IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 28 August 2023, accepted 11 October 2023, date of publication 20 October 2023, date of current version 21 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3326374

RESEARCH ARTICLE

Characterization of the Quality of Experience and Immersion of Point Cloud Videos in Augmented Reality Through a Subjective Study

MINH NGUYEN[1], SHIVI VATS[1], SAM VAN DAMME[2,3], JEROEN VAN DER HOOFT[2], (Member, IEEE), MARIA TORRES VEGA[2,3], (Senior Member, IEEE), TIM WAUTERS[2], (Member, IEEE), FILIP DE TURCK[2], (Fellow, IEEE), CHRISTIAN TIMMERER[1], (Senior Member, IEEE), AND HERMANN HELLWAGNER[1], (Senior Member, IEEE)

[1]Department of Information Technology, Alpen-Adria-Universität (AAU) Klagenfurt, 9020 Klagenfurt, Austria
[2]IDLab, Department of Information Technology, Ghent University-imec, 9025 Ghent, Belgium
[3]eMedia Research Laboratory, Department of Electrical Engineering (ESAT), KU Leuven, 3000 Leuven, Belgium

Corresponding author: Minh Nguyen (minh.nguyen@aau.at)

This work was supported by the European Union (Scalable Platform for Innovations on Real-Time Immersive Telepresence Project, https://www.spirit-project.eu/) under Grant 101070672. The work of Sam Van Damme and Jeroen van der Hooft was supported by the Research Foundation Flanders (FWO) under Grant 1SB1822N and Grant 1281021N, respectively.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Point cloud streaming has recently attracted research attention as it has the potential to provide six degrees of freedom movement, which is essential for truly immersive media. The transmission of point clouds requires high-bandwidth connections, and adaptive streaming is a promising solution to cope with fluctuating bandwidth conditions. Thus, understanding the impact of different factors in adaptive streaming on the Quality of Experience (QoE) becomes fundamental. Point clouds have been evaluated in Virtual Reality (VR), where viewers are completely immersed in a virtual environment. Augmented Reality (AR) is a novel technology and has recently become popular, yet quality evaluations of point clouds in AR environments are still limited to static images. In this paper, we perform a subjective study of four impact factors on the QoE of point cloud video sequences in AR conditions, including encoding parameters (quantization parameters, QPs), quality switches, viewing distance, and content characteristics. The experimental results show that these factors significantly impact the QoE. The QoE decreases if the sequence is encoded at high QPs and/or switches to lower quality and/or is viewed at a shorter distance, and vice versa. Additionally, the results indicate that the end user is not able to distinguish the quality differences between two quality levels at a specific (high) viewing distance. An intermediate-quality point cloud encoded at geometry QP (G-QP) 24 and texture QP (T-QP) 32 and viewed at 2.5 m can have a QoE (*i.e.*, score 6.5 out of 10) comparable to a high-quality point cloud encoded at 16 and 22 for G-QP and T-QP, respectively, and viewed at a distance of 5 m. Regarding content characteristics, objects with lower contrast can yield better quality scores. Participants' responses reveal that the visual quality of point clouds has not yet reached an immersion level as desired. The average QoE of the highest visual quality is less than 8 out of 10. There is also a good correlation between objective metrics (*e.g.*, color Peak Signal-to-Noise Ratio (PSNR) and geometry PSNR) and the QoE score. Especially the Pearson correlation coefficients of color PSNR is 0.84. Finally, we found that machine learning models are able to accurately predict the QoE of point clouds in AR environments. The subjective test results and questionnaire responses are available on GitHub: https://github.com/minhkstn/QoE-and-Immersion-of-Dynamic-Point-Cloud.

The associate editor coordinating the review of this manuscript and approving it for publication was Chin-Feng Lai.

INDEX TERMS Point clouds, quality of experience, subjective tests, augmented reality.

I. INTRODUCTION

In recent years, immersive video delivery has improved significantly [46]. This type of video content can be viewed in Extended Reality (XR), including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), to provide near-lifelike three-dimensional (3D) objects and scenes. To represent these 3D entities, Point Clouds (PCs) are commonly used as the format providing high-fidelity representations without any constraint on the viewpoint and interaction. A PC comprises thousands or even millions of points, including information about colors (*e.g.* RGB) and geometry (x,y,z coordinates) of each point. Thus, the usage of PCs costs a large amount of storage and network bandwidth. A single raw PC frame can reach hundreds of Mbit in size, which leads to several Gbit/s of bandwidth requirement for an uncompressed 30 frames per second (fps) video [20], [48]. One promising solution is adaptive streaming techniques (*i.e.*, HTTP Adaptive Streaming (HAS) [10]) combined with point cloud compression (PCC) [42].

In HAS, the content is encoded into various quality levels, then temporally split into multiple segments with the same duration. These segments are stored on one or multiple servers in a content delivery network [16]. The end users' media player selects the quality level for each segment based on the network conditions (*i.e.*, observed throughput), the devices' characteristics (*i.e.*, resolution and buffer), and the users' preferences [9]. HAS adapts to variations in the network conditions in order to prevent rebuffering events while providing the highest quality possible by changing the quality of the content, which leads to quality switches. The quality switches are well-known to affect the Quality of Experience (QoE) in traditional video streaming [29], [41]. However, they have not been fully considered in dynamic point cloud streaming.

Due to extremely high data volumes of uncompressed PCs, point cloud compression (PCC) is necessary to reduce both storage requirements and the amount of data delivered through networks. Initial studies started with compression of static 3D objects [17], [21], [27]. However, recent work focused on dynamic scenarios [35]. The Moving Picture Experts Group (MPEG) developed standardized solutions for point cloud compression by leveraging their existing codecs, such as High Efficiency Video Coding (HEVC) [39]. Their video point cloud encoder implementation can achieve a compression rate of 125:1. This means that a dynamic PC with 1 million points can be compressed at a bitrate of 8 Mbit/s [1], which is feasible for delivery over current networks. However, PCC comes at the cost of visual quality, determined by the quantization parameter (QP). A higher QP provides a lower bitrate but leads to a lower quality.

Therefore, understanding the impact of different factors in HAS and PCC on the QoE is of importance. PCs have been evaluated in different viewing conditions (*i.e.*, VR head-mounted displays (HMDs) and 2-dimensional (2D) screens) [6], [47]. However, research on the subjective quality assessment of PCs in AR environments is still limited. AR enhances people's perception of physical and virtual environments [9]. Thus, it is an interesting setting for immersive telepresence applications, which we develop and assess in the European project Scalable Platform for Innovations on Real-time Immersive Telepresence (SPIRIT) [2].

In this paper, we study the impact of various factors on the QoE for PC-based video streaming in AR environments, *i.e.*, *(i)* encoding parameters (*i.e.*, QPs), *(ii)* quality switches, *(iii)* viewing distance, and *(iv)* content characteristics. We also consider the immersion level of PC objects in the real environment. In addition, cybersickness, which is a common symptom in VR, is investigated with the user wearing an AR HMD. The contributions of this paper are fourfold:

- We provide quantitative results on the perceptual quality and the impacts of various factors on the QoE of the user, including encoding parameters (*i.e.*, QPs), quality switches, viewing distance, and content characteristics.
- We provide qualitative results regarding the presence of life-size digital humans as dynamic point clouds in the physical world and the cybersickness issue in AR environments.
- We evaluate the correlation between objective metrics and subjective quality for dynamic PCs.
- We assess state-of-the-art machine learning-based models in predicting the QoE for PC-based video streaming in AR environments.

The remainder of this paper is organized as follows. Section II provides an overview of existing literature related to PCC, HAS, subjective quality assessment, and cybersickness. Section III describes our subjective study, followed by Section IV containing the results and discussions. Section V describes the evaluations of the QoE prediction ability of common machine learning models. Finally, Section VI concludes this paper and suggests future work.

II. RELATED WORK

A. POINT CLOUD COMPRESSION

MPEG started working on point cloud compression (PCC) in 2014. After a call for proposals in 2017, three technologies were selected, including LIDAR PCC (L-PCC) for dynamically acquired data, Surface PCC (S-PCC) for static content, and Video-based (V-PCC) for dynamic point clouds [35]. The final standard now comprises two technologies: *(1) Geometry-based PCC (G-PCC)*, which combines L-PCC and S-PCC because of their similarities [15], and *(2) Video-based PCC (V-PCC)*. G-PCC encodes the point cloud 3D positions directly to create the compressed point cloud. Tiles and slices are introduced to encode parts of point clouds independently. A slice is a group of points that can be independently encoded, and a tile consists of multiple slices. However, the current version of G-PCC only supports intra prediction. Motion estimation and inter prediction will be considered in the next version of the standard [20]. V-PCC, on the other hand, projects the 3D points onto 2D images and then uses traditional encoders, *e.g.*, H.265/HEVC, to encode these images and thus can benefit from the encoders' efficient coding and simplify deployment.

B. HTTP ADAPTIVE STREAMING FOR POINT CLOUDS

Hosseini and Timmerer were among the first to investigate dynamic adaptive point cloud streaming techniques that extend the concepts of Dynamic Adaptive Streaming over HTTP (DASH) [38]. However, instead of a dedicated encoder, they used sampling to reduce the data usage and create different quality levels. They also proposed to stream PCs on a per-frame basis, leading to a significant overhead of 30 GET requests per second (for a 30 fps sequence). In a later work [48], van der Hooft et al. used MPEG's reference encoder to generate quality levels and proposed different heuristic rate adaptation techniques for dynamic point cloud objects. They also introduced a DASH-compliant framework for point cloud streaming. In a fairly recent work, Wang et al. proposed a QoE-optimized rate adaptation algorithm for point cloud transmission [52]. The authors leverage the 3D tiling technique [30] to divide the point cloud into small cubes, *i.e.*, 3D tiles. Each tile is encoded and decoded independently. Thus, only the tiles in the user's viewport are requested, and their bitrates are allocated separately under a given network bandwidth. This technique can significantly reduce the amount of data transferred to the end user.

C. SUBJECTIVE QUALITY ASSESSMENT

Wu et al. [56] evaluated the quality of PCs with different quality levels. However, quality switching in the test sequences was not considered, and the device used in the subjective test was a VR device (*i.e.*, HTC VIVE[1]) rather than an AR device. The work in [47] considered the quality switching of PC videos in the context of HAS. It was found that the texture of PC objects is an essential factor in determining the QoE and that content with fewer contrast differences can provide higher QoE. However, the PCs were displayed on a 2D screen that cannot offer a truly immersive experience. The work in [12] considered different quality levels and viewing distances while watching two 3D representation formats — PCs and meshes. The experiments were conducted on a flat screen, and the results showed that a closer viewing distance led to lower QoE for a given quality level. In a similar work, Van Damme et al. also used 3D objects in the two different formats but showed the objects in a VR environment [44]. These experiments also found that a closer viewing distance leads to lower QoE.

D. CYBERSICKNESS

Cybersickness can be defined as the feeling of dizziness, nausea, or headache when people are watching content on 2D screens or through XR devices. It has been commonly investigated in XR environments, especially VR [13], [34], [36], [43]. Tran et al. [43] considered cybersickness a factor contributing to QoE in a subjective test with different 360° videos. They found that cybersickness is a critical issue in VR, especially in videos with fast camera motion. Caserman et al. [13] compared the impact of different HMDs, including HTC VIVE, Oculus Rift DK1, and DK2,[2] on cybersickness. The results showed that HTC VIVE mitigates the level of nausea symptoms thanks to its accurate positional tracking. However, the studies of cybersickness in AR environments are still limited. Recently, Kirollos et al. [25] compared the cybersickness in VR and AR HMDs by rendering the entire scene of the physical environment viewed in a Microsoft HoloLens 2[3] (AR HMD) into an Oculus Rift S[4] (VR HMD). VR was found to cause more sickness than AR due to more virtually rendered elements. However, a limitation of this work is that the objects used in the evaluation are static furniture inside a room.

In this paper, we conduct a subjective test in which participants wear the Microsoft HoloLens 2 to watch and rate dynamic point clouds. We ask the participants about their cybersickness symptoms at the end of the subjective test to understand the level of cybersickness in an AR environment.

III. SUBJECTIVE TEST FOR POINT CLOUD ASSESSMENT

This section gives an overview of the influence factors considered, followed by a description of our subjective test methodology, including dataset, equipment, environment, and subjects' tasks.

A. INFLUENCE FACTORS

The focus of this study is on four main influence factors on the QoE while watching PC videos in AR environments.

Encoding parameters: QoE is clearly affected by encoding parameters, most notably by the QPs. These parameters reduce the amount of data in the video at the cost of distorting the perceptual quality. As there are two attributes (*i.e.*, geometry and texture) in a PC, a pair of QPs, namely geometry QP (G-QP) and texture QP (T-QP), are used in the encoding. A higher G-QP causes points to deviate more from their original position. Similarly, when T-QP increases, some color information is combined [56].

Quality switches: In the context of HAS, the video quality can be changed due to fluctuation of the throughput [29], [41] to minimize content rebuffering. This change in video quality is referred to as a quality switch. Quality switches can be classified as *switching up*, when quality increases and *switching down* when quality decreases.

Viewing distance: As six degrees of freedom (6DoF) interaction allows end users to move freely in their space, the viewing distance from end users to the object can vary

[1]https://www.vive.com/. Accessed 16 August 2023.

[2]https://developer.oculus.com/blog/open-source-release-of-rift-dk2/. Accessed 16 August 2023.

[3]https://www.microsoft.com/en-us/hololens. Accessed 16 August 2023.

[4]https://www.oculus.com/rift-s/. Accessed 16 August 2023.

(a) Loot (b) RedAndBlack (c) LongDress (d) Soldier

FIGURE 1. Tested objects in 8i Voxelized full bodies database [19].

FIGURE 2. Raw points (left) and square shader (right) representations of ***Loot*** at 2.5 m.

depending on their movement. In this work, we examine the impact of viewing distance on the user's QoE.

Content characteristics: The perspective of viewers can vary depending on the content [43], [47]. In this paper, four videos with different characteristics are used.

B. DATASET PREPARATION

As this work focuses on the usage of PCs in telepresence applications, we used four PC objects from the 8i Voxelized Full Bodies Database [19]: *Loot*, *RedAndBlack*, *LongDress*, and *Soldier* as shown in Fig. 1. The first two have a lower contrast content than the other two [47]. Each sequence captures a complete object using 42 RGB cameras operating at 30 fps for 10 seconds.

We use the MPEG V-PCC reference software Test Model Category 2 (TMC2) [4] to create compressed PCs by varying the quantization parameters (QPs). This software already includes five sets of QPs defined in MPEG's Common Test Conditions (CTC) [3] with the geometry QP (G-QP) and texture QP (T-QP) ranging from 16 to 32 and from 22 to 42, respectively. Three such pairs from the MPEG PCC software with the lowest, middle, and highest QPs (*i.e.*, the rates R5, R3, and R1 in MPEG's software [4]) are selected as follows:

- Q1 (R1): (G-QP, T-QP) = (32, 42)
- Q2 (R3): (G-QP, T-QP) = (24, 32)
- Q3 (R5): (G-QP, T-QP) = (16, 22)

TABLE 1. Bitrates in Mbit/s of different quality levels of the PC objects.

Video	Quality		
	Q1	Q2	Q3
Loot	2.28	5.63	16.68
LongDress	4.64	14.05	46.78
RedAndBlack	3.39	7.55	22.90
Soldier	4.38	11.58	35.29

TABLE 2. Notation and description of the test sequences.

Task	Notation	Description	Sequences
1	Qij	The video starts with quality Qi, then switches to Qj after 5 s. $i, j \in \{1, 2, 3\}$.	Loot and LongDress
2	Qi_Dk	The video is watched at quality Qi at distance Dk. $i, k \in \{1, 2, 3\}$.	RedAndBlack and Soldier

$Q3$ is thus the best quality level. The bitrate of the objects decreases with increasing QPs (see Table 1). For instance, *Loot*'s bitrates are 2.3 Mbit/s, 5.6 Mbit/s and 16.7 Mbit/s for $Q1$, $Q2$, and $Q3$, respectively.

We developed a Unity project [49] using the Pcx Point Cloud importer [40] and the PointXR square shader [6] to import and render PCs, respectively. Fig. 3 illustrates the architecture of the platform. Our platform consists of a workstation, a Wi-Fi router, and HoloLens 2. A workstation running on Windows 10 with an Intel Core i9-13900K processor, 64 GB memory, and an NVIDIA RTX 4070 Ti GPU stores the compressed point cloud frames in the Unity project. A *Sequence Configuration* is created to list all the point cloud objects with their configurations (*i.e.*, quality levels and viewing distance). Then, the Unity software in the workstation generates the test sequences via *Test Sequences Generation* and sends them to HoloLens 2 through a Wi-Fi router using Holographic Remoting.[5] The participants watch and rate the visual quality of the test sequences through interactions with the HoloLens 2. We collect the rating scores and store them at the workstation.

Fig. 2 compares a PC in raw points and square shader representations. Pre-tests conducted by comparing the subjective visual quality of the point and square shaders showed that the square shader of [6] is visually superior to raw PCs; therefore, this format is used in our test. The C# scripts in [50] are used to control the quality and distance to design the sequences mentioned in the sequel. These scripts are part of the testing platform, which is available on GitHub [49].

C. EQUIPMENT AND ENVIRONMENT

The test subjects use Microsoft HoloLens 2 to interact with our experiments. HoloLens 2 includes two displays with 2K resolution and a diagonal field of view (FoV) of 52°

[5]https://learn.microsoft.com/en-us/windows/mixed-reality/develop/native/holographic-remoting-overview. Accessed: 20 August 2023.

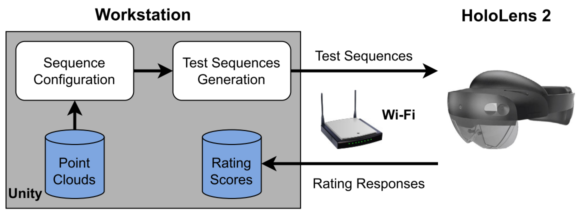


FIGURE 3. Platform architecture.

(43° horizontal and 29° vertical) [33], [51]. The HoloLens 2 allows users to see and interact (*e.g.*, by scaling, rotating, and moving) with virtual objects while being visualized in the real world. However, in our experiments, we only allowed users to view the PC objects, and the interaction was only with the feedback user interface (UI) to rate the visual quality of the test sequences.

Following the recommendations of ITU-R BT.500-15 [22], our experiments are carried out in a room with grey walls and low illumination. The tested PCs are placed in the room and scaled to be life-sized (1.8 m height) to simulate realistic telepresence scenarios.

D. EXPERIMENT TASKS

We design two tasks for each participant, with both tasks consisting of 18 sequences of length 10 s. Table 2 describes the sequences. Before the experiment, participants are asked to provide some background information, including age, gender, eyesight, and experience in viewing VR, AR, and MR content.

1) TASK 1: IMPACT OF VIDEO ENCODING AND QUALITY SWITCHES

The participant watches nine sequences for each of the two objects, including three sequences with static quality (Q1, Q2, and Q3) and six sequences with a quality switch in the middle of each sequence. The objects are placed 5 m from the participant so that the whole body can be viewed. *Loot* and *LongDress* are used in this task as they belong to low and high contrast levels, respectively.

2) TASK 2: IMPACT OF VIEWING DISTANCE

The participant watches static-quality sequences of the other two objects at quality levels $Q1$, $Q2$, and $Q3$ at three distances:

- D1: 1.25 m (only face and shoulder in FoV)
- D2: 2.5 m (only upper body in FoV)
- D3: 5 m (full body in FoV)

The order of tasks and sequences for each task are randomized for each participant. After watching each sequence, the participant is asked to rate the perceptual quality (*i.e.*, 1, 2 – very bad, 3, 4 – bad, 5, 6 – fair, 7, 8 – good, 9, 10 – very good) through the immersive slider shown in Fig. 4.

After the experiment, participants are asked to provide feedback on their experience regarding levels of general discomfort, nausea, sweating, headache, or dizziness that

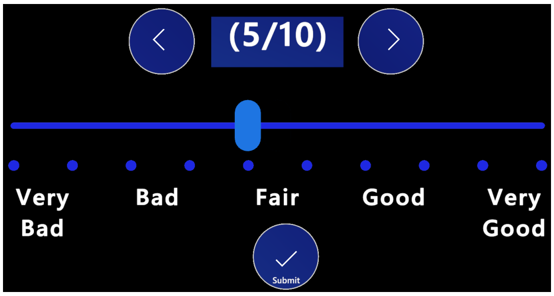


FIGURE 4. Immersive rating slider within the user interface of the HoloLens 2 as used during the subjective tests.

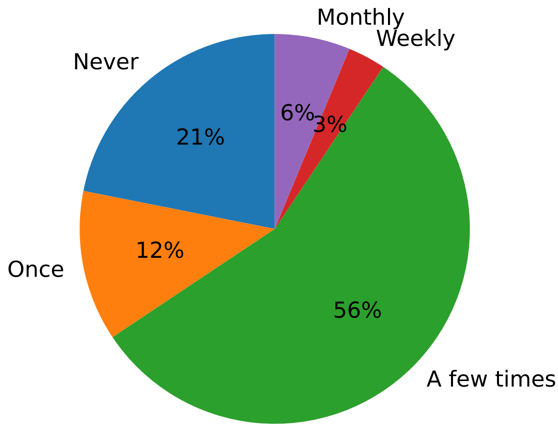


FIGURE 5. Frequency of participants watching XR contents.

they may have experienced. Participants also answer a question of whether they feel the PC objects are part of the real environment by selecting one of five options: *(i)* strongly disagree, *(ii)* disagree, *(iv)* neutral, *(i)* agree, and *(v)* strongly agree. The total duration of a single experiment is approximately 25 minutes.

IV. RESULTS AND DISCUSSIONS

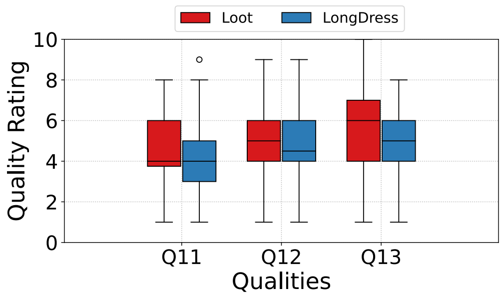
A. PARTICIPANTS

A total of 36 participants, who were recruited from AAU Klagenfurt, attended the subjective test, including 22 (61%) males, 13 (36%) females, and 1 (3%) non-binary. 3 (8%) were in the age group of 18 to 24 years, 18 (50%) were between 25 and 34, 12 (33%) between 35 and 44, 2 (6%) between 45 and 54, and 1 (3%) between 55 and 64. The color vision of the participants is evaluated using the Ishihara test [11]. Four participants failed this test, so their ratings are excluded. Hence, the results in this section are gathered from 32 participants which is compliant with ITU-R BT. 500 [22].

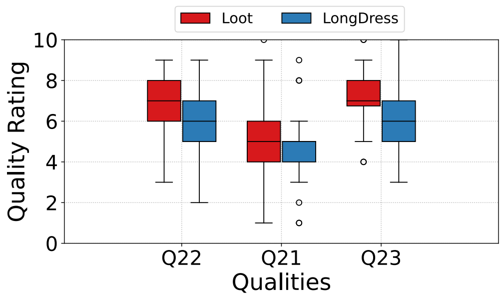
Fig. 5 shows how often the participants experienced XR content. It is clearly seen that most of them have experienced XR before this subjective test. Only 21% of them have never watched XR content.

B. TASK 1 – IMPACT OF VIDEO ENCODING

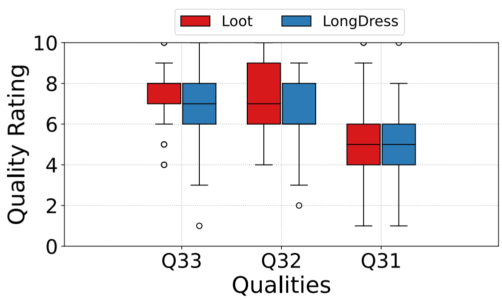
Fig. 6 shows the quality ratings of the participants for the test sequences at different quality levels and quality switches. Regarding the sequences with *static* quality levels (*i.e.*, $Q11$, $Q22$, $Q33$ in Fig. 6(a)-(c)), it can be seen that objects encoded



(a) Quality $Q1$ and switching from $Q1$.



(b) Quality $Q2$ and switching from $Q2$.



(c) Quality $Q3$ and switching from $Q3$.

FIGURE 6. Quality ratings for different quality levels and quality switches.

with lower qualities have lower scores. At least 75% of the viewers gave *Loot* (*LongDress*) a rating of 6 (5) or less for the lowest-quality sequence, $Q11$. Their medians are both 4, which means a bad experience. With a higher-quality sequence, $Q22$, the median quality scores improve remarkably to 7 (*i.e.*, good) and 6 (*i.e.*, fair) for *Loot* and *LongDress*, respectively. For the highest-quality sequence, $Q33$, there is an improvement in quality ratings, but it is less remarkable than for $Q22$. *Loot* still receives good ratings from participants, with a median of 8 (*i.e.*, good), while *Longdress* achieves ratings ranging from fair (median of 6) for $Q22$ to good (median of 7) for $Q33$. Furthermore, though $Q33$ can achieve very good ratings (9 or 10), the majority (at least 75%) of the participants rate this sequence at no more than 8. To statistically validate these claims, we used one-way analysis of variance (ANOVA) [37] and post-hoc comparison analysis using Tukey's honestly significant difference (HSD) test [5]. According to the ANOVA results shown in Table 3, there is a significant difference ($p < 0.001$) between the three quality levels. Post-hoc pairwise, Tukey's HSD (see Table 4) reveals that quality ratings do not differ significantly ($p > 0.05$) between $Q22$ and $Q33$ for *Loot*, but do for

TABLE 3. ANOVA results. $p < 0.001$ (***).

Sequences	Objects	f	p
Q11, Q22, Q33	Loot	33.1863	***
	LongDress	20.5979	***
Q11, Q12, Q13	Loot	2.7075	0.0719
	LongDress	2.0969	0.1286
Q21, Q22, Q23	Loot	13.4655	***
	LongDress	7.9238	***
Q31, Q32, Q33	Loot	18.4664	***
	LongDress	9.6920	***

LongDress ($p < 0.05$). Furthermore, there are significant p-values ($p < 0.001$) between $Q11$ and the others.

C. TASK 1 – IMPACT OF QUALITY SWITCHES

Fig. 6 also describes the participant ratings for different quality switches, including *switching up* when the quality is increased and *switching down* when the quality is decreased. There is no remarkable improvement in the quality scores when the sequence starts at quality $Q1$ (*i.e.*, $Q11$, $Q12$, and $Q13$). ANOVA analysis indicates no significant difference ($p > 0.05$) among the quality scores for both *Loot* ($p = 0.07$) and *LongDress* ($p = 0.13$). This can be attributed to the severe distortion of $Q1$ in the initial 5 s that affects the QoE when watching the entire 10 s video.

Regarding *switching down*, when the quality changes from $Q2$ or $Q3$ to $Q1$ (*i.e.*, $Q21$ or $Q31$), the quality ratings are markedly reduced compared to the constant-quality sequences ($Q22$ and $Q33$). However, there are no significant differences when the quality changes between $Q2$ and $Q3$. We conducted a paired samples t-test [24] to further validate this observation. It shows non-significant p-values between $Q22$ and $Q23$ (*e.g.*, $p = 0.6136$ for *Loot*) as well as between $Q33$ and $Q32$ (*e.g.*, $p = 0.1162$ for *LongDress*). More details can be found in Table 5. Combined with the results in the previous section, we claim that the end user hardly recognizes the quality differences between $Q2$ and $Q3$. Thus, we recommend that it is unnecessary to change the quality from $Q2$ to $Q3$ when the object is viewed at a distance of 5 m. This can remarkably reduce the amount of transferred data.

D. TASK 2 – IMPACT OF VIEWING DISTANCE

Fig. 7 shows the quality ratings of the test sequences at different viewing distances. It is noticeable that distance significantly impacts the visual quality of the objects: the higher the viewing distance, the higher the quality scores. The reason is that, at a higher distance, it is harder for the viewers to recognize some quality distortions; thus, they give higher quality scores, which is comparable to what has been reported for traditional video sequences [7]. Additionally, we observe that to achieve the same visual quality, the object should be encoded with lower QPs (*i.e.*, more data) if viewed closer. For example, *RedAndBlack* at quality $Q1$ is rated on average

TABLE 4. Tukey's HSD results. $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***).

		Loot		LongDress	
Group 1	Group 2	q-value	p-value	q-value	p-value
Q11	Q22	8.644	***	5.592	***
Q11	Q33	10.918	***	8.988	***
Q22	Q33	2.275	0.247	3.395	*
Q11	Q12	2.099	0.304	1.636	0.483
Q11	Q13	3.244	0.062	2.888	0.108
Q12	Q13	1.145	0.682	1.251	0.639
Q21	Q22	6.015	***	4.413	**
Q21	Q23	6.649	***	5.234	***
Q22	Q23	0.633	0.888	0.821	0.812
Q31	Q32	7.124	***	3.874	*
Q31	Q33	7.726	***	6.158	***
Q32	Q33	0.602	0.900	2.285	0.244

TABLE 5. T-test results. $p < 0.01$ (**), $p < 0.001$ (***).

		Loot		LongDress	
Group 1	Groups 2	t-statistic	p-value	t-statistic	p-value
Q21	Q22	4.046	***	3.150	**
Q22	Q23	-0.507	0.6136	-0.5894	0.5576
Q23	Q21	-4.449	***	-3.6135	***
Q31	Q32	4.7038	***	2.8026	**
Q32	Q33	0.4710	0.6393	1.5934	0.1162
Q33	Q31	5.3661	***	4.3183	***

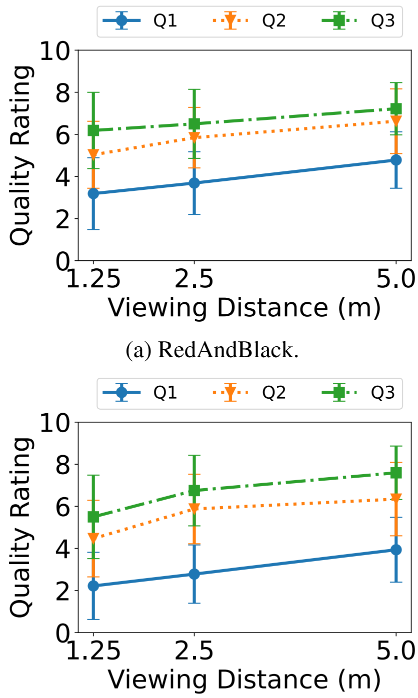
4.8 at 5 m, and this object must be encoded at $Q2$ to gain a similar score (*i.e.*, 4.9) if it is viewed at 1.25 m ($p = 0.5$ in a paired t-test).

E. IMPACT OF CONTENT CHARACTERISTICS

In this paper, we also evaluate the impact of content characteristics on the visual perception of participants for both tasks, as shown in Fig. 8. *Loot* and *RedAndBlack* achieve higher quality ratings in most cases. For example, the quality scores of *Loot* and *RedAndBlack* with quality $Q1$ viewed at distance $D3$ (*i.e.*, 5 m) are 4.5 and 4.8, respectively. Under the same conditions, these figures for *LongDress* and *Soldier* are 4.2 and 3.9, respectively. This can be explained by the fact that participants are less sensitive to quality distortion/changes for the content with fewer contrast differences. This finding extends the results presented in the work [47] on 2D screens to an AR environment with AR HMDs, in which the texture of the objects is a crucial factor for viewers.

F. CYBERSICKNESS IN AUGMENTED REALITY

The cybersickness levels of the participants are illustrated in Fig. 9. Fig. 9a shows that most of the participants did not feel symptoms of cybersickness in their experiment session that lasted about 25 minutes. 84% and 81% of



(a) RedAndBlack.

(b) Soldier.

FIGURE 7. Average quality ratings for different distances.

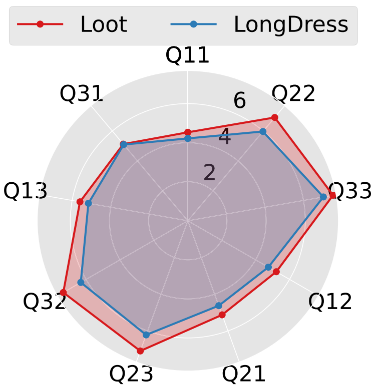
them did not sweat or feel nauseated, respectively. The most common symptom is dizziness, but only 21% of the participants reported feeling dizzy during the test. Fig. 9b provides more details about the symptoms of the participants who received cybersickness. No one suffers from all the symptoms mentioned above. There is only one person who experiences three symptoms, including sweating, headache, and dizziness. Three participants felt two symptoms, and six others received one symptom.

On the contrary, in a similar-duration subjective test [43] where participants were watching videos with four characters in a room and dolphins in the ocean with VR HMDs, cybersickness was a serious problem that affected more than 90% of the viewers.

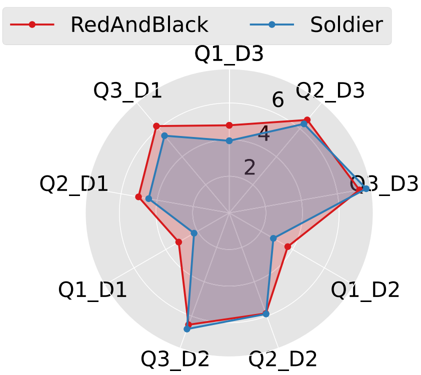
G. OBJECTS' IMMERSION LEVELS

Fig. 10 shows the immersion levels of digital objects in the physical world rated by the participants. It can be seen from the figure that 39% of the participants (strongly) agreed that the objects were part of the real environment. Only 27% of the participants (strongly) disagreed with this feeling. Therefore, the tested objects and HoloLens 2 provide the feeling of telepresence to some extent. However, some participants complained about the quality of some parts of the objects, even at the highest quality level. For example, the hair of *RedAndBlack* was perceived as blocky, and the heels of *LongDress* were missing in some frames (see Fig. 1).

When we consider the impact of the participants' frequency of watching XR contents on the immersion level rating of the tested objects, there are two findings to be

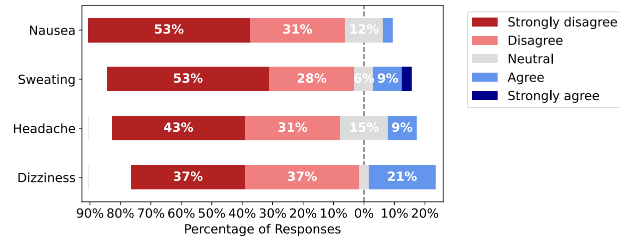


(a) Task 1.

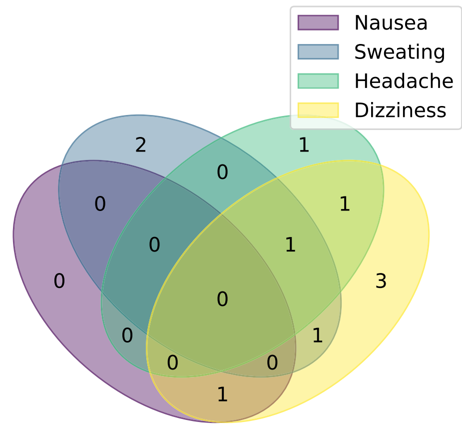


(b) Task 2.

FIGURE 8. Average quality ratings of participants. It should be noted that the sequence *Qii* in Task 1 is equivalent to *Qi_D3* ($i \in \{1, 2, 3\}$) in Task 2 as they are encoded at quality *Qi* and viewed at 5 m.

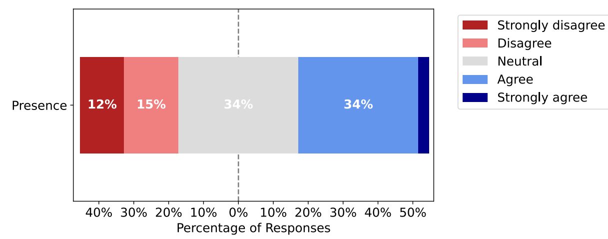


(a) Overall responses.

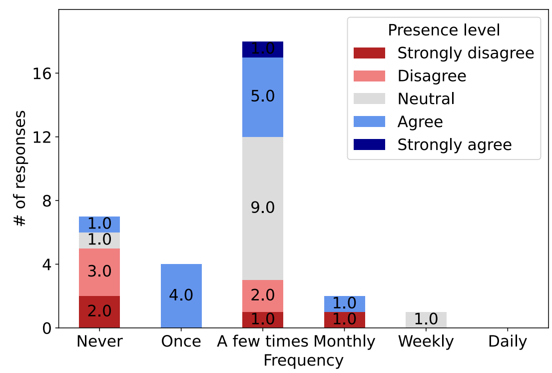


(b) Symptoms from the participants who got cybersickness.

FIGURE 9. Cybersickness levels of the participants.



(a) Level of feeling the presence of objects in the real environment.



(b) Frequency of watching XR contents vs. level of feeling the presence of objects.

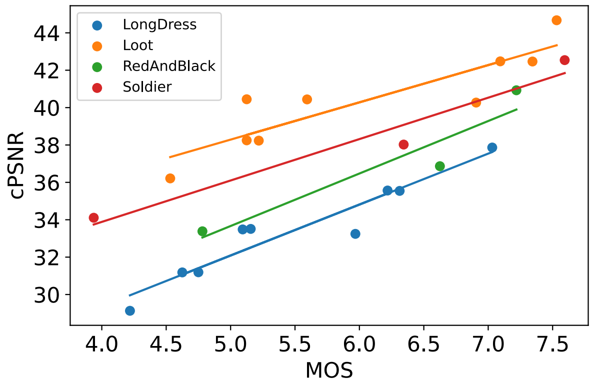
FIGURE 10. Objects' immersion levels.

noted. First, most participants (5 out of 7 people) who have never watched XR content do not feel that the test sequences are real. Second, most experienced participants felt neutral or agreed that the objects were real. These people may have a good understanding of how 3D objects look in such environments, and thus may have lower expectations in terms of feeling the presence of these digital objects.

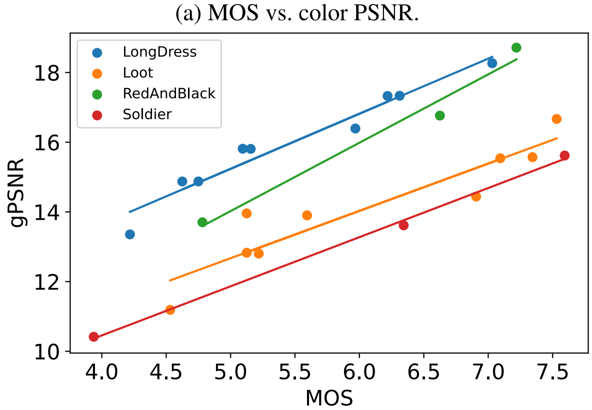
H. CORRELATION OF SUBJECTIVE AND OBJECTIVE METRICS

Fig. 11 shows the scatter plots of the participants' ratings represented by the mean opinion score (MOS) versus objective metrics, *i.e.*, geometry PSNR (gPSNR) and color PSNR (cPSNR) [23], [32]. Geometry PSNR is calculated from the distance between each point in the compressed point cloud and the corresponding point in the reference point cloud. Color PSNR is based on the error of color in YCbCr color space [14] of associated points between the compressed and reference point clouds. These metrics are computed using the software supplied by Working Group (WG) 11 of MPEG [28]. Each dot represents a test sequence viewed at 5 m. The lines are fitted using linear regression, minimizing the squared error. It can be clearly seen that the considered objective metric is linearly correlated with the subjective ratings.

The correlation coefficients calculated using the Spearman and Pearson methods for the MOS with the considered objective metrics are shown in Fig. 12. With a coefficient of more than 0.80 in both methods, color PSNR is highly correlated with the MOS. For geometry PSNR, this value is 0.62 and 0.69 for the Spearman and Pearson methods,



(a) MOS vs. color PSNR.



(b) MOS vs. geometry PSNR.

FIGURE 11. Correlation between subjective and objective metrics at 5 m viewing distance.

respectively, which means a moderate correlation with the MOS. We conclude that color PSNR achieves a better correlation with the MOS than geometry PSNR.

It should be noted from Fig. 11 that the correlation of color and geometry PSNRs is dependent on the content. That means two objects with similar PSNRs might have significant MOS differences. This motivates us to model the QoE with machine learning techniques that take into account the content characteristics.

V. EVALUATION OF MACHINE LEARNING-BASED QoE MODELS

Previous studies have found that machine learning methods can be used to effectively model the QoE of traditional, 360° and PC videos viewed on 2D screens or VR HMDs [8], [18], [26], [45]. In this work, we evaluate the performance of supervised machine learning techniques in predicting the QoE of PCs in the context of HAS in AR environments.

A. DATA PREPARATION

As described in Section III, we consider four influence factors, including encoding parameters, quality switching, viewing distance, and content characteristics. The first two factors are represented by the values of start and end QPs of sequences. The content characteristics can be represented by the bitrate of the encoded bitstream. As can be seen in Table 1,

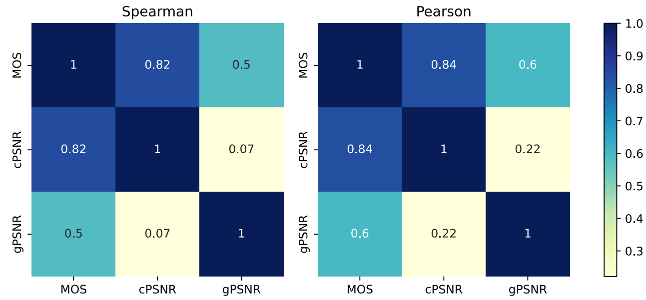


FIGURE 12. Spearman and Pearson correlation coefficients of the participants' ratings (MOS) and the considered objective metrics (cPSNR and gPSNR).

the objects have different bitrates even at the same quality level (same QPs).

For this evaluation, each input data record comprises six features: start G-QP, start T-QP, end G-QP, end T-QP, viewing distance, and bitrate. The corresponding ratings of the participants are used as the learning targets. We received 1152 responses in total from 32 participants. After omitting outliers defined by the interquartile range (IQR) method [54], 1107 responses are used as the input data.

To receive a reliable and unbiased estimate of model performance, we use leave-one-out cross-validation [55]. The input data is split into k groups, in which $k-1$ groups are used as the training dataset, and the remaining one as the testing dataset. The process of splitting the data is repeated k times so that every group is used as a testing dataset once. There are, in total, 36 test sequences; hence, we have $k = 36$ so that the ratings for 35 test sequences are used for training, and the others are for testing.

B. EVALUATION RESULTS

The work in [53] reported the results of five top-performance machine learning models in predicting the QoE of point clouds viewed on a 2D screen. Here, we will apply these models for QoE prediction in AR environments. Their performance is reported in Table 6. These models are implemented using the Python scikit-learn library [31], [53]. There are two classes of the considered models: *i) regression* (Gradient Boosting Regressor, Random Forest Regressor, Decision Tree Regressor, and Polynomial Regression), and *ii) classification* (Decision Tree Classifier). While the former directly predicts the MOS, the latter anticipates the probability of each class (*i.e.*, rating scores from 1 to 10) of the QoE distribution. The R2 score and the mean squared error (MSE) are calculated to evaluate the results. A better model should retrieve a higher R2 score and a lower MSE. It can be clearly seen that Gradient Boosting Regressor outperforms the others with an R2 score of 0.8582 and MSE of 0.2874.

Fig. 13 shows the correlation of the MOS (from our subjective test) with respect to the predicted MOS using the Gradient Boosting Regressor. The predicted MOS is highly correlated (Pearson correlation coefficient = 0.93) with the perceived MOS, which was gathered from our subjective test.

TABLE 6. Performance of machine learning models in predicting the MOS of point clouds in AR environments. The bold entry signifies the best performance.

QoE Model Type	R2 Score ↑	MSE ↓
Gradient Boosting Regressor	**0.8582**	**0.2874**
Random Forest Regressor	0.8384	0.3273
Decision Tree Classifier	0.8155	0.3738
Decision Tree Regressor	0.7954	0.4146
Polynomial Regression (Degree 2)	0.7650	0.4761

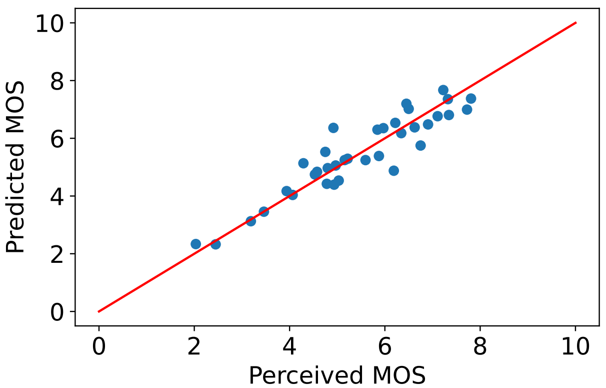


FIGURE 13. Perceived MOS (from our subjective test) versus predicted MOS using the gradient boosting regressor. The red line represents the $y = x$ line.

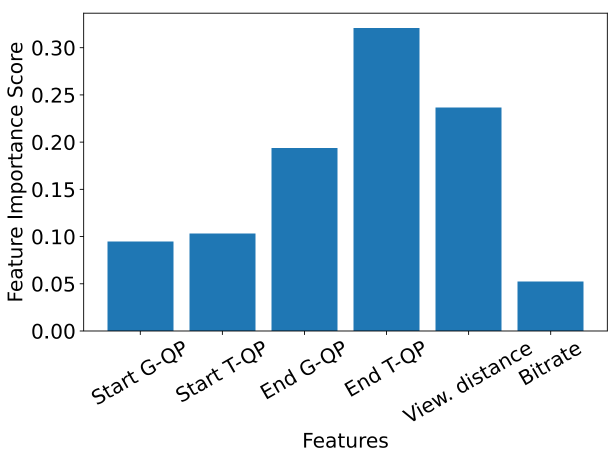


FIGURE 14. Feature importance scores of input features in gradient boosting regressor.

Fig. 14 presents feature importance scores of input features in the Gradient Boosting Regressor model using the Python scikit-learn library [31]. A higher score means more importance when building a predictive model. It is highlighted that the end T-QP plays the most crucial role for Gradient Boosting Regressor in predicting the QoE, followed by viewing distance and end G-QP. Their importance scores are 0.32, 0.24, and 0.19, respectively. The content characteristics represented by the bitrate are the least relevant feature of the prediction model, with an importance score of 0.05.

Additionally, the examination reveals that classification models are able to predict the MOS of PCs in AR environments, though regression models provide better performance. The best classification model, Decision Tree Classifier, achieves an R2 score of 0.8155 but is only ranked fifth among the tested models.

VI. CONCLUSION AND FUTURE WORK

In this paper, we investigate the impact of different factors on the QoE of point cloud videos in AR environments, including encoding parameters, quality switches, viewing distance, and content characteristics. We performed subjective tests with two separate tasks. The first task evaluates the impact of encoding parameters, quality switches, and content characteristics. The second task focuses on the impact of viewing distance and content characteristics. A common point cloud dataset was encoded using MPEG's V-PCC reference encoder and shown via an AR HMD (Microsoft HoloLens 2). We also investigate the correlation between the objective metrics and the participants' ratings. In addition, several machine learning models were trained, and we evaluated their performance in terms of QoE prediction for point clouds in AR environments.

The experimental results show that all of the considered parameters significantly impact the QoE of point clouds. We conclude that the QoE will be decreased if the sequence is encoded at high QPs and/or switches to lower quality and/or is viewed at a shorter distance, and vice versa. Additionally, lower contrast contents in the tested dataset are able to provide higher QoE. We also suggest that the sequence should be encoded at lower QPs (*i.e.*, better quality) to maintain a good quality score when viewed at a shorter distance. Regarding the correlation of objective metrics and subjective results, both color and geometry PSNRs show a positive correlation with the participants' scores, but the former is more correlated with a coefficient of more than 0.8 for both Spearman and Pearson methods. Additionally, cybersickness does not seem to be a major concern for point cloud-based AR applications, as less than 22% of the participants felt considered symptoms. However, the visual quality of point clouds has not yet reached the level that viewers expect. Finally, machine learning-based models perform reasonably well in the prediction of the participants' ratings. In particular, the Gradient Boosting Regressor provides the best QoE prediction among the models considered, with an R2 score of almost 0.86.

Overall, the results of this study provide valuable insights into the factors that impact the QoE of point clouds in the context of HAS in an AR environment. These insights can be used to improve the quality and performance of point cloud-based applications. Our results show that the current quality of PCs compressed by the MPEG V-PCC reference software TMC2 does not meet viewers' expectations, suggesting a need for improved compression algorithms to make PCs appear more realistic. Additionally, our analysis of cybersickness is limited to AR telepresence applications with slow-moving human objects. Further research is needed to investigate cybersickness in AR applications with fast-moving objects, such as first person shooter games. Finally, a QoE model for PCs should take into account the viewing distance, as it is one of the major factors in QoE.

In the future, we plan to extend our work in three directions: *(1)* develop machine learning-based compression approaches for PCs, *(2)* investigate cybersickness in AR applications with fast-moving objects, and *(3)* develop a QoE model for PCs that takes into account the viewing distance.

REFERENCES

[1] *ICIP 2020*. Accessed: Jul. 13, 2023. [Online]. Available: https://2020.ieeeicip.org/program/tutorials/point-cloud-compression-in-mpeg/index.html

[2] *Scalable Platform for Innovations on Real-Time Immersive Telepresence (SPIRIT) Project. Funded by the European Union. Grant Agreement 101070672*. Accessed: Jul. 13, 2023. [Online]. Available: https://www.spirit-project.eu/

[3] *Common Test Conditions for Point Cloud Compression, Author=MPEG 3DG*, Standard ISO/IEC JTC1/SC29/WG1 N17229, 2017.

[4] (2022). *Video Point Cloud Compression—VPCC—MPEG-PCC-TMC2 Test Model Candidate Software*. Accessed: Jul. 27, 2023. [Online]. Available: https://github.com/MPEGGroup/mpeg-pcc-tmc2

[5] H. Abdi and L. J. Williams, "Tukey's honestly significant difference (HSD) test," *Encyclopedia Res. Design*, vol. 3, no. 1, pp. 1–5, 2010.

[6] E. Alexiou, N. Yang, and T. Ebrahimi, "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality," in *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–6.

[7] H. Amirpour, R. Schatz, C. Timmerer, and M. Ghanbari, "On the impact of viewing distance on perceived video quality," in *Proc. Int. Conf. Vis. Commun. Image Process. (VCIP)*, Dec. 2021, pp. 1–5.

[8] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad, and Z. Fei, "Subjective QoE of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp. 148084–148099, 2020.

[9] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain, "A survey of augmented, virtual, and mixed reality for cultural heritage," *J. Comput. Cultural Heritage*, vol. 11, no. 2, pp. 1–36, Jun. 2018.

[10] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann, "A survey on bitrate adaptation schemes for streaming media over HTTP," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 562–585, 1st Quart., 2019.

[11] J. Birch, "Efficiency of the ishihara test for identifying red-green colour deficiency," *Ophthalmic Physiological Opt.*, vol. 17, no. 5, pp. 403–408, Sep. 1997.

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

[13] P. Caserman, A. Garcia-Agundez, A. Gámez Zerban, and S. Göbel, "Cybersickness in current-generation virtual reality head-mounted displays: Systematic review and outlook," *Virtual Reality*, vol. 25, no. 4, pp. 1153–1170, Dec. 2021.

[14] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," in *Proc. Intell. Syst. Technol. New Millennium (TENCON)*, vol. 2, Sep. 2000, pp. 421–424.

[15] A. Chen, S. Mao, Z. Li, M. Xu, H. Zhang, D. Niyato, and Z. Han, "An introduction to point cloud compression standards," *GetMobile, Mobile Comput. Commun.*, vol. 27, no. 1, pp. 11–17, May 2023.

[16] G. Cofano, L. De Cicco, T. Zinner, A. Nguyen-Ngoc, P. Tran-Gia, and S. Mascolo, "Design and experimental evaluation of network-assisted strategies for HTTP adaptive streaming," in *Proc. 7th Int. Conf. Multimedia Syst.*, May 2016, pp. 1–12.

[17] R. L. de Queiroz and P. A. Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3947–3956, Aug. 2016.

[18] T. N. Duc, C. T. Minh, T. P. Xuan, and E. Kamioka, "Convolutional neural networks for continuous QoE prediction in video streaming services," *IEEE Access*, vol. 8, pp. 116268–116278, 2020.

[19] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, *8I Voxelized Full Bodies, version 2—A Voxelized Point Cloud Dataset*, document ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG), Input Document M40059/M74006, 2017.

[20] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Trans. Signal Inf. Process.*, vol. 9, no. 1, p. e13, 2020.

[21] Y. Huang, J. Peng, C.-C.-J. Kuo, and M. Gopi, "A generic scheme for progressive point cloud coding," *IEEE Trans. Vis. Comput. Graphics*, vol. 14, no. 2, pp. 440–453, Mar. 2008.

[22] *Methodologies for the Subjective Assessment of the Quality of Television Images*, ITU, document Recommendation ITU-R BT. 500-15 (05/2023), 2023.

[23] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Joint geometry and color projection-based point cloud quality metric," *IEEE Access*, vol. 10, pp. 90481–90497, 2022.

[24] T. K. Kim, "T test as a parametric statistic," *Korean J. Anesthesiol.*, vol. 68, no. 6, pp. 540–546, 2015.

[25] R. Kirollos and W. Merchant, "Comparing cybersickness in virtual reality and mixed reality head-mounted displays," *Frontiers Virtual Reality*, vol. 4, Feb. 2023, Art. no. 1130864.

[26] G. Kougioumtzidis, V. Poulkov, Z. D. Zaharis, and P. I. Lazaridis, "A survey on multimedia services QoE assessment and machine learning-based prediction," *IEEE Access*, vol. 10, pp. 19507–19538, 2022.

[27] M. Krivokuća, M. Koroteev, and P. A. Chou, "A volumetric approach to point cloud compression," 2018, *arXiv:1810.00484*.

[28] MPEG. *MPEG-3DG MPEG-PCC-Dmetric Metric Software*. Accessed: Aug. 16, 2023. [Online]. Available: http://mpegx.int-evry.fr/software/MPEG/PCC/mpeg-pcc-dmetric

[29] M. Nguyen, C. Timmerer, and H. Hellwagner, "H2BR: An HTTP/2-based retransmission technique to improve the QoE of adaptive video streaming," in *Proc. 25th ACM Workshop Packet Video*, 2020, pp. 1–7.

[30] J. Park, P. A. Chou, and J.-N. Hwang, "Rate-utility optimized streaming of volumetric media for augmented reality," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 149–162, Mar. 2019.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2012.

[32] S. Perry, *JPEG Pleno Point Cloud Coding Common Test Conditions V3.6*, Standard ISO/IEC JTC1/SC29/WG1 N 91058, 2021.

[33] A. Pose-Díez-de-la-Lastra, R. Moreta-Martinez, M. García-Sevilla, D. García-Mato, J. A. Calvo-Haro, L. Mediavilla-Santos, R. Pérez-Mañanes, F. von Haxthausen, and J. Pascau, "HoloLens 1 vs. holoLens 2: Improvements in the new model for orthopedic oncological interventions," *Sensors*, vol. 22, no. 13, p. 4915, Jun. 2022.

[34] L. Rebenitsch and C. Owen, "Estimating cybersickness from virtual reality applications," *Virtual Reality*, vol. 25, no. 1, pp. 165–174, Mar. 2021.

[35] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuca, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging MPEG standards for point cloud compression," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 133–148, Mar. 2019.

[36] J.-C. Servotte, M. Goosse, S. H. Campbell, N. Dardenne, B. Pilote, I. L. Simoneau, M. Guillaume, I. Bragard, and A. Ghuysen, "Virtual reality experience: Immersion, sense of presence, and cybersickness," *Clin. Simul. Nursing*, vol. 38, pp. 35–43, Jan. 2020.

[37] L. Stahle and S. Wold, "Analysis of variance (ANOVA)," *Chemometrics Intell. Lab. Syst.*, vol. 6, no. 4, pp. 259–272, 1989.

[38] T. Stockhammer, "Dynamic adaptive streaming over HTTP: Standards and design principles," in *Proc. 2nd Annu. ACM Conf. Multimedia Syst.*, Feb. 2011, pp. 133–144.

[39] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[40] K. Takahashi. (2021). *Pcx—Point Cloud Importer/Renderer for Unity*. Accessed: Jul. 27, 2023. [Online]. Available: https://github.com/keijiro/Pcx

[41] B. Taraghi, M. Nguyen, H. Amirpour, and C. Timmerer, "Intense: In-depth studies on stall events and quality switches and their impact on the quality of experience in HTTP adaptive streaming," *IEEE Access*, vol. 9, pp. 118087–118098, 2021.

[42] C. Timmerer and A. C. Begen, "A journey towards fully immersive media access," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 2703–2705.
[43] H. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on QoE of 360 video for VR communication," in *Proc. IEEE 19th Int. workshop multimedia signal Process. (MMSP)*, Jul. 2017, pp. 1–6.
[44] S. Van Damme, I. Mahdi, H. K. Ravuri, J. van der Hooft, F. De Turck, and M. T. Vega, "Immersive and interactive subjective quality assessment of dynamic volumetric meshes," in *Proc. 15th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2023, pp. 141–146.
[45] S. Van Damme, M. Torres Vega, J. van der Hooft, and F. De Turck, "Clustering-based psychometric no-reference quality model for point cloud video," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 1866–1870.
[46] J. v. d. Hooft, H. Amirpour, M. T. Vega, Y. Sanchez, R. Schatz, T. Schierl, and C. Timmerer, "A tutorial on immersive video delivery: From omnidirectional video to holography," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1336–1375, 2nd Quart., 2023.
[47] 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 *Proc. 12th Int. Conf. Quality Multimedia Exper. (QoMEX)*, May 2020, pp. 1–6.
[48] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner, "Towards 6DOF HTTP adaptive streaming through point cloud compression," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 2405–2413.
[49] S. Vats. *A Platform for Subjective Quality Assessment in Mixed Reality Environments*. Accessed: Aug. 15, 2023. [Online]. Available: https://github.com/shivivats-aau/MR-Subjective-Testing-Platform
[50] S. Vats, M. Nguyen, S. Van Damme, J. van der Hooft, M. T. Vega, T. Wauters, C. Timmerer, and H. Hellwagner, "A platform for subjective quality assessment in mixed reality environments," in *Proc. 15th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2023, pp. 131–134.
[51] VRCompare. *Microsoft Hololens 2*. Accessed: Feb. 10, 2023. [Online]. Available: https://vr-compare.com/headset/microsofthololens2
[52] L. Wang, C. Li, W. Dai, S. Li, J. Zou, and H. Xiong, "QoE-driven adaptive streaming for point clouds," *IEEE Trans. Multimedia*, vol. 25, pp. 2543–2558, 2022.
[53] J. Weil, Y. Alkhalili, A. Tahir, T. Gruczyk, T. Meuser, M. Mu, H. Koeppl, and A. Mauthe, "Modeling quality of experience for compressed point cloud sequences based on a subjective study," in *Proc. 15th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2023, pp. 135–140.
[54] D. L. Whaley III, "The interquartile range: Theory and estimation," Ph.D. thesis, Dept. Math., East Tennessee State Univ., Johnson City, TN, USA, 2005.
[55] T.-T. Wong, "Performance evaluation of classification algorithms by K-fold and leave-one-out cross validation," *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.
[56] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong, "Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4630–4644, Dec. 2021.

MINH NGUYEN received the engineering degree from the Hanoi University of Science and Technology, in 2018, and the Ph.D. degree from the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt (AAU), in 2023. He is currently a Postdoctoral Researcher in the scalable platform for innovations on real-time immersive telepresence (SPIRIT) project with AAU. His research interests include adaptive video streaming, multimedia networking, augmented and virtual reality, and QoS/QoE evaluation. More information can be found at https://minhstreaming.com.

SHIVI VATS received the Bachelor of Computer Science degree from IIIT Una, India, in 2018, and the master's degree in game studies and engineering from Klagenfurt University, in 2022. He was on the 5G Playground Research Project on behalf of Klagenfurt University. He is currently a Project Assistant with the SPIRIT Project. His research interests include volumetric video, virtual, augmented, and mixed reality, and low latency video streaming. More information can be found at https://shivivats.github.io/.

SAM VAN DAMME was born in Zottegem, Belgium, in 1995. He received the B.Sc. and M.Sc. degrees in computer science engineering from Ghent University, Belgium, in 2017 and 2019, respectively. He is currently active as a Joint Ph.D. Researcher affiliated with the Internet Technology and Data Science Laboratory (IDLab), Department of Information Technology (INTEC), Ghent University, and the e-Media Research Laboratory, KU Leuven, Belgium. He has expertise in quality of experience (QoE) modeling and assessment, human–computer interaction (HCI), and haptic feedback, with a main focus on immersive multimedia technologies, such as virtual and augmented reality (VR/AR).

JEROEN VAN DER HOOFT (Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science engineering from Ghent University, Belgium, in 2014 and 2019, respectively. Funded by a grant provided by FWO, he is currently a Postdoctoral Fellow with Ghent University. His main research interests include the end-to-end quality of experience optimization in adaptive video streaming, and low-latency delivery of immersive video content.

MARIA TORRES VEGA (Senior Member, IEEE) received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Madrid, Spain, in 2009, and the Ph.D. degree from the Eindhoven University of Technology, The Netherlands, in 2017. She is currently a Tenure Track Assistant Professor with KU Leuven, Belgium, where her research focuses on devising human-driven control and management mechanisms for enhancing the perception of immersive systems. Her research interests include quality of service and QoE in immersive multimedia systems and autonomous networks management.

TIM WAUTERS (Member, IEEE) received the M.Sc. and Ph.D. degrees in electro-technical engineering from Ghent University, in 2001 and 2007, respectively. He has been a Postdoctoral Fellow with the FWO-V, Department of Information Technology (INTEC), Ghent University, where he is currently a Senior Researcher with imec. His work has been published in more than 160 scientific publications. His main research interests include the design and management of networked services, covering multimedia distribution, cybersecurity, big data, and smart cities.

FILIP DE TURCK (Fellow, IEEE) currently leads the Network and Service Management Research Group, Ghent University, Belgium, and imec. He is involved in several research projects with industry and academia. He (co)authored over 750 peer-reviewed papers. His research interests include design of efficient softwarized networks and video systems. He serves as a Steering Committee Member for the IM, NOMS, CNSM, and NetSoft Conferences. He received the IEEE ComSoc Dan Stokesberry Award, in 2021. He served as the Chair for the IEEE Technical Committee on Network Operations and Management (CNOM). He served as the Editor-in-Chief for IEEE Transactions on Network and Service Management.

CHRISTIAN TIMMERER (Senior Member, IEEE) has co-founded Bitmovin to provide professional services around MPEG-DASH, in 2012, where he holds the position of the Chief Innovation Officer (CIO)-the Head of Research and Standardization. He is currently a Full Professor of computer science with the Institute of Information Technology (ITEC), Alpen-Adria-Universität Klagenfurt (AAU), and also the Director of the Christian Doppler (CD) Laboratory ATHENA. He has participated in several EC-funded projects, notably DANAE, ENTHRONE, P2P-Next, ALICANTE, SocialSensor, COST IC1003 QUALINET, ICoSOLE, and SPIRIT. His research interests include multimedia systems, immersive multimedia communication, streaming, adaptation, and the quality of experience, where he has coauthored more than 300 articles and holds seven patents. He was the General Chair of WIAMIS 2008, QoMEX 2013, MMSys 2016, and PV 2018. He also participated in ISO/MPEG work for several years, notably in the area of MPEG-21, MPEG-M, MPEG-V, and MPEG-DASH, where he also served as a Standard Editor. More information can be found at http://blog.timmerer.com.

HERMANN HELLWAGNER (Senior Member, IEEE) is currently a Full Professor of computer science with Alpen-Adria-Universität Klagenfurt (AAU), where he leads the Research Group Multimedia Communication (MMC), Institute of Information Technology (ITEC). Earlier, he held a position as an Associate Professor with the Technical University of Munich (TUM) and a Senior Researcher with Siemens Corporate Research, Munich. He has published widely on parallel computer architecture, parallel programming, and multimedia communication and adaptation. His current research interests include distributed multimedia systems, multimedia communication and adaptation, QoS/QoE, and communication in multi-UAV networks. He was a member of the Scientific Board of the Austrian Science Fund (FWF), from 2005 to 2016, and the FWF Vice President, from 2013 to 2016. He is a Senior Member of ACM. He is also a member of the CD Senate of Christian Doppler Forschungsgesellschaft (CDG). More information can be found at https://www.itec.aau.at/ hellwagn/.