Subjective and Objective Testing in Support of the JPEG Pleno Point Cloud Compression Activity

Stuart Perry
University of Technology Sydney
Sydney, Australia
Stuart.Perry@uts.edu.au

Luis A. da Silva Cruz
Instituto de Telecomunicações (IT)
Universidade de Coimbra
Coimbra, Portugal
lcruz@deec.uc.pt

João Prazeres, António M. G. Pinheiro
Instituto de Telecomunicações (IT)
Universidade da Beira Interior
Covilhã, Portugal
joao.prazeres@ubi.pt pinheiro@ubi.pt

Emil Dumic
Department of Electrical Engineering
University North, Croatia
Varazdin, Croatia
edumic@unin.hr

Davi Lazzarotto, Touradj Ebrahimi
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
davi.nachtigalllazzarotto@epfl.ch touradj.ebrahimi@epfl.ch

***Abstract*—Point clouds have many applications in today's society ranging from entertainment to autonomous driving. With these new applications comes the need to compress the growing volume of point cloud data in a manner that is both suitable for human visualization and machine processing applications. The JPEG Pleno Point Cloud activity has been working toward a learning-based coding standard for point clouds, offering a single-stream, compact compressed domain representation, supporting advanced flexible data access functionalities targeting both interactive human visualization, and effective performance for 3D processing and machine-related computer vision tasks. As part of this activity, the JPEG Committee has been performing a number of exploration studies to evaluate existing coding standards as well set up baseline anchors and examine objective metrics against which new learning-based solutions may be compared. This article provides an overview of the JPEG Pleno Point Cloud activity and discusses challenges and solutions to the problem of evaluating and comparing cloud coding solutions. Experimental results will be presented demonstrating methodologies used by the JPEG Committee for point cloud compression assessment as well as outlining the performance of current state of the art compression standards on point clouds as well as the sensitivity of the objective metrics used for this activity to various adjustable parameters.**

***Index Terms*—JPEG, point cloud, compression, machine learning**

I. Introduction

Point cloud applications have become more numerous in the last 10 years and look to continue on an accelerated trajectory of adoption by society. Applications derived from 3D scanning, analysis and visualisation as well as augmented, mixed and virtual reality applications look to have a dramatic effect on society in the near future. These emerging applications create new challenges and demand new technologies to unlock their potential. One of the major emerging challenges is the massive volume of 3D data that needs to be collected, stored, analysed and displayed to enable the use of point clouds in practical applications. A high quality scan of even a small object can require millions of points to represent the object shape, while the unrestricted positions of the points in space together with the need to store attributes mean representation cost of the full point cloud can be large, easily reaching gigabytes. If one considers emerging applications such as autonomous driving that involve the capture and processing of streams of point cloud data in real time, the need for efficient and powerful compression technologies for point clouds becomes urgent. The JPEG Committee has been working on coding standards for plenoptic data as part of its JPEG Pleno activity for a number of years. Plenoptic data in this context is considered to cover holography, light fields and point clouds, all of which are different representations of the plenoptic capture function [1], [2]. The scope of the JPEG Pleno Point Cloud activity is the development of standards for point cloud representation that not only involve efficient coding, but also support machine vision applications. This activity will advance through a series of stages:

- Stage 1: A learning-based coding standard addressing human visualization and decompressed/reconstructed domain 3D processing and computer vision tasks;
- Stage 2: A learning-based coding standard additionally supporting compressed domain 3D processing such as visual enhancement and super-resolution;
- Stage 3: A learning-based coding standard additionally supporting compressed domain computer vision tasks such as classification, recognition and segmentation.

During the 94th JPEG meeting, the JPEG Committee released

António M. G. Pinheiro and João Prazeres thank the Portuguese FCT-Fundação para a Ciência e Tecnologia under the project UIDB/50008/2020, PLive X-0017-LX-20, and operation Centro-01-0145-FEDER-000019 - C4 - Centro de Competencias em Cloud Computing for funding this research.
Davi Lazzarotto and Touradj Ebrahimi thank the Swiss National Foundation for Scientific Research (SNSF) under the grant number 200021-178854 for funding this research.

978-1-6654-6623-3/22/$31.00 ©2022 IEEE

a Final Call for Proposals on JPEG Pleno Point Cloud Coding. This call addresses Stage 1 of the activity [3]. In early 2022 a study was performed to support the Call for Proposals. The goal of this study was to:

1) Evaluate the performance of current state of the art point clouds coding solutions tested on samples of the point cloud training set to be supplied to proponents for the Call for Proposals.
2) Understand to which extent differences between laboratories may affect the subjective quality assessment of submissions to the upcoming Call for Proposals.
3) Determine the impact of point cloud normal estimation methods and parameters on the computation of objective metrics intended to be used during the Call for Proposals.

In Section II the experimental methodology followed will be presented including the selection of point clouds and state of the art point cloud codecs for use in the study, as well as the objective metrics and the subjective testing methodology to be used. Section III will detail the results of the study, while Section IV will provide a discussion of the results.

II. Experimental Setup

A. Content

To benchmark the performance of the chosen state of the art codecs, a set of seven point clouds were chosen. The point clouds used in this investigation are shown in Fig. II-A. The *longdress*, *guanyin* and *rhetorician* point clouds were sourced from the JPEG Pleno Database [4]. The *camera*, *car*, *plantanopote* and *suzuki* point clouds are sampled from meshes obtained from the ShapeNetCore Database [5]. The dataset is publicly available[1]. The sampling process followed Lazzarotto and Ebrahimi's methodology [6], which involved the exclusion of internal faces prior to the sampling process in order to avoid obtaining colors from different faces at similar positions.

B. Anchor Codecs

In order to establish a baseline for the future comparison of learning-based point cloud codecs, the JPEG Committee chose two common non-learning-based codecs developed by the MPEG Standardisation group; G-PCC [7] and V-PCC , [8], [9] as anchor codecs. These codecs will form the base level of performance for the subsequent Call for Proposals on JPEG Pleno Point Cloud Coding [3], so it is imperative that the performance on the training set is well understood.

G-PCC uses an octree encoding method. It has two encoding modes for the deepest level of geometrical information; Octree and Triangle Soup. In this work the Octree encoding mode was selected, with compression factor controlled by the *positionQuantizationScale* parameter to obtain five encoding rates (R01-R05) from low to high quality. For each of the rates, the Lifting parameters *seq_lod* and *seq_dist2* were set to 12 and 3 respectively.

[1]http://webx.ubi.pt/∼pinheiro/euvip2022pcdb.html

Fig. 1. Point Clouds used in this investigation. The *longdress*, *guanyin* and *rhetorician* point clouds were sourced from the JPEG Pleno Database [4], while *camera*, *car*, *plantanopote* and *suzuki* point clouds are sampled from meshes obtained from the ShapeNetCore Database [5] using the technique of Lazzarotto and Ebrahimi [6].

V-PCC uses a projection based method wherein the point cloud is projected as a set of patches onto multiple planes (usually six). The projection patches represent point cloud texture and color, depth information and an occupancy map. Each projected set of patches is compacted and the resulting sequence of images compressed using traditional 2D video techniques. MPEG V-PCC test model TMC2 version 8 [9] with VVC was used in All Intra (AI) coding mode with the encoding condition being *C2, Lossy Geometry - Lossy Attributes*.

C. Objective Metrics

Currently, there is already a wide variety of point cloud quality metrics available. Based on a previous study [10], the JPEG Committee has found that the PSNR D1 and PSNR D2 [11] quality metrics display consistent performance in terms of point cloud quality evaluation. Since PSNR D1 and PSNR D2 only measure geometrical accuracy of point clouds, there is a need to include additional metrics that use both color and geometry. For the purpose of supporting the Call for Proposals, the authors considered some recent point cloud quality measures, PCQM [12] and PointSSIM [13] [14]. For each point cloud/codec/rate combination, the objective quality metrics PSNR D1, PSNR D2, PCQM and PointSSIM were computed. The PSNR D1 and PCQM measure point to point distances, whereas PSNR D2 requires normal information to measure surface to point distances and PointSSIM can also be employed with normal-based features or color-based features. To compute the PointSSIM metric, the variance (VAR) was used as a statistical estimator, and a neighborhood size of 12 was used as recommended in the original work [13]. Both normal-based and color-based features were considered. The use of normal information in objective metric calculations can lead to inconsistent results, particularly for sparse point clouds. Depending on spatial sparsity of the points and the method of normal calculation, the obtained metrics can be subject to unwanted variation. In this work, an investigation was conducted to inquire whether the number of neighbouring points used to compute the normals had an effect on the accuracy of PSNR D2 and PointSSIM. To do so, Cloud Compare [15] was used to fit a quadric local surface based on 5, 10 and 20 neighbour points from which the normals were computed. The estimated normals were then used to compute PSNR D2 and PointSSIM values.

D. Subjective Testing Methodology

For the subjective quality component of this experiment, a set of 12 second stimulus videos at 4096x2160 resolution were created with reference and processed (encoded by a codec at a particular rate point and then decoded to create a reconstruction) point clouds shown side by side. The videos were shown to subjects at a frame rate of 30fps using a customised version of the MPV video player [16]. During the 12 second period, the reference and processed point clouds were rotated synchronously about their respective central vertical axes, to complete a full 360° path. Subjects were instructed to judge visual quality of the processed with respect to the reference point clouds according to a Double Stimulus Impairment Scale protocol with 5 possible impairment ratings (1 - *very annoying*, 2 - *annoying*, 3 - *slightly annoying*, 4 - *perceptible, but not annoying* and 5 - *imperceptible*). To mitigate potential bias, each subject was only shown videos with the reference on the same side of the display, with half of subjects shown videos with the reference on the left and the remaining half of the subjects shown videos with the reference on the right. The content presentation order was random, but adjusted so that subjects did not at any point see the same content as that shown in the immediately preceding video. Each session started with a training session using a point cloud from the JPEG Pleno Database [4] that was not used for subsequent data collection. Following the training session, subjects were shown seven different content types processed by two codecs at five different rates together with seven reference-reference pairs (one for each content point cloud) for a total of 77 double stimuli videos. The reference-reference pairs were included to understand subject behaviour in the case when no artefacts were present and to determine if non-attentive subjects were present. Data from two test labs is described in this work: University of Beira Interior (UBI), Covilhã, Portugal and University North (UNIN), Varaždin, Croatia and test environments were set up according to ITU-R Recommendation BT.500-13 [17] as shown in Table I. The display resolution used by UNIN is smaller than the videos, but as no video scaling was allowed, the videos were displayed in true resolution. After careful check, was observed that the information of the point cloud was not cropped. This means that subjects in UBI and UNIN saw the same information. The cropped area did not show any point cloud information in all cases. Outlier detection was performed according to BT.500-13 [17] on each laboratory set of data separately with no outliers found. Finally mean opinion scores (MOS) and 95% confidence intervals were computed. Table II presents the gender and age breakdowns of the subjects for the two laboratories.

TABLE I
EXPERIMENTAL SETUP AT TEST LABORATORIES

Laboratory	Display Type	Resolution	Viewing Distance
UBI	Eizo ColorEdge CG318-4K	4096x2160 (31.1″)	1.2m (FV ±15cm)
UNIN	Sony TV 55” KD-55x8505C	3840x2160 (55″)	1.5m (FV ±15cm)

TABLE II
TEST SUBJECT INFORMATION AT TEST LABORATORIES

	Males	Females	Total	Age span	Average age
UBI	10	8	18	21-34	26.0
UNIN	17	1	18	19-59	26.4

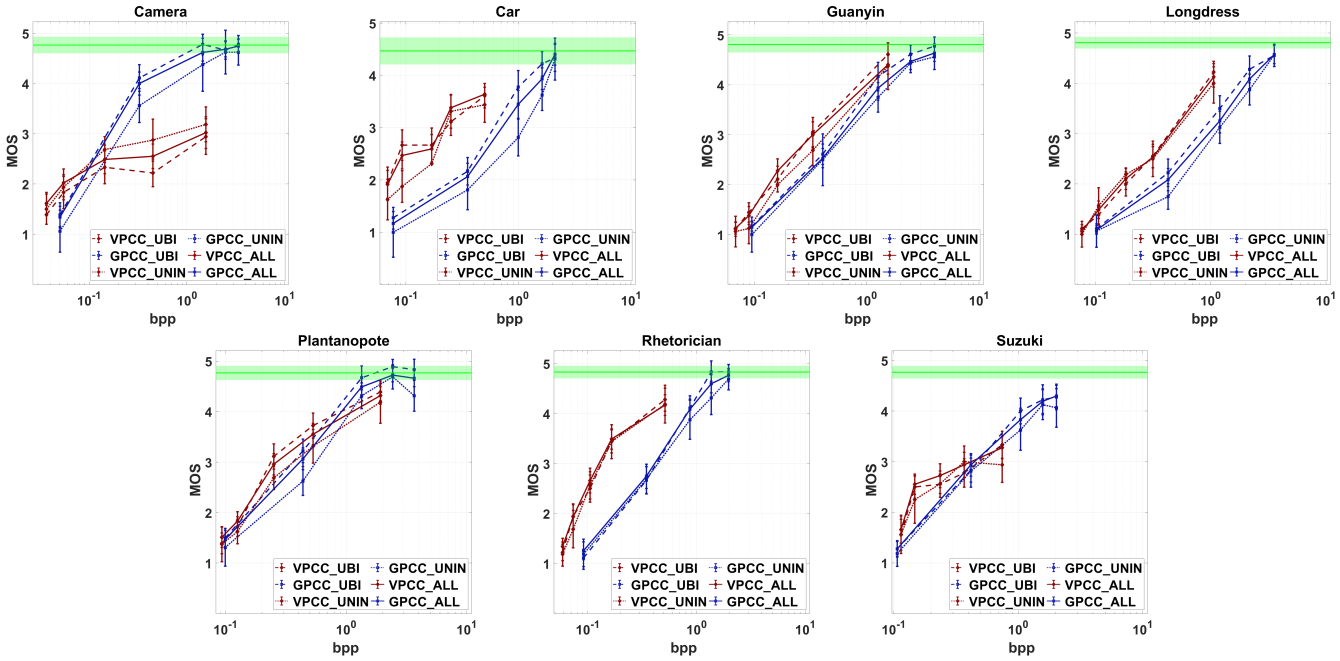

Fig. 2. MOS results for the seven tested point clouds. The red lines represent results for the V-PCC codec, while the blue lines represent results for the G-PCC codec. The error bars are 95% confidence intervals, while the green bars represent the 95% confidence intervals for the reference-reference stimuli.

TABLE III
CORRELATION OF MOS RESULTS ACROSS LABS

	PCC	SROCC	RMSE	OR
UBI vs UNIN	0.983	0.979	0.062	0.143

III. RESULTS

A. Subjective Results

Figure 2 shows the MOS plotted against bitrate for individual labs and aggregated across all the subjects from all of the participating laboratories. Bitrate is measured as bits per point (bpp) and is computed as the ratio of the total number of bits of the encoded content divided by the number of input points in the encoded point cloud. Based on the high degree of correlation found between the different laboratories, as will be demonstrated in Section III-B, the authors considered the consolidation of the scores from all the laboratories to be valid.

B. Correlation Across Labs

To determine the degree of correspondence of MOS between the different test laboratories, the Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), Root-Mean Squared Error (RMSE) and Outlier Ratio (OR) were computed. The results are presented in Table III. Figure 3 shows the linear fitting across all laboratories. In general, the correlation between the test laboratories is quite high with both Pearson and Spearman correlation coefficients above 0.97.

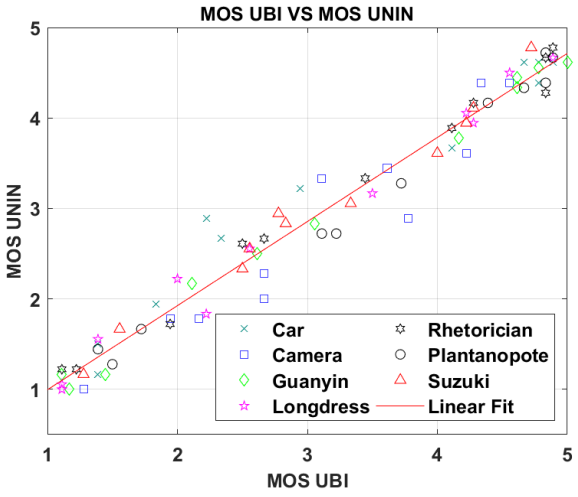

Fig. 3. Linear fitting for correlation between MOS obtained from different test laboratories.

C. Objective Metric Results

To measure the ability of the objective metrics to predict subjective scores, we employed the methodology from Recommendation ITU-T P.1401 [18]. This involves the computation of PCC, SROCC, RMSE and OR on the original and predicted MOS values. The predicted MOS values were obtained following the fitting of a logistic function to the objective scores. The results are shown in Table IV, while the individual MOS-objective quality pairs and fitted curves are shown in Fig. 4.

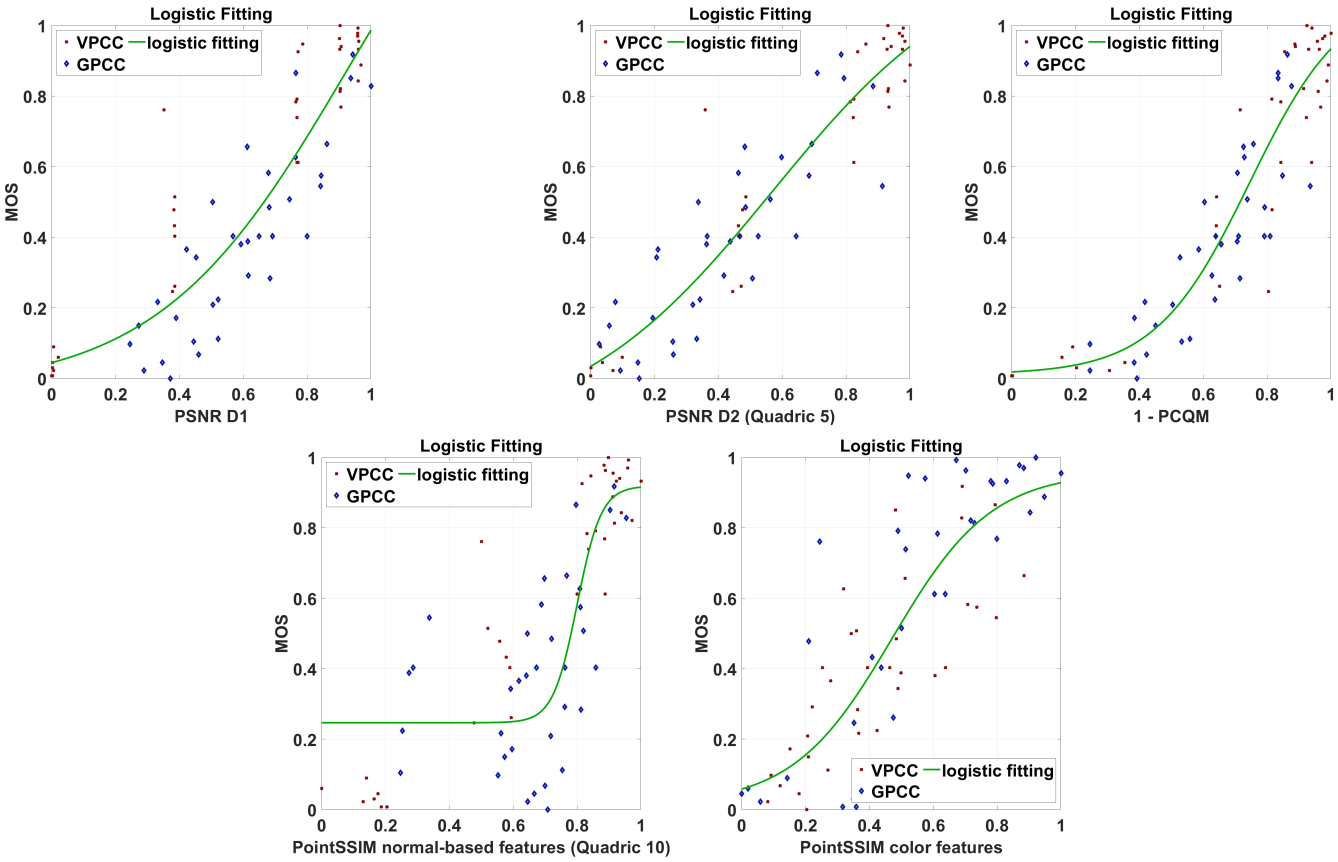

Fig. 4. MOS results plotted against objective metrics and regression curve between metric and MOS.

TABLE IV
PERFORMANCE MEASURES OF OBJECTIVE METRICS AGAINST MOS RESULTS.

Metric	PCC	SROCC	RMSE	OR
PSNR D1	0.889	0.879	0.151	0.529
PSNR D2 Quadric 5	0.928	0.921	0.123	0.457
PSNR D2 Quadric 10	0.928	0.918	0.123	0.443
PSNR D2 Quadric 20	0.926	0.916	0.125	0.443
PointSSIM Quadric 5	0.761	0.765	0.213	0.671
PointSSIM Quadric 10	0.824	0.800	0.186	0.700
PointSSIM Quadric 20	0.765	0.763	0.212	0.686
PointSSIM Color-based	0.830	0.827	0.184	0.600
PCQM	0.916	0.913	0.132	0.586

IV. DISCUSSION AND CONCLUSIONS

Based on the results described in Section III, a number of conclusions can be drawn. The results from the two laboratories were highly consistent despite the use of different displays and different resolutions. This robustness has been observed in previous studies [10], [19] and is encouraging as this is crucial to the ability of the JPEG Committee to accurately ascertain the performance of proposals. Examining Fig. 2 it can be observed that although different laboratories have similar MOS for the same content, there are clear differences between the performance G-PCC and V-PCC dependent on content. For example, for most of the content, V-PCC outperforms or performs as well as G-PCC with the exception of the *camera* point cloud where at higher bitrates G-PCC performs better. It is unclear as to what aspects of the *camera* point cloud might be responsible. The point cloud has a number of large flat surfaces that may have been difficult for V-PCC to encode accurately when not aligned precisely with the projection surfaces. For the objective metric results, we observe in Fig. 4 that PSNR D1, PSNR D2 and PCQM show a good relationship between the objective metrics and MOS, however PointSSIM appears to have reduced accuracy in MOS prediction at lower quality levels for the version that makes use of normal features. Higher compression levels for point clouds are generally associated with an increased sparsity of the reconstructed point cloud. The reduced accuracy of the PointSSIM metric may be related to the increased sparsity of

the point clouds at lower quality levels. From Table IV we can observe a decreased correlation and an increased Outlier Ratio for PointSSIM compared to the other metrics. In regard to the effect that normal calculation has on the accuracy of the metrics, we can observe that while PSNR D2 appears to be relatively robust to the number of neighbouring points used in the normal calculation, PointSSIM appears more sensitive to this factor. Pearson correlation values range from 0.761 to 0.824 with the addition of normal information, below the value of 0.830 when color-based features are employed. This increased sensitivity of PointSSIM to the normal vectors is probably due to the fact that the estimation has to be performed for both the reference and the degraded models, contrary to PSNR D2 which only requires normal vectors for the reference.

ACKNOWLEDGMENT

The authors thank Nhung Thi Hong Nguyen for assistance with the preparation of stimuli for the experiment. We also thank University of Beira Interior (PT), Vrije Universiteit Brussel (BE), University North (HR) and University of Patras (GR) for participating in the subjective experiment. Unfortunately, due to low numbers or differences in display resolution data from Vrije Universiteit Brussel (BE) and University of Patras (GR) could not included in this work.

REFERENCES

[1] P. Schelkens, Z.Y. Alpaslan, T. Ebrahimi, K.-J. Oh, F.M.B. Pereira, A.M.G. Pinheiro, I. Tabus, Z. Chen, "JPEG Pleno: a standard framework for representing and signaling plenoptic modalities," Proc. SPIE 10752, Applications of Digital Image Processing XLI, 107521P (17 September 2018)

[2] P. Astola, L. da Silva Cruz, E. da Silva, T. Ebrahimi, P. Freitas, A. Gilles, K. Oh, C. Pagliari, F. Pereira, C. Perra, S. Perry, A. Pinheiro, P. Schelkens, I. Seidel, I. Tabus, "JPEG Pleno: Standardizing a coding framework and tools for plenoptic imaging modalities", ITU Journal: ICT Discoveries, vol. 3, no. 1, June 2020.

[3] ISO/IEC JTC1/SC29/WG1, "Final Call for Proposals on JPEG Pleno Point Cloud Coding," Doc. WG1N100097, Jan 2022.

[4] JPEG Pleno Database, https://jpeg.org/plenodb/. [Online]. Available: https://jpeg.org/plenodb/.

[5] ShapeNet, https://shapenet.org/. [Online]. Available: https://shapenet.org/.

[6] D. Lazzarotto and T. Ebrahimi, "Sampling color and geometry point clouds from ShapeNet dataset", arXiv:2201.06935, Jan 2022.

[7] MPEG 3DG, "G-PCC Codec Description v5," ISO/IEC JTC1/SC29/WG11 N18891, Geneva, CH, October 2019.

[8] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging mpeg standards for point cloud compression," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 1, pp. 133–148, March 2019.

[9] MPEG 3DG, "V-PCC Test Model v8," ISO/IEC JTC1/SC29/WG11 W18884, Geneva, CH, October 2019.

[10] S. Perry, H.P. Cong, L.A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, T. Ebrahimi, "Quality Evaluation of Static Point Clouds Encodec Using MPEG Codecs," Proceedings of the 27th IEEE International Conference on Image Processing 2020 (ICIP2020), Abu Dhabi, United Arab Emirates, 25-28 October 2020.

[11] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Evaluation metrics for point cloud compression," ISO/IEC JTC m74008, Geneva, Switzerland, Tech. Rep., January 2017.

[12] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: a full-reference quality metric for colored 3d point clouds," in the Proceedings of the 2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX), 2020, pp. 1–6.

[13] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in the Proceedings of the 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW), 2020, pp. 1–6.

[14] D. Lazzarotto; E. Alexiou; T. Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression", in the Proceedings of the 2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP), 2021, pp. 1-6.

[15] Cloud Compare, https://cloudcompare.org/doc/wiki/. [Online]. Available: https://cloudcompare.org/doc/wiki/.

[16] MPV video player, https://mpv.io. [Online]. Available: https://mpv.io.

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

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

[19] S. Perry, L.A. da Silva Cruz, E. Dumic, N.H.T. Nguyen; A. Pinheiro, E. Alexiou, "Comparison of Remote Subjective Assessment Strategies in the Context of the JPEG Pleno Point Cloud Activity", Proceedings of the 2021 IEEE International Workshop on Multimedia Signal Processing (MMSP 2021), Tampere, Finland, 6-8 October 2021.