EE-550 | **Laboratory 4**
Point Cloud subjective quality assessment

Gloria Dal Santo

November 11^{th}, 2020

1 Subjective quality assessment

In this laboratory session we have compared the pcc_geo_color and CWI-PCL codecs for point clouds using both subjective and objective quality metrics. This first section will discuss about the subjective assessment, which uses the Absolute Category Rating with Hidden Reference (ACR-HR) scores, gathered during the lab session, to evaluate the quality of the codecs.

1.1 Outliers detection

The detection and removal of the outliers was performed according to the algorithm described in ITU-T Recommendation P.913 which employs the correlation coefficients between the scores of one subject and the Mean Opinion Score (MOS) over all the subjects. To compute the linear Pearson correlation coefficient (PLCC) the `'corr'` function was used, with `('Type','Pearson')` as extra arguments. For the MOS, the `'mean'` function was used. After computing the coefficients $r_1(x,y)$, relative to the individual stimuli, and $r_2(x,y)$, relative to all the stimuli in a content, and applying the rejection rule that discards the scores of one subject if $r_1(x,y) < 0.75$ and $r_2(x,y) < 0.8$ are verified, it has been obtained that there are no outliers in the recorded ratings: all the assigned scores are consistent with themselves and there are no drastic deviations from the mean.

1.2 Processing of data

The processing of the subjective data was performed by computing the Differential Mean Opinion Score (DMOS) and the Confidence Interval (CI). To compute the DMOS, the mean of the Differential Viewer scores (DVs) was evaluated according to the definition of the latter: $DV = V - V(\text{ref}) + 5$, where V is the ACR score given by a subject to an encoded model and $V(\text{ref})$ is the ACR given to the reference model. It can be noticed that the individual DV scores reach the maximum value when $V \geq V(\text{ref})$, that is when a subject does not perceive any negative difference between the distorted and the original model, indicating that the applied codec reproduced the original model with high accuracy and yielded a good quality result. The CIs were computed using the Student's t-distribution as $\text{CI} = t(1-\alpha/2, N-1) \cdot \sigma \cdot N^{-1/2}$, where N is the number of subjects, σ is the standard deviation of the scores assigned to each stimulus, computed with `'std'` function, and parameter $\alpha = 0.05$. To compute the t-distribution, `'icdf'` function with parameter `'T'` was used.

The obtained DMOS ad CIs are depicted in Fig.(1). It can be observed that throughout the different contents and codecs, the DMOS is always increasing with the bitrate, meaning that the quality of a distorted model perceived by the subjects increases with the bitrate. The increase in DMOS is initially steep while it becomes almost flat at higher bitrates. For the `'rhetorician'` model, encoded with pcc_geo_color, this trend is violated when the bitrate increases from 1.59 to 3.85 bpp. Indeed, for this specific case the DMOS decreases by 0.045, which can be interpreted as a an overall similarity between the quality perceived by the models encoded with 1.59 and 3.85 bpp, with the former slightly more favorable than the latter. For what concerns the CIs, a global trend is difficult to be detected, however it can be seen that the CIs span wider ranges in the case of `'longdress'` and `'phil'` models since the assigned scores differ the most along the different subjects.

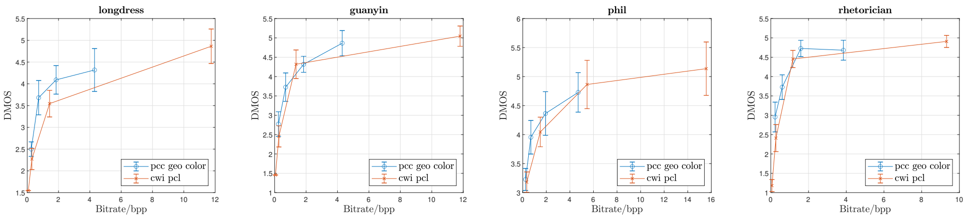

Figure 1: DMOS with confidence intervals

In terms of perceived quality, the CWI-PLC codec performed better as it achieved higher DMOS values, always around 5 for the higher bitrates. However, for bitrates in the range [1, 4] bpp the curves linking the DMOS values of CWI-PLC codec are almost always below the ones of pcc_geo_color, indicating that for a given bitrate it performed worse than the other codec.

2 Objective quality assessment

This section will discuss about the different point-based and color-based objective metrics that were used for the quality assessment.

2.1 Point-to-point

The point-to-point metric is a objective quality metric for geometric-only distortions and it depends on the Euclidean distance between each point of the distorted model and its nearest neighbor from the reference model. In order to find the point in a reference model that is the nearest to a specific point in a model under analysis, the `'knnsearch'` function was used. For computing the Euclidean distance it was directly applied the formula $||\bar{v}_a^b||_2 = \sqrt{(b_x - a_x)^2 + (b_y - a_y)^2 + (b_z - a_z)^2}$. To evaluate the symmetric point-to-point metric for each model, the implemented function sets both models as reference and estimates the Mean Squared Error (MSE) and Hausdorff distances for both cases. Since the point-to-point is a geometric distance, the function keeps only the higher error values.

The values of the point-to-point objective metric vs the bitrate are illustrated in Fig.(2). It can be seen that the trend is always decreasing, meaning that as the bitrate increases the geometric displacement of the points in the distorted model from the reference position reduces. As it was observed in the subjective quality assessment, the CWI-PLC codec for the highest bitrates allows to achieve higher perceived quality. The MSE and Hausdorff distances end up giving similar trends, with the latter resulting in values that are around five orders of magnitude lower. However, when Hausdorff distance is evaluated the average difference between the scores of the two codecs is much higher and it shows also that the CWI-PLC codec performed always better.

2.2 Plane-to-plane

The plane-to-plane metric is a different objective quality metric for geometric-only distortions that depends on the angular similarity of planes tangent to associated points in the distorted and reference models. To compute the scores with this metric, the normals of the models are needed. To this purpose the function `'pcnormals'` was used, with `'K'` parameter set to 128 nearest neighbors. The symmetric plane-to-plane metric with MSE was computed by the given function `'angularSimilarity'` whose results are reported in Fig.(3).

From the figure, it can be observed that the global trend is increasing, owing to the fact that the plane-to-plane is a similarity metric. Consistently with the previous results, for higher bpp the scores increase indicating that the distorted model is much more similar to the reference one when higher bitrates values are employed. In general, CWI-PLC codec performed better only for stimuli with the highest bpp value in each content.

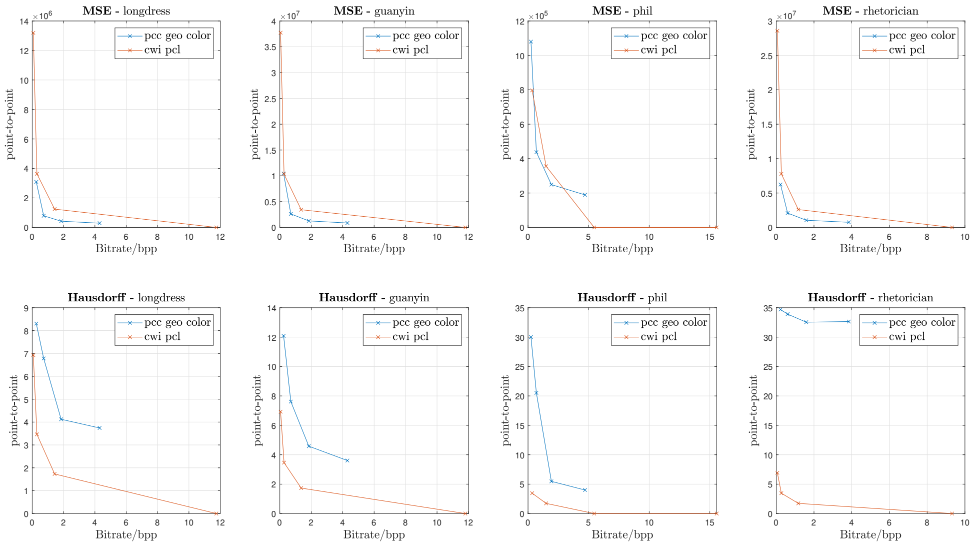

Figure 2: Point-to-point vs bitrate

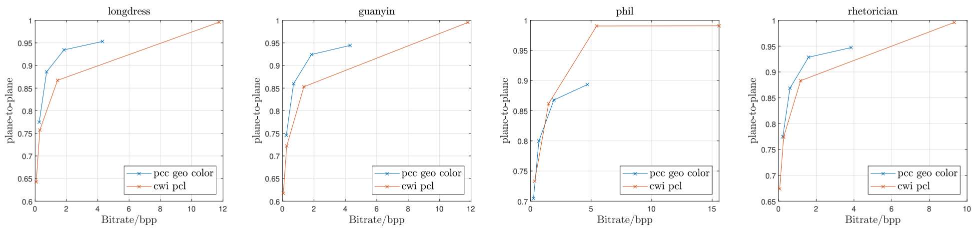

Figure 3: Plane-to-plane vs bitrate

2.3 PSNR

The Peak Signal-to-Noise Ratio (PSNR) is a color-only metric applied between each point of the distorted model and its nearest neighbor from the reference model, which can be found as in section 2.1 were function `'knnsearch'` was used. Three different color-metrics were used:

- along all the three RGB channels: $\text{PSNR}_{RGB} = 10\log_{10}\left(255^2/\text{MSE}\right)$
- by averaging the PSNRs from the three channels: $\text{PSNR}_{RGB} = \left(\text{PSNR}_R + \text{PSNR}_G + \text{PSNR}_B\right)/3$
- for the YUV colorspace: $\text{PSNR}_{YUV} = \left(6 \cdot \text{PSNR}_Y + \text{PSNR}_U + \text{PSNR}_V\right)/8$

To compute the PSNRs, function `'psnr'` was used. Regarding PSNR_{YUV}, to convert the colorspace the given `'rgb2yuv'` function was emplyed.

The scores obtained from the symmetric color-metrics are reported in Fig.(4), where it is clear that for higher bitrates the similarity score increases. The different metrics reported similar results for the same model and no major differences are present across different models for the same applied metric. Also in this case it can be concluded that the CWI-PLC outperformed the pcc_geo_color codec at higher bitrates.

2.4 Comments

The scores of all the tested objective metrics improved with increasing bitrate and they exhibit a steep decrease (or increase for similarity metrics) followed by a rather constant behaviour. The results obtained for a given content with a particular metric is in general in agreement with the scores achieved by the other metrics for the same content and the trends are coherent also along the different contents. CWI-PLC codec outperformed pcc_geo_color only for higher bitrates, according to all the metrics.

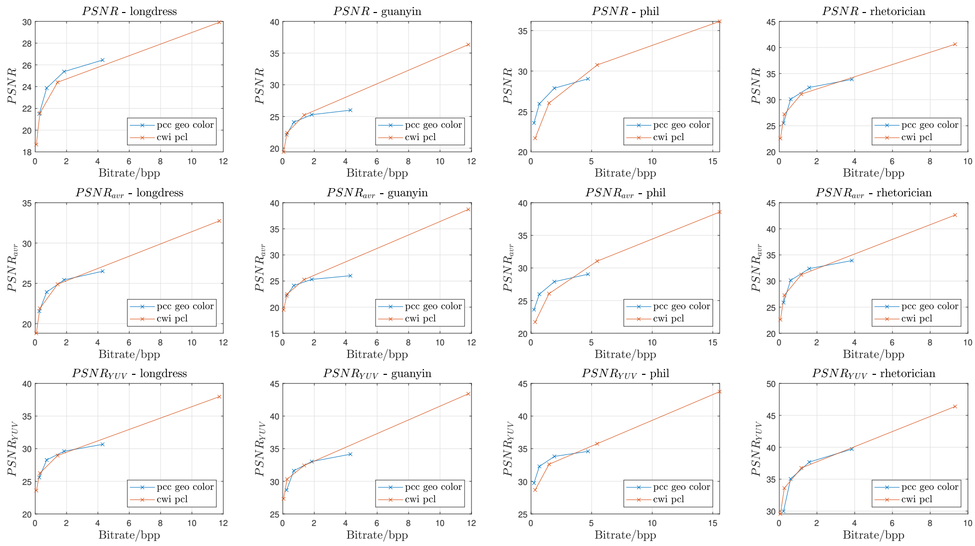

Figure 4: PSNR vs bitrate

3 Benchmarking of objective quality metrics

In this last section the objective scores are compared with the subjective ones, considered as ground truth, in order to assess the correlation between them. To this purpose, Pearson, Spearman correlation coefficients and the Root-Mean-Square-Error (RMSE) have been used before and after linear and cubic fitting of the objective scores on the subjective ones. The fitting operation is needed to associate the ranges of values taken by the subjective and objective scores. The Pearson and Spearman correlation coefficients of each objective metric with the MOS of the subjective scores was computed using `'corr'` function with `('Type','Pearson')` and `('Type','Spearman')` arguments. The RMSE was directly computed with the formula RMSE $= \sqrt{(R_i - \mathrm{MOS})^2/m}$, where R_i is the vector of objective metric scores and $m = 32$, the number of distorted models. For the fitting opration the function `'polyfit'` was used.

Pearson correlation coefficient examines the strength and direction of the linear relationship between the MOS values and the objective scores while the Spearman correlation is a measure of the strength of a monotonic relationship between them. They can take values between 1, strongest positive correlation, and -1, strongest negative correlation. On the other hand, the RMSE is a measure of accuracy of an objective metric to predict the subjective ratings with minimum average error. A high accuracy is given by a low value of RMSE, meaning that the standard deviation between the objective metric and MOS is lower.
Comparing the results reported in Table (1), (2) and (3), the objective metric that achieved the highest performance in predicting the objective scores is the plane-to-plane MSE metric, whose correlation coefficients and standard deviation take more favorable results when cubic fitting is applied. This can also visually seen in Figures (5), (6) and (7) where the plots relative to plane-to-plane metric show an arrangement of the points around the diagonal.
Focusing on the result in Table (3), it can be noticed that the the color metrics give similar results, with the averaged color-PSNR performing slightly better. Point-to-point achieved good results when MSE was applied. Indeed, the results obtained with point-to-point metric with Hausdorff distance are the less

favorable among all the other indexes. This behavior can be seen also in Figures (5), (6) and (7) where the subplot relative to the Hausdorff distance show a more dispersed and nonlinear behaviour. Due to this less favorable results, it can be concluded that on average the color metrics performed better.

One limitation of the objective metrics in fitting the subjective results could be that they do not take into account the environmental conditions or the test methodologies used during the subjective assessment hence the correlation indices are always lower than the maximum value. Moreover, subjects' opinions on the stimuli have to be taken into account since they might have impact on the subjective results. In addition to this, it must be remarked that each objective metric seen in the laboratory is based on a particular aspect of the model, thus a combination of them might be needed to enhance the performance of the objective assessment.

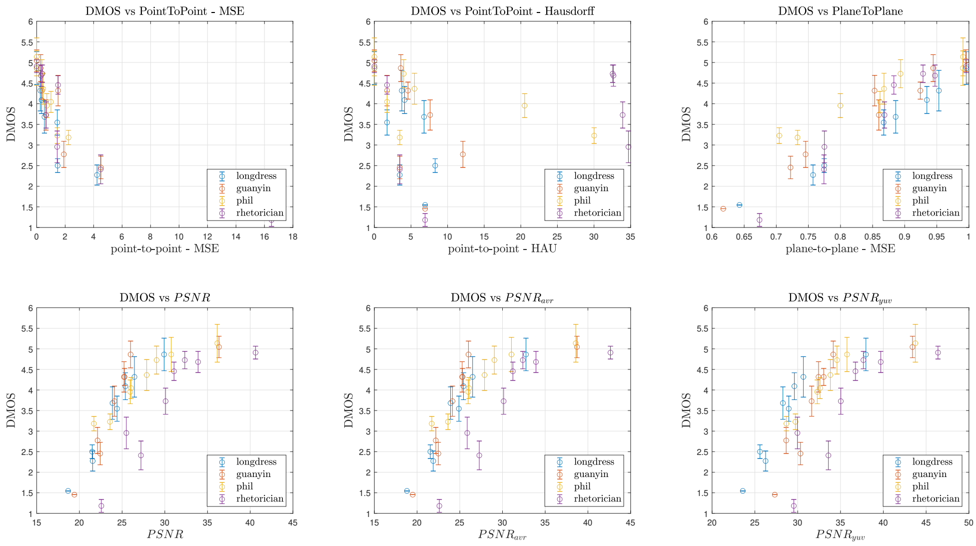

Figure 5: DMOS vs objective metrics

Without fitting			
Metric	Pearson	Spearman	RMSE
point-to-point MSE	-0.6227	-0.7187	5.3772
point-to-point Hausdorff	0.0304	-0.2266	12.7630
plane-to-plane MSE	0.8574	0.8383	2.3950
color-PSNR	0.7452	0.7530	24.6064
color-PSNR$_{avr}$	0.7307	0.7568	25.0906
color-PSNR$_{YUV}$	0.7186	0.7394	30.7464

Table 1: Correlations between objective and subjective scores - without fitting

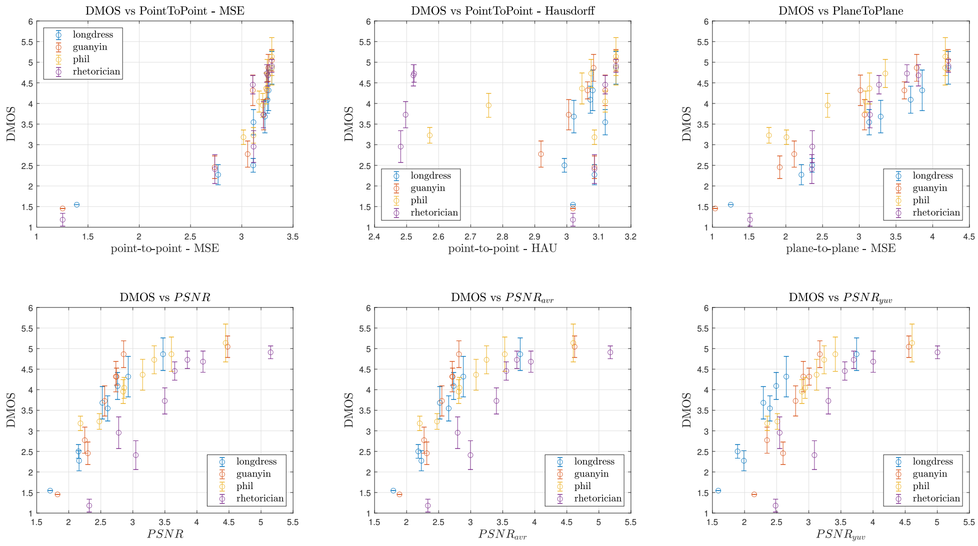

Figure 6: DMOS vs objective metrics - linear fit

Linear fitting			
Metric	Pearson	Spearman	RMSE
point-to-point MSE	0.6227	0.7187	0.9434
point-to-point Hausdorff	0.0304	-0.2266	1.2051
plane-to-plane MSE	0.8574	0.8383	0.6205
color-PSNR	0.7452	0.7530	0.8040
color-PSNR$_{avr}$	0.7307	0.7568	0.8232
color-PSNR$_{YUV}$	0.7186	0.7394	0.8385

Table 2: Correlations between objective and subjective scores - linear fitting

Cubic fitting			
Metric	Pearson	Spearman	RMSE
point-to-point MSE	0.7357	0.7187	0.8166
point-to-point Hausdorff	0.5606	0.5403	0.9984
plane-to-plane MSE	**0.8664**	**0.8383**	**0.6020**
color-PSNR	0.7830	0.7530	0.7500
color-PSNR$_{avr}$	0.7849	0.7568	0.7472
color-PSNR$_{YUV}$	0.7411	0.7334	0.8095

Table 3: Correlations between objective and subjective scores - cubic fitting

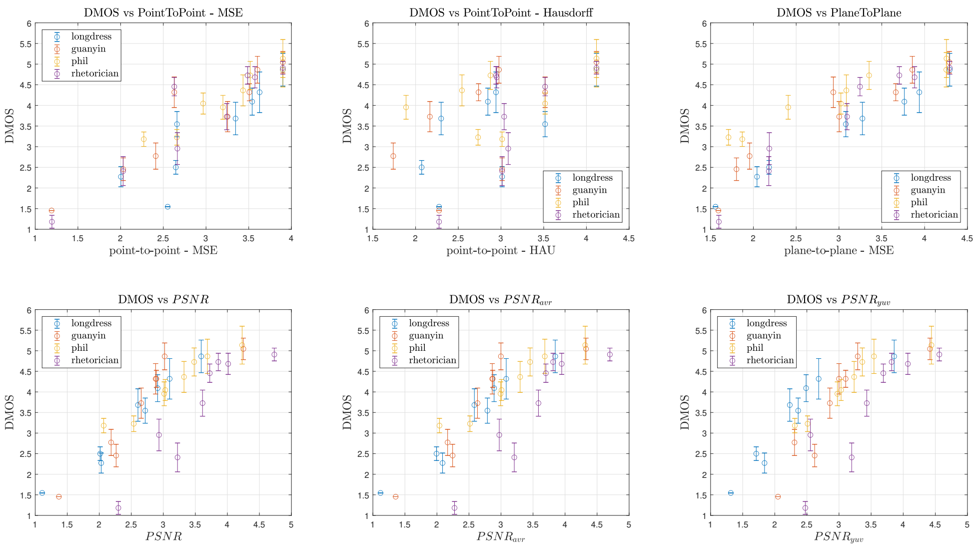

Figure 7: DMOS vs objective metrics - cubic fit