Towards a multiscale point cloud structural similarity metric

Davi Lazzarotto, Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
davi.nachtigalllazzarotto@epfl.ch, touradj.ebrahimi@epfl.ch

***Abstract*—Point clouds are effective data structures for the representation of three-dimensional media and hence adopted in a wide range of practical applications. In many cases, the portrayed data is expected to be visualized by humans. After acquisition, point clouds may undergo different processing operations such as compression or denoising, potentially affecting their perceived quality. Although subjective experiments are still the most reliable form of assessing the intensity of degradation, they are expensive and time-consuming, pushing many systems to depend on objective metrics. Such algorithms are used to model the human visual system, and their performance is usually assessed through their correlation with subjective visual quality scores. In this paper, an objective quality metric capable of evaluating distortions between a reference and a distorted point cloud at multiple scales is presented. The proposed metric is based on the point cloud structural similarity metric (PointSSIM), which computes a score based on the difference between statistical estimators obtained on the distribution of the luminance attribute over local neighborhoods. A collection of PointSSIM scores is produced for multiple scales obtained through the voxelization of both models at different bit depth precisions. These scores are then pooled through a weighted sum, with the importance of each scale being defined through logistic fitting to subjective mean opinion scores, producing one MS-PointSSIM score. Three datasets were employed for fitting and performance assessment, demonstrating a clear advantage of the proposed metric when compared to the single-scale baseline. Moreover, the presented MS-PointSSIM is shown to be the best predictor according to the average Pearson correlation coefficient across the three datasets when compared to state-of-the-art metrics.**

***Index Terms*—Point cloud, objective metric, quality**

I. Introduction

The use of point clouds to represent three-dimensional objects and scenes has been growing in recent years. Due to the large amount of data needed to represent them, efforts have been devoted to developing efficient compression solutions. Lossy methods are capable of achieving higher compression ratios by excluding details that are not easily perceived by humans, inducing a trade-off between visible degradation and compression ratio. During the development of compression algorithms, it is essential to correctly estimate the impact of added distortions on the quality of decoded point clouds. For applications where humans are the end users, quality is defined subjectively and is more accurately measured through human observation. It is common for such measures to be obtained during subjective experiments where subjects are asked to observe a distorted point cloud and its reference, attributing a score to the impairment observed between the models.

Even if subjective experiments are the most reliable method for point cloud quality estimation, many applications require faster and less expensive techniques. Objective quality metrics are usually used for this purpose by computing a score based on the geometric and color features of the model. Their goal is to accurately predict the subjective score associated with a point cloud, and are often evaluated through the correlation between their result and mean opinion scores (MOS) from subjective experiments. The most recent metrics proposed in the state of the art can be classified as either point-based or image-based. While metrics from the first category are based on operations applied directly in the three-dimensional space, image-based metrics first project the point cloud into planes and then estimate the quality of the obtained projections. Such metrics have the advantage of leveraging prior research developments on image quality assessment but are highly dependent on the position selected for the projection planes as well as on the rendering method.

For that reason, point-based metrics have been mostly preferred for the evaluation of point cloud compression methods. Early metrics such as point-to-point and point-to-plane [1] are applied directly in the geometry, while color PSNR computes the difference between color values of neighboring points. Although these metrics have been leveraged during the development of MPEG compression standards G-PCC [2] and V-PCC [3], recent studies indicate that they do not correlate well with human perception across different types of compression artifacts [4], especially those generated by codecs based on neural networks.

The fact that these metrics concentrate only either on geometry or color while neglecting the other is another disadvantage since they are not able to fully model the human visual system. Recently proposed objective quality metrics attempt to solve this issue by either explicitly combining geometry and color features [5] or implicitly considering geometry distortions by pooling color attributes across local neighborhoods [6], [7]. In particular, PointSSIM [7] is inspired by the image-based

The authors would like to thank the Swiss National Foundation for Scientific Research (SNSF) under grant number 200020_207918 for funding this research.

structural similarity metric (SSIM) [8], computing statistical features on the luminance channel and comparing feature maps from both point clouds. The correlation between image-based SSIM and human perception is usually higher when scores are computed at multiple scales and pooled together [9], even when considering artifacts created by learning-based codecs [10].

Inspired by the multiscale SSIM (MS-SSIM) metric, this paper evaluates the use of the PointSSIM metric at multiple scales, which are obtained through voxelization at different bit depth precisions. Single-scale scores are produced through the PointSSIM metric at each scale and pooled together through a weighted sum, with weights obtained from logistic fitting to subjective scores. In addition, a modification is proposed to the original PointSSIM implementation by using range searching rather than k-Nearest Neighbors (kNN) from local neighborhoods. A software implementation of the metric is made publicly available [1].

II. Related work

As one of the simplest methods to evaluate the distortion of a point cloud, the point-to-point metric computes the average distance between each point in the evaluated point cloud and its nearest neighbor in the reference. Similarly, point-to-plane [1] considers the normal vectors of the reference and only computes the distance perpendicular to the normal plane. On the other hand, the angular similarity metric [11] produces a quality score equivalent to the average angular distance between normal vectors. The point-to-distribution [12] is computed through the Mahalanobis distance to achieve a scale-invariant measure, and the same authors also proposed improvements to traditional PSNR-based metrics [13]. Newer approaches [14] attempt to leverage 3D convolutional networks in an autoencoder architecture and compute the distance in the latent space. Such metrics rely only on geometry distortions to achieve a final quality score.

The color-based PSNR metrics compute a metric value similar to its image-based counterpart, establishing correspondence between points on both point clouds through the nearest neighbor. PCQM [5] computes color-based and geometry-based features and obtains a single score through a weighted sum with weights obtained after fitting on a subjective dataset [15]. GraphSIM [6] constructs graphs around key points and computes three color gradient features over each graph, which are fused into one similarity score. PointSSIM forms neighborhoods around each point and generates a score from statistical estimators of the distribution of either geometry-based or color-based attributes. The GQI [16] uses convolutional layers to compute structural distortion on geometry, curvature, and color around randomly placed patches. Apart from the simple color-based PSNR, the remaining metrics are usually better suited to model the human visual system.

[1] https://github.com/mmspg/ms-pointssim

III. Single-scale structural similarity

The structural similarity used to compute single-scale scores is based on the local distribution of attributes. In the original work where this metric was first presented [7], four attributes were proposed: geometry, normal vectors, curvature values, and luminance. While the three first attributes are derived only from spatial coordinates, the latter is based on color attributes. For both the reference and the evaluated point clouds, different estimators are computed to represent the statistical distribution of the selected attribute over local neighborhoods. In the original PointSSIM metric, these neighborhoods are defined as the k-Nearest Neighbors of each point. A statistical estimator is calculated over each neighborhood and assigned as a feature F to each target point. In this paper, the following statistical estimators of an attribute A are evaluated: the mean μ_A, the standard deviation σ_A, the variance σ_A^2, the coefficient of variance COV_A, the mean absolute deviation μAD_A, the median absolute deviation $m\text{AD}_A$ and the quartile coefficient of dispersion QCD_A. The last four estimators were computed using Equations 1a to 1d.

$$\text{COV}_A = \frac{\sigma_A}{\mu_A} \tag{1a}$$

$$\mu\text{AD}_A = \mathbb{E}[A - \mu_A] \tag{1b}$$

$$m\text{AD}_A = \mathbb{E}[A - m_A] \tag{1c}$$

$$\text{QCD}_A = \frac{Q_A(3) - Q_A(1)}{Q_A(3) + Q_A(1)} \tag{1d}$$

After feature computation, the nearest neighbor q from the evaluated point cloud D to each reference point p in the reference R is detected. A similarity value $S(p)$ is then attributed to each point p through Equation 2, where ϵ is a small constant defined as the machine rounding error.

$$S(p) = \frac{|F(q) - F(p)|}{\max\{|F(q)|, |F(p)|\} + \epsilon} \tag{2}$$

The similarity score S_R between reference R and evaluated D point clouds is then obtained through the average between the similarity values of all points in R. The entire process is then repeated using the point cloud D as the reference, resulting in another similarity score S_D. The symmetrical PointSSIM value S_{RD} is finally selected as $S_{RD} = \min\{S_R, S_D\}$.

Although four different attributes were implemented on the PointSSIM metric [7], in this paper, only the luminance was retained. While geometry-based features are left untouched if only the color attributes of a point cloud are distorted, luminance-based features capture geometry distortions indirectly because point displacements alter the statistical distribution of color over local neighborhoods, being the only attribute affected by all types of distortion. Moreover, luminance was already found to be the attribute with predictions better correlated with subjective scores [7]. A metric based on color attributes can also be used as the loss function for

training learning-based compression algorithms for point cloud attributes in future works. While recent image compression algorithms [17] currently use different objective metrics for training such as the multiscale structural similarity [9] and learning-based metrics such as LPIPS [18], point cloud compression methods [19] still estimate distortion with the PSNR, which correlates poorly with human perception.

A modification to the algorithm used to define neighbors for feature computation is proposed in this paper, with range search being studied in addition to the original implementation with kNN. The proposed change mainly aims to better deal with differences in point density between the reference and distorted point cloud: if both models have widely different point densities for a given region in space, the neighborhoods formed by the kNN algorithm span over much larger volumes for the sparser point cloud relatively to the denser. The computed features may therefore diverge even if the spatial distribution of the selected attribute is similar, only due to the extent of the surface over which the distribution is sampled. In contrast, with range search, all points lying within a sphere with a fixed radius centered on the target point are selected for the neighborhood, solving the aforementioned problem and potentially leading to better correlation with subjective scores.

As an alternative, both neighboring search methods could have been combined, where candidate neighborhoods are formed with each method and the final neighborhood being either the union or the intersection between the two candidate neighborhoods. This analysis is not conducted in this paper and is deferred for future work.

IV. Multiscale structural similarity

The objective quality metric presented in this paper combines PointSSIM scores computed for multiple scales. The motivation behind this design is that subjective perception can be affected both by distortions at coarser and finer scales. Therefore, better correlation can be achieved by assigning one single score that pools together degradations at multiple scales, similar to the image-based MS-SSIM metric.

The metric computation starts with the voxelization of both the reference point cloud R and the evaluated D at different bit depth precisions to achieve different scales. For a precision value p, the coordinates of the point clouds are scaled to fit into a bounding box of size 2^p and then rounded to the nearest integer. If more than one point is quantized to the same position, their color attributes are averaged in the voxelized point cloud. In this paper, voxelization is applied with p values ranging from $p_i = 6$ to $p_f = 10$. Each point cloud pair is then served as input to the single-scale metric, producing one score S for each scale. The final MS-PointSSIM score M is then generated as the weighted sum over the scores obtained for each scale, as depicted in Equation 3, with v_p representing the voxelization operation with precision p.

$$M(R, D) = \sum_{p=p_i}^{p_f} w_p \times S(v_p(R), v_p(D)) \tag{3}$$

The weight values w_p are obtained by fitting the metric scores to a subjectively annotated dataset. In order to keep the multiscale metric value bound to the same range as the single-scale scores, the weights w_p are constrained to have their sum equal to 1. In order to ensure the constraint, the vector $\mathbf{w}$ is parametrized by Equation 4.

$$\mathbf{w} = \text{softmax}(\mathbf{b}) \tag{4}$$

A logistic function is used to obtain PMOS values (predicted MOS) from the MS-PointSSIM scores. Let us consider a subjectively annotated dataset containing N point cloud pairs $(D_i,\ R_i)\ \forall\ 1 \leq i \leq N$, where D_i is a distorted point cloud and R_i its corresponding reference. Each pair received MOS_i equal to the average attributed score across all subjects scaled to the range $[0, 1]$. The PMOS value for each stimulus is given by Equation 5, where the score $M(D_i, R_i)$ is computed using Equation 3.

$$\text{PMOS}_i = \frac{1}{1 + \exp\{\alpha \cdot [M(D_i, R_i) - \delta]\}} \tag{5}$$

The set of parameters $\beta = \{\mathbf{b}, \alpha, \delta\}$ is obtained through the minimization of the least-squares error between the MOS and PMOS values. An implementation from the *scipy* python library of the trusted region reflective algorithm is used for the optimization, with initial values of $\mathbf{b} = \mathbf{0}$, $\delta = 0$ and $\alpha = 1$.

V. Evaluation conditions

Three subjectively annotated datasets containing point clouds distorted using different compression algorithms were selected for fitting and evaluation. The IST Rendered Point Cloud Quality Assessment Dataset (**IRPC**) [15] contains six point clouds extracted from the MPEG repository compressed with three codecs: V-PCC, G-PCC using the *TriSoup* module for geometry coding, and the compression method from the Point Cloud Library (PCL) [20]. Two point clouds from the dataset are originally represented with a bit depth precision of 10, and while the remaining models use a bit depth precision of 12, they were also reduced to a precision of 10 for V-PCC encoding. The **ICIP2020** dataset [21] contains six point clouds compressed with both G-PCC and V-PCC at five quality levels. The two geometry coding modules from G-PCC, namely *octree* and *TriSoup* were employed. In this paper, the *Sarah* model was excluded from the dataset due to its lower voxelization precision. Finally, the **SR-PCD** [22] was selected, containing six models compressed with two codecs, namely G-PCC with the *octree* coding module and a learning-based geometry compression method [23] combined with the *lifting* module from G-PCC for color coding. The combined evaluation dataset consists of a total of 177 distorted point clouds, 54 from IRPC, 75 from ICIP2020, and 48 from SR-PCD.

The single-scale PointSSIM metric was first computed on the entire combined dataset with different sets of parameters. The seven proposed statistical estimators were used, with a voxelization bit depth precision p ranging from 6 to 10.

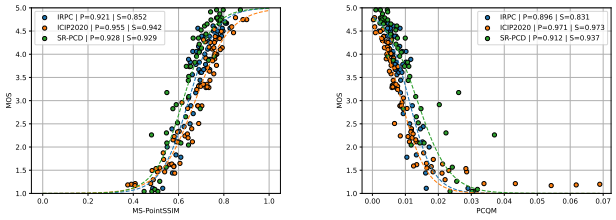
TABLE I: Correlation coefficients for single-scale PointSSIM

Combination	PLCC			SROCC			Average	
	IRPC	ICIP2020	SR-PCD	IRPC	ICIP2020	SR-PCD	PLCC	SROCC
8-bit, mAD, $k = 24$	0.859	0.906	0.904	0.790	0.878	0.925	0.890	0.864
8-bit, σ^2, $k = 6$	0.838	0.926	0.910	0.738	0.902	0.925	0.891	0.855
8-bit, σ^2, $k = 48$	0.771	0.916	0.915	0.682	0.887	0.931	0.867	0.834
8-bit, mAD, $r = 3$	**0.883**	0.914	0.892	**0.800**	0.893	0.926	**0.896**	**0.873**
10-bit, σ^2, $r = 1$	0.551	**0.957**	0.679	0.458	**0.943**	0.684	0.729	0.695
10-bit, μAD , $r = 2.5$	0.751	0.911	**0.934**	0.656	0.880	**0.936**	0.865	0.824

Both kNN and range searching were used to establish local neighborhoods, the first with $k = 6, 12, 24$ and 48, and the latter with ranges set to $r = 1, 1.5, 2, 2.5, 3, 3.5, 4$ and 4.5. All possible combinations of estimator, scale, and neighborhood were used, and PMOS values were produced through least-squares logarithmic fitting for each combination, with separate parameters for each dataset. Finally, the Pearson and Spearman correlation coefficients were computed between MOS and PMOS.

The single-scale scores were used to obtain the weight vector $\mathbf{w}$ using three different configurations to combine scores from different scales. At first, only single-scale scores from the same estimator and the same neighborhood size were combined. In the second configuration, different estimators were allowed to be combined at different scales, while maintaining the same neighborhood size. Finally, single-scale scores computed with distinct estimators and neighborhood sizes were allowed to be pooled together, with the constraint that the same search method had to be used, i.e. kNN and range search. These three approaches are further denominated as SESN (Single-estimator Single-neighborhood), MESN (Multi-estimator Single-neighborhood), and MEMN (Multi-estimator Multi-neighborhood), respectively. Each configuration allows for a higher amount of combinations than the previous one, with 84 for SESN, approximately 201 thousand for MESN, and more than 560 million for MEMN. For both SESN and MESN, one weight vector $\mathbf{w}$ was obtained for all possible combinations separately for each dataset. Given the massive amount of possible combinations for MEMN, only 20 thousand random combinations were tested per dataset. Similarly to the single-scale PointSSIM, correlation coefficients were obtained to evaluate the performance of the metric.

Since MESN and MEMN configurations contain larger parametric spaces, it is expected that their best-performing combinations will improve in relation to SESN. However, such correlation values can be the result of specific combinations that are overfitted to the data which will not generalize well for unseen distortions. For this reason, the best-performing combinations for each configuration and each dataset were also tested on the other datasets. In this case, although the weight vector $\mathbf{w}$ was kept the same as the dataset to which the combination was originally fit, a new logistic fitting of the parameters α and δ of Equation 5 was conducted in order to account for differences on the protocol used during the subjective experiments. For SESN, only the 10 best-performing combinations were tested for other datasets, while 100 were used for MESN and 200 for MEMN.



(a) MS-PointSSIM (b) PCQM

Fig. 1: Scatter plots of objective scores against mean opinion scores for different metrics

Finally, in order to compare the performance of MS-PointSSIM with other state-of-the-art metrics, the following objective quality metrics were computed on the three evaluated datasets: point-to-point PSNR (D1 PSNR), point-to-plane PSNR (D2 PSNR) [1], Y and YUV PSNR, PCQM [5] and GraphSIM [6]. Y PSNR was computed only on the luminance channel, and YUV was computed through a weighted average across all channels using a $[6, 1, 1]$ weighting scheme between luminance and the two chrominance channels. All evaluated metrics were separately fitted to the subjective scores of each dataset using a logistic function, and the correlation coefficients were obtained.

VI. Results and Discussion

The Pearson and Spearman correlation coefficients (PLCC and SROCC) computed for the single-scale PointSSIM combinations achieving higher Pearson correlation coefficient for each dataset are presented in Table I. The upper half of the table includes only combinations based on kNN, while combinations of the lower half group neighborhoods using range search. It can be seen that the best-performing combination is different for each dataset. For this reason, the average correlation values across all datasets are also displayed in the columns at the right. Combinations using range search are found to consistently outperform those based on kNN, both for individual datasets as well as for the average. In particular, the combination with the best performance for the IRPC dataset is also the best single-scale combination overall. In general, correlation values for this dataset are lower than the remaining

TABLE II: Correlation coefficients for multiscale PointSSIM: SESN (Single-estimator Single-neighborhood) and MESN (Multi-estimator Single-Neighborhood) configurations

	PLCC			SROCC			Average		Scale weights				
Combination	IRPC	ICIP2020	SR-PCD	IRPC	ICIP2020	SR-PCD	PLCC	SROCC	6-bit	7-bit	8-bit	9-bit	10-bit
mAD, $k = 24$	0.861	0.903	0.904	0.791	0.878	0.925	0.890	0.865	-	0.296	0.704	-	-
mAD, $k = 6$	0.585	0.946	0.815	0.663	0.928	0.869	0.782	0.820	0.527	-	-	-	0.473
σ^2, $k = 12$	0.819	0.919	0.921	0.744	0.893	**0.934**	0.886	0.857	-	0.741	-	0.259	-
mAD, $r = 1$	0.921	0.955	0.928	0.852	0.942	0.929	**0.935**	**0.908**	0.789	-	0.053	0.070	0.088
mAD, $r = 1$	0.765	0.967	0.900	0.747	0.961	0.893	0.877	0.867	0.689	-	-	-	0.311
mAD, $r = 1.5$	0.780	0.946	0.956	0.786	0.931	**0.934**	0.894	0.884	0.590	-	-	0.118	0.292
$r = 1$	**0.928**	0.901	0.913	**0.856**	0.886	0.911	0.914	0.884	0.533 σ	0.300 μ	0.047 μAD	0.051 QCD	0.068 mAD
$r = 1$	0.679	**0.971**	0.863	0.622	**0.963**	0.865	0.837	0.817	0.304 σ^2	0.285 μ	-	0.015 σ^2	0.396 mAD
$r = 1.5$	0.795	0.945	**0.958**	0.783	0.928	0.933	0.899	0.881	0.491 σ^2	-	-	0.160 QCD	0.349 mAD
$r = 1$	0.925	0.920	0.918	0.855	0.901	0.911	0.921	0.889	0.459 σ^2	0.349 μ	0.055 μAD	0.055 μAD	0.082 mAD

TABLE III: Average correlation across datasets

	MS-PointSSIM	PCQM	GraphSIM	D1 PSNR	D2 PSNR	Y PSNR	YUV PSNR
PLCC	**0.935**	0.926	0.870	0.857	0.870	0.761	0.768
SROCC	0.908	**0.914**	0.832	0.808	0.828	0.771	0.769

two, indicating that there are characteristics from those point clouds that hinder the performance of this objective metric. While the employed codecs are very similar to those used in ICIP2020, the point clouds from IRPC have highly varying point densities when compared with the models from the former, which are uniformly dense. This is likely the reason for the particularly lower performance of the combinations using 10-bit precision on IRPC: for the sparser point cloud models, most local neighborhoods are empty since most points are more distant from each other than the range used to form the neighborhoods. The combination with higher performance for ICIP2020 displays low performance also for SR-PCD, possibly due to its lack of compression artifacts from learning-based codecs.

Table II depicts the correlation coefficients achieved by the MS-PointSSIM metric. The upper and lower halves contain results for the SESN and MESN configurations, respectively. The former is further divided according to the method used to form neighborhoods, with the combination producing higher coefficients for each dataset using both kNN and range searching being displayed on top and on the bottom, respectively. Moreover, the five columns at the right contain the weights w_p assigned to each scale after logistic fitting, where all values lower than 10^{-4} are omitted. It is observed that MS-PointSSIM achieves higher correlation values than the baseline for both neighborhood search methods. Similarly to the previous case, range searching is also found to allow for better performance for all datasets, with the average Pearson correlation going up to 0.935 for the best configuration. The median absolute deviation is found to be particularly effective as an estimator, and the minimum range of 1 produces the best results for two datasets. These results may seem counter-intuitive since local neighborhoods formed with this range contain only points adjacent to the target, being mostly empty for sparser point clouds. This effect is however attenuated at lower scales, where the points of the voxelized models are arranged closer together, allowing the metric to extract meaningful information. This is likely the reason why the combination with the highest correlation for IRPC was optimized with the highest weight for the coarser scale voxelized at 6-bit depth, with the finest scale at 10-bit depth receiving significantly smaller importance.

Since for both the single-scale metric and the SESN configuration range search was found to be the best method for forming neighborhoods, results obtained with kNN were omitted for the MESN configuration. The first three rows correspond to the single configuration with the highest Pearson correlation for each dataset. In order to compute the last row, the 100 best combinations for each dataset were tested on the remaining two, and the combination with the highest average Pearson correlation was retained. Since the set of MESN combinations is a superset of the previous SESN configuration, higher performance is naturally achieved separately for each dataset due to the higher number of options to choose from. However, results suggest that the large parametric space makes the metric overfit to specific datasets, adapting to noise in the data rather than to general features of the human visual system on point clouds. This is the likely reason why the average correlation values are lower than for the SESN configuration, even when evaluating the best 100 combinations for each dataset, as in the last row of the table. Since results for the MEMN configuration followed a similar trend, with higher correlation values for single datasets but not reaching an average correlation as high as SESN, results for that configuration are not displayed in this paper.

Therefore, the SESN combination using the mAD estimator and unitary range is selected, with scale weights obtained from the logistic fitting on the IRPC dataset. This combination assigns the highest weight for the coarsest scale, while the 7-bit scale is completely ignored. These results do not mean that the single-scale score at this precision is the least correlated with human perception, but rather that it does not provide enough additional information to what can be captured by the remaining scales. The average correlation coefficients across the three datasets obtained by this combination are compared to other metrics in Table III, and scatter plots of the metric values against the MOS are presented in Figure 1 for the two best-performing estimators, namely MS-PointSSIM and PCQM. Pearson and Spearman correlation values for individ-

ual datasets are also reported in the legends of the plots from Figure 1. It is observed that the proposed metric outperforms the remaining ones in terms of average Pearson correlation. If PCQM maintains good performance for all datasets, it is still not able to reach a correlation higher than 0.9 for IRPC. MS-PointSSIM is the only evaluated metric able to achieve such correlation values, obtaining a 0.935 average Pearson correlation against 0.926 for PCQM and 0.870 for GraphSIM. While it could be argued that the leading performance of MS-PointSSIM on the IRPC dataset is due to the fact that this dataset was directly used for the optimization of its scale weights, the same holds for PCQM. Therefore, such results corroborate the conclusion that objective evaluations at multiple scales allow the metric to better model the human visual system.

VII. Conclusion

In this paper, an objective quality metric computing the structural similarity of the luminance attribute at multiple scales is proposed and evaluated. The PointSSIM metric is used as the baseline, and modifications to the method used to form local neighborhoods are presented. Multiscale scores are obtained through a weighted sum of single-scale metric values computed at five scales, with weights being the result of logistic fitting to subjective scores extracted from three different sources. Range searching allows for better performance than the previously proposed k-Nearest Neighbors method for the single-scale metric. The multiscale PointSSIM is found to achieve a higher correlation with human perception, outperforming the baseline even when the same statistical estimator and range value are fixed across all scales. Further experiments showed that allowing for multiple estimators to be combined at different scales can achieve an even higher correlation with subjective scores for single datasets. However, such combinations achieve lower average correlation across datasets, indicating that the larger parametric space allows for the metric to overfit and lose its generalization power. Comparison to state-of-the-art metrics reveals that the proposed metric achieves the highest average Pearson correlation across the three evaluated subjective datasets. Future works may focus on evaluation in more subjective datasets in order to select the combination and scale weights that obtain the best performance across a wider range of distortions, as well as comparison to a wider range of metrics including learning-based quality estimators.

References

[1] 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)*. IEEE, 2017, pp. 3460–3464.

[2] 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.

[3] 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.

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

[5] 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)*. IEEE, 2020, pp. 1–6.

[6] 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*, vol. 44, no. 6, pp. 3015–3029, 2020.

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

[8] 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.

[9] 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. Ieee, 2003, pp. 1398–1402.

[10] M. Testolina, E. Upenik, J. Ascenso, F. Pereira, and T. Ebrahimi, "Performance evaluation of objective image quality metrics on conventional and learning-based compression artifacts," in *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2021, pp. 109–114.

[11] 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.

[12] 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.

[13] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Improving psnr-based quality metrics performance for point cloud geometry," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 3438–3442.

[14] M. Quach, A. Chetouani, G. Valenzise, and F. Dufaux, "A deep perceptual metric for 3d point clouds," *arXiv preprint arXiv:2102.12839*, 2021.

[15] 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.

[16] A. Chetouani, M. Quach, G. Valenzise, and F. Dufaux, "Convolutional neural network for 3d point cloud quality assessment with reference," in *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2021, pp. 1–6.

[17] F. Mentzer, G. D. Toderici, M. Tschannen, and E. Agustsson, "High-fidelity generative image compression," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11 913–11 924, 2020.

[18] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[19] J. Wang and Z. Ma, "Sparse tensor-based point cloud attribute compression," in *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2022, pp. 59–64.

[20] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 1–4.

[21] 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.

[22] 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, pp. 1–6.

[23] N. Frank, D. Lazzarotto, and T. Ebrahimi, "Latent space slicing for enhanced entropy modeling in learning-based point cloud geometry compression," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 4878–4882.