Subjective and Objective Quality Assessment for Dynamic Point Cloud with Visual Attention in 6 DoF

XUEMEI ZHOU, Centrum Wiskunde & Informatica, Amsterdam, TU Delft, Delft, The Netherlands
IRENE VIOLA, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
EVANGELOS ALEXIOU, Xiaomi Technology, The Hague, The Netherlands
JACK JANSEN, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
PABLO CESAR, Centrum Wiskunde & Informatica, Amsterdam, TU Delft, Delft, The Netherlands

Perceptual quality assessment of Dynamic Point Cloud (DPC) contents plays an important role in various Virtual Reality (VR) applications that involve human beings as the end user. Understanding and modeling perceptual quality assessment is greatly enriched by insights from visual attention. However, incorporating aspects of visual attention in DPC quality models is largely unexplored, as ground-truth visual attention data are scarcely available. Besides, testing methods and procedures for collecting visual attention data are still to be agreed on. This article presents a dataset containing subjective opinion scores and visual attention maps of DPCs, collected in a VR environment using eye-tracking technology. Both the quality score and eye-tracking data were collected during a subjective quality assessment experiment, in which subjects were instructed to watch and rate DPCs at various degradation levels under 6 Degrees of Freedom (DoF) inspection, using a head-mounted display. Qualitative interview analysis was also conducted after the experiment. The dataset consists of 50 DPCs, including 5 reference DPCs, with each reference encoded at 3 distortion levels using 3 different codecs (namely G-PCC, V-PCC, CWI-PCL), amounting to a total of 9 degraded version per reference. Additionally, it incorporates 1,000 gaze trials from 40 participants, yielding a total of 15,000 visual attention maps across all the DPCs.

We additionally benchmark objective quality metrics originally designed for static point clouds, evaluating their performance in our dataset using two temporal pooling strategies. Furthermore, we employ the visual attention data that are retrieved during our experiment to evaluate whether the performance of widely used objective quality metrics is improved by considering subjective measurements of visual attention. This dataset establishes a link between quality assessment and visual attention within the context of DPC. Moreover, thematic analysis of the interviews helps uncover user behavior and factors impacting perceptual quality for DPC in 6 DoF. This work deepens our understanding of DPC quality assessment and visual attention, driving progress in the realm of VR experiences and perception.

CCS Concepts: • **Human-centered computing** → **Visualization design and evaluation methods**; • **Computing methodologies** → **Perception**; *Model development and analysis*; **Image processing**; **Virtual reality**; **Interest point and salient region detections**;

This work was supported through the NWO WISE grant and the European Commission Horizon Europe program, under the grant agreement 101070109, TRANSMIXR https://transmixr.eu/. Funded by the European Union.
Authors' Contact Information: Xuemei Zhou (corresponding author), Centrum Wiskunde & Informatica, Amsterdam, TU Delft, Delft, The Netherlands; e-mail: xuemei.zhou@cwi.nl; Irene Viola, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands; e-mail: irene.viola@cwi.nl; Evangelos Alexiou, Xiaomi Technology, The Hague, The Netherlands; e-mail: alexiou@xiaomi.com; Jack Jansen, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands; e-mail: jack.jansen@cwi.nl; Pablo Cesar, Centrum Wiskunde & Informatica, Amsterdam, TU Delft, Delft, The Netherlands; e-mail: p.s.cesar@cwi.nl.

ACM 1551-6865/2025/8-ART230
https://doi.org/10.1145/3731759

Additional Key Words and Phrases: Volumetric video, Dynamic point cloud, Visual saliency, Visual attention, Subjective quality assessment, Objective quality metrics, Eye tracking, 6DoF

ACM Reference format:
Xuemei Zhou, Irene Viola, Evangelos Alexiou, Jack Jansen, and Pablo Cesar. 2025. Subjective and Objective Quality Assessment for Dynamic Point Cloud with Visual Attention in 6 DoF. *ACM Trans. Multimedia Comput. Commun. Appl.* 21, 8, Article 230 (August 2025), 24 pages.
https://doi.org/10.1145/3731759

1 Introduction

Volumetric video has become available to represent real-world objects due to the rapid development of capture devices, transmission technologies, and computational capabilities. Point cloud has emerged as one of the most popular formats for volumetric video representation. Specifically, **Dynamic Point Clouds (DPCs)** can be used for automotive/robotic navigation [74], medical imaging [9], and virtual video conferencing [30, 61], among other applications. A point cloud frame can be defined as a set of points in space represented in a 3D coordinate system. Each point cloud frame requires a large number of points to faithfully represent the content and achieve a good **Quality of Experience (QoE)**. A DPC is essentially a sequence of individual point cloud frames played in succession. Therefore, effective compression is essential before transmission, storage, rendering, and display. Quality degradation will be inevitably introduced during this end-to-end pipeline, which deteriorates the visual quality and affects human perception. Exploring the distortion characteristics of DPCs and effectively measuring them is a challenge in both subjective and objective quality assessment [3].

Subjective quality assessment leads to ground-truth ratings for visual impairments that appear in a stimulus. Subjective quality assessment for DPC has been explored in passive desktop viewing conditions [70, 71] or in immersive environments with users consuming the contents through a **Head-Mounted Display (HMD)** under 6 **Degrees of Freedom (DoF)** [53, 63]. In the latter case, information about users' movements can be captured in addition to subjective quality ratings, to understand how users navigate and observe objects in **Virtual Reality (VR)** space. A more accurate representation of the user's consumption is given by gaze data, which highlight the specific areas of content being viewed with focused attention. This information aids in the creation of visual attention maps. Incorporating visual attention into quality assessment has demonstrated potential improvement for predicting the visual quality of 2D/3D image/video [33, 73]. Nonetheless, visual attention for DPCs is unexplored, thus hindering the utilization of its outcomes in aiding visual quality assessment.

A summary of existing subjective quality assessment and visual attention datasets for point clouds is shown in Table 1. Most of the studies in the literature involving DPCs are conducted in desktop setups with 2D monitors, where the DPCs are pre-recorded and the playback is conducted using conventional video software [68, 70, 71]. However, such desktop setups limit user interactions (if any) to mouse and keyboard, restricting user freedom. In contrast, in immersive HMD-based setups, a more natural consumption with 6 DoF through user movements is enabled, allowing for a more realistic representation of the entire DPC. The technical challenges faced with HMD-based setups (e.g., real-time playback of DPCs [53]), though, lead to experimental sessions that typically involve a smaller number of DPCs (20), usually static or with shorter time duration (5 seconds). Maintaining the accuracy of the eye-tracker in 6 DoF is yet another challenge [40]. Due to such constraints, no visual attention dataset specifically designed for DPCs consumed by HMD-based setups in 6 DoF has been released so far; existing research has primarily explored the attention of

Table 1. Publicly Available Subjective Quality Assessment and Visual Attention Datasets for Point Clouds

Dataset	Type	Degradation	Stimuli	Time	Display	Interaction	Opinion Score	Visual Attention
VsenseVVDB [70]	Dynamic	Down-sampling, V-PCC	32	6.6 s	2D monitor	✗	✓	✗
VsenseVVDB2 [71]	Dynamic	*Mesh*: Draco+JPEG *Point Clouds*: G-PCC, V-PCC	28 136	10 s	2D monitor	✗	✓	✗
Owlii [68]	Dynamic	*Mesh*: TFAN, FFmpeg *Point Clouds*: V-PCC, FFmpeg	20	20 s	2D monitor	✗	✓	✗
Marvie et al. [37]	Dynamic	Position, Texture coordinate HEVC, Triangle holes	176	10 s	2D monitor	✗	✓	✗
VOLVQAD [12]	Dynamic	V-PCC	376	10 s	2D monitor	✗	✓	✗
Subramanyam et al. [53]	Dynamic	CWI-PCL, V-PCC	72	5 s	HMD	✓	✓	✗
ComPEQ-MR [42]	Dynamic	V-PCC, G-PCC	52	10 s	Augmented reality	✓	✓	✓
ViAtPCVR [5]	Static	Only reference	8	-	HMD	✓	✗	✓
QAVA-DPC (Ours)	Dynamic	V-PCC, G-PCC, CWI-PCL	50	10 s	HMD	✓	✓	✓

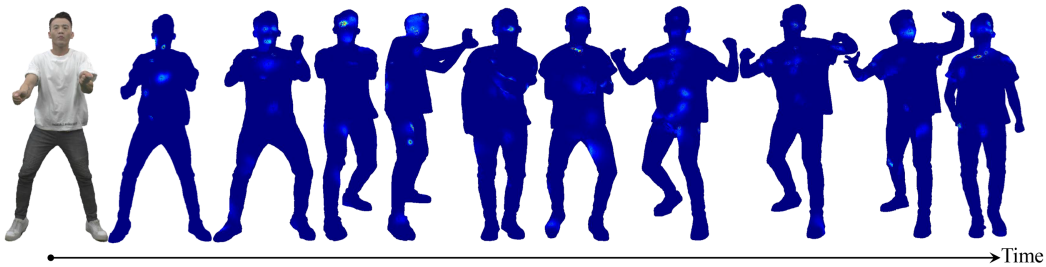


Fig. 1. Fixation maps of *dancer* sequence with uniform temporal sampling every 30 frames.

static point clouds [5], confining the scope to a few undistorted contents. More generally, there is currently a lack of studies that connect visual attention and visual quality specifically for DPC.

Visual attention plays a crucial role in various vision tasks, such as segmentation, localization, and registration [22]. Notably, leveraging visual attention maps to weight quality maps has shown improvements in perceptual quality prediction [35]. By connecting visual attention and visual quality for DPCs, quality allocation between salient regions and surrounding areas, saliency-aware compression and streaming, and saliency-aided objective quality metrics can be further investigated and optimized.

In our previous paper [75], we created an eye-tracking-based **Quality Assessment and Visual Attention dataset for DPCs (QAVA-DPC)**. The dataset includes diverse content and various compression distortions using the MPEG standard codecs: **Video-Based Point Cloud Compression (V-PCC)**, **Geometry-Based Point Cloud Compression (G-PCC)**, and the MPEG reference codec, referred to here as CWI-PCL. With this work, we aim to extend our previous efforts by conducting an in-depth analysis of semi-structured interviews conducted after the subjective experiments and by benchmarking objective point cloud quality assessment metrics with collected visual attention data. The added value of associated visual attention maps and results from the semi-structured interviews can thereby enhance our understanding of human behavior within 6 DoF environments, ultimately contributing to the optimization of QoE. Our contributions can be summarized as follows:

— We propose a new dataset, namely, QAVA-DPC, which contains five reference DPCs; each DPC is encoded by three codecs, with each codec configured at three distortion levels. Fixation maps are constructed, collected, and presented for both the reference and distorted sequences as heatmaps overlaid on top of the stimuli frames. To the best of our knowledge, this is the first time connecting visual attention and visual quality for DPCs in VR.

— We benchmark state-of-the-art metrics, initially intended for static point clouds, along with two temporal pooling methods on the QAVA-DPC dataset. In addition, we validate the performance of these metrics using ground-truth visual attention maps.
— We release all raw data, containing the opinion scores and gaze samples collected in our study, alongside the software used to perform the experiment, and the scripts used to export visual attention maps, at the following link: https://github.com/cwi-dis/ISMAR_PointCloud_EyeTracking.

2 Related Work

2.1 Subjective Quality Assessment for DPC

Whereas subjective quality assessment of static point clouds has been explored in more detail in the literature [4], analogous research on DPCs is still a sophisticated and challenging problem, owing to numerous factors such as the evaluation methodology, rendering method, display equipment, and so forth. Subjective quality scores, such as **Mean Opinion Score (MOS)** or **Differential MOS (DMOS)**, are commonly used to quantify the subjective perception of visual artifacts. Zerman et al. [70] conduct a subjective experiment on two DPCs from the VsenseVVDB dataset that are distorted using V-PCC compression [49]. They argue that certain geometric distortion metrics are incongruent with the expected quality. Hooft et al. investigate how and to what extent various aspects impact the user's QoE, via extensive subjective evaluation of volumetric 6 DoF streaming [58]. Mekuria et al. evaluate the subjective quality of the CWI-PCL codec performance using a realistic 3D tele-immersive system in a virtual room scenario, in which users are represented and interact as 3D avatars and/or 3D point clouds [38]. The subjective study shows that introduced prediction distortions are negligible compared with the original reconstructed point clouds. Cao et al. [8] study the perceptual quality of compressed 3D sequences, for both point cloud compression and mesh-based compression. They explore the impact of bit rate and observation distance on perceptual quality. Cox et al. [12] present VOLVQAD, a volumetric video quality assessment dataset with 376 video sequences. The volumetric video sequences are first encoded with MPEG V-PCC using 4 different avatar models and 16 quality variations, and then rendered into test videos for quality assessment using 2 different background colors and 16 different quality switching patterns with a 2D display. However, these experiments are performed in a desktop setup. Viola et al. [63] compare two different VR viewing conditions enabling 3/6 DoF, along with a desktop setup, to understand how interaction in the virtual space affects the perception of quality. Results show no statistical difference between scores given in a desktop and VR setup; however, qualitative results highlighted the added value of interactive evaluations. One limitation of the study lies in the time duration (5 seconds) of the sequences used for the evaluation, as the authors use 150 frames. Subramanyam et al. [54] evaluate the performance of several adaptive streaming solutions in an interactive VR experiment. They compare the performance of V-PCC with respect to CWI-PCL, using various adaptive streaming strategies. Quantitative subjective results and qualitative insights indicate that V-PCC has a more favorable performance than the CWI-PCL, especially at low bit rates. Damme et al. [56] conducted an in-depth subjective study on the impact of converting point clouds to meshes with varying-quality representations. Additionally, while end-users demonstrate awareness of quality switches, the effect on their perception remains limited. Gutiérrez et al. [20] present a subjective study on DPCs considering different compression rates using the MPEG standard V-PCC. Results on users' exploration behavior show no significant differences when visualizing point clouds with different qualities, no changes in the behavior during the test session, and no correlation between exploration activity and quality assessments. Nguyen et al. [42] provide an open source compressed point cloud dataset with eye-tracking data and quality assessment

score in mixed reality with Hololens 2, including 4 DPCs. Eye-tracking data and opinion scores are collected under different experimental settings. While numerous studies have explored the subjective quality assessment of DPCs, there is a gap in research focusing on visual attention and quality assessment in VR with 6 DoF.

2.2 Objective Quality Assessment for DPC

Objective quality assessment of 2D/3D video has achieved remarkable progress in recent years. However, few specific objective quality metrics have been designed for DPCs so far. Ak et al. [1] explore the possibility of temporal sub-sampling of the content under evaluation for objective quality evaluation without sacrificing the correlation with the subjective opinion. Thirty different objective quality metrics are tested on the VsenseVVDB2 dataset, combined with temporal sub-sampling and temporal pooling methods. Results show that the performance of objective metrics for point cloud compression is minimally affected by the temporal sub-sampling rate. Freitas et al. [18] investigate the added value of incorporating temporal pooling into the DPC's quality assessment model using metrics designed for static point clouds. They find that the performance of temporal pooling is consistently better when a temporal variation model [43] is used. The same authors investigate the suitability of geometric-aware texture descriptors to blindly assess the quality of colored DPCs [19] on top of the same temporal pooling strategy, leading to similar conclusions. Yang et al. [69] conduct a subjective user study to understand the effectiveness of different perceptual quality metrics for volumetric video and design an objective metric called Volu-FMAF. Volu-FMAF combines point-based and pixel-based metrics with viewpoint-related features. They further propose a distortion-aware rendered image super-resolution network in a volumetric video streaming framework, which exploits the insights obtained from their user study. Damme et al. [57] present a thorough correlation analysis of both full-reference and no-reference objective metrics to subjective MOS with a double purpose. Additionally, they investigate how region of interest selection and weighting procedures impact the accuracy to enhance it further. The study shows that the classical video quality metric VMAF [32] is very well suited as an objective benchmark for volumetric media streaming in terms of correlation to subjective scores, and a combination of no-reference features could provide a good real-time assessment. Fan et al. [16] propose a deep-learning-based no-reference volumetric video quality assessment method based on multi-view learning. They first project volumetric videos to 2D video sequences from various viewpoints. Then, ResNet 3D is utilized to extract quality-aware features, and a quality regression module is designed to fuse the features learned from the multiple viewpoints and jointly regress them into quality scores. Marvie et al. [37] benchmark and calibrate several objective quality metrics on a challenging volumetric video dataset represented as textured meshes. They evaluate two model-based approaches—MPEG PCC and PCQM—by converting meshes into point clouds using a devised mesh surface sampling method, as well as an image-based approach (IBSM), for which they introduce two new features specifically designed to detect holes and temporal defects. For each metric, the optimal selection and combination of features are determined by logistic regression through cross-validation. The performance analysis, combined with MPEG experts' requirements, leads to recommendations on the features of highest importance through learned feature weights, such as temporal pooling, integrating an attention model.

2.3 Visual Attention-Based Objective Quality Assessment

Recent literature in eye-tracking-based visual saliency for immersive contents has mainly focused on task-free experiments to gather visual attention maps [45, 51]; no study has been conducted to link visual attention to visual quality assessment for volumetric videos. The literature suggests that visual attention might be beneficial for understanding the process of perception of visual

quality for 2D images/videos; in fact, different metrics for **Image Quality Assessment (IQA)** have been extended with a computational model of visual attention [33], but the resulting gain on the metrics' performance is so far unclear. To better understand the added value of including visual attention in the design of objective metrics for 2D images, some works in the literature have taken advantage of recorded visual attention data. Lin and Heynderickx [34] perform two eye-tracking experiments: one with a free-looking task and one with a quality assessment task. They found a tendency that adding saliency to a metric yields a larger amount of gain in performance. The extent of the performance gain tends to depend on the specific objective metric and the image content. In addition, the gain is small for objective metrics that already show a high correlation with perceived quality for a given distortion type. Zhang and Liu [73] propose a new methodology to eliminate the inherent bias due to the involvement of stimulus repetition. The refined methodology results in a new eye-tracking dataset with a large degree of stimulus variability. Based on ground-truth labeling, the statistical evaluation shows that the visual attention information of both the reference and the distorted scene is beneficial for IQA metrics, but the latter tends to further boost the effectiveness of integrating attention in IQA metrics. Jin et al. [27] utilize an eye-tracker to create foveation-compressed VR datasets and evaluate both the foveated and non-foveated objective image/video quality assessment algorithms.

To better understand whether the findings regarding visual saliency and quality assessment on 2D images/videos can hold for volumetric videos, ad-hoc datasets, and benchmarking validation for objective metrics that combine the two aspects are needed. That is the research gap we aim to fill with this article.

In this study, our objective is to enrich the existing literature by conducting a subjective experiment that compares the visual quality of several state-of-the-art compression techniques for DPC. This experiment is interactive, using an HMD-based VR rendering of 10-second DPC sequences from various datasets, a methodology not previously explored in conjunction with eye-tracking in the literature. In addition, we delve into the analysis of interviews conducted during the subjective experiment to gain deeper insights into user experiences and the factors that influence the perceptual quality of DPCs. Furthermore, we validate widely used objective metrics and pooling methods reported in the literature for DPC objective quality assessment.

3 QAVA-DPC Construction

3.1 Content Selection

For the creation of the dataset, we selected five DPCs from three public datasets, namely VsenseVVDB2 [71], 8i [14], and Owlii [68]. To show the diversity of DPCs, we considered the **Spatial Information (SI)** and **Temporal Information (TI)** for each content [24]. We projected the source point cloud into four views, which are the left, right, front, and back view, of its bounding box to apply SI and TI separately [67], then obtain the maximum value among the four views over all the first 300 frames as the final SI/TI for one sequence. Their distribution for all DPCs can be seen in Figure 2. The dispersion in SI (horizontal axes)/TI (vertical axes) shows the diversity of our contents in the spatial/temporal domain. We finally chose *dancer*, *exercise*, *long dress*, *Rafa2*, and *soldier* as the contents of our dataset.

3.2 Stimuli Processing

Prior to the subjective experiment on DPCs, specific procedures, such as pre-processing, encoding, and rendering, are necessary due to codec implementations, with the goal of minimizing additional influencing factors.

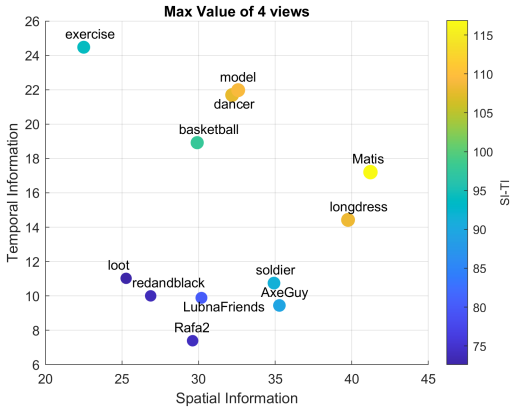


Fig. 2. Distribution of SI and TI of 12 source DPCs from three datasets, the color value is computed by $\sqrt{(SI^2 + TI^2)}$. SI, Spatial Information; TI, Temporal Information.

3.2.1 Pre-Processing. The sequences mentioned above are selected from different datasets, which means that the resolution, position, and orientations vary. To create a realistic tele-immersive scenario, the DPCs should be displayed as life-size models. To do so, we normalized the DPCs to a similar bounding box. For *dancer* and *exercise*, the geometry precision is voxelized from 11 to 10 bits to ensure consistency with the other contents. The source models were processed with rotation, translation, and scaling. Additionally, since the V-PCC encoder operates with integer values, the coordinates of all DPCs were rounded before V-PCC compression. CWI-PCL encoder has specific requirements for the resolution of DPCs, so before CWI-PCL compression, the coordinates went through a scaling operation.

3.2.2 Encoding. Distorted versions were generated using the state-of-the-art MPEG PCC reference software Test Model Category 2 Version 18 (TMC_2V-18.0) and Category 1&3 Version 14 (TMC_{13}V-14.0) [49], which implement V-PCC and G-PCC, respectively. We also adopt the CWI-PCL [38] codec as a comparison, which served as a base reference software platform in MPEG. G-PCC was originally proposed to compress static point clouds, although more recent versions are capable of handling DPCs; it operates in the 3D, point cloud domain. V-PCC was developed for DPC compression; it operates in the 2D, video domain. CWI-PCL was designed as a lightweight codec that complies with real-time requirements; the geometry is encoded in the 3D and the attributes in the 2D domain. To compare them in a fair way, we set the G-PCC encoder to use **Region-Aptive Hierarchical Transform (RAHT)** and Octree for compressing color attributes and geometry, respectively; the V-PCC encoder to use **All Intra (AI)** mode, which applies intra-prediction to all frames; and the CWI-PCL encoder to use intra-prediction in all frames, with octree subdivision and JPEG codec to compress geometry and color attributes, respectively.

To define the configuration parameters for the encoders, the MPEG **Common Test Conditions (CTC)** [52] are followed. Moreover, we select three distortion levels that represent comparable low, medium, and high-quality ranges for each encoder. Specifically, for G-PCC, we select R02, R04, and R05 from the MPEG CTC, which are realized by adjusting the *positionQuantizationScale* and the QP parameters. For V-PCC, we select R01, R03, and R05 by adjusting the geometry QP, the attribute QP, and the *occupancyPrecision* parameters. For CWI-PCL, we choose three combinations of octree depth with JPEG QP parameters to match a similar quality range, by looping over octree depth from 7 to 9 and JPEG QP from 25 to 95 (step size = 10). When tested on the dataset, the specified parameter settings for the three codecs yielded subjectively similar quality ranges. The specific

Table 2. Parameter Sets for the Selected Encoders

Encoders	Distortion Level		
G-PCC (Octree-RAHT)	R02 (0.125, 46)	R04 (0.5, 34)	R05 (0.75, 28)
V-PCC (AI)	R01 (32,42, 4)	R03 (24, 32, 4)	R05 (16, 22, 2)
CWI-PCL	R01 (7, 25)	R02 (8, 95)	R03 (9, 95)

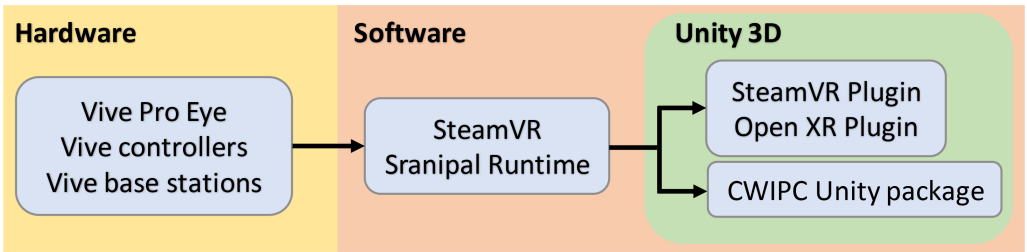


Fig. 3. Schematic diagram with the hardware and software modules together with their inter-dependencies.

parameter settings are shown in Table 2. Each reference DPC is compressed using 3 encoders, and each encoder has 3 distortion levels, for a total of 45 distorted DPCs.

3.2.3 Rendering. Rendering is the process of producing a visual representation that can be consumed by users using an available display. In the case of point clouds, different rendering methods may have a significant impact on perceived quality [26]. In our experiment, we chose to render the point clouds without any additional processing (e.g., surface reconstruction), directly using the point cloud data (point-based).

Our experiment software is developed in Unity (version 2021.3.10.f1), exploiting the SteamVR plugin (version 1.24.7) to connect with VR headsets and controllers. **CWI Point Cloud (CWIPC)**-supported unity package (version 0.10.0) helps us import the DPCs and play them back inside Unity [46]. A high-level diagram indicating the hardware/software dependencies is provided in Figure 3. Notably, a large size of DPC file might take up too much memory and cause a system hang. We converted the DPC data into the CWIPC-supported point cloud playback format to enhance software stability without introducing additional distortion. To ensure smooth playback of DPCs, we took advantage of the Unity Coroutine scheme to preload each DPC into memory before the user switches to the next DPC. It should be noted that for each DPC sequence, we only chose the first 300 frames from the source model. The frame rate for playback was set to 30 frames per second, hence, each DPC sequence lasts for 10 seconds.

For the same stimuli, both the reference and distorted versions were made to appear watertight by adjusting the point size based on the average distance to each point's 10 nearest neighbors across the entire point cloud [53]. Within a DPC, we utilized the same point size for all frames. All the point clouds were rescaled to a similar size, around 1.8 m in height, to mimic a realistic tele-immersive scenario. The VR scene was illuminated by a virtual lamp on the ceiling, centralized above the models. The lamp was set as an area light with intensity values of 2 in Unity to simulate ordinary lighting in a room. We used HTC Vive Pro Eye devices with eye-tracking capabilities and Vive hand controllers for participants to interact in our experiment. To develop eye-tracking

Fig. 4. Playback scene.

applications for the Vive Pro Eye, we used the native HTC Vive SRanipal SDK. The sampling frequency (binocular) of the eye tracker was 120 HZ.

3.3 Experimental Procedure

Since there is no specific recommendation for designing subjective quality assessment experiments for DPCs in VR, we have made an effort to adhere to existing ITU recommendations on images/videos [21, 23, 25] to establish an appropriate assessment methodology for DPCs. In our subjective study, we opted for the Absolute Category Rating with **Hidden Reference (HR)** using five-level ratings (*1-Bad, 2-Poor, 3-Fair, 4-Good, and 5-Excellent*). Each time, only a single DPC was shown to the observer; test materials included impaired DPCs with randomly inserted intact HRs, represented as any other test stimulus. To avoid the effects of contextual or memory comparisons, we randomly generated a playlist for each subject, and care was given to avoid displaying the same content consecutively.

Before the experiment, the visual acuity and color vision of every subject were tested using Snellen [17] and Ishihara [11] charts. Each subject was informed in advance about the manner and purpose of the study as part of the informed consent procedure. At the beginning of the session, the inter-pupillary distance was measured and the headset was adjusted by the subject accordingly. Then, a training session was conducted to help familiarize the subjects with the system, including the controllers and the naming of each button to communicate more easily. One training sequence, namely *loot*, was used, which was not included in the dataset. The quality range of *loot* was similar to the quality range of the test videos, giving the subjects a sense of what they would see in the formal sessions. The subjects always started at the same location, which is 1.5 m away from the center of the virtual room, but could move freely from there onward.

A DPC was located in the center of the virtual room, as shown in Figure 4, and each DPC was randomly rotated between $[0°, 360°]$ to avoid bias. During the experiment, the subjects were instructed to view each model carefully in the VR environment, by moving freely during the playback of each DPC. The subjects were also required to stand still while doing the calibration and error profiling. A fixed distance was set between the subjects and the error profiling scene, which was a circle composed of nine eye-ball markers.

After feeling comfortable with the set-up, the participants were informed about the task that was assigned to them: "we ask you to examine a set of human DPC models, each model will be looped three times, and each loop lasts for 10 seconds; after visualization, we will ask you to rate the quality of the stimuli you are looking at, and at the same time, we will record your gaze-related data." To determine the number of loops, we referred to related papers on video quality assessment and eye-tracking-based visual saliency detection [13, 31, 59, 72]. Furthermore, in [44], the effect of exposition time was explored by repeating the same video from one to four loops, concluding that more loops do not necessarily result in more unique fixation points for most videos. Hence, we chose three loops instead of one or an unlimited number. There were two dummy objects at the beginning of each session, without participants being aware, to familiarize them with the testing procedure and the rating scale. For each subject, the test was split into two rounds, lasting around 30 minutes each, with a mandatory 5-minute break in between. Before and after each round, participants were requested to fill in a **Simulator Sickness Questionnaire (SSQ)** on a 1–4 discrete scale (1 = none to 4 = severe) [28]. After participants finished the two rounds, the researchers conducted a semi-structured interview. The interview was performed individually in a non-VR setting, and the entire conversation was recorded for analysis purposes.

For every model and subject, a round was split into four consecutive steps:

1. *Calibration* is the process of determining the unique geometric features of a participant's eyes, which serves as the foundation for precise gaze point calculations. This process optimizes the eye-tracking algorithm for accuracy. Calibration was done at the beginning of the experiment, and only when calibration was successful, the users could enter into the DPC playback stage.
2. *Inspection of models* is the step where the participants are visualizing a DPC, while their trajectory and gaze-related information are recorded.
3. *Quality evaluation of models* requires the subject to rate a DPC. The rating button was marked with labels ranging from "Poor" to "Excellent" to facilitate anchoring the rating process, and subjects could use their controllers to select and submit a score without removing the headset.
4. *Error profiling* is issued as the last step to estimate the accuracy of the gaze measurements due to calibration inaccuracies, or HMD displacements. A regular circle of nine markers at pre-defined positions in the virtual scene was presented to the user. Based on the recorded gaze measurements, the average angular error was computed per marker in real time. This procedure allowed us to decide whether the gaze data obtained from a certain session was accurate or not.

A total of 40 participants took part in the subjective tests of this study, with a diverse composition that includes 1 non-binary individual, 19 males, and 20 females. The participants' ages ranged from 20 to 34, with an average age of 26.90 and a standard deviation of 3.51. Each participant observed half of the DPCs among all stimuli, leading to 20 opinion scores per sequence. In terms of occupation, the majority (80%) of the participants were students, ranging from bachelor to PhD levels. The remaining 20% were researchers, postdoctoral fellows, and one auditor. Regarding familiarity with VR devices, 7 participants had never experienced VR before the experiment, 26 participants had intermediate experience (using VR 1 to 3 times), and 7 were considered experts, having backgrounds as VR designers or researchers. Additionally, 22 out of 40 participants wore glasses during the experiment.

3.4 Data Processing

After collecting raw data from participants during the subjective quality assessment experiments described in Section 3.3, we process the most relevant data to deepen our understanding of the

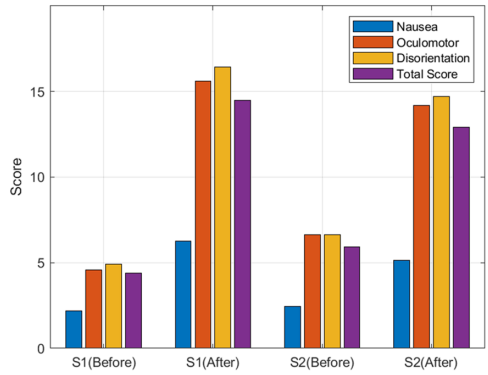


Fig. 5. SSQ score for two test sessions.

relationship between gaze behavior and perceived quality of DPCs while ensuring data integrity and reliability. This involves analyzing SSQ responses to evaluate the physiological impact of the VR experience, processing opinion scores to quantify subjective quality ratings, and refining gaze data to generate accurate saliency maps for further analysis.

3.4.1 Processing of SSQ Data. SSQ comprises 16 symptoms, which are further grouped into three different categories: Oculomotor, Nausea, and Disorientation; we also compute the total score. Figure 5 suggests that simulator scores are increasing after performing the experiment. However, breaks typically help in reducing simulator sickness.

3.4.2 Processing of Opinion Scores. After removing the scores of the first two dummy objects, outlier detection was performed according to the ITU-T Recommendation P.913 [25]. The suggested threshold values $r_1 = 0.75$ and $r_2 = 0.8$ were used. No outliers were found in our test. After outlier detection, the MOS was computed for each DPC, and the associated 95% **Confidence Intervals (CIs)** were obtained assuming a Student's t-distribution. Additionally, the DMOS was calculated after applying HR removal, following the procedure described in the ITU-T Recommendation P.913 [25].

3.4.3 Processing of Gaze Data. One subject walked into the body of two DPCs in the VR environment while observing them, thus, the corresponding gaze data were not included. We ignored the initial 400 ms gaze data of each user to avoid unintentional gaze because of the unexpected appearance of the DPC. Then, only the valid gaze samples provided by the SRanipal SDK were selected. Each valid gaze sample was processed as follows:

(1) *Verify the Validity of Gaze Data*: We used GazeMetrics [7], an open source tool for measuring the data quality of HMD-based eye trackers. A barycentric interpolation with weights equal to corresponding angular errors obtained from the error profiling step for the corresponding model was applied. During error profiling, a threshold of 7.5° was used to discard unintentional gaze. Moreover, after displaying each target, 0.8 seconds were waited before the samples were included in actual calculations. This delay accounts for calibration stabilization, participant adaptation, and gaze analysis during fully engaged periods [50]. Finally, we applied a compensatory weighted average angular error to each gaze sample. This was repeated for every user gaze sample to maintain high-quality estimations while avoiding discarding useful data. Figure 6 illustrates the estimation of angular error for gaze data in 2D, g represents the intersection between the gaze ray and the plane formed by nine markers denoted as $m1$ to $m9$. These markers were positioned at a distance of 1.25 m relative to the camera within the VR environment.

(2) *Identifying Fixation Points from Gaze Data*: Taking into account the dynamic nature of our content, we chose the **Dispersion-Threshold Identification (I-DT)** [47] method. I-DT leverages the fact that fixation points, owing to their reduced velocity, tend to cluster in close proximity [65]. It identifies fixation points as groups of consecutive points within a particular dispersion, or maximum separation. The I-DT algorithm requires two parameters, the dispersion threshold and the duration threshold. We set the dispersion threshold equal to 3° and the duration threshold equal to 100 ms, respectively. For consecutive gaze samples that meet the dispersion threshold, we considered their average coordinates within the duration threshold as the coordinates of the fixation point.

(3) *Mapping Gaze Data to DPC Frames*: We proceeded by associating the valid gaze data, with each DPC frame and translating the gaze data (x, y, z) that correspond to a DPC frame from world space into fixation points. As a result, we got gaze data for 300 frames in total. We adopted the truncated-cone-sector algorithm to assign weights to points in a given DPC frame [5].

(4) *Fusing Multiple Users' Gaze Data to DPC Frames*: A fixation map is the aggregation of fixation points from all users viewing the same DPC frame, which can mark the region of interest. In our experiment, unintentional observation could cause isolated fixation points on DPC frames after mapping. Thus, it is necessary to filter out these noisy fixation points. For that purpose, we chose the **Density-Based Spatial Clustering of Applications with Noise (DBSCAN)** [15] algorithm. Based on the density of fixation points on the point cloud, the DBSCAN was configured to remove the noisiest fixation points in clusters with high density while at the same time being able to retain certain core fixation points in clusters with less density [41]. Figure 7 illustrates the effect of filtering noisy fixations. DBSCAN requires two parameters: ϵ is the radius of the circle to be created around each data point to check the density, and θ is the minimum number of points required inside that circle for that data point to be classified as a core point. θ should increase as the point size α of a point cloud becomes small, which means a high-density point cloud. The minimum number of points is computed as

$$\theta = \left[\frac{2^7}{1 + 20 \cdot \alpha} \right]. \tag{1}$$

ϵ is decided by a k-distance graph [48]. We took the average of fixation maps generated by multiple subjects, which is defined as

$$VS_f = \frac{1}{N} \sum_{n=1}^{N} (VS_n), \tag{2}$$

where VS_f is the averaged fixation map for each DPC frame, VS_n is the fixation map for that frame for a single subject (averaged across the number of times the frame is viewed by that subject), and N denotes the total number of subjects who have viewed that specific frame. After we got the averaged fixation map for one DPC frame, we applied the DBSCAN filtering operation to get the final fixation map. Figure 1 shows the final saliency map of *dancer* DPC.

4 Experimental Result

4.1 Analysis of Opinion Scores

Figure 8 shows the results of the subjective quality assessment of the contents in our experiment. In particular, the MOS ranging from 1 to 5, associated with the distorted DPCs, are shown with solid lines, along with corresponding CIs, whereas the dashed lines represent the respective DMOS scores. The HR scores with the corresponding CIs are represented by a shaded area.

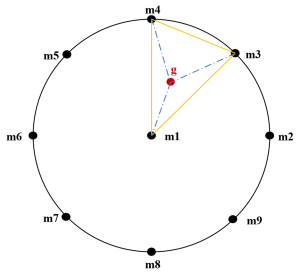
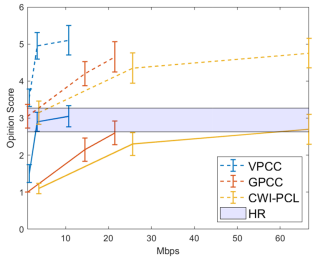


Fig. 6. Estimation of the angular error for one gaze data.

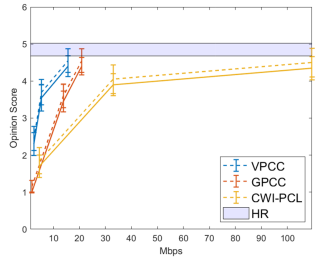
(a) Fixation points of *dancer*, front view (original fixations)

(b) Fixation points of *dancer*, front view (with filtering)

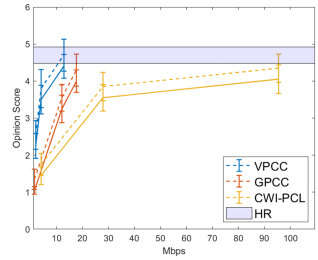
Fig. 7. Fixation map comparison with/without filtering by DBSCAN.



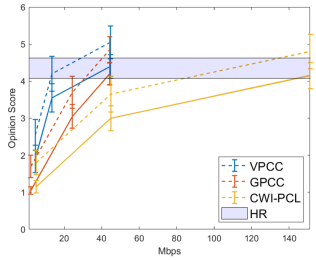
(a) *Rafa2*



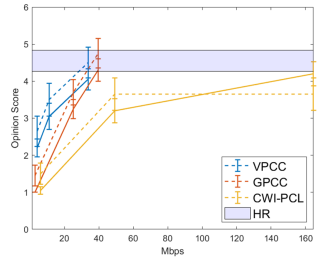
(b) *dancer*



(c) *exercise*



(d) *long dress*



(e) *soldier*

Fig. 8. MOS (solid line) and DMOS (dashed line) against bit-rate, expressed in Mbps. HR scores are shown using a shaded purple plot.

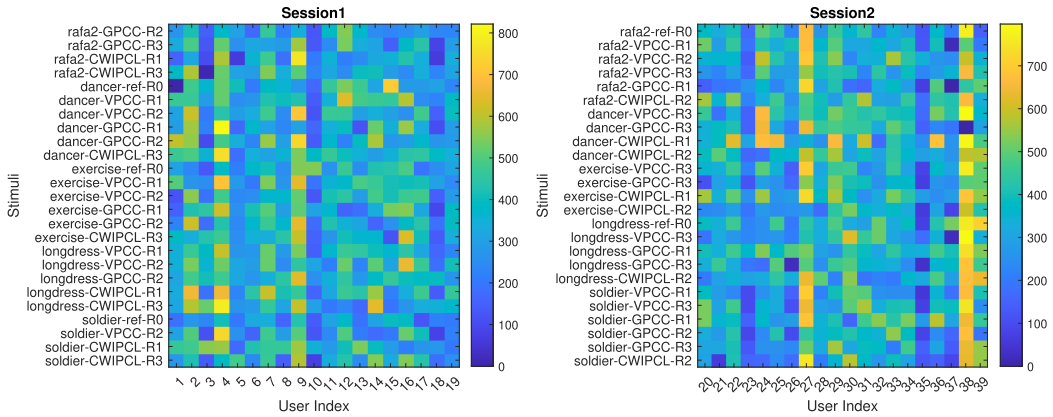


Fig. 9. The fixations for each subject and for each content. Each row denotes the fixations on a specific content, and each column denotes the fixations for each subject, respectively. $R1$ (low), $R2$ (medium), and $R3$ (high) indicate the bitrates of each codec, while $R0$ denotes the reference DPC.

While evaluating the point cloud codecs, we observe that, under similar bitrates, V-PCC exhibits the best perceptual quality, G-PCC is the second, and CWI-PCL is the last codec for all five contents. This observation reinforces prior results [53, 71] and extends similar trends for DPCs. From the perspective of contents, MOS and DMOS present similar trends, as the MOS for the HR contents is between 4 and 5. The only exception to the trend is *Rafa2*, for which the MOS for the reference content remains at around 3. This is likely related to the reconstruction error: compared with other contents captured in more professional studio settings, the reference version of *Rafa2* does not offer a satisfactory quality. The calculated DMOS is between [3, 5], because the reference content was rated so low.

4.2 Analysis of Gaze Data

To understand how and what users explore while consuming DPCs in VR, we analyze the relationship between fixations and bitrates. Figure 9 presents the number of fixations of each subject on each DPC. Figure 10 depicts the exact number of fixations across all subjects on different bitrates. Combining Figures 2, 9, and 10, we have the following observations:

- —Subjects are more interested in the high-motion DPCs (i.e., with higher TI) compared to the low-motion ones. For example, the average number of fixations on *dancer* and *long dress* is higher than *Rafa2* and *soldier*, which have lower TI on average.
- —There is no indication that low-quality content will receive less attention. In fact, we do not observe any particular trend regarding the number of fixations changing with varying quality levels.
- —Certain subjects consistently exhibit a higher number of gaze fixations (e.g., user 27 and 38 in Figure 9), possibly due to the individual differences of the participants or the accuracy of the device during the experiment.

We also explore where the subjects are looking at the DPCs in VR, and how the quality degradation will impact the visual attention in a dynamic scene. Subjects pay attention to unrealistic rendering artifacts, such as the high-heeled shoes and the hair of *long dress*. Figure 11 depicts the fixation map focusing on these two areas. Note that in this DPC, certain frames miss the heelpiece, while there are also frames that exhibit unnatural hair rendering. Figure 12 shows the fixation maps of

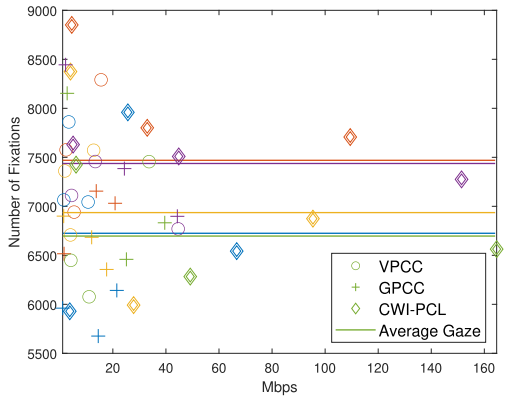


Fig. 10. Fixations against bitrates, expressed in Mbps. The average number of fixations is expressed with a line. Each color denotes a content, specifically, *Rafa2* is in blue, *dancer* is in red, *exercise* is in yellow, *long dress* is in purple, and *soldier* is in green.

the same frame among reference and all distorted *long dress* point clouds. We remark that subjects are interested in the head and areas with high motion. For all five contents, subjects tend to focus on the faces and the front view of the DPCs, despite the random rotation of the DPCs themselves. No differences are observed for the salient areas across different distortion levels. The heat values on the face are consistent across all the distortion levels; the heat value in high-motion areas floats with the motions; the heat value on the remaining areas has no pattern. This randomness may stem from unintentional fixations or the random rotation of DPCs.

We also explore where the subjects are looking at the DPCs in VR, and how the quality degradation will impact the visual attention in a dynamic scene. Subjects pay attention to unrealistic rendering artifacts, such as the high-heeled shoes and the hair of *long dress*. Figure 11 depicts the fixation map focusing on these two areas. Note that in this DPC, certain frames miss the heelpiece, while there are also frames that exhibit unnatural hair rendering. Figure 12 shows the fixation maps of the same frame among reference and all distorted *long dress* point clouds. We remark that subjects are interested in the head and areas with high motion. For all five contents, subjects tend to focus on the faces and the front view of the DPCs, despite the random rotation of the DPCs themselves. No differences are observed for the salient areas across different distortion levels. The heat values on the face are consistent across all the distortion levels; the heat value in high-motion areas floats with the motions; the heat value on the remaining areas has no pattern. This randomness may stem from unintentional fixations or the random rotation of DPCs.

4.3 Benchmarking of Objective Quality Metrics

Evaluation Criteria. Three evaluation criteria are commonly used to reflect the relationship between objective scores and subjective scores: (1) **Pearson Linear Correlation Coefficient (PLCC)**, which measures the linearity of prediction; (2) **Spearman Rank-order Correlation Coefficient (SRCC)**, which measures the monotonicity of prediction; (3) **Root Mean Square Error (RMSE)**, which measures the error of prediction. Higher values of PLCC and SRCC indicate better performance in terms of correlation with human opinion, while lower RMSE indicates better consistency.

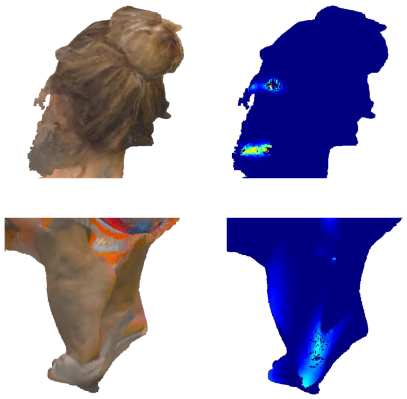


Fig. 11. Fixation map on the hair and heel of *long dress*.

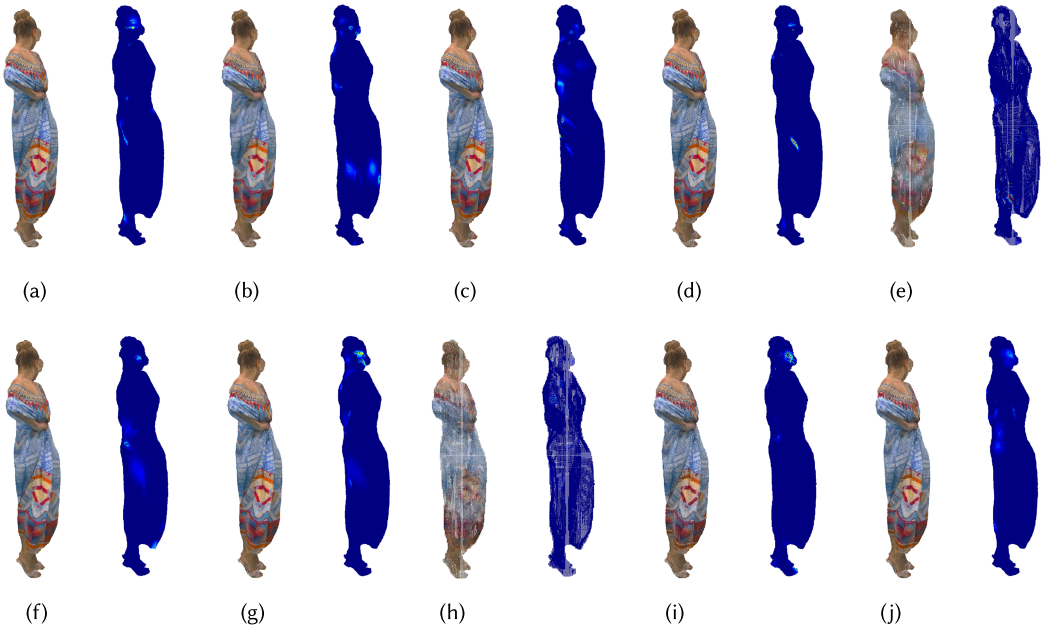


Fig. 12. The referenced and distorted versions of point cloud *long dress* (frame 128) with corresponding visual attention maps based on the proposed processing protocol. (a) The reference version. (b)–(j) The distorted versions of *long dress* from low to high quality. Specifically, (b)–(d): V-PCC, (e)–(g): G-PCC, (h)–(j): CWI-PCL.

A five-parametric logistic regression is adopted to fit the relationship between the subjective scores and the objective scores before calculating the PLCC and RMSE [6].

Performance Evaluation of Objective Quality Metrics without Visual Attention. With the subjective scores collected in our experiments, we conduct an evaluation and comparison of existing objective metrics for the task of DPC quality assessment. Only point-based metrics are considered. We chose metrics adopted by the MPEG group, namely, point-to-point and point-to-plane with MSE and Hausdorff distances, with and without using **Peak Signal to Noise Ratio (PSNR)**, and the

weighted average color differences for Y, U, and V channels in terms of PSNR, which is defined as [55]:

$$PSNR_{YUV} = \frac{6 \cdot PSNR_Y + PSNR_U + PSNR_V}{8}. \tag{3}$$

Furthermore, we choose another four state-of-the-art metrics, namely, histY [62], pointSSIM [2], PCM_RR [60], and PCQM[39]. Since the metrics are originally designed for static contents and do not explicitly consider the temporal aspect, we apply the metrics to each frame and then pool all the 300 frames' scores as the final score for one DPC sequence. Based on the findings of [19], we choose the mean as the baseline and the variation model of [43] as the temporal pooling method to obtain the objective score for each DPC sequence.

Performance Evaluation of Objective Quality Metrics with Visual Attention. Since the visual saliency is point-based, we integrated the importance weight with the quality score computed by the existing metric for each point per frame, and adopted the same pooling strategy, to obtain the final quality score. The importance weight is normalized between 0 and 1 based on the computation result in Equation (2). We adopted two ways of weighting the point-based metrics with visual saliency. The first considers only the salient region and excludes the unsalient region, which is defined as:

$$Q_v^1 = M \cdot VS_f, \tag{4}$$

where Q_v^1 is the quality score with only the salient region for one frame belonging to a DPC, and M is the quality score for the corresponding frame from the point-based metrics. The second method retains the unsalient region but assigns relatively higher weights to the salient region, termed as normalized saliency, which is defined as:

$$Q_v^2 = M \cdot (VS_f + 1), \tag{5}$$

where Q_v^2 is the quality score with the normalized saliency for one frame belonging to a DPC.

The results of the performance indices for the 13 objective metrics, with mean and variation as the temporal pooling methods, are presented in Tables 3 and 4. Consistent with findings in [1], we note that altering the temporal pooling method does not significantly impact high-performing quality metrics (PLCC higher than 0.5). In the original implementation, PCM_RR achieves the highest PLCC/SRCC performance with average pooling, p2point_Hausdroff demonstrates the best performance when only considering the salient region with average pooling, hist_Y achieves the highest SRCC performance after applying normalized salient weighting, and pointSSIM achieves the best PLCC/RMSE, and PCQM achieves the best SRCC after normalized saliency. The significant performance increase of p2point_Hausdroff metric may be attributed to the exclusion of the non-salient region, which helps mitigate the sensitivity to outliers. This refinement ensures that only errors within the salient region are retained. Generally, several studies report that modern video quality assessment models experience only slight improvements from incorporating saliency [64]. However, for DPC quality assessment, metrics that focus exclusively on salient regions show a decline in performance. Conversely, metrics using a normalized weighting strategy achieve similar results to the original implementation, with performance also depending on the temporal pooling methods employed.

4.4 Qualitative Results

Thirty-two interview audio recordings were transcribed into texts and coded using Dovetail,[1] with eight participants declining to participate. Following Maguire's guideline on thematic analysis [36], we initially reviewed and labeled the text, organized these labels into themes, and subsequently

[1] https://dovetail.com/

Table 3. Performance Evaluation of State-of-the-Art Quality Metrics on QAVA-DPC with Average Pooling

Metrics	PLCC			SRCC			RMSE		
	OI	SPO	NSW	OI	SPO	NSW	OI	SPO	NSW
p2point_MSE	0.782	0.613	0.781	0.738	0.544	0.738	0.728	0.902	0.730
p2point_MSE_PSNR	0.513	0.580	0.513	0.465	0.555	0.465	1.002	0.929	1.003
p2point_Hausdroff	0.200	**0.854**	0.434	0.240	**0.687**	0.323	1.144	**0.594**	1.052
p2point_Hausdroff_PNSR	0.521	0.661	0.364	0.146	0.641	0.204	0.997	0.856	1.088
p2plane_MSE	0.710	0.571	0.710	0.690	0.546	0.689	0.822	0.937	0.823
p2plane_MSE_PSNR	0.628	0.597	0.484	0.483	0.571	0.483	0.908	0.915	1.026
p2plane_Hausdroff	0.207	0.767	0.428	0.257	0.634	0.281	1.142	0.732	1.055
p2plane_Hausdroff_PNSR	0.514	0.699	0.532	0.163	0.617	0.183	1.005	0.816	0.994
YUV_PSNR	0.661	0.606	0.662	0.654	0.561	0.652	0.876	0.907	0.875
hist_Y	0.840	0.204	0.827	0.820	0.086	0.781	0.633	1.143	0.657
PCM_RR	**0.844**	0.058	0.444	**0.822**	0.046	0.398	**0.626**	1.166	1.046
pointSSIM	0.836	0.652	**0.836**	0.772	0.649	0.771	0.641	0.885	**0.640**
PCQM	0.813	0.334	0.833	0.758	0.256	**0.815**	0.681	1.101	0.646

Columns represent Original Implementation (OI), Salient-Part-Only (SPO), and Normalized Saliency Weighting (NSW). The best performance for PLCC/SRCC/RMSE is highlighted in bold and marked with red/blue/orange color, respectively.

Table 4. Performance Evaluation of State-of-the-Art Quality Metrics on QAVA-DPC with Variation Pooling

Metrics	PLCC			SRCC			RMSE		
	OI	SPO	NSW	OI	SPO	NSW	OI	SPO	NSW
p2point_MSE	0.769	0.604	**0.769**	0.731	0.536	**0.730**	0.747	**0.909**	**0.751**
p2point_MSE_PSNR	0.267	0.108	0.267	0.135	0.180	0.135	1.125	1.134	1.125
p2point_Hausdroff	0.527	0.593	0.515	0.251	0.536	0.223	0.993	0.919	1.001
p2point_Hausdroff_PNSR	0.248	0.114	0.319	0.232	0.069	0.286	1.131	1.133	1.107
p2plane_MSE	0.662	0.570	0.662	0.621	0.539	0.625	0.875	0.937	0.875
p2plane_MSE_PSNR	0.210	0.103	0.210	0.119	0.134	0.119	1.141	1.134	1.142
p2plane_Hausdroff	0.453	**0.605**	0.501	0.163	0.556	0.217	1.041	**0.909**	1.010
p2plane_Hausdroff_PNSR	0.361	0.054	0.327	0.242	0.005	0.299	1.089	1.139	1.103
YUV_PSNR	0.643	0.601	0.643	0.625	**0.583**	0.625	0.894	0.912	0.894
hist_Y	0.735	0.249	0.669	0.727	0.054	0.617	0.792	1.133	0.868
PCM_RR	0.552	0.393	0.293	0.485	0.216	0.020	0.974	1.074	1.146
pointSSIM	0.705	0.594	0.707	0.710	0.539	0.714	0.828	0.940	0.825
PCQM	**0.807**	0.357	0.732	**0.738**	0.256	0.552	**0.690**	1.091	0.796

Columns represent OI, SPO, and NSW. The best performance for PLCC/SRCC/RMSE is highlighted in bold and marked with red/blue/orange color, respectively.

convened to establish the connection between perceptual quality and visual attention during the subjective test. Each participant is denoted as P1–P32, with the number of participants concurring with each statement indicated in parentheses. The qualitative results of this thematic analysis are presented in this section, with detailed findings discussed in the following subsections.

4.4.1 *Factors That Impact Quality Assessment and Visual Attention.*

Temporal Information. Participants (13) pointed out that the flickering effect is the most bothersome artifact in our DPC playback scene, often leading to lower ratings. (P21: “If it’s very like vague

or like with big dots, then it's ok, it's fake, but if it flickers all the time, that could be a bit annoying, it's sort of a lot to see.") Additionally, participants (20) reported that they tended to explore more during the observation of high-motion point cloud sequences.

Geometry and Texture. Distortions in geometry (12) and texture (11) are identified as the second and third factors influencing the subjective rating of point clouds under scrutiny. (P3: "I was observing precisely two things, the edges of the body and how distorted they are, and also some distortions inside the costume.") For DPC quality assessment, TI emerges as more critical than either geometric or texture distortions, with geometry and texture exhibiting nearly equal importance.

Distance Impacts the Quality Rating. Participants (8) discovered that the viewing distance can impact the subjective rating of the same content. Five of them noted that the appearance of holes is determined by the viewing distance, resulting in visually distinct point clouds even when inspecting the same sequence. This finding complements the conclusion in [56] that objects viewed from a greater distance tend to receive lower ratings compared to their closer counterparts.

Relationship between Visual Attention and Quality Assessment. Participants favored the *longdress* (15), *soldier* (11), and *dancer* (9) point cloud sequences among all the contents. The two primary reasons cited were their high quality (19) and the presence of cues aiding in quality score determination (15). The characteristics of the point cloud itself influenced participants' visual attention. For instance, individuals tended to focus more on content characterized by high motion (*dancer*), realism (*longdress, soldier*), and intricate details (*soldier*) to facilitate the quality assessment task.

4.4.2 *Subjective Quality Experiment Design for Dataset Construction.*

Procedure. Participants (9) recommend streamlining the calibration and error profiling process to enhance user-friendliness, while also acknowledging the importance of maintaining data accuracy. This suggests a need for striking a balance between data precision and ease of use. Such a balanced approach is crucial for advancing the development of eye-tracking techniques, particularly in terms of calibration procedures.

Interaction with Content. Participants (5) highlighted the importance of allowing individuals to determine the number of loops for each content themselves. They emphasized that for tasks requiring quality assessment, they may already have instant results in mind for certain content. Furthermore, participants expressed a desire for increased interaction between themselves and the objects within DPCs playback scenes in VR for better evaluation of the quality, underscoring the importance of customizable interaction for effective evaluation. (P4: "... I like to rotate it to whatever angle I want and then go and see it.")

4.4.3 *Content Characteristics and Quality Assessment Task: Influencing User Interaction.*

Impacts of the Quality Assessment Task for Interaction in VR. Participants (21) attributed most of their movement to the need to observe the front face to determine or recheck their quality score ratings. Additionally, VR cues presented in the human figures also prompted them to move extensively, enabling them to identify more cues for evaluating point cloud quality. (P5: "When I was seeing the quality, I was seeing the helmet and it had like a small thing on top, and there is a difference in the quality of that as well, and even the gun, you could see like different features on the gun. So there are more things to look at.")

Impacts of Contents Characteristics for Interaction in VR. Participants (21) expressed the view that if the quality level is easy to discern, they prefer to remain stationary until the sequence completes its loop, such as in cases of excellent and poor quality scales. However, if the quality

level is relatively challenging to distinguish, individuals opt to move around for a comprehensive observation, even without engaging in random rotation operations. This observation aligns with Damme's conclusion [56] that human perception of underlying quality representations is intricately linked with the content and its geometric properties under examination.

5 Discussion

5.1 Advancing and Enhancing the Explanation of DPC Quality Assessment Metrics

There is potential for improving the performance of point-based metrics through a strategic combination of visual attention and pooling methods. However, the visual attention data available for DPCs typically cover only a small region of the dense object, limiting its efficacy in advancing DPC quality assessment metrics. Our investigation reveals that employing temporal variation pooling methods leads to decreased performance, prompting further exploration into suitable temporal pooling techniques or spatial-temporal pooling for DPC quality metrics. It is crucial to account for the intrinsic characteristics of DPCs beyond merely relying on video quality indicators. Additionally, careful consideration should be given to matching different distortion types with appropriate temporal pooling methods.

5.2 The Influence of Task for DPC Visual Attention

Our experiment, which focuses on evaluating the visual quality of DPCs, likely influenced participants' attention toward specific content areas that facilitated the task: for example, areas with patterns on which distortions would easily be spotted. That does not necessarily mean that the same area would be a salient region had the test been administered with a different task or task-free. Insights from the semi-structured interviews confirm that the identified salient regions are not only inherently attention-grabbing but also offer cues that aid participants in the quality assessment process. This dual influence highlights the contextual nature of visual attention, driven by both intrinsic content characteristics and the task's demands. Moreover, the accuracy of visual attention map predictions varies depending on the display environment and associated tasks [29, 66]. Future research should explore how visual attention varies across different tasks or task-free scenarios to uncover more generalized patterns of saliency in DPCs and their implications for real-world applications [10].

5.3 Visual Attention Applications for DPC

The insights derived from gaze data for DPC extend far beyond quality assessment. Accurate prediction of visual attention provides a robust foundation for optimizing compression and streaming strategies. For example, saliency-driven bit allocation can enhance encoding efficiency by prioritizing the fidelity of regions that capture user attention while allocating fewer resources to non-salient areas. Additionally, integrating saliency maps with temporal motion information, such as motion vectors, enables adaptive streaming parameters that dynamically adjust to user focus, ensuring a seamless and engaging experience. Furthermore, leveraging these insights can support semantic segmentation of motion-dominant and motion-static regions, which has significant potential for applications such as object recognition and interactive XR environments.

5.4 Dataset Applications and Prospective Extensions

QAVA-DPC, encompassing MOS/DMOS, users' gaze data, and our meticulously processed visual attention maps, hold significant potential as a foundational reference for the multiple aspects. First, since the dataset includes the raw data alongside the visual attention maps, it provides researchers and practitioners with valuable resources to develop and test novel algorithms for post-processing

gaze data and creating visual attention maps. Additionally, QAVA-DPC facilitates the comparison of visual saliency maps across different devices (e.g., screen-based or XR-based), highlighting the need for similarity metrics tailored to DPC. Furthermore, the dataset enables the development of objective quality metrics and visual attention prediction models for DPC without requiring resource-intensive user studies. Moreover, insights derived from the qualitative analysis and visual attention design paradigms can drive advancements in DPC-related research and applications. Finally, existing point-based objective quality metrics can be refined and tailored for DPC to explore how to incorporate visual attention and assess its added value in DPC quality assessment.

6 Conclusion

This article has presented a dataset containing subjective opinion scores and visual attention maps of DPCs in 6 DoF. An experimental protocol was designed for the subjective quality assessment and visual attention detection of the DPCs in a VR environment. Additionally, we evaluated existing objective quality metrics with and without the integration of visual attention maps to study the added value of visual attention in objective DPC quality metrics. Results from semi-structured interviews and the captured visual attention maps provided deeper insights into human behavior in VR environments. While previous literature indicates that visual attention may be helpful for image/video streaming, compression, as well as quality assessment, evaluations and benchmarking for DPCs in our work are just the beginning. In our future work, we will focus on predicting the visual attention of DPCs and designing objective metrics to further improve the metrics' correlations between predicted quality scores and ground-truth subjective scores.

References

[1] Ali Ak, Emin Zerman, Suiyi Ling, Patrick Le Callet, and Aljosa Smolic. 2021. The effect of temporal sub-sampling on the accuracy of volumetric video quality assessment. In *2021 Picture Coding Symposium (PCS)*, 1–5. DOI: https://doi.org/10.1109/PCS50896.2021.9477449

[2] Evangelos Alexiou and Touradj Ebrahimi. 2020. Towards a point cloud structural similarity metric. *In 2020 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, 1–6. DOI: https://doi.org/10.1109/ICMEW46912.2020.9106005

[3] Evangelos Alexiou, Yana Nehmé, Emin Zerman, Irene Viola, Guillaume Lavoué, Ali Ak, Aljosa Smolic, Patrick Le Callet, and Pablo Cesar. 2023. Subjective and objective quality assessment for volumetric video. In *Immersive Video Technologies*. Elsevier, 501–552.

[4] Evangelos Alexiou, Irene Viola, Tomás M. Borges, Tiago A. Fonseca, Ricardo L. de Queiroz, and Touradj Ebrahimi. 2019. A comprehensive study of the rate-distortion performance in MPEG point cloud compression. *APSIPA Transactions on Signal and Information Processing* 8, 1 (2019), e27. DOI: https://doi.org/10.1017/ATSIP.2019.20

[5] Evangelos Alexiou, Peisen Xu, and Touradj Ebrahimi. 2019. Towards modelling of visual saliency in point clouds for immersive applications. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 4325–4329.

[6] Jochen Antkowiak, T. D. F. Jamal Baina, France Vittorio Baroncini, Noel Chateau, France FranceTelecom, Antonio Claudio França Pessoa, F. U. B. Stephanie Colonnese, Italy Laura Contin, Jorge Caviedes, and France Philips. 2000. Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000. Retrieved from https://www.academia.edu/2102517/FINAL_REPORT_FROM_THE_VIDEO_QUALITY_EXPERTS_GROUP_ON_THE_VALIDATION_OF_OBJECTIVE_MODELS_OF_VIDEO_QUALITY_ASSESSMENT_March_

[7] Isayas B. Adhanom, Samantha C. Lee, Eelke Folmer, and Paul MacNeilage. 2020. Gazemetrics: An open-source tool for measuring the data quality of HMD-based eye trackers. In *ACM Symposium on Eye Tracking Research and Applications*, 1–5.

[8] Keming Cao, Yi Xu, and Pamela Cosman. 2020. Visual quality of compressed mesh and point cloud sequences. *IEEE Access* 8 (2020), 171203–171217.

[9] Qiangqiang Cheng, Pengyu Sun, Chunsheng Yang, Yubin Yang, and Peter Xiaoping Liu. 2020. A morphing-based 3D point cloud reconstruction framework for medical image processing. *Computer Methods and Programs in Biomedicine* 193 (2020), 105495.

[10] Shiwei Cheng, Jing Fan, and Yilin Hu. 2023. Visual saliency model based on crowdsourcing eye tracking data and its application in visual design. *Personal and Ubiquitous Computing* 27, 3 (2023), 613–630.

[11] J. H. Clark. 1924. The Ishihara test for color blindness. *American Journal of Physiological Optics* 5 (1924), 269–276.

[12] Samuel Rhys Cox, May Lim, and Wei Tsang Ooi. 2023. VOLVQAD: An MPEG V-PCC volumetric video quality assessment dataset. In *Proceedings of the 14th Conference on ACM Multimedia Systems*, 357–362.
[13] Erwan J. David, Jesús Gutiérrez, Antoine Coutrot, Matthieu Perreira Da Silva, and Patrick Le Callet. 2018. A dataset of head and eye movements for 360 videos. In *Proceedings of the 9th ACM Multimedia Systems Conference*, 432–437.
[14] Eugene d'Eon, Bob Harrison, Taos Myers, and Philip A. Chou. 2019. JPEG Pleno Database: 8i Voxelized Full Bodies (8iVFB v2)—A Dynamic Voxelized Point Cloud Dataset.
[15] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 226–231.
[16] Yu Fan, Zicheng Zhang, Wei Sun, Xiongkuo Min, Jiaman Lin, Guangtao Zhai, and Ning Liu. 2023. MV-VVQA: Multi-view learning for no-reference volumetric video quality assessment. In *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 670–674.
[17] Frederick L. Ferris, III, Aaron Kassoff, George H. Bresnick, and Ian Bailey. 1982. New visual acuity charts for clinical research. *American Journal of Ophthalmology* 94, 1 (1982), 91–96.
[18] Pedro Garcia Freitas, Mateus Gonçalves, Johann Homonnai, Rafael Diniz, Mylène, and C. Q. Farias. 2022. On the performance of temporal pooling methods for quality assessment of dynamic point clouds. In *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.
[19] Pedro G. Freitas, Giovani D. Lucafo, Mateus Gonçalves, Johann Homonnai, Rafael Diniz, Mylène, and C. Q. Farias. 2022. Comparative evaluation of temporal pooling methods for no-reference quality assessment of dynamic point clouds. In *Proceedings of the 1st Workshop on Photorealistic Image and Environment Synthesis for Multimedia Experiments*, 35–41.
[20] Jesús Gutiérrez, Gulzhanat Dandyyeva, Matteo Dal Magro, Carlos Cortés, Michele Brizzi, Marco Carli, and Federica Battisti. 2023. Subjective evaluation of dynamic point clouds: Impact of compression and exploration behavior. In *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 675–679.
[21] Jesus Gutierrez, Pablo Perez, Marta Orduna, Ashutosh Singla, Carlos Cortes, Pramit Mazumdar, Irene Viola, Kjell Brunnstrom, Federica Battisti, Natalia Cieplinska, et al. 2021. Subjective evaluation of visual quality and simulator sickness of short 360° videos: ITU-T rec. P. 919. *IEEE Transactions on Multimedia* 24 (2021), 3087–3100.
[22] Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11 (1998), 1254–1259.
[23] ITU-R BT.500-13. 2012. *Methodology for the Subjective Assessment of the Quality of Television Pictures*. International Telecommunications Union.
[24] ITU-T P.910. 2008. *Subjective Video Quality Assessment Methods for Multimedia Applications*. International Telecommunication Union.
[25] ITU-T P.913. 2016. *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*. International Telecommunication Union.
[26] Alireza Javaheri, Catarina Brites, Fernando Pereira, and João Ascenso. 2021. Point cloud rendering after coding: Impacts on subjective and objective quality. *IEEE Transactions on Multimedia* 23 (2021), 4049–4064. DOI: https://doi.org/10.1109/TMM.2020.3037481
[27] Yize Jin, Meixu Chen, Todd Goodall, Anjul Patney, and Alan C. Bovik. 2021. Subjective and objective quality assessment of 2D and 3D foveated video compression in virtual reality. *IEEE Transactions on Image Processing* 30 (2021), 5905–5919.
[28] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220.
[29] Haksub Kim, Sanghoon Lee, and Alan Conrad Bovik. 2014. Saliency prediction on stereoscopic videos. *IEEE Transactions on Image Processing* 23, 4 (2014), 1476–1490. DOI: https://doi.org/10.1109/TIP.2014.2303640
[30] Sergi Fernández Langa, Mario Montagud, Gianluca Cernigliaro, and David Rincón Rivera. 2022. Multiparty holomeetings: Toward a new era of low-cost volumetric holographic meetings in virtual reality. *IEEE Access* 10 (2022), 81856–81876.
[31] Olivier Le Meur, Alexandre Ninassi, Patrick Le Callet, and Dominique Barba. 2010. Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric. *Signal Processing: Image Communication* 25, 7 (2010), 547–558.
[32] Zhi Li, Anne Aaron, Ioannis Katsavounidis, Anush Moorthy, and Megha Manohara. 2016. Toward a Practical Perceptual Video Quality Metric. Retrieved from https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652
[33] Weisi Lin and C.-C. Jay Kuo. 2011. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation* 22, 4 (2011), 297–312.
[34] Hantao Liu and Ingrid Heynderickx. 2011. Visual attention in objective image quality assessment: Based on eye-tracking data. *IEEE Transactions on Circuits and Systems for Video Technology* 21, 7 (2011), 971–982. DOI: https://doi.org/10.1109/TCSVT.2011.2133770

[35] Zhongkang Lu, Weisi Lin, Xiaokang Yang, EePing Ong, and Susu Yao. 2005. Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation. *IEEE Transactions on Image Processing* 14, 11 (2005), 1928–1942.
[36] Moira Maguire and Brid Delahunt. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education* 9, 3 (2017).
[37] Jean-Eudes Marvie, Yana Nehmé, Danillo Graziosi, and Guillaume Lavoué. 2023. Crafting the MPEG metrics for objective and perceptual quality assessment of volumetric videos. *Quality and User Experience* 8, 1 (2023), 4.
[38] Rufael Mekuria, Kees Blom, and Pablo Cesar. 2017. Design, implementation, and evaluation of a point cloud codec for tele-immersive video. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 4 (2017), 828–842. DOI: https://doi.org/10.1109/TCSVT.2016.2543039
[39] Gabriel Meynet, Yana Nehmé, Julie Digne, and Guillaume Lavoué. 2020. PCQM: A full-reference quality metric for colored 3D point clouds. In *2020 12th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.
[40] Jesús Moreno-Arjonilla, Alfonso López-Ruiz, J. Roberto Jiménez-Pérez, José E. Callejas-Aguilera, and Juan M. Jurado. 2024. Eye-tracking on virtual reality: A survey. *Virtual Reality* 28, 1 (2024), 38.
[41] Anh Nguyen and Zhisheng Yan. 2019. A saliency dataset for 360-degree videos. In *Proceedings of the 10th ACM Multimedia Systems Conference*, 279–284.
[42] 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 *Proceedings of the 15th ACM Multimedia Systems Conference (MMSys '24)*. ACM, New York, NY, 367–373. DOI: https://doi.org/10.1145/3625468.3652182
[43] Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, and Dominique Barba. 2009. Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE Journal of Selected Topics in Signal Processing* 3, 2 (2009), 253–265.
[44] Cagri Ozcinar and Aljosa Smolic. 2018. Visual attention in omnidirectional video for virtual reality applications. In *2018 10th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.
[45] Yashas Rai, Jesús Gutiérrez, and Patrick Le Callet. 2017. A dataset of head and eye movements for 360 degree images. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, 205–210.
[46] Ignacio Reimat, Evangelos Alexiou, Jack Jansen, Irene Viola, Shishir Subramanyam, and Pablo Cesar. 2021. CWIPC-SXR: Point cloud dynamic human dataset for social XR. In *Proceedings of the 12th ACM Multimedia Systems Conference*, 300–306.
[47] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications*, 71–78.
[48] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems* 42, 3 (2017), 1–21.
[49] Sebastian Schwarz, Marius Preda, Vittorio Baroncini, Madhukar Budagavi, Pablo Cesar, Philip A. Chou, Robert A. Cohen, Maja Krivokuća, Sébastien Lasserre, Zhu Li, et al. 2019. Emerging MPEG standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 9, 1 (2019), 133–148. DOI: https://doi.org/10.1109/JETCAS.2018.2885981
[50] Ludwig Sidenmark, Mathias N. Lystbæk, and Hans Gellersen. 2023. GE-simulator: An open-source tool for simulating real-time errors for HMD-based eye trackers. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications (ETRA '23)*. ACM, New York, NY, Article 8, 6 pages. DOI: https://doi.org/10.1145/3588015.3588417
[51] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1633–1642. DOI: https://doi.org/10.1109/TVCG.2018.2793599
[52] S. Schwarz. 2019. Common test conditions for PCC. ISO/IEC JTC1/SC29/WG11 Doc. N18665.
[53] Shishir Subramanyam, Jie Li, Irene Viola, and Pablo Cesar. 2020. Comparing the quality of highly realistic digital humans in 3dof and 6dof: A volumetric video case study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 127–136.
[54] Shishir Subramanyam, Irene Viola, Jack Jansen, Evangelos Alexiou, Alan Hanjalic, and Pablo Cesar. 2022. Subjective QoE evaluation of user-centered adaptive streaming of dynamic point clouds. In *2022 14th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.
[55] Eric M. Torlig, Evangelos Alexiou, Tiago A. Fonseca, Ricardo L. de Queiroz, and Touradj Ebrahimi. 2018. A novel methodology for quality assessment of voxelized point clouds. In *Applications of Digital Image Processing XLI*. SPIE, 174–190.
[56] Sam Van Damme, Imen Mahdi, Hemanth Kumar Ravuri, Jeroen van der Hooft, Filip De Turck, and Maria Torres Vega. 2023. Immersive and interactive subjective quality assessment of dynamic volumetric meshes. In *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 141–146.

[57] Sam Van Damme, Maria Torres Vega, and Filip De Turck. 2021. A full-and no-reference metrics accuracy analysis for volumetric media streaming. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 225–230.

[58] Jeroen van der Hooft, Maria Torres Vega, Christian Timmerer, Ali C. Begen, Filip De Turck, and Raimund Schatz. 2020. Objective and subjective QoE evaluation for adaptive point cloud streaming. In *2020 12th International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. DOI: https://doi.org/10.1109/QoMEX48832.2020.9123081

[59] Anouk van Kasteren, Kjell Brunnström, John Hedlund, and Chris Snijders. 2022. Quality of experience of 360 video–subjective and eye-tracking assessment of encoding and freezing distortions. *Multimedia Tools and Applications* 81, 7 (2022), 9771–9802.

[60] I. Viola and P. S. César Garcia. 2020. PCM_RR: A reduced reference metric for visual quality evaluation of point cloud contents.

[61] Irene Viola, Jack Jansen, Shishir Subramanyam, Ignacio Reimat, and Pablo Cesar. 2023. VR2Gather: A collaborative social VR system for adaptive multi-party real-time communication. *IEEE MultiMedia* 30, 2 (2023), 48–59. DOI: 10.1109/MMUL.2023.3263943

[62] Irene Viola, Shishir Subramanyam, and Pablo Cesar. 2020. A color-based objective quality metric for point cloud contents. In *2020 12th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[63] Irene Viola, Shishir Subramanyam, Jie Li, and Pablo Cesar. 2022. On the impact of vr assessment on the quality of experience of highly realistic digital humans. *Quality and User Experience* 7, 1 (2022).

[64] Xinyi Wang, Angeliki Katsenou, and David Bull. 2023. UGC quality assessment: Exploring the impact of saliency in deep feature-based quality assessment. In *Applications of Digital Image Processing XLVI*. SPIE, 351–365.

[65] Heino Widdel. 1984. Operational problems in analysing eye movements. In *Advances in Psychology*. Vol. 22, Elsevier, 21–29.

[66] Nan Wu, Kaiyan Liu, Ruizhi Cheng, Bo Han, and Puqi Zhou. 2024. Theia: Gaze-driven and perception-aware volumetric content delivery for mixed reality headsets. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*, 70–84.

[67] Xinju Wu, Yun Zhang, Chunling Fan, Junhui Hou, and Sam Kwong. 2021. Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 12 (2021), 4630–4644. DOI: https://doi.org/10.1109/TCSVT.2021.3101484

[68] Yi Xu, Yao Lu, and Ziyu Wen. 2017. Owlii dynamic human textured mesh sequence dataset. In *ISO/IEC JTC1/SC29/WG1 1 input document m41658*.

[69] Mengyu Yang, Di Wu, Zelong Wang, Miao Hu, and Yipeng Zhou. 2023. Understanding and improving perceptual quality of volumetric video streaming. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1979–1984.

[70] Emin Zerman, Pan Gao, Cagri Ozcinar, and Aljosa Smolic. 2019. Subjective and objective quality assessment for volumetric video compression. *Electronic Imaging* 10 (2019), 323–321.

[71] Emin Zerman, Cagri Ozcinar, Pan Gao, and Aljosa Smolic. 2020. Textured mesh vs coloured point cloud: A subjective study for volumetric video compression. In *2020 12th International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. DOI: https://doi.org/10.1109/QoMEX48832.2020.9123137

[72] Wei Zhang and Hantao Liu. 2017. Study of saliency in objective video quality assessment. *IEEE Transactions on Image Processing* 26, 3 (2017), 1275–1288. DOI: https://doi.org/10.1109/TIP.2017.2651410

[73] Wei Zhang and Hantao Liu. 2017. 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.

[74] Yihuan Zhang, Liang Wang, and Yifan Dai. 2023. PLOT: A 3D point cloud object detection network for autonomous driving. *Robotica* 41 (2023), 1–17.

[75] Xuemei Zhou, Irene Viola, Evangelos Alexiou, Jack Jansen, and Pablo Cesar. 2023. 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)*, 69–78. DOI: https://doi.org/10.1109/ISMAR59233.2023.00021

Received 4 August 2024; revised 30 January 2025; accepted 11 March 2025