# Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming

# Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming

Jeroen van der Hooft[1], Maria Torres Vega[1], Christian Timmerer[2,3], Ali C. Begen[4], Filip De Turck[1] and Raimund Schatz[2,5]
[1]IDLab, Department of Information Technology, Ghent University - imec, jeroen.vanderhooft@ugent.be
[2]Institute of Information Technology, Alpen-Adria-Universität Klagenfurt [3]Bitmovin
[4]Computer Science Department, Özyeğin University [5]AIT Austrian Institute of Technology

***Abstract*—Volumetric media has the potential to provide the six degrees of freedom (6DoF) required by truly immersive media. However, achieving 6DoF requires ultra-high bandwidth transmissions, which real-world wide area networks cannot provide today. Therefore, recent efforts have started to target efficient delivery of volumetric media, using a combination of compression and adaptive streaming techniques. It remains, however, unclear how the effects of such techniques on the user perceived quality can be accurately evaluated. In this paper, we present the results of an extensive objective and subjective quality of experience (QoE) evaluation of volumetric 6DoF streaming. We use PCC-DASH, a standards-compliant means for HTTP adaptive streaming of scenes comprising multiple dynamic point cloud objects. By means of a thorough analysis, we investigate the perceived quality impact of the available bandwidth, rate adaptation algorithm, viewport prediction strategy and user's motion within the scene. We determine which of these aspects has more impact on the user's QoE, and to what extent subjective and objective assessments are aligned.**

***Index Terms*—Volumetric media, HTTP adaptive streaming, 6DoF, MPEG V-PCC, QoE assessment, objective metrics.**

## I. Introduction

Six degrees of freedom (6DoF) allows an immersive media user to move freely within the virtual environment. Enabling volumetric 6DoF, however, requires sophisticated methods for media representation and delivery. One plausible solution is to combine point cloud compression (PCC) with adaptive streaming techniques that are accustomed to two-dimensional and 360° video content. PCC significantly reduces the storage amount at the expense of complex preprocessing and rendering at the client. HTTP adaptive streaming (HAS) copes with dynamic network conditions while attempting to deliver the highest quality possible under the given circumstances.

Even though various approaches have already proposed to apply HAS to point clouds, this is still a niche area of research. Notable studies include the ones by Hosseini and Timmerer [1] and van der Hooft *et al.* [2]. While these studies agree on the advantages of using HAS to nicely trade off the streaming quality with the bandwidth consumption, the impact of this trade-off on the user perception, *i.e.*, on the quality of experience (QoE), is yet to be thoroughly examined. Thus, a subjective QoE study is necessary in order to shed light on to the initial findings, which is our primary goal in this paper.

This paper presents an analysis of the effects of PCC and HAS on the perceived quality. Starting from the state of the art, we assess the quality by means of both subjective and objective evaluations. Our *three primary contributions* are as follows: (*i*) we perform a subjective quality assessment study on 6DoF adaptive point cloud streaming (PCS), (*ii*) we analyze the impact of different network conditions and configurations on the QoE, and (*iii*) we provide a benchmark for objective quality metrics. Before we get to the details of the study and results, we provide background on point cloud compression and streaming, and review related work.

## II. Background and Related Work

### A. Point Cloud Compression and Streaming

HAS allows the client to adapt the video quality based on the network conditions, playback buffer status, user preferences and the considered video content. When it comes to point clouds, several PCC techniques exist to compress the original data (*e.g.*, the encoder by Mekuria *et al.* [3] and the recent V-PCC encoder [4]). Changing the quantization parameters results in multiple representations, each of which comes with a different bit rate and quality. These compressed objects can then be retrieved by a client and used to render a three-dimensional scene with 6DoF.

Hosseini and Timmerer are the first to propose a standards-compliant approach for on-demand PCS of a single object [1]. The authors sample different points to generate versions of lower quality, and request objects on a per-frame basis. He *et al.* consider view-dependent single PCS, using a cubic projection to create six two-dimensional images that are then compressed using traditional compression techniques [5]. The proposed approach relies on a (hybrid) broadband and broadcast network and in-network optimizations such as caching. Li *et al.* introduce a framework for PCS, balancing communication and computational resources to maximize a proprietary QoE metric [6]. Simulation results are promising, but do not consider important factors such as latency. Van der Hooft *et al.* present PCC-DASH, a framework for streaming scenes consisting of multiple point cloud objects [2]. PCC is used to prepare multiple quality versions of the objects, and several rate adaptation heuristics that take into account the user's position and viewing angle are proposed.

### B. Objective Metrics for Point Cloud Streaming

Multiple objective metrics have been used to represent the quality of point clouds in the past. In this regard, a distinction must be made between the quality of a point cloud object compared to a derived version of the same object, and the quality of the rendered field of view (*i.e.*, what a user observes when looking through a head-mounted display). To compare the quality of derived point clouds, two Peak Signal-to-Noise Ratio (PSNR)-based metrics have been proposed by MPEG, referred to as point-to-point and point-to-plane geometry distortion metrics [7]. The former calculates the mean square error (MSE) between the reference and reconstructed points (both for geometry in terms of $x$, $y$ and $z$, and for color in terms of $YUV$). The latter calculates the MSE between the surface plane and reconstructed points. PSNR values are obtained based on the volume resolution for geometry and on the color depths for each color channel.

Although relevant to assess the performance of compression techniques for volumetric media, these metrics give no indication of how the user visually perceives the corresponding point cloud object(s) from a given viewpoint or angle. Recently, well-known metrics for traditional video streaming have been applied to assess the visual quality of the rendered point cloud content, compared to a given benchmark (*e.g.*, uncompressed point cloud objects). These metrics include the PSNR, the structured similarity index (SSIM) [8] and the multiscale SSIM [9], to name a few. Although these metrics give an indication of the visual quality of the rendered content, it should be noted that they take the background of the consumed scenes into account. This background contributes less to the perceived quality, since it is expected that users mainly focus on the objects in the front. One work has considered background removal for images generated from point cloud content, using a MATLAB-based tool for assisted removal [10]. Although useful, this feature comes with a significant computational overhead when video is considered.

### C. Subjective Evaluation of Point Cloud Quality

Subjective assessment of the quality of point cloud rendering has been performed in a number of settings and configurations featuring different attributes (colored, colorless), rendering types (raw points, cube, mesh) and degradation types (compression, noise, octree-pruning), see [11], [12] for a comprehensive overview. The majority of studies so far has focused on static models (except, *e.g.*, [3]), using a passive evaluation protocol (except, *e.g.*, [11], [13]) and double-stimulus testing (except, *e.g.*, [3]). Recent work has confirmed the viability of using pre-rendered 2D imagery for point cloud subjective quality testing (*cf.*, [11], [14]) and investigated alternatives to double-stimulus testing (*cf.*, [15]).

In short, two common key aspects are found in the literature: encoding evaluation and double-stimulus assessment. First, almost all studies so far have focused on the quality degradation due to the encoding. However, this degradation does not cover the effect of networks on the perceived quality, which will be fundamental given the bandwidth required by PCS. Second, subjective evaluations were predominantly performed using double-stimulus testing, where the users rate the degradation of one video comparing to the unimpaired source. While double-stimulus provides the means to assess the perceived degradation of the content, it comes short to assess the overall perceived quality. The aim of this work is thus to provide an experimental subjective and objective evaluation of the effects of adaptive PCS on the user's QoE.

TABLE I: Parameter configurations considered in this work.

| Parameter | Configurations |
|---|---|
| Point cloud objects | loot, redandblack, soldier, longdress |
| Quality representations | R1, R2, R3, R4, R5 |
| Segment duration | 1 second |
| Scene / camera path | 1, 2, 3 |
| Bandwidth | 20, 60, 100 Mb/s |
| Latency | 37 ms (reference for 4G) |
| Buffer size | 4 seconds |
| Object priority | $A_{vis}$ |
| Bit rate allocation | uniform, view-focused |
| Prediction | most recent, clairvoyant |

## III. Experimental Approach

To address aforementioned gaps in the existing research, we conducted a subjective and objective evaluation of the QoE of adaptive PCS. Our aim was to assess the QoE impact of relevant settings and parameters in a setup that considers the dynamic nature of PCS (animated models and moving camera). Further, we wanted to benchmark common objective metrics against subjective results in order to identify their alignment with human perception. In particular, the purpose of this research is to answer the following two questions:

**RQ1:** What is the impact of available network bandwidth, viewport prediction, bit rate allocation, and 3D scene type on adaptive PCS quality perception?

**RQ2:** How do objective image-based metrics correlate with the subjective quality for different adaptive PCS delivery settings and scenarios?

The remainder of this section presents the required steps and approaches used to answer these research questions.

### A. Evaluation Space and Content Generation

We implemented the setup and algorithms proposed by van der Hooft *et al.* [2] and deployed them on a dedicated network testbed[1]. This allowed us to replicate their experiments and analyze the observed results ourselves. Multiple parameters are defined by the authors, of which the ones in Table I were considered in this work.

First, the four point cloud objects from the 8i dataset [16] were encoded using the V-PCC encoder with MPEG's five reference quality representations [4]. As an example, Fig. 2 shows the lowest and highest quality representation of one of the frames of the *soldier* object, which has an uncompressed bit rate of $5.7\,\mathrm{Gb/s}$. The original content comes with 300 frames per object, at a frame rate of 30 frames per second. A segment duration of one second (corresponding to 30 frames) is considered in this work.

[1] https://doc.ilabt.imec.be/ilabt/virtualwall/

Fig. 1: Screenshot of an example field of view.

TABLE II: Selected parameter configurations for each SRC.

| Bandwidth [Mb/s] | Content | Allocation | Prediction |
|---|---|---|---|
| 20 | compressed | view-focused | most recent |
| 60 | compressed | view-focused | most recent |
| 100 | compressed | view-focused | most recent |
| 20 | compressed | view-focused | clairvoyant |
| 60 | compressed | view-focused | clairvoyant |
| 100 | compressed | view-focused | clairvoyant |
| 60 | compressed | uniform | most recent |
| $\infty$ | original | N/A | N/A |

Second, the point cloud objects were merged together to form different scenes. We considered two different types of scenes: one in which the objects (humans) are positioned in a line, and one in which they are placed on a semi-circle. To allow the evaluation of a larger number of conditions, three excerpts of the original content (120 s, playing out the 300 frames forward and backward six times each) were extracted. For the remainder of this paper, these excerpts are referred to as Reference Source Sequences (SRC). The first SRC, with a duration of 24 s, belongs to the first scene and pans the four objects. The other two, with a duration of 18 s, are taken from the second scene and zoom in and out of two different objects (*loot* and *redandblack* in SRC 2, *soldier* and *longdress* in SRC 3). With regard to the user's movement and focus, it is worth noting that the same programmatically generated traces were used as the ones in [2]. This allowed us to generate and render the field of view of different videos using the MPEG point cloud compression renderer[2]. An example screenshot can be found in Fig. 1. The three resulting videos (with the original point cloud objects) have been made available online[3].

Third, a network was emulated in which a single client was connected to an HTTP server. Using traffic control (tc) on a shared network link, the available bandwidth was fixed to discrete values (20, 60 and 100 $\mathrm{Mb/s}$). The latency was set to 37 $\mathrm{ms}$, a reference value for 4G networks. The server contained all point cloud objects at different quality representations, and offered a manifest file containing all the metadata required during streaming. The client used a Python-based player, which allowed to set multiple configurations. In this work, the

[2]http://mpegx.int-evry.fr/software/MPEG/PCC/mpeg-pcc-renderer/

[3]https://users.ugent.be/~jvdrhoof/pcc-dash/

Fig. 2: Two representations of the *soldier* object: R1 at $4.5\,\mathrm{Mb/s}$ (left) and R5 at $40.4\,\mathrm{Mb/s}$ (right).

buffer size was set to four seconds, while the size of the visual area (defined as $A_{vis}$ in [2]) was used to prioritize objects within the scene. Once these objects were ranked according to their priority, the available bandwidth was allocated to the four point cloud objects in a uniform way (*i.e.*, increase the quality of all objects one by one, starting with the object of the highest priority), or in a view-focused way (*i.e.*, increase the quality of the visible objects first, before considering the objects outside of the field of view). Finally, there was an option to use either the most recent information on the user's position and focus at the time of buffering, or use clairvoyant prediction (*i.e.*, the client assumes perfect prediction on the user's position and focus when the content is being played out). Once all configurations had been set, the client started a streaming session by retrieving the different point cloud segments one by one, adapting the video quality based on the observed throughput and the configurations above. Decisions on the quality representation for each of the point cloud objects were logged on a per-segment basis, so that the resulting field of view of the entire streaming session could be generated once the experiment had finished.

For evaluation purposes, eight configurations were selected for each SRC (see Table II). This results in a total number of 24 processed video sequences (PVS).

### B. *Subjective Quality Experiments*

For the test setup and procedure, we used ITU-R BT.500-13 and ITU-T P.913 as general guidelines. A screen with 2K resolution was used to passively show the content to the user, who was sitting at a distance of approximately four times the height of the screen. The applied test protocol was as follows:

1) Welcome (3 min): Briefing and informed consent.
2) Setup (2 min): Screening and demographic data.
3) Training (1 min): Evaluation of a single video example not related to point clouds.
4) Evaluation (16 min): 24 PVSs, post-stimulus questionnaire.
5) Debriefing (2 min): Feedback and remarks.

Prior to the experiment, subjects were screened for correct visual acuity using Snellen charts (20/20) and for color vision using Ishihara charts. A short training was performed at the beginning of each test session to familiarize the subjects with the test procedure. During the evaluation

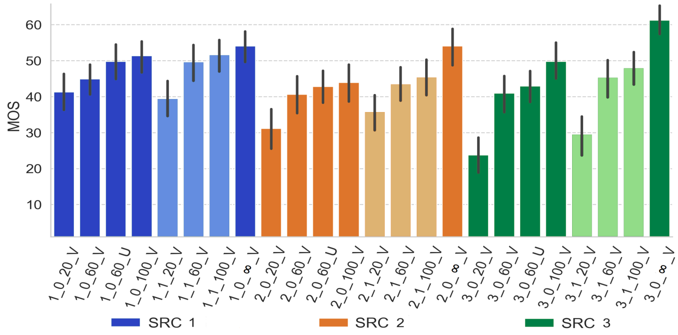


Fig. 3: MOS scores for the 24 PVSs. MOS is normalized to 1-100, with 1-20 equating "bad" and 81-100 equating "excellent" quality. (MOS CI = 0.95). Each color designates an SRC, with conditions further grouped by viewport prediction and ordered by increasing network bandwidth. Infinite bandwidth designates reference conditions. Condition code: SRC, prediction [0: most recent, 1: clairvoyant], bandwidth [Mb/s], rate allocation [V: view-focused, U: uniform].

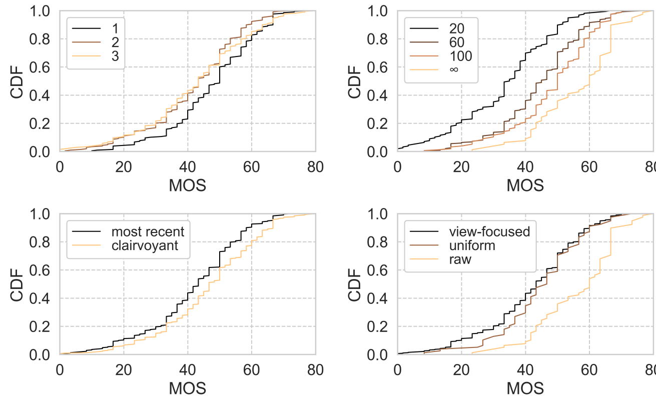


Fig. 4: CDF plots of subjective quality ratings grouped by SRC (top left), available bandwidth [Mb/s] (top right), viewport prediction (bottom left) and bit rate allocation (bottom right).

sessions, 24 PVSs were shown, corresponding to the three SRCs with different configurations. The post-stimulus rating questionnaire prompted participants to rate the quality of each PVS on an ACR-7 continuous scale.

### C. *Objective Quality Evaluation*

To objectively evaluate the generated fields of view, the considered PVSs were compared with the SRCs containing the original, non-compressed point cloud objects. Three full-reference video metrics were considered in this paper:

- Weighted PSNR [17]: Using reference weights of 0.75, 0.125 and 0.125 to the YUV components, respectively;
- SSIM [8]: Averaged over all frames in the video;
- VQM [18]: The standardized NTIA General Model, using full-reference calibration.

Results for these metrics were compared with those of the subjective evaluation procedure. Note that the distortion metrics, specifically developed for point clouds, cannot be applied, because the considered scenes consist of multiple point clouds, which are shown either sequentially or in parallel throughout the considered SRCs.

## IV. Evaluation Results

In this section, we first answer **RQ1** discussing the results of the subjective study and the objective evaluation. We then address **RQ2** via a comparative analysis of both result sets.

### A. *Impact of Content/Streaming-Related Factors (**RQ1**)*

*1) Subjective Results:* A total of 30 subjects participated in our subjective experiment: 7 subjects were female and 23 were male, while 19, 10 and 1 subjects were between 20-29, 30-39 and 40-49 years old, respectively. Subjects were recruited from Ghent University, offering participants a chance of winning a movie ticket in a raffle as incentive. As the result of our outlier removal procedure, according to ITU-R BT.1788 Annex 2 [19], we eliminated four subjects from the study dataset resulting in a final subject count of $N = 26$.

Fig. 3 shows the normalized MOS scores (0-100) for each test condition. An important observation is that the range of the average MOS scores for all configurations is limited to an interval between 23.8 and 61.3. Even though the subjects were shown the uncompressed content, the majority of the people rated the videos on the lower end of the quality spectrum. This shows that the considered content, which was captured using 42 cameras and requires $19.7\,\mathrm{Gb/s}$, was not to the subjects' standards. This was also clear from the debriefing interviews with feedback and remarks, where the subjects often mentioned that their expectations of the visual content were higher because of their familiarity with full-HD and 4K resolutions for traditional video.

Fig. 4 shows the cumulative distribution function (CDF) of the MOS scores given by the 26 subjects, for different configurations. Results show that SRC 1 (line-up of the objects, zoomed out) is rated higher than SRC 2 and 3 (zoom in on a specific object). This can be related to the visual size and area of the considered point cloud objects, which is significantly lower in SRC 1. Also the texture of the point cloud objects turned out to be an important factor to the subjects. Multiple subjects indicated to have given higher scores to SRC 2 than to SRC 3, because the *loot* and *redandblack* objects show less contrast differences than the *soldier* and *longdress* objects in SRC 3. This explains the higher range of MOS scores for the latter content. We also observe that higher bandwidth values result in higher MOS scores, which was to be expected. Viewport prediction leads to better results as well, which can be explained by the lack of quality switches when shifting the focus from one object to the other (the client was able to anticipate this change during buffering). The bit rate allocation scheme also has an impact on the subjects' perception of the clip, with a preference toward uniform bandwidth allocation. Finally, the results show that even the uncompressed content is rated less than 60 almost half of the time. This again indicates that the subjects were generally underwhelmed by the provided video sequences.

The above observations are confirmed by our mixed model ANOVA results, which identify all factors except the

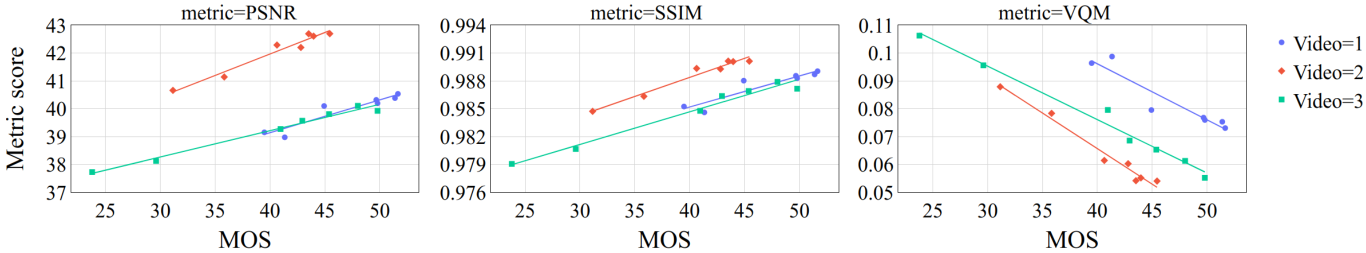


Fig. 5: Scatter plots for the different objective metrics (PSNR, SSIM, VQM) with MOS scores as the ground truth. Each dot represents a test condition, each color a video sequence. Lines are fitted via linear regression using ordinary least squares.

TABLE III: Mixed-model ANOVA results for fixed (F-Test) and random effects (likelihood-ratio test). Asterisks indicate levels of significance.

| Fixed Effects | F | p |
|---|---|---|
| Bandwidth | 73.177 | <0.001 *** |
| Prediction | 4.830 | 0.035 * |
| Allocation | 2.844 | 0.092 |
| SRC | 12.472 | <0.001 *** |
| **Random Effects** | **ChiSq** | **p** |
| Bandwidth:User | 6.499 | 0.011 * |
| Prediction:User | 3.937 | 0.047 * |
| Allocation:User | 0.000 | 0.998 |
| SRC:User | 12.644 | <0.001 *** |
| User | 32.494 | <0.001 *** |

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

bandwidth allocation strategy as exerting significant influence on the MOS (see Table III), with the prediction component being on the borderline ($p$=0.035). In addition, we also tested for user-related influences by analyzing the random effects part of our mixed-effects model. Indeed, the model suggests the presence of significant "assessor effects" on quality ratings ($p$<0.001). In particular, the influence of the shown SRC on quality rating behavior varied significantly across participants ($p$<0.001). However, we could not detect any systematic influence of subject variables such as age or gender.

*2) Objective Results:* Table IV shows the results obtained for the objective metrics. Three observations can be made. First, even though the configurations are significantly different (especially regarding the available bandwidth), differences in terms of metrics are limited. Looking at the obtained PSNR values, for instance, the range among all PVSs equals 4.97, while the highest range among the three SRCs merely equals 2.38. This can mostly be attributed to the fact that the considered metrics are calculated based on the whole field of view, which includes the (static) background. A different point cloud representation thus affects these metrics in a less pronounced manner than is the case for traditional video.

Second, although differences are small, the observed trends are evident: ($i$) with increasing network bandwidth, better scores are always observed, ($ii$) viewport prediction has a positive effect on the resulting video quality, and ($iii$) uniform allocation of the available bandwidth leads to better results than prioritizing objects visible at the time of buffering. The latter is related to our prior observation that a change of focus can result in a negative impact on the observed video quality.

Third, a strong linear correlation between the considered metrics is observed. As shown in Fig. 6, the highest correlation is achieved for the PSNR and SSIM metrics, with a Pearson correlation coefficient of 0.85. Similar results are observed for other correlation metrics, such as Spearman's rank correlation.

TABLE IV: Subjective MOS and objective metrics for the different test conditions. Condition code: SRC, prediction, bandwidth, bit rate allocation (see Fig. 3).

| Condition | MOS | PSNR | SSIM | VQM |
|---|---|---|---|---|
| 1_0_20_V | 41.3462 | 38.9713 | 0.9846 | 0.0987 |
| 1_0_60_V | 44.9365 | 40.0958 | 0.9880 | 0.0795 |
| 1_0_100_V | 51.4000 | 40.3799 | 0.9887 | 0.0753 |
| 1_1_20_V | 39.4873 | 39.1501 | 0.9852 | 0.0964 |
| 1_1_60_V | 49.6907 | 40.3158 | 0.9885 | 0.0768 |
| 1_1_100_V | 51.6669 | 40.5320 | 0.9890 | 0.0730 |
| 1_0_60_U | 49.8088 | 40.1941 | 0.9883 | 0.0759 |
| 2_0_20_V | 31.1538 | 40.6554 | 0.9847 | 0.0879 |
| 2_0_60_V | 40.6415 | 42.2854 | 0.9893 | 0.0614 |
| 2_0_100_V | 43.9754 | 42.6054 | 0.9901 | 0.0551 |
| 2_1_20_V | 35.8327 | 41.1388 | 0.9863 | 0.0783 |
| 2_1_60_V | 43.5254 | 42.6909 | 0.9901 | 0.0541 |
| 2_1_100_V | 45.4488 | 42.6925 | 0.9901 | 0.0540 |
| 2_0_60_U | 42.8208 | 42.1978 | 0.9893 | 0.0602 |
| 3_0_20_V | 23.7823 | 37.7223 | 0.9790 | 0.1062 |
| 3_0_60_V | 40.9615 | 39.2676 | 0.9848 | 0.0796 |
| 3_0_100_V | 49.8077 | 39.9239 | 0.9871 | 0.0551 |
| 3_1_20_V | 29.6150 | 38.1217 | 0.9807 | 0.0956 |
| 3_1_60_V | 45.3858 | 39.8047 | 0.9869 | 0.0653 |
| 3_1_100_V | 48.0119 | 40.1055 | 0.9879 | 0.0612 |
| 3_0_60_U | 42.9485 | 39.5652 | 0.9864 | 0.0685 |

### *B. Subjective vs. Objective Results (**RQ2**)*

In this section, we answer **RQ2** by comparing subjective assessment results with objective metrics. To this end, Fig. 5 shows the scatter plots for the considered objective metrics with the MOS scores as the ground truth. Each dot represents a test condition, as defined in Table II. Lines are fit through linear regression, using ordinary least squares. From these graphs, we observe that there is a clear linear correlation between the MOS scores and the considered objective metrics. This corresponds to the results observed for the Pearson correlation coefficients in Fig. 6, with values of 0.40, 0.82 and -0.59 for PSNR, SSIM and VQM, respectively. Comparing results with the observed MOS scores, the SSIM metric thus seems to best reflect the increasing/decreasing trend of the subjects' preferences toward the considered content. However, given the observed offsets between the different videos, it should be noted that the relation between the subjective scores and the objective metrics is strongly SRC-dependent. Indeed, although the MOS scores overlap, objective results for SRC 2 are significantly better than for SRC 1 and 3. Thus, we

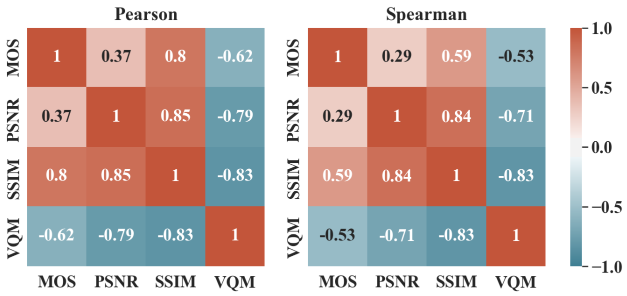


Fig. 6: Pearson and Spearman correlation coefficients for the MOS scores and the three objective metrics, over all 24 PVSs.

conclude that it is not possible to accurately deduce QoE scores solely from the objective results. We believe this can be attributed to two factors.

First, the scores for the objective metrics strongly depend on the properties of the point cloud object(s). Objects that are smaller, for instance, contribute less to the PSNR of the field of view. Real users, however, tend to strongly focus on (the quality of) the objects, without taking the background into account. Furthermore, it is harder for users to evaluate (differences between) the quality of the objects that are farther away, resulting in a lower range of MOS scores.

Second, the objective scores do not reflect the impact of quality switches within the video. These metrics are based on (weighted) averages and do not take dynamic behavior into account. A typical user, however, pays attention to quality switches and rate the video as such. Psychological factors play an important role here, in that a subject remembers that what was bad (*e.g.*, switching to a lower quality when a new object is being focused on). This shows that accurate prediction models for 6DoF user movement are important to improve the QoE in these applications.

## V. Conclusions and Future Work

In this paper, we evaluated the impact of adaptive streaming optimizations on the quality of experience (QoE) of point clouds. Based on the state of the art, we prepared a set of 24 impaired volumetric test videos that were analyzed both subjectively, through a single stimulus approach, and objectively, through the full reference metrics PSNR, SSIM and VQM. First, we found out that users are able to provide accurate and consistent responses when assessing quality without the presence of the ground truth, even though the most common point cloud quality assessment approach in the literature so far has been double stimulus. However, given the different nature of volumetric media, subjects tend to give lower ratings than those for traditional HD or 4K videos. Second, high correlation between objective and subjective metrics was shown for the case of adaptive point cloud streaming. Nonetheless, objective metrics need to be rescaled or adjusted to properly match the human perception. Thus, there is a need for more representative metrics and QoE models that more accurately reflect the quality perceived by the user. In future work, we aim to extend our experimental test set with additional scenes, conditions as well as a comparative analysis of the performance of single versus double stimulus testing for this type of media. Furthermore, we plan to use alternative video-based point cloud compression techniques that allow for real-time decoding of the considered point cloud objects. This way, we can evaluate truly interactive 6DoF video scenarios.

## Acknowledgments

This research is part of a collaborative project between Huawei and Ghent University, funded by Huawei Technologies, China, and has been supported in part by the Christian Doppler Laboratory ATHENA (https://athena.itec.aau.at/). Maria Torres Vega is funded by the Research Foundation Flanders (FWO), grant number 12W4819N.

## References

[1] M. Hosseini and C. Timmerer. Dynamic Adaptive Point Cloud Streaming. In *Packet Video*, 2018.
[2] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner. Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression. In *ACM Multimedia*, 2019.
[3] R. Mekuria, K. Blom, and P. Cesar. Design, Implementation, and Evaluation of a Point Cloud Codec for Tele-Immersive Video. *IEEE Trans. CSVT*, 27(4), 2017.
[4] S. Schwarz et al. Emerging MPEG Standards for Point Cloud Compression. *Jour. Emerging and Selected Topics in Circuits and Systems*, 9(1), 2018.
[5] L. He, W. Zhu, K. Zhang, and Y. Xu. View-Dependent Streaming of Dynamic Point Cloud over Hybrid Networks. In *Advances in Multimedia Information Processing*, pages 50–58. Springer, 2018.
[6] Jie Li, Cong Zhang, Zhi Liu, Wei Sun, and Qiyue Li. Joint Communication and Computational Resource Allocation for QoE-driven Point Cloud Video Streaming. *arXiv e-prints*, page arXiv:2001.01403, 2020.
[7] MPEG. MPEG 3DG and Requirements - Call for Proposals for Point Cloud Compression V2. https://bit.ly/2RSWdWe, 2017.
[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing*, 13(4), 2004.
[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale Structural Similarity for Image Quality Assessment. In *The 37th Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402, 2003.
[10] E. Alexiou and T. Ebrahimi. Exploiting User Interactivity in Quality Assessment of Point Cloud Imaging. In *QoMEX*, 2019.
[11] 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 Trans. Signal and Information Proc.*, 8, 2018.
[12] E. Dumic, C. R. Duarte, and L. A. da Silva Cruz. Subjective Evaluation and Objective Measures for Point Clouds - State of the Art. In *Int. Colloquium on Smart Grid Metrology (SmaGriMet)*, page 1–5, 2018.
[13] E. Alexiou and T. Ebrahimi. On Subjective and Objective Quality Evaluation of Point Cloud Geometry. In *QoMEX*, 2017.
[14] E. Alexiou et al. Point Cloud Subjective Evaluation Methodology Based on 2D Rendering. In *QoMEX*, 2018.
[15] E. Zerman, P. Gao, C. Ozcinar, and A. Smolic. Subjective and Objective Quality Assessment for Volumetric Video Compression. *Image Quality and System Performance XVI*, 2019(10), 2019.
[16] E. d'Eon, T. Myers, B. Harrison, and P. A. Chou. Joint MPEG/JPEG Input. 8i Voxelized Full Bodies - A Voxelized Point Cloud Dataset. https://jpeg.org/plenodb/pc/8ilabs/, 2017.
[17] J. Ohm, G. J. Sullivan, H. Scharz, T. K. Tan, and T. Wiegand. Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC). *IEEE Trans. CSVT*, 22(12), 2012.
[18] M. H. Pinson and S. Wolf. A New Standardized Method for Objectively Measuring Video Quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, 2004.
[19] ITU-R Rec. BT.1788. Methodology for the Subjective Assessment of Video Quality in Multimedia Applications, 2007.