Benchmarking of objective quality metrics for point cloud compression

Davi Lazzarotto, Evangelos Alexiou, Touradj Ebrahimi
École Polytechnique Fédérale de Lausanne (EPFL)
Multimedia Signal Processing Group (MMSPG)
Lausanne, Switzerland

***Abstract*—Point cloud is a promising representation format for 3D media. The vast volume of data associated with it requires efficient compression solutions, with lossy algorithms leading to larger bit-rate savings at the expense of visual impairments. While conventional encoding approaches rely on efficient data structures, recent methods have incorporated deep learning for rate-distortion optimization, while inducing perceptual degradations of different natures. To measure the magnitude of such distortions, subjective or objective quality evaluation methodologies are employed. Lately, a remarkable amount of efforts has been devoted to development of point cloud objective quality metrics, which have been reported to attain high prediction accuracy. However, their performance and generalization capabilities still haven't been evaluated in the presence of artifacts from learning-based codecs. In this study, we tackle this matter by conducting the first crowdsourcing experiment for point cloud quality reported in the literature, in order to obtain subjective ratings for point cloud models whose topology and color attributes are encoded by both conventional and data-driven methods. Using the subjective scores as ground truth, the performance of a large pool of state-of-the-art quality metrics is rigorously benchmarked, drawing useful insights regarding their efficacy.**

***Index Terms*—Point Cloud, Quality Assessment, Subjective Quality Evaluation, Objective Quality Metrics, Compression.**

I. Introduction

Point cloud has emerged as one of the main representations of 3D content for a multitude of use-cases. This imaging modality typically requires a large amount of data, which in turn demands effective lossy compression. For this task, different methods have been proposed, which may rely on different structures for data representation. The majority of available solutions make use of handcrafted techniques to increase the compression efficiency at the expense of added distortions. Recently deep learning-based encoders provide a powerful alternative, already achieving comparable performance to the state of the art.

In applications that target point cloud consumption by human end-users, particular care must be devoted to the perceptual quality of the displayed content. The latter can be assessed by either subjective or objective means. Subjective quality evaluations rely on human opinions that are recruited to judge the visual quality of degraded stimuli, denoting the most reliable methods, albeit being expensive and time consuming. Objective quality evaluations are realized by metrics that quantify signal degradations; yet, their performance in predicting perceived impairments should be benchmarked against subjective scores.

The visual artifacts introduced by different compression methods can greatly vary, which poses a challenge for objective quality metrics to correlate well with human perception. Numerous related studies have been conducted in the past, using ground-truth subjective quality scores of compressed point clouds for performance evaluation of objective metrics. In [1], static and dynamic point clouds were evaluated after compression using the codec proposed in [2]. A study on the quality assessment of models subject to graph-based and octree-based compression algorithms is reported in [3]. In [4], [5], subjective quality evaluation experiments for point clouds compressed using the codec described in [2] were conducted, under different rendering schemes. In [6], an octree-based and a projection-based encoder were evaluated by three independent laboratories using both small and large scale models. In [7], a subjective experiment was carried out for quality assessment of the state-of-the-art MPEG G-PCC [8] and V-PCC [9] test models. Quality evaluations on a volumetric video dataset was performed in [10], under compression artifacts from V-PCC. In [11], different types of degradation were assessed, including Gaussian noise in both topology and texture, octree down-sampling, and compression artifacts from the MPEG codecs. In [12], geometric artifacts from octree-based coding and the MPEG encoding engines were evaluated using different rendering strategies. Finally, in [13], the performance of MPEG codecs was assessed in four dislocated laboratories that participated in the JPEG Pleno Point Cloud Exploration Study activities.

Despite the above-mentioned efforts to benchmark objective metrics, no previous study has accounted for distortions caused by both conventional and learning-based encoders. This paper describes the design of a crowdsourcing experiment for subjective quality assessment of point clouds compressed with three standardized codecs and two recent deep learning methods. The collected subjective scores are used to evaluate the performance of well-established and newly-introduced objective quality predictors, with separate analysis issued on the entire dataset and each subset that occurs after splitting the stimuli per codec. User reliability was ensured by a post-screening

This work was supported by the Swiss National Foundation for Scientific Research (SNSF) under the grant number 200021-178854.

978-1-6654-3288-7/21/$31.00 © 2021 IEEE

(a) *guanyin* (b) *longdress* (c) *phil*

(d) *rhetorician* (e) *ricardo* (f) *soldier*

Fig. 1: Point cloud contents of the evaluated dataset.

method based on hidden reference scores. Our results show that two specific metrics are ranked among the best options across all tested cases.

The main contributions of our work can be summarised as follows:

1) To the best of our knowledge, this is the first study benchmarking objective quality metrics on subjective scores collected on point cloud models degraded with both learning-based and conventional codecs. Moreover, we evaluate the largest pool of metrics in a single experiment up to date.
2) We present the first crowdsourced subjective experiment for point cloud quality assessment, employing a post-screening method for detection of unreliable users based on hidden reference scores.
3) The subjectively annotated dataset will be released, and in case of acceptance, the corresponding link will be provided in the final version of the paper.

II. Experimental setup

A. Content selection

The point cloud models used during the evaluation session were taken from the JPEG Pleno Point Cloud Common Test Conditions (JPEG CTC) document [14]. A total of six models were retained, namely, *longdress* and *soldier*, which denote two full body models [15]; *phil* and *ricardo*, which represent two upper bodies from MVUB [15]; *rhetorician* and *guanyin*, which represent cultural heritage assets collected by the PointXR dataset [16]. The models are voxelized at a bit depth of 10, except for *phil* that has a bit depth of 9. Figure 1 illustrates the above point cloud contents.

B. Encoding engines

A total of five different codecs were recruited in this experiment, namely, V-PCC [9], two distinct configurations of G-PCC [8], and two learning-based compression schemes.

While V-PCC compresses 2D maps corresponding to projections of a point cloud, G-PCC encodes the model's topology first and then uses the decoded geometry to compress color attributes. Two alternatives can be used for geometry coding, namely octree and TriSoup. In this study, we select both geometry codecs in conjunction with the Lifting color-encoding module to compress the models, which are referred to as *octree-lifting* and *trisoup-lifting*.

The next algorithm, proposed in [17], is selected as a representative of geometry-only learning-based codecs. Models compressed with this geometry-only method were recolored using the nearest neighbour attribute of the reference content, and later color-encoded with the Lifting module from G-PCC. This combination is referred to, hereafter, as *geo_cnn-lifting*.

The second learning-based method [18] is a color-only encoder that maps the point cloud attributes in 2D using a learned folding operation, and then compresses the color information using already established image codecs. Among the learning-based color-only encoders, it was the only candidate with open source code. The topology of models is compressed with the octree module, being then recolored with the same process previously described for G-PCC. After folding, the 2D color maps are encoded with BPG [19]. This combination is henceforth referred to as *octree-folding*.

The configurations reported in the JPEG CTC document [14] were used for encoding with V-PCC and G-PCC. Specifically, four out of the five compression levels defined in [14] were selected, excluding the rate point with lowest quality. Regarding *geo_cnn-lifting*, the c6 configuration was adopted for geometry coding as described in [17], using the following values of λ to obtain different compression levels: $\{2\times10^{-5}, 5\times10^{-5}, 1\times10^{-4}, 3\times10^{-4}\}$. For color encoding, the same configuration as *octree-lifting* and *trisoup-lifting* was used. When encoding with the folding-based method, the point cloud models were first partitioned into patches and then, for each patch, a neural network was trained to map the color attributes into a 2D map. Each map was then encoded with one of the following quality parameters: $\{40, 35, 30, 25\}$. For geometry encoding, the same configuration as *octree-lifting* was used.

C. Rendering

For display purposes, the point cloud web renderer released in [7] was employed. Every stimulus was displayed using screen-faced splats of adaptive size, based on local sparsity. In particular, the splat size of every point was set equal to the local mean distance of its 12 nearest neighbors, if it wasn't identified as an outlier; in the latter case, the global mean distance, computed over the same neighborhood population was used instead, to avoid magnification of isolated points. Every splat was additionally multiplied by a scaling factor of

Fig. 2: Screenshot of the quality evaluation testbed. The reference *longdress* is displayed on the left and its encoded version using *octree-folding* at the lowest quality on the right.

1.05, which was determined after expert viewing. No shading was enabled in order to avoid masking of visual artifacts.

D. Testing stimuli

Video sequences illustrating different views of the rendered point clouds were formed and evaluated by human observers. The models were centered in the virtual scene of the renderer with the background color set to black. The virtual camera was placed at a fixed distance after ensuring that each model would be visible at its entirety, and was programmed to perform a complete circle around its vertical axis capturing views in steps of 1°. In every step, a snapshot was exported, leading to a total of 360 frames. Note that for *phil* and *ricardo*, only the frontal part was examined, with the 180° semi-circle traversed twice to keep the same total number of snapshots. The captured images were encoded using a visually lossless configuration (i.e., CRF equal to 18) of H.264/AVC codec in FFmpeg at 30 fps, leading to animated videos of 12 seconds duration. Each frame was of resolution 880×880 pixels, which is equal to the canvas that was employed to display the models in the renderer.

E. Test method

A simultaneous DSIS protocol was adopted, showing videos of the distorted and the corresponding reference models to the observers side-by-side. The subjects were asked to passively inspect the entire video duration before scoring the impairment of the distorted model with respect to the pristine, using the following 5-grading scale: “5 - Imperceptible”, “4 - Perceptible, but not annoying”, “3 - Slightly annoying”, “2 - Annoying” and “1 - Very annoying”. A hidden uncompressed reference was included among the testing stimuli.

F. Experiment design

The experiment was split in a training and a testing phase. During training, the subjects were instructed about the task at hand. To get familiarized with the types of artifacts on point cloud contents, two distorted versions of *ricardo* were shown. Hence, this model was excluded from the rest of the test.

During testing, the entire set of stimuli was consumed and evaluated by every subject; that is, a total of 105 stimuli, considering 5 contents compressed with 5 codecs at 4 compression levels, including a hidden reference. The stimuli were presented to the participants in a random order, after ensuring that the same content was never shown consecutively. To avoid biases induced by preferences to a particular side of the screen, half of the sequences portrayed the reference at the right side and half at the left side of the screen.

G. Subject recruitment

The subjective assessment was conducted using a crowdsourcing online platform [20], with participants recruited using the Amazon Mechanical Turk service. Human Intelligence Tasks (HIT) were created, with each task constrained to a single worker at a time. Upon acceptance of the task, each worker was redirected to the server where the evaluation platform was hosted. The screen size was then detected and only participants with resolution above 1920×1080 were allowed to continue. This step can be regarded as the first of a two-stage process, narrowing the pool of subjects to only those who have proper display equipment for the test.

The amount of active HITs limited the number of simultaneous accesses on the server, ensuring sufficient bandwidth for the video streaming. Users with low internet speed were instructed to not accept the task, warned that the test could take much longer. A reward of 5 U$ was granted to the participants after verification that all stimuli had been rated.

A total of 48 subjects (42 males and 6 females), with an age span of 20 - 63 years and average of 35.1, participated in this subjective quality assessment experiment.

III. Data processing

A. Objective quality metrics

Point cloud objective quality assessment methodologies can be distinguished as: (a) point-based, and (b) image-based. From the former category, we choose all point-to-point and the point-to-plane [21] variations, and the color PSNR of the luminance and each chrominance component computed separately, which comprise the MPEG PCC metrics. Moreover, we consider the point-to-distribution [22], plane-to-plane [23], PCQM [24], PointSSIM [25], and PCM_RR [26] metrics. From the latter category, we recruit the PSNR, SSIM [27], MS-SSIM [28], and VIFp [29] (i.e., multi-scale in pixel domain).

B. Subjective quality scores

According to [30], traditional outlier detection algorithms that operate only on subjective scores attributed to distorted stimuli are unable to efficiently filter unreliable subjects in crowdsourcing experiments. In this study, in order to confirm that subjects were focused on the task at hand, we make use of their ratings with respect to the 5 hidden references that were randomly presented to them. In particular, we exclude subjects that rated even one hidden reference below an imposed threshold of 4, which corresponds to the description “Perceptible, but not annoying”. Thus, we rule out subjects who either didn’t

properly understand the instructions, had inappropriate screen or lighting conditions, or answered carelessly to the questions.

In a second stage, the outlier detection algorithm described in the ITU-R Recommendation BT.500 [31] was applied to the scores of the remaining subjects. Similarly, all scores of participants identified as outliers were discarded. Using the remaining ratings, the Mean Opinion Score (MOS) and the 95% Confidence Interval (CI) assuming Student's t-student distribution is finally computed for each stimulus.

C. Computation of objective quality metrics

The point-to-point and point-to-plane metrics are implemented by the software version 0.13.5 that is presented in [32]. Both the Mean Square Error (MSE) and the Hausdorff distance (HSD) were used as pooling methods, with the corresponding geometric PSNR variants additionally considered. For the color PSNR metrics, the color attributes were converted from the original RGB to the YCbCr color space, following the ITU-R Recommendation BT.709-6 [33], using the same software release. For each aforementioned metric, the symmetric error was adopted. Following the JPEG CTC document [14], the CloudCompare software was used in order to estimate the normal vectors of the pristine models for the computation of the point-to-plane metric, using a quadratic surface fitting with range search of radius 20.

For the point-to-distribution metric, the source code provided in [22] was executed, using the default neighborhood size of 31. To compute a global degradation score, both the mean and the MSE pooling methods were used, which are referred to as MMD and MSMD, respectively. The corresponding PSNR versions are additionally assessed. In all cases, the symmetric error provided the final predictions. For the plane-to-plane metric, the version 1.0 of the scripts released with [23] was employed. The normal vectors were estimated using CloudCompare's quadric fitting with radius search of 20, as per [14].

For the PointSSIM, the scripts published with [25] were employed. The PointSSIM was computed on the luminance channel, using variance as dispersion estimator, neighborhood size of 12, and target voxel resolution of 9, which equals the minimum voxel bit depth among the contents. To obtain a global degradation score, the symmetric error was used. For PCQM and PCM-RR, the software released in [24] and [26], respectively, was used as such.

For image-based metrics, the same rendering settings that were adopted during subjective evaluation were employed. The models were captured from 8 viewpoints, starting from the frontal view and rotating by 45°, with a simple average providing the total degradation score. The metrics were applied on the union of foregrounds of the projected pristine and impaired models, and were computed on the luminance channel, after converting the RGB color values to the YCbCr color space using the ITU-R Recommendation BT.709-6 [33].

D. Benchmarking of objective quality metrics

Objective quality metrics are typically benchmarked after applying a regression model in order to map the objective scores to the subjective quality range, while accounting for biases, non-linearities and saturations that might appear in subjective testing. In our analysis, we follow the Recommendation ITU-T J.149 and use the logistic fitting function that is given in (1),

$$f(\mathrm{x}) = a + \frac{b}{1 + \exp^{-c\cdot(\mathrm{x}-d)}} \tag{1}$$

where a, b, c and d denote parameters of the function that are determined using a least squares method, after ensuring monotonicity in order to maintain the ranking order, and $f(\mathrm{x})$ indicates a predicted MOS. The Pearson linear correlation coefficient (PLCC), the Spearman rank order correlation coefficient (SROCC), the root-mean-square error (RMSE) and the outlier-ratio (OR) are then computed between the MOS and the predicted MOS in order to examine the linearity, monotonicity, accuracy and consistency of an objective metric, respectively. The regression and the performance indexes were computed across the entire dataset and for each codec separately.

IV. Results

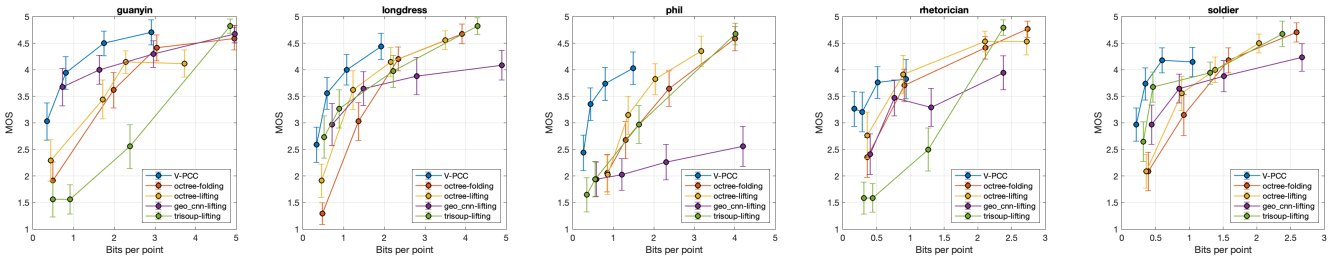
A. Subjective evaluation

A total of 14 subjects were identified as outliers based on hidden reference ratings. The subsequent method from ITU-R BT.500 recommendation found no outliers. We observed that the discarded subjects rated the least compressed models with lower scores than the remaining participants. Therefore, the average range of the MOS between the first and the last compression levels was increased from 1.77 to 2.02 as a results of applying the outlier detection method.

Based on the remaining ratings, the MOS and the CIs were computed and corresponding plots against the bit-rate are provided in Figure 3, for all codecs and compression levels, separated by content. It can be observed that V-PCC tends to have better performance at low bit-rates for all contents, without reaching highest quality scores at the assessed compression levels. The *octree-folding* method has a similar performance to *octree-lifting*, with a small disadvantage on intermediate bit-rates, but always achieving slightly higher MOS on the last compression level at a cost of bit-rate increase. The results for *geo_cnn-lifting* indicate that while its performance is clearly better than *octree-lifting* and *trisoup-lifting* for *guanyin*, the subjective results are comparable for the rest of the contents at low bit-rates, and deteriorate at higher bit-rates. It is evident that the performance of this codec is poor for *phil*, which might be explained by the inability of learning-based solutions to generalize well to content of different characteristics (i.e., the geometric resolution of *phil* is sparser than the rest of the contents). Finally, *trisoup-lifting* is found to be superior to *octree-lifting* at low quality levels with very similar performance after mid-range bit-rates for *longdress* and *soldier*, whereas for the rest of the contents, the *octree-lifting* is outperforming.

B. Performance of objective quality metrics

The performance indexes of the objective metrics under consideration using logistic regression, are reported in Table I.

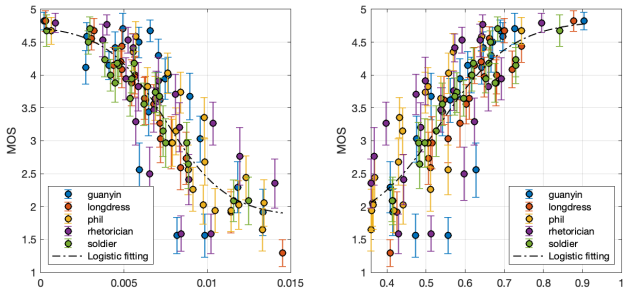


(a) *guanyin* (b) *longdress* (c) *phil* (d) *rhetorician* (e) *soldier*

Fig. 3: Subjective scores against bit-rates achieved by each codec, clustered per content.

TABLE I: Performance indexes computed over the entire dataset and after splitting per codec.

	entire dataset				**V-PCC**		**octree-folding**		**octree-lifting**		**geo_cnn-lifting**		**trisoup-lifting**	
Objective metrics	PLCC	SROCC	RMSE	OR	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
point-to-point HSD	0.452	0.370	0.858	0.76	0.242	0.349	0.875	0.778	0.858	0.803	0.648	0.554	0.930	0.903
point-to-point MSE	0.730	0.641	0.657	0.62	0.846	0.841	0.872	0.789	0.856	0.754	0.854	0.830	0.894	0.887
point-to-plane HSD	0.352	0.493	0.900	0.80	0.129	0.173	0.870	0.802	0.854	0.759	0.422	0.369	0.956	0.893
point-to-plane MSE	0.730	0.688	0.658	0.66	**0.886**	0.860	0.869	0.837	0.857	0.810	0.823	0.799	0.881	0.887
PSNR point-to-point HSD	0.483	0.521	0.845	0.82	0.563	0.379	0.803	0.743	0.775	0.748	0.793	0.707	0.921	0.869
PSNR point-to-point MSE	0.726	0.681	0.663	0.63	0.746	0.715	0.831	0.882	0.923	0.883	0.924	0.887	0.763	0.729
PSNR point-to-plane HSD	0.579	0.524	0.784	0.65	0.562	0.396	0.806	0.826	0.786	0.813	0.723	0.619	0.929	0.861
PSNR point-to-plane MSE	0.739	0.739	0.649	0.63	0.753	0.700	0.815	0.883	0.906	0.873	0.921	0.870	0.772	0.761
PNSR_Y	0.585	0.575	0.780	0.71	0.463	0.414	0.809	0.813	0.785	0.778	0.531	0.483	0.822	0.767
PNSR_U	0.400	0.390	0.881	0.76	0.454	0.301	0.587	0.592	0.643	0.582	0.353	0.170	0.583	0.502
PNSR_V	0.413	0.401	0.876	0.77	0.667	0.534	0.581	0.552	0.680	0.566	0.352	0.225	0.527	0.467
MMD	0.660	0.558	0.722	0.70	0.871	**0.866**	0.741	0.489	0.755	0.466	0.773	0.811	0.856	0.889
MSMD	0.663	0.544	0.720	0.68	0.851	0.824	0.740	0.487	0.755	0.505	0.713	0.763	0.858	0.891
PSNR MMD	0.711	0.683	0.676	0.68	0.773	0.755	0.884	0.828	0.921	0.792	**0.953**	**0.957**	0.796	0.785
PSNR MSMD	0.729	0.682	0.659	0.67	0.760	0.733	0.886	0.780	0.921	0.740	0.933	0.933	0.833	0.789
Plane-to-plane	0.687	0.593	0.699	0.69	0.874	0.845	0.757	0.530	0.775	0.582	0.789	0.744	0.855	0.800
PCQM	**0.870**	**0.875**	**0.473**	**0.42**	0.852	0.836	**0.963**	**0.931**	**0.953**	0.865	0.831	0.699	0.921	0.852
PointSSIM	0.827	0.846	0.540	0.56	0.853	0.859	0.931	0.925	0.922	0.882	0.886	0.875	0.907	0.856
PCM-RR	0.630	0.623	0.747	0.68	0.652	0.640	0.941	0.868	0.791	0.686	0.340	0.321	0.791	0.533
PSNR	0.609	0.564	0.763	0.76	0.686	0.504	0.731	0.713	0.782	0.730	0.295	0.142	0.862	0.815
SSIM	0.644	0.637	0.736	0.68	0.492	0.496	0.809	0.816	0.844	0.836	0.033	-0.050	0.871	0.822
MS-SSIM	0.812	0.765	0.561	0.62	0.601	0.546	0.907	0.870	0.950	**0.895**	0.446	0.330	**0.963**	**0.916**
VIFp	0.710	0.689	0.677	0.65	0.551	0.539	0.858	0.848	0.908	0.862	0.368	0.209	0.907	0.857



(a) PCQM. (b) PointSSIM.

Fig. 4: Subjective against objective scores obtained from the two best-performing quality metrics.

Scatter plots indicating the behaviour of best-performing objective quality metrics are presented in Figure 4.

According to the performance indexes computed over the entire dataset, PCQM was found the most accurate predictor, with PointSSIM the second best-performing alternative, followed by the image-based MS-SSIM. Our results indicate that luminance-based features are effective in quantifying visual impairments of point cloud contents. The higher performance of PCQM can be justified by the additional geometric measurements that are incorporated in its final prediction, which were found beneficial. For PointSSIM, geometric distortions are implicitly captured by the formulation of local neighborhoods to compute local statistics, while the voxelization operation leads to reduction of cross-content density differences, improving the generalization capabilities of the method. The benefits of local pooling to compute features against point-to-point attribute comparisons are evident when comparing the performance of PointSSIM with PSNR_Y.

The analysis of the metrics' performance per codec indicates consistent behaviour of those two predictors in most cases, even if they are outperformed by other metrics such as MMD and MSMD, variants of point-to-point and point-to-plane, and MS-SSIM, under visual impairments from single codecs.

The majority of metrics from the MPEG suite show poor performance under distortions from V-PCC, which is however not the case for the MSE variants of point-to-point and plane-to-plane. Image-based metrics do not correlate well with the subjective scores for this codec neither, showing even worse performance against *geo_cnn-lifting*. However, they provide more accurate predictions under the remaining codecs, with MS-SSIM outperforming all metrics for *trisoup-lifting*.

Finally, we observe that lower-complexity metrics, such as

point-to-point and point-to-plane variants, or even PSNR_Y can perform well under degradations occurring from a single encoder. In particular, the MSE version of the first two methods shows rather consistent performance across all tested codecs.

V. Conclusion

In this study, state-of-the-art point cloud objective quality metrics are benchmarked against subjective scores collected from a crowdsourced experiment for evaluation of learning-based and conventional codecs. Performance indexes show that most objective metrics perform poorly when tested against the entire dataset, with PCQM and PointSSIM providing the best correlation results. This can be explained by the diversity of the evaluated set, which contains point clouds of different characteristics, compressed by a large number of methods that introduce visual impairments of different natures. These results suggest that the underlying mechanisms of the aforementioned predictors used for combining luminance-based features with geometric measurements are particularly efficient for correlation with human perception. A separate analysis per codec showed good results for several other metrics, suggesting that under similar visual degradations and in a more limited context, they can be useful.

References

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

[2] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, 2017.

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

[4] 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.

[5] 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.

[6] L. A. da Silva Cruz *et al.*, "Point cloud quality evaluation: Towards a definition for test conditions," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–6.

[7] 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.

[8] MPEG Systems, "Text of ISO/IEC DIS 23090-18 Carriage of Geometry-based Point Cloud Compression Data," ISO/IEC JTC1/SC29/WG03 Doc. N0075, Nov. 2020.

[9] MPEG 3D Graphics Coding, "Text of ISO/IEC CD 23090-5 Visual Volumetric Video-based Coding and Video-based Point Cloud Compression 2nd Edition," ISO/IEC JTC1/SC29/WG07 Doc. N0003, Nov. 2020.

[10] E. Zerman, P. Gao, C. Ozcinar, and A. Smolic, "Subjective and objective quality assessment for volumetric video compression," *Electronic Imaging*, vol. 2019, no. 10, pp. 323–1, 2019.

[11] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual quality assessment of 3d point clouds," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3182–3186.

[12] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point cloud rendering after coding: impacts on subjective and objective quality," *arXiv preprint arXiv:1912.09137*, 2019.

[13] 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.

[14] S. Perry, "JPEG Pleno Point Cloud Coding Common Test Conditions v3.1," ISO/IEC JTC1/SC29/WG1 Doc. N86044, Jan 2020.

[15] "JPEG Pleno Database, https://jpeg.org/plenodb/."

[16] 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.

[17] M. Quach, G. Valenzise, and F. Dufaux, "Improved Deep Point Cloud Geometry Compression," in *IEEE International Workshop on Multimedia Signal Processing (MMSP'2020)*, Sep. 2020.

[18] ——, "Folding-based compression of point cloud attributes," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 3309–3313.

[19] F. Bellard, "Bpg image format."

[20] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "QualityCrowd - A framework for crowd-based quality evaluation," in *2012 Picture Coding Symposium*. Krakow: IEEE, May 2012, pp. 245–248. [Online]. Available: http://ieeexplore.ieee.org/document/6213338/

[21] 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.

[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] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, Nov 2003, pp. 1398–1402 Vol.2.

[29] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb 2006.

[30] T. Hoßfeld, C. Keimel, M. Hirth, B. Gardlo, J. Habigt, K. Diepold, and P. Tran-Gia, "Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541–558, 2014.

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

[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] ITU-R BT.709-6, "Parameter values for the HDTV standards for production and international programme exchange," International Telecommunication Unionn, Jun. 2015.