Subjective Quality Evaluation of Point Clouds using Remote Testing

Ashutosh Singla*
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
ashutosh.singla@cwi.nl

Shuang Wang
Audiovisual Technology Group, TU-Ilmenau
Ilmenau, Germany
shuang.wang@tu-ilmenau.de

Steve Göring
Audiovisual Technology Group, TU-Ilmenau
Ilmenau, Germany
steve.goering@tu-ilmenau.de

Rakesh Rao Ramachandra Rao
Audiovisual Technology Group, TU-Ilmenau
Ilmenau, Germany
rakesh-rao.ramachandra-rao@tu-ilmenau.de

Irene Viola
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
irene.viola@cwi.nl

Pablo Cesar
Centrum Wiskunde & Informatica
TU Delft
Amsterdam, The Netherlands
pablo.cesar@cwi.nl

Alexander Raake
Audiovisual Technology Group, TU-Ilmenau
Ilmenau, Germany
alexander.raake@tu-ilmenau.de

ABSTRACT

Subjective quality assessment serves as a method to evaluate the perceptual quality of 3D point clouds. These evaluations can be conducted using lab-based or remote or crowdsourcing tests. The lab-based tests are time-consuming and less cost-effective. As an alternative, remote or crowd tests can be used, offering a time and cost-friendly approach. Remote testing enables larger and more diverse participant pools. However, this raises the question of its applicability due to variability in participants' display devices and environments for the evaluation of the point cloud. In this paper, the focus is on investigating the applicability of remote testing by using the Absolute Category Rating (ACR) test method for assessing the subjective quality of point clouds in different tests. We compare the results of lab and remote tests by replicating lab-based tests. In the first test, we assess the subjective quality of a static point cloud geometry for two different types of geometrical degradations, namely Gaussian noise, and octree-pruning. In the second test, we compare the performance of two different compression methods (G-PCC and V-PCC) to assess the subjective quality of coloured point cloud videos. Based on the results obtained using correlation and Standard deviation of Opinion Scores (SOS) analysis, the remote testing paradigm can be used for evaluating point clouds.

IXR '23, October 29, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0280-8/23/10.
https://doi.org/10.1145/3607546.3616803

CCS CONCEPTS

• **Human-centered computing → User studies**.

KEYWORDS

Remote testing, point clouds, quality assessment, coloured point clouds

ACM Reference Format:
Ashutosh Singla†, Shuang Wang, Steve Göring, Rakesh Rao Ramachandra Rao, Irene Viola, Pablo Cesar, and Alexander Raake. 2023. Subjective Quality Evaluation of Point Clouds using Remote Testing. In *Proceedings of the 2nd International Workshop on Interactive eXtended Reality (IXR '23), October 29, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3607546.3616803

1 INTRODUCTION

In recent years, due to the advancements in acquisition and display devices, the adoption of 3D data to represent immersive applications has increased. The 3D data allow users to have more immersive experiences compared to 2D videos as it allows a user to explore content in 6 DoF (Degrees of Freedom) and in a more natural way [1–3]. To represent 3D content, polygon meshes, and point clouds are the predominant approaches [4–6]. A point cloud consists of points in 3D space. A typical point cloud is defined by its X, Y, and Z coordinates with multiple attributes such as colour, and surface normal [1]. A polygon mesh consists of edges, vertices, and surfaces. Point clouds are preferred over meshes as they are easy to capture, store, and transmit and do not need connectivity information [1, 6, 7].

Most of the experiments published in the literature were conducted in lab-based experiments under controlled conditions (see

† A part of the research presented in this paper was undertaken while Ashutosh Singla was affiliated with the Audiovisual Technology Group at TU Ilmenau.

Sec. 2). However, due to its time-consuming, and less cost-effective nature, conducting lab-based tests is not always feasible. Crowdsourcing or remote testing has become a popular approach for assessing quality online using a relatively large number of participants. Furthermore, these tests represent a time- and cost-friendly method and it is one possible solution for conducting quality tests.

Crowdsourcing has been used for the assessment of the visual quality of colored point cloud as an alternative solution, using the Double Stimulus Impairment Scale (DSIS) test method [8, 9]. In [8], Nehmé et al. compared the results obtained in a crowd-based study with a lab-based study for the case of colored point clouds. Their results demonstrated that the crowdsourcing paradigm is equally accurate when compared to lab-based tests. However, no crowdsourcing study using the ACR test method for quality evaluation of point clouds has been conducted. Furthermore, there is a lack of studies comparing the performance of state-of-the-art encoders (Geometry-based Point Cloud Compression (G-PCC), and Video-based Point Cloud Compression (V-PCC)) between lab-based and crowd-based tests. Hence, it is necessary that the crowdsourcing or remote testing approach should be validated using the ACR test method for the case of assessing the point cloud geometry and for comparing the performance of different (G-PCC and V-PCC) encoders.

Hence, in this paper, we address the following research questions:

(1) Can the remote testing approach be used effectively to assess the visual quality of point cloud geometry?
(2) How does the performance of the G-PCC and V-PCC encoders from remote testing compare to the results obtained using the lab tests?

To address these research objectives, we conducted two remote tests (Test 1 and Test 2) using the ACR test method. In Test 1, we investigate the research question if remote testing can be used to evaluate the visual quality of point cloud geometry. In Test 2, the performance of G-PCC and V-PCC encoders is assessed using a remote testing approach. Furthermore, we compare our results with the results obtained in the lab as reported from other studies, to analyze the applicability of the remote testing paradigm for the case of colorless point clouds (Test 1).

This paper is organized as follows. Section 2 describes the related work in the area of assessing the visual quality of point clouds. The remote testing framework is described in Sec. 3. The results of colorless and colored point clouds are analyzed and compared with the lab-based results in Sec. 4 and Sec. 5, respectively. Concluding remarks are given in Sec. 6.

2 RELATED WORK

There are several works on the visual quality assessment of point clouds reported in the literature, as also listed in Table 1 [1–3, 6–18]. These studies can be classified into two categories, among other criteria. The first type of studies evaluates geometrical degradations [1, 2, 8, 10–12], while the second type focuses on assessing the impact of compression algorithms [3, 6, 7, 13–17]. A more comprehensive overview of subjective studies for point cloud contents can be found in [19].

In [1], Alexiou et al. compared the point cloud geometry using the ACR (typically using a 5-point scale, with the average over subjects referred to as mean opinion score, MOS) and DSIS test methods. They compared two types of geometrical degradations namely, Gaussian noise and Octree-pruning. Their experimental results showed that the DSIS test method provides better results for both of the degradations. Furthermore, Alexiou et al. [18] compared the performance of different objective metrics available in the state-of-the-art for different types of geometrical degradations. Their experimental results showed that objective metrics perform well in the case of Gaussian noise, but not for the compression-like distortions.

Besides, comparing different test methods, the viewing environment plays a crucial role in the assessment of the visual quality of point clouds. In [10], Alexiou et al. used an AR-HMD (Augmented Reality Head Mounted Display) to assess the visual quality of point clouds for two geometrical degradations. Their analysis showed that objective metrics correlate well with the subjective scores in the case of Gaussian noise, in agreement with the results obtained in [18]. For the compression artifacts, the performance of objective metrics depends on the content. The authors concluded that there is a need for better objective metrics. The impact of different display environments (Desktop vs. AR-HMD) was investigated in [2] using the DSIS test method. Experimental results showed that in the case of Gaussian noise, the scores obtained from different setups were found to be statistically equivalent. For Octree-pruning, scores obtained from both environments were statistically distinguishable.

The impact of several denoising algorithms on point clouds was compared by Javaheri et al. [11] in their subjective test. Their results showed that a graph-based denoising algorithm can improve the quality of point cloud data. Further, they found that p2plane metrics have a better correlation with MOS.

Many other factors such as coding artifacts, rendering solutions, different viewing conditions, and packet loss were investigated in [3, 6, 7, 13–16]. The influence of different rendering and coding solutions on the perceived visual quality of point clouds was studied by [3] using the DSIS test method. Their main conclusion was that different types of coding artifacts were not equally perceivable for all three rendering approaches. Zerman et al. [7] compared the two different representations of 3D data (Meshes and point cloud) for different state-of-the-art encoders using the ACR test method. They found that textured mesh provides the best visualization quality, especially at higher bitrates, and the point cloud performed better at lower bitrates. Cao et al. [6] explored the effect of observation distance and bitrate on the perceived quality for point cloud and mesh-based compression. Their experimental results were also in-line with the ones in [7] especially at lower bitrates, where it was found that point cloud is better at low bitrates, whereas a mesh representation is preferred when observation distance is close and bitrate is high.

The different viewing conditions on the perceived quality of dynamic point clouds were investigated in [13, 16]. Subramanyam et al. [13] compared the quality in 3 DoF and 6 DoF of dynamic point clouds in virtual reality HMDs using the ACR-HR test method. Their results showed that the V-PCC codec has better performance than the MPEG anchor, especially at lower bitrates. Further, they showed that there is a need for a new source point cloud dataset, as the sequences they used in the test had a significant impact on the MOS. Viola et al. [16] extended the work and compared the quality

Table 1: List of Studies on the Subjective Evaluation of Point Clouds.

S.No	Authors	Static or Dynamic	Geometry or Compression	Test Method	Test Paradigm	Degradations	Display Device
1	Alexiou et al. [1]	Static	Geometry	DSIS & ACR	Lab	Octree-prunning & Gaussian noise	Apple Cinema Display (2D screen)
2	Alexiou et al. [2]	Static	Geometry	DSIS	Lab	Octree-prunning & Gaussian noise	2D monitor & AR-HMD
3	Alexiou et al. [10]	Static	Geometry	DSIS	Lab	Octree-prunning & Gaussian noise	Occipital Bridge AR headset
4	Javaheri et al. [11]	Dynamic	Geometry	DSIS	Lab	Gaussian Noise	2D screen
5	Nehmé et al. [8]	Dynamic	Geometry	DSIS	**Remote**	Geometry and Color distortions	2D screen
6	Seufert et al. [12]	Static	Geometry	ACR	**Remote**	Point reduction techniques	2D screen
7	Javaheri et al. [3]	Static	Compression	DSIS	Lab	3 rendering & codec solutions	ASUS VH238 monitor
8	Cao et al. [6]	Dynamic	Compression	ACR & Pair Comparison	Lab	Observation distance & bitrates	LED Monitor
9	Zerman et al. [7]	Dynamic	Compression	ACR	Lab	Coding artifacts	LCD Display
10	Subramanyam et al. [13]	Dynamic	Compression	ACR-HR	Lab	Coding artifacts	Oculus Guardian System & Oculus Rift
11	Hooft et al. [14]	Dynamic	Compression	ACR	Lab	Bandwidth & Coding artifacts	2D screen
12	Dumic et al. [15]	Dynamic	Compression	DSIS	Lab	Packet-loss, Compression & Corrupted bitstream types	2D Screen
13	Viola et al. [16]	Dynamic	Compression	ACR-HR	Lab	Coding artifacts	2D screen and Oculus Rift
14	Perry et al. [9]	Dynamic	Compression	DSIS	**Remote**	Coding artifacts	2D screen
15	Weil et al. [17]	Dynamic	Compression	ACR	**Remote**	Coding artifacts, Framerate, & Viewing Distance	2D screen

of dynamic point clouds on a 2D screen and in VR with 3DoF and 6DoF using the ACR-HR test method. Their results showed that viewing paradigms have a significant impact on the MOS, but not between 3DoF and 6DoF.

Different network-related parameters such as packet loss ratio and bandwidth were investigated in [14, 15]. Dumic et al. [15] investigated the impact of packet loss ratio with different levels of compression using the DSIS test method in different labs. Their results showed that the subjects provided the lowest MOS at a higher packet loss ratio irrespective of the compression level. Moreover, a strong correlation between the results of the two laboratories was shown. Van der Hooft et al. [14] investigated the impact of network bandwidth, viewport prediction, and bitrate allocation on the subjective quality of volumetric video streaming using the ACR test method. Their results indicated that higher bandwidth results in a higher MOS, which is expected. Their bitrate allocation scheme also has an impact on the perceived quality. Further, they pointed out that in order to match human perception precisely, advanced metrics and QoE models are needed.

There are only a few studies in the literature that used a crowdsourcing approach for the assessment of quality in the case of point clouds [8, 9, 12, 17, 20]. Nehmé et al. [8] conducted a crowdsourcing test using a DSIS test method and compared the results with the lab-based test. Their results showed that the crowdsourcing paradigm is as accurate as a lab-based test. However, the lab-based test was conducted using a HTC Vive Pro HMD. The crowdsourcing results should be compared with the lab-based test conducted on a 2D screen to nullify the effects of an HMD [8]. Perry et al. [9] compared the online subjective testing using two different approaches. In the first approach, participants were instructed to download the whole data and execute the experiment by running some MATLAB scripts. In the second approach, participants accessed the server using a web browser and performed the experiment. The authors addressed the first approach as the Direct Download Option and the second approach as Web Browser Option which is remote testing. Their experimental results showed a preference for the direct download option over the web browser option. However, there is a high correlation between direct downloads and web browser options. Their results lack in comparison with the lab-based tests.

Seufert et al. [12] investigated the impact of two different point reduction techniques on the QoE from crowdsourcing users and experts using the ACR test method. Their results showed that experts found a bigger difference between the reduced and original point clouds. However, they did not study whether the crowdsourcing test is comparable to a lab test by using naive (non-expert) subjects. Herfort et al. [20] used crowdsourcing in different experiments in the classification of point clouds and found that the accuracy of the results was affected by the characteristics of the point cloud data.

Weil et al. [17] conducted a crowdsourcing study using the ACR test method to study the impact of viewing distances, framerates, and two compression methods (Draco and V-PCC). Their main observation was at a higher viewing distance, the degradations were less noticeable. Furthermore, they developed QoE models for predicting the perceived quality, and based on the performance, their model accurately predicted the MOS. The limitation of their study is the absence of a comparison between crowdsourcing and lab-based results.

From the studies reviewed above [1, 2, 6–14, 18], it can be noted that the impact of different test methods, rendering solutions, degradation types, display devices and compression algorithms were studied using lab-based tests. However, the general applicability of the crowdsourcing or remote testing paradigm on the visual quality of colorless point clouds has not been extensively investigated. Hence, in this paper, the remote testing paradigm will be applied to the assessment of the visual quality of point cloud geometry and the obtained results will be compared with lab-based studies. Furthermore, the performance of different (G-PCC and V-PCC) encoders will be analyzed using remote testing and comparing results with the lab-based test.

3 REMOTE TESTING FRAMEWORK

To conduct the remote tests, a client-server architecture website was developed and integrated using the publicly available tool AVRateVoyager [21] for collecting the ratings. The recommended hosting environment includes Apache web server version 2.4, MySQL version 5.7, and PHP version 7.4.13.

For Test 1, a web-based point cloud renderer was developed by using the Three.js 3D library[1] that allows visualizing the point clouds. The 3D geometry is centered and rotated around the local Y-axis, while the camera focuses on the center of the geometry.

For Test 2, the colored point cloud videos were played using the HTML5 $< video >$ element. For both tests, the background colour of the scene was set to black as recommended by [22]. A passive approach, i.e., with subjects unable to interact with the point clouds in both tests 1 and 2, was selected to minimize the between-subject variation. We preloaded the videos into the browser cache to avoid the effects of stalling as done in [9].

A web browser (e.g., Firefox or Chrome) was needed to start the test at a user's end. The website integrated point cloud visualization, and a rating function. The website is a cross-platform and cross-operating system and can be easily used for remote testing without requiring any special installations. The ratings were stored online in a SQLite 3 database and can be exported in a CSV file.

4 TEST 1

The goal of this test was to evaluate the subjective quality assessment of point cloud geometry using remote testing and compare these findings with lab-based results [1].

This section provides the details of the selected dataset, and the test method used. Subsequently, we present the results obtained from the remote testing and compare them with the lab-based results.

4.1 Selection and Preparation of Sequences

For Test 1, five different static point clouds (1: Bunny, 2: Cube, 3: Dragon, 4: Sphere, and 5: Vase) without colour attributes were selected from [1]. To allow the subjects to view a point cloud from multiple views, we rotated the camera path along the vertical (Y) axis that enables the viewer to visualize the whole point cloud[2] for 10 s. By having a passive mode of inspection, it was ensured that each point cloud is viewed in the exact same way by all subjects. This may result in minimizing variations that may arise due to interactivity [19].

In this test, our objective was to assess the quality for geometrical errors. Therefore, two types of geometric distortions were introduced. Gaussian noise is used to model position errors. The coordinates of each point are affected by an error with a target standard deviation from the set $\sigma \in$ (0.0005, 0.002, 0.008, 0.016). The other type of degradation is based on Octree and is used to represent point clouds in an efficient way and in addition yielding high error [1, 2], which is desirable when testing for the adequacy of crowd-based or remote testing. The compression ratio of a point cloud was adjusted by changing the LoD (Level of Details) value. LoD values are selected for each content in order to obtain the target percentage (p) to the original points with an acceptable deviation of ± 2% (p = 30%, 50%, 70% and 90%) [1]. We used the exact Processed Video Sequences (PVSes) used in [1].

4.2 Test Method

The ACR [23] test method is used to rate the colorless point cloud on a five-point scale (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent). The duration of each PVS is 10 s. Each subject had to rate 40 point clouds (5 Contents and 8 degradations). A total of 90 participants took part in the test. The participants were recruited from the university via email reflectors. Out of the 90, only 71 participants completed the test[3]. The duration of the test was ≈ 15 minutes.

To yield stable results, test participants were recommended to use a desktop, laptop, or computer with a minimum resolution of 1280×720. Only a mouse or trackpad is needed to rate the videos. Prior to the experiment, instructions were provided to the participants explaining the aim of the experiment and their tasks. To familiarize with the test procedure, five training samples were used.

4.3 Results and Discussions

For outlier detection, the procedure according to ITU guidelines specified in ITU-R BT.500-13 was performed [24]. No subject was found to be an outlier in this test.

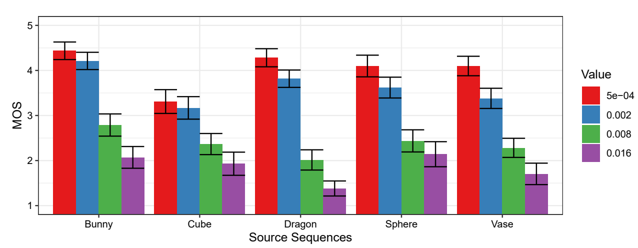

Figure 1: MOS for Gaussian Noise for different SRCs.

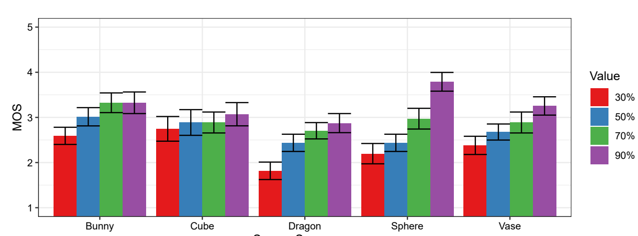

Figure 2: MOS for Octree-Pruning for different SRCs.

[1] https://github.com/mrdoob/three.js/

[2] The size of point cloud was adapted based on the screen size.

[3] The subjects were not paid for this test.

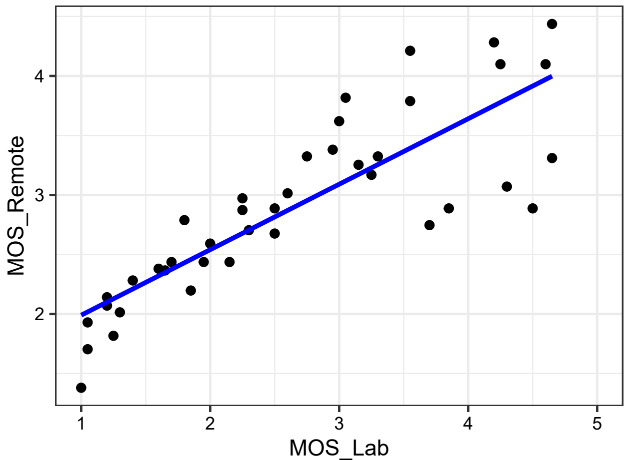

Figure 3: Correlation plot between MOS ratings from Remote vs Lab for test 1.

Figures 1 and 2 show the Mean Opinion Scores (MOS) with associated 95% confidence intervals (CIs) for different source point clouds for different levels of Gaussian Noise and Octree-pruning settings. As expected, as the level of impairment increases, the MOS decreases, which can be observed irrespective of the Source Sequences (SRCs) and degradation type. This trend can be seen for the lab-based experiment as well [1]. One of our main objectives was to investigate if the relative quality-impact of position errors and density of the point clouds as a result of the Octree-pruning are similar when assessed in remote testing as they are when assessed in the lab. Therefore, the noisy and octree-based compressed contents were assessed in a single test session. The experimental results showed that noisy content is perceived as more annoying to the participants than the density of the point clouds. We observed that the users are able to perceive differences in position errors more easily than in Octree-pruning. Those results were also achieved in the lab-based test in [1] for the ACR test method. We compared our results with the one obtained by Alexiou et al. in [1] for the ACR test method and observed that their results indicated a similar trend for Gaussian noise at higher standard deviation, where users rated the sequences more critically than in Octree-pruning at lower point cloud density. Fig. 3 shows the scatter plot of the MOS ratings from remote and lab tests. The Pearson correlation coefficient (PCC) between these two tests is high ($r = 0.840$).

In Figs. 4 and 5, the standard deviation is plotted over the respective MOS rating as described in [25]. The SOS parameter for the lab and remote test is $a_{Lab} = 0.25$ and $a_{Remote} = 0.266$, respectively. The a value is in the same order of magnitude, indicating that these two test results have similar accuracy. The $a_{\{Lab,Remote\}}$ values are similar to video streaming and cloud gaming tests which are reported in [25].

5 TEST 2

Following the colorless point cloud study (Test 1), a second use case focuses on the comparison of different encoders (G-PCC and V-PCC) for the case of colored point cloud videos in a remote testing setup. This section describes the details of the selected sequences and the test method used. Additionally, we present the results obtained from our own test and compare them with the lab-based results [7].

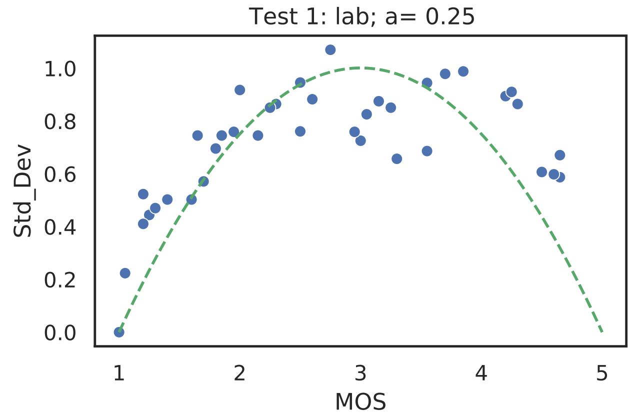

Figure 4: Standard deviation vs. MOS plot, where the line represents a quadratic fitted curve derived from the SOS hypothesis [25] for test 1 (Lab).

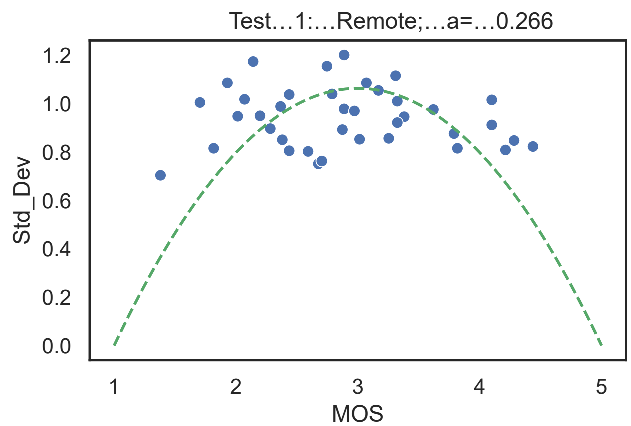

Figure 5: Standard deviation vs. MOS plot, where the line represents a quadratic fitted curve derived from the SOS hypothesis [25] for test 1 (Remote).

5.1 Selection and Preparation of Sequences

In Test 2, we selected four dynamic clouds (1: AxeGuy, 2: LubnaFriends, 3: Matis, and 4: Rafa2) from [7]. The duration of each sequence is 10 s with 30 frames per second. The original lab-based test conducted by Zerman et al. [7] consisted of a total of 164 PVSs. These included both colored and textured meshes. In this test, our aim is to compare the performance of the two state-of-the-art encoders in the context of remote testing. Hence, we used the exact PVSes provided by [7] for G-PCC Region-Adaptive Hierarchical Transform (RAHT) and V-PCC Random Access (RA).

5.2 Test Method

The ACR [23] test method is used to rate the point cloud sequences on a five-point scale (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent). Each PVS is shown to the subjects for 10 s. In this test, each subject had to rate 40 PVS (4 Contents, 2 codecs and 5 bitrates). A total of 82 participants took part in the test. These participants were recruited from the university via email reflectors. Out of 82, only 62 participants completed the test[4]. The duration of the test was under 15 minutes.

Similar to Test 1, test participants were asked to use a desktop, laptop, or computer with a minimum resolution of 1280×720. To rate the videos, a mouse or a trackpad is needed. Before, starting the experiment, participants were explained the aim of the experiment and their task. Five training samples were used to make them familiarize with the test procedure.

5.3 Results and Discussions

To check the reliability of subjects, outlier detection was performed based on [24]. In this test, no outlier was found.

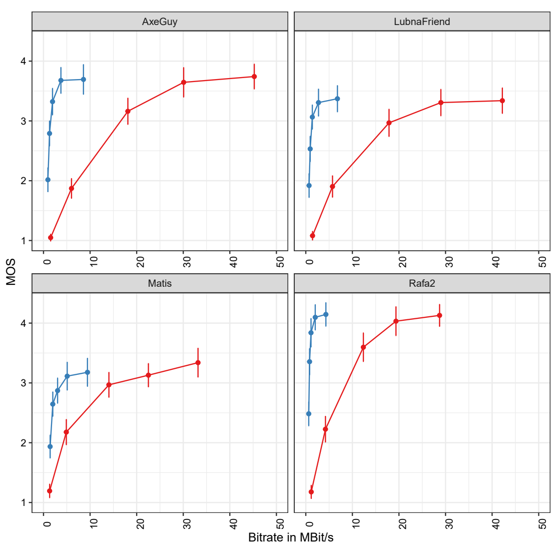

Figure 6: MOS vs bitrates for G-PCC and V-PCC encoders for different SRCs.

In Test 2, we compare the performance of the state-of-the-art encoders: G-PCC (RAHT) and V-PCC (RA) using colored point cloud sequences used in [7]. Figure 6 shows the MOS with associated 95% CIs for different source point clouds for G-PCC (RAHT) and V-PCC (RA). It can be clearly seen from the results that V-PCC (RA) is better than G-PCC (RAHT), irrespective of the sources. This observation is also confirmed by the lab-based tests in [7]. The results obtained in Test 2 and in [7] are understandable, as G-PCC does not consider any temporal redundancy and is targeted to compress static point clouds, whereas V-PCC was developed to compress dynamic point clouds. Furthermore, we compute the PCC between the MOS obtained in the remote test and lab-based test, see

[4]The subjects were not paid for this test.

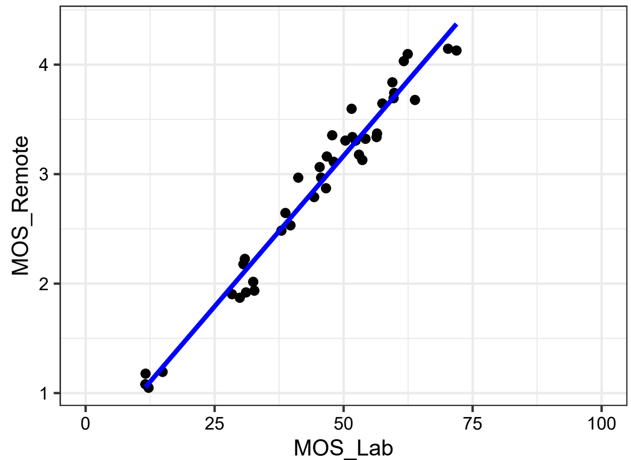

Figure 7: Correlation plot between MOS ratings from Remote vs Lab for test 2.

Fig. 7. Our MOS is highly correlated (0.982) with the lab-based MOS. This high correlation is similar to the reported cross-lab correlation of 2D video tests [26]. From Fig. 7, it can be observed that the range of ratings in the lab test is restricted as compared to the crowd or online test where the participant seems to use the total range of ratings. The smaller range of ratings in the lab test can be attributed to the fact that in the lab test, the subjects were shown both textured meshes and colored point clouds together and hence the subjects may have been more critical while rating the point clouds. Whereas in the remote test, the subjects rated only colored point clouds.

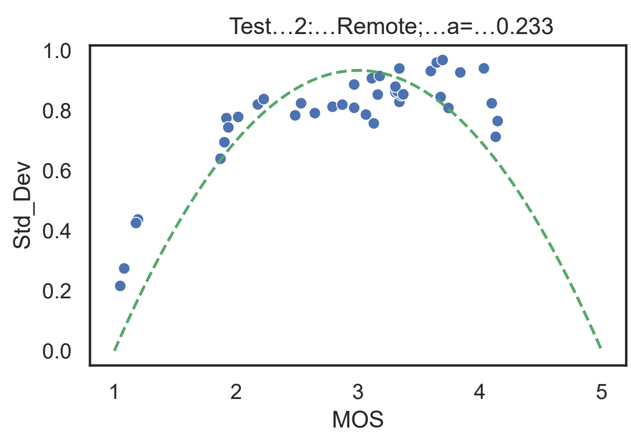

Figure 8: Standard deviation vs. MOS plot, where the line represents a quadratic fitted curve derived from the SOS hypothesis [25] for test 2 (Remote).

In Fig. 8, the standard deviation is plotted over the respective MOS rating as described in [25]. The SOS parameter for the remote test is $a_{Remote} = 0.233$. The a value is in the same order of magnitude as obtained in Test 1 (lab and remote), indicating similar validity and reliability. The SOS plot could not be plotted for the lab-based test, as the ACR (0-100) point scale is used. The magnitude of a value would be different and would not be appropriate to compare it with a_{Remote}.

5.4 Reliability Test

As an additional reliability test to find out if crowd-based tests are equally reliable as compared to the lab-based for Tests 1 and 2, the statistical reliability was evaluated based on [27]. For both these tests, $\mathrm{MCI_{norm}}$ was calculated as shown in Eq. 1.

$$\mathrm{MCI_{norm}} = \frac{\mathrm{MCI}}{\mathrm{MOS\ Range}} \tag{1}$$

Here, MCI is the *Mean Confidence Interval*. *MOS Range* is the absolute difference between the highest and lowest MOS for each test. Table 2 shows the MOS, MOS Range and $\mathrm{MCI_{norm}}$ for Test 1 and 2. Results indicate that all these tests are equally reliable.

Table 2: MCI, MOS Range and $\mathrm{MCI_{norm}}$ for Lab and Remote tests.

	Test 1		Test 2	
	Lab	Remote	Lab	Remote
MCI	0.329	0.222	4.21	0.199
MOS Range	3.65	3.056	60.11	3.096
$\mathrm{MCI_{norm}}$	0.090	0.072	0.070	0.064

6 CONCLUSIONS

In this paper, we conducted two different subjective tests using remote testing and the ACR test method. In Test 1, we compared the different levels of Gaussian Noise and compression-like distortions using Octree-pruning. We found a similarity between our results and the findings obtained through lab-based test in [1], where users also find noisy sequences at a higher standard deviation more annoying than in Octree-pruning at lower point cloud density. In Test 2, we compared the subjective quality G-PCC (RAHT) and V-PCC (RA) encoders on coloured point cloud volumetric videos. The findings obtained in the lab by Zerman et al. [7] could be replicated also in the remote test, showing that V-PCC (RA) outperforms the G-PCC (RAHT). Hence, it seems to be feasible with some assumptions to conduct remote tests for the subjective quality assessment of point clouds. To confirm this initial finding, further analysis is required, e.g. considering different coding parameters.

In future work, we propose to conduct subjective tests where participants will have the possibility to interact with point clouds using a mouse, and their interaction will be recorded for analysis. Furthermore, we intend to conduct further remote subjective tests to compare the results with lab-based tests performed using an HMD.

ACKNOWLEDGMENTS

This work was partly supported by the European Commission's Horizon Europe program under grant 101070109 TRANSMIXR (https://transmixr.eu/).

REFERENCES

[1] Evangelos Alexiou and Touradj Ebrahimi. On the performance of metrics to predict quality in point cloud representations. In *Applications of Digital Image Processing XL*, volume 10396, page 103961H. International Society for Optics and Photonics, 2017.

[2] E. Alexiou and T. Ebrahimi. Impact of visualisation strategy for subjective quality assessment of point clouds. In *2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pages 1–6, 2018.

[3] Alireza Javaheri, Catarina Brites, Fernando Manuel Bernardo Pereira, and Joao M. Ascenso. Point cloud rendering after coding: Impacts on subjective and objective quality. *IEEE TMM*, 2020.

[4] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. High-quality streamable free-viewpoint video. *ACM Trans. Graph.*, 34(4), July 2015.

[5] R. Mekuria, K. Blom, and P. Cesar. Design, implementation, and evaluation of a point cloud codec for tele-immersive video. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(4):828–842, 2017.

[6] K. Cao, Y. Xu, and P. Cosman. Visual quality of compressed mesh and point cloud sequences. *IEEE Access*, 8:171203–171217, 2020.

[7] Emin Zerman, C. Ozcinar, Pan Gao, and A. Smolić. Textured mesh vs coloured point cloud: A subjective study for volumetric video compression. *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, 2020.

[8] Y. Nehme, P. Le Callet, F. Dupont, J. Farrugia, and G. Lavoue. Exploring crowdsourcing for subjective quality assessment of 3d graphics. In *IEEE MMSP*, pages 1–6, 2021.

[9] S. Perry, L. Cruz, E. Dumic, N. Nguyen, A. Pinheiro, and E. Alexiou. Comparison of remote subjective assessment strategies in the context of the jpeg pleno point cloud activity. In *IEEE MMSP*, pages 1–6, 2021.

[10] E. Alexiou, E. Upenik, and T. Ebrahimi. Towards subjective quality assessment of point cloud imaging in augmented reality. In *IEEE MMSP*, pages 1–6, 2017.

[11] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso. Subjective and objective quality evaluation of 3d point cloud denoising algorithms. In *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pages 1–6, 2017.

[12] M. Seufert, J. Kargl, J. Schauer, A. Nüchter, and T. Hoßfeld. Different points of view: Impact of 3d point cloud reduction on qoe of rendered images. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, 2020.

[13] S. Subramanyam, J. Li, I. Viola, and P. Cesar. Comparing the quality of highly realistic digital humans in 3dof and 6dof: A volumetric video case study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 127–136, 2020.

[14] J. van der Hooft, M. Torres Vega, C. Timmerer, A. C. Begen, F. De Turck, and R. Schatz. Objective and subjective qoe evaluation for adaptive point cloud streaming. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6, 2020.

[15] Emil Dumic and Luis A. da Silva Cruz. Subjective quality assessment of v-pcc-compressed dynamic point clouds degraded by packet losses. *Sensors*, 23(12), 2023.

[16] I. Viola, S. Subramanyam, J. Li, and P. Cesar. On the impact of vr assessment on the quality of experience of highly realistic digital humans. *Qual User Exp*, 7(3), 2022.

[17] Jannis Weil, Yassin Alkhalili, Anam Tahir, Thomas Gruczyk, Tobias Meuser, Mu Mu, Heinz Koeppl, and Andreas Mauthe. Modeling quality of experience for compressed point cloud sequences based on a subjective study. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 135–140, 2023.

[18] E. Alexiou and T. Ebrahimi. On subjective and objective quality evaluation of point cloud geometry. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3, 2017.

[19] E. Alexiou, Y. Nehme, E. Zerman, I. Viola, G. Lavoué, A. Ak, A. Smolic, P. Le Callet, and P. Cesar. Subjective and objective quality assessment for volumetric video. In *In Immersive Video Technologies, Academic Press*, 2023.

[20] B. Herfort, B. Höfle, and Carolin Klonner. 3d micro-mapping: Towards assessing the quality of crowdsourcing to support 3d point cloud analysis. *Isprs Journal of Photogrammetry and Remote Sensing*, 137:73–83, 2018.

[21] Steve Göring, Rakesh Rao Ramachandra Rao, Stephan Fremerey, and Alexander Raake. Avrate voyager: an open source online testing platform. In *MMSP*, pages 1–6. IEEE, 2021.

[22] S Perry. Jpeg pleno point cloud coding common test conditions v3.1". in: Iso/iec jtc1/sc29/wg1 n 86044, 2020.

[23] ITU-T. Recommendation itu-t p.910: Subjective video quality assessment methods for multimedia applications, April 2008.

[24] ITU-R. Recommendation itu-r bt.500-13: Methodology for the subjective assessment of the quality of television pictures, January 2012.

[25] Tobias Hoßfeld, Raimund Schatz, and Sebastian Egger. Sos: The mos is not enough! In *QoMEX*, pages 131–136, 2011.

[26] Margaret H Pinson and Stephen Wolf. Comparing subjective video quality testing methodologies. In *Visual Communications and Image Processing 2003*, volume 5150, pages 573–582. International Society for Optics and Photonics, 2003.

[27] T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi. Performance comparisons of subjective quality assessment methods for mobile video. In *2010 Second International Workshop on QoMEX*, pages 82–87, June 2010.