ComPEQ–MR: Compressed Point Cloud Dataset with Eye Tracking and Quality Assessment in Mixed Reality

Minh Nguyen[§*]
Fraunhofer FOKUS
Germany

Shivi Vats[*]
Alpen-Adria-Universität Klagenfurt
Austria

Xuemei Zhou
Centrum Wiskunde & Informatica
TU Delft
Netherlands

Irene Viola
Centrum Wiskunde & Informatica
Netherlands

Pablo Cesar
Centrum Wiskunde & Informatica
TU Delft
Netherlands

Christian Timmerer
Alpen-Adria-Universität Klagenfurt
Austria

Hermann Hellwagner
Alpen-Adria-Universität Klagenfurt
Austria

ABSTRACT

Point clouds (PCs) have attracted researchers and developers due to their ability to provide immersive experiences with six degrees of freedom (6DoF). However, there are still several open issues in understanding the Quality of Experience (QoE) and visual attention of end users while experiencing 6DoF volumetric videos. First, encoding and decoding point clouds require a significant amount of both time and computational resources. Second, QoE prediction models for dynamic point clouds in 6DoF have not yet been developed due to the lack of visual quality databases. Third, visual attention in 6DoF is hardly explored, which impedes research into more sophisticated approaches for adaptive streaming of dynamic point clouds. In this work, we provide an open-source **Com**pressed **P**oint cloud dataset with **E**ye-tracking and **Q**uality assessment in **M**ixed **R**eality (**ComPEQ–MR**). The dataset comprises four compressed dynamic point clouds processed by Moving Picture Experts Group (MPEG) reference tools (*i.e.*, VPCC and GPCC), each with 12 distortion levels. We also conducted subjective tests to assess the quality of the compressed point clouds with different levels of distortion. The rating scores are attached to **ComPEQ–MR** so that they can be used to develop QoE prediction models in the context of MR environments. Additionally, eye-tracking data for visual saliency is included in this dataset, which is necessary to predict where people look when watching 3D videos in MR experiences. We collected opinion scores and eye-tracking data from 41 participants, resulting in 2132 responses and 164 visual attention maps in total. The dataset is available at **https://ftp.itec.aau.at/datasets/ComPEQ-MR/**.

[*]Both authors contributed equally to the paper.

[§]Minh Nguyen conducted this work when he was a Postdoctoral Researcher at the Alpen-Adria-Universität Klagenfurt, Austria.

MMSys '24, April 15–18, 2024, Bari, Italy

ACM ISBN 979-8-4007-0412-3/24/04.
https://doi.org/10.1145/3625468.3652182

CCS CONCEPTS

• **Information systems** → **Multimedia databases**; **Multimedia content creation**; • **General and reference** → **Evaluation**.

KEYWORDS

Point Cloud, Augmented Reality, Dataset, Adaptive Video Streaming, Metaverse

ACM Reference Format:
Minh Nguyen[§], Shivi Vats, Xuemei Zhou, Irene Viola, Pablo Cesar, Christian Timmerer, and Hermann Hellwagner. 2024. ComPEQ–MR: Compressed Point Cloud Dataset with Eye Tracking and Quality Assessment in Mixed Reality. In *ACM Multimedia Systems Conference 2024 (MMSys '24), April 15–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3625468.3652182

1 INTRODUCTION

Recent advancements in immersive video have enabled the creation of six-degrees-of-freedom (6DoF) experiences using Extended Reality (XR) technologies like Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). Point clouds (PCs) are a widely adopted format for presenting immersive videos because of their high-fidelity and viewpoint-independent nature. Dynamic PCs (DPCs) can be used for applications in telepresence (*i.e.*, video conferencing [32]), medical and health (*i.e.*, anatomic pathology [10, 16]), and autonomous driving [9].

A PC is a set of thousands or even millions of points in space with information about 3D coordinates (*i.e.*, x, y, z) and/or other attributes such as RGB. However, the storage and bandwidth requirements of PCs are significant, with a single raw PC frame potentially reaching hundreds of megabits in size. This translates to a bandwidth demand of several gigabits per second for an uncompressed 30 frames-per-second (fps) video. Thus, efficient PC compression is crucial for storage, delivery, and rendering. However, it comes at the cost of visual quality degradation that can negatively impact the Quality of Experience (QoE) of end users.

Measuring the impact of PC compression on the QoE is of importance when evaluating the performance of different compression methods and when selecting a suitable distortion level to transmit and render to end users under specific circumstances (*e.g.*, current network conditions and device's storage capacity). Subjective quality assessment is often selected to gain insight into these impacts. Dynamic PCs (DPCs) have been subjectively evaluated under various viewing conditions, including a 2D screen [31], a VR Head-Mounted Display (HMD) [30], and a MR HMD [19, 20]. With VR and AR HMDs, end users can move freely in 6DoF conditions and their eye-tracking data can be collected to give more insight into end users' attention. Eye-tracking data is capable of supporting the selection of video quality [14]. Some open datasets for eye tracking in VR environments have been published [4, 36], but those for MR environments where end users can watch 3D content in familiar physical environments are still limited.

In this work, we provide a **Com**pressed **P**oint cloud dataset with **E**ye-tracking and **Q**uality assessment in **M**ixed **R**eality (**ComPEQ–MR**) that includes both eye-tracking data and quality rating scores for DPCs under MR conditions from subjective tests. The tested DPCs are also made publicly available for reproducibility. The contributions of this paper are thus threefold:

- We provide a compressed DPC database processed by state-of-the-art compression codecs: VPCC, GPCC Octree, and GPCC Trisoup. This database comprises 52 sequences that can be used to run subjective tests to consider QoE impact factors such as quality levels, quality switches, and stall events.
- We provide a visual saliency dataset from 41 observers while exploring four point cloud humans in the context of telepresence in MR environments. The visual saliency is collected in a task-free scenario where observers see the raw versions of point clouds. This dataset can help develop and compare approaches that predict where people look in DPCs.
- We conducted subjective tests to evaluate the QoE performance of the compression codecs. The rating scores are made publicly available to help train and validate QoE prediction models as well as develop objective quality metrics.

We summarize some existing datasets for point clouds, quality assessment, and visual attention in Table 1.

The remainder of the paper is organized as follows. Section 2 gives an overview of point cloud compression, eye-tracking experiments, and subjective quality assessment for 3D contents. Section 3 describes our data acquisition methods (original DPCs, eye-tracking data, and QoE scores). Section 4 presents a dataset analysis (eye-tracking data and subjective test results), the dataset structure, and usage scenarios. Finally, Section 5 concludes the paper.

2 RELATED WORK

2.1 Point Cloud Compression

Point cloud compression (PCC) has received much attention in the literature [25]. The Moving Picture Experts Group (MPEG) is working on two technologies: *(1) Geometry-based PCC (GPCC)* and *(2) Video-based PCC (VPCC)* [2]. GPCC directly encodes the 3D positions of PCs to generate the compressed representation using either the Octree or Trisoup (triangle soup) methods. The color of the PC can be encoded by the Region Adaptive Hierarchical Transform (RAHT) or the Predicting/Lifting (Predlift) transform. In contrast to GPCC, VPCC adopts a more indirect approach by projecting the 3D points onto 2D images. Subsequently, it utilizes conventional encoders, *e.g.*, HEVC, to compress these projected images. This strategy enables VPCC to leverage existing efficient coding mechanisms and simplify deployment.

2.2 Eye-tracking Data

Visual saliency has been widely used for perception-related optimization algorithms. Visual saliency models are typically evaluated based on ground truth fixations that are collected from eye-tracking experiments [15]. Fixation is defined as the maintenance of the gaze at a single location for a specific period [1]. A public dataset containing both the mean opinion score (MOS) and eye-tracking information is crucial for the research community to evolve in the development of efficient techniques for coding, transmitting, and rendering volumetric content. Zhang *et al.* [35] propose a new experimental methodology to obtain reliable eye-tracking data to provide insight into the optimal use of visual attention in image quality assessment. David *et al.* [6] obtain the head and eye movements for 360° videos in a free-viewing experiment; saliency maps, scanpaths, and users' behaviors are presented. Alexiou *et al.* [4] conduct an eye-tracking experiment in VR on static PCs for saliency modeling. Zhou *et al.* [36] recruited participants to watch compressed DPCs via a VR HMD. The authors provided eye-tracking data with heat maps to boost the saliency research. They focused on the impact of quality distortions on the eye gaze of the viewers. The limitation of this work is that the distorted DPCs due to the compression process are not published. However, an eye-tracking dataset for DPCs in AR environments has not been reported in the literature.

2.3 Subjective Quality Assessment

Many works have been focusing on subjective quality assessment for PCs. Ak *et al.* [2] conducted a subjective test for static PCs through a crowdsourcing platform. More than 1200 test sequences (*i.e.*, stimuli) were evaluated by more than 3000 participants. However, as the participants watch PCs on their 2D monitors, there is no interaction between the viewers and the 3D objects. Eye-tracking data is not included either. Subramanyam *et al.* [27] focused on dynamic PCs for VR environments but the compressed PCs are not published and the eye-tracking data is not taken into account. In addition, GPCC is omitted.

Our previous work [19, 20] considered the impact of quality levels, quality switches, viewing distance, and content characteristics on the QoE for PCs in AR environments through a subjective study. Subjective tests were performed with the distorted PCs compressed by VPCC and watched on an AR HMD. However, the participants were asked to stand still; thus, there was no interaction.

Unlike these related works, we provide opinion scores and eye-tracking data collected in interactive subjective tests in an AR environment. Our subjective data can benefit the validation of QoE prediction models, objective quality metrics development, and help researchers understand the visual attention of participants. We also publish the quality distortions of PCs for reproducibility.

Table 1: Existing datasets for point clouds, quality assessment, and visual attention.

Dataset	Type	Compression	Stimuli	Participants	Duration	Display	Interaction*	Opinion Score	Eye-Tracking Data	Public PCs
VsenseVVDB2 [34]	Dynamic	VPCC, GPCC	136	23	10 s	2D	✗	✓	✗	✗
BASICS [2]	Static	VPCC, GPCC	1494	60	-	2D	✗	✓	✗	✓
Subramanyam *et al.* [27]	Dynamic	VPCC, CWI-PCL	72	52	5 s	VR	✓	✓	✗	✗
QAVA-DPC [36]	Dynamic	VPCC, GPCC, CWI-PCL	50	40	10 s	VR	✓	✓	✓	✗
Nguyen *et al.* [19]	Dynamic	VPCC	36	32	10 s	AR	✗	✓	✗	✗
ComPEQ–MR (Ours)	Dynamic	VPCC, GPCC (Octree and Trisoup)	52	41	20 s	AR	✓	✓	✓	✓

* Interaction here refers to being able to move around and observe the point cloud from different angles.

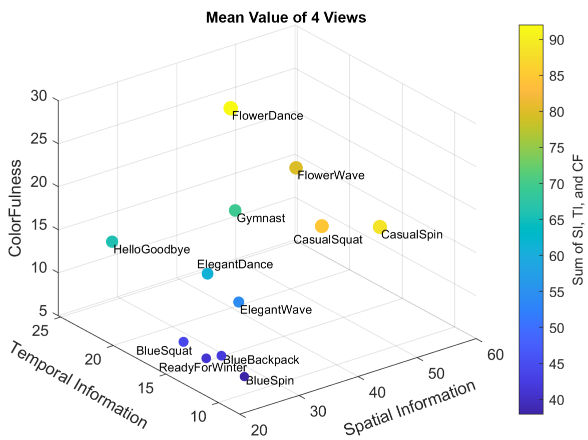
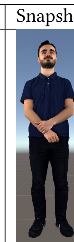


Figure 1: SI, TI, and CF characteristics of the UVG-VPC dataset.

Table 2: UVG-VPC sequences [8] used in this open dataset. Red and green numbers indicate that the sequence has low and high values, respectively, for corresponding metrics.

Name	Description	Snapshot
BlueSpin	A person wearing a blue t-shirt and spinning at a consistent rate. SI: 20.8 TI: 8.0 Colorfulness: 8.6	
CasualSquat	A person wearing a striped shirt and jeans in the performance of a squat exercise. SI: 53.5 TI: 19.0 Colorfulness: 11.5	
FlowerDance	A person in a long, flowing dress spinning and twirling. SI: 43.9 TI: 22.3 Colorfulness: 25.3	
ReadyForWinter	A person donning a beanie and scarf. SI: 20.6 TI: 11.5 Colorfulness: 7.8	

3 DATA ACQUISITION

3.1 Original Sequences

In this work, we use a state-of-the-art uncompressed point cloud dataset and cutting-edge codecs for point cloud compression. One of the latest voxelized 10-bit point cloud datasets is UVG-VPC [8]. This dataset comprises 12 sequences of human objects with various content characteristics and number of points. These dynamic point clouds are captured by 96 cameras at a frame rate of 25 fps and are each 10 s long.

Fig. 1 shows the spatial information (SI), temporal information (TI), and colorfulness (CF) of the UVG-VPC sequences. We select four sequences that have large differences in these criteria to cover wide diversity: BlueSpin, CasualSquat, FlowerDance, and ReadyForWinter. Table 2 describes the chosen sequences.

This dataset focuses on MPEG's tools for point cloud compression, including GPCC and VPCC. GPCC is suitable for static and dynamically acquired point clouds and VPCC is typically used for dynamic point clouds. GPCC includes 3D point and color compression modes. The former includes two approaches, namely Octree and Trisoup, and the latter comprises RAHT and Predlift. Previous research [3, 13] found that the performance of Octree is equivalent to or better than that of Trisoup, for the same color encoding module, and that the Lifting color encoding module is marginally better than the RAHT module. In this work, we use VPCC and the combinations GPCC-Oct-Pred (Octree and Predlift modes) and GPCC-Tri-RAHT (Trisoup and RAHT modes) to process the PCs.

We used MPEG's reference software tools to encode the objects, including the test model category (TMC) 2 version v.22.1[1] for VPCC and TMC13 version v.23.0[2] for GPCC. The quality levels are based

[1]https://mpeg.expert/software/MPEG/3dgh/VPCC/software/mpeGPCC-tmc2. Accessed 18 September 2023.
[2]http://mpegx.int-evry.fr/software/MPEG/PCC/TM/mpeGPCC-tmc13. Accessed 18 September 2023.

Table 3: Encoder parameters to generate compressed dynamic PCs.

Compression		Quality Levels				
		r01	r02	r03	r04	r05
VPCC	Geometry QP	36	32	28	20	16
	Texture QP	47	42	37	27	22
GPCC-Oct-Pred	QP	-	-	40	34	28
	Depth	-	-	0.5	0.75	0.875
GPCC-Tri-RAHT	QP	40	34	28	22	-
	Level	5	4	3	2	-

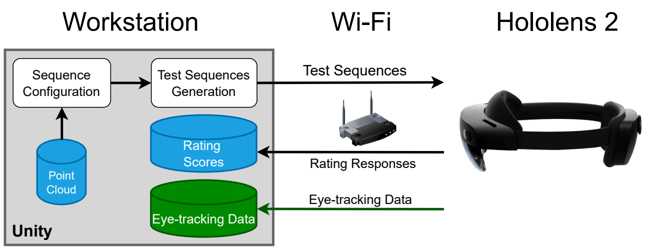


Figure 2: Platform architecture to conduct subjective test. The component highlighted in green color is added in this work, compared to the original version in [33].

on Common Test Conditions (CTC) [17, 18] for each tool, which are described in Table 3.

3.2 Subjective Tests

We upgraded our subjective test platform[3] [33], as shown in Fig. 2, to perform two separate tasks:

- Task 1 – Eye-tracking data acquisition: The eye-tracking data is collected while the participants watch the raw versions of the tested DPCs.
- Task 2 – QoE scores acquisition: The scores are collected while participants watch various quality levels of the tested DPCs.

The participants are asked to start watching every sequence from the same position in the real world. Then, they can freely move around the room. The test room has gray walls with low illumination, which is in accordance with the recommendations of ITU-R BT.500-15 [12]. The area for the participants is a 4 m × 4 m square; thus, the participant has enough space to move freely while watching the objects. We place DPCs 2 m away from the participants at the beginning of each sequence to emulate the trigger distance to start a conversation between two people [21], then the participants can move freely in the room.

3.2.1 Participants. A total of 41 participants, who were recruited from Alpen-Adria-Universität Klagenfurt, participated in the subjective test, including 19 (46%) females and 22 (54%) males. 18 (45%) were in the age group of 18 to 24 years, 14 (35%) were between 25 and 34, 7 (17.5%) between 35 and 44, and 1 (2.5%) between 55 and 64.

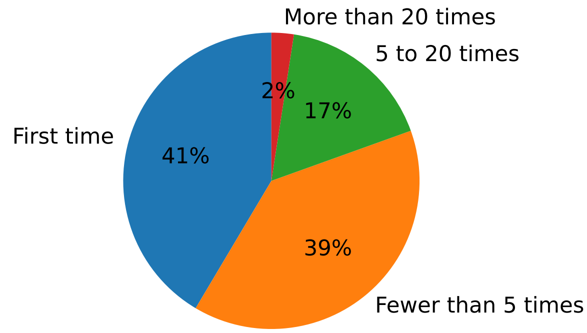


Figure 3: Frequency of participants using AR HMDs.

Fig. 3 shows the experience of the participants with AR HMDs. People who have never used AR HMDs are the most dominant group with 41% of the total participants, followed by those who have experienced AR fewer than five times.

3.2.2 Eye-tracking Data Acquisition. Before this task begins, the in-built eye calibration[4] of the HoloLens 2 is performed for each participant, since the eye-tracking services of the HoloLens do not function without calibration. The actual Task 1 consists of two subtasks: *(i)* error measurement and *(ii)* watching the PC videos.

Since the state-of-the-art tool GazeMetrics [5] is not available for HoloLens 2, we implemented our own system to show the calibration targets and store the user gaze data. Nine 5 cm large spherical targets arranged in a rectangular format are shown 2 m in front of the participant. Each target is visible for 3 s, and the participants are asked to look at them while staying stationary. The participant's gaze origin and gaze direction are stored 60 times per second to match the HoloLens 2 framerate[5].

The second subtask consists of the participant watching 20 s long uncompressed (voxelized 10-bit format) PC sequences. The participants are allowed to move freely in the space of the test room (6DoF), but are required to return to the starting point before starting the next sequence. The participant's gaze origin and gaze direction are stored once per DPC frame. The order of the DPCs is randomized among the participants to avoid bias.

The subtasks are alternated until the participant has watched all four sequences. The duration of this task is around 5 min.

3.2.3 QoE Scores Acquisition. We follow the subjective methodology Absolute Category Rating (ACR) based on ITU-T Recommendation P.919 [11]. ACR is a single-stimulus methodology where the observer sees one video at a time before spending some time to rate that video. We do not use the double-stimulus method, where the observer is presented with two stimuli, because these stimuli may not be viewed from the same viewpoint under MR conditions. As we follow the ACR method, a five-level rating scale is used as follows:

- 5: excellent;

[3]https://github.com/shivivats-aau/MR-Subjective-Testing-Platform. Accessed 20 October 2023.

[4]https://learn.microsoft.com/en-us/windows/mixed-reality/design/eye-tracking. Accessed 21 November 2023.

[5]https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/performance/perf-getting-started?view=mrtkunity-2022-05. Accessed 30 November 2023.

- 4: good;
- 3: fair;
- 2: poor;
- 1: bad.

Each participant watched 52 DPCs (48 compressed and four raw sequences), each of which is 20 s long (*i.e.*, two loops). Similar to Task 1, the participant can move freely in the test room, and the sequences are randomized. The total duration of this task is around 30 min.

4 TEST RESULTS AND DATASET

4.1 Eye Tracking

4.1.1 Processing Error Data. Similar to GazeMetrics [5], we process the error data and obtain the average accuracy (error) per marker for each user using the difference between the actual gaze ray and the projected gaze ray from the origin to the marker. For each marker, the initial 1.5 s are discarded to account for initial movements. A threshold of 7.5° is used to discard unintentional gaze. Barycentric interpolation with the corresponding angular error as weights is applied to the user gazes, and based on the interpolation results, a compensatory weighted average angular error is applied to each gaze sample.

4.1.2 Identifying Fixations. The Dispersion Threshold Identification (I-DT) [23] method was used to identify the fixations within angular and temporal constraints. The angular dispersion and interval thresholds were set to 3° and 120 ms (3 frames) [36], respectively. The average of the gaze points within a valid fixation set is considered a fixation point. Barycentric interpolation is performed again for the fixation point, and if found to be valid, a compensatory error is applied similar to the previous step.

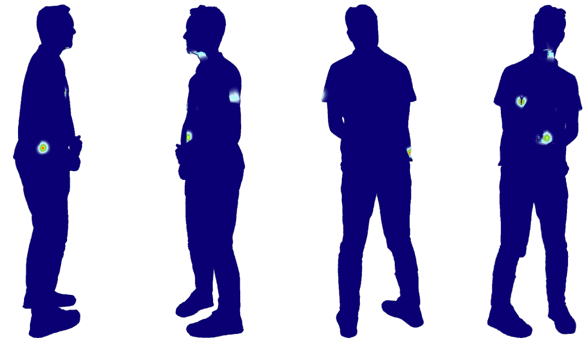
4.1.3 Mapping Gaze Data to DPC Frames. The truncated-cone-sector algorithm [4] is applied to assign weights to the DPC frames associated with a fixation. The algorithm can identify PC points corresponding to the fixation gazes. Points are assigned weights, and these weights are stored for assessment.

4.1.4 Filtering Gaze Data. The DBSCAN algorithm [7, 24] is used to filter out noisy fixation data. The minimum number of points for DBSCAN is determined using the point size [36], and the search distance is calculated using the k-distance graph [24][6].

4.1.5 Generating Fixation Maps. The weights are stored for every frame for every user. The sum of the weights of all users viewing a DPC frame gives us the fixation map for said frame. The heatmaps for every frame from four points of view (front, back, left, and right) can be found in the dataset[7]. 25 fps videos of the heatmaps are provided with the dataset as well.

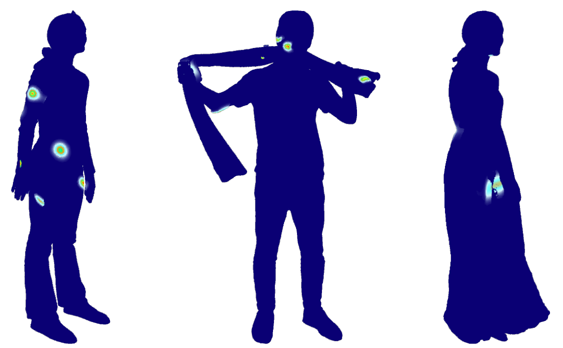
4.1.6 Analyzing Gaze Data. From the heatmaps, we can infer that the users fixate mostly on the faces of the objects and the parts with high degrees of motion. We can only show a selected number of heatmap frames here due to limited space (Fig. 4 and Fig. 5). Thus, we recommend going through the heatmap images and videos to get a better idea of users' viewing preferences.

[6]https://nl.mathworks.com/help/radar/ref/clusterdbscan.clusterdbscan.estimateepsilon.html. Accessed 25 January 2024.
[7]https://ftp.itec.aau.at/datasets/ComPEQ-MR/. Accessed 24 January 2024.



(a) Back view (b) Front view (c) Right view (d) Left view

Figure 4: The heatmap of fixation of a BlueSpin frame in four views.



(a) CasualSquat (b) ReadyForWinter (c) FlowerDance

Figure 5: The heatmap of fixation of the front view.

4.2 Subjective Tests

4.2.1 Opinion Scores. First, we show the raw opinion scores of the participants for all tested DPCs in Fig. 6. Some video sequences receive consistent scores from all participants. For example, the 18^{th} and 41^{st} videos (*i.e.*, BlueSpin and CasualSquat encoded by GPCC-Tri-RAHT at quality r01, respectively) have mostly all low scores (*i.e.*, scores 1 or 2), while the 5^{th} video gets high scores (*i.e.*, scores 4 and 5) from most of the participants. Regarding the participants, some of them are not satisfied with the quality of the video sequences (*e.g.*, participants 2 and 30), while others feel the opposite (*e.g.*, participant 17).

We calculate the Mean Opinion Score (MOS) by following the recommendations in ITU-R BT.500-15 [12]. No outliers are detected in our subjective tests. We also compute 95% confidence intervals (CIs). MOS and CIs are included in our dataset.

4.2.2 QoE Performance of the Compression Algorithms. Fig. 7 compares the QoE performance of the compression algorithms used in the dataset. We use the number of bits per point (bpp) for the rate. It can be clearly seen that VPCC achieves the best visual quality, which validates the findings of the predecessors [2]. However, GPCC-Oct-Pred provides worse MOS than GPCC-Tri-RAHT, which is opposite to [2].

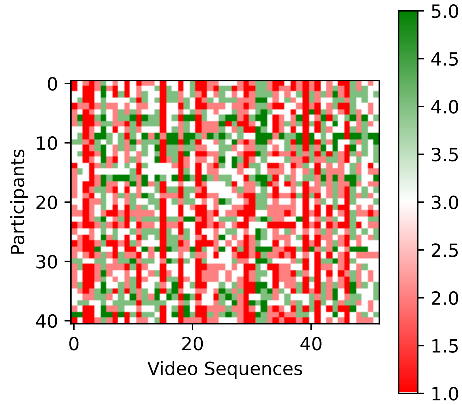
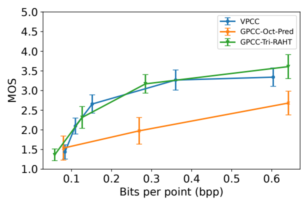
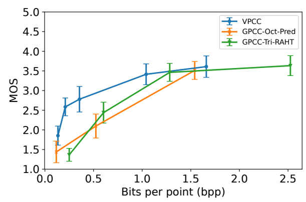


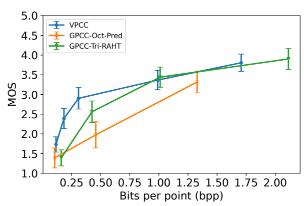
Figure 6: Raw opinion scores



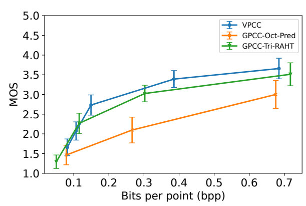
(a) BlueSpin



(b) CasualSquat



(c) FlowerDance



(d) ReadyForWinter

Figure 7: Bits per point vs. MOS for each video. Each point represents the MOS score computed by ITU.R BT.500 and the error bar is the 95% CI.

4.3 Dataset

The dataset comprises three top-level folders:

- **Compressed-point-cloud**: This folder comprises the compressed DPCs. It includes two sub-folders, namely *VPCC*, and *GPCC*, each with DPCs processed by corresponding MPEG tools. The *GPCC* folder is sequentially constituted by *octree-predlift* and *trisoup-raht* folders.
- **Eye-tracking**: This folder also comprises two sub-folders: *heatmap-images* and *heatmap-weights*. The former contains the heatmap images for every frame and 25 fps videos generated by concatenating these images. The latter contains text files with a fixation weight for every point of the PC frames.
- **Rating-score**: This folder contains a Comma Separated Value (CSV) file that stores all rating scores of participants for every DPC, and another CSV file storing MOS and the 95% CI.

4.4 Applications and Usage Scenarios

Our open dataset can benefit diverse research directions that explore the perceptual interaction and QoE of end-users in immersive environments. This includes, but is not limited to, the following applications:

- The compressed DPCs can be used to conduct further subjective tests with various impact factors to the QoE of the user such as quality levels, quality switches, and stall duration, in the context of HTTP Adaptive Streaming (HAS) [26, 29].
- The rating scores collected from participants with a wide variety of backgrounds (*i.e.*, age, gender, and previous experience with AR HMDs) can be used to train and validate QoE prediction models for DPCs in XR environments.
- The dataset is also useful for the development and evaluation of novel objective quality metrics.
- Our visual saliency database from 41 observers while watching humans in the format of DPCs can be used to develop and compare foveated rendering approaches [22] in the context of telepresence in MR environments. These approaches are necessary for dynamic streaming on the client side to provide higher quality of the focus areas where the end user is looking and lower quality for the others to optimize the delivered data [28].

5 CONCLUSIONS

In this paper, we propose an open dataset of dynamic PCs with compressed PCs, rating scores, and eye-tracking data for an MR experience. The compressed PCs include 12 quality levels of four dynamic PCs processed by three different compression algorithms (*i.e.*, VPCC, GPCC-Oct-Pred, GPCC-Tri-RAHT). We collect rating scores and eye-tracking data from 41 participants covering various age groups and experience in using AR HMDs. Our dataset can be utilized for a wide range of purposes: *(i)* testing XR streaming systems and algorithms, *(ii)* training and validating future QoE models, and *(iii)* developing and comparing approaches that predict visual attention.

The dataset is available at:
https://ftp.itec.aau.at/datasets/ComPEQ-MR/

ACKNOWLEDGEMENTS

This work has been funded by the European Union (SPIRIT project, Grant Agreement 101070672, https://www.spirit-project.eu/, and TRANSMIXR project, Grant Agreement 101070109, https://transmixr.eu/), and German Federal Ministry for Research and Education (6G_NeXt project, Grant Agreement 11-04133-3150-00001, https://6gnext.de).

REFERENCES

[1] Adler, F. H., and Fliegelman, M. Influence of Fixation on the Visual Acuity. *Archives of Ophthalmology 12*, 4 (1934), 475–483.

[2] Ak, A., Zerman, E., Quach, M., Chetouani, A., Smolic, A., Valenzise, G., and Callet, P. L. BASICS: Broad Quality Assessment of Static Point Clouds In Compression Scenarios. *arXiv preprint arXiv:2302.04796* (2023).

[3] Alexiou, E., Viola, I., Borges, T. M., Fonseca, T. A., De Queiroz, R. L., and Ebrahimi, T. A Comprehensive Study of the Rate-distortion Performance in MPEG Point Cloud Compression. *APSIPA Transactions on Signal and Information Processing 8* (2019), e27.

[4] Alexiou, E., Xu, P., and Ebrahimi, T. Towards Modelling of Visual Saliency in Point Clouds for Immersive Applications. In *2019 IEEE International Conference on Image Processing (ICIP)* (2019), IEEE, pp. 4325–4329.

[5] B. Adhanom, I., Lee, S. C., Folmer, E., and MacNeilage, P. GazeMetrics: An Open-source Tool for Measuring the Data Quality of HMD-based Eye Trackers. In *ACM Symposium on Eye Tracking Research and Applications* (2020), pp. 1–5.

[6] David, E. J., Gutiérrez, J., Coutrot, A., Da Silva, M. P., and Callet, P. L. A Dataset of Head and Eye Movements for 360° Videos. In *Proceedings of the 9th ACM Multimedia Systems Conference* (2018), pp. 432–437.

[7] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *2nd International Conference on Knowledge Discovery and Data Mining* (1996), pp. 226–231.

[8] Gautier, G., Mercat, A., Fréneau, L., Pitkänen, M., and Vanne, J. UVG-VPC: Voxelized Point Cloud Dataset for Visual Volumetric Video-based Coding. In *15th International Conference on Quality of Multimedia Experience (QoMEX)* (2023), IEEE, pp. 244–247.

[9] Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., and Tabatabai, A. An Overview of Ongoing Point Cloud Compression Standardization Activities: Video-based (V-PCC) and Geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing 9* (2020), e13.

[10] Hanna, M. G., Ahmed, I., Nine, J., Prajapati, S., and Pantanowitz, L. Augmented Reality Technology Using Microsoft HoloLens in Anatomic Pathology. *Archives of Pathology & Laboratory Medicine 142*, 5 (2018), 638–644.

[11] ITU. Subjective Test Methodologies for 360º Video on Head-Mounted Displays, Recommendation ITU-T P.919 (10/2020), 2020.

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

[13] Javaheri, A., Brites, C., Pereira, F., and Ascenso, J. Subjective and Objective Quality Evaluation of Compressed Point Clouds. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)* (2017), pp. 1–6.

[14] Lin, W., and Kuo, C.-C. J. Perceptual Visual Quality Metrics: A Survey. *Journal of Visual Communication and Image Representation 22*, 4 (2011), 297–312.

[15] Mahdi, A., Su, M., Schlesinger, M., and Qin, J. A Comparison Study of Saliency Models for Fixation Prediction on Infants and Adults. *IEEE Transactions on Cognitive and Developmental Systems 10*, 3 (2017), 485–498.

[16] Moro, C., Phelps, C., Redmond, P., and Stromberga, Z. HoloLens and Mobile Augmented Reality in Medical and Health Science Education: A Randomised Controlled Trial. *British Journal of Educational Technology 52*, 2 (2021), 680–694.

[17] MPEG. Common Test Conditions for G-PCC. *ISO/IEC JTC1/SC29/WG7 N722* (2021).

[18] MPEG 3DG. JPEG Pleno Point Cloud Coding Common Test Conditions v3.6. *ISO/IEC JTC1/SC29/WG1 N91058* (2021).

[19] Nguyen, M., Vats, S., van Damme, S., van der Hooft, J., Vega, M. T., Wauters, T., De Turck, F., Timmerer, C., and Hellwagner, H. Characterization of the Quality of Experience and Immersion of Point Cloud Videos in Augmented Reality Through a Subjective Study. *IEEE Access 11* (2023), 128898–128910.

[20] Nguyen, M., Vats, S., Van Damme, S., Van Der Hooft, J., Vega, M. T., Wauters, T., Timmerer, C., and Hellwagner, H. Impact of Quality and Distance on the Perception of Point Clouds in Mixed Reality. In *15th International Conference on Quality of Multimedia Experience (QoMEX)* (2023), IEEE, pp. 87–90.

[21] Novick, D., and Rodriguez, A. E. A Comparative Study of Conversational Proxemics for Virtual Agents. In *13th International Conference on Virtual, Augmented and Mixed Reality* (2021), Springer, pp. 96–105.

[22] Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., and Lefohn, A. Towards Foveated Rendering for Gaze-tracked Virtual Reality. *ACM Transactions on Graphics (TOG) 35*, 6 (2016), 1–12.

[23] Salvucci, D. D., and Goldberg, J. H. Identifying Fixations and Saccades in Eye-tracking Protocols. In *2000 Symposium on Eye Tracking Research & Applications* (2000), pp. 71–78.

[24] Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., and Xu, X. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS) 42*, 3 (2017), 1–21.

[25] Schwarz, S., Preda, M., Baroncini, V., Budagavi, M., Cesar, P., Chou, P. A., Cohen, R. A., Krivokuća, M., Lasserre, S., Li, Z., Llach, J., Mammou, K., Mekuria, R., Nakagami, O., Siahaan, E., Tabatabai, A., Tourapis, A. M., and Zakharchenko, V. Emerging MPEG Standards for Point Cloud Compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9*, 1 (2019), 133–148.

[26] Seufert, M., Egger, S., Slanina, M., Zinner, T., Hossfeld, T., and Tran-Gia, P. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys & Tutorials 17*, 1 (2014), 469–492.

[27] Subramanyam, S., Li, J., Viola, I., and Cesar, P. Comparing the Quality of Highly Realistic Digital Humans in 3DOF and 6DOF: A Volumetric Video Case Study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), IEEE, pp. 127–136.

[28] Subramanyam, S., Viola, I., Hanjalic, A., and Cesar, P. User Centered Adaptive Streaming of Dynamic Point Clouds with Low Complexity Tiling. In *28th ACM International Conference on Multimedia* (2020), pp. 3669–3677.

[29] Taraghi, B., Nguyen, M., Amirpour, H., and Timmerer, C. INTENSE: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming. *IEEE Access 9* (2021), 118087–118098.

[30] Van Damme, S., Mahdi, I., Kumar Ravuri, H., van der Hooft, J., De Turck, F., and Torres Vega, M. Immersive and Interactive Subjective Quality Assessment of Dynamic Volumetric Meshes. In *15th International Conference on Quality of Multimedia Experience (QoMEX)* (2023), IEEE.

[31] van der Hooft, J., Torres Vega, M., Timmerer, C., Begen, A. C., De Turck, F., and Schatz, R. Objective and Subjective QoE Evaluation for Adaptive Point Cloud Streaming. In *12th International Conference on Quality of Multimedia Experience (QoMEX)* (2020), IEEE.

[32] Van Holland, L., Stotko, P., Krumpen, S., Klein, R., and Weinmann, M. Efficient 3D Reconstruction, Streaming and Visualization of Static and Dynamic Scene Parts for Multi-client Live-telepresence in Large-scale Environments. In *2023 IEEE/CVF International Conference on Computer Vision Workshops* (2023), pp. 4260–4274.

[33] Vats, S., Nguyen, M., Van Damme, S., van der Hooft, J., Vega, M. T., Wauters, T., Timmerer, C., and Hellwagner, H. A Platform for Subjective Quality Assessment in Mixed Reality Environments. In *15th International Conference on Quality of Multimedia Experience (QoMEX)* (2023), pp. 131–134.

[34] Zerman, E., Ozcinar, C., Gao, P., and Smolic, A. Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression. In *12th International Conference on Quality of Multimedia Experience (QoMEX)* (2020), IEEE, pp. 1–6.

[35] Zhang, W., and Liu, H. Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications. *IEEE Transactions on Image Processing 26*, 5 (2017), 2424–2437.

[36] Zhou, X., Viola, I., Alexiou, E., Jansen, J., and Cesar, P. QAVA-DPC: Eye-Tracking Based Quality Assessment and Visual Attention Dataset for Dynamic Point Cloud in 6 DoF. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2023), IEEE.