2023 15th International Conference on Quality of Multimedia Experience (QoMEX)

Immersive and Interactive Subjective Quality Assessment of Dynamic Volumetric Meshes

Sam Van Damme[1,3], Imen Mahdi[2], Hemanth Kumar Ravuri[1], Jeroen van der Hooft[1], Filip De Turck[1], and Maria Torres Vega[1,3]
[1]IDLab, Department of Information Technology (INTEC), Ghent University - imec, Belgium
[2]Institut Polytechnique de Paris, France
[3]eMedia Research Lab, KU Leuven, Belgium

***Abstract*—Dynamic point cloud delivery can provide the required interactivity and realism to six degrees of freedom (6DoF) interactive applications. However, dynamic point cloud rendering imposes stringent requirements (*e.g.*, frames per second (FPS) and quality) that current hardware cannot handle. A possible solution is to convert point cloud into meshes before rendering on the head-mounted display (HMD). However, this conversion can induce degradation in quality perception such as a change in depth, level of detail, or presence of artifacts. This paper, as one of the first, presents an extensive subjective study of the effects of converting point cloud to meshes with different quality representations. In addition, we provide a novel in-session content rating methodology, providing a more accurate assessment as well as avoiding post-study bias. Our study shows that both compression level and observation distance have their influence on subjective perception. However, the degree of influence is heavily entangled with the content and geometry at hand. Furthermore, we also noticed that while end users are clearly aware of quality switches, the influence on their quality perception is limited. As a result, this has the potential to open up possibilities in bringing the adaptive video streaming paradigm to the 6DoF environment.**

***Index Terms*—Quality of Experience, Interactive Virtual Reality, Dynamic Volumetric Media, Meshes**

I. Introduction

Extended reality (XR) has seen an increase on popularity in terms of content and applications, with network and content providers already envisioning to offer immersive content with 6DoF. 6DoF allows the user of immersive media to move freely within the virtual environment. As such, it is expected to enable a plethora of opportunities to support interactive application domains, such as immersive training, immersive surgery, or multi-user interactive gaming.

Point cloud delivery [1] has the potential to provide the required interactivity and realism to 6DoF applications [2]. Point cloud objects are composed by a dense set of 6D points (x, y, z, and three color channels) which are presented to the user's HMD. Then, the users can move around, interact with the figures, and explore them from different sides and angles. In order to ensure a good quality of experience (QoE), a stable and high FPS and high resolution must be maintained. These impose stringent requirements on the end-user devices, which can result in low quality rendering, *i.e.* a reduced number of points in the cloud, blurriness, or freezes. However, given the current hardware limitations, such conditions cannot be guaranteed while rendering point clouds in HMDs. Thus, there is a need for point cloud content conversion to less computationally intensive options, such as meshes. However, the point cloud conversion can result in unexpected quality perception degradation such as a change in depth, level of detail, and presence of artifacts. Moreover, higher quality meshes result in higher conversion times and more data, which can influence the speed of the rendering and storage requirements of the HMD. While some studies have appeared with the focus on assessing the quality of rendered meshes [3], [4], these are mostly limited to specific use cases and fixed content quality conditions. Moreover, the assessment has mostly been done in a non-interactive manner, *i.e.*, the subject fills in a quality questionnaire after the session. This circumstance could be affected by biased memory.

Herein, we propose a systematic, general immersive, and interactive study of the perceived QoE of rendered meshes on an HMD, with the objective to evaluate the impact of different quality representations of volumetric video on the user experience in a virtual environment. In particular, our aim is to answer the following research questions:
1) How do quality level and distance affect user perception?
2) How do dynamic changes of quality level affect perception?
3) Is the user's quality assessment content-dependent?

To perform the subjective study, both a novel interactive subjective methodology as well as an immersive rendering testbed were devised. Through this study we found out that the degree to which compression and observation distance have an influence on end-user perception depends heavily on the content under scrutiny. In addition, we noticed that while end-users are aware of quality switches, the impact on their experience is minimal. As such, this opens possibilities to translate the adaptive streaming paradigm to 6DoF multimedia.

The remainder of this paper is organized as follows. Section II presents an overview of the related work. Section III discusses the adopted methodology, which was applied in an experimental setup to perform the subjective study, as detailed in Section IV. Section V presents the results, while Section VI lists the most important conclusions.

II. Related Work

Subjective quality evaluation for volumetric media, such as point clouds or meshes, is still at an early stage. As such,

979-8-3503-1173-0/23/$31.00 ©2023 European Union

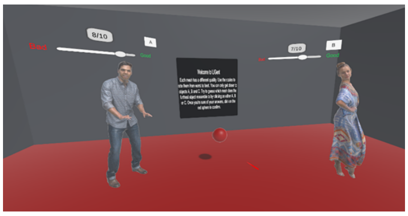

Fig. 1: Illustration of in-session quality rating.

the standards for testing methods and procedures are still to be agreed upon. Current evaluations of volumetric content are based on standards for other (immersive) multimedia such as Rec. ITU-T P.919 [5] on subjective test methodologies for omnidirectional video. The approach closest to a standard is the proposal by the JPEG committee [6] based on the work of *da Silva Cruz et al.* [7]. It consists of creating a two-dimensional (2D) video from projections of the point cloud object on a virtual camera along a predetermined path in the virtual environment. This video is then presented to the subjects in a double-stimulus test on a 2D screen, after which subjective assessment standards for traditional video can be applied. This does, however, not allow for 6DoF end-user interaction as would be the case using an HMD [8].

Key parameters of volumetric media quality evaluation study design mainly relate to presentation (interactive vs. passive, single-stimulus vs. double-stimulus), viewing technology (2D/three-dimensional (3D) screen vs. virtual reality (VR) HMD) and rendering scheme (*e.g.*, raw points vs. cubes/ellipsoids) [9]. Studies that align with this work, *i.e.*, subjectively evaluate 6DoF volumetric content in an immersed multimedia environment, mainly do this by placing the (often static) distorted and reference point cloud objects side by side in the center of the virtual environment. *Wu* et al. [10] used this approach to subjectively evaluate the impact of V-PCC encoding distortions. Their results showed that observers are sensitive to the compression distortion when QP pairs arise from (28, 37) to (36, 47).

Alternatively, a single object with absolute category rating (ACR) can be used [11]. Therefore, the user is placed in front of this object directly facing it, but is allowed to walk around to inspect the figure. So-called *guardians* are defined to make sure the user stays within a safe, predetermined area in the real environment. *Subramanyam* et al. used this method to quantify the gains adaptive tile selection strategies can bring with respect to non-adaptive solutions [11]. Their results confirmed that considerable gains can be obtained with a user-adaptive streaming solution.

To obtain subjective mean opinion scores (MOS), objects are typically rated using 5-point or 10-point Likert scales [10]–[12], which are typically presented in the virtual environment as a fixed canvas on the wall [11] or a floating, virtual tablet [10]. In this work, we present a novel in-session evaluation methodology in which the user can rate the quality of dynamic volumetric video whilst visually interacting with the content. This methodology will be elaborated on below.

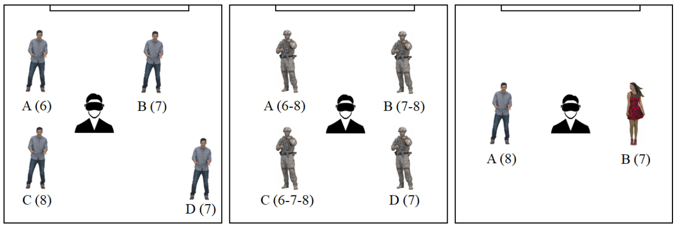

Fig. 2: Considered tests, with two to four dynamic point cloud objects A-D shown at a specific video quality between brackets: (left) an object is shown at three different qualities, with an additional object shown further away; (middle) an object is shown with regular quality switches, and (right) two different objects are shown at specific quality levels.

III. Interactive Subjective Assessment

The participant first fills in a demographics questionnaire, polling about gender, age group, eye sight and the use of technologies such as a desktop, smartphone, tablet, augmented reality (AR), VR, and a HMD. Subsequently, the participant is asked to put on the HMD and the interactive subjective evaluation session begins. In a sequential manner, the user is asked to rate and rank a set of meshes according to certain instructions. Note that no explicit training session was included, but that users were not limited in time to provide their ratings, were given an oral introduction to the practicalities of the VR controls and were informed that they could ask for any further clarification at any time. When the evaluation is over, they are asked to fill in a post-session questionnaire with their general impressions. Details of the interactive evaluation as well as the post-session questionnaires are provided next.

A. *In-Session Evaluation*

To achieve a quantitative evaluation of the perceived quality of 3D dynamic video, the participant is placed in a virtual room by means of an HMD. They can move with 6DoF within this room, bounded only by the length of the cable that connects the HMD to a personal computer (PC) that serves the content. As such, subjects had freedom of movement to inspect figures from all angles. Once immersed, the user is introduced to two to four 3D dynamically moving objects placed inside the virtual environment. The user is asked to rate the quality of each object by moving a scale displayed above the respective object (Figure 1). To guide the experiment, a blackboard displays specific instructions for each of the tests, each of which is linked to a specific research question:

How do quality level and distance affect user perception?

In the first test, the impact of two aspects is studied: quality and distance. First, we want to determine if there is a threshold above which the user can no longer differentiate between quality levels. If such a threshold exists, it can be concluded that there is no need to generate meshes with a higher quality representation. Therefore, the user is introduced to three versions of the same object, each at one of three different quality representations, referred to as A, B and C (see Figure 2). During the session, the user is asked to rate each object on a scale from 0 to 10 with 10 referring to the highest quality.

Note that we decided to deviate from the more common 5-point Likert scale in order to provide sufficient granularity to assess the sometimes subtle differences between quality representations. Second, to assess the effects of distance, a fourth object D is added further away from the participant. This object is of the same quality as one of the three other closer objects. During the evaluation, the participant is asked to also provide a quality rating for D, and to match its to either object A, B or C. To ensure the participant respects this distance boundary, a guardian in the virtual environment is used. Participants are explicitly informed of its meaning. Our hypothesis is that a user will rate an object with the same quality at a larger distance higher than an object close by. If this is the case, it can be concluded that objects at a larger distance can be retrieved at a lower quality representation without impacting the perceived quality.

How do dynamic changes of quality level affect perception?

In real-time VR applications, dynamic adjustments of the quality representation of an object over time are expected to reduce the data load and adapt to network limitations. For this reason, the user's tolerance to dynamic quality changes must be studied. During the second test, the same object is displayed four times at an equal distance from the user's initial position within the scene (see Figure 2). The quality of each instance of the object changes dynamically over time, according to a predefined scheme in a continuously repeated loop. The user is asked to rate the quality of each object. Note that we decided to collect only one score as we wanted to focus on the influence of the visibility of quality switching rather than individual qualities.Furthermore, they were asked to indicate how noticeable the quality changes are on a scale from 0 (unnoticeable) to 10 (extremely noticeable).

Is the user's quality assessment content-dependent?

At different quality levels, some 3D objects might present more details than others, or suffer from more visible artifacts. This test aims to determine if the user's quality assessment is content-dependent. To this end, the participant is introduced to a pair of two different objects encoded at specific quality levels (see Figure 2). They are asked to select the object of the highest quality, effectively comparing one object to the other.

B. Post-Session Questionnaires

Participants answer a survey following each test to further understand how each individual qualitatively experienced the task. Depending on the completed task, participants are asked about realism, artifacts, distance and differences of quality between objects. At the end, participants also score their general experience, grading statements such as "I felt I was part of the virtual environment", and "The quality of the image of the HMD was optimal". In this way, we aim to gain more insights in how the participants perceived the tasks, and how we can improve the experimental setup for future use.

IV. Experimental Setup

To perform the interactive subjective study, an experimental setup is devised in Unity 2022.1.6f1. Both the XR Interaction Toolkit 2.0.2 and XR Plugin Management 4.2.1 packages are included. As illustrated in Figure 1, a virtual room is created in which multiple 3D objects are placed. The user can explore this room with 6DoF, performing tasks that are detailed on on a blackboard on the wall.

(a) $d = 6$ (b) $d = 7$ (c) $d = 8$

Fig. 3: A screenshot of the *soldier* object [13] shown for three values of the tree depth d.

The 8i Voxelized Full Bodies (8iVFB v2) dynamic point cloud dataset [13] is used for subjective evaluation. This dataset is composed of four voxelized point clouds with $1024 \times 1024 \times 1024$ (RGB) points. Each point cloud comes with 300 frames, resulting in ten seconds of video at $30\,\mathrm{FPS}$. In our setup, half of these are used for evaluation, resulting in a 5-second video that plays out on repeat.

Because the Unity framework requires significant computational resources to render multiple point clouds with 1M points per frame, the objects are first converted to meshes. MeshLab is used to generate filter scripts, which are used with the PyMeshLab library to convert the point clouds into meshes using the *Poisson surface reconstruction algorithm*. Different quality representations are obtained through different values for the depth parameter d, an integer that defines the maximum depth of the tree that will be used for surface reconstruction. Running at depth d corresponds to solving on a voxel grid whose resolution is no larger than $2^d \times 2^d \times 2^d$. Aside from the default value of 8, values of 6 and 7 are used as well (see Figure 3). When the quality is fixed (test 1, 2 D and 3), all frames are rendered at the same quality; when two quality representations are considered (test 2 A, B), 75 consecutive frames are loaded of each quality; and when three representations are considered (test 2 C), 50 consecutive frames are loaded of each quality.

A gaming laptop with an Intel core i7 processor, 16GB of RAM and a dedicated Nvidia Geforce RTX 2070 are used for the rendering of the scenes. The Meta Quest 2 is chosen as HMD, because of its ease of use and high resolution. The HMD is connected to the laptop by means of a cable link (USB-C to USB-C, $5\,\mathrm{Gb/s}$). This way, the content can be rendered on the laptop and then streamed to the HMD, allowing a stable frame rate of $90\,\mathrm{FPS}$. The HMD's associated controllers are used to operate the interactive sliders.

The full source code of the setup is available on Github [14], along with instructions on how to deploy the implementation.

V. Results

This section presents the results of the subjective study. First, a description of the participants of the study is provided.

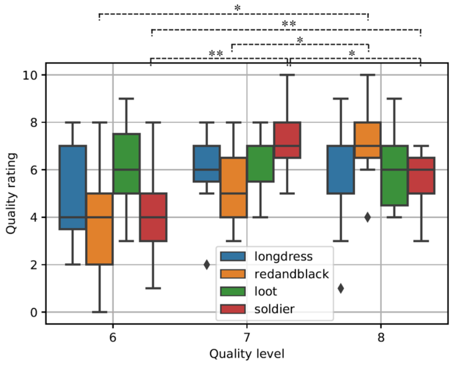

Fig. 4: Comparison of quality ratings for different quality levels of the point cloud objects in test 1. Significant differences are indicated by * (p<0.05), ** (p<0.01) and *** (p<0.001). Note that only the scores of objects A, B and C are considered for fair comparison.

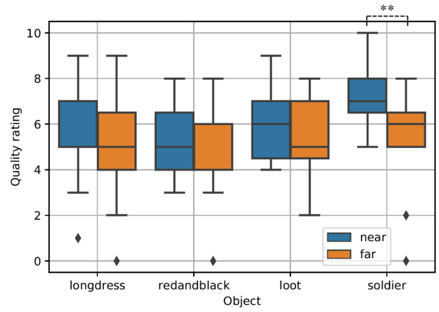

Fig. 5: Comparison of quality ratings for near and far objects of the same quality level in test 1.Significant differences are indicated by * (p<0.05), ** (p<0.01) and *** (p<0.001).

TABLE I: Relative frequency of each near object to be matched correctly (*i.e.*, same quality level) with its far counterpart in test 1.

Object	Relative frequency
longdress	80.0%
loot	53.3%
redandblack	80.0%
soldier	73.3%

In the remaining three sections, the three research questions in the form of the tests are answered. To conclude, the participants' perception of the system is evaluated based on the post-session questionnaire.

A. Participants Description and Demographics

A total of 30 participants were gathered for testing. 24 of them (80%) identified as male, 5 (16.7%) as female and 1 (3.3%) as non-binary. 3 (10%) were between 18 and 24 years old, 21 (70%) between 25 and 34, 4 (13.3%) between 35 and 44 and 2 (6.7%) between 45 and 54. All users were tested for correct color vision using *Ishihara Tests* and were instructed to optimize their visual capabilities for the test w.r.t. other, self-reported eyesight issues (*e.g.*, keep glasses on in case of farsightedness). To avoid user fatigue, the participants were randomly split in two groups of 15 participants each. The first group was presented with *longdress* and *redandblack* in Test 1 and *soldier* and *loot* in Test 2. The opposite holds for the second group. The seven one-to-one comparisons of Test 3 were the same for both groups.

B. Test 1: Quality vs Distance

The first purpose of the study was to understand if and how quality levels and distance affect the perception of dynamic meshes.Figure 4 shows the boxplots for the quality ratings given to each quality representation of each point cloud figure. Based on visual inspection, there can be noticed that the quality ratings of both *soldier* and *redandblack* tend to vary depending on their quality representation while *longdress* and *loot*, in contrary, show more or less consistent behaviour over the different qualities. To statistically validate this claim, to avoid accumulating Type-I errors due to pair-wise comparison (as we have more than two groups) and to distinguish it from sampling coincidence, a *Kruskal-Wallis Rank Test* for ordinal data was performed. It confirms that no significant (p>0.05) differences in quality rating exist for *longdress* (p=0.21) and *loot* (p=0.88). For *redandblack* and *soldier* (both p<0.001), post-hoc pairwise *Wilcoxon Signed Rank Tests* with *Bonferroni correction* were performed to further investigate pair-wise differences between quality levels not resulting from pure chance. These show significant differences between each pair of quality levels for *soldier* (p<0.01 for 6-7 and 6-8 and p<0.05 for 7-8) as well as significant differences between quality levels 6-8 and 7-8 of *redandblack* (both p<0.05). This points at the assumption that the sensitivity of the human perception towards the underlying quality representation is heavily entangled with the specific object and its geometrical properties under scrutiny. For *soldier*, the quality level is highly pronounced in the fine-grained end of the weapon, which can even appear floating when heavily compressed. A similar observation can be made for the hair of *redandblack*. For the other two objects, differences between quality levels are often more nuanced (*e.g.* the fingers of *loot*), making them less susceptible to human perception.

Figure 5 studies the influence of the distance of an object by comparing it with quality ratings of the objects in their near and far representation at the same quality level. At first, there seems to be no pronounced difference between the quality ratings for near and far representations apart from *soldier*. Remarkable enough, subjects seem to give lower ratings to the more distant object in this case than to its nearby counterpart. To further validate this observation, we applied Wilcoxon Signed Rank Tests on each near-far object pair. The analysis reveals that there only exists a significant difference in perception for *soldier* (p<0.01). This is counter-intuitive, as it would be expected that an increased distance between the object and the subject to make it more difficult to spot any impairments in the geometry of the figure and would thus result in higher quality ratings.

Finally, Table I shows the relative frequency of users matching the correct quality level of the 'near' objects with the quality level of the 'far' object. More or less the same

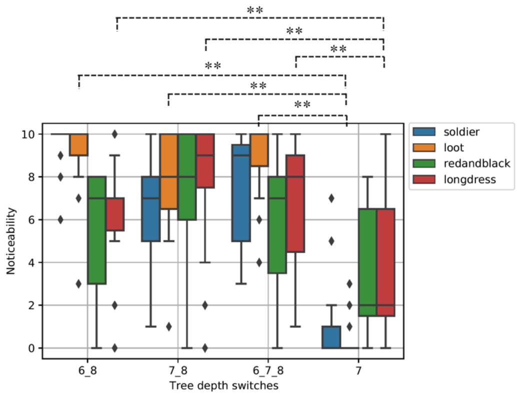

Fig. 6: Noticeability (with 0 being unnoticeable and 10 extremely noticeable) of the respective quality changes for each of the point cloud models in test 2. Significant differences are indicated by * ($p<0.05$), ** ($p<0.01$) and *** ($p<0.001$).

success ratios can be observed for *longdress*, *redandblack* and *soldier* (80.0, 80.0 and 73.3, respectively), while for *loot* a percentage of only 53.3 is obtained. This once again hints at the dependence of subjective perception on the particular content at hand. Where we previously noticed that the quality levels of both *longdress* and *loot* were difficult to assess, the addition of a distant representation seems to be easing this task for *longdress*. For *loot*, however, this is still not the case. This could indicate that *loot* has the most robust geometry when it comes to quality degradation artifacts.

In the post-session qualitative evaluation, 45% of participants agreed that they can clearly tell the differences between the quality representations, while 3% strongly agreed. 23% remained neutral, and 29% disagreed. This illustrates that the quality level does not impact all users in the same way.

C. Test 2: Dynamic changes of the quality

This evaluation had the purpose of understanding the effects of dynamic adjustments of quality on the perception.

First, Figure 6 shows the level of noticeability for four quality switches over time, for each of the four point cloud objects (with 0 unnoticeable and 10 extremely noticeable). The objects switch between quality levels 6 and 8 (6_8), 7 and 8 (7_8), all three quality levels (6_7_8) or show no quality switches at all (fixed at level 7). Note that these representations were distributed at random over the four spots in the virtual environment. As one could expect, the noticeability scores for the non-changing figures are clearly lower than is the case for the other representations. Remarkable, however, is the fact that the scores for *soldier* and *loot* are much lower in both spread and value than is the case for the other figures in quality level 7. Furthermore, these objects also show a clear difference in quality scores compared to the quality-changing representations, while for *redandblack* and *longdress* this difference is far less pronounced. To quantify this, Kruskal-Wallis Rank Tests are performed. These show significant differences for each of the objects ($p<0.05$ for *redandblack*, $p<0.01$ for *longdress* and $p<0.001$ for *loot* and

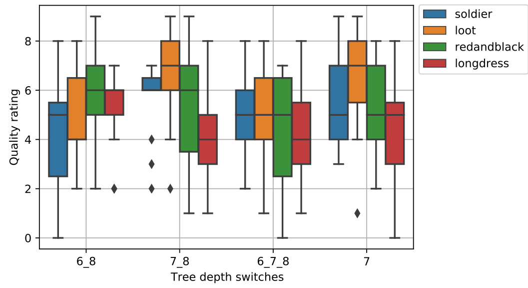

Fig. 7: Quality rating of the respective quality changes for each of the point cloud models in test 2.

soldier). Post-hoc pairwise Wilcoxon Signed Rank Tests with Bonferroni correction show significant pairwise differences between the non-changing configuration (7) and the three configurations switching in tree depth ($p<0.01$). However, this conclusion can only be made for *soldier* and *loot* as was already suspected from visual analysis. It has to be mentioned, though, that small but non-significant p-values are observed between the *7_8* and *7* configurations of *longdress* ($p=0.056$) and between *6_7_8* and *7* of *redandblack* ($p=0.050$). This shows that for all objects, dynamic changes in quality still have a high chance of being noticed by an end-user.

Figure 7 shows the ratings of the perceived quality ratings for the same tree depth switches as in Figure 6. As such, we can investigate whether the noticeability of a change in quality also has its influence on the subjective quality perception of an end-user. At first sight, no specific differences in quality ratings can be noticed between the multiple representations for any of the figures, although some varying behavior can be observed for *loot*. To further investigate this, a Kruskal-Wallis Rank Test was once again conducted. This confirmed that no significant differences between the quality representations for any of the objects can be found.

These results imply that while end users are aware of the quality switches, the influence on their quality perception shows to be limited. As such, this can potentially open the door to translating the video adaptive streaming paradigm to a 6DoF point cloud counterpart. Still, the post-session qualitative evaluation showed that 45% of participants disagreed that they are not bothered by the change in quality, while 13% disagreed strongly. 16% remained neutral, 23% agreed and 3% agreed strongly. Since the majority of participants is bothered by quality switches, they should be avoided where possible.

D. Test 3: Content dependency

Finally, the third test aims at confirming the hypothesis that the quality perception is dependent on the object at hand. Already, in test 1 and 2 we have seen the deep variability of assessment dependent on the object. Here, we went one step further, where the participant was exposed to two different objects at different quality levels. Then, the test participant was asked to indicate the object they perceived to have the highest quality, or 'equal' if no difference could be spotted.

The results of the comparisons provided to the subjects are shown in Table II. these pairs are chosen such that every

TABLE II: Overview of the six comparisons provided to the subjects in test 3. "Frac." indicates the fraction of users correctly indicating the object with the highest quality level as most visually appealing (or "equal" if applicable).

Scene	Obj. 1	Qual. 1	Obj. 2	Qual. 2	Frac.
1	soldier	7	longdress	6	70.0%
2	loot	8	redandblack	7	50%
3	soldier	7	redandblack	7	43.3%
4	loot	6	longdress	6	66.7%
5	longdress	6	redandblack	8	83.3%
6	loot	7	soldier	6	43.3%

figure was paired once with every other. Quality presentations were chosen to represent a balanced mix of combinations. It is interesting to see that, on the one hand, comparisons against *longdress* (scenes 1, 4 and 5) show to be rather straightforward (70%, 66.7% and 83.3%, respectively). The mutual comparisons of *soldier*, *loot* and *redandblack*, on the other hand, show to be a lot less pronounced (50%, 43.3%, and 43.3%). This might be explained by the fact that *longdress* tends to suffer from a "balloon head" when heavily compressed, which could provide a straightforward indication of its lower quality when compared to another object of higher quality level. Distortion artifacts in other objects are often more similar, making it difficult to distinguish between quality representations. This again highlights the impact of the selected content in the perception of point cloud meshed quality. This can also be deferred from the participants' reactions to the statement "I can clearly tell the quality difference between two objects": 6% disagreed strongly, 29% disagreed, 29% remained neutral, 32% agreed and 3% agreed strongly.

E. Post-session questionnaires: Overall perception

To evaluate the representativeness of the setup for immersive multimedia, the post-session questionnaire polled users about immersiveness, visual quality and wearing comfort on a 5-point Likert scale. On the question "I felt I was part of the virtual environment", 19 out of 30 users (63.3%) agreed or strongly agreed, therefore showing the immersive potential of the system. Only 6 out of 30 participants (20%) disagreed with the statement. When asked whether "the quality of the image of the HMD was optimal", 13 (43.3%) participants agreed while 11 (36.7%) remained neutral. This shows that the current visual quality HMDs is sufficient to deliver such immersive experiences, but that there is still room for further improvement. To the question "Was the HMD comfortable?" 21 (70%) users (strongly) agreed, while only 3 (10%) disagreed. As such, wearability of current HMDs does not seem to have a notable impact on the experience.

VI. Conclusion

This paper has presented a novel subjective evaluation methodology for point cloud content in a 6DoF environment. It allows users to rate the content in real-time while in the immersive environment, providing a more accurate assessment and decreased bias compared to post-study questionnaires. This methodology was used to perform an extensive subjective study on the effects of point-cloud-to-mesh conversion with different quality representations. Our results shows that both compression level and observation distance have their influence on subjective perception, but that the degree to which they play a role is heavily entangled with the content and its geometry at hand. In addition, it was observed that while end-users are clearly aware of quality switches, the influence on their quality perception is limited. As such, this has the potential to open up possibilities in bringing the video adaptive streaming paradigm to the 6DoF point cloud environment.

Acknowledgments

Sam Van Damme and Jeroen van der Hooft are funded by the Research Foundation Flanders (FWO), grant numbers 1SB1822N and 1281021N respectively. Part of this research was funded by the ICON project INTERACT, realized in collaboration with imec, with project support from Flanders Innovation and Entrepreneurship (VLAIO). Project partners are imec, Rhinox, Pharrowtech, Dekimo and TEO. This research is partially funded by the FWO WaveVR project (Grant number: G034322N). This research is also partially funded by the European Union (SPIRIT project, Grant Agreement 101070672). The authors also want to acknowledge the contributions of Gilles Christiaens, Júnior Ednailson, Gwendolyn Heyninck, Harshit Kaushik, Wout Meert, Ayoub Sassi, Gianni Vertriest and Daniel Wieczorek, who have assisted in the development of the framework discussed in this paper.

References

[1] A. Aaron *et al.*, "Challenges in cloud based ingest and encoding for high quality streaming media," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 1732–1736.
[2] J. van der Hooft *et al.*, "Towards 6DoF HTTP Adaptive Streaming Through Point Cloud Compression," in *ACM International Conference on Multimedia*, 2019.
[3] E. Zerman *et al.*, "Textured Mesh vs Coloured Point Cloud: a Subjective Study for Volumetric Video Compression," in *International Conference on Quality of Multimedia Experience*, 2020.
[4] E. Zerman *et al.*, "User Behaviour Analysis of Volumetric Video in Augmented Reality," in *International Conference on Quality of Multimedia Experience*, 2021.
[5] International Telecommunication Union, "ITU-T Rec. P. 919(10/2020) Subjective test methodologies for 360º video on head-mounted displays."
[6] P. Astola *et al.*, "JPEG Pleno: Standardizing a Coding Framework and Tools for Plenoptic Imaging Modalities," *ITU Journal: ICT Discoveries*, vol. 3, 2020.
[7] L. A. da Silva Cruz *et al.*, "Point Cloud Quality Evaluation: Towards a Definition for Test Conditions," in *International Conference on Quality of Multimedia Experience*, 2019.
[8] J. van der Hooft *et al.*, "Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming," in *International Conference on Quality of Multimedia Experience*, 2020.
[9] E. Dumic and L. A. da Silva Cruz, "Point Cloud Coding Solutions, Subjective Assessment and Objective Measures: A Case Study," *Symmetry*, vol. 12, no. 12, 2020.
[10] X. Wu *et al.*, "Subjective quality database and objective study of compressed point clouds with 6dof head-mounted display," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 12, pp. 4630–4644, 2021.
[11] S. Subramanyam *et al.*, "Subjective QoE Evaluation of User-Centered Adaptive Streaming of Dynamic Point Clouds," in *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*, 2022.
[12] A. Joshi *et al.*, "Likert scale: Explored and explained," *British journal of applied science & technology*, vol. 7, no. 4, p. 396, 2015.
[13] E. d'Eon *et al.*, "8i Voxelized Full Bodies - A Voxelized Point Cloud Dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006*, 2017.
[14] I. Mahdi, "Subjective tests," https://github.com/MahdiImen/SubjectiveTests, 2022.