Subjective QoE Evaluation of User-Centered Adaptive Streaming of Dynamic Point Clouds

Subramanyam, Shishir; Viola, Irene; Jansen, Jack; Alexiou, Evangelos; Hanjalic, Alan; Cesar, Pablo

DOI
10.1109/QoMEX55416.2022.9900879

Publication date
2022

Document Version
Final published version

Published in
2022 14th International Conference on Quality of Multimedia Experience, QoMEX 2022

2022 14th International Conference on Quality of Multimedia Experience (QoMEX)

Subjective QoE Evaluation of User-Centered Adaptive Streaming of Dynamic Point Clouds

Shishir Subramanyam*†, Irene Viola*, Jack Jansen*, Evangelos Alexiou*, Alan Hanjalic† and Pablo Cesar*†
*Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
†TU Delft, Delft, The Netherlands

***Abstract*—Technological advances in head-mounted displays and novel real-time 3D acquisition and reconstruction solutions have fostered the development of 6 Degrees of Freedom (6DoF) teleimmersive systems for social VR applications. Point clouds have emerged as a popular format for such applications, owing to their simplicity and versatility; yet, dense point cloud contents are too large to deliver directly over bandwidth-limited networks. In this context, user-adaptive delivery mechanisms are a promising solution to exploit the increased range of motion offered by 6DoF VR applications to yield gains in perceived quality of 3D point cloud user representations, while reducing their bandwidth requirements. In this paper, we perform a user study in VR to quantify the gains adaptive tile selection strategies can bring with respect to non-adaptive solutions. In particular, we define an auxiliary utility function, we employ established methods from the literature and newly-proposed schemes for distributing the bit budget across the tiles, and we evaluate them together with non-adaptive streaming baselines through subjective QoE assessment. Results confirm that considerable gains can be obtained with user-adaptive streaming, achieving bit rate gains of up to 65% with respect to a non-adaptive approach to deliver comparable quality. Our analysis provides useful insights for the design and development of social VR applications.**

***Index Terms*—6DoF, virtual reality, point cloud, adaptive streaming, QoE assessment, teleimmersion**

I. Introduction

In recent years, advances in head-mounted displays (HMDs) and 3D capturing devices have enabled a plethora of immersive virtual reality (VR) experiences and applications, where users are free to navigate within the 3D scene with 6 Degrees of Freedom (6DoF). In particular, real-time dynamic point cloud representations of users have enabled social immersive 6DoF VR applications, such as teleimmersion [1], [2]. In such platforms, photo-realistic reconstructions of users have been shown to improve immersion and communication [3] as compared to avatar representations. In this imaging modality, the geometry of the object is defined by point coordinates, while associated attributes, such as color and transparency, are stored on a point basis, ensuring photo-realistic representations. Moreover, no additional computational overhead is needed for triangulation, making them suitable for real-time applications. However, dense, high precision point clouds require a large volume of data, requiring significant compression to be transmitted over bandwidth-limited networks. This has prompted active work in the topic of point cloud compression, both from academia and industry players, and has led to the definition of standardization activities and new standards for point cloud contents [4], [5]. Alongside efficient compression solutions, user navigation can be exploited to further optimize delivery of volumetric data. As only parts of the content are visible at any given time, user-adaptive streaming solutions can be deployed to reduce the bandwidth allocation for parts of the content that are outside of the field of view, ensuring a better quality for the parts that are visible.

Adaptive streaming techniques for point clouds have received significant research interest in recent years. For scenes with multiple point clouds, Hosseini et al. [6] propose DASH-PC, where they present three algorithms to spatially sub-sample point cloud objects in a scene. The client selects an appropriate density based on human visual acuity irrespective of the orientation of the underlying surface. Hooft et al. [7] propose PCC-DASH, a standards-compliant means for HTTP-adaptive streaming of scenes, comprised of multiple point cloud reconstructions. The authors propose rate adaptation heuristics to select representations for each object in the scene, based on viewport position, available bandwidth, and current buffer status. This approach uses a single quality for each object. Subramanyam et al. [8] build on the ideas presented in [7] to tile individual point cloud objects based on low-complexity surface estimation, carrying out objective quality evaluation using image distortion metrics and prerecorded navigation paths.

The aforementioned solutions have been mostly relied on objective evaluation. User evaluation of adaptive streaming strategies for point cloud contents in VR have largely been absent from the literature, due to the real-time rendering requirements that such an evaluation entails [9]. In this work, we perform a subjective evaluation study to demonstrate that user-centered tiled streaming can consistently deliver a higher QoE with respect to non-adaptively encoded content, under the same codec and bit rate. In contrast to existing studies on adaptive streaming for point clouds, which used pre-recorded videos displayed on 2D screens [10], we evaluate real-time adaptation to user movements in VR, in order to assess the quality of the tiled point cloud streams in realistic conditions. To do so, we define a utility function and employ different tile selection strategies to optimize the allocation of available bandwidth. We implement a live playback environment to render synchronized tiled streams of point clouds, which supports adaptive quality selection based on user movements in 6DoF. By comparing

2022 14th International Conference on Quality of Multimedia Experience (QoMEX) | 978-1-6654-8794-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/QOMEX55416.2022.9900879

different tile selection strategies, we evaluate acceptable settings for quality differences among adjacent tiles. Moreover, we present an analysis of user interactions and navigation patterns during the quality assessment task, drawing useful insights.

II. Tile Rate Allocation

In this section, we present the utility function and tile rate allocation strategies we devised for user-adaptive point cloud streaming. In our scenario, we aim at optimizing the delivery of a single point cloud object, based on the relative position and rotation of the viewer.

A. Utility Function Definition

Let us consider a point cloud P, partitioned into N non-overlapping segments (tiles) T_i, $i = \{1,2,...,N\}$. For simplicity, we assume the segmentation to be operated on the XZ floor plane; that is, every segment spans the entire Y axis. We assume N to be even, and that each tile T_i has an associated tile centroid $T_i^{(c)}$ and a tile orientation vector $\vec{T}_i$; that is, $T_i = \left[T_i^{(c)}, \vec{T}_i\right]$. Such a vector could be estimated using the surface normal at each point, or obtained after surface reconstruction. For real-time capturing system, an approximation using the transformation matrix of each camera to construct an orientation vector was proposed and validated in [8]. For our tiling allocation strategies, we assume the orientation vectors to be such that $\forall i, \exists j : \vec{T}_i \cdot \vec{T}_j = -1$, with $i \neq j$; i.e., for each tile, there exists an opposite facing tile. For any given user visualizing the point cloud from an external vantage point, with associated viewport $V = \left[V^{(c)}, \vec{V}\right]$ in which $V^{(c)}$ defines the center of the viewport, and $\vec{V}$ the orientation, we define the absolute utility $\left|u\left(V,T_i\right)\right| = \left|\vec{V} \cdot \vec{T}_i\right|$. Such a quantity considers the viewing angle, assigning higher utility to tiles that are facing the users, and lowest utility to tiles that are orthogonal with respect to the user's viewing direction. However, it fails to consider the impact of the users' location on the visibility and importance of the tiles. To incorporate the location information in the tiling utility, for each set of tiles with equal absolute utility (i.e., opposite-facing) (T_i, T_j), we compute the Euclidean distances $d(\cdot)$ between the viewport location $V^{(c)}$ and the tile centroids $T_i^{(c)}$ and $T_j^{(c)}$. The utility is the computed as follows:

$$u(V,T_i) = \begin{cases} \left|\vec{T}_i \cdot \vec{V}\right|, & \text{if } d(V^{(c)},T_i^{(c)}) < d(V^{(c)},T_j^{(c)}) \\ -\left|\vec{T}_i \cdot \vec{V}\right|, & \text{otherwise.} \end{cases} \quad (1)$$

The resulting utility can then be used to divide the bit rate budget among the available tiles; a larger utility will correspond to higher visibility, whereas a smaller utility indicates lower visibility. As we compute utility using unit vectors, all values lie between -1 and 1. A representation for each tile can then be selected, and the final representation for a frame is retrieved from the server.

B. Tile Rate Allocation Schemes

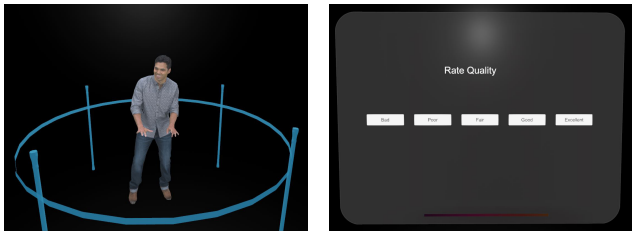
In this work, we consider three allocation schemes to select representations for each of the tiles. The first two were originally proposed by Hooft et al. [7] for multiple point cloud objects in a scene rather than surfaces of a single point cloud. This approach was adapted and used for point cloud segments in [8]. In addition, the authors originally proposed a *Hybrid* allocation scheme, where tiles are separated into visible and not visible tiles. The representation of visible tiles is uniformly increased. Then, the remaining bandwidth budget is spent by uniformly increasing the representation of the remaining tiles. This approach, while leading to better results in terms of objective evaluation as seen in [8], can lead to differences amongst tiles at the boundary of user visibility. So, in this work, we propose a *Weighted Hybrid* tile allocation scheme where the budget is allocated to each tile based on utility. The representation of each tile is then maximized based on this allocation, and the remaining bandwidth budget is subsequently used to uniformly increase the representation of each tile in order of utility. After pilot testing, we chose to use the *Weighted Hybrid* allocation scheme, as this approach ensured smoother differences amongst adjacent tiles and achieved higher perceived quality as compared to the *Hybrid* tile allocation. By introducing one additional novel allocation scheme, we aim to increase the variation in quality differences amongst adjacent tiles to identify acceptable limits for perceived quality. Below, we provide a description of the bit rate allocation strategies that are employed.

1) **Greedy (W1)** [7]: The highest quality representation is first set for the highest utility tile and then we move on the next highest-ranked tile until the bit rate budget is spent.
2) **Uniform (W2)** [7]: The representation of tiles is increased one step at a time, starting with the highest utility tile.
3) **Weighted Hybrid (W3)**: The bit rate budget is first split based on the weighted utility of each tile, defined as $\frac{u(V,T_i)+1}{\sum(u(V,T_i)+1)}$. The highest possible representation is then set for each tile based on the budget allocated to each tile. The remaining bit rate budget is then used to uniformly increase the representation of each tile in the order of their utility.

III. Experimental Approach

A. Subjective Evaluation Platform

Real-time user interactions in 6DoF VR applications require low latency systems for selecting and rendering content at the client side. While previous research has focused on objective evaluation [6], [8] or offline loading of point clouds on physical memory for subjective evaluation [9], [11], in this study, we conduct the subjective assessment of tiled streams of point clouds under realistic playback conditions and user interactions in real-time. We implemented the playback environment using the Unity game engine and a serialized binary point cloud reader. Each point is rendered as a camera facing quad with the offset determined by the point size. We determine the point size for each point cloud frame offline, based on the average distance of each point to its 5 nearest neighbors, following previous research on the field [11], [12]. The evaluation environment consists of two 3D scenes to playback and evaluate tiled point



(a) *Playback scene* (b) *Rating scene*

Fig. 1: Scenes from the playback application.

(a) *Longdress* (b) *Loot* (c) *Red and Black* (d) *Soldier*

Fig. 2: Sequences in the 8i Voxelized Full Body Dataset [13].

cloud content. The *playback scene* places the point cloud sequence at the centre and spawns the user at a location in front of the point cloud. Players are free to walk and look around as they inspect the point clouds while performing the evaluation. To ensure participant safety and consistent boundaries for movement within the scene, we included a blue ring of 2.3 meters diameter to mark the extent of the physical play area as shown in Figure 1 (a). The *rating scene* places a fixed canvas on the wall with the rating scale. During playback all interactions and decisions are logged, this includes user position, user orientation, playback performance statistics, adaptation decisions and bandwidth usage.

B. Dataset Preparation

In this work, we use 4 contents from the 8i Voxelized Full Body Dataset [13], namely *Longdress*, *Loot*, *Red and Black*, and *Soldier* (see Figure 2). These sequences have been used extensively in previous research on adaptive streaming of point clouds [6]–[8] and for the development of new standards in point cloud compression [5]. Similarly to [8], we create four non-overlapping tiles by placing virtual cameras around the object on the XZ (floor) plane, at (1,0,0), (0,0,1), (-1,0,0) and (0,0,-1). We draw a vector from the centroid of the point cloud to every point on the surface, and we use the minimum vector dot product of these vectors with the 4 virtual cameras to assign a tile number to each point in the cloud.

The MPEG Anchor codec [1] is used to evaluate the impact of user adaptation, as it has real-time encoding and decoding capabilities. For each adaptation set, the tiles are encoded using the all-intra configuration, with octree depths from 7 to 10, and the JPEG quantization parameter varying from 25 to 95 in increments of 10. To measure the performance of viewport adaptive streaming, the source point clouds are additionally encoded under the same codec configurations. Moreover, we use V-PCC to provide a baseline for our tiling approaches with state-of-the-art rate-distortion performance. In particular, we encode the source point clouds using the Release 9.0 of V-PCC, and the configurations provided in the Common Test Conditions (CTC) for Category 2 All Intra (C2AI) encoding [14]. We select the rate points 2, 3 and 5 and extend it to an additional rate point, using a Texture QP of 9, a geometry QP of 12 and an occupancy precision of 2. The size of the resulting encoded bitstreams is used to set target rate points and bit rate budgets labelled as R1-R4, separately for each of the 4 sequences in the dataset.

C. Subjective Evaluation Methodology

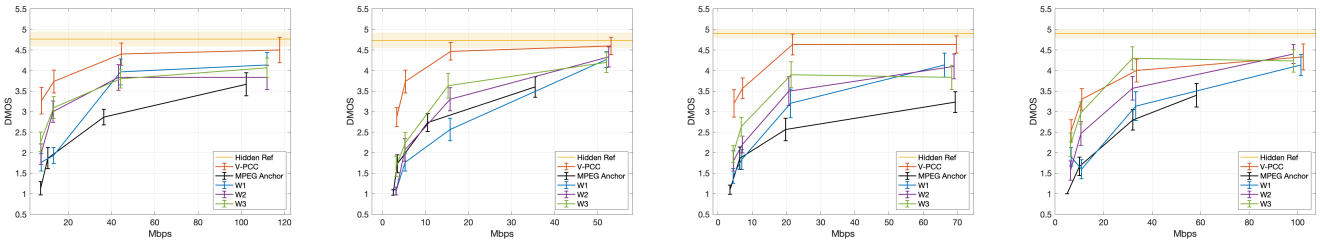
A total of 84 stimuli were generated, considering all combinations of contents (4), the tile allocation strategies with the baselines (3+2), and the target rate points (4), including hidden references (4). The test was divided into two sessions of 42 stimuli, each separated by a 10 minute break to reduce fatigue and motion sickness. The participants were requested to fill in the Simulator Sickness Questionnaire (SSQ) on a 1-4 discrete scale (1=none to 4=severe) [15] before beginning the experiment and after each session. A total of 30 participants were recruited (16 males, 14 females), with 15 reporting 1-3 prior VR experiences, 6 participants never used a VR HMD before, and 9 participants declaring to be very experienced with VR.

For the subjective experiment, we chose the Absolute Category Rating test method with Hidden References (ACR-HR), following the ITU-T Recommendation P.910 [16]. The point cloud sequences were presented to the participants one at a time and were rated independently. The participants were asked to observe a loop of 300 frames of each dynamic point cloud sequence, played back at 30 fps for a minimum of 10 seconds, and rate the corresponding visual quality on a scale from 1 to 5 (*1-Bad*, *2-Poor*, *3-Fair*, *4-Good*, and *5-Excellent*). Each sequence was rendered with a randomized initial rotation to prevent bias and encourage user movements.

Every experiment was split into three stages, namely, screening, training and testing. During the screening, the color vision of the participants was checked using the Ishihara chart, according to the ITU-T Recommendation P.910 [16]. During training, 3 versions of content not shown at the test were employed, depicting examples of *1-Bad*, *5-Excellent* and *3-Fair*. During testing, the order of the displayed stimuli was randomized per participant and per session, and the same content was never displayed twice in a row to avoid temporal referencing bias. Three dummy samples were added at the beginning of each testing session in order to ease participants into the task, with the corresponding scores subsequently discarded.

D. Data Processing

Outlier detection was performed on the individual quality scores collected by the subjects, according to the ITU-T Recommendation P.913 [16], with the recommended threshold



(a) *Longdress* (b) *Loot* (c) *Red and Black* (d) *Soldier*

Fig. 3: DMOS (solid line) and Hidden Reference (shaded area) against achieved bit rate, expressed in Mbps.

values $r_1 = 0.75$ and $r_2 = 0.8$. After outlier detection, a Mean Opinion Score (MOS) was computed for each stimulus, independently per viewing condition. The associated 95% Confidence Intervals (CIs) were computed assuming a Student's t-distribution. Additionally, the Differential MOS (DMOS) was obtained by applying hidden reference removal, following the procedure described in the ITU-T Recommendation P.913 [16].

IV. Results

A. Subjective Scores

Based on the collected scores of this experiment, no outliers were identified. Thus, the entirety of the subjective scores is employed in the subsequent analysis. In Figure 3, the results of the subjective quality assessment of the contents, are illustrated. In particular, the DMOS associated with the compressed contents are shown with solid lines, along with the relative CIs. The hidden reference scores for each content are represented with a solid yellow line to indicate the average score, along with a shaded area to represent the corresponding CIs. From the charts in Figure 3, we observe that the V-PCC codec generally achieves the best quality. Among the MPEG Anchor-based solutions, W3 usually achieves the best quality, whereas W1 has generally the lowest quality. At R1-R3, W1 results in large quality differences amongst adjacent tiles, often at different octree depths; participants reported that they sometimes found the resulting reconstruction unpleasant. Using a utility-weighted distribution of the bit-budget across tiles, as in W3, appears to result in acceptable quality differences across adjacent tiles, while optimizing the representation of tiles facing the user. In general, we observe that, at lower bit rates, the MPEG Anchor achieves similar scores to W1, indicating that participants prefer lower-quality, uniform content over large quality differences amongst adjacent tiles. W3 achieves the best quality at rate points R1-R3 for most content under test. For all sequences except *Soldier*, we observe that none of the tile allocation strategies achieve the same quality as V-PCC, independently of the rate point. Moreover, for contents *Red and Black* and *Soldier*, none of the compression solutions used in the study reach transparent quality with respect to the reference content. For contents *Longdress* and *Loot*, V-PCC is able to achieve statistically equivalent quality with respect to the uncompressed reference at the highest bit rate.

A Shapiro-Wilk normality test issued on the entirety of the subjective scores indicates that they don't follow a normal distribution ($W = 0.9093$, $p < 0.001$). Thus, non-parametric statistical tools were applied to perform an exploratory data analysis and understand whether statistical differences could be found amongst the different conditions being evaluated. To compare the different codecs and bit rate allocation strategies being tested, we first conduct a Friedman's test to check if there are any groups with significant differences ($\chi^2 = 888.69$, $p < .001$). We then conduct a pairwise post-hoc analysis with the Wilcoxon signed rank test with Bonferroni correction. The results are shown in Table I.

TABLE I: Pairwise post-hoc test codecs and bite rate allocation strategies, using Wilcoxon signed-rank test with Bonferroni correction.

Codec	Z	p	r
V-PCC – MPEG Anchor	18.261	<.001	0.589
V-PCC – W1	15.665	<.001	0.506
V-PCC – W2	14.412	<.001	0.465
V-PCC – W3	12.557	<.001	0.405
MPEG Anchor – W1	-9.846	<.001	0.318
MPEG Anchor – W2	-14.000	<.001	0.452
MPEG Anchor – W3	-16.016	<.001	0.517
W1 – W2	-6.347	<.001	0.205
W1 – W3	-10.392	<.001	0.335
W2 – W3	-6.110	<.001	0.197

It can be observed that all pairwise codec comparisons exhibit statistically significant differences with varying effect sizes. V-PCC shows a large effect size with respect to all other codecs ($r = 0.405$ to 0.589), as expected, due to its superior rate distortion performance and high encoder complexity. All three adaptive tile selection strategies show a medium to large effect size with respect to the baseline naive MPEG Anchor codec ($r = 0.318$ to $r = 0.517$). W3 in particular shows the largest effect size and Z value with respect to the naive baseline codec. This, combined with results shown in Figure 3, indicates that W3 yields significantly better visual quality as compared to the baseline across all sequences in the dataset.

In order to further confirm the impact of our tiling adaptive strategies on the scores with respect to the baseline MPEG Anchor, we perform a Kruskal-Wallis ANOVA test on the two unmatched groups. A significant effect was found on the scores ($\chi^2 = 93.11$, $p < .001$). This demonstrates the statistically significant overall performance gain across all tiling allocation strategies. In general, W3 yields the best results for content

TABLE II: Pairwise post-hoc test on contents, using Wilcoxon signed-rank test with Bonferroni correction.

Content	Z	p	r
Longdress – Loot	6.086	<.001	0.176
Longdress – Red and Black	7.060	<.001	0.204
Longdress – Soldier	6.933	<.001	0.200
Loot – Red and Black	0.717	0.473	0.021
Loot – Soldier	0.564	0.573	0.016
Red and Black – Soldier	-0.343	0.732	0.01

TABLE III: Target bit rates expressed in terms of bits per reference point averaged across all frames in a sequence.

Content	R1	R2	R3	R4
Longdress	0.32	0.56	1.87	4.93
Loot	0.15	0.24	0.70	2.34
Red and Black	0.22	0.35	1.05	3.35
Soldier	0.22	0.36	1.09	3.33

encoded with the baseline MPEG Anchor codec, as shown in Figure 3 and Table I.

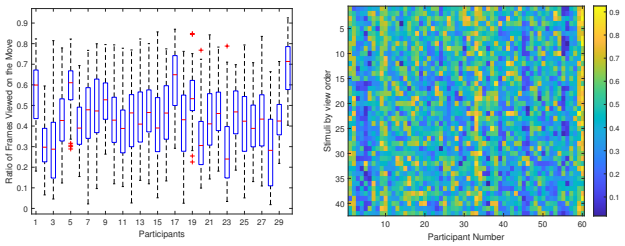
B. Content Preference

To check if there are any statistically significant differences amongst the four sequences in the dataset, we first ran a Friedman test. The results confirmed that content has a significant effect on the recorded scores ($\chi^2 = 70.24$, $p < .001$). Post-hoc analysis using the Wilcoxon signed-rank test with Bonferroni correction further confirmed that the *Longdress* sequence had statistically significant differences in scores, as shown in Table II, albeit with small to medium effect sizes ($r = 0.176$ to $r = 0.2$). The remaining three contents do not show statistically significant differences. This can be explained by the fact that the V-PCC encoding parameters lead to larger values in terms of bits per reference points for *Longdress* with respect to the other contents, as shown in Table III.

C. User Navigation

We recorded the position and rotation of the user's viewport for each participant and stimuli in the user study. In order to analyze user movement patterns, we defined a motion threshold to quantify the total frames viewed *on the move* for each stimulus. Frames were classified as viewed *on the move* if participants either translated more that 0.05 cm from the previous frame or rotated their viewport by more than 0.573 degrees along any axis from the previous frame, based on the findings of Rossi et al. [17].

The overall distribution of the ratio of frames viewed *on the move* by each participant across all 84 stimuli, is shown in Figure 4 (a). We observe a lot of variation in the median and the distribution of overall movement, essentially indicating that every participant's motion is unique. In order to see if participants' previous VR experience might have affected their navigation behaviour, we performed a comparison between the reported VR familiarity with respect to the average ratio of frames *on the move*; however, no significant correlation at 5% significance level was found (Spearman's $\rho = 0.264, p = 0.160$).



(a) Box plot (b) Viewing order

Fig. 4: Ratio of frames viewed *on the move*, across all participants.

To check if fatigue had an impact on participant movement over time, we compare the motion ratio for each combination of participant and session number in chronological order, as shown in Figure 4 (b). Here we treat each session as separate, resulting in a total of 60 participant sessions. The results show that no particular patterns of reduction in motion ratio can be observed over time that could be due to participant fatigue.

To check if adaptive playback using any of the three bit rate allocation strategies has a statistically significant effect on motion ratios, we compare it against naive playback; this includes the MPEG Anchor codec, V-PCC, and the hidden references. We performed a Kruskal-Wallis ANOVA test and observed a significant affect on motion ratios ($\chi^2 = 22.54$, $p < .001$). We observed that participants exhibit higher motion ratios while viewing adaptive content (median: 0.472, min: 0.018, max: 0.925) as compared to non-adaptive content (median: 0.407, min: 0.022, max: 0.921). When we compare the median motion ratios based on the scores assigned by participants, we observe that all scores have similar median motion ratios of 0.44. However, for stimuli rated as *fair* and *good*, we observe a larger range of motion ratios (median: 0.442, min: 0.025 and max: 0.925) as compared to all other scores (median: 0.439, min: 0.066 and max: 0.857), indicating that users might have interacted more to distinguish between these two quality ratings.

D. Simulator Sickness

Based on the SSQ filled by every subject, we observe low post-exposure total severity scores for cybersickness, as defined in [15], [18] ([mean, median, std] baseline: [5.24, 0, 10.29], after session 1: [16.83, 13.09, 16.89], after session 2: [22.32, 18.70, 19.20]). Looking at some of the individual symptoms reported by participants in the course of the experiment using the three SSQs, we found no statistically significant differences in fatigue ($\chi^2 = 4.31$, $p = 0.116$). We do observe statistically significant differences in eyestrain ($\chi^2 = 19$, $p < .001$) and general discomfort ($\chi^2 = 19.18$, $p < .001$). However, no participants reported severe symptoms on any of the SSQ questions after completing the experiment. No participants dropped out of the experiment on account of cybersickness.

V. Discussion

Results of our subjective evaluation showed clear benefits in using adaptive streaming strategies when compared with the

non-adaptive baseline as shown in Figure 3. Our results seem to indicate that smoother quality transitions between adjacent tiles are to be preferred with respect to a greedy approach, which leads to obvious discontinuities in the appearance of the content. As our evaluation focuses on point cloud contents depicting humans, boundary artifacts might be more annoying if they lie on regions of interest, such as faces, or if they influence the structural integrity of the reconstruction (e.g., disappearing fingers). We observe the best quality gains using W3, which at high bit rates approaches the quality offered by the state of the art codecs such as V-PCC. Such strategy ensures that transitions among adjacent tiles are not too drastic, while maximizing the quality of high utility tiles.

In general, segmenting the point cloud spatially leads to larger encoded payload sizes due to a loss of entropy and lower compression efficiency. While this trade-off appears to yield significant improvements in perceived quality using the MPEG Anchor codec, further assessment is required once more low-latency point cloud codecs become available, to ensure performance gains are maintained.

Finally, to evaluate the proposed adaptive streaming strategies, the network conditions and available bandwidth were set based on the CTC defined by the MPEG standardization activity [14]. While these cover a wide range of bit rates (3 - 117 Mbit/sec), similar to [8], the bit rate budget was constant for the duration of the playback sequence. The constant bit rate budget was selected to avoid introducing biasing factors in the subjective evaluation, as a variable bit rate with adaptive tiling might have been a confounding factor for both DMOS and SSQ. In order to adequately assess the performance of the system, further analyses in adverse varying network conditions are required to ensure the performance gains are maintained.

VI. Conclusion

In this paper we playback tiled streams of dynamic point clouds in 6DoF VR and adapt to user interactions in real-time for a quality evaluation task. We use tile allocation strategies that consider the user's position with respect to the point cloud to assign different quality levels to independently encoded tiles. In order to evaluate the gains in QoE, we performed a user study comparing tiled adaptive streaming of point clouds to non-adaptive approaches. A utility-weighted bit rate allocation was observed to provide acceptable quality differences amongst adjacent tiles at the target bit rates used for evaluation, and yielded the highest gains over baseline non-adaptive playback across bit rates and contents in the dataset under test.

Acknowledgment

This work is funded by the European Commission H2020 program, under the grant agreement 762111, *VRTogether*, http://vrtogether.eu/.

References

[1] 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*, January 2016.

[2] G. Cernigliaro, M. Martos, M. Montagud, A. Ansari, and S. Fernandez, "PC-MCU: Point Cloud Multipoint Control Unit for Multi-User Holoconferencing Systems," in *Proceedings of the 30th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. New York, NY, USA: Association for Computing Machinery, 2020, p. 47–53.

[3] R. Mekuria, P. Cesar, I. Doumanis, and A. Frisiello, "Objective and subjective quality assessment of geometry compression of reconstructed 3D humans in a 3D virtual room," *Proc. SPIE 9599, Applications of Digital Image Processing XXXVIII, 95991M*, September 2015.

[4] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an efficient representation of visual reality," *IEEE Multimedia*, vol. 23, no. 4, pp. 14–20, 2016.

[5] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging MPEG Standards for Point Cloud Compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, March 2019.

[6] M. Hosseini and C. Timmerer, "Dynamic Adaptive Point Cloud Streaming," in *Proceedings of the 23rd Packet Video Workshop*, ser. PV '18. New York, NY, USA: ACM, 2018, pp. 25–30.

[7] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 2405–2413.

[8] S. Subramanyam, I. Viola, A. Hanjalic, and P. Cesar, "User Centered Adaptive Streaming of Dynamic Point Clouds with Low Complexity Tiling," in *Proceedings of the 28th ACM International Conference on Multimedia*, ser. MM '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 3669–3677.

[9] 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)*. IEEE, 2020, pp. 127–136.

[10] J. van der Hooft, M. T. 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)*, 2020, pp. 1–6.

[11] E. Alexiou, N. Yang, and T. Ebrahimi, "PointXR: A Toolbox for Visualization and Subjective Evaluation of Point Clouds in Virtual Reality," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.

[12] A. Javaheri, C. Brites, F. M. B. Pereira, and J. M. Ascenso, "Point Cloud Rendering after Coding: Impacts on Subjective and Objective Quality," *IEEE Transactions on Multimedia*, pp. 1–1, 2020.

[13] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i Voxelized Full Bodies - A Voxelized Point Cloud Dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva, CH*, January 2017.

[14] MPEG 3DG and Requirements, "Complementary PCC Test Material," *ISO/IEC JTC1/SC29 WG11 Doc. N16716, Geneva, CH*, January 2017.

[15] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[16] ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, April 2008.

[17] S. Rossi, I. Viola, J. Jansen, S. Subramanyam, L. Toni, and P. Cesar, "Influence of Narrative Elements on User Behaviour in Photorealistic Social VR," in *Proceedings of the International Workshop on Immersive Mixed and Virtual Environment Systems (MMVE '21)*, ser. MMVE '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 1–7.

[18] P. Bimberg, T. Weissker, and A. Kulik, "On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 464–467.