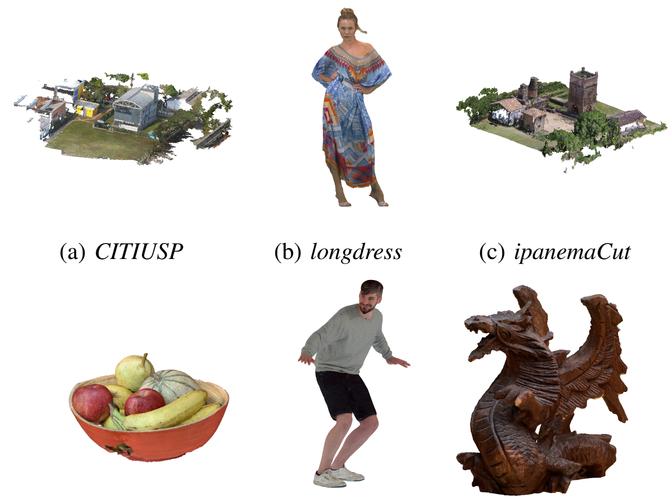
Influence of spatial rendering on the performance of point cloud objective quality metrics

Anonymous EUVIP 2022 submission

***Abstract*—The use of point clouds to digitally represent three-dimensional objects with both geometry and color attributes is rapidly increasing in several applications. Since storage and transmission of uncompressed point cloud data are often impractical, several lossy compression algorithms have been proposed, each exhibiting specific types of visual distortion. This creates a challenging environment for objective quality metrics, which might be effective only on a restricted number of distortion types and contexts. To evaluate the performance of the most recent objective quality metrics in predicting distortions as perceived by humans, a benchmarking study is conducted using subjective scores from observers examining models distorted with a conventional as well as a learning-based compression method, while rendering them on both a traditional flat monitor and a eye-sensing light field display. The results are then analyzed and conclusions are drawn on the correlation between recent state-of-the-art objective quality metrics and the subjective perception of human subjects while viewing point cloud data on different types of display devices.**

***Index Terms*—point clouds, immersive imaging, objective quality assessment, perceptual visual quality, eye-sensing light field display**



(a) *CITIUSP* (b) *longdress* (c) *ipanemaCut*

(d) *fruits* (e) *mitch* (f) *wooden_dragon*

Fig. 1: Point cloud contents included in the dataset.

I. Introduction

The way humans interact with digital media is constantly evolving. While the main public has been regularly consuming images and videos for decades, other imaging modalities that allow for an enhanced and more immersive experience are being increasingly adopted and further developed. Point clouds have been identified as a very significant alternative to represent three-dimensional objects and scenes for this purpose, partly because they represent shapes as sets of disconnected points similarly to how most acquisition devices output data.

However, due to the vast amount of data needed for its representation, point clouds need to be efficiently compressed prior to transmission or storage in the majority of applications. Lossy algorithms are often employed to achieve an higher compression ratio, at the cost of partially loosing information from the input data and thus generating distortions. While minor distortions may pass unnoticed to most human observers, larger degradation induce perceptual losses that harm user quality of experience. Moreover, different compression methods can add distortions of distinct natures, and accurately predicting their impact is not a trivial task. Although hand-crafted algorithms have been largely studied and employed in the MPEG compression standards G-PCC and V-PCC, learning-based methods have received increased attention due to their high compression performance. Therefore, being able to accurately estimate the impact that added distortions have on the subjective opinion is paramount to the development of compression algorithms with minimal impact on the visual quality. Although subjective experiments are a reliable method to evaluate this impact, they can be expensive and time consuming, not being appropriate for applications where quality evaluation needs to be performed quickly. Objective quality metrics are alternatives which aim at estimating the quality of a degraded point cloud through the computational analysis of its geometry and color attributes, generally comparing a distorted object to its corresponding reference. The performance of such metrics is benchmarked through the computation of performance indexes against mean opinion scores obtained from subjective experiments, determining if they properly model the human perception. Although many benchmarking experiments have been reported in the literature, the performance of existing solutions has been shown to fluctuate according to types of point clouds and distortions, making the search for objective quality metrics with good performance is still an open problem.

While flat monitors are used in the majority of subjective experiments reported in the literature, other works have explored devices that take better advantage of the immersion allowed by point clouds, such as augmented reality glasses and virtual reality headsets. In a recent study [1], authors conducted the first point cloud subjective experiment using the novel eye-sensing light field display (ELFD) [2], motivated by the attention that such technology recently attracted in various domains and applications [3], [4]. The study emphasized that, while the ELFD monitor allows for enhanced immersion and naturalness, its reduced size decreases the discriminatory power, which makes the subjective scores collected on such display statistically different from the ones collected on a flat monitor for the mid-range qualities. For this reason, an analysis on the performance of common objective quality metrics on both display types is necessary to assess the influence of spatial rendering on such methods.

In this paper, we perform a benchmarking of multiple state-of-the-art objective quality metrics against subjective scores obtained in an experiment using both a flat monitor and an eye-sensing light field display. The experiment included six test point clouds distorted with both conventional and learning-based compression methods. The impact of the rendering strategy on the performance of these metrics is evaluated and discussed. Performance indexes are also computed separately for each content of the dataset, as well as for each compression method.

The main contribution of this paper are:

- We perform the first benchmarking experiment of objective quality metrics with subjective scores collected from an eye-sensing light field display.
- We include in the experiment both a conventional and a learning-based codec, as the impact of learning-based artifacts on the performance of objective quality metrics is still understudied.
- We evaluate the performance of a large number of objective quality metrics, some of which have not yet been included in a benchmarking experiment so far.

The remaining of this paper is structured as follows: in section II, previous benchmarking experiments are reported and compared. Section III describes in detail the subjective experiment where the scores used as a basis for this study were collected. The following section IV gives a summary of the evaluated objective quality metrics, their computation as well as the statistical analysis that compared them to the subjective scores. In section V the obtained results are presented and discussed. Finally, section VI elaborates the main conclusions of this paper and outline possible directions for the future.

II. Previous Work

Assessing the performance of objective quality metrics is essential in order to define the most suitable compression method to be used in each scenario. Several previous studies have devoted efforts to this task.

Preliminary studies on the correlation between subjective and objective scores were presented in [5], [6], where the authors benchmarked objective quality metrics against subjective scores collected using a regular flat monitor and a head mounted display, on a dataset composed of geometry-only point clouds. The target distortions were octree-prunning and Gaussian noise, in which the former was found to be harder to model by the employed predictors.

Subsequent studies focused on models containing both geometry and color distortions, most of them generating distorted stimuli with MPEG compression standards G-PCC or V-PCC. The studies from [7], [8] evaluated point-to-point, point-to-plane, plane-to-plane [9] and color-based metrics against such distortions. [8] also included image-based metrics, which were applied in the projections of point clouds onto planes and pooled through the average. Other studies [10]–[13] employed distortions including one of MPEG compression methods or others similar in principle, also evaluating other factors such as the impact of the rendering [10], [12], dynamic models [13] and inter-laboratory consistency [11]. However, such studies don't incorporate modern objective quality metrics that are able to take into account both geometry and color attributes, and were all conducted on a conventional desktop setup for display.

A later study was conducted in [14], adopting a dataset targeting the V-PCC codec and with subjective scores collected using a head-mounted display. Metrics such as PCQM [15], PointSSIM [16] and PC-MSDM [17] were incorporated to the evaluation, while projection-based metrics were maintained. GraphSIM [18] was additionally evaluated in [19], while [20] included the PCM-RR metric. Both studies applied G-PCC and V-PCC compression to obtain distorted stimuli. [21] applied, downsampling and PCL [22] compression to point clouds, focusing on the performance of image quality metrics applied on projections. Finally, the performance of objective quality metrics against a subjective dataset with distortions generated by the MPEG compression standards together with two learning-based methods was evaluated in [23], using a crowdsourcing environment. This study observed low general performance on the whole dataset when the distortions are highly diverse, being the PCQM and PointSSIM the metrics showing the highest correlation values.

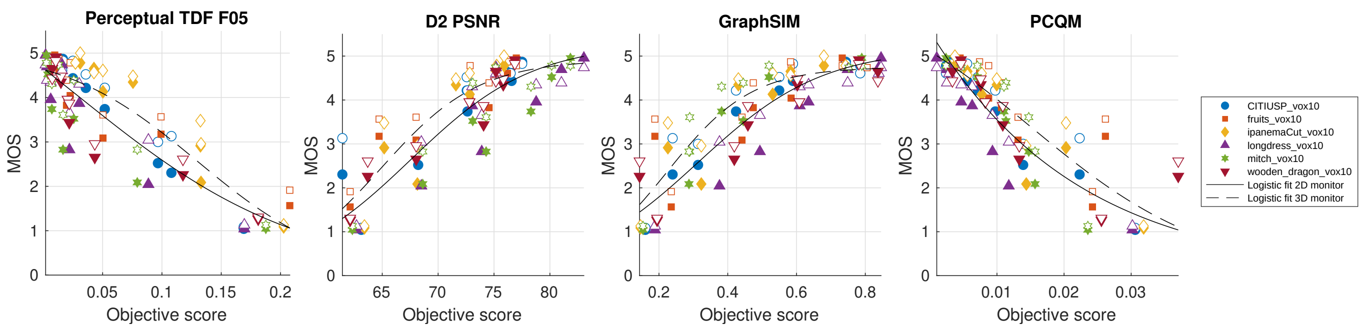
Regardless the large number of studies on the topic, no previous benchmarking study was conducted employing subjective scores collected on autostereoscopic light filed displays.

III. Subjective Experiment

The subjective dataset collected in [1] is employed for benchmarking of the objective quality metrics. The subjective visual scores in the form of mean opinion scores (MOS) were collected in a controlled environment by employing two distinct visualization devices, namely a DELL UltraSharp U3219Q *flat monitor* and a Sony Spatial Reality *eye-sensing light field display (ELFD)*. The native resolution of the DELL flat monitor is of 3840 x 2160 pixels with 31.5 inches diagonal size, while the Sony ELFD has resolution of 3840 x 2160 pixels and 15.6 inches diagonal size. The viewing conditions were set according to the ITU-R Rec. BT.500 [24] for the DELL flat monitor, while the recommendations from the manufacturer [1] were used for the Sony ELFD.

Six test point clouds were selected for the experiment, including representative contents from different use cases and applications, i.e. large-scale outdoor scenes (*CITISUP* and

[1] https://www.sony.net/Products/Developer-Spatial-Reality-display/en/develop/Specifications.html



(a) Perceptual TDF F05 (b) D2 PSNR (c) GraphSIM (d) PCQM

Fig. 2: Scatter plots of MOS scores against objective quality scores, along with the logistic fitting curve. The scores collected on the flat monitor and on the ELFD are denoted by filled and hollow symbols, respectively.

ipanemaCut), full-body human figures (*longdress* [25] and *mitch*) and small objects (*wooden_dragon* and *fruits*). Figure 1 illustrates the selected test point clouds. The different point cloud contents were distorted using two distinct compression methods, each one targeting four different quality levels, and selected to represent both conventional and learning-based compression artifacts. Specifically, the selected compression technologies for geometry data are the octree module of the MPEG coding standard Geometry-based Point Cloud Compression (G-PCC) [26] and the learning-based compression method presented in [27]. Both algorithms were combined with the lifting module from G-PCC for color compression and are referred here as *octree-lifting* and *slicing-lifting*, respectively.

The employed subjective protocol is the Simultaneous Double-Stimulus Impairment Scale (DSIS) with 5-scale rating and hidden reference, where all subjects carried out a short training session prior the beginning of the experiment. The order of the test stimuli was randomized, and the scores of the first four dummy stimuli were excluded. A total of 23 suitable subjects participated in the experiment, being 11 females and 12 males, all having normal or corrected-to-normal vision capability. The subjects had age span between 18 and 25 years, being their average and median age respectively 21.35 and 21 years.

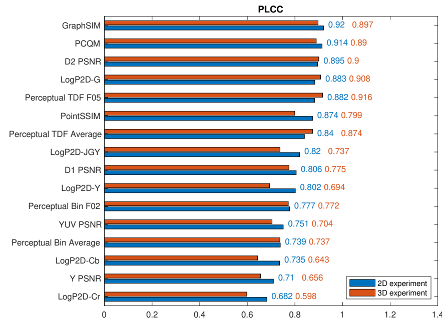
IV. Benchmarking

The subjective opinion scores obtained in [1] are here used to evaluate the performance of several objective quality metrics. The selected metrics were computed over the distorted dataset and their ability to predict the mean opinion scores was assessed through performance indexes in different settings, as explained in the following subsections.

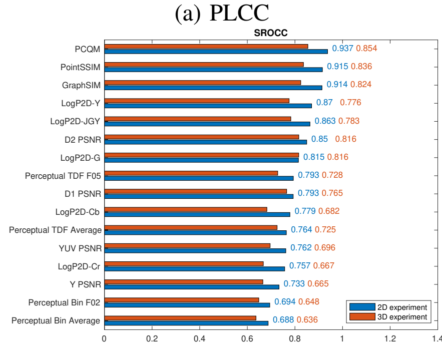
A. Objective quality Metrics

This study evaluates metrics from three different categories: geometry-only, which take into account only topology degradation; color-only that consider solely color attributes; and joint metrics that simultaneously receive as input both geometry and color. From the first category, D1 (point-to-point) and D2 (point-to-plane) PSNR [28], point-to-distribution [29] and a learning-based metric [30], which is here referred to as perceptual loss, were selected. D1 and D2 PSNR are widely used to evaluate geometry compression algorithms, and rely on euclidean distances computed between points of the reference and distorted models, either taken directly on D1 or only over local normal planes in D2. The distance values per point are pooled through the mean squared error, and the PSNR value is finally used. The point-to-distribution metric leverages the Mahalanobis distance to take into consideration local neighbourhoods during computation. Finally, the perceptual loss computes the difference between feature vectors generated with an autoencoder model applied on point cloud blocks, which are represented either as binary voxels or as truncated distance fields (TDF).

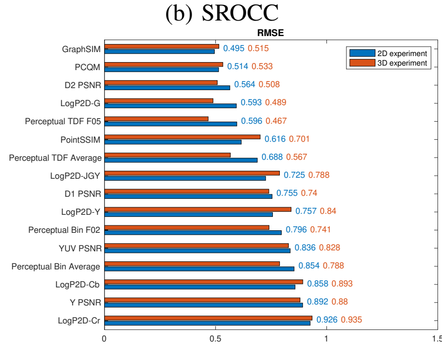
Even if such metrics are able to predict subjective scores accurately in some settings, they are not capable of modeling human perception regarding color distortion and are therefore an incomplete model of our visual system. PCQM [15] proposes a solution by computing a set of features based on either curvature or lightness, which were then pooled into a single score through a weighting vector optimized on the subjective dataset from [12]. PointSSIM [16] compares luminance-based features computed through statistical estimators over local neighbourhoods from both the distorted and the reference models. The best set of parameters is also optimized against subjective scores from [8] and [12]. Similarly, GraphSIM [18] computes the color-based local significance feature over graphs built around keypoints, and pools a final score averaging across local graphs. The previously mentioned point-to-distribution metric was extended to deal with color distortions in [31] using luminance and chrominance channels separately. A joint metric is also proposed by averaging the geometry and luminance-based point-to-distribution. Finally, the color-based PSNR metric extensively used for image quality assessment can be also computed for point clouds by establishing correspondence between points in the reference and distorted models through the nearest neighbour and comparing their color values. In this study, we consider the PSNR computed only in the Y channel, as well as a weighted average over the three YUV channels following the equation:



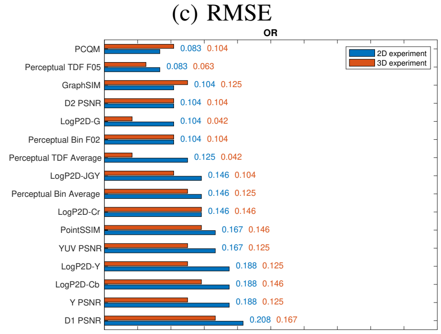
(a) PLCC



(b) SROCC



(c) RMSE



(d) OR

Fig. 3: Performance indexes calculated over the entire dataset.

$$PSNR_{YUV} = \frac{6 * PNSR_Y + PSNR_U + PSNR_V}{8} \quad (1)$$

In this work, for the computation of D1, D2 and color-based PSNR, the software version 0.13.5 of the MPEG suite was employed. The source code made available by the authors was used for the perceptual loss metric, using both binary and TDF representation. Point clouds were partitioned into blocks of resolution 64 and the final value was averaged among all occupied blocks. Both the average value across all features and the best performing feature over the whole dataset were used, which was found to be feature 2 for the metric for the training with binary voxels and feature 5 for the training with the TDF representation. For PointSSIM, voxelization to a target bit depth 9 was employed as a pre-processing step. Luminance-based features were employed using variance as a statistical estimator, with neighbourhood size 12. The logarithmic version was selected for the point-to-distribution metrics, which were computed for geometry only, for each color channel separately and finally with the joint metric using both color and geometry. The source code provided by the authors of PCQM and GraphSIM were used with default configuration.

B. Statistical Analysis

The performance of the objective quality metrics in Section IV-A are assessed by computing the Pearson Linear Correlation Coefficient (PLCC) and Spearman's Rank Correlation Coefficient (SROCC) between the objective and the subjective scores. Correlation values are expected to be in the range ± 1, where ± 1 denotes perfect positive or negative correlation and 0 no correlation.

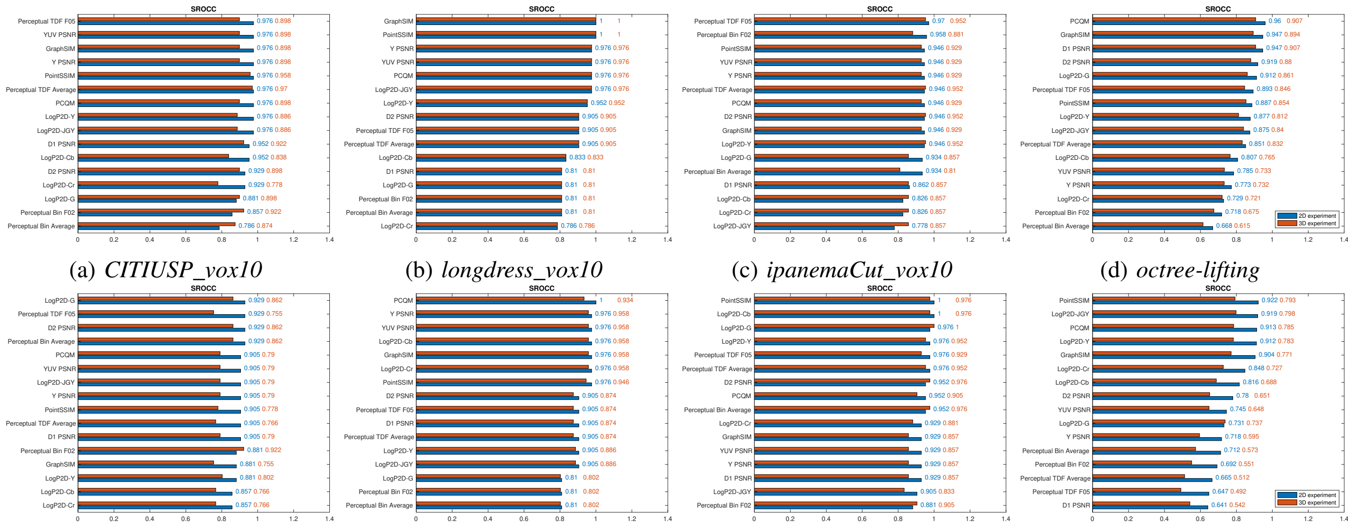
Additionally, the performance of the objective quality metrics is assessed with two further performance indexes, i.e., the Root Mean Square Error (RMSE) and the Outlier Ratio (OR) [32], which are based on the standard deviation of the prediction error and on the number of "outlier" points which exceeds the 95% confidence interval (CI) respectively. In this case, lower RMSE or OR values indicate better performance.

As subjective quality scores usually follow a non-linear behaviour, a common practice is to remove this non-linearity through a least-squares regression procedure [33]. Specifically, to remove the non-linearity, a logistic function without offset [24] is fitted to the data:

$$y = \frac{a}{1 + exp(-b * (x - c))} \quad (2)$$

where the a,b and c parameters are initialized to zero. Following the recommendations in [33], the Matlab functions `nlinfit` and `nlpredci` are adopted for the fitting procedure. Figure 2 shows the MOS against the objective scores after the fitting procedure.

The computation of the performance indexes were performed separately for the subjective scores obtained with the flat monitor and with the ELFD. These values were first computed across all the distorted stimuli from the experiments, in order to evaluate the performance of the metrics on the entire experiment. This whole process was also repeated after grouping the stimuli into the different compression methods, as well as into the different contents, in order to evaluate the influence of these factors.



(a) *CITIUSP_vox10* (b) *longdress_vox10* (c) *ipanemaCut_vox10* (d) *octree-lifting*

(e) *fruits_vox10* (f) *mitch_vox10* (g) *wooden_dragon_vox10* (h) *slicing-lifting*

Fig. 4: Spearman correlation computed for each codec and content.

V. Performance and Analysis

Figure 2 shows the scatter plots along with the curve of the fitting functions for four representative objective quality metrics. Likewise, the performance indexes values for the entire dataset can be visualized in Figure 3.

Figure 2 reveals that the subjective MOS scores are generally higher in the experiment using the ELFD, which is made explicit by the comparison between the curves of the fitting function. As noted in [1], this is mainly due to the fact that the ELFD display has a smaller size than the flat monitor, and thus it is harder for subjects to differentiate between the reference and distorted point clouds when the degradation is light. Therefore, more stimuli were classified with very similar MOS scores. Since these scores are harder to differentiate, this naturally leads to lower correlation values in the experiment with the ELFD, as shown in Figure 3.

The correlation values allow also to discern that the best performing metrics are GraphSIM and PCQM, which rank very highly for both PLCC and SROCC. The latter metric has been already reported to display high performance in the presence of learning-based artifacts [23], as well as with other distortions such as noise, downsampling and octree compression [21]. Moreover, even if the dataset contained point clouds with variation in the point density, no large performance drop of the objective metrics was observed due to compression such as in the study from [19]. Other metrics such as PointSSIM, point-to-distribution, D2 PSNR and the perceptual loss also achieved satisfying performance indexes. For the PointSSIM metric, this result is also in line with [23]. The correlation coefficients don't allow to define a clear ranking between the tested versions of the point-to-distribution metric, since the luminance-based, geometry-based and joint metrics are ranked in different orders for the PLCC and the SROCC. The D2 PSNR metric is shown to perform consistently better than D1 PSNR, showing the benefit of implicitly considering local neighbourhoods through the normal vector estimation. Finally, the performance of the perceptual loss indicates the potential of neural networks for modeling subjective perception. This result is important because this netweork is based in an autoencoder that was not trained to distinguish added distortion or predict subjective scores. Rather, the only learning goal of the encoder is to learn meaningful features that allow a faithful reconstruction by the decoder. These findings also corroborate that the representation of the point cloud TDF is also useful for this purpose. Moreover, not all the learned features carry the same correlation with human perception, and selecting the best features improves the prediction power of the metric.

Figure 4 depicts the Spearman correlation values between the objective quality metrics and the MOS scores separated by codec and by content. Since the correlation is computed within a smaller number of stimuli, the obtained values are naturally higher. For that reason, all the analyzed objective quality metrics present comparable performance in the content-based scenario. It is however possible to observe that these values are slightly lower for *fruits_vox10*, possibly due to its lower point density.

The separate evaluation by codec reveals the difference in performance of the objective metrics when considering different types of distortions. In general, the correlation values with *octree-lifting* are higher, suggesting that the artifacts generated by this codec are more easily captured by the considered objective metrics. Indeed, compressing geometry with the octree module result in a uniform downsampling of the point cloud, while distortion caused by the learning-based codec are usually less predictable. This might also explain why geometry-based metrics rank lower with *slicing-lifting*, for which all the top ranking metrics account for color distortions. This is particularly the case for the perceptual loss metric, which doesn't seem to capture well learning-based artifacts, probably caused by the lack of generalization power of the network to artifacts unseen during its training.

VI. Conclusions

In this study, a set of point-based objective quality metrics are benchmarked against subjective scores obtained in a experiment that employed both a flat monitor and an eye-sensing light field display as rendering devices. Useful insights are provided related to the influence on the performance of the predictors by not only the visualization strategy, but also by conventional and learning-based compression artifacts as well as different content types. In particular, metrics that effectively combine geometry and color information, such as PCQM and GraphSIM, are reported to achieve the highest performance. Future studies may further evaluate these metrics with a larger variety of learning-based compression artifacts, including different coding algorithms for color attributes.

Acknowledgments

The acknowledgments are removed for double blind purpose.

References

[1] D. Lazzarotto, M. Testolina, and T. Ebrahimi, "On the impact of spatial rendering on point cloud subjective visual quality assessment," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2022.

[2] K. Aoyama, K. Yokoyama, T. Yano, and Y. Nakahata, "Eye-sensing light field display for spatial reality reproduction," in *SID Symposium Digest of Technical Papers*, vol. 52, no. 1. Wiley Online Library, 2021, pp. 669–672.

[3] "Spatial Reality Display ELF-SR1 Technical Background," https://www.sony.com/en/SonyInfo/sony_ai/siggraph2021/assets/img/event_session04-01.pdf, accessed: 2022-03-12.

[4] T. Itamiya, M. To, T. Oguchi, S. Fuchida, M. Matsuo, I. Hasegawa, H. Kawana, and K. Kimoto, "A novel anatomy education method using a spatial reality display capable of stereoscopic imaging with the naked eye," *Applied Sciences*, vol. 11, no. 16, p. 7323, 2021.

[5] 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)*. IEEE, 2017, pp. 1–6.

[6] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL*, vol. 10396. International Society for Optics and Photonics, 2017, p. 103961H.

[7] S. Perry, H. P. Cong, L. A. da Silva Cruz, J. Prazeres, M. Pereira, A. Pinheiro, E. Dumic, E. Alexiou, and T. Ebrahimi, "Quality evaluation of static point clouds encoded using mpeg codecs," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3428–3432.

[8] 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, 2019.

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

[10] 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*, vol. 10752. International Society for Optics and Photonics, 2018, p. 107520I.

[11] L. A. da Silva Cruz, E. Dumić, E. Alexiou, J. Prazeres, R. Duarte, M. Pereira, A. Pinheiro, and T. Ebrahimi, "Point cloud quality evaluation: Towards a definition for test conditions," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–6.

[12] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Point cloud rendering after coding: Impacts on subjective and objective quality," *IEEE Transactions on Multimedia*, vol. 23, pp. 4049–4064, 2020.

[13] 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.

[14] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong, "Subjective quality database and objective study of compressed point clouds with 6dof head-mounted display," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4630–4644, 2021.

[15] 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.

[16] 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.

[17] G. Meynet, J. Digne, and G. Lavoué, "Pc-msdm: A quality metric for 3d point clouds," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–3.

[18] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[19] Y. Liu, Q. Yang, Y. Xu, and Z. Ma, "Which one is better: Assessing objective metrics for point cloud compression," *arXiv preprint arXiv:2109.07158*, 2021.

[20] Q. Liu, H. Su, Z. Duanmu, W. Liu, and Z. Wang, "Perceptual quality assessment of colored 3d point clouds," *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[21] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3d-to-2d projection-based exploration," *IEEE Transactions on Multimedia*, vol. 23, pp. 3877–3891, 2020.

[22] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1–4.

[23] D. Lazzarotto, E. Alexiou, and T. Ebrahimi, "Benchmarking of objective quality metrics for point cloud compression," in *IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, no. CONF, 2021.

[24] Recommendation ITU-R BT.500-14, "Methodologies for the subjective assessment of the quality of television images," *International Telecommunication Union*, 2019.

[25] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies - a voxelized point cloud dataset," ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva, Switzerland, Jan. 2017.

[26] 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.

[27] N. Frank, D. Lazzarotto, and T. Ebrahimi, "Latent space slicing for enhanced entropy modeling in learning-based point cloud geometry compression," in *2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

[28] 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.

[29] 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.

[30] M. Quach, A. Chetouani, G. Valenzise, and F. Dufaux, "A deep perceptual metric for 3d point clouds," *Electronic Imaging*, vol. 2021, no. 9, pp. 257–1, 2021.

[31] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "A point-to-distribution joint geometry and color metric for point cloud quality assessment," *arXiv preprint arXiv:2108.00054*, 2021.

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

[33] ITU-R Tutorial, "Objective perceptual assessment of video quality: Full reference television," *International Telecommunication Union*, 2004.