ϕ_{j_1} and ϕ_{j_2} indicating corresponding statistical features obtained from the reference and the distorted point clouds:

$$AD = |\phi_{j_1} - \phi_{j_2}| \quad (12)$$

$$SD = |\phi_{j_1} - \phi_{j_2}|^2 \quad (13)$$

$$RD1 = \frac{|\phi_{j_1} - \phi_{j_2}|}{\max\{|\phi_{j_1}|, |\phi_{j_2}|\} + \varepsilon} \quad (14)$$

$$RD2 = \frac{2\phi_{j_1}\phi_{j_2}}{\phi_{j_1}^2 + \phi_{j_2}^2 + \varepsilon} \quad (15)$$

$$RD3 = \frac{|\phi_{j_1} - \phi_{j_2}|}{|\phi_{j_1} + \phi_{j_2}| + \varepsilon} \quad (16)$$

$$RD4 = \frac{\max\{|\phi_{j_1}|, |\phi_{j_2}|\}}{|\phi_{j_1} - \phi_{j_2}|} \quad (17)$$

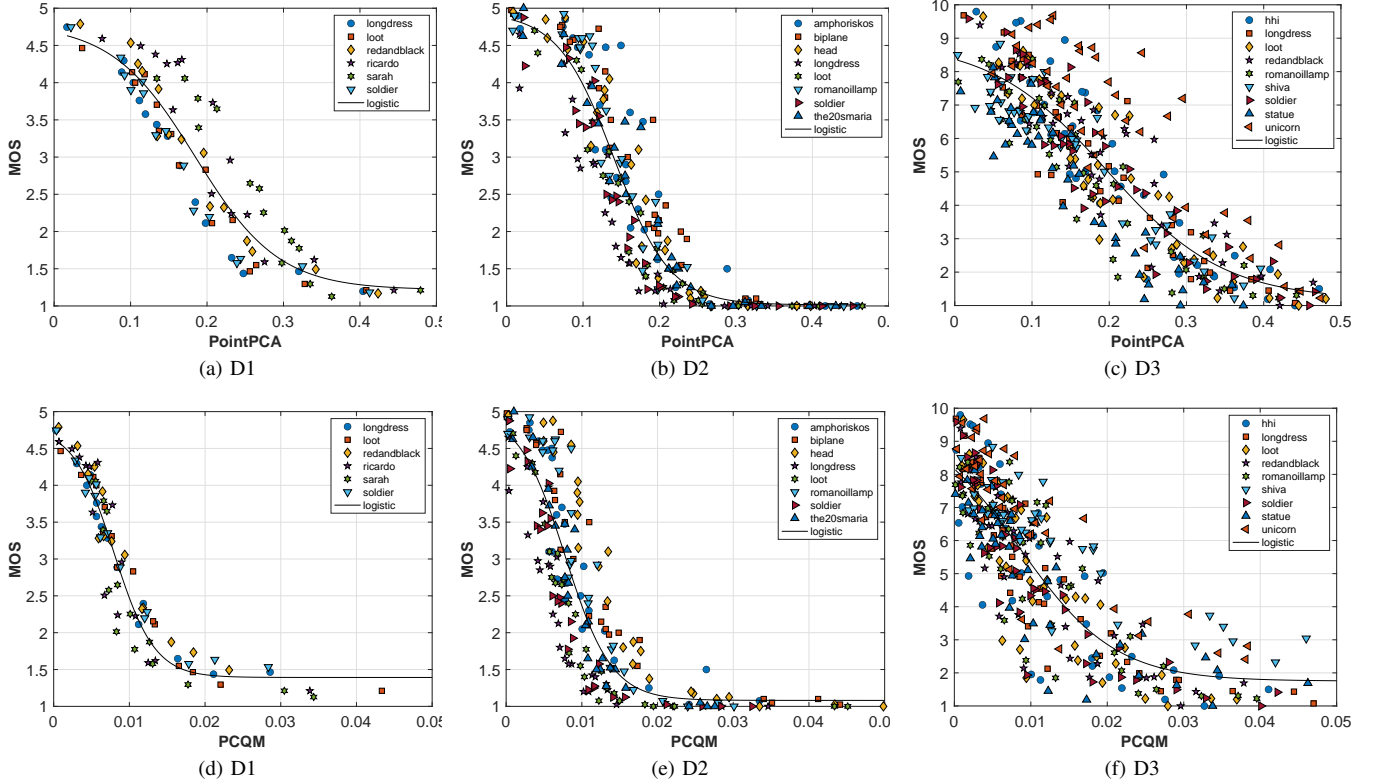


Fig. 9. Scatter plots of subjective scores against objective predictions from PointPCA (first row) and PCQM (second row), per dataset (column). Note that for PCQM we provide a zoomed version for better illustration; the max error was 0.069, 0.1126 and 0.095, for D1, D2 and D3, respectively.

with ε set to machine rounding error for floating point numbers. In particular, in the first two equations we account for the absolute (AD) and the squared difference (SD) of the measurements, which are straightforward comparison methods. The subsequent RD1, RD2 and RD3 denote three different implementations of the relative difference, which lead to different penalization strategies; that is, the RD1 has been used in all point cloud metric implementations [7], [13], [23]–[25], the RD2 has been widely used in conventional image metrics, such as in [26], [52], whereas the RD3 is a simple alternative. Finally, RD4 is an ad-hoc implementation following the opposite behaviour than the Weber’s law; namely, penalizing equal differences more at larger magnitudes.

In Figure 8, we show the SROCC attained using every predictor s_j and corresponding comparison methods, per dataset. We observe that RD1, RD2 and RD3 are providing overall better performance, with marginal differences between them favoring RD1. Evidently, these results hold for both geometric and textural descriptors, and are in principle generalized across all datasets. For RD4, obvious performance reductions are noted for all predictors, excluding s_{16} (i.e., mean of textural descriptor) in D1 and D2. However, this behaviour cannot be generalized in D3, while also, using RD4 and s_{32} (i.e., standard deviation of textural descriptor), also leads to lower accuracy in all datasets. Thus, despite the individual gains, the overall performance of RD4 is judged limited, as expected.

Considering the AD and SD, we note that they bring some advantages to predictors that capture surface orientation (i.e., s_{13} – s_{15} and s_{29} – s_{31}) in D3. Moreover, gains are sporadically

noted for the eigenvalue-based predictors linearity, planarity, sphericity and anisotropy (i.e., s_5 – s_8 and s_{21} – s_{24}); however, no consistent trends are reported, while the differences remain minor in all cases. It is noteworthy that both AD and SD perform consistently better with textural predictors (i.e., s_{16} and s_{32}) in D1. Their superior performance in comparison to the RD1–RD3 variants is explained by the color distribution of the contents of this dataset. In particular, “ricardo” and “sarah” denote two point clouds from the Microsoft Voxalized Upper Bodies dataset [53] with large surfaces covered in dark, which leads the RD1–RD3 variants to penalize more exhibiting distortions. Hence, their generalization capabilities across the contents of this dataset are affected negatively. This observation justifies our previous findings, regarding the low performance of the textural descriptor under the proposed RD1 variant integrated in our metric, against this dataset. Future developments should consider mechanisms to compensate such cross-content variations, for better performance.

APPENDIX B PERFORMANCE EVALUATION

In Figure 9, we provide scatter plots illustrating the prediction accuracy of the proposed metric and the PCQM as a competing baseline against all datasets. We note that the objective scores from PointPCA are spanning in a larger range, showing higher discrimination power. In principle, though, very similar performance is observed between the two metrics, which is also reflected in the indicators that have been reported

in the paper. Regarding PointPCA, we observe that the correlation achieved in D1 is high. Based on previous findings, in this dataset geometric predictors excel, with textural ones being sub-optimal. From Figure 9a, it is evident that these discrepancies are effectively introduced by the “ricardo” and “sarah” contents, for the reasons mentioned in section A-C. In D2, the performance remains high, achieving the highest linearity, monotonocity and accuracy, compared to other existing solutions, based on the reported PLCC, SROCC and RMSE performance indexes. In D3, the performance reduces, with a larger spread of data points. However, this is mainly caused from the “unicorn” content, which shows an outlying behavior. In particular, when excluding it, the performance is reaching a PLCC of 0.910, SROCC of 0.891 and RMSE of 0.988. In the same setting, PCQM achieves 0.865, 0.858 and 1.195 for the above indicators, respectively.