UVG-CWI-DQPC: Dual-Quality Point Cloud Dataset for Volumetric Video Applications

Guillaume Gautier
guillaume.gautier@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Xuemei Zhou
xuemei.zhou@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
Delft University of Technology
Delft, The Netherlands

Thong Nguyen
thong.nguyen@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Jack Jansen
jack.jansen@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Louis Fréneau
louis.freneau@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Marko Viitanen
marko.viitanen@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Uyen Phan
uyen.phan@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Jani Käpylä
jani.kapyla@tuni.fi
Tampere University
Tampere University, Finland

Irene Viola
irene.viola@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Alexandre Mercat
alexandre.mercat@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Pablo Cesar
p.s.cesar@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
Delft University of Technology
Delft, The Netherlands

Jarno Vanne
jarno.vanne@tuni.fi
Ultra Video Group
Tampere University
Tampere, Finland

Abstract

Volumetric video is a key enabler of immersive extended reality (XR) experiences and is often represented using point clouds for their structural simplicity. However, capturing volumetric content through multi-view acquisition and depth sensing poses many challenges, such as occlusions and depth mismatches. To foster research in this field, we introduce a unique dual-quality point cloud dataset, named UVG-CWI-DQPC, which is designed to support the development of point cloud enhancement, compression, and quality assessment. Our dataset includes 12 dynamic sequences captured simultaneously by: 1) a high-end capture system producing high-fidelity point clouds with extensive processing; and 2) a consumer-grade capture system relying on affordable RGB-D cameras, lightweight processing, and open-source tools. For each sequence, our dataset provides ground-truth point clouds from the high-end capture system and raw RGB-D footage from the consumer-grade capture system, along with calibration data and tools for point cloud generation. This dual-quality setup enables direct comparison and benchmarking of algorithms for densification, occlusion removal, registration, and quality enhancement. Our dataset is publicly available under a permissive license to support reproducible research and standardization work in Moving Picture Experts Group (MPEG) and 3rd Generation Partnership Project (3GPP).

MM '25, Dublin, Ireland.

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758263

CCS Concepts

• **Computing methodologies** → **Virtual reality**; **Point-based models**.

Keywords

Point cloud; Mesh; RGB-D; Volumetric video; Registration; Densification; Occlusion removal; Dataset

ACM Reference Format:
Guillaume Gautier, Xuemei Zhou, Thong Nguyen, Jack Jansen, Louis Fréneau, Marko Viitanen, Uyen Phan, Jani Käpylä, Irene Viola, Alexandre Mercat, Pablo Cesar, and Jarno Vanne. 2025. UVG-CWI-DQPC: Dual-Quality Point Cloud Dataset for Volumetric Video Applications. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3746027.3758263

Table 1: Overview of Existing Point Cloud Video Datasets

[ref] Dataset	#Seq	FPS	#Frames	#Person	Format	Capture System
[4] 8iVFB v2	4	30	300	1	point cloud	High-end
[17] Owlii	4	30	600	1	point cloud/mesh	High-end
[5] UVG-VPC	12	25	250	1	point cloud/mesh	High-end
[6] FSVVD	26	N/A	127–2192	≥ 1	point cloud	Consumer-grade
[12] CWIPC-SXR	21	30	592–2768	≥ 1	point cloud	Consumer-grade
[9] PCVD	20	30	600–900	≥ 1	point cloud/mesh	Consumer-grade
[1] HUMAN4D	19	30	1356–6192	≥ 1	point cloud/mesh	Consumer-grade
[18] vsenseVVDB2	3	30	300	1	point cloud/mesh	Consumer-grade
Our UVG-CWI-DQPC	**12**	**15/30**	**157–201**	≥ 1	**point cloud/mesh**	**High-end & Consumer-grade**

1 Introduction

Recent advances in visual volumetric media technologies have enabled the creation of detailed 3D digital representations of real-world objects and scenes. These state-of-the-art solutions lay the foundation for immersive *extended reality* (*XR*) experiences with *six degrees of freedom* (*6DoF*), reshaping the application landscape, e.g., in communication, remote supervision, gaming, broadcasting, distance education, and telemedicine.

High-end capture systems can produce realistic 3D representations, but their substantial cost and complex setup limit widespread adoption. In contrast, more affordable consumer-grade capture systems equipped with a limited set of lower-resolution cameras tend to output sparser point clouds with a higher number of occlusions and noise. These quality limitations can be mitigated through techniques like occlusion removal, point cloud densification, and camera registration, but designing them calls for suitable datasets.

To our knowledge, all public point cloud datasets have been captured using either a high-end [4, 5, 17] or a consumer-grade [1, 6, 9, 12, 18] capture system. We aim to bridge this gap by introducing a dataset, named UVG-CWI-DQPC, which comprises 12 point cloud sequences recorded simultaneously with two different systems:

1) A high-end capture system [16], which produces high-quality point clouds with expensive hardware and time-consuming processing.
2) A consumer-grade capture system [2], which produces lower-quality point clouds in practical time, built on off-the-shelf components and open-source software.

UVG-CWI-DQPC is distributed at

https://ultravideo.fi/UVG-CWI-DQPC/

under a permissive license [15] that makes it accessible for academic research and *Moving Picture Experts Group* (*MPEG*) and *3rd Generation Partnership Project* (*3GPP*) standardization.

UVG-CWI-DQPC provides both the ground truth target and the synchronized low-quality version to support research on 1) capture quality enhancement (registration, calibration, holes, occlusion); 2) user-perceived quality analyses (artifacts and level of detail); and 3) human behavior studies (gaze and movement). The sequences cover diverse scenarios, from static objects to interacting people, with varying motion and surface complexity. The dataset can be used, among others, to develop quality metrics, enhance point cloud fidelity, and analyze compression effects.

For each sequence, UVG-CWI-DQPC provides:

- A high-quality point cloud sequence from the high-end capture system.
- A lower-quality point cloud sequence from the consumer-grade capture system.
- Intermediary files such as RGB-D images, calibration data, and software for reproducibility and enhancement methods.

This paper is outlined as follows: Section 2 provides an overview of existing datasets. Section 3 details our dual-capture setup and workflow. Section 4 introduces our dataset and its key features. Section 5 explores use cases and Section 6 concludes the paper.

2 Related Work

State-of-the-art solutions increasingly focus on capturing dynamic point clouds in unconstrained, real-world environments involving multi-person interactions and complex settings. However, using high-end studio systems to create datasets for such rich scenarios is often cost-prohibitive and technically challenging. As a result, many recent datasets have been captured using consumer-grade sensors, which trade some geometric/textural precision for scalability, realism, and accessibility.

Table 1 summarizes the point cloud datasets available in the literature. They can be categorized as datasets captured using either high-end or consumer-grade systems.

2.1 High-End Captured Point Cloud Datasets

The 8i Voxelized Full Bodies (8iVFB v2) [4] dataset consists of four sequences recorded with a 42-camera RGB setup, each lasting 10 seconds. Similarly, the Owlii [17] dataset includes four sequences with mesh representations and high-resolution texture maps (2048×2048). The UVG-VPC [5] dataset contains 12 voxelized sequences with associated normals, specifically designed to support the development and benchmarking of *Video-based Point Cloud Compression* (*V-PCC*) technologies. These datasets have been widely utilized for evaluating point cloud compression, interpolation, and reconstruction techniques.

2.2 Consumer-Grade Captured Point Cloud Datasets

FSVVD [6] is a full-scene volumetric video dataset that contains both multiple target people and the interacting environment. Four

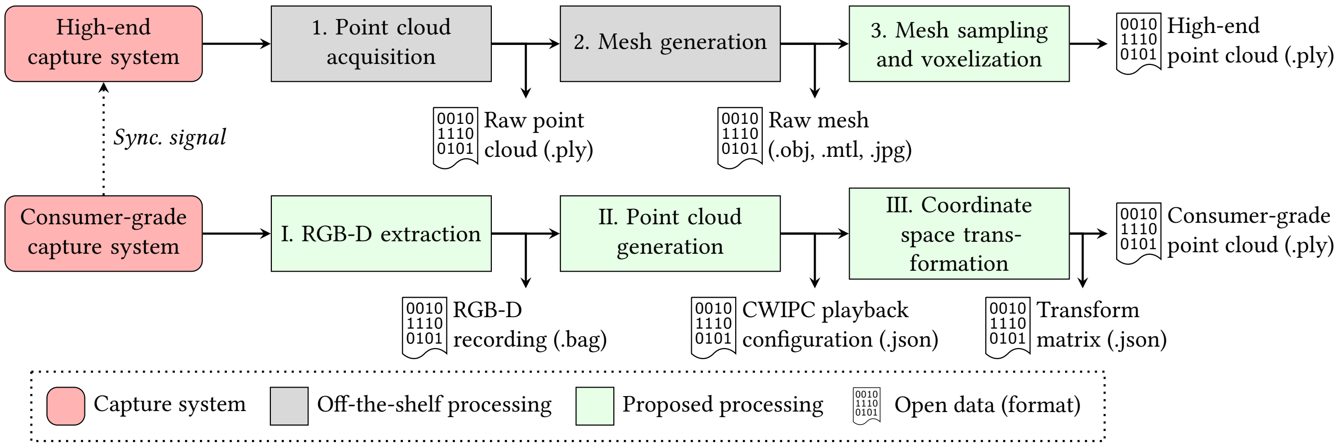


Figure 1: Point cloud capture and processing steps implemented to create the UVG-CWI-DQPC dataset.

real-life scenarios: 'Education', 'Exercise', 'Daily life', and 'Entertainment' are included. CWIPC-SXR [12] focuses on social XR use cases using commodity sensors, depicting multi-user interactions. PCVD [9] captures a broad range of real-world scenarios for behavior analysis and human action research. HUMAN4D [1] offers a multimodal perspective by combining multi-RGBD, volumetric reconstruction, and synchronized audio data to capture 'physical', 'daily' and 'social' human activities. The volumetric depth maps have high synchronization precision because of the intra- and inter-sensor hardware synchronization. This dataset is especially suitable for tasks requiring spatio-temporal alignment across modalities, such as human pose estimation, activity recognition, and volumetric video quality assessment. VsenseVVDB2 [18] is the only dataset that relies on a lightweight capture system based on commodity cameras [11] to produce highly-dense point-clouds. It includes not only point cloud sequences but also their subjective scores and distorted versions.

The UVG-CWI-DQPC dataset proposed in this paper bridges the gap between high-fidelity and real-time point cloud capture, offering valuable benchmarking data to improve and validate processing pipelines for consumer-grade systems.

3 Capture Methodology

The proposed dataset is made up of volumetric point cloud videos captured using two distinct systems:

(1) a high-end capture system, optimized for the highest possible quality without constraints on time or cost, and
(2) a consumer-grade capture system that uses affordable hardware to provide quality comparable to live capture scenarios.

Figure 1 illustrates the implemented capture and processing steps needed to comprise the final dual-quality dataset.

3.1 High-End Capture System

The high-quality dataset was captured using the volumetric capture studio developed by Mantis Vision [16]. The studio setup was composed of 32 (19 long and 13 short) camera units with different stereo distances. Each camera unit included two RGB cameras, an *infrared* (*IR*) projector, an IR camera, and an Intel *Next Unit of Computing* (*NUC*), as specified in Table 2. The camera units were connected through a tree topology using 10Gbps switches, with four camera units connected to each switch, and two switches connected to the rendering computer. The studio was set up to a height of 2.5 m, with a diameter of 3 m, and it allowed scanning of a scene with a height of 2.2 m and a diameter of 1.6 m. To enhance the capture quality of faces, most cameras were located on the upper part of the body.

Table 2: Specification of Short and Long Camera Units

Camera unit type		Long (×19)	Short (×13)
RGB camera ×2	Specification	UI-328xCP-C	UI-308xCP-C
	Resolution	2456×2054	2456×2054
	Stereo distance	∼30cm	∼10cm
IR camera	Specification	UI-314xCP-M	
	Resolution	640×512	
Intel NUC	Processor	Intel(R) Core(TM) i7-8665U CPU @ 1.90GHz	
	Memory	32 GB	
	Hard drive	Samsung 970 EVO Plus SSD 1Tb	

The studio can capture volumetric video at up to 25 *frame per second* (*fps*). For this dataset, the frame rate was set to 15 fps to meet the camera constraints of the consumer-grade capture system. The studio featured 40 LED tubes of 50 W and a Sync LED that flashed and triggered at the recording frame rate. Cameras captured images at slightly different times to avoid IR interference, with opposite cameras exposing at the same time.

3.2 Consumer-Grade Capture System

To acquire synchronized color and depth data, eight Intel RealSense D455 cameras [8] were positioned around the capture environment. Each camera recorded RGB-D images at a resolution of 1280 × 720 pixels, and operated at 30 fps. Seven cameras were set to synchronization mode 2 (slave), and one to mode 1 (master). The signal from the master camera was used to synchronize the other seven cameras and the high-end capture system.

The open-source CWIPC library (version 7.6.5) [2] was used to process the captured data. This library is built upon the *Point Cloud Library* (*PCL*) and integrates various vendor-specific capture libraries. The CWIPC framework enables the creation of end-to-end pipelines for capturing, transmitting, receiving, and rendering dynamic point clouds in real time. Additionally, CWIPC offers tools for aligning point clouds from multiple cameras and includes tunable filters for coarse and fine calibration, point cloud frame grabbing, and visualization.

Despite the advantages of using commodity depth-sensing devices, several limitations remain. Visual artifacts often appear due to inaccuracies in depth measurements, and they become more severe when combining data from multiple cameras. These issues stem from behaviour discrepancy between the cameras, variations in scanning accuracy caused by differing angles or distances to common points within overlapping fields of view, and temporal discrepancies due to asynchronous frame captures. These limitations have been well documented in previous studies [12] highlighting the need for meticulous calibration and synchronization to ensure acceptable volumetric reconstructions.

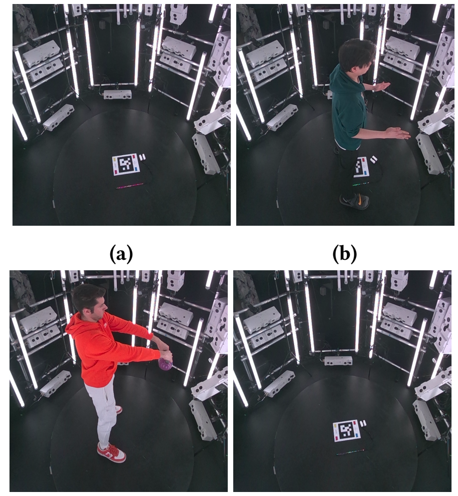
3.3 Capture Systems Synchronization

Synchronizing the two independent capture systems posed a significant challenge, as each system controlled its camera array using a distinct synchronization mechanism. To achieve unified timing, the hardware synchronization signal from the consumer-grade capture system was used to replace the original sync signal of the high-end capture system. This configuration ensured that both systems were triggered simultaneously at the hardware level. To refine the temporal alignment across the recorded sequences, an LED strip was placed within the field of view of both systems. It flashed at capture at 15 Hz, providing a visual marker for frame-level alignment during processing.

3.4 Capture Protocol

Each sequence was recorded according to a standardized capture protocol consisting of four distinct phases as illustrated in Figures 2 (a)–(d):

(a) **Initialization**: An ArUco marker is placed in view, and an LED strip blinks at 15 Hz. The blinking sequence begins with all LEDs illuminated in pink, followed by a gray code counter. The marker and LEDs are used to calibrate the consumer-grade capture system and facilitate the temporal and spatial alignment of the two systems.
(b) **Posing**: A human subject remains stationary for several seconds to establish a reference pose.
(c) **Action**: The human subject performs a predefined action, which forms the content of the final sequence.
(d) **Control**: The initialization phase is repeated to assess a potential synchronization drift at the end of the sequence.



(a) (b) (c) (d)

Figure 2: Different phases of the capture protocol: (a) Initialization, (b) Posing, (c) Action, and (d) Control.

3.5 Dataset Generation

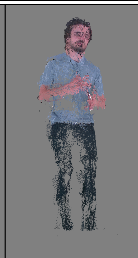
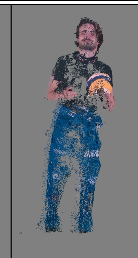
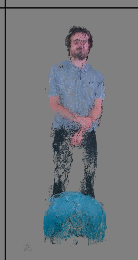
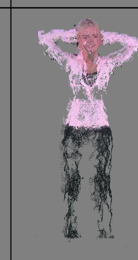
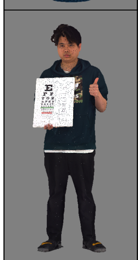
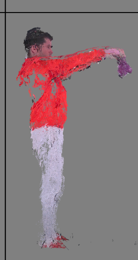
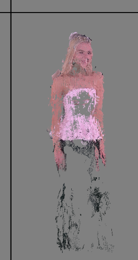
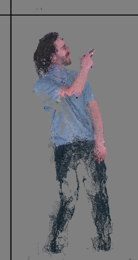
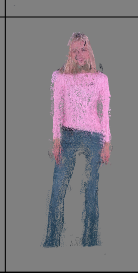
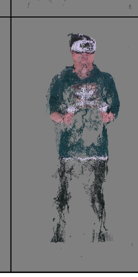
Figure 1 illustrates the processing pipelines for both capture systems. The processing steps of the high-end (1–3) and the consumer-grade (I–III) capture systems are described as follows:

(1) **Point cloud acquisition** is executed by the camera units and the rendering computer. The camera units capture both RGB and IR data, which are fused by the NUCs into a point cloud structure. Both camera data and generated point clouds are sent over the network to the rendering computer that merges them into a single raw point cloud.
(2) **Mesh generation** is used to create a mesh from the raw merged point cloud with off-the-shelf Poisson surface reconstruction algorithm provided by the high-end capture system.
(3) **Mesh sampling and voxelization** generates the high-end point cloud. After the triangle point picking from the trimesh Python library [3], the point cloud is voxelized and color attributes are averaged within the same voxel. For this step, the sampled point number is set to 6 million and the voxel size is set to one millimeter.
(I) **RGB-D extraction** is carried out using CWIPC software to record the RGB-D sequences and store them into bag files.
(II) **Point cloud generation** transforms the RGB-D footages from the cameras into a merged point cloud. The CWIPC playback configuration is created to store the registration parameters used to generate the sequences.
(III) **Coordinate space transformation** is performed to transfer point cloud output from `cwipc` into the same coordinate space as the high-end point cloud. The transform matrix parameters are stored into a JSON file.

4 UVG-CWI-DQPC Dataset

UVG-CWI-DQPC consists of 12 sequences, with a length ranging from 157 to 201 frames. Table 3 lists all sequences and characterizes them with snapshots, names, length, content descriptions, and specific features.

Table 3: Characteristics of the Dual-Quality Point Cloud Sequences in the Proposed UVG-CWI-DQPC Dataset

Snapshot		Name and description	Snapshot		Name and description
High end	**Cons. grade**		**High end**	**Cons. grade**	
		**Name:** BlueSpeech **Length:** 169 frames **Description:** A person delivers a speech while using hand gestures. **Specific features:** Moderate movement; Human; Simple textures			**Name:** BlueVolley **Length:** 171 frames **Description:** A person plays with a volleyball, passing the ball behind the back and between the legs. **Specific features:** Fast movement; Human; Object; Interactions; Complex texture; High occlusion
		**Name:** BouncingBlue **Length:** 157 frames **Description:** A person sits and bounces on a gym ball. **Specific features:** Global movement; Human; Object; Interactions; Simple textures			**Name:** FitFluencer **Length:** 201 frames **Description:** A person stretches their back sideways from one side to the other. **Specific features:** Global movement; Human; Simple textures
		**Name:** GoodVision **Length:** 168 frames **Description:** A person conducts an eye exam, presenting letters on a chart one at a time. **Specific features:** Little movement; Human; Objects; Text; Simple textures			**Name:** Mannequin **Length:** 188 frames **Description:** A mannequin wearing a head-mounted display (HMD) and a T-shirt with various logos stands still. The T-shirt shifts slightly. **Specific features:** Static sequence; Objects; Text; Complex textures
		**Name:** OrangeKettlebell **Length:** 170 frames **Description:** A person performs multiple repetitions of the kettlebell swing exercise. **Specific features:** Global movement; Human; Object; Simple textures			**Name:** PinkNoir **Length:** 201 frames **Description:** A person adopts various poses facing the camera. **Specific features:** Little movement; Human; Simple textures
		**Name:** TicTacToe **Length:** 165 frames **Description:** Two persons play Tic Tac Toe using plastic building bricks. **Specific features:** Fast movement; Humans; Objects; Interactions			**Name:** TrumanShow **Length:** 171 frames **Description:** A person greets cameras all around with smiles and gestures. **Specific features:** Fast movement; Human; Simple textures
		**Name:** VictoryHeart **Length:** 197 frames **Description:** A person greets the camera with friendly hand gestures and forms a heart shape. **Specific features:** Moderate movement; Human; Simple textures			**Name:** VirtualLife **Length:** 196 frames **Description:** A person wearing a head-mounted display (HMD) engages in virtual reality gameplay, involving arm movements and body rotations. **Specific features:** Fast movement; Human; Object; Simple textures

As illustrated in Figure 1, each sequence comes with the high-quality point clouds from the high-end capture system, along with the lower-quality point clouds captured with the consumer-grade capture system. All intermediary files from the consumer-grade capture system such as RGB-D videos, registration information, frame correspondence with the high-end capture system, and software [2] are also provided, allowing broader use of UVG-CWI-DQPC.

The structure of the dataset files is shown in Figure 3. Provided data is regrouped per sequence then per capture systems allowing independent use of high-end and consumer-grade captured sequences. Consumer-grade cameras are limited in capture framerate option, coupled with the limitation from the high-end capture system, the dataset is recorded at 15 fps for the high-end capture system and 30 fps for the consumer-grade capture system. The point cloud sequences at 15 fps, `HE/15fps` and `CG/15fps`, are matched frame-by-frame. In this case, `HE/15fps` can serve as the ground truth of `CG/15fps`. The intermediary files located in the `RGBD` folder can be coupled with the `cwipc` software [2] to regenerate the `CG` sequences. The open-source nature of `cwipc` enables users to directly modify the source code and implement innovative solutions to enhance `CG` sequences.

UVG-CWI-DQPC is available on our website [14] and is distributed under a permissive license [15] that allows non-commercial research and standardization activities. Further instructions are available and can be found on the dataset website.

5 Dataset Applications

The uniqueness of UVG-CWI-DQPC lies in its synchronous capture using both high-end and consumer-grade capture systems, enabling direct comparison between high-fidelity and noisy point clouds. This dual-quality setup provides a valuable benchmark for a wide range of research areas, from computer graphics to human-computer interaction:

- **Point cloud post-processing:**
  By providing high-fidelity ground truth along with intermediate RGB-D data, UVG-CWI-DQPC supports the development and evaluation of advanced post-processing techniques such as denoising, super-resolution, inpainting, and calibration.
- **Perceptual quality assessment:** UVG-CWI-DQPC can serve as a testbed to evaluate objective quality metrics on data captured or generated with different setups. The variation in visual quality introduces diverse real-world distortions beyond typical compression and synthetic artifacts. This diversity enriches training datasets for machine learning metrics and supports subjective quality studies.
- **Behavior in immersive 6DoF environments:** Several datasets [7, 10, 13] provide user behavior data for applications like adaptive rendering and streaming optimization. UVG-CWI-DQPC enables further study of eye tracking, navigation, attention modeling, engagement, and interaction in immersive 6DoF environments [19]. Using a dual-capture system, it allows analysis of how capture quality affects these behaviors.

In addition, the independent dual-quality capture allows these sub-datasets to be used separately. The diverse scenarios, motion dynamics, and rich textures in the high-quality point clouds make them ideal for challenging tasks in volumetric video compression.

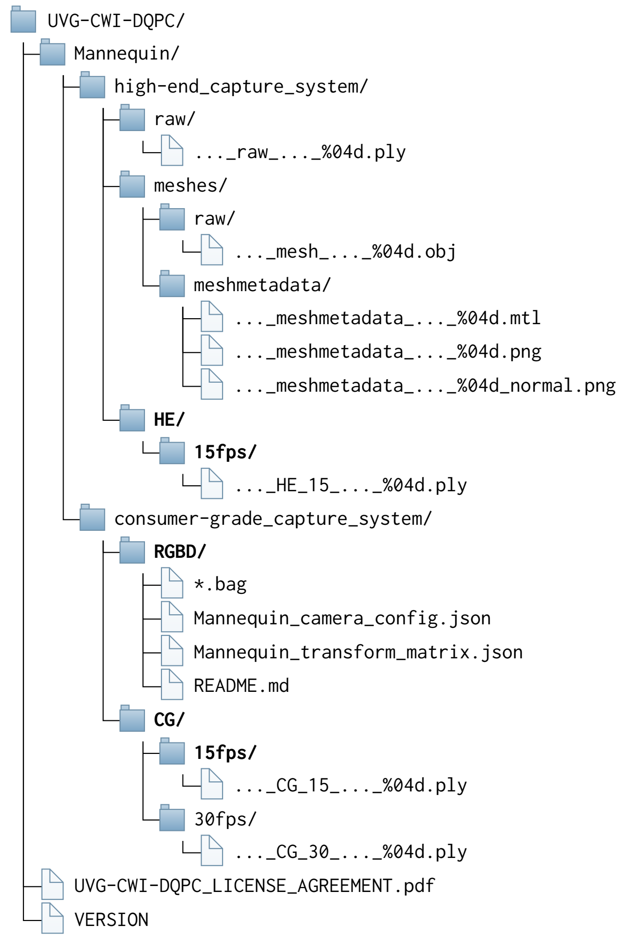


Figure 3: Folder structure of the UVG-CWI-DQPC dataset, illustarted with the Mannequin sequence.

Meanwhile, the RGB-D images enable multi-modal learning and support the generation of point clouds, meshes, and Gaussian splats.

6 Conclusion

This paper presented UVG-CWI-DQPC, a publicly available, dual-quality dataset designed to advance research in volumetric video processing. By synchronizing the capture of 12 dynamic sequences using both a high-end and a consumer-grade capture systems, UVG-CWI-DQPC enables the development and benchmarking of point cloud enhancement, compression, and quality assessment under varying fidelity conditions. In addition to providing high-quality reference data and raw RGB-D captures, the dataset includes comprehensive calibration metadata and intermediate files, supporting tasks such as point cloud reconstruction, temporal upscaling, and cross-modality learning.

Given its scope, content, and accessibility, UVG-CWI-DQPC offers a valuable foundation for future benchmarking efforts or grand challenges aimed at improving consumer-grade volumetric video quality. It paves the way for bridging the gap between practical constraints and immersive user experiences.

Acknowlegment

This work was carried out with the support of Centre for Immersive Visual Technologies (CIVIT) research infrastructure, Tampere University, Finland. This work was funded in part by the XR Simulation and Presence at the Cloud Edge (XR-SPACE) project led by Nokia and funded by Business Finland, and the Research Council of Finland (decision no. 349216). 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.

References

[1] Anargyros Chatzitofis, Leonidas Saroglou, Prodromos Boutis, Petros Drakoulis, Nikolaos Zioulis, Shishir Subramanyam, Bart Kevelham, Caecilia Charbonnier, Pablo Cesar, Dimitrios Zarpalas, et al. 2020. Human4d: A human-centric multimodal dataset for motions and immersive media. *IEEE Access* 8 (2020), 176241–176262.

[2] CWI DIS. 2025. *CWIPC Github repository*. https://github.com/cwi-dis/cwipc

[3] Dawson-Haggerty et al. 2019. *trimesh*. https://trimesh.org/

[4] Eugene d'Eon, Bob Harrison, Taos Myers, and Philip A Chou. 2017. 8i voxelized full bodies-a voxelized point cloud dataset. *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006* 7, 8 (2017), 11.

[5] Guillaume Gautier, Alexandre Mercat, Louis Fréneau, Mikko Pitkänen, and Jarno Vanne. 2023. UVG-VPC: Voxelized point cloud dataset for visual volumetric video-based coding. In *2023 15th international conference on quality of Multimedia experience (QoMEX)*. IEEE, 244–247.

[6] Kaiyuan Hu, Yili Jin, Haowen Yang, Junhua Liu, and Fangxin Wang. 2023. FSVVD: A dataset of full scene volumetric video. In *Proceedings of the 14th Conference on ACM Multimedia Systems*. 410–415.

[7] Kaiyuan Hu, Haowen Yang, Yili Jin, Junhua Liu, Yongting Chen, Miao Zhang, and Fangxin Wang. 2023. Understanding user behavior in volumetric video watching: Dataset, analysis and prediction. In *Proceedings of the 31st ACM International Conference on Multimedia*. 1108–1116.

[8] Intel Corporation. 2025. Intel RealSense depth camera D455. https://www.intelrealsense.com/depth-camera-d455/. Accessed: 2025-05-17.

[9] Jie Li, Shujiao Chen, Qiyue Li, and Zhi Liu. 2025. PCVD: A Dataset of Point Cloud Video for Dynamic Human Interaction. In *Proceedings of the 16th ACM Multimedia Systems Conference*. 284–290.

[10] Jérémy Ouellette, Jashanjot Singh Sidhu, and Abdelhak Bentaleb. 2025. MazeLab: A large-scale dynamic volumetric point cloud video dataset with user behavior traces. In *Proceedings of the 16th ACM Multimedia Systems Conference*. 298–304.

[11] Rafael Pagés, Konstantinos Amplianitis, David Monaghan, Jan Ondřej, and Aljosa Smolić. 2018. Affordable content creation for free-viewpoint video and VR/AR applications. *Journal of Visual Communication and Image Representation* 53 (2018), 192–201.

[12] 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.

[13] Shishir Subramanyam, Irene Viola, Alan Hanjalic, and Pablo Cesar. 2020. User centered adaptive streaming of dynamic point clouds with low complexity tiling. In *Proceedings of the 28th ACM international conference on multimedia*. 3669–3677.

[14] UVG-CWI-DQPC Dataset. 2025. https://ultravideo.fi/UVG-CWI-DQPC/. Accessed: 2025-05-30.

[15] UVG-CWI-DQPC Licence. 2025. https://ultravideo.fi/UVG-CWI-DQPC/UVG-CWI-DQPC_LICENSE_AGREEMENT.pdf. Accessed: 2025-05-30.

[16] Mantis Vision website. 2025. https://mantis-vision.com. Accessed: 2025-05-30.

[17] Yi Xu, Yao Lu, and Ziyu Wen. 2017. Owlii dynamic human mesh sequence dataset. ISO/IEC JTC1/SC29/WG11 m41658, 120th MPEG Meeting.

[18] 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 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 1–6.

[19] Xuemei Zhou, Irene Viola, Silvia Rossi, and Pablo Cesar. 2025. Comparison of visual saliency for dynamic point clouds: task-free vs. task-dependent. *IEEE Transactions on Visualization and Computer Graphics* 31, 5 (2025), 2964–2974.