A novel methodology for quality assessment of voxelized point clouds

*Eric M. Torlig^a, *Evangelos Alexiou^b, Tiago A. Fonseca^c,
Ricardo L. de Queiroz^d, Touradj Ebrahimi^b

^{a}Electrical Engineering Department, Universidade de Brasília, Brazil
^{b}Multimedia Signal Processing Group, École Polytechnique Fédérale de Lausanne, Switzerland
^{c}Gama Engineering College, Universidade de Brasília, Brazil
^{d}Computer Science Department, Universidade de Brasília, Brazil;

ABSTRACT

Recent trends in multimedia technologies indicate a significant growth of interest for new imaging modalities that aim to provide immersive experiences by increasing the engagement of the user with the content. Among other solutions, point clouds denote an alternative 3D content representation that allows visualization of static or dynamic scenes in a more immersive way. As in many imaging applications, the visual quality of a point cloud content is of crucial importance, as it directly affects the user experience. Despite the recent efforts from the scientific community, subjective and objective quality assessment for this type of visual data representation remains an open problem. In this paper, we propose a new, alternative framework for quality assessment of point clouds. In particular, we develop a rendering software, which performs real-time voxelization and projection of the 3D point clouds onto 2D planes, while allowing interaction between the user and the projected views. These projected images are then employed by two-dimensional objective quality metrics, in order to predict the perceptual quality of the displayed stimuli. Benchmarking results, using subjective ratings that were obtained through experiments in two test laboratories, show that our framework provides high predictive power and outperforms the state of the art in objective quality assessment of point cloud imaging.

Keywords: Voxelized point clouds, quality assessment, subjective evaluations, objective metrics

1. INTRODUCTION

New information and communication systems are progressively adopting 3D visual modalities for content representation in order to better approximate the perception of real-world sceneries. To this aim, point cloud technology provides a viable solution to represent richer visual stimulations, due to the low complexity and high efficiency in capturing, encoding and rendering of 3D contents. An overview of the wide range of target applications for which point clouds can be adopted is detailed in a recent JPEG document "Use cases and requirements".[1] When it comes to immersive applications and communication technologies, the visual quality of a content is of crucial importance, as it directly affects the user experience. Yet, the quality assessment for this type of imaging remains a challenge. Subjective evaluations are extremely valuable for their reliability; however, they are expensive in terms of time and cost. Hence, the development of an objective metric that would accurately predict the perceptual quality of distorted point cloud contents, is essential. Such an advance would benefit several technologies, such as sophisticated compression and transmission mechanisms, which typically depend on fast and inexpensive ways to assess introduced degradations.

A number of studies has been recently focused on subjective quality assessment of point clouds, under different types of degradations and using different content representations. In particular, Zhang et al.[2] conducted experiments with point clouds, whose geometry and color were degraded using various levels of uniform noise.

Alexiou et al.[3,4] performed subjective evaluation of geometry-only contents subject to Gaussian noise and octree-based compression. The distorted models were visualized as raw point clouds and assessed using an interactive protocol in a desktop setup and an augmented reality scenario, respectively. Mekuria et al.[5] conducted a study involving compression of dynamic colored point clouds, which were assessed in a tele-immersive environment. Javaheri et al.[6] performed subjective evaluations of octree- and graph-based encoding schemes, while the color attributes of the original models remained uncompressed. The point clouds were displayed using cube primitives, whose size was automatically adjusted based on nearest neighbors and assessed in a passive way. Javaheri et al.[7] performed quality assessment of point cloud denoising algorithms, with the subjects visualizing the resulted contents after applying Screened Poisson surface reconstruction. Similar approach was followed in another study,[8] where colorless point cloud models were encoded using octree-pruning and then visualized as polygonal meshes using the same surface reconstruction algorithm.

As implied by the numerous alternative content representations that were adopted in the aforementioned studies, one main issue towards defining an adequate solution for quality assessment of point clouds is the way this visual data is displayed. In particular, cube primitives have been proposed, whose size is either fixed, or adaptively determined based on the intrinsic geometry resolution of the contents.[7,9] In several studies,[2-4] raw point clouds were adopted, which is rather uncommon approach for visualization of 3D models. In other studies,[7,8] a mesh reconstruction was employed as a rendering algorithm to account for a more natural way of consumption. Notably, the latter study showed poor correlation between visualization of raw point clouds and reconstructed meshes as rendering means for identically degraded point clouds. Furthermore, benchmarking results of the state-of-the-art objective quality metrics, which was conducted as part of the majority of the aforementioned studies, show that their performance is not sufficiently accurate. In fact, there are two main reasons why poor correlation is observed: (a) The current metrics do not consider the rendering approach, which naturally affects the visual perception of artifacts. (b) The current metrics are focused on providing an objective score based on either the geometric or the color degradation of an impaired object, without defining a formula that properly combines the influence of each type of distortion in the overall perceived quality.

In this study, we aim at tackling these issues by proposing a novel framework that provides an alternative to quality assessment of point clouds. As a first step, we devise a rendering scheme for visualization of point clouds in typical 2D screens. In particular, to enable manual handling and display configuration for the rendering of the contents under assessment, a prototype software implementation was developed based on real-time mapping of the 3D volumes to 2D images using orthographic projection. As a second step, we propose the use of conventional 2D metric techniques, which are applied on the projected views as they are visualized in the proposed renderer. This way, we combine geometric and color degradations of the point cloud contents in a single objective score, while also taking under consideration the specific way in which the objects are displayed.

Although perspective projection provides a more accurate model of how the human visual system works, the use of orthographic projections comes with a number of advantages, stemming from the fact that processing is independent of the distance between the viewer and the object. In fact, perspective projection behaves identically to orthographic when the point of view goes to infinity. In other terms, the latter model approximates the effect of seeing an object from a distance, without causing any discomfort to the viewers for visualization of contents in mid- or closer-ranges. Moreover, orthographic projections come with a number of positive consequences. Firstly, they do not suffer from the effect of perspective distortion, in which different parts of the same object can appear to warp or change their relative sizes based on the viewing position. In perspective projections, objects closer to the observer appear larger than objects that are farther away. By using an orthographic projection, this effect disappears, which reduces both the computational cost of visualizations and a number of artifacts that are introduced in the projected images, simplifying the analysis. A second advantage is that rendered elements can have a fixed, predetermined size while remaining watertight, even when the rendered image is magnified. This is not the case with other rendering approaches, such as the use of fixed sized splats, which can reveal empty spaces between image elements while zooming into the object.

Rendering costs can be further reduced when the point clouds are voxelized. Voxelization is the process by which 3D models are converted from a continuous space to a discrete spatial representation. In non-voxelized contents, the coordinates of a point can be found at any position in 3D space. When points are voxelized, they are represented as cubic volumetric elements, namely voxels, in a regularly spaced grid. This lattice grid is

typically accessed by integer indexes. Voxels can be either occupied or not. When they are occupied, a color value is displayed, given by the corresponding point. In case more than one points fall within the same area, the color of the voxel is given by the average of the color values of these points. An example, is shown in Figure 1.
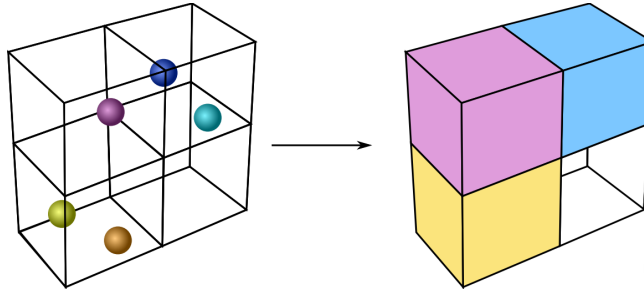

Figure 1. Example of the voxelization process.

When voxels are projected frontally, i.e. when their faces are parallel to the projection plane, there is a one-to-one relationship between visible voxels and pixels that belong to the projected image. This indicates that there is no need for interpolation, or any other type of post-processing, when the projection plane is perpendicular to one of the reference axes that define the coordinate system where the object is placed. In the case of a three-dimensional Cartesian coordinate system, there are 6 projection planes that fulfil this condition. We exploit this property by issuing the 2D objective metrics on the 6 orthographic projections of the voxelized contents that correspond to their reference axes. Then, a total degradation value is obtained, which represents the level of visual impairment of this content as perceived from the corresponding 6 viewing angles.

To assess the performance of our approach, subjective evaluations were conducted in two different test laboratories using the proposed rendering scheme on a diverse set of contents. The point clouds were assessed after being encoded by a wide range of geometry and color degradations. Correlation results show the superiority of our methodology with respect to state-of-the-art objective quality metrics.

2. DESIGN OF EXPERIMENTS

2.1 Rendering Software

The rendering software used in experiments was developed in C++. Two point cloud files are read and rendered in separate windows, shown side by side in a graphical user interface (GUI). The GUI also includes a set of rating scores that users can choose from, and a button to load the next content to be displayed.

Rendering is done in two steps. In the first step, each point cloud is read point by point. The upper and lower bounds in each dimension are recorded, along with the average position of all the points (i.e. the centroid). By subtracting the centroid from each set of coordinates, the 3D model is centered in the viewing volume. Then, both point clouds are scaled by multiplying the spatial coordinates of each point by the biggest power of two that would still let the model fit in a cubic volume of side 1024. In other terms, spatial coordinates are multiplied by a scaling component, s, given by Equation 1

$$s = z_f \times 2^{(10 - \lceil \log_2(w_{max} - w_{min}) \rceil)} \tag{1}$$

where z_f is a zoom factor that is inversely proportional to the virtual distance between the user and the content, and is updated based on the user's scrolling of the mouse wheel, whilst w_{max} and w_{min} correspond to the largest and smallest coordinate values of the point cloud. Essentially, this procedure scales the content appropriately as a function of the current virtual distance, and then projects the content (or part of it) in a pixel grid of 1024 by 1024 resolution. Subsequently, the points go through a rigid rotation, as a function of the viewing angle. This is done by multiplying the spatial coordinates by a rotation matrix. This rotation matrix is calculated using angles in two axes, which are determined dynamically by the user by clicking and dragging with the right mouse button, across the screen in the X and Y directions. Rotations in these angles are equivalent to incremental changes to the yaw and pitch, respectively, of the object being rendered. Both contents, namely the reference

and the distorted point clouds, are rotated identically, in order to facilitate simultaneous visualization from the same angle.

In the second step, the spatial coordinates of both lists of points are quantized to integer values, after an exhaustive iteration. As each point's spatial position is read, an extra shift can be added to the x and y coordinates. This shift, or panning, similarly to rotation and zoom level, is determined by the user, this time after dragging the mouse while holding the left mouse button, and is issued on both the reference and the distorted contents.

The color value of every point with spatial coordinates (x, y, z), as resulted from the aforementioned procedure, is associated with an image pixel $(\hat{x}, \hat{y})$ in the respective projected image. During the iteration, if another point with identical (x, y) coordinates and a smaller distance from the projection plane is identified, the first point is ignored, and the color value of the pixel $(\hat{x}, \hat{y})$ is given by the second point. In the special case where multiple points have coinciding coordinates after rotation, quantization and panning, the associated pixel value is derived as the average of the color values of the points. This procedure is repeated for every point of both the reference and the distorted point clouds, providing content projections that are finally rendered to the viewer. Unoccupied pixels in the rendered images are given a default value of (127, 127, 127) in the RGB color space, which corresponds to neutral gray.
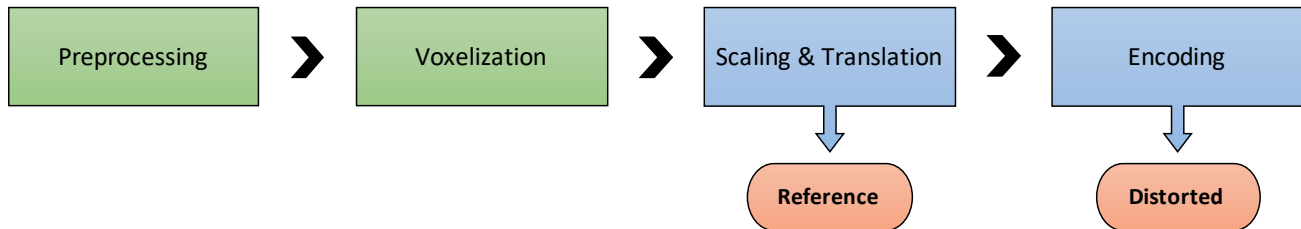
2.2 Content Selection and Preparation

Figure 2. Pre-visualization processing workflow.

A total of 7 static contents with rather diverse characteristics were selected for experiments. In particular, both human bodies and inanimate objects were considered, each having different levels of geometry and color details. The *longdress_vox10_1300* (from now on cited as *longdress*), *loot_vox10_1200* (*loot*), *redandblack_vox10_1550* (*redandblack*),[10] and *statue_Klimt* contents were chosen from the MPEG repository* and belong to the first class. The *romanoillamp*[11] and *biplane* models were selected from the JPEG repository†, while the *amphoriskos*[12] point cloud was found in the online platform Sketchfab‡. Such point clouds are typically acquired when objects are scanned by sensors that provide either directly or indirectly a cloud of points representing their 3D shape. Typical use cases involve applications where the models are consumed from the outside, and viewers may interact with them by virtual rotations in a desktop set up, or by moving around them in the physical world, assuming an augmented or a virtual reality scenario.

A point cloud compression scheme was applied on the selected objects, and the resulting stimuli were assessed by test subjects. Before encoding, the test contents were prepared based on the work-flow indicated in Figure 2, in order to reduce a number of influencing factors in our experiments, namely, number of points and geometric structure. Please note that with green color we annotate preparation steps that were issued on a subset of the test contents, whilst with blue color we specify processing that was enabled on the whole data set. Below, we provide implementation details for every stage of our content preparation approach.

Pre-processing: This step was enabled in order to ensure similar number of points for every test content. Specifically, *biplane* is provided in multiple versions that correspond to different scans. In this experiment,

---

(a) *amphoriskos*   (b) *biplane*   (c) *longdress*   (d) *loot*

(e) *redandblack*   (f) *romanoillamp*   (g) *statue_Klimt*

Figure 3. Reference test contents.

we used a combined version that provides a fully reconstructed model (i.e., *1x1_Biplane_Combined_000*), which consists of approximately $106 \cdot 10^6$ points. To reduce this number to acceptable limits, we applied subsampling using the CloudCompare[13] software, by setting a maximum allowed distance between nearest neighbors equal to 0.009. Another content that was pre-processed was *amphoriskos*. In particular, the original model was represented by approximately $200 \cdot 10^3$ points. To increase its resolution, we initially applied the Poisson Surface Reconstruction algorithm,[14] as implemented in CloudCompare, using default configurations and 1 samples per node. Moreover, the original normal vectors that were associated to the coordinates of this point cloud were employed. From the reconstructed mesh, $1 \cdot 10^6$ points were sampled by randomly picking a given number on each triangle, using the same software. For the rest of the selected contents, no pre-processing was applied.

Voxelization: With this operation, we ensure a regular-spaced geometric structure for the point clouds, in order to avoid biases that may be introduced by either the encoder or the proposed renderer. In particular, given that a subset of our data set (i.e., human bodies) was already voxelized, we converted the continuous geometric representations of the rest of the contents (i.e., inanimate objects) into sets of voxels that lie in three-dimensional lattices of 10-bit depth, to remove this influencing factor (i.e., arbitrary geometric structure) from our results.

Scaling & Translation: This step ensures that the geometry of both the reference and the distorted contents lies in the same dynamic range. Specifically, the selected codec produces point clouds with output coordinates that are proportionally displaced in the range $[-0.5, 0.5]$ with respect to the input. Considering our subjective evaluation protocol that demands the simultaneous display of the reference and the distorted contents, there is need for identical dimensions in both versions. Thus, appropriate scaling on the aforementioned range was applied on every test content before encoding, and the models were translated to the origin $(0, 0, 0)$. The output of this processing step, as indicated in Figure 2, produces the reference contents of this experiment, which are depicted in Figure 3. Information about their intrinsic geometric characteristics is provided in Table 1.

Table 1. Geometric description of every reference content.

Contents:	amphoriskos	biplane	longdress	loot	redandblack	romanoillamp	statue_Klimt
Points:	828,820	773,447	857,966	805,285	757,691	636,097	482,941
Min NN:	0.000977501	0.000977516	0.00101107	0.00101936	0.00103515	0.000977516	0.000977516
Max NN:	0.00239442	0.0470835	0.00226096	0.00203872	0.00253568	0.0693761	0.0100166
X/Y/Z:	0.60/1/0.68	0.65/0.23/1	0.40/1/0.20	0.35/1/0.41	0.44/1/0.30	1/0.45/0.51	0.30/1/0.29

Encoding: This step essentially produces the distorted versions of the test contents, which were assessed by the subjects. The reference point clouds were encoded using an open source software that was used as the anchor in the recent call for proposals issued by MPEG on point cloud compression.[15] It is a typical octree-based compression scheme, with the color attributes encoded using the JPEG algorithm, after they are mapped to an image grid using a depth first tree traversal. A detailed description can be found in.[5] To obtain a wide range of impairments, 3 quality levels for geometry and 3 quality levels for color degradations were defined. Specifically, the test contents were encoded using 8-bit, 9-bit and 10-bit octree depth (OD) that correspond to low, medium, and high geometry quality, while for different levels of color fidelity the JPEG quality parameter (QP) was set to 10, 50 and 90, respectively. The rest of the encoding options were identically set to the default values of the configuration file provided with the software release.[§] The point clouds were encoded using all possible combinations of geometry and color quality levels, leading to a total of 9 degradations per content, which were assessed by every subject. The output of this processing step produces the distorted contents of this experiment, as indicated in Figure 2. In Table 2, we report the bits per input points (bpp) for every degraded model, along with the corresponding percentage of remaining points. As expected, for the same OD and QP values the distribution of bits for both geometry and color varies remarkably depending on the content.

2.3 Equipment and Environment

The experiments were conducted in two test laboratories across two different countries, namely, MMSPG at EPFL in Lausanne, Switzerland and EED at UNB in Brasilia, Brazil. In both cases, a typical desktop setup involving an Apple Cinema Display of 27-inches and 2560x1440 resolution (Model A1316) was installed. The subjects visualized the contents under assessment through the renderer described in Section 2.1, and interacted with them by zooming, rotation, and translation using the mouse. Radio buttons that correspond to the selected grading scale were also provided in the GUI, and to submit a score, the subjects had to select the option that would correspond to their judgement.

Regarding the environmental conditions, in MMSPG, the experiments were conducted in a room that fulfills the ITU-R Recommendation BT.500-13[16] for subjective evaluation of visual data representations. Specifically, the room is equipped with neon lamps of 6500 K color temperature, while the color of the walls and the curtains is mid gray. The brightness of the screen was set to 120 cd/m^2 with a D65 white point profile, while the lighting conditions were adjusted for ambient light of 15 lux, as was measured next to the screen, according to the ITU-R Recommendation BT.2022.[17] In UNB, the test room was isolated with no exterior light affecting the assessment. The wall color was white, and the lighting conditions involved two fluorescent lamps of 4000 K color temperature.

2.4 Subjective Evaluations

The simultaneous double-stimulus impairment (DSIS) with 5-grading scale (5: *Imperceptible*, 4: *Perceptible, but not annoying*, 3: *Slightly annoying*, 2: *Annoying*, 1: *Very annoying*) was selected for its high accuracy and consistency in subjective quality assessment of point clouds;[3] thus, the reference and the distorted contents were visualized side-by-side by subjects, while being clearly annotated. In Figure 4, a screenshot of the visualization of a pair of stimuli in our GUI is indicatively presented. In order to remove contextual effects, the position of the reference was randomly selected for every subject, and remained fixed across the whole session. Thus, for half of the subjects, the position of the reference was set at the right side of the screen with the distorted content placed

---

[§]https://github.com/cwi-dis/cwi-pcl-codec

Table 2. Output points, geometry, and color bpp, for every encoded test content.

Content	Octree-depth	Percentage of remaining points	Geometry bpp	Color bpp		
				$QP = 10$	$QP = 50$	$QP = 90$
amphoriskos	*OD = 08*	16.61%	0.400	0.078	0.234	0.652
	OD = 09	53.92%	1.561	0.188	0.612	1.764
	OD = 10	100%	5.006	0.301	1.004	2.889
biplane	*OD = 08*	8.04%	0.142	0.069	0.191	0.430
	OD = 09	32.69%	0.618	0.209	0.686	1.623
	OD = 10	100%	2.890	0.589	2.101	4.926
longdress	*OD = 08*	7.76%	0.169	0.047	0.134	0.358
	OD = 09	29.63%	0.649	0.125	0.414	1.178
	OD = 10	100%	2.520	0.347	1.169	3.423
loot	*OD = 08*	7.84%	0.173	0.034	0.078	0.210
	OD = 09	29.99%	0.662	0.073	0.213	0.636
	OD = 10	100%	2.556	0.182	0.561	1.716
redandblack	*OD = 08*	8.13%	0.182	0.039	0.093	0.258
	OD = 09	31.09%	0.699	0.084	0.249	0.773
	OD = 10	100%	2.694	0.199	0.632	2.037
romanoillamp	*OD = 08*	12.14%	0.282	0.055	0.159	0.447
	OD = 09	42.47%	1.059	0.136	0.491	1.488
	OD = 10	100%	3.827	0.289	1.124	3.492
statue_Klimt	*OD = 08*	15.00%	0.324	0.098	0.286	0.722
	OD = 09	50.56%	1.384	0.240	0.792	2.147
	OD = 10	100%	4.552	0.413	1.392	3.889

at the left, and vice versa for the other half. Furthermore, the order of the test contents under assessment was randomly picked for every session. To reduce temporal references, we intentionally avoided showing the same reference content consecutively. The subjects were free to interact with the contents under assessment without imposing any time limitation; thus, participants were able to spend as much time as needed for every individual assessment, before making their judgment. Moreover, the physical distance between the viewer and the monitor was adjusted as per each subject's preferred position.

The experiments were split in two stages: (a) the training and (b) the actual test. In the training, the subjects got familiarized with this type of visual data representation and the types of artifacts that would be assessed during the actual test. Additionally, the training served the purpose of letting participants adapt with the interaction part of our proposed framework. For this purpose, the *statue_Klimt* content was selected; thus, it was excluded from the actual tests. The training was performed identically in both test laboratories, as specific instructions and descriptions were mutually agreed and delivered by the corresponding trainers. Subjects were instructed to explicitly rate the visual quality of the degraded stimuli with respect to the reference, in terms of how annoying is for them the level of impairment.

In each session, 6 contents and 9 degradations were assessed along with a hidden reference for sanity check, leading to 60 stimuli per session. A total of 20 subjects participated in the experiments in UNB, involving 11 females and 9 males, with an average of 28 years of age. In EPFL, another 20 subjects took part, comprised of 6 females and 14 males, with an average age of 28 years old.

3. QUALITY ASSESSMENT METHODOLOGY

In this section we report the methods that were used to assess, both subjectively and objectively, the visual quality of the distorted contents. Furthermore, a description of the performance indexes employed for the benchmarking of the objective quality metrics is provided.

Figure 4. A screenshot of the GUI presenting the reference *longdress* content at the left, and its compressed version using octree-depth equal to 10 and JPEG quality parameter equal to 10.

3.1 Subjective Quality Metrics

The outlier detection algorithm defined in the ITU-R Recommendation BT.500-13[16] was separately issued on the collected subjective scores from each test laboratory, in order to exclude subjects whose ratings deviated drastically from the rest of the scores. Based on this methodology, it is initially determined whether the distribution of scores for a particular content is normal or not. Specifically, for each test content, if the kurtosis coefficient of the scores is between 2 and 4, the distribution can be considered as normal. Then, a confidence interval is defined and based on the number of occurrences of scores being outside of this range, a subject is rejected or not. Specifically, if the scores are distributed normally, for each score larger than $2 \cdot \sigma$ from the mean of the scores (upper limit) of a content i, a counter P_i is incremented. For each score smaller than $2 \cdot \sigma$ from the mean of the scores (lower limit) of a content i, a counter Q_i is incremented. In case of non-normal distributions, the upper and lower limits are set as $\sqrt{20} \cdot \sigma$ from the mean of the scores of a content. Assuming a total of N_s number of stimuli, the scores of a subject are removed if the following conditions are met:

$$\frac{\sum_{i=1}^{N_s}(P_i + Q_i)}{N_s} > 0.05 \text{ and } \left| \frac{\sum_{i=1}^{N_s}(P_i - Q_i)}{\sum_{i=1}^{N_s}(P_i + Q_i)} \right| < 0.3. \tag{2}$$

No outliers were identified, thus, leading to 20 ratings per test content at every lab which is denoted by k, with $k \in \{A, B\}$. Then, for every stimulus, the Mean Opinion Score (MOS) was computed on the collected scores based on Equation 3

$$\text{MOS}_j^k = \frac{\sum_{i=1}^{N} m_{ij}}{N} \tag{3}$$

where $N = 20$ and m_{ij} is the score by subject i for the stimulus j. Furthermore, for every test content, the 95% confidence interval (CI) of the estimated mean was computed assuming a Student's t-distribution, based on Equation 4

$$\text{CI}_j^k = t(1 - \alpha/2,\ N - 1) \cdot \frac{\sigma_j}{\sqrt{N}} \tag{4}$$

where $t(1 - \alpha/2,\ N - 1)$ is the t-value corresponding to a two-tailed Student's t-distribution with $N - 1$ degrees of freedom and a significance level $\alpha = 0.05$, and σ_j is the standard deviation of the scores for content j.

3.2 Objective Quality Metrics

Objective quality assessment in point cloud representations is typically performed by full reference metrics, which can be distinguished in two main categories: (a) point-based, and (b) projection-based.

Point-based metrics: The state-of-the-art point-based methods that assess the geometry of a content can be distinguished in four classes:[18] (i) point-to-point, (ii) point-to-plane, (iii) plane-to-plane and (iv) point-to-mesh. The point-to-point metrics are based on Euclidean distances of associated points between the reference and the content under assessment; the error value essentially reflects the geometric displacement from the reference position. The point-to-plane metrics[19] are based on the projected error of a point that belongs to a content under assessment to the normal vector of a reference point; the error value indicates the deviation of the point from the reference local surface. The plane-to-plane metrics[20] are based on the angular similarity of tangent planes that correspond to associated points between the reference and the content under assessment; the error value provides an approximation of the dissimilarity between the corresponding local surfaces. The point-to-mesh metrics are based on projected distances between a content under assessment and the reconstructed reference object; however, they are considered sub-optimal, as the objective scores strongly depend on the selected surface reconstruction algorithm. Thus, these metrics will not be considered in this study. Using any of these point-based approaches, to obtain a value that represents the total geometric degradation of an impaired content, the Mean Squared Error (MSE), the Root-Mean-Squared (RMS), or the Hausdorff distance are commonly used. Finally, to account for differently scaled contents, the geometric Peak-Signal-to-Noise-Ratio (PSNR) is proposed, which, in this study, is defined as the ratio of the squared maximum distance of nearest neighbors of the original content divided by the squared error value.

The state-of-the-art point-based methods that assess the color of a distorted model are based on conventional formulas that are employed in two-dimensional content representations. In particular, points that belong to the content under assessment are associated to points that belong to the reference model, using typically the nearest neighbor algorithm, and the total color degradation value is computed by using common formulas. In this study, the well-known PSNR is used, after converting the default RGB color values associated to the coordinates of the points to the YUV color space using the ITU-R Recommendation BT.709-3.[21] Then, a weighted average on the luma and color components is realized based on Equation 5.[22]

$$\text{PSNR}_{\text{YUV}} = \left(6 \cdot \text{PSNR}_{\text{Y}} + \text{PSNR}_{\text{U}} + \text{PSNR}_{\text{V}}\right)/8 \tag{5}$$

For both geometry and color degradations, the symmetric error is commonly calculated. The symmetric error is obtained by setting both the original and the distorted content as a reference, computing both error values, and keeping the maximum.

To compute the objective scores, the software ver. 0.12[23, 24] was employed for point-to-point, point-to-plane, and color distortions metrics. For the latter, the software, essentially, provides the PSNR_{Y}, PSNR_{U}, and PSNR_{V} values, which were then combined using Equation 5 to derive a total color degradation score. For point-to-point and point-to-plane metrics, the geometry degradation values were based on the MSE and the Hausdorff distance. The plane-to-plane metrics were computed using the software release ver. 1.0[¶]. Since both the reference and the distorted contents had in principle no associated normals assigned to their coordinates, the methodology proposed by Hoppe et al.[25] was selected for normal estimation as implemented in Point Cloud Library (PCL).[26] This method is based on best fitting planes in the least-square sense, issued on local neighborhoods around every point of interest. In our case, we selected a neighborhood of 12 nearest points.

[¶]https://github.com/mmspg/point-cloud-angular-similarity-metric

Projection-based metrics: In the projection-based approaches, the 3D point cloud contents are mapped onto conventional two-dimensional planes by orthographic projection. It is desirable that projections cover as much area as possible, thereby incorporating more information from the object. While any number of points of view (and projection planes) can be selected, using too many projections is both impractical and unnecessary, as the same area can be covered by an arbitrary number of planes, while cost grows linearly. We propose using a total of 6 projected planes, which can be interpreted as setting a bounding cube around the object under analysis on whose faces the content is projected. This ensures that when voxels are orthographically projected onto each plane, there is a one-to-one correspondence between voxels and pixels in the projected images, and no further processing is required.

After the projected images are obtained, any 2D image metric can be used to compare the projected images of the content under assessment with their respective image pair of the projected reference content. We have investigated the performance of several 2D metrics applied to the projected images, as implemented in the VQMT software.[27] In particular, PSNR, PSNR-HVS, PSNR-HVS-M, SSIM, MSSIM and VIFP are employed in our test. For each metric, one score for each of the 6 pairs of images is obtained, and then the average across the projected planes is computed.

3.3 Comparison of Objective against Subjective Quality Metrics

To evaluate how well an objective quality metric is able to predict the perceptual visual quality of a test content, the human ratings are commonly set as the ground truth and compared with predicted MOS values that correspond to objective scores obtained from a particular metric. Specifically, let us assume that the result of executing an objective metric provides a Point cloud Quality Rating (PQR). A predicted MOS value, denoted as MOS_P, that is assigned to a specific distorted content, is estimated after regression analysis on each [PQR, MOS] pair. In our case, we used the monotonic cubic function as a regression model, and the predicted MOS was computed based on Equation 6

$$MOS_p = a \cdot x^3 + b \cdot x^2 + c \cdot x + d \tag{6}$$

where a, b, c and d were determined using a least squares method. Then, following the Recommendation ITU-T P.1401,[28] the Pearson linear correlation coefficient (PCC), the Spearman rank order correlation coefficient (SROCC), the root-mean-square error (RMSE), and the outlier ratio based on standard error (OR) were computed between MOS and MOS_P, to assess the linearity, monotonicity, accuracy and consistency of the predictions of perceptual quality offered by every objective quality metric.

4. RESULTS

In this section we analyse the subjective scores collected in our experiments, followed by the benchmarking of the state-of-the-art objective quality metrics.

4.1 Analysis of the Subjective Scores

The subjective results of the 6 contents that were involved in the tests after outlier detection are shown in Figure 5, with the caption of each sub-figure indicating the test laboratory from which they were collected. Specifically, we provide the MOS along with the CI for every test content in form of histograms, clustered per type of degradation. The naming convention is as follows: ODXX_QPYY, where OD and QP stand for octree-depth and JPEG quality parameter, respectively, while XX $\in$ {08, 09, 10} and YY $\in$ {10, 50, 90} denote the geometry and color quality levels.

Based on the results indicated in Figure 5, the subjective ratings vary per type of degradation for the same content. In particular, it is obvious that for the sparsest versions (i.e., OD = 08), the mean score is increasing slowly as the color quality level is getting better, independently of the model. Higher rates of increase are observed as the geometry resolution becomes higher. This outcome essentially indicates that, when the geometry resolution of a content remains low, the overall perceptual quality is severely affected, regardless of color improvements. It is important to note that, in the proposed renderer, the absence of geometry details is expressed by the presence of missing pixels. In particular, following the description of Section 2.1, a single voxel
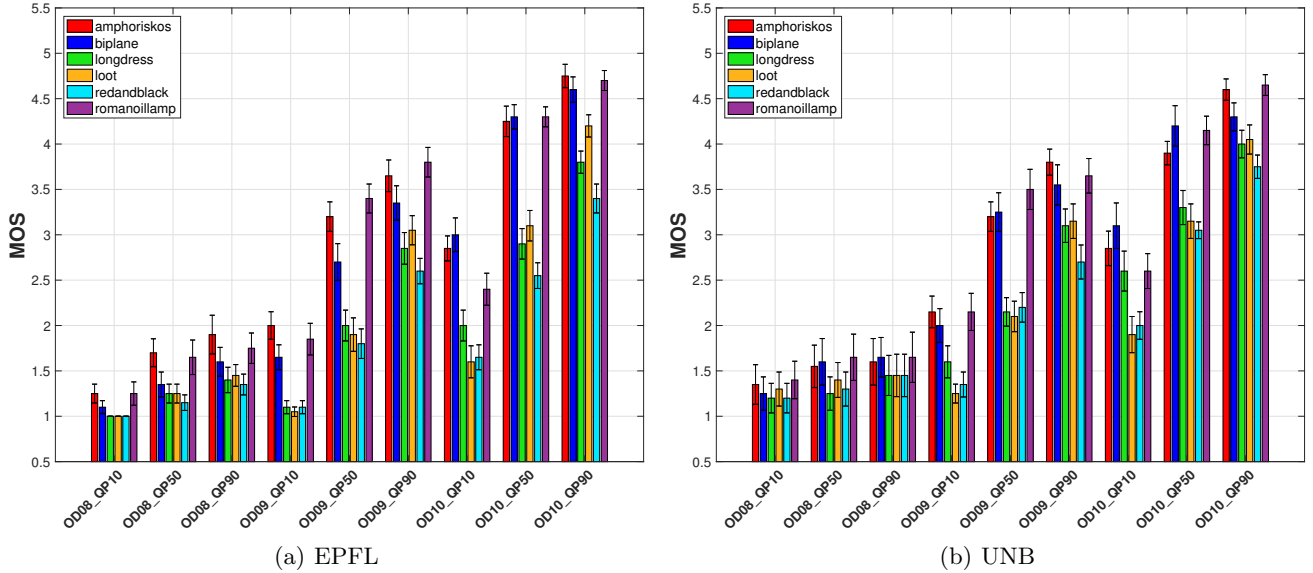
Figure 5. Subjective ratings for every content against degradations.

is projected in a limited neighborhood of pixels, as a function of the zooming applied by user interaction, in order to reduce blurring effects on the reference stimuli. As such, missing color information is mainly observed in contents encoded at an octree-depth of 8, while for octree-depth equal to 9, such artifacts become visible only when a viewer inspects the object from very close virtual distances after zooming. Another reason that explains the rating behaviors of contents with lowest geometry quality levels is the usage of the octree structure as basis for point cloud compression. In particular, by reducing the geometry resolution of an octree, an increasing number of points that belong to the original model naturally fall within the leaf nodes. Considering that the color of a representative output point is given by blending the colors of input points that are placed in the same leaf node, color degradations are by default bound by high geometry degradation.

Another conclusion that can be drawn based on Figure 5 is that, for a specific type of degradation, the perceptual quality notably differs depending on the content. In fact, subjects seem to be more critical with point clouds that represent humans, when compared to point clouds that represent inanimate objects. Smaller rating deviations are observed between contents that belong to the same class (i.e., humans, or objects), indicating that similar rating behavior can be observed within the groups.

Furthermore, by inspecting the total bitrates of the encoded contents, as reported in Table 2, and after comparison with the corresponding scores indicated in Figure 5, we conclude that higher bitrates do not necessarily lead to better visual quality. For instance, for every test content, subjects from both test laboratories showed their clear preference in the combination of best color quality (i.e., QP = 90) with medium geometry (i.e., OD = 09), when compared to best geometry quality level (i.e., OD = 10) with the worst color quality (i.e., QP = 10); the latter combination requires higher bitrates for every model. Although the obtained bpp values are obviously codec dependent, such observations suggest that savings may be achieved by appropriate allocation of bits between geometry and color. To verify these observations, a one-tail t-test at a 5% significance level was issued on the collected data, separately, per test laboratory. The null hypothesis assumes that a mean score, obtained by averaging the ratings over all contents for a particular color and geometry level, is the same with the average score of all contents for any another combination of degradations. The results are presented in Figure 6. It can be seen that based on the subjective scores obtained in EPFL, the combinations OD09_QP50 and OD09_QP90 are preferred 1 and 5 times in a total of 6 contents, against the combination OD10_QP10. Based on the UNB scores, the same trend is observed for these particular cases with 2 and 5 times, respectively, while similar preferences are obtained even for lower geometry levels, although not that strongly (i.e., OD08_QP50 and OD08_QP90 was preferred 1 and 3 times against OD09_QP10).

Finally, rating variations are noticed between subjects from different test laboratories. Based on Figure 5, it
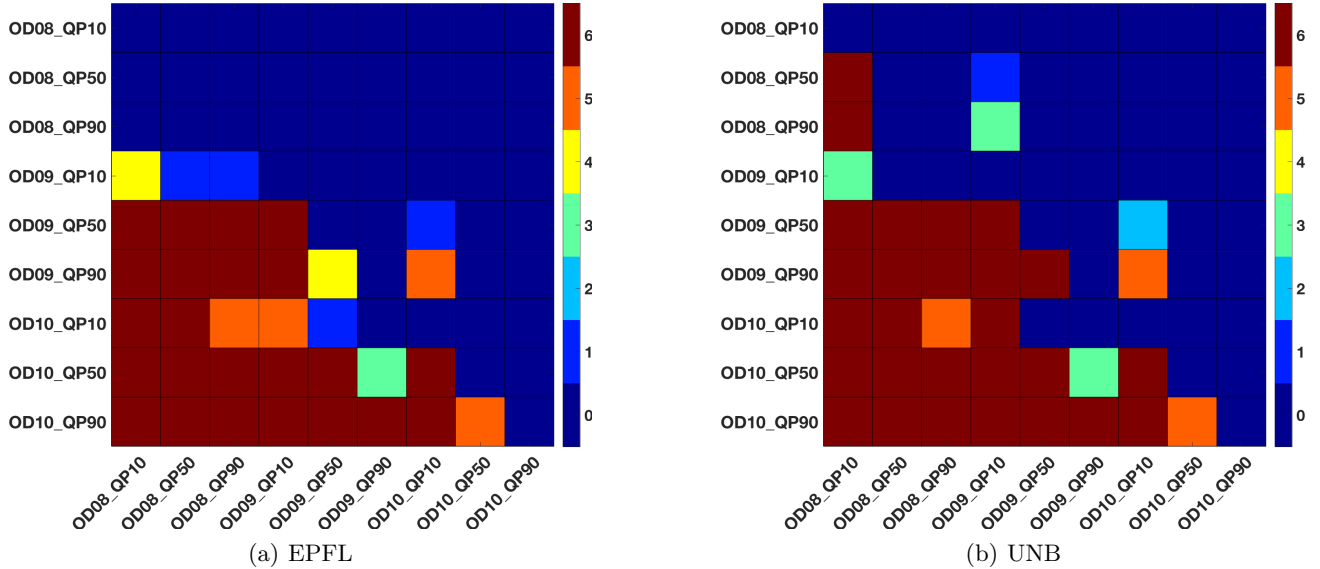
Figure 6. Significance difference matrices at a 5% level, indicating the preference of subjects on a particular degradation against all others, considering the whole data set.

seems that the subjects in EPFL were in principle more critical with highest color degradations (i.e., QP = 10), especially for point clouds that represent human bodies, with respect to the participants of UNB. To examine whether statistical differences on the rating behaviors of the subjects of each laboratory per test content are obtained, in Figure 7, we provide the results of a one-tail t-test at a 5% significance level. In this case, the null hypothesis suggests that the MOS values, as computed from the scores of the two labs, are equal for each degraded content. Our results show that, when the null hypothesis is rejected, it is because the subjects in EPFL give lower quality scores. This scoring difference is mainly observed for lower and medium color quality levels. The only exception is noted for *amphoriskos* with lowest geometry and medium color quality, which was rated lower by UNB. It is also worth remarking that the content *redandblack* obtained higher ratings in UNB with respect to EPFL, for most of the cases.

These observations are verified by the results of a multi-way ANOVA performed on the scores, which are reported in Table 3. The obtained *p*-values suggest that every influencing factor, namely, test laboratory (i.e., EPFL and UNB), type of content (i.e., human bodies and inanimate objects), geometry (i.e., OD = {08, 09, 10}) and color (QP = {10, 50, 90}) quality levels, leads to ratings that are statistically different at a 5% significance level.

Table 3. Multi-way ANOVA.

Source	‖	SS	DF	MS	F	p
Test laboratory	‖	8.82	1	8.817	13.92	0.0002
Content type	‖	249.42	1	249.424	393.81	0
Geometry degradation	‖	1460.84	2	730.422	1153.24	0
Color degradation	‖	618.86	2	309.429	488.55	0
Error	‖	1363.64	2153	0.633		
Total	‖	3701.58	2159			

4.2 Benchmarking of Objective Quality Metrics

In this section we provide benchmarking results for the state-of-the-art projection-based and point-based objective quality metrics. Please note that, although the geometry PSNR values for point-to-point and point-to-plane metrics were computed considering both MSE and Hausdorff distances, their performance was found to be
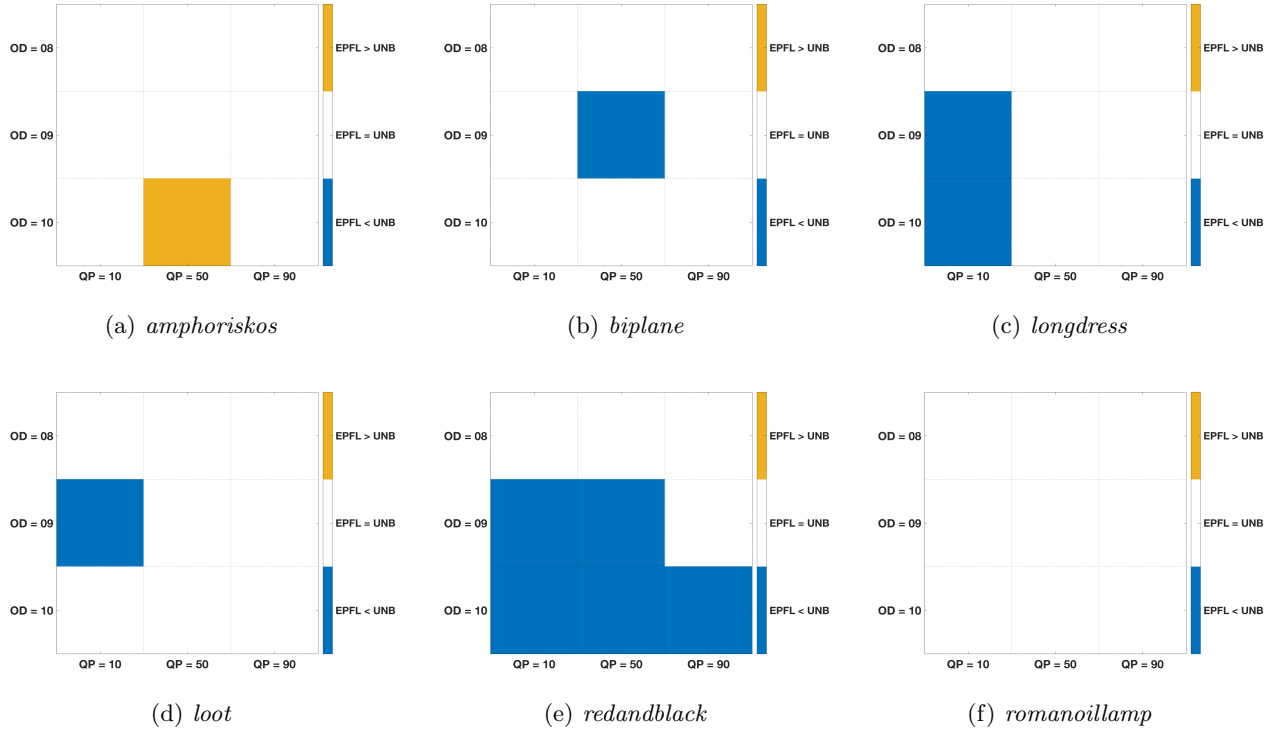
Figure 7. Significance difference matrices at a 5% level, indicating if subjects of one lab perceived significantly higher or lower the visual quality of a particular degradation with respect to the other lab, for every test content.

worse. Thus, to reduce the load, they were not reported. In the tables and figures of this section we use the abbreviations po2point, po2plane, and pl2plane to indicate the point-to-point, point-to-plane, and plane-to-plane metrics, respectively, for spacing reasons.

Considering that the subjective scores obtained from the two test laboratories are statistically distinguishable, as shown in Section 4.1, the benchmarking is conducted against the two sets of scores separately. Furthermore, given that the subjective scores were found to be statistically different per type of content, the same analysis is repeated on 3 sets of contents: (i) the whole data set, (ii) a subset of point clouds that represents human bodies, and (ii) a subset of point clouds that represents inanimate objects.

In Table 4, we provide the performance indexes for every objective quality metric that was tested against the ground truth subjective scores collected in EPFL, for every set of contents individually. In general, it is evident that the projection-based objective quality metrics perform better than the point-based. The best-performing metric on the whole data set was found to be VIFP, although the correlation remains weak. The correlation is remarkably improved when the objective metrics are benchmarked with the subjective ratings obtained for inanimate objects and human bodies, separately. In particular, both MSSIM and VIFP attain high predictive power in both sets of contents, with the first being better in the case of objects, while the second being superior in the case of human models.

In Figures 8 and 9, we present scatter plots indicating the correlation of the best-performing projection-based and point-based approaches, for the 3 sets of contents. It is obvious that the point-based approaches are limited by the fact that they either examine the geometry or the color degradation of a content. In particular, in Figure 9, it is clear that every content is associated with a predicted MOS based on the octree-depth, and although the subjective scores are increasing as the color quality is improved, the objective scores remain identical. On the contrary, the projection-based methods attain higher discrimination power and are able to better predict the visual quality perceived by subjects.

In Table 5, and Figures 10, 11 the correlation between the objective quality metrics and the subjective

Table 4. Benchmarking of objective quality metrics using subjective scores from EPFL as ground truth.

	Metric	All				Inanimate objects				Human bodies			
		PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Projection-based	PSNR	0.520	0.497	0.981	0.741	0.797	0.786	0.735	0.630	0.744	0.739	0.633	0.704
	PSNR-HVS	0.570	0.564	0.943	0.741	0.845	0.841	0.650	0.630	0.797	0.773	0.572	0.667
	PSNR-HVS-M	0.601	0.585	0.918	0.741	0.866	0.851	0.609	0.593	0.822	0.795	0.539	0.667
	SSIM	0.494	0.497	0.998	0.778	0.873	0.838	0.593	0.704	0.847	0.815	0.503	0.630
	MSSIM	0.677	0.682	0.845	0.685	**0.929**	**0.934**	**0.451**	**0.556**	0.814	**0.861**	0.550	0.667
	VIFP	**0.754**	**0.717**	**0.754**	**0.648**	0.906	0.932	0.516	0.593	**0.905**	**0.861**	**0.402**	**0.519**
Point-based	po2point$_{MSE}$	0.672	0.597	0.850	0.667	**0.795**	0.822	**0.738**	**0.630**	0.651	0.702	0.719	0.704
	po2point$_{Hausdorff}$	**0.683**	**0.725**	**0.839**	**0.648**	0.793	**0.824**	0.741	**0.630**	0.651	0.707	0.719	0.704
	po2plane$_{MSE}$	0.656	0.598	0.866	0.704	0.763	0.755	0.786	0.741	0.637	0.689	0.730	0.741
	po2plane$_{Hausdorff}$	**0.683**	0.686	**0.839**	**0.648**	0.792	0.778	0.743	0.667	0.652	0.686	0.718	0.741
	pl2plane$_{RMS}$	0.679	0.676	0.843	0.759	0.707	0.702	0.861	0.778	**0.756**	0.653	**0.620**	0.630
	pl2plane$_{MSE}$	0.675	0.676	0.847	0.759	0.662	0.753	0.912	0.852	0.701	**0.715**	0.676	**0.593**
	Color - PSNR$_{YUV}$	0.539	0.491	0.967	0.833	0.669	0.753	0.904	0.852	0.702	**0.715**	0.675	**0.593**

(a) Total set of contents.   (b) Inanimate objects.   (c) Human bodies.
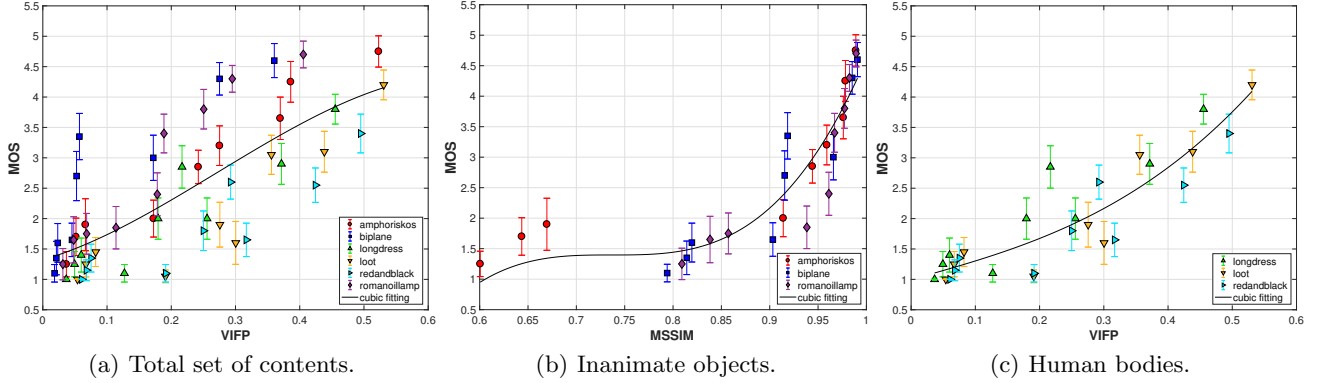
Figure 8. Best performing projection-based metric on EPFL subjective scores.

(a) Total set of contents.   (b) Inanimate objects.   (c) Human bodies.
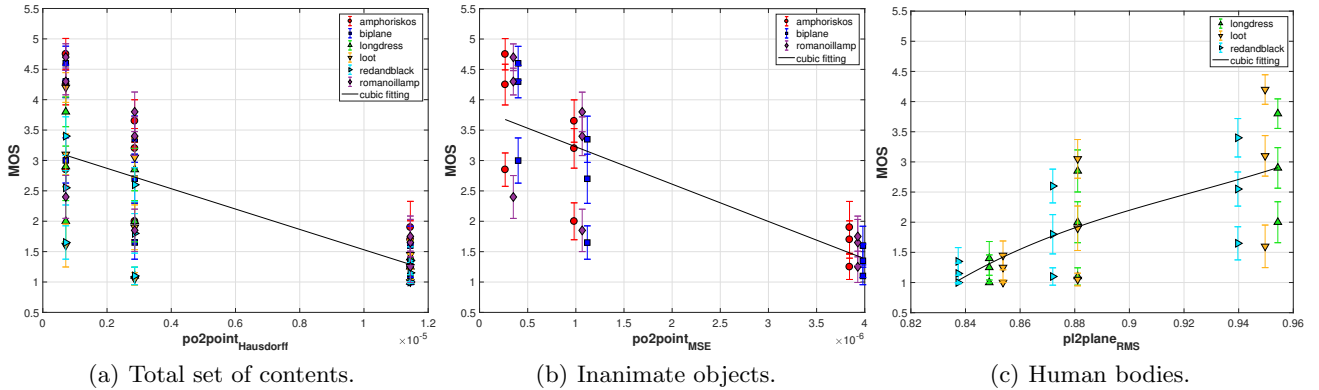
Figure 9. Best performing point-based metric on EPFL subjective scores.

scores obtained by UNB is indicated. It is clear that the projection-based metrics are superior to point-based algorithms using this set of subjective scores. The MSSIM and VIFP metrics were found to outperform the rest, independently of the set of contents they were applied to. The RMSE and OR coefficients remain lower with respect to the values achieved with the EPFL scores. This phenomenon can be explained considering that larger CIs were obtained with UNB scores. In fact, the average CI of the UNB scores was 26.95% larger with respect to the average CI of the EPFL scores.

Table 5. Benchmarking of objective quality metrics using subjective scores from UNB as ground truth.

	Metric	All				Inanimate objects				Human bodies			
		PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR	PCC	SROCC	RMSE	OR
Projection-based	PSNR	0.582	0.545	0.874	0.667	0.799	0.794	0.683	0.481	0.756	0.747	0.616	0.444
	PSNR-HVS	0.623	0.608	0.840	0.648	0.835	0.850	0.625	0.519	0.805	0.783	0.558	0.407
	PSNR-HVS-M	0.652	0.629	0.814	0.630	0.853	0.862	0.592	0.444	0.830	0.806	0.524	0.444
	SSIM	0.566	0.570	0.886	0.667	0.880	0.893	0.539	0.593	0.865	0.831	0.471	0.370
	MSSIM	0.739	0.738	0.724	0.537	**0.940**	**0.961**	**0.389**	**0.222**	0.859	0.886	0.482	0.370
	VIFP	**0.784**	**0.740**	**0.667**	**0.519**	0.877	0.884	0.545	0.444	**0.919**	**0.890**	**0.370**	**0.296**
Point-based	po2point$_{MSE}$	0.747	0.652	0.714	0.556	0.843	0.792	0.610	**0.481**	0.728	0.758	0.645	0.519
	po2point$_{Hausdorff}$	0.757	**0.775**	0.702	0.537	**0.844**	**0.839**	**0.609**	**0.481**	0.728	0.757	0.645	0.519
	po2plane$_{MSE}$	0.736	0.670	0.727	**0.500**	0.824	0.798	0.643	0.519	0.713	0.740	0.659	0.556
	po2plane$_{Hausdorff}$	**0.758**	0.749	**0.701**	0.537	**0.844**	0.806	0.610	0.481	0.730	0.762	0.643	0.519
	pl2plane$_{RMS}$	0.520	0.461	0.918	0.815	0.654	0.596	0.859	0.741	0.685	0.607	0.685	0.593
	pl2plane$_{MSE}$	0.666	0.664	0.801	0.704	0.629	0.678	0.882	0.778	0.771	**0.781**	0.599	**0.444**
	Color - PSNR$_{YUV}$	0.672	0.664	0.795	0.704	0.629	0.678	0.883	0.778	**0.773**	**0.781**	**0.597**	**0.444**

(a) Total set of contents.   (b) Inanimate objects.   (c) Human bodies.
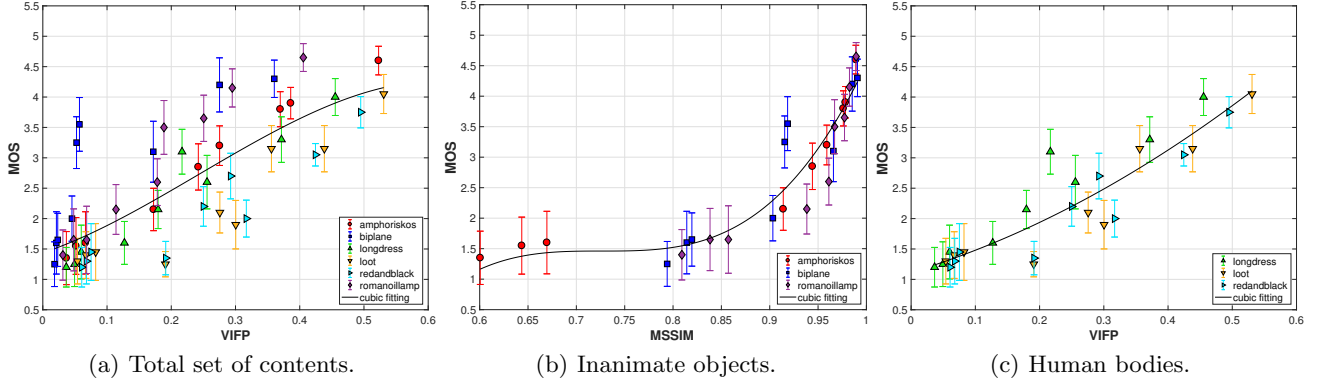
Figure 10. Best performing projection-based metric on UNB subjective scores.

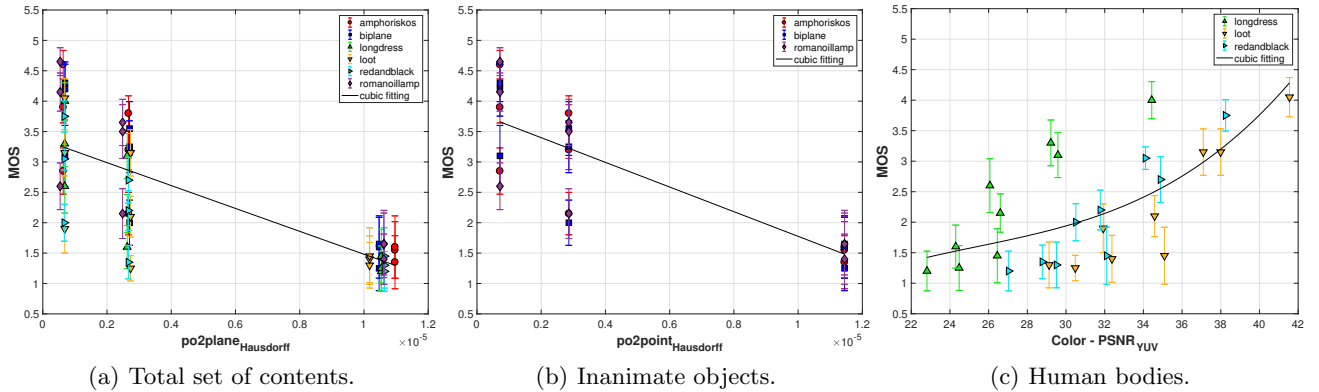(a) Total set of contents.   (b) Inanimate objects.   (c) Human bodies.

Figure 11. Best performing point-based metric on UNB subjective scores.

A general remark of this analysis is that higher accuracy is obtained for point clouds that represent inanimate objects. It is known that subjects tend to rate people in a different way with respect to objects, as they are more sensitive to degradations that affect facial characteristics of the displayed images. Such differences are not taken under consideration by the objective quality metrics, thus leading to different rating behaviors. This is confirmed by the results of the ANOVA issued on the total set of scores, reported in Table 3, which shows that the scores are not statistically equivalent.

It is noteworthy that, although an interactive approach was selected for the subjective evaluation protocol, the objective metric is only computed on the 6 orthographic projections of the initial view that is displayed before the interaction. However, the 6 projections seem to be sufficient to predict the visual quality of the content. Further improvements in terms of correlation may be achieved by considering more viewing angles, and more virtual distances between the content and the subject. Additional refinements may be observed by excluding the natural grey background of the projected images from the computations. In fact, the gray background could be the reason why we observe high differences in the objective scores between different contents, although the subjective scores are very similar (i.e., *amphoriskos* and *romanoillamp* encoded at OD = 08 in Figure 8 (b) and 10 (b)). Finally, a natural extension of our proposed framework would be the real-time and continuous computation of an objective score, based on the projected view the observer is inspecting while interacting with the contents under assessment. We aim to incorporate such extensions and improvements in our framework as a future work.

5. CONCLUSIONS

In this study, a quality assessment methodology for point clouds is proposed. The 3D volumes are voxelized and orthographically projected on 2D image planes in real-time. The 2D images are visualized by subjects using a software implementation that was developed by the authors. In this testbed, the usage of conventional 2D image quality metrics on a set of 6 viewing angles was tested, and found to be superior than the state of the art, using as validation subjective evaluations that were conducted in two test laboratories. Our approach incorporates geometric distortions, color degradations, and the rendering approach. Despite its simplicity, high predictive power is achieved, providing an adequate alternative for quality assessment of point clouds.

ACKNOWLEDGMENTS

REFERENCES

[1] Perry, S., "JPEG Pleno Point Clouds – Use Cases and Requirements." ISO/IEC JTC 1/SC29/WG1 input document M80043 (July 2018).

[2] Zhang, J., Huang, W., Zhu, X., and Hwang, J. N., "A subjective quality evaluation for 3D point cloud models," in [*2014 International Conference on Audio, Language and Image Processing*], 827–831 (July 2014).

[3] Alexiou, E. and Ebrahimi, T., "On the performance of metrics to predict quality in point cloud representations," in [*Proceedings of SPIE*], *Applications of Digital Image Processing XL* **10396** (August 2017).

[4] Alexiou, E., Upenik, E., and Ebrahimi, T., "Towards subjective quality assessment of point cloud imaging in augmented reality," in [*2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*], 1–6 (October 2017).

[5] Mekuria, R., Blom, K., and Cesar, P., "Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video," *IEEE Transactions on Circuits and Systems for Video Technology* **27**, 828–842 (April 2017).

[6] Javaheri, A., Brites, C., Pereira, F., and Ascenso, J., "Subjective and objective quality evaluation of compressed point clouds," in [*2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*], 1–6 (October 2017).

[7] Javaheri, A., Brites, C., Pereira, F., and Ascenso, J., "Subjective and objective quality evaluation of 3D point cloud denoising algorithms," in [*2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*], 1–6 (July 2017).

[8] Alexiou, E., Bernardo, M. V., da Silva Cruz, L. A., Dmitrovic, L. G., Duarte, R., Dumic, E., Ebrahimi, T., Matkovic, D., Pereira, M., Pinheiro, A., and Skodras, A., "Point cloud subjective evaluation methodology based on 2D rendering," in [*2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*], (May 2018).

[9] Mekuria, R., Laserre, S., and Tulvan, C., "Performance assessment of point cloud compression," in [*2017 IEEE Visual Communications and Image Processing (VCIP)*], 1–4 (Dec. 2017).

[10] d'Eon, E., Harrison, B., Myers, T., and Chou, P. A., "8i Voxelized Full Bodies, version 2 – A Voxelized Point Cloud Dataset." ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m40059/M74006 (January 2017).

[11] Zuffo, M., "U. S. Paulo Point Cloud Database description." ISO/IEC JTC 1/SC29/WG1 input document M79100 (April 2018).

[12] "Amphoriskos - clean point cloud." https://sketchfab.com/models/85cba491e0a84ce58dc4a75715073ad2.

[13] "CloudCompare." http://www.danielgm.net/cc/.

[14] Kazhdan, M. and Hoppe, H., "Screened Poisson Surface Reconstruction," *ACM Trans. Graph.* **32**, 29:1–29:13 (July 2013).

[15] MPEG 3DG and Req., "Call for proposals for Point Cloud Compression V2." ISO/IEC JTC1/SC29/WG11 input document MPEG2017/N16763 (April 2017).

[16] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures." International Telecommunications Union (January 2012).

[17] ITU-R BT.2022, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays." International Telecommunications Union (August 2012).

[18] Alexiou, E. and Ebrahimi, T., "Benchmarking of objective quality metrics for colorless point clouds," in [*2018 Picture Coding Symposium (PCS)*], 1–5 (June 2018).

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

[20] Alexiou, E. and Ebrahimi, T., "Point cloud quality assessment metric based on angular similarity," in [*2018 IEEE International Conference on Multimedia and Expo (ICME)*], 1–6 (July 2018).

[21] ITU-R BT.709-6, "Parameter values for the HDTV standards for production and international programme exchange." International Telecommunications Union (June 2015).

[22] Ohm, J.-R., Sullivan, G. J., Schwarz, H., Tan, T. K., and Wiegand, T., "Comparison of the Coding Efficiency of Video Coding Standards-Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Cir. and Sys. for Video Technol.* **22**(12), 1669–1684 (2012).

[23] Tian, D., Ochimizu, H., Feng, C., Cohen, R., and Vetro, A., "Evaluation Metrics for Point Cloud Compression." ISO/IEC JTC1/SC29/WG11 input document MPEG2016/M39966 (January 2017).

[24] Tian, D., Ochimizu, H., Feng, C., Cohen, R., and Vetro, A., "Updates and Integration of Evaluation Metric Software for PCC." ISO/IEC JTC1/SC29/WG11 input document MPEG2017/M40522 (April 2017).

[25] Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W., "Surface Reconstruction from Unorganized Points," in [*Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*], *SIGGRAPH '92*, 71–78, ACM (July 1992).

[26] "Point cloud library (PCL)." http://pointclouds.org/.

[27] "VQMT: Video Quality Measurement Tool." https://mmspg.epfl.ch/vqmt.

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