

# QUALITY EVALUATION OF STATIC POINT CLOUDS ENCODED USING MPEG CODECS

Stuart Perry<sup>†</sup> Huy Phi Cong<sup>†</sup> Luís A. da Silva Cruz<sup>\*‡</sup> João Prazeres<sup>‡</sup>  
Manuela Pereira<sup>‡◦</sup> Antonio Pinheiro<sup>‡◦</sup> Emil Dumic<sup>§</sup> Evangelos Alexiou<sup>¶</sup> Touradj Ebrahimi<sup>¶</sup>

<sup>†</sup>University of Technology Sydney, Australia

<sup>\*</sup>University of Coimbra, Portugal

<sup>‡</sup>Instituto de Telecomunicações, Portugal

<sup>◦</sup>Universidade da Beira Interior, Portugal

<sup>§</sup>Department of Electrical Engineering, University North, Croatia

<sup>¶</sup>Multimedia Signal Processing Group, École Polytechnique Fédérale de Lausanne, Switzerland

## ABSTRACT

This paper presents a quality evaluation study of point cloud codecs that have been recently standardised by the MPEG committee. In particular, a subjective experiment to assess their performance in terms of bitrate against visual quality is designed and realized in four independent laboratories. The experimental setup of each laboratory varies; yet, the obtained subjective scores exhibit high inter laboratory correlation, confirming that the adopted assessment protocol is robust to equipment selection and viewing conditions, ensuring reliability and facilitating repeatability. Our study confirms the superior compression performance of the MPEG V-PCC, when compared to MPEG G-PCC, in the case of static contents. Finally, results from a benchmark of the most popular objective quality metrics using the obtained subjective scores as ground truth, reveal that the point2plane with mean square error is the most accurate quality predictor, closely followed by the point2point also using mean square error as distance measure.

**Index Terms**— Point Cloud, Subjective Quality Evaluation, Benchmarking, Coding

## 1. INTRODUCTION

Point clouds have recently attracted a strong interest in volumetric representation of visual information. Such contents commonly demand huge amounts of data; thus, efficient compression solutions are essential. Considering lossy encoding, reliable quality models are vital to assess the perceived distortions that ordinarily result from bitrate reductions.

In the literature, there are several studies targeting at the definition of frameworks and protocols for subjective quality evaluation of point clouds. Simplistic degradations with geometry-only contents are attempted in [1, 2], compression artifacts using prior encoding schemes are evaluated in [3–5], whereas in [6, 7], the contents are compressed using a sin-

gle codec. Current efforts account for a wider range of high-performing codecs, such as the ones reported in [8, 9].

In this paper, a subjective and objective quality evaluation campaign using the state-of-the-art MPEG point cloud codecs, namely, Video-based Point Cloud Compression (V-PCC) and Geometry-based Point Cloud Compression (G-PCC) is conducted, in four remote test laboratories. The objective is the establishment of a reproducible and reliable quality evaluation framework for static point cloud contents.

## 2. SUBJECTIVE QUALITY ASSESSMENT

### 2.1. Contents

In this study, a dataset of 6 static point clouds is used with both texture and geometry information, namely, *Long Dress*, *Loot*, *Soldier*, *Red and Black*, which can be found in [10, 11] and *Ricardo10* and *Sarah9*, available at [10]. The models represent human figures with up to about 1 million points. Frontal views of the reference contents are illustrated in Fig. 1. These point clouds are encoded using the MPEG G-PCC and V-PCC codecs, as described below.

### 2.2. MPEG G-PCC encoding

In MPEG G-PCC [12], there are two encoding modules integrated to compress geometry information, namely, Octree and Triangle Soup (TriSoup) [13, 14]. The first approach is based on an Octree decomposition, which is regulated through the positionQuantizationScale parameter (herein referred as *Depth*). The second approach is based on a surface reconstruction using triangular primitives, after enclosing the model in an octree structure. The octree can be adjusted through the *Depth* parameter, while the size of the block on which the triangular surface approximation is applied, is determined through the *trisoup\_node\_size\_log2* parameter (herein referred as *Level*). Octree-based encoding leads



**Fig. 1:** Frontal views of each point cloud.

to regular down-sampling of the input points, whereas the TriSoup-based encoding leads to the presence of holes in the form of triangles. The latter approach is advised to be used only with high-density point clouds.

The geometry of a model is initially encoded, using any of the above solutions, and then decoded in order to define the shape over which the color will lie. In particular, the color attributes are associated to the output points (after geometric compression) through a re-colouring step that takes place and uses the color values of the original model. The color information can then be encoded using two different approaches, namely, RAHT [15], and Prediction-plus-Lifting (i.e., Lifting). The first alternative is based on the 3D Haar transform, whereas the second is based on prediction of a color value from its neighbors. In [8], it was shown that by using the MPEG point cloud coding Common Test Conditions (CTC) [16], the observers tend to prefer Lifting over the RAHT counterpart. Thus, in this study we only consider the Lifting codec, in order to reduce the parameter space of our experiment, while retaining the most diverse types of artifacts.

Five out of six encoding configurations specified in the CTC (R01-R06) are used to encode the contents using the G-PCC version 7 [17]. Specifically, R01, R02, R04, R05 and R06 are employed, and define 5 quality levels (Q01-Q05) that span from very low to very high. Every content is encoded using the exact settings specified in the CTC, except of *Sarah9* with the TriSoup module. In the latter case, a *Depth* value of 0.5 is used to account for the higher sparsity of this particular content, instead of 1 that is applied on the rest of the contents. Yet, the same *Level* values are employed for the corresponding quality levels. Moreover, it should be noted that the *seq\_lod* and the *seq\_dist2* parameters of the Lifting are set

as 12 and 3, respectively, for every content.

### 2.3. MPEG V-PCC encoding

In MPEG V-PCC [13, 18], point clouds are encoded after being projected onto a set of planes (usually six), followed by 2D encoding of the projections. The projections are represented by three sets of information; texture patches, depth information and occupancy maps. The texture patches are packed to create a 2D image with as few pixels as possible which is then encoded using legacy video encoding methods. The distances from the projection plane to the points that project onto each texture patch pixel are represented by depth maps, also structured as patches matching the texture patches. The depth information is encoded using 2D video encoding methods too. The occupancy map is a 2D binary field that indicates which pixels in the 2D patch composite images (texture and depth) contain meaningful information and is encoded using a form of spatial quantization together with raster scanning and entropy encoding. The current MPEG V-PCC test model, TMC2 [19], uses HEVC to encode the sequence of projection images but other video encoders could be used.

Version 8 of the reference software is used with encoding parameters defined in the MPEG CTC document [16]. The selected encoding condition is C2, *Lossy Geometry – Lossy Attributes*, and since we are encoding static point clouds, the coding mode is All Intra (AI). Six rate points are selected, five borrowed from the MPEG CTC (R01-R05 from low to high quality) and an additional one defined on purpose for this study to yield a lower quality than R01 and comparable to the lowest quality rate point used in the G-PCC.

### 2.4. Evaluation Methodology

The subjective experiments were conducted in 4 different laboratories: University of Beira Interior (UBI), Covilhã, Portugal, University of Coimbra (UC), Coimbra, Portugal, University North (UNIN), Varaždin, Croatia and University of Technology Sydney (UTS), Australia. The conditions of every test environment were adjusted to follow the ITU-R Recommendation BT.500-13 [20]. The equipment and viewing conditions used in each laboratory are described in Table 1. A passive subjective evaluation methodology was applied using a customized version of the MPV video player [21]. The evaluated point clouds were rotated around a vertical axis passing through the center of each model in video sequences of 60 fps with a total duration of 12 seconds. Participants provided their scores after the completion of the playback animation.

The simultaneous Double Stimulus Impairment Scale (DSIS) test method was adopted with a 5-level rating scale (1 - *very annoying*, 2 - *annoying*, 3 - *slightly annoying*, 4 - *perceptible, but not annoying*, 5 - *imperceptible*), including a hidden reference for sanity check. Both the reference and the degraded stimuli were simultaneously shown to the observer

side-by-side, and every subject rated the visual quality of the processed with respect to the reference, which was clearly identified. To avoid biases, in half of the individual evaluations, the reference was placed on the right and vice-versa for the rest. Also, particular care was taken to not present the same content consecutively. At the beginning of each individual evaluation, a training session took place in order to familiarize the subjects with the artifacts under assessment.

A total of 96 scores were obtained per evaluation session, considering that each subject assessed 6 test models degraded by 3 compression schemes at 5 quality levels, plus hidden references. Outlier detection based on the ITU-R Recommendation BT.500-13 [20] was applied on the quality scores of each test laboratory, separately. In every case, no outliers were found. The Mean Opinion Scores (MOS) and the 95% Confidence Intervals (CIs), assuming a Student's t-distribution were computed on every set of scores. In Table 2, we report observers demographical information, per test laboratory.

### 3. RESULTS

#### 3.1. Subjective Scores

In Figure 2, the MOS against the achieved bitrates for all encoded stimuli are depicted per content, after aggregating the scores collected from each test laboratory. The bitrate is measured in bpp and is computed as the total number of bits required for a particular stimulus divided by the number of input points from the corresponding content. The scores are pooled together, given that the MOS obtained from the test laboratories are strongly correlated, as will be confirmed in the next section.

#### 3.2. Correlation between Subjective Scores

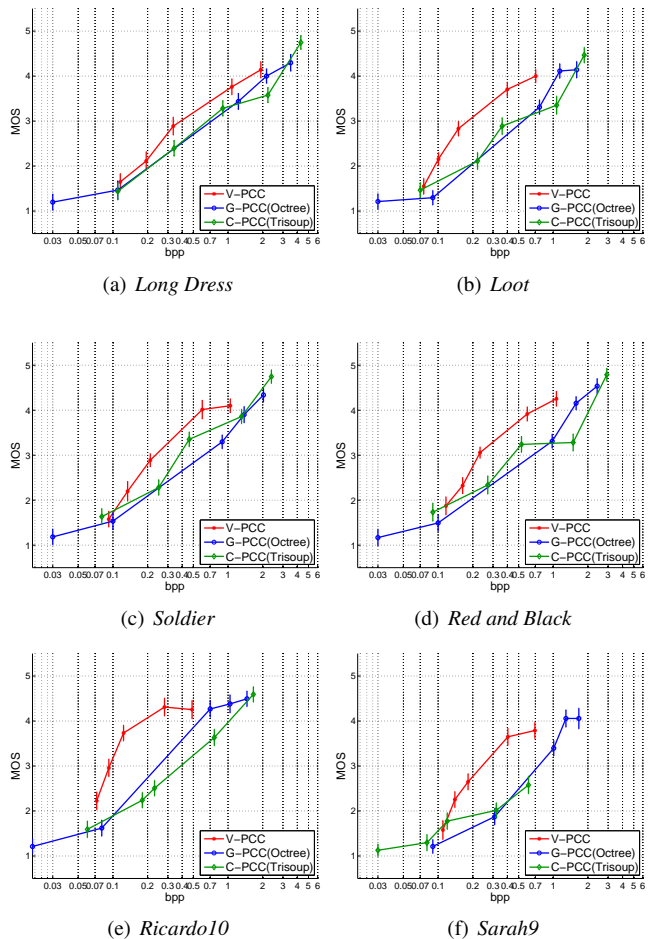
The Pearson Correlation Coefficient (PCC), the Spearman Rank Order Correlation Coefficient (SROCC), the Root-

**Table 1:** Equipment information and viewing distance per test laboratory.

	Monitor	Inches	Resolution	View Distance
UBI	Eizo ColorEdge CG318-4K	31.1"	4096x2160	1.2m ( $\pm 15$ cm)
UC	Monitor: Sony KD-49X8005C	49"	3840x2160	1.8 m (FV $\pm 30$ cm)
UNIN	Sony TV KD-55x8505C	55"	3840x2160	1.5 m (FV $\pm 15$ cm)
UTS	Eizo ColorEdge CG318-4K	31.1"	4096x2160	1.2 m ( $\pm 15$ cm)

**Table 2:** Subjects information per test laboratory.

	Males	Females	Overall	Age span	Average age
UBI	7	9	16	19-32	22
UC	7	8	15	18-54	28
UNIN	10	5	15	19-59	29
UTS	21	6	27	21-47	32



**Fig. 2:** Subjective scores pooled across test laboratories against bitrates for every codec.

Mean Squared Error (RMSE) and the Outlier Ratio (OR) are used to measure the linearity, monotonicity, accuracy, and consistency between the subjective scores obtained from each test laboratory, respectively. In particular, performance indexes considering every pair combination are presented in Table 3, without applying any fitting function and after linear regression. High correlation is observed in each case, with PCC and SROCC values always higher than 0.97. A sample of these results can be visually confirmed in Figure 3, where scatter plots of the MOS between different test laboratories is presented for 3 out of the 12 possible combinations.

#### 3.3. Benchmarking of Objective Metrics

The point-to-point (po2point) and point-to-plane (po2plane) metrics are employed to estimate geometric distortions [22], using the Mean Square Error (MSE) and the Hausdorff distance measures, as computed in [23]. The geometry PSNR ratio is also computed for each combination following the [16].

The color MSE is employed to quantify the color degradations of an encoded stimulus, after converting the default

RGB to the YCbCr colorspace, according to the implementation of [23]. Based on these measurements, the color PSNR values are also obtained and considered in the study. Moreover, the plane-to-plane (pl2plane) metric [24] is employed using a simple average, and the MSE distance.

For each stimulus, the normal vectors are estimated using a quadric fitting function on neighborhoods of fixed radius (i.e., 5), as implemented in CloudCompare [25] software. Finally, the symmetric error is used to obtain a total distortion value for each metric; that is, the maximum error after setting both the original and the distorted point cloud as a reference.

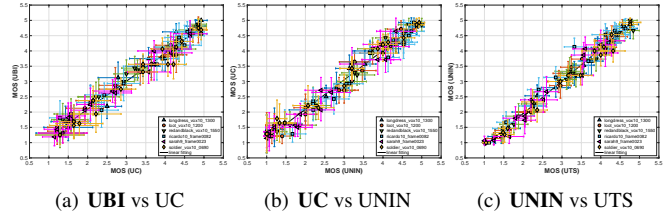
The PCC, SROCC, RMSE and OR performance indexes are computed between pairs of MOS and predicted MOS, to measure the performance of each metric against the subjective ground truth, as defined in the Recommendation ITU-T P.1401 [26]. The predicted MOS is obtained after applying the logistic fitting function on the objective quality

**Table 3:** Consistency of Subjective Scores across test laboratories (Bold text represents the assuming Ground Truth).

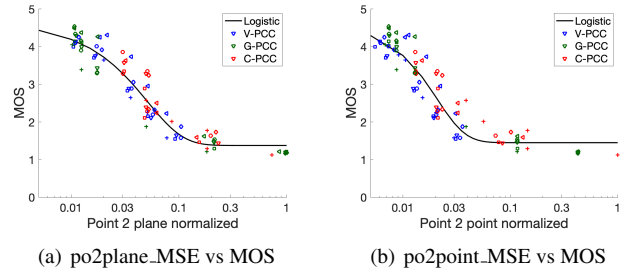
	Fitting	PCC	SROCC	RMSE	OR
<b>UBI vs UC</b>	no	0.987	0.980	0.213	0.041
	linear	<b>0.987</b>	<b>0.980</b>	0.183	0.052
<b>UBI vs UNIN</b>	no	0.984	0.982	0.254	0.093
	linear	<b>0.984</b>	<b>0.982</b>	0.201	0.052
<b>UBI vs UTS</b>	no	0.984	0.978	0.225	0.093
	linear	<b>0.984</b>	<b>0.978</b>	0.206	0.072
<b>UC vs UBI</b>	no	0.987	0.980	0.213	0.104
	linear	<b>0.987</b>	<b>0.980</b>	0.198	0.093
<b>UC vs UNIN</b>	no	0.986	0.982	0.219	0.072
	linear	<b>0.986</b>	<b>0.982</b>	0.206	0.072
<b>UC vs UTS</b>	no	0.989	0.987	0.200	0.145
	linear	<b>0.989</b>	<b>0.987</b>	0.183	0.072
<b>UNIN vs UBI</b>	no	0.984	0.982	0.254	0.322
	linear	<b>0.984</b>	<b>0.982</b>	0.222	0.260
<b>UNIN vs UC</b>	no	0.986	0.982	0.219	0.187
	linear	<b>0.986</b>	<b>0.982</b>	0.212	0.177
<b>UNIN vs UTS</b>	no	0.990	0.987	0.193	0.166
	linear	<b>0.990</b>	<b>0.987</b>	0.177	0.156
<b>UTS vs UBI</b>	no	0.984	0.978	0.225	0.270
	linear	<b>0.984</b>	<b>0.978</b>	0.212	0.208
<b>UTS vs UC</b>	no	0.989	0.984	0.218	0.261
	linear	<b>0.989</b>	<b>0.984</b>	0.183	0.196
<b>UTS vs UNIN</b>	no	0.988	0.986	0.215	0.177
	linear	<b>0.988</b>	<b>0.986</b>	0.184	0.140

**Table 4:** Performance indexes for Objective Metrics using Symmetric error and Logistic fitting.

Metric	PCC	SROCC	RMSE	OR
po2point_MSE	0.946	0.934	0.368	0.666
po2plane_MSE	<b>0.959</b>	<b>0.951</b>	<b>0.321</b>	<b>0.577</b>
PSNR_po2point_MSE	0.868	0.855	0.540	0.752
PSNR_po2plane_MSE	0.913	0.910	0.443	0.588
po2point_HAU	0.401	0.531	1.045	0.844
po2plane_HAU	0.534	0.613	0.966	0.877
PSNR_po2point_HAU	0.548	0.456	0.911	0.870
PSNR_po2plane_HAU	0.580	0.547	0.887	0.847
color_Y_MSE	0.876	0.892	0.551	0.766
color_Cb_MSE	0.683	0.694	0.834	0.844
color_Cr_MSE	0.594	0.616	0.918	0.844
color_Y_PSNR	0.887	0.892	0.525	0.688
color_Cb_PSNR	0.693	0.694	0.822	0.844
color_Cr_PSNR	0.626	0.617	0.890	0.855
pl2plane_AVG	0.922	0.910	0.439	0.600
pl2plane_MSE	0.925	0.912	0.432	0.611



**Fig. 3:** Linear fitting, for correlation between MOS obtained from different test laboratories (Bold text represents the assuming Ground Truth).



**Fig. 4:** Logistic fitting, for correlation between MOS and best-performing metrics. Each symbol  $\nabla, \square, \circ, \diamond, \triangleleft, +$  represents a different content in the display order of Figure 1.

scores. Based on our results, presented in Table 4, the best-performing metrics found to be the po2plane\_MSE, closely followed by the po2point\_MSE. Scatter plots with the MOS against these two metrics are presented in Fig. 4, including the logistic fitting function. In this result, though, it should be accounted the fact that the point clouds we use represent only one type of contents, i.e., human figures. As shown in [8], the performance of metrics can be remarkably improved when point clouds of the same category are clustered together. We aim at expanding our efforts in the near future, adopting the described evaluation framework with more diverse conditions and contents.

#### 4. CONCLUSIONS

A quality evaluation of the MPEG point cloud codecs is reported, unveiling the high compression performance of the MPEG V-PCC, when used for static point cloud coding. The adopted methodology is confirmed to be reliable and repeatable by producing consistent results in four independent test laboratories, despite the differences of equipment and viewing conditions. Moreover, several objective quality metrics are found to provide accurate predictions when compared to subjective scores obtained from our evaluation experiments. Among them, the po2plane metric with MSE is found to be the best-performing with PCC of 0.959 and SROCC of 0.951. The subjective data and test material are available online<sup>1</sup>.

<sup>1</sup><http://emerging.di.ubi.pt/icip2020PC.html>

## 5. REFERENCES

- [1] E. Alexiou, E. Upenik, and T. Ebrahimi, "Towards subjective quality assessment of point cloud imaging in augmented reality," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, 2017, pp. 1–6.
- [2] E. Alexiou et al., "Point cloud subjective evaluation methodology based on 2d rendering," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–6.
- [3] A. Javaheri et al., "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] 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.
- [5] S. Perry et al., "Study of subjective and objective quality evaluation of 3d point cloud data by the jpeg committee," *Electronic Imaging*, vol. 2019, no. 10, pp. 312–312–7, 2019.
- [6] E. M. Torlig et al., "A novel methodology for quality assessment of voxelized point clouds," in *Proceedings of SPIE*, ser. Applications of Digital Image Processing XLI, vol. 107520I, Sep 2018.
- [7] E. Zerman et al., "Subjective and objective quality assessment for volumetric video compression," *Electronic Imaging*, vol. 2019, no. 10, pp. 323–323–7, 2019.
- [8] E. Alexiou et al., "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.
- [9] H. Su et al., "Perceptual quality assessment of 3d point clouds," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3182–3186.
- [10] JPEG Pleno Database. [Online]. Available: <https://jpeg.org/plenodb/>
- [11] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies – a voxelized point cloud dataset," ISO/IEC JTC1/SC29/WG11/WG1 Doc. m40059/M74006, Geneva, Switzerland, Jan 2017.
- [12] MPEG 3DG, "G-PCC Codec Description v5," ISO/IEC JTC1/SC29/WG11 Doc. N18891, Geneva, Switzerland, Oct 2019.
- [13] S. Schwarz et al., "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2019.
- [14] E. Pavez et al., "Dynamic polygon clouds: representation and compression for vr/ar," *APSIPA Transactions on Signal and Information Processing*, vol. 7, p. e15, 2018.
- [15] R. L. de Queiroz and P. A. Chou, "Compression of 3d point clouds using a region-adaptive hierarchical transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016.
- [16] MPEG 3DG, "Common test conditions for point cloud compression," ISO/IEC JTC1/SC29/WG11 Doc. N18474, Geneva, Switzerland, Mar 2019.
- [17] —, "G-PCC Test Model v7," ISO/IEC JTC1/SC29/WG11 Doc. N18664, Gothenburg, Sweden, Jul 2019.
- [18] —, "V-PCC Codec Description," ISO/IEC JTC1/SC29/WG11 Doc. N18892, Geneva, Switzerland, Nov 2019.
- [19] —, "V-PCC Test Model v8," ISO/IEC JTC1/SC29/WG11 Doc. W18884, Geneva, Switzerland, Oct 2019.
- [20] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunications Union, Jan 2012.
- [21] MPV video player. [Online]. Available: <https://mpv.io>
- [22] D. Tian et al., "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3460–3464.
- [23] —, "Evaluation metrics for point cloud compression," Geneva, Switzerland, Jan 2017.
- [24] 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.
- [25] 3D point cloud and mesh processing software Open Source Project. CloudCompare. [Online]. Available: <http://www.danielgm.net/cc/>
- [26] ITU-T P.1401, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models," International Telecommunication Union, Jul 2012.