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

Minh Nguyen§, Shivi Vats§, Xuemei Zhou, Irene Viola, Pablo Cesar, Christian Timmerer, and 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 Compressed Point cloud dataset with Eye-tracking and Quality assessment in Mixed Reality (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/.

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:

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 Compressed Point cloud dataset with Eye-tracking and Quality assessment in Mixed Reality (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.

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

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

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.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Snapshot</th>
</tr>
</thead>
<tbody>
<tr>
<td>BlueSpin</td>
<td>A person wearing a blue t-shirt and spinning at a consistent rate.</td>
<td><img src="image" alt="BlueSpin" /></td>
</tr>
<tr>
<td></td>
<td>SI: 20.8</td>
<td>SI: 8.0</td>
</tr>
<tr>
<td></td>
<td>Colorfulness: 8.6</td>
<td></td>
</tr>
<tr>
<td>CasualSquat</td>
<td>A person wearing a striped shirt and jeans in the performance of a squat exercise.</td>
<td><img src="image" alt="CasualSquat" /></td>
</tr>
<tr>
<td></td>
<td>SI: 53.5</td>
<td>TI: 19.0</td>
</tr>
<tr>
<td></td>
<td>Colorfulness: 11.5</td>
<td></td>
</tr>
<tr>
<td>FlowerDance</td>
<td>A person in a long, flowing dress spinning and twirling.</td>
<td><img src="image" alt="FlowerDance" /></td>
</tr>
<tr>
<td></td>
<td>SI: 43.9</td>
<td>TI: 22.3</td>
</tr>
<tr>
<td></td>
<td>Colorfulness: 25.3</td>
<td></td>
</tr>
<tr>
<td>ReadyForWinter</td>
<td>A person donning a beanie and scarf.</td>
<td><img src="image" alt="ReadyForWinter" /></td>
</tr>
<tr>
<td></td>
<td>SI: 20.6</td>
<td>TI: 11.5</td>
</tr>
<tr>
<td></td>
<td>Colorfulness: 7.8</td>
<td></td>
</tr>
</tbody>
</table>

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

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 for VPCC and TMC13 version v.23.0 for GPCC. The quality levels are based on the opinion scores.


Table 3: Encoder parameters to generate compressed dynamic PCs.

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

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].

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.


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;


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 gaze. 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]Ω.

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 datasetΩ. 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.

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.

4.2 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 18th and 41st 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 5th 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].
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 telesubscription 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


