UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN ICT FOR INTERNET AND MULTIMEDIA

"DESIGN AND DEVELOPMENT OF A SUBJECTIVE TEST METHODOLOGY FOR QUALITY OF EXPERIENCE EVALUATION OF POINT CLOUDS"

Relatore: Dott Federica Battisti

Laureando/a: Gulzhanat Dandyyeva

Correlatore: Dott Jesus Gutierrez

ANNO ACCADEMICO 2021-2022

Data di laurea 05.10.2022

Table of contents

List of Figures

List of Tables

List of Abbreviations

- Absolute Category Rating (ACR)
- Double Stimulus Impairment Scale (DSIS)
- Geometry-based point cloud compression (G-PCC)
- Head-mounted display (HMD)
- International Telecommunication Union (ITU)
- Level of Details (LoD)
- Mean Opinion Score (MOS)
- Network Abstraction Layer (NAL)
- Video-based point cloud compression (G-PCC)
- Virtual Reality (VR)
- Point clouds (PC)
- Polygon File Format (PLY)
- Simulator Sickness Questionnaire (SSQ)
- Quality of Experience (QoE)
- 6 Degrees of Freedom (6DoF)

Abstract

Nowadays, point clouds are considered to be a promising tool to represent visual data for immersive applications. A Point Cloud (PC) is defined as a 3D representation of a given object or scene, composed of multiple points in a coordinate system. Apart from spatial coordinates, they usually have some attributes as color, opacity, reflectance and surface normal. PCs can represent items with wide variety of scales from narrow blood vessels to large objects as building or even an entire city. There is also a possibility to compress PCs in a scale required for real-time and portable use for Virtual /Augmented Reality (VR/AR).

Recently developed imaging sensors enable to acquire richer and denser point clouds, approximately with millions of points, and emphasize the need of efficient point cloud compression and transmission solutions. It is essential to assess the influence and performance of processing algorithms in a point cloud communication system, that can cause degradations and artifacts. Therefore, the most common artifacts caused by different types of processing should be investigated considering their impact on the Quality of Experience (QoE) of the end users. In this context, the experience of the users in a multimedia environment is influenced by human, technical, and context factors. QoE evaluation is essential and subjective tests provide the most reliable assessment compared to a sole use of objective metrics, and provide useful insights to improve current technologies. There are several tests available for 2D images/video, but only a limited number of them is applicable to immersive technologies, which offer new opportunities to the users to explore the content and interact with it. As the field of immersive applications is rapidly expanding, there is a need of more studies and datasets containing subjective quality data. For this purpose, subjective tests is the main focus of this thesis work.

The three primary contributions of this work are as follows: (i) a software for dynamic point cloud visualization was developed, (ii) subjective quality assessment study was performed on 3D point clouds, and (iii) the impact of standard compression rates on the QoE was statistically analysed. Before providing the details of the research and results, a literature review on point cloud compression and streaming, and previous findings are covered in the following sections.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Professor Jesús Gutiérrez and Professor Federica Battisti for their guidance, support and valuable feedback during the implementation of this research project. The feedback and encouragements given by both professors helped me to aspire my research goals and enabled me to learn and improve my technical and soft skills. I also extend my appreciation to all members of GTI Laboratory group for assisting me in experiments and for sharing their experience in the lab. Especially, I would like to thank PhD student Mr. Carlos Cortés for teaching me the fundamental skills working in Unity and assisting in solving encountered engineering problems.

I am sincerely grateful for Polytechnic University of Madrid and, especially to GTI group, to giving the opportunity to conduct my thesis project in their lab with up-to- date facilities.

1 Introduction

1.1 Motivation

Point Cloud (PC) is defined as a 3D representation of the given object or scene, composed of multiple points in a coordinate system by floating point values. 3D coordinates can be also converted into integer representation based a required spatial precision. The application of PCs involves various aspects of life including architecture, medical imaging, 3D printing, video gaming and augmented reality. PCs can represent items with wide variety of scales from narrow blood vessels to large objects as building or even an entire city. There is also a possibility to compress PCs in a scale required for real-time and portable use as for autonomous navigation and Virtual /Augmented Reality (VR/AR). For example, in autonomous navigation point clouds are used to gather data about the surrounding environment in order to avoid collisions.

PCs can be classified according to acquisition methods such direct and indirect. Direct acquisition involves imaging modalities aimed at collecting 3D information in the form of series of scattered points or dense PC where each pixel is associated with a depth value (e.g., LiDARs and time-of-flight cameras). As for the indirect acquisition, methodology involves algorithms that do not directly measure 3D information. This can be exemplified by algorithms that generate PC information from a set of 2D images which match points according to the images of an object, taken at different angles (e.g. photogrammetry).

Previous investigations have proposed several families of 3D point cloud compression methods. In this paper, an overview of the general techniques will be presented as well as existing subjective and objective methods for evaluation of the quality of point clouds. It should be noted that point clouds do not carry topological or connectivity information compared to 3D mesh representations and this is the main obstacle to overcome through finding the suitable compression and transmission technique [1]. In this context, the focus of this paper is to examine in a subjective way, the impact of the different artifacts produced by state-of-the-art point cloud codecs for different types of compression techniques. It has been emphasized that rendering and coding steps take an important role on the final perceived quality, and thus it should be jointly evaluated for the point clouds. This paper considers the impact of the compression artifacts on the perceived quality.

Research on the subjective quality of different compression techniques for point clouds and analysis of the visibility of the distortions caused by coding will contribute to the design of a suitable subjective assessment methodology and choice of the most relevant method. In other words, comparison of results will allow to the design more reliable subjective metrics for point clouds coding mechanisms. Before providing the details of the research and results, a literature review on point cloud compression and streaming, and previous finding will be covered in the following sections.

1.2 Objectives

The main goal of this research work is to evaluate the impact of standard Video-based point cloud compression (V-PCC) technique on the Quality of Experience of the users;

The main objectives of this thesis project are as follows:

- to perform a literature review on existing methodologies for compression of the point clouds and subjective tests
- to implement encoding step for the dataset preparation
- to develop the software for visualization of the dynamic point clouds in Virtual reality (VR) environment.
- to perform a subjective quality assessment study on 3D dynamic point clouds ;
- to perform statistical data analysis based on Mean Opinion Score (MOS) and Simulation Sickness

1.3 Structure

The contents of the paper is organized as follows. Section 3 introduces standard methods involved in the compression of 3D point clouds . Apart from that, the existing objective and subjective quality evaluation methods for Point cloud images are reviewed and Quality of Experience is defined under the effect of possible artifacts. Then, Section 4 describes experimental design including dataset preparation,the equipment and software developments, methods for subjective tests and subsequent data analysis. Section 5 presents the results of the conducted studies and comparison analysis between different conditions. The results' interpretation is given in the Section 6 and, eventually, Section 7 concludes this work by providing summary of the results obtained and future work.

2 Background and Related Work

This section presents the key concepts of Quality of Experience and previous findings on the quality assessment of the point clouds. Apart from that, different metrics are listed for the evaluation of the quality of the point clouds in the areas of subjective and objective tests. Then, the detailed explanation is provided for the encoding and decoding techniques and the available standards are discussed. In addition, the set of the most common point cloud datasets is depicted for subjective tests considerations.

2.1 Quality of Experience

It was stated [2] that the most important criteria in the PCs selection include the following i) the PC density, ii) the semantic content (e.g. PCs from Buildings and People), iii) the PC geometry attributes (e.g. PCs with holes), and iv) the colour characteristics. Apart from these characteristics, during acquisition step, some original point clouds can contain significant amount of noise, as it was illustrated by cultural buildings list which tend to have holes, outliers or not precise positioning. The research emphasized density characteristics of the original point cloud can substantially affect the perceived quality of the encoded and rendered point clouds. Thus, it was concluded that noise and density are essential factors involved in the subjective variables given by users.

It has been concluded in previous studies that the rendering and encoding processes may considerably influence the perceived quality by the users. There are standard assessment tools for perceived quality on a use of immersive technology and perceived differences between rendering methods provided by the International Telecommunication Union (ITU). Some compression artifacts are manually simulated by adding noise or coding and this shows lower efficiency than realistic distortions generated by relevant coding methods (e.g. MPEG PC codecs). Additionally, previous research work on PC quality assessment generally lack of consistent and coherent experimental conditions, given by MPEG and JPEG standardization organizations. It has been observed that previous studies mostly apply a single type of objective metrics, which is insufficient to accurately study the impact of the rendering process on the perceived quality. Therefore, future developments should be considered in the areas of subjective and objective PC quality assessment.

Previous studies have already shown benefits of HTTP adaptive streaming (HAS) application methods as well as G-PCC and V-PCC to balance the streaming quality with the bandwidth consumption [3],[4]. Nevertheless, the influence of these encoding and further rendering on the quality of experience (QoE) should be investigated. QoE can be defined as experience of users in a multimedia environment consisted of the psycho-physical responses to colour, motion, texture, audio, and context. In this definition, experience is referred as an individual's path of perception and interpretation of the given stimulus or event, while quality is defined as a result of an person's comparison and judgment process or ,in other words, the degree of need fulfilment. It involves perception, reflection about the perception, and the further description of the results [5]. Thus, QoE is generated from the level of fulfilment of user's expectations in accordance with utility and satisfaction level of the object or service in consideration of the user's personality and current state.

In accordance with a given user context, it would be feasible to employ real-time estimation of QoE to get feedback and make more relevant decisions on coding tools and infrastructure [6]. It has been suggested that the availability of models for QoE enables to simplify the planning process for the use of the given immersive media by modelling the user experience instead of focusing just on the objective performance metrics. It should be noted that the description of QoE features involve influence Factors with categories such as Human, System and Context, and they are interrelated with each other [5].

A Human IF category is related to any property or characteristic of a human user. These characteristic can show the demographic and socio-economic background, the physical and mental abilities, or emotional state of the user. Human IFs are sophisticated and they affect perception at two levels in terms of early sensory (low-level processing) or higher-level cognitive processing (interpretation) [7]. A System IF category is related to features and characteristics that describe the technical quality of an application [8]. System IFs can be exemplified by capturing, coding, transmission, rendering, and communication of data. A Context IF category refers to factors that support any situational characteristics to define the user's environment in the form of physical, temporal, social, economic and technical characteristics [8]. These factors generally appear in the combined forms. To sum up, QoE is a very complex criteria for evaluation as it involves various interconnected subjective factor on different levels of human perception.

It has been stated that the most accurate indicator of the Quality of Experience is obtained through direct observation of a group of observers [9]. During these experiments, the Mean Opinion Score (MOS) is measured in accordance with an evaluation of perceived quality from a pool of observer. Hooft et al. [4] concluded that namely the subjective analysis contribute to the most reliable results compared to objective methodologies, although subjective analysis tend to be more expensive and time-consuming. It should be highlighted that observers are unaware of the various processing stages that led to the changes in content and several factors can cause the same artifact. Thus, a subjective QoE study is required to examine the PCs and the artifacts associated with them after compression stage, and this would be a primary goal in this research work.

2.2 Subjective Evaluation of the Quality

The quality of point cloud rendering has been assessed through subjective evaluation in a number of configurations with different attributes in terms of color (colored, colorless), rendering (raw points, cube, mesh) and degradation (compression, noise, octree-pruning). Recent literature features two common key aspects as encoding evaluation and double-stimulus assessment. As for the definition of a Subjective assessment, the term describes systematic data collection and consequent analysis of participants' thoughts about an immersive experience and given stimulus. This refers to obtaining an explicit feedback from users in accordance with a specific area of interest of an immersive media. The main tools used during a described procedure involves questionnaires, rating devices or interviews that raise necessary questions about experiences and lead to qualitative and quantitative results.

There are common methodologies employed in the subjective assessment of the quality of images. These image assessment methods are using measurements derived from the direct reactions of users who view the images under the test. As it has been discussed previously, objective metrics might not always completely describe performance of the tested systems; subsequently, it is necessary to conduct both objective tests together with subjective tests.

Overall, the classification of subjective assessments distinguished by the lack or presence of the optimum conditions for tests. The first class is Quality Assessments, which conduct the performance tests under optimum conditions. The second class is Impairment Assessments which involve the tests on the capability of systems to recover quality under non-optimum conditions related to transmission [10]. For the proper procedure of subjective assessments, it is required firstly to choose an appropriate methodology that matches with given circumstances and then objectives of the tests should be identified. In single stimulus methods, a single point cloud is shown to the observer who gives afterwards an opinion on the whole presentation [10]. The materials for test consist of only test point clouds, or both the test stimuli and the reference. As for the reference , it can be presented as a freestanding stimulus for rating purposes. In this categorical judgements, participants view images and allocate an image to one of the categories, which may be show judgements or a presence/lack of the impairment. It should be noted that categorical scales tend to be the most frequently used technique to evaluate the point cloud's quality and its impairment. Thus, this method reflects a set of judgements for scale categories and generate results which can be analysed in terms of the judgement and the information identified (detection of a threshold, ranks of conditions).

A common single stimulus subjective assessment method is Absolute Category Rating (ACR). In this technique each point cloud is assessed individually based on the ACR scale. The labels assigned on each scale are as follows: "bad", "poor", "fair", "good", and "excellent". For calculation of MOS values, these labels are converted into the values 1, 2, 3, 4 and 5. ACR method may also include a hidden reference, which is original unimpaired sequence to be presented along with the impaired sequences, without awareness of the subjects of its presence. Results in this case are based on ratings derived form differential scores between the reference and the impaired types[10].

Another class of the commonly used methodologies in subjective tests refer to DSIS (Double Stimulus Impairment Scale) or, in other words, DCR (Degradation Category Rating. This technique is characterized by observers who first see an unimpaired reference video, then the same video, but in impaired version, and subsequently, they vote on the second video according to the 'impairment scale' (i.e., from "impairments are imperceptible" to "impairments are very annoying"). The double-stimulus method is considered to be cyclic as the viewer is first presented with an unimpaired reference, then with the same image impaired [10]. In usual sessions with duration of approximately of 30 minutes, a a series of image is given to a participant in random order and with impairments covering all necessary combinations. Importantly, an unimpaired image must be listed in the image or its sequences under the test. Once a session is finished, the mean score for each test condition and test image is measured based on the scales judgement from viewers. It should be noted for the impairment scale, the stability of the results is higher for small impairments than for large impairments.

2.3 Objective Metrics

As it was mentioned before,the assessments on the efficiency of an encoding algorithms in terms of quality of decompressed images is prioritized for the purpose of trade-off between amount of data and visual quality. Visual quality of point clouds can be evaluated through objective methodologies. The objective means are performed by algorithms that generate predictions. Previous studies have shown several objective metrics to describe the quality of point clouds [4], such as delay, memory use, and signal-to-noise ratio. It should be emphasized, there is a distinction between the quality of a point cloud object in comparison with an object's derived image, and the rendered field of view.

Objective metrics have shown relevant results in the performance assessment of compression techniques for volumetric data, although these metrics cannot explain how the user visually perceives point cloud objects from a given perspective. The list of these metrics includes common metrics for traditional video streaming such as PSNR, the structured similarity index (SSIM) and the multiscale SSIM [4]. The main function of this list of metrics is to indicate levels of the visual quality of the rendered content, and importantly they consider a background of the scenes. In turn, these backgrounds lead to the lower perceived quality, as users tend to mainly focus on the objects in the front. One of the studies [11] has worked on a removal of the background for images generated from point cloud data using a MATLAB-based tool, but the results have shown a substantial computational complexity for videos. This means that a sole use of objective metrics may not be sufficient for assessment of the point clouds' quality and subsequent QoE.

2.4 Point Cloud Compression and Streaming

Point clouds are 3D scenes and objects with given geometry characteristics (e.g., shape, size, position) and additional attributes (e.g., colour, reflectance and even temporal changes), commonly represented by volumetric visual data [9]. These data can be either generated from computer-based 3D models, or captured from real-world settings using multiple cameras or especial ones, such as LIDARs. It should be noted that volumetric data are mostly represented in the form of polygon meshes or point clouds. Furthermore, volumetric video is a major technology for Augmented/Virtual/Mixed Reality technology and for production of Six Degrees of Freedom (6DoF) viewing possibilities. The main purpose of immersive media is to expand the audiovisual quality and to improve the user experience and interactivity.

The acquisition and consequent steps applied on point clouds are depicted in Figure 1. The captured data (e.g., from a LIDAR-camera), with attributes including reflectance and RGB, colour are fused in a post-processing processes. This fusing step can also involve data filtering, thus reducing possible redundancies or outliers. In result, a point cloud representation is obtained with (x, y, z) coordinates and corresponding attributes with each point.

The main issues associated with acquisition technologies of PCs are related to high storage and transmission costs. The reason is that PCs require a storage of a large number of points for object representation and a high amount of attributes, such as colour, surface normal and reflectance, measured for each point. To cope with this issues, the compression of these data is essential. In this sense, some approaches have been already presented. For example, van der Hooft et al. [4] proposed to use a combination of point cloud compression (PCC) with HTTP adaptive streaming (HAS) technique that are adjusted to 2D and 360° video content. In this case PCC considerably decreases the storage amount through multistep pre-processing and rendering stages. At the same time, HAS enables to obtain the higher quality under the given circumstances through overcoming dynamic network conditions. This methodology has shown efficient results with 6DoF. However, complex methods for media representation and delivery are necessary for volumetric 6DoF and a convincing solution is given by combination of PCC with HTTP as stated before.

2.4.1 Octree pruning and Projection-based encoding

Two common techniques, such as octree pruning and projection-based encoder implemented in the 3DTK toolkit are used to degrade the original PC. The octree pruning

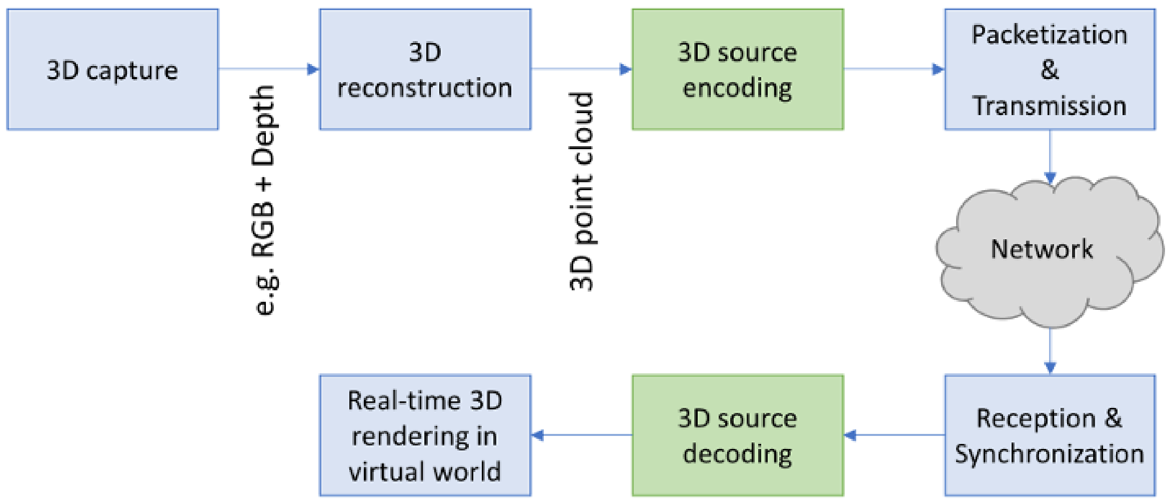


Figure 1: Data path for 3D geometry based tele-immersion application.

method is implemented using the Point Cloud Library (PCL) and it is chosen to show artifacts caused by a standard removal of points. This compression technique involves the contents arranged in an octree structure, which is a 3D information structure with nodes representing a 3D bounding box decomposed into eight children leaves in a recursive way.

In this method a neighbourhood-based predictor is applied and octree's depth is defined in the beginning [12]. Firstly, one root node depicts a whole entire space with the point cloud and afterwards at each level 8 leaf nodes are generated as a result of even division of 3D cell. The binary occupancy code is assigned and used to identify the varying points in a coding process. Level of Details (LoD) usually represents the size of the leaf nodes, and its modification adjusts a resolution of the content accordingly. This can be illustrated by the case if the LoD, increases, the quantity of points of the compressed object tend to decrease [13]. The octree development terminates once a pre-defined maximum depth level has been reached, and this in turn corresponds to the amount of quantization bits for the geometry representation. For the application purposes, the octree technique is known for generating a stable structure of the 3D space, that can be employed in sequential frames to estimate movements [14]. There are three quality levels of model compression: high, medium and low. The quality level of each method was identified from the percentage of remaining points after PCL compression. This indicator is derived through expert viewing per content and tests with different octree depths for every model. Having defined the LoD corresponding to the quality level needed, a target percentage of remaining points can be found [13].

The second method is defined as a projection-based encoder implemented in the 3DTK toolkit. This method generates wide range of panorama resolutions and equirectangular projections in the software used to encode the observed contents [13]. The technique consists of a binary search tree which is applied to identify a suitable panorama size to achieve the required target percentage in 3DTK. This technique begins with 32768x32768 pixels as constraints set by the OpenCV software. A resolution of the panorama was set in order to obtain similar amount of points as PCL compression for each level of quality. Therefore, the final visual quality can vary from that of the target quality level. Results of the studies, presented by Silva Cruz et. al.[13] and based on observers' opinion, have shown through a subjective assessment that the artifacts caused by the octree-based methodology are more preferable over that of caused by the 3DTK projection-based implementation.

2.4.2 Video-based and Geometry-based encoding

There are ongoing point cloud compression standardization strategies provided by MPEG-3DG (3D Graphics group), such as video-based (V-PCC) and geometry-based (G-PCC) which utilize in their principle the above mentioned algorithms. The first class is video-based, V-PCC, and employs 2D video technologies through projecting the points into 2D frames. V-PCC architecture includes following functional tools: patch generation, packing, occupancy map, geometry, attributes, atlas image generation, image padding, and video compression [15]. In other words, V-PCC model produces 3D surface segments through division of the point clouds into a set of connected regions known as 3D patches. Afterwards, each of these patches are separately projected onto a 2D patch without self-occlusions and with limited distortion. The main goal is to obtain a coherent mapping by correlating each point of the 3D point cloud with each cell of the 2D grid [16]. This results in a 2D image representing a geometry and attributes of point clouds, which then compressed with a given video codec [16].

It should be emphasized that the mapping between the point cloud and the 2D grid is not bijective, as an additional binary image called the occupancy map, is required to differentiate the filled (related to a point) and the empty (not correlated with any point) cells of the grid [16]. Encoding process in V-PCC involves the empty space between patches to be filled through a padding function, which produces an image with a relevant smoothness and suitability for video compression. V-PCC technique contributed to the decrease in projection problems including a self-occlusion and hidden surface [15]. Additionally, another benefit of the V-PCC is a high encoding efficiency that allows to transmit a point cloud video over a band-limited network. Therefore, it can be applied for tele-presence purposes: a user with a head-mounted display can interact with the virtual space remotely by the transmission of compressed point clouds. As for limitations, a V-PCC shows some gaps in temporal correlations [17].

V-PCC is suitable for point clouds with an approximately uniform distribution of points and is considered to be the most efficient compared to other existing alternatives [17]. If the clouds are distributed in a more sparse way, a geometry-based class, G-PCC, is commonly employed. This approach is derived from the combination of LIDAR point cloud compression (L-PCC) for dynamically acquired data and Surface point cloud compression (S-PCC) for static point cloud data, due to their similarities. G-PCC approach includes decomposition of the 3D space into a cubical hierarchical structure and assigning every point to a corresponding index of the cube. While V-PCC coding method focused on 3D to 2D projections, G-PCC, in contrast, compresses the content of 3D space in a direct way. For this purpose G-PCC employs the octree pruning method which has been described previously. It should be noted that there is no initial assumption on the input point cloud coordinates, but there is internal integer-based value, converted from a floating point value [15]. Another important feature of G-PCC is characterized by tiles and slices, composed of a set of points (geometry and attributes) that enable a parallel coding [15]. Similar to V-PCC, G-PCC method is limited with lack of temporal prediction tool. Other methods for attribute coding in G-PCC, listed as follows: (a) RAHT [18]; (b) Predicting Transform [15]; and (c) Lifting Transform [19]. The main mechanism of RAHT is to employ the attribute values in a lower octree level to anticipate the values in the upcoming level. As for the Predicting Transform, the main idea is to predict according to a distance for attribute coding. The main idea behind it is a use of a Level of Detail (LoD) form that divides the input points in sets of refinements levels (R) employing a deterministic Euclidean distance criterion. Based on research results by Graziosi [15], the colour encoding efficiency given by G-PCC, the Lifting colour encoding module is considered to be significantly better than the RAHT module [15]. Thus, there is a preference for the Lifting encoding scheme over the RAHT module.

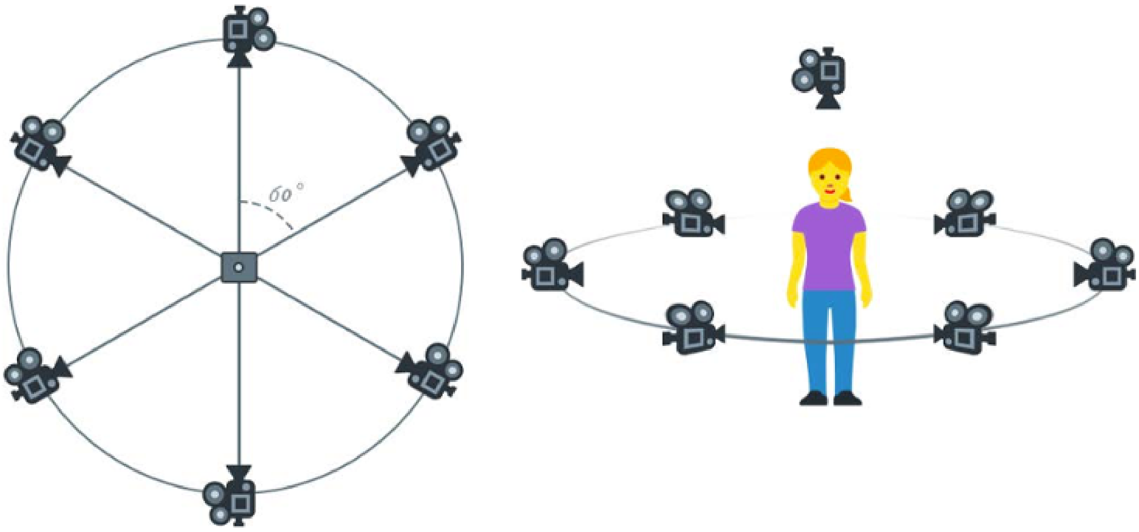


Figure 2: Geometrical camera arrangement during acquisition process of the images [24]

2.5 List of available Point Clouds

There are multiple sets of 3D/volumetric representations widely available for various applications and their capability of reproducing real-life objects from various perspectives is crucial. Therefore, this section focuses on the description of the datasets available and their characteristics, which define their application. Huge amount of data that is necessary for volumetric representations induces challenges, and a need in solutions in acquisition, compression, delivery and rendering. Thus, accessibility to real-life acquired datasets is encouraged, because they enable to evaluate innovative solutions immediately and form a common foundation for comparison objectives. In previous research, volumetric human datasets were acquired using complex techniques such 3D or 4D registration. In turn, this involves Multiview camera settings as shown in Figure 2, for example, the 8i Voxelized Surface Light Field (8iVSLF) dataset, the HUMBI dataset [20], the DFAUST [21], V- SENSE volumetric video quality database [22], TotalCapture dataset [23], and the recently obtained CWI Point Cloud Social XR Dataset [24].

Dynamic point clouds captured by CWI [24] in Figure 3 were captured using multiple synchronized Azure Kinect DK devices, showing people performing usual social activities in real-time communication cases. For the dataset, 4 key scenarios for social experience were chosen, such as “Education and Training”, “Healthcare”, “Communication and Social interactions”, and “Performance and Sports”. All scenarios involved a single actor

Figure 3: Sample frames from CWI Point Cloud Social XR Dataset [24]

who was placed to the scene for obtaining the highest quality possible for each scan and then synchronization was achieved across different actors in multi-person plays.

It should be mentioned that apart from 3D datasets, an acquisition of 4D scan data is being developed recently in computer vision research. The scans of 4D involves 3D scans of moving non-rigid objects, captured over period of time. Registration to common techniques and their extension to 4D is important during acquisition step of 4D scans. The dataset is presented by [21], as shown in Figure 4, consists of high-resolution 4D scans of human subjects captured in motion at 60 fps. The methodology in this research is based on the registration of 3D geometry and texture information for all scans in a sequence. The technique enables to resolve temporal offsets between shape and texture capture. This dataset involves 40,000 raw and aligned meshes, thus extending the FAUST dataset to dynamic 4D data. As for the setup, it was organized by scanning sequences of the object 3D scans at 60 fps using 22 pairs of stereo cameras, 22 colour cameras, 34 speckle projectors and arrays of white-light LED panels. The dataset is represented by dynamic performances of 10 subjects both males and females of different shapes and ages.

The dataset Point XR provided by EPLF group [25] is represented by an assembly of a high-quality point cloud archive of cultural heritage models. The set includes 20 high-quality meshes that were chosen from the platform Sketchfab and then the models were obtained through photogrammetry tools. It should be noted that the acquisition process of the point clouds was implemented in non-controlled environments, and thus, the contents without extra amount of acquisition-induced artifacts (e.g, shading and shadows) were carefully selected. Having loaded the meshes in Meshlab, a texel sampling was

Figure 4: Dynamic FAUST: 4D Dataset for dynamic performances of 10 subjects(3D raw scans and 60fps with shape and appearance) [21]

Figure 5: Frontal view of the point cloud samples from PointXR dataset [25]

implemented with texture resolution of 4096 × 4096 (width × height), and thus, point clouds with high density were obtained and formed into PointXR dataset(Figure 5).

There is another dataset Provided by 8i Labs which is widely used for immersive tests as it has been approved by JPEG Pleno standards [26]. This dataset includes dynamic voxelized point cloud sequences with the name 8i Voxelized Full Bodies (8iVFB) consisting of longdress, loot, redandblack, and soldier, as it is shown in Figure 6. The process of capturing each sequence was implemented using 42 RGB cameras configured in 14 clusters, at 30 fps, over duration of 10 seconds. The spatial resolution for each sequence is a cube of 1024x1024x1024 voxels and the attributes of a voxel include are the red, green, and blue details of the surface colour.

Owlii dynamic human textured mesh sequence dataset by Owlii Inc.[27] is represented by the set of mesh files, each one containing a .obj file for geometry, a .mtl file for material and a .png file for texture. There are in total four sequences in the dataset,

Filename	Frames or fps	# of points	Format, texture	Attributes, Noise	Image
Long Dress	300	800K	.ply dynamic	Color (RGB)	
Loot	300	780K	.ply dynamic	Color (RGB)	
Red and Black	300	700K	.ply dynamic	Color (RGB)	
Soldier	300	1.5M	.ply dynamic	Color (RGB)	

Figure 6: 8i Voxelized Full Bodies (8iVFB) dataset
[26]

Filename	Frames or fps	# of points	Format, texture	Attributes, Noise	Image
Rafa	~40 polys/frame		Dynamic,5s dancing OBJ+JPG 4069x4069.	Color (RGB)	
Levi	~40 polys/frame		Dynamic,5s dancing OBJ+JPG 4069x4069.	Color (RGB)	
Sir Frederick	~40 polys/frame		Dynamic, 1 min monologue OBJ+JPG 4069x4069.	Color (RGB)	

Figure 7: Volumetric video datasets by Invictus [28]

such as basketball player, dancer, exercise, and model. The framerate is 30 fps within 20 seconds period for each sequence containing around 40K triangles. The resolution of texture maps is 2048×2048.

The Volumetric video dataset (Figure 7) released by Volograms has been used for testing purposes in MPEG standardisation efforts [28]. The datasets consists of 3 dynamic sequences known as Rafa, Levi and Sir Frederick. Rafa is a sequence of five seconds with some dance moves. Rafa was scanned at V-SENSE's 12 camera studio in Dublin, Ireland' resulting in meshes of 40 polys/frame and texture images of 4069x4069. Levi' sequence also lasts for five second and includes dancing performance. Levi was scanned in a 60 camera studio in California, US. As for the sequence of Sir Frederick, it consists of 1 minute monologue, and it was captured in a 12 camera studio in Dublin. It should be noted that the number of frames and resolution is same for all sequences.

The Stanford 3D Scanning Repository was scanned with a Cyberware 3030 MS scanner which is swept-stripe, laser triangulation range. Each scan is represented by a range image, shown in the local coordinate system of the scanner. To align these range images, a modified Iterative closest point (ICP) algorithm was used to minimize the difference between two point clouds and further stored in ".conf" file [29]. Afterwards, a combination of the aligned range images was implemented to produce a single triangle mesh using methods by Stanford group. The repository involves 3D models representing some artifacts which have historical, religious or cultural significance as depicted in Figure 8.

3 Experimental Design

This Section presents all the necessary steps for the implementation of the project starting from the preparation of the dataset consisting of the selected point clouds, development of the software for visualization, the procedure of the subjective tests and further data analysis considerations.

3.1 Dataset Preparation

At this stage, the set of point clouds for the experiment was selected, encoded and received by the lab. Afterwards, pre-processing steps were implemented consisted of decompression software implementation and preparation of the test stimuli. The detailed procedures are described in the following paragraphs.

3.1.1 Decompression

In this research work, the dataset was compressed by V-PCC technique. The philosophy behind V-PCC is to make use of existing video codecs in order to compress the geometry and texture information of a point cloud. This process is implemented by converting the point cloud into a set of video sequences, which captures the geometry information and the texture information of the point cloud [30]. It should be noted, that metadata is necessary for interpreting the two video sequences, and it involves an occupancy map and auxiliary patch data.The point cloud V-PCC bitstream is obtained by multiplexing together a video generated bitstreams and the metadata. The metadata itself takes up to 5-20 % of the overall bitstream and it is considered to be a small amount. Figure 9 and

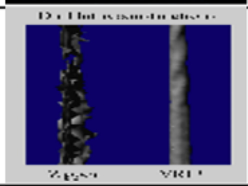
Filename	Number of scans	# of points	Format, texture	Attributes, Noise	Image
Stanford Bunny	10	362,272	Static, object, tar.gz	Color (RGB) contains 5 holes in the bottom	
Drill bit	12	50643 points	Static, object, tar.gz	Color (RGB)	
Happy Buddha	~60	4,586,124	Static, object, tar.gz	hole-free, but contains small bridges	
Dragon	~70	2,748,318	Static, object, tar.gz	numerous small holes	
Armadillo	~ 114	3,390,515	Static, object, tar.gz+ply.gz		
Lucy	47	58,241,932	Static, object, tar.gz+ply.gz	hole-free, but contains small bridges	

Figure 8: The Stanford 3D Scanning Repository
[29]

Figure 10 provide an overview of the V-PCC compression and decompression processes accordingly.

Initially, point clouds are segmented and orthogonally projected onto a set of planes. This results in texture patches, containing the attributes. Sequentially, the patches are packed together to generate a 2D image with the smallest possible number of pixels. At the same time, a depth map is produced to characterize a distances from the projection plane to the points that project onto each pixel of the texture patch. 2D video compression (HEVC) is applied to encode the texture and geometry. As it was mentioned in previous sections, the occupancy map is obtained and is represented by a 2D binary field that shows pixels with composite images (texture and depth) in the 2D patch. Thus, a patch is compressed via spatial quantization together with raster scanning and entropy encoding as it contains meaningful information.

In accordance with Figure 9, the encoder considers an input in a PLY (Polygon File Format) file for each frame of the point cloud, which involves integer positions and colour attributes. The result of the encoder is represented by a binary bitstream captured through the V3C annex-B format (Geneva, CH, 2021). Each V-PCC bitstream contains the following information: (i) V-PCC header, (ii) V3C Parameter Set (VPS), (iii) Atlas Data (AD), (iv) Occupancy Video Data (OVD), (v) Geometry Video Data (GVD), and (vi) Attribute Video Data (AVD). The metadata are given by VPS and the AD, that indicate the point cloud sequence-level and patch-level configurations accordingly. Additionally, each V-PCC bitstream has a sequence of Network Abstraction Layer (NAL) Units (NALUs), which are composed by a header and a payload through a specific syntax that enables transmission over a packet-oriented network.

During the transmission process, a bitstream is demultiplexed into its main components at the server side. VPS is transmitted first and acknowledged by the client. It should be noted, VPS is necessary for the decoding step. The data is transmitted in parallel over the network via the Real-time Transport Protocol (RTP) over User Datagram Protocol (UDP). The decoder in Figure 10, considers the received bitstreams at the client side, reconstructing the V3C format, and producing PLY files, which eventually are sent to the application layer and further visualized by users.

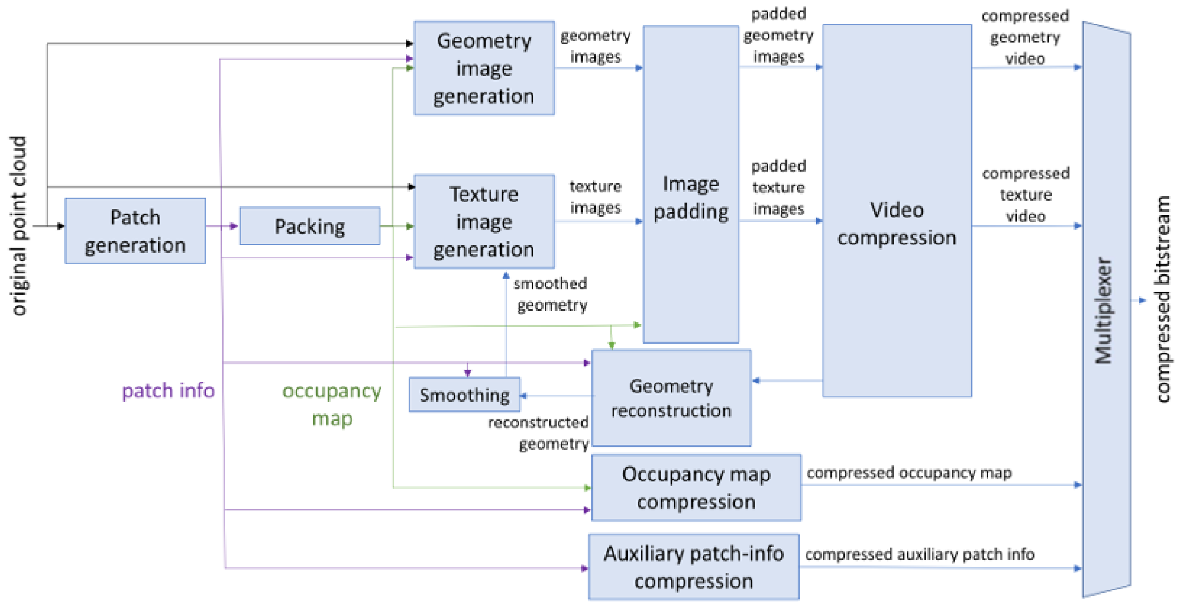


Figure 9: Encoding path for V-PCC methodology [17]

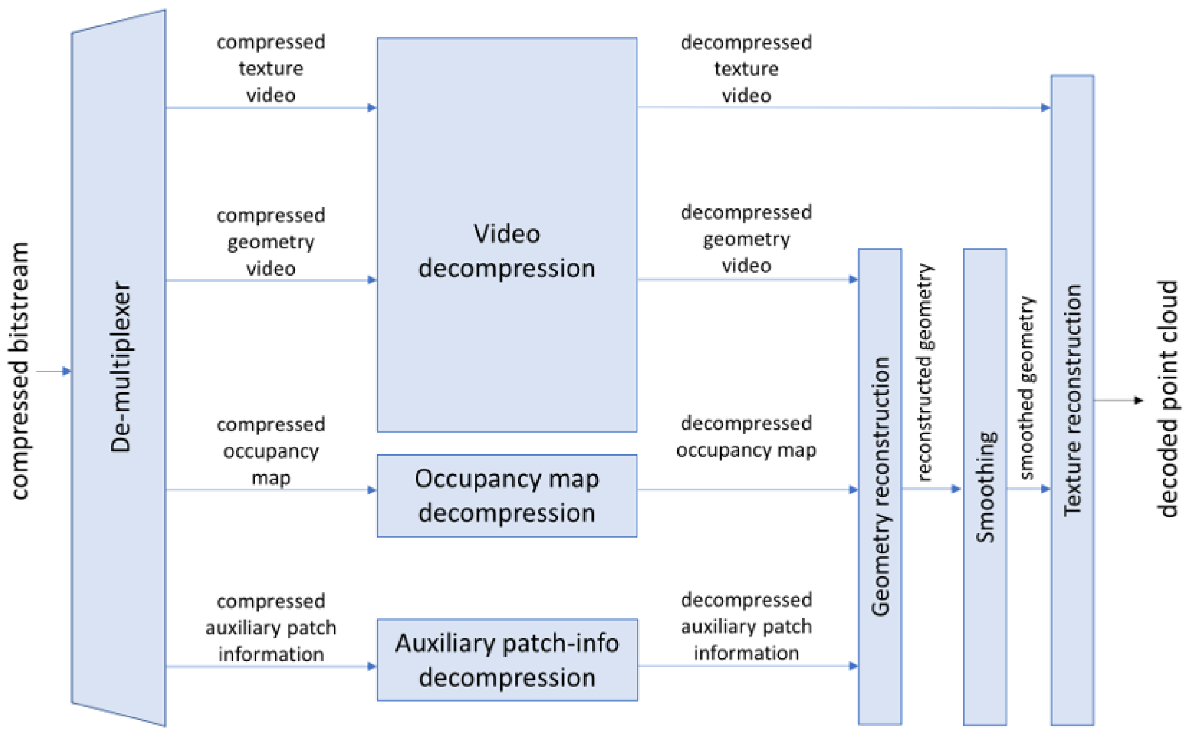


Figure 10: Decoding path for V-PCC methodology [17]

3.1.2 Test stimuli

The encoding procedures were applied on four point clouds by the previous research project developed in collaboration with the Information Engineering of the University of Padova and the University of Rome.

In this work, the encoded point clouds are decompressed and reconstructed first. The decoder source obtained from open-source software available in GitHub [31]. The code applied in the project for decoding step is indicated as follows in Figure 10. The reconstructed point cloud might have some distortions in terms of its visualization depending on the level of encoding configurations.

First of all, a pre-test with experts is conducted and this process includes visualizing various test stimuli of the dataset, and aligning the objectives of the current study with a duration of the whole test. Four point clouds were selected for the analysis, covering a variety of texture and geometry properties as presented in Figure 12 for the experiments. The point clouds were adopted from the dynamic point cloud sequences [32], whose characteristics are summarized in Table 1. The list consist of the following point clouds: longdress (S26C2AIR), loot(S23C2AIR), redandblack (S24C2AIR), and soldier(S25C2AIR) which are demonstrated accordingly in the Figure 12. This dataset was acquired by capturing the whole human body with 42 RGB cameras set in 14 clusters, at 30 fps, for a period of 10 seconds for each sequence. A spatial resolution for each sequence is 1024x1024x1024 voxels (depth 10), provided that it's characterized by a cube. First, the point cloud sequences have been encoded using the V-PCC methodology, through TMC2 version available in Github [31]. The point clouds characterized by five rate points from R1 to R5, where R1 shows the highest compression factor and lowest quality, as opposed to R5 with lowest compression factor and highest quality according to the MPEG CTC. The configurations were applied for All Intra (AI) encoding. It should be noted that the rate point R5 is as close as possible to the original point cloud, as it has the highest quality among other distorted point clouds. Thus, it might be taken as a reference for further subjective tests.

Thus, different levels of distortions for each point cloud were considered and included into dataset. 5 types of the standard distortion rates by MPEG [17] defined by compression of the point clouds resulted in 5 different quality versions of the original point cloud. For the subjective test no reference was set, generating overall 20 Point clouds for sub-

Rate	Geometry QP	Texture QP	Occupancy Map Precision
R01	36	47	4
R02	32	42	4
R03	28	37	4
R04	20	27	4
R05	16	22	2

Figure 11: Standard distortion levels
[17]

jective tests. The details of the selected point clouds are given in Table 1, illustrating the attributes of 5 different dynamic objects.

Filename	Frames or fps	No. of points	Format	Attributes
Long Dress S26C2AIR	300	800K	ply	Color (RGB)
Loot S23C2AIR	300	780K	ply	Color (RGB)
Red and Black S24C2AIR	300	700K	ply	Color (RGB)
Soldier S25C2AIR	300	1.5M	ply	Color (RGB)

Table 1: 8i Voxelized Surface Light Field (8iVSLF) Dataset

The command illustrated in Figure 13 was used in order to decode point clods. As a result, 300 frames for each point cloud at quality factor from R1 to R5 are obtained. Different types of distortions might occur due to the loss of texture, geometry, or occupancy in different frame.

3.2 Equipment

3.2.1 Experimental Setup

Point clouds were visualized using Pico Neo 3 in a test room where the observers could move appropriately. The reason why this device was selected is that Pico Neo 3 allows eye tracking function. This can be used as one of the metrics for QoE assessment of the point clouds visualization. Apart from that, the selected device provides the user with six degrees of freedom and high quality. For the protection purposes from COVID-19, HMDs were covered with a disposable mask, and disinfected before each use. The

Figure 12: Dataset from MPEG Common Test Conditions
[33]

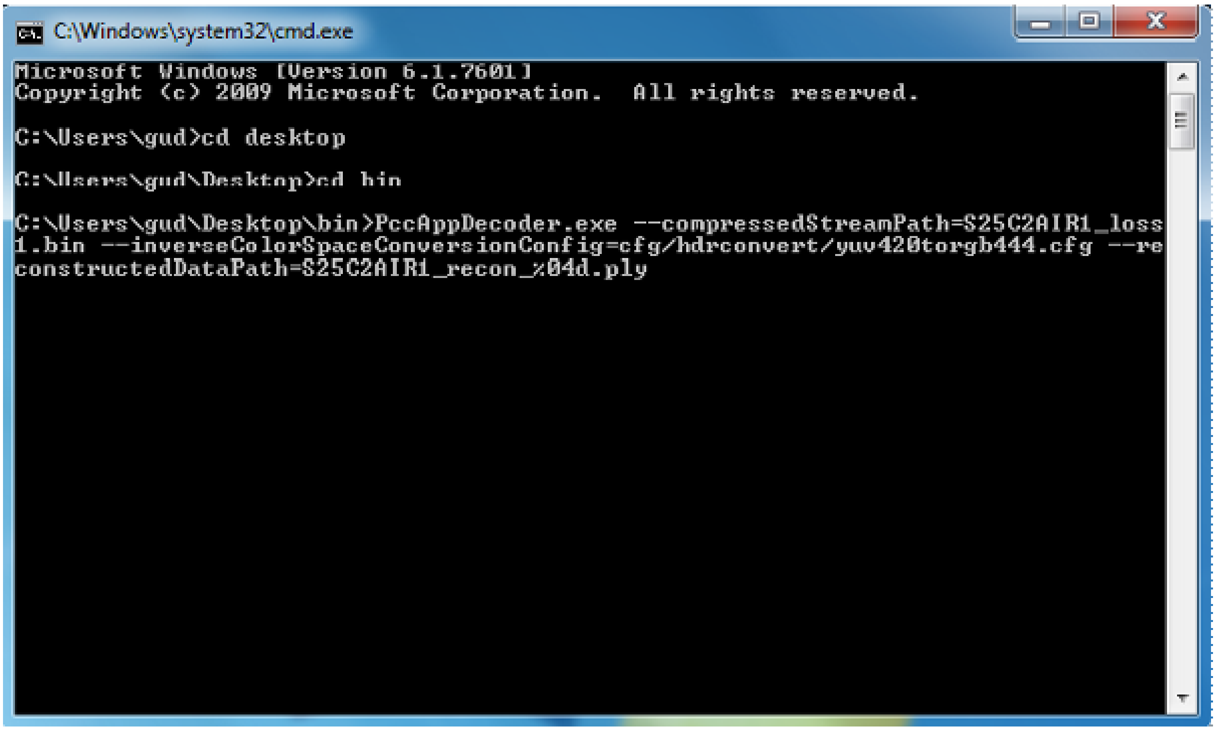


Figure 13: Code used for decoding purposes of previously compressed point clouds

Figure 14: Pico headset and its controllers

environment where the experiments were conducted was a quiet square shaped room, relatively spacious to allow observers to move freely. It should be noted, the tests were conducted in summer and the room temperature was higher that average (30°C), thus ventilator was used.

The virtual environment for experiment contained a virtual parallelepiped-shaped room 10 × 10 × 7 (virtual unit meters) and medium gray walls ([128,128,128]) with low reflectivity, without any distraction. The point clouds were dynamic and they were demonstrated at a distance of 1 meter, in front of the observer (X=0, Y=0.1, Z=1). Observers were able to move around a physical space and visualise point clouds from different sides Figure 15 ,Figure 16.

3.2.2 Unity project development

The tool for their visualization was prepared using Unity3D software (version 2017.3.1f1). This environment is a cross-platform graphics engine developed by Unity Technologies that allows to develop video games and other interactive content, such as real-time 3D animations. The project for visualization of the point clouds and their quality assessment was provided by the University of Rome. The main functionalities have been created through the VisualStudio2017 integrated development environment, using C# as a programming language. In order to process point clouds in Unity, the PCX importer and renderer was used previously and the supported format for visualization is characterized by the binary PLY objects.

It should be mentioned that the project was developed from an already existing

Figure 15: Observer looking within a point cloud

Figure 16: Observer looking at the point cloud from the frontal area

project provided by a student of University of Rome. It should be emphasized, that the Unity project was developed for the HTC Vive HMD, and subsequently the problems occurred with Pico Neo controllers' functions to select the buttons. They were unresponsive initially, then the built-in configurations for the controllers were modified for each Scene in Unity and the problems were solved.

Apart from the encountered challenges mentioned earlier, there were three main issues regarding to the development of the project and changing the setup was required in order to comply the project with the goals set.

First of all, the Unity project was developed to visualize only static point clouds. The project was able to show only one frame of the point clouds. However, each point cloud in the dataset which was applied in this work consists of 300 frames. To overcome this challenge the code was modified through VisualStudio Code. The principle was based on creating a loop which starts with the first frame of the point cloud and then it goes accordingly until it reaches frame 300. The loop was implemented for each point cloud making it possible to connect the frames and to visualize them in a dynamic way. The program was returning to the first frame of the following point cloud. All of the first frames of the point clouds were stored in one folder, and once program instantiate the first frame, the following frames were appended accordingly. Thus, a dynamic point cloud visualization in Unity was implemented through the modification of the codes to show them sequentially with time function.

Another challenge faced with Unity interface is that the project was not able to read the decompressed point clouds, even though their format was 'ply' files. However, the MeshLab software was opening the files without facing any issues. The problem turned out to be in the decompression parameters which resulted in covering the higher scale of the point cloud, but the Unity was not able to process the files with some potential points scattered along the point cloud area. This issue was resolved by saving each frame as a ply file through MeshLab, thus having to save separately 20PC*300frames, which results in a high number of frames. Thus, Python code was applied through integrated platform as PyMeshLab in order to make the process of conversion in efficient way.

For the PC with a standard hard drive, to load the 6000 frames taking up significant amount of storage (each frame of the point clouds occupies more than 20 MB) and to display them in Unity, it would take much longer time. For example, to load 1 frame, at least 1 min was required. In order to solve this issue, the solid-state drive (SSD), which enabled to display dynamic point clouds in the correct frame per second (fps) rate.

Thirdly, the built in platform of the project in Unity was Windows intended for use by HTC Vive. This made it impossible to run the software in the Pico Neo 3, because this type of headset can run the applications built on Android platform. To overcome this issue first of all, the project was attempted to change the platform to Android system through the Software Development Kit (SDK). The Unity project was not displayed in the HMD, despite the fact that the platform was changed and the corresponding application was developed. However, once it was running, the grey or black screen was displayed in the headset. As it was found the documentation provided by the Pico Neo and through discussion with colleagues, the possible reason for the incapability to run the application is that it should have been developed from the beginning at the Android platform in Unity. Therefore, switching to the Android platform from the built-in Unity environment was not feasible for already developed project. An alternative solution is proposed in this work : to use VR streaming system through connection to Mobile hot-spot provided by Wi-Fi adapter. For the streaming purpose, the personal computer was changed to the one with latest operating system as Windows 10 was required for streaming.

SteamVR is a run-time included with the Steam software that provides virtual reality experiences. Once Steam detects a headset connected to a user's PC, SteamVR is automatically opens and it allows to run the Unity project in the headset. The SteamVR SDK's benefit is that it is not bound to be used with a particular headset, but instead it can used with any HMD that supports SteamVR. Subsequently, another issue occurred: steam VR could not detect the headset and the reason was incompatible versions of Pico VR streaming assistant. Thus it should be highlighted that the compatibility problems tend to be frequent if several devices should be inter-operated.

Thus, these three main challenges were resolved during the project implementation.

3.3 Methodology for subjective tests

This section describes the methodology for subjective tests in the experiment. The templates of questionnaires used for the tests will be included in the Appendix section.

3.3.1 Quality

The quality of the point clouds for visualisation criteria was assessed using the ACR (Absolute Category Rating) methodology [10]. The principle behind is to rate stimuli independently based on a five-grade category scale: "Bad", "Poor", "Fair", "Good", and "Excellent".

In order to set the experiment on the even basis for every participant, the same information paper was provided and training was conducted. Each participant evaluated the set of point clouds which were randomized and counterbalanced in order to prevent learning effects.

3D point clouds were accessible from all sides, thus observers were able to freely examine the displayed objects by moving around and taking a closer look. Afterwards, participants were asked to rate the perceived quality of the 3D objects. Each object was displayed for 10 seconds, and then 5 seconds were given to give the score within virtual reality environment. Then, observers clicked on the button 'Next' to access the following object using the Pico controller. Also, they were instructed to go back to the starting position before proceeding with the next point cloud. It should be emphasized that the order of displaying point clouds was randomized for each observer according to the ITU-R Recommendation BT.500-13 [34]. The reason behind it is to avoid the same content to be displayed consecutively as the results can be biased by temporal references.

3.3.2 Simulator Sickness

To assess the discomfort, visual fatigue, and sickness and overall to monitor the physical state of participants, the Simulator Sickness Questionnaire (SSQ) [35] was selected. It enables to measure 16 symptoms which are categorized in groups as oculomotor, nausea, and disorientation. The four-level scale was used to rate physical state(where 0 = "None", 1 = "Slight", 2 = "Moderate", and 3 = "Severe").

3.3.3 Test session structure

For the test protocol, the ITU-R BT.500- 13[34] and ITU-T P.919 [36] general guidelines were followed. The test protocol can be described as follows:

- 1. Welcome (2 min): Briefing and informed consent.

- 2. Setup (2 min): Screening, demographic data and SSQ.
- 3. Training (1 min): Evaluation of a 2 point clouds.
- 4. Evaluation. Session 1 (10 min): 10 point clouds, SSQ.
- 5. Break (5 min): SSQ
- 6. Evaluation. Session 2 (10 min): 10 point clouds, SSQ.
- 7. Debriefing (2 min): SSQ and feedback.

The structure of the whole experiment was divided into three sessions. First of all, the conditions and procedures of the experiment were explained to the participants. The welcome session involved also signing the informed consent by participant for processing his/her data according to the General Data Protection Regulation (GDPR) of the European Union. Afterwards, the Simulator Sickness Questionnaire (SSQ) was filled. This was followed by the Snellen test and the Ishihara test.

Subsequently, a training session was conducted before beginning the subjective test to give instructions to participants, and to provide examples of 3D objects using 2 dynamic point clouds as shown in Figure 17 and Figure 18 (S23C2AIR1 and S23C2AIR5, with the lowest and highest quality levels correspondingly) defined by the test conditions and following voting procedure. These point clouds were not eliminated from the test material. At the same time, participants familiarized themselves with the headset, the interaction area, and the rating methodology. Then, after clarifying doubts and questions by participants, the Session 1 was conducted by visualizing and evaluation 10 dynamic point clouds.Using the controller it was possible for the user to select the buttons present in the scene. There time limit applied on each point cloud is around 10 seconds. The duration of the first session was approximately 10 minutes. Once it finished, the small break of 5 minutes was provided in between session 1 and session 2 for a participant to rest and fulfil SSQ.

After this break, participants were asked if they want to proceed with the Session2. After confirmation, the rest of the test stimuli (10 more point clouds) were displayed in VR environment for evaluation by observers.The duration of this session was approximately 10 minutes. The end of the session was notified by instructor and the last SSQ was filled by the participant. Observers were remunerated for the participation in this study and their feedback was collected at the last test stage.

Figure 17: For training session - frame 1 of S23C2AIR1, lowest quality and highest compression factor

Figure 18: For training session - frame 1 of S23C2AIR5, highest quality and lowest compression factor

3.3.4 Observers

Data collected from each observer included: age, gender, vision (corrected or normal) and experience with Virtual Reality (first time, less than 5 times, between 5 and 20 times, more than 20 times). Thus, this information will be used to characterize the participants and indicate the diversity of the results.

A total number of the test participants were accounted for 20 (10 women, 10 men), with ages ranging between 19 and 29 (the mean of 22.7 and the standard deviation of 2.6). One participant was eliminated because the hardware shut down during collection of the results for the Session 2.This resulted in reduction of the considered observers with a number of 19 (47% females and 53% males). All observers were assessed on the normal or corrected- to-normal vision.Following the ITU-R BT.500-13 Recommendation [34], the Snellen test for vision screening was conducted before the subjective test, to ensure that they have normal or corrected-to-normal vision, while for correct colour perception - the Ishihara test was conducted. Apart from that, participants were requested to fill a questionnaire about their experience in using VR headsets, and according to the results : 74 % of the participant were using it the first time, 10 % of them had used it less than 5 times, and 16% had used it more than 20 times.

This statistical data was expected as the most of the participants were students from the university in different areas from Engineering, such as Biotechnology, Politics and Languages. Participants were recruited through the announcement made in social media platform as Instagram. Thus, 47% of participants are international students during their Erasmus+ for studies mobility. Also, students from the GTI UPM department were recruited through face-to-face invitation.

4 Results

One of the main goals of the subjective test is to compare standard compression distortions applied on dynamic point clouds and thus to identify if there is a threshold below which the artifacts, caused by compression process, are not visible.

During the experiment, the data is automatically collected through the Unity platform. Each participant is associated with a random identification number and 4 text files are created at the start of the application containing the information related to the evaluation session:

- User configurations: text file where demographic information related to the user was inserted(age,gender, name and associated identification number)
- Log (Testing): log files that save the information related to the user, such as ID number, login date. It contains the votes given by users in the various scenes, with the ID of the scene in which the stimulus was displayed.
- Recording: log file containing information related to the instance in time, the position of the viewer in the 3 spatial coordinates (x, y, z), the rotation of the viewer in the direction of the 3 Cartesian axes (rotation with respect to the x axis, rotation with respect to the y axis, rotation with respect to the z axis), the date and the time of the experiment, and the ID of the scene.
- ShuffleList (Testing): text file containing the order with which the user visualizes the stimuli. It is unique to each session as the set of the displayed stimuli is changed randomly at each application initialization.

The raw data of the users' votes, is stored in the "Log_Testing" file. The Pearson correlation screening was carried out on the basis of the ITU-T P.910 [37] recommendation, from which 1 observer was found to be an outlier (r=0.13), therefore this observer's results could be discarded and compared with the initial results in order to identify the impact of the scores given by this participant. However, it should be noted that it is the first time for the participant to wear HMD and being exposed to the VR environment. Therefore, the data provided by this participants still could be analyzed separately and it has been indicated in the standards by ITU-T P.919 [36]. The standards stated that post-screening of subjects can be appropriate or inappropriate according to the purposes of the experiment.For this purpose, questionnaires or interviews after the experiment can be conducted to identify if the subject understood the task. It has been emphasized by ITU-T P.919 [36] the data can be analyzed both with and without the screened subjects.Thus, both cases will be compared in the Results section.

To meet the goals of the tests, the MOS scores were calculated for every dynamic point cloud. In addition, the position of the headset and viewing angle of users were recorded in real time, at the rendering frequency of 90 Hz, in order to analyze movement trajectory of users in VR. The data obtained was grouped by various factors of interest. In particular, three types of analysis was performed:

- Analysis of Quality of Experience metrics: MOS scores given by participants per each distortion level;
- Analysis of the artifacts visibility to identify threshold for each point cloud in accordance with a distortion level;
- Analysis of SSQ: impact of the VR on the physical state of the participant after each session;

4.1 MOS trends

The perceived quality of the point clouds was examined with the Mean Opinion Score (MOS) depicted in Figure 19. The main goal was to analyze differences between 5 different distortion levels and quality score given by participant.

The Analysis of Variance (ANOVA) was used to study the significant differences. In addition, the Pearson correlation coefficient was calculated to identify outliers and the results are given with and without considering the identified outlier. It should be reported, the considered level of significance was 0.05.

Table 2 summarizes the results of the ANOVA test given by the p values for each point cloud with different conditions (coding rate points). According to the results, level of significance requirements are met between different conditions within each point cloud quality evaluation and thus results, can be considered for further analysis.

Point cloud	p value
S23C2AIR	7.27E-10
S24C2AIR	0.012
S25C2AIR	0.00014
S26C2AIR	0.00007

Table 2: ANOVA test results for each dynamic point cloud with different conditions

The correlation function was obtained between the scores given by observers and mean values (Figure 20). Through the comparison of the Pearson coefficient, 1 outlier (with r=0.1) has been detected and removed. Thus, the new results for MOS trends are indicated in Figure 21. According to analysis, the trends for MOS have been conserved

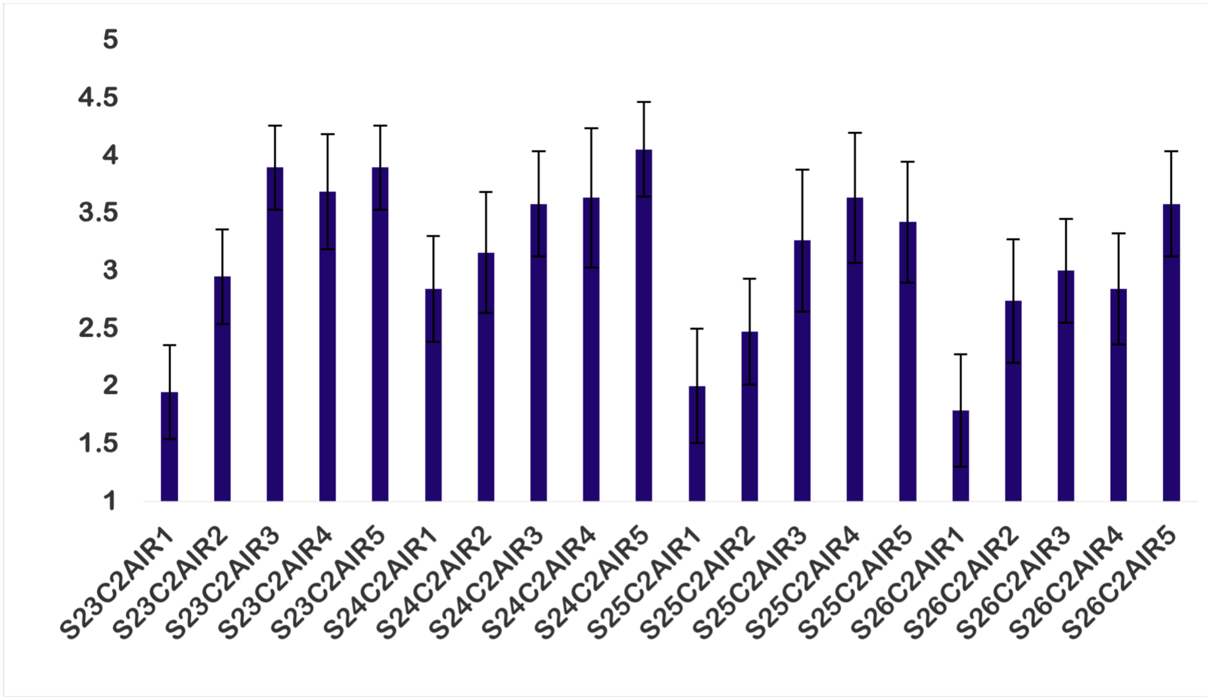


Figure 19: Mean Opinion Scores given by observers for each dynamic point cloud

and no significant changes in the graphs are observed.Taking into consideration, the ITU-T P.919 recommendations [36], the outlier's data has been added back into the results. Furthermore, 1 participant was identified with a slight protanomaly, but the Pearson correlation r=0.6652 has shown that these results are reliable, provided that the observer has given the feedback of clearly differentiating the different quality conditions during the test.

The analysis of the MOS trends for all considered V-PCC distorted point clouds are discussed in the following paragraphs. It should be noted that for the point cloud S23C2AIR (loot) there are significant differences between groups with first two distortion level (e.g., R1 and R2) and the rest of three levels.These results are expected as the quality of the point clouds are much higher for the distortion levels of R3,R4 and R5. It is essential to point out, that the relatively lower results of MOS for distortion level 4 were obtained, probably, due to the error of loading only the first 200 frames of the point cloud and the subsequent point cloud 'freeze' issue. Thus the test probably should be repeated for this specific point cloud with a distortion level of R4.

As for the S24C2AIR point cloud (redandblack), the trends of MOS are slightly different compared to the S23C2AIR (loot). The highest MOS value was obtained for the distortion level R5, which is, at the same time, the highest of obtained values among all

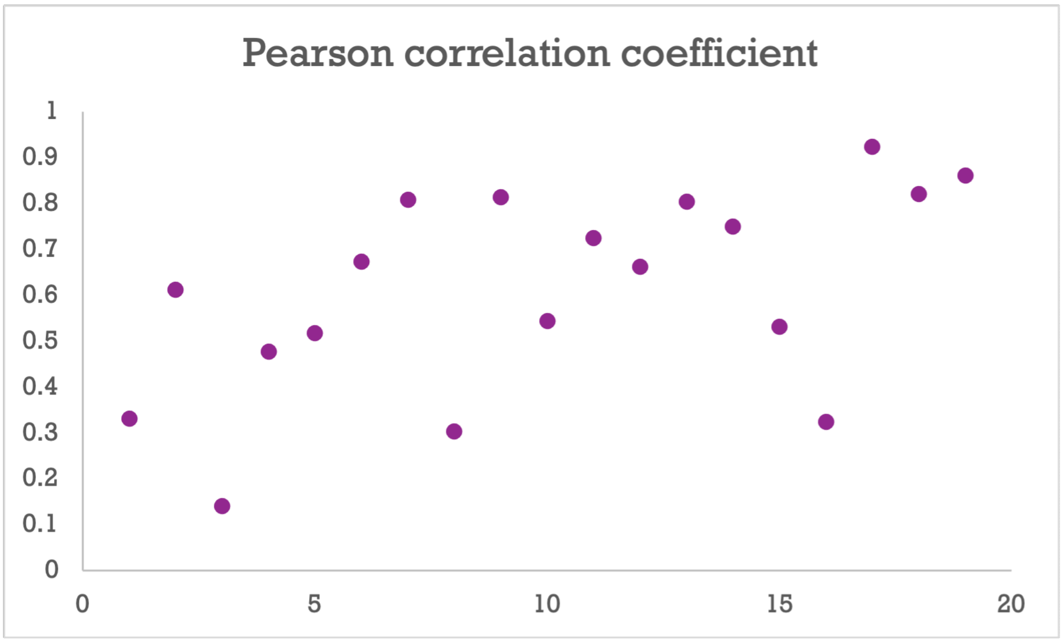


Figure 20: Detection of outliers

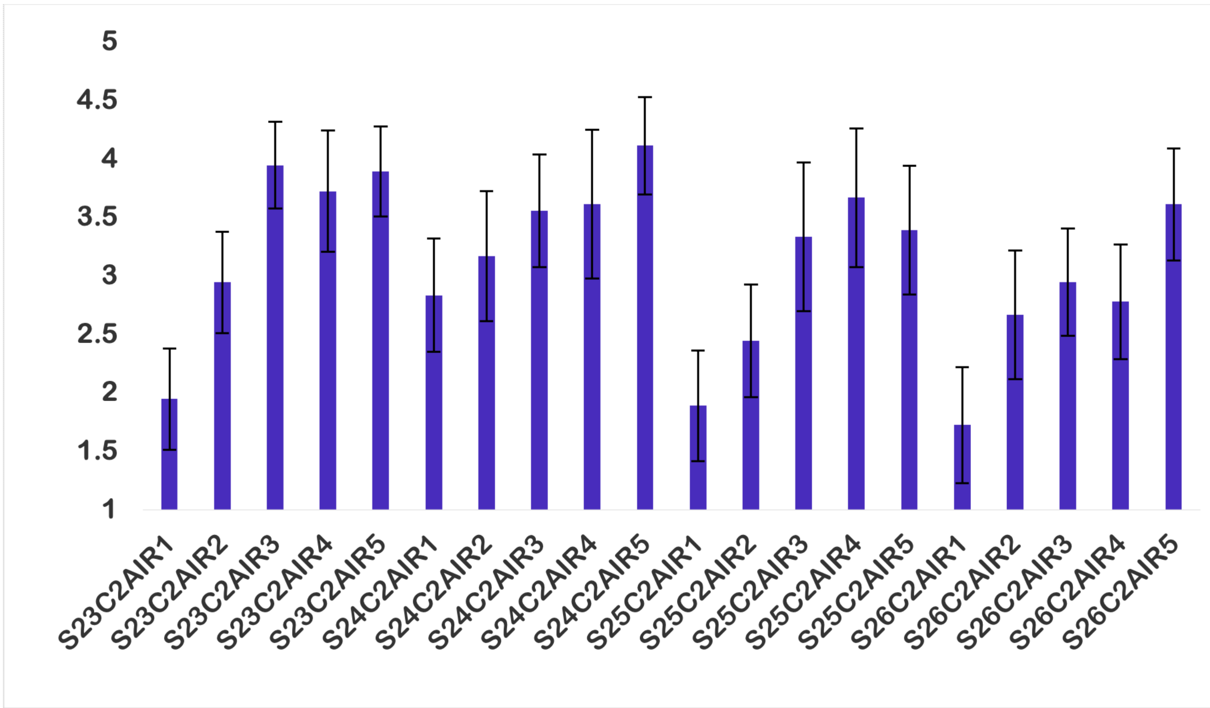


Figure 21: MOS results after removing an outlier

other point clouds. It should be noted, that the average maximum value for the highest quality point cloud is 4 (characterized as ’Good’) , despite the quality of the level R5 point clouds is supposed to be ’Excellent’.

As for the S25C2AIR point cloud (soldier), the MOS value for the level R5 distortion is lower than for the level R4, and similar with level R3. This unexpected trend can be explained by the massive amount of points that the ’soldier’ contains, which is the highest among other point clouds (as depicted in the patterned uniform shown on the point cloud). Therefore, it might be more difficult to evaluate the quality of the given point cloud. However, it should be noted that the general lower trends for compression levels R1 and R2 are conserved.

As for the S26C2AIR point cloud (longdress), the MOS values for the first four levels is much lower than for other point clouds. This trend is expected as, the point cloud contains various colours (highest number among others) and different patterns. Once compressed, those attributes might be indistinguishable and vague for observers. Therefore, it might be challenging for observers to assess the quality of the given point cloud. The highest score was obtained for distortion level R5, and is accounted for 3,5.

Thus, distortion that caused least discomfort was distortion level 3 for S23C2AIR,level R4 for S24C2AIR, and level R5 for S24C2AIR and S26C2AIR. Overall, the highest quality of all point clouds was perceived as "good".

4.2 Threshold for QoE

The graphs for separate point clouds Figure 22, Figure 23, Figure 24 reveal that starting from to the third level of distortion,the differences between between trends become insignificant. For instance, the confidence intervals of three distortion rates for S23C2AIR are almost overlapping, with very close average values (t value is 1.73 between R3 and R5). The similar pattern can be observed for the case of S24C2AIR (t value is -1.83 between R3 and R5) and S25C2AIR (t value is 1.73 between R3 and R5). The general trend of the increase of the QoE metrics with increase of the distortion level is conserved. However, for the S26C2AIR point cloud (longdress), the perception rates for the compression level starting from R2 to R4 were almost equal, thus those qualities might be similarly perceived and lead to the lower QoE among observers. As discussed earlier, the similar factors regarding the colour and pattern might cause this trend in this case too.

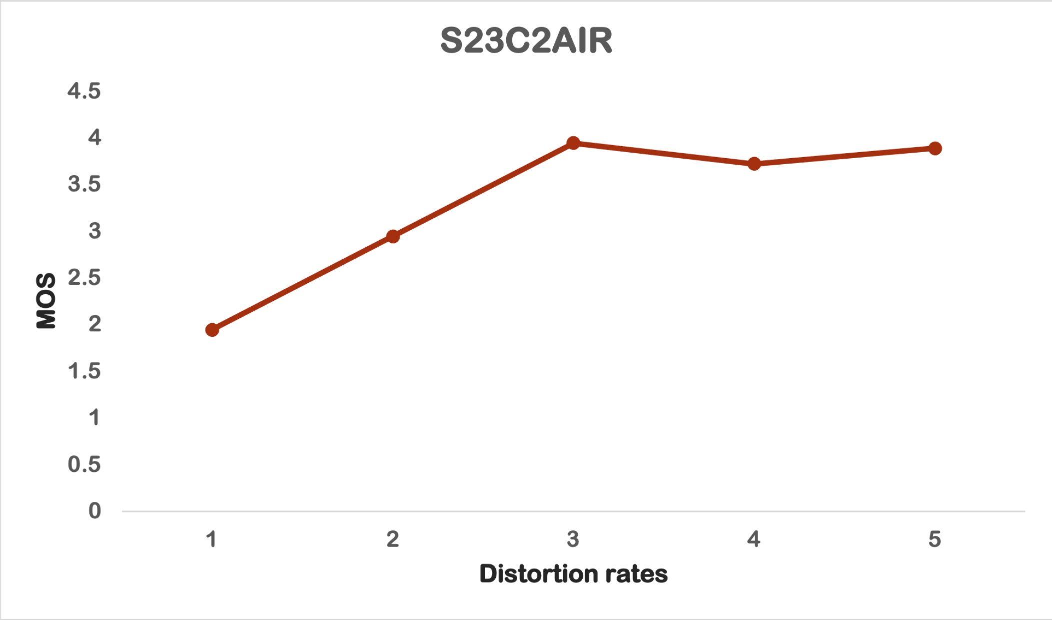


Figure 22: MOS scores given for S23C2AIR point cloud in accordance with the distortion rate

Overall, according to the results, the threshold should be set at the distortion level R3, because starting from this level a user might not differentiate significant artifacts among three groups. Furthermore, by evaluating the trends for the lower compression levels as R1 and R2 , it is possible to note that the trends tend to be relatively low and then from level three a spike can be noticed among three point clouds except S26C2AIR 'longdress'.

4.3 Simulator Sickness

As previously presented, the SSQ is a tool to evaluate the simulator sickness and general discomfort of the participants. The Figure 26 presents the distribution of the Total Score (TS) taking into consideration the two sessions of the subjective tests and the reference before the tests. As can be observed, the TS values obtained during the session are relatively low and only one participant has indicated relatively significant amount of TS equal to 48.62, but it should be noticed that the reference value calculated for the user was already at 18.7. The majority of participants did not have any discomfort before the tests (57%); after the Session 1 - the prevailed TS was accounted for 0 or 3.74 (1 symptom with a level 'slight'); and after the Session 3 the most of the scores were same as during the break. Therefore,it could be stated that point cloud visualization in VR environment has not led to significant changes in physical state of the users and did not

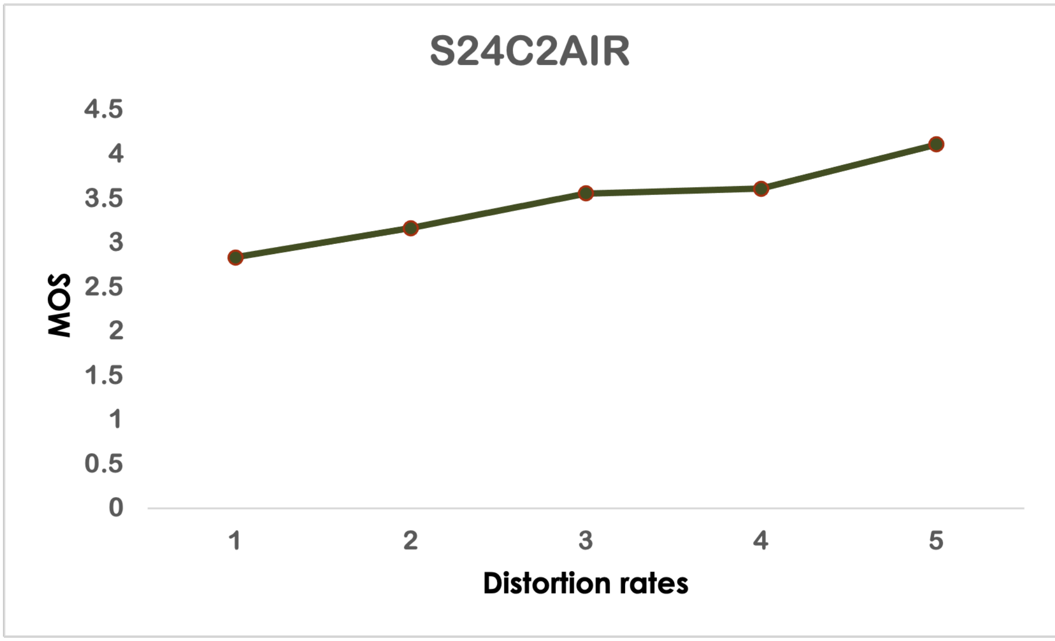


Figure 23: MOS scores given for S24C2AIR point cloud in accordance with the distortion rate

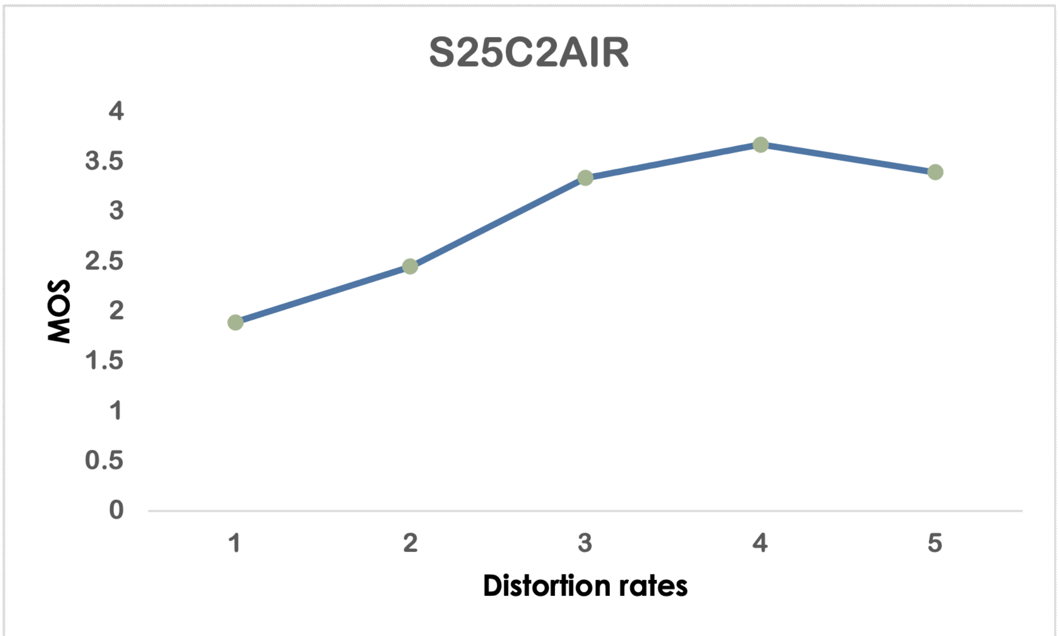


Figure 24: MOS scores given for S25C2AIR point cloud in accordance with the distortion rate

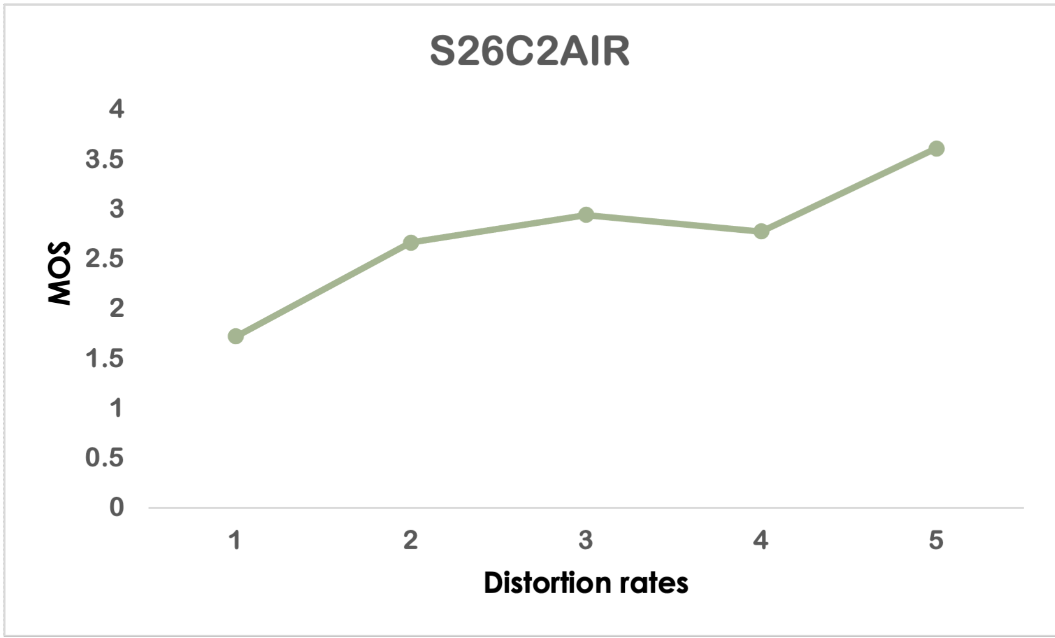


Figure 25: MOS scores given for S26C2AIR point cloud in accordance with the distortion rate

cause any discomforts.

5 Discussion

This research project elaborated on the importance of impact of the compression artifacts on the Quality of experience of users while using immersive environments.

- QoE is still low for the high quality point clouds

  Despite the R5 level was relatively close to the original point cloud in terms of the quality, the maximum scores obtained for this condition are considered to be 'Good'. Participants were prone to give the score 4 to the highest quality rather than 5. This can be explained by subjective factors related to the use of TV, monitors as well as the lack of the clear set of criteria given to the observer to assess a point cloud. In addition, the possible explanation might be that 74% of the participants were using VR for the first time. Therefore, the lack of experience in immersive media might decrease the trends for the visualization of the point clouds with a relative high quality. It must be highlighted, different point clouds are perceived differently in terms of ethical norms and aesthetic preferences. For example, an observer might give a lower score for S25C2AIR (soldier) due to the fact that the dynamic point

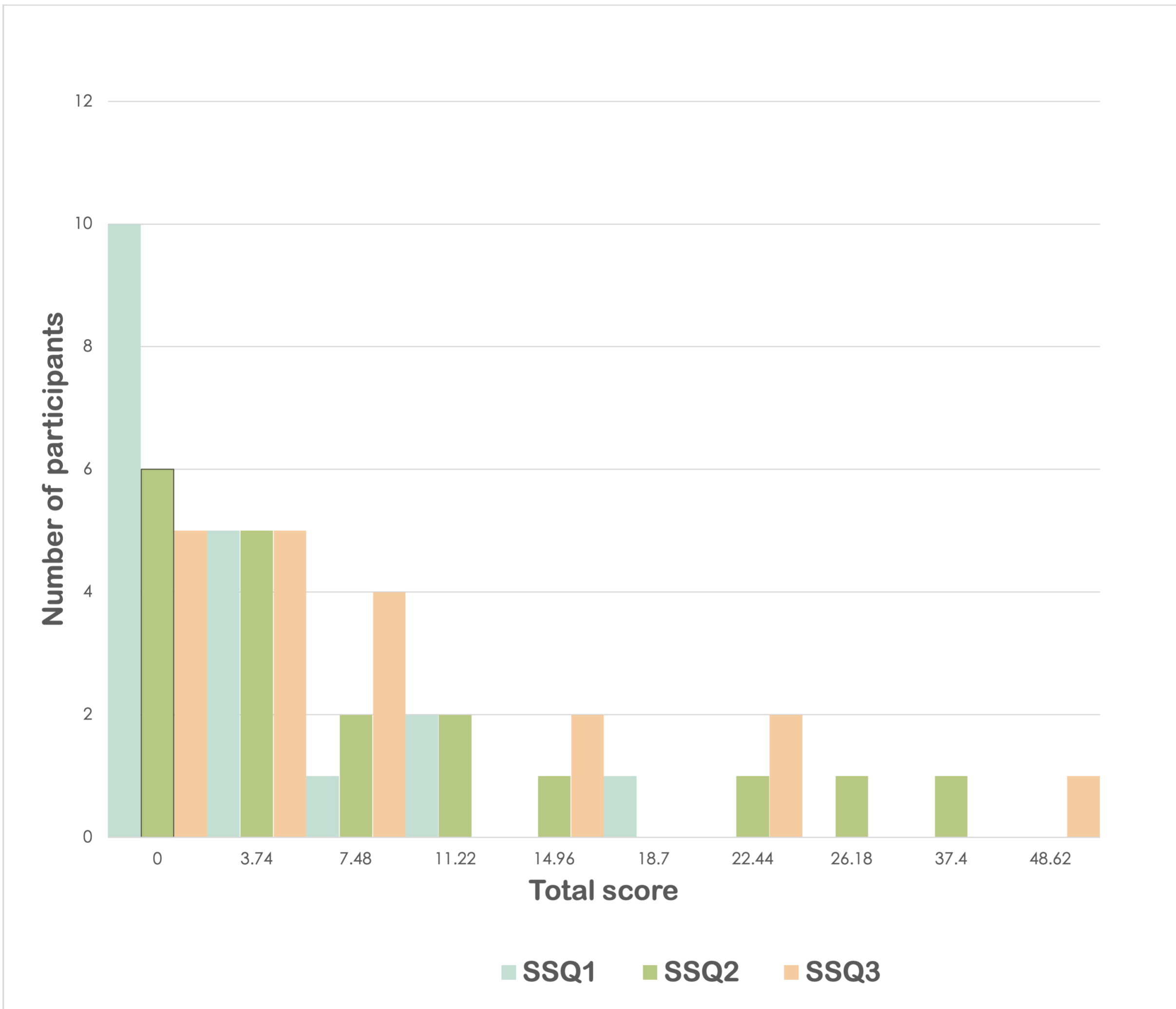


Figure 26: SSQ indicator given by each participant before, during and after experimental session

cloud is using a gun and thus, some ethical questions may affect the observer's opinion. Thus, psychological reasons tend to be involved in the assessment of the point clouds.

- Threshold for higher QoE is characterized by V-PCC distortion level 3

  As expected, significant differences were observed in the overall perception of the point clouds with various quality factors. However, some common patterns have been observed among different conditions as it can be interpreted from the pattern observed for the three dynamic point clouds from R1 to R5. This could possibly set a value of threshold to obtain a proper QoE and initiate new standards for future compression techniques. It should be mentioned, that the results cloud be interpreted in the way that compression level R3 and compression level R5 are perceived in a similar way by observers. However, the data is not sufficient for this conclusion to be made. Apart from MOS values, the eye-tracking and motion trajectory analysis should be performed to analyze behaviour of the user in a response to the specific distortions. At the same time, the test should be repeated several times among groups with different demographic data. The objective metrics might be needed to support the results obtained by subjective tests through comparing the technical details between two conditions.

- VR envioronment did not cause significant discomfort

  The SSQ results have confirmed that the exposure to the VR for a short periods of time did not cause any significant effects on the overall state of the participants. The only common discomfort reported was the eyes strain, but it tends to be a general consequence of using the technology. Moreover, it should be noted that most of the participants were interested in attending the subjective tests and wearing the HMD to be exposed to the Virtual environment. The interest of the participants is highlighted by the fact recruitment of the participants was done without facing any difficulties during 3 consecutive days of the subjective tests. Apart from that, participant show positive approach in exploring the technology and indicated interests in future studies. The high level of involvement of participants can be explained probably by the age range from 19 to 29, and this means, the next subjective tests should cover broader range of participants.

6 Conclusion

This research work accomplished to implement subjective tests on standard V-PCC distortion rates. ACR experiment was adopted to evaluate the QoE metrics of on 4 dynamic point clouds. The possible outcomes of this study are as follows: the lowest MOS values were obtained for the sequence with high number of patterns and colour range; threshold for the higher QoE is characterized by V-PCC distortion level 3; and VR environment did not cause significant discomfort among paticipants.

The MOS trends are not enough for the assessment of QoE, the movement trajectory should be taken into account. Thus, as the future work, the two-dimensional histogram of the overall movement pattern will be generated to evaluate the most common observation areas by users.

As it has been mentioned before, one of the reasons why Pico Neo 3 was selected for the research is that it allows eye-tracking capabilities. For these purposes, the application should be build in the Android platform through packages available from Pico for Unity. However, the change to the Android system was not implemented. Thus, due to the time constraints and incompatibility issues, the activation of the Eye-tracking functions will be postponed for the future work.

Apart from compression related distortions, the dataset with transmission issues characterized by packet loss rates, was planned to be considered for this study. However, the issues with encoding stage were reported by the research group of the University of Rome, from which the current dataset is obtained from. Thus, only the compression distortions were used for the dataset. The future work should consider the V- PCC encoded dynamic point cloud streams in error-prone network environments and tested using subjective methodology. It is essential to understand how network congestion would affect the transmission and the quality of the decoded point clouds in order to update standards on the emerging contents.

A Impact

This section of the work describes the possible impact of this research in several aspects of life.

• Environmental impact: during the project development, several devices were used and all of them were plugged into the main power to charge or to allow the device to operate. As the implementation of the research project required high amount of energy consumption, it may adversely affect the environment.

• Ethical and professional responsibility: ethical guidelines are essential in this project, since participants might experience discomfort, dizziness, sickness and other physical discomfort. Therefore,the subjects were informed of possible discomforts which can be resulted from exposure to the stimuli. The subjects were given the opportunity to be withdrawn fro the test at any moment. Also, participants were asked to be the part of visual screening required for the subjective test, and this, in turn, involves collection of the health data, which requires protection and responsibility of its processing by the instructor. Thus, the respect for data protection legislation is a key consideration during implementation of the research project.

• Social impact: the research could affect society in various areas. Possible impacts could cover the following areas in medicine, civil engineering, education and entertainment industry:

a) high-quality visualization of the organs could lead to precise diagnosis of diseases and higher prevention rate;

b) visualization of the engineered constructions and identification of possible defects may prevent safety problems in buildings;

c) training for specific positions, such as medical doctors, teachers and lab assistants through application of Virtual reality

d) development of the VR/AR games and its integration with entertainment and/or education fields

• Economic impact: use of VR and better visualization techniques would allow to organize events for companies from distance mode, thus saving the transportation costs (consequently, reducing environmental consequences described by transport emissions). Also, the high quality visualization of the point clouds would be applicable in the field of the industry development. This can be described by application of the VR techniques to monitor the proper work of the plants and technologies from distant mode and an efficient detection of the operational issues. Thus, it could reduce the labour input, requiring less financial spendings on the labour.

Costs for the project implementation:

- Computer (Intel Core i9-11900K 3.5 Ghz,Samsung 980 PRO 1TB SSD PCIe 4.0 NVMe M.2 GeForce RTX 308) - 2.777 EU
- Pico Neo 3 Pro Eye - 907.5 EU;
- Remuneration for 11 participants - 110 EU

The total costs, thus, are accounted for 3857,5 EU.

B Materials used for subjective tests

Information sheet

Design and development of Subjective test methodology for Quality of Experience of point clouds

Information sheet provides the basic information about the purpose of a study and the importance of the data collected during the experiment.

Main researcher	Gulzhanat Dandyyeva
Institution	Grupo de Tratamiento de Imágenes (GTI) – Universidad Politécnica de Madrid (UPM)
Address	E.T.S. Ingenieros de Telecomunicación, Edificio C, Desp: C-306, Av. Complutense, 30, 28040 Madrid
Contact information	Jesús Gutiérrez Sánchez, jgs@gti.ssr.upm.es (Post-doctoral researcher)
Project	Project supported by the Spanish Administration Agency CDTI (IDI-20200225)

The aim of this test is to observe and evaluate Quality of Experience of different compression conditions performed on Point clouds.

Point Cloud (PC) is defined as a 3D representation of the given object or scene, composed of multiple points in a coordinate system by floating point values

The test has a duration of approximately 30 minutes. The experiment starts with the pre-questionnaires: a personal information, and the initial SSQ (Simulator Sickness Questionnaire). The rest of the experiment is divided into 2 parts, where you will watch a series of dynamic point clouds of different qualities. Each sequence has a duration of 10 seconds. Each session has a duration of approximately 10 minutes. We ask you to observe each point cloud and, once it is finished, rate its overall quality using the rating interface. You should evaluate the quality of each point cloud using the following quality scale: "5. Excellent", "4. Very good", "3. Fair", "2. Poor" and "1. Bad". After quality evaluation, the next button should be pressed **once instructed by researcher**. The end of each session will be notified by instructor and then you should remove the HMD and complete a post-questionnaire. During the test you will be standing, so you can freely rotate to explore the whole 360-degree content.

Before starting the formal test, you will do a preliminary perceptual test to check your vision (visual acuity, color vision, etc.). Then, you will do a training session with some point cloud example to familiarize with the evaluation method, the interface, and to have a reference of the highest and lowest of available qualities. During this training session, do not hesitate to ask the experimenter to adjust the HMD (clean the screen and lenses, etc.) and any other question or doubt you may have to fully understand the test.

After each session, we will ask you to fill SSQ. If during the test you feel any persistent problems (headache, dizziness, etc.) do not hesitate to indicate it to the experimenter. If you decide to participate in the experiment, it will be voluntary and you will have the right to withdraw from the study at any time.

Impact of the experiment

QoE evaluation is essential and subjective tests provide the most reliable assessment compared to a sole use of objective metrics. Research on the subjective quality impact of different compression combinations for point clouds and analysis of the visibility of the distortions caused by coding will contribute to the design of a suitable subjective assessment methodology and choice of the most relevant method.

Data Treatment

Your information will be processed in accordance with the General Data Protection Regulation (GDPR) 2016/679 of the European Union.

Figure 27: Information sheet for participants

Part:
ID:
Age:
Experience with VR use:

- first time
- less than 5 times
- 5 - 20 times
- more than 20 times.

Simulator Sickness Questionnaire

General discomfort	None	Slight	Moderate	Severe
Fatigue	None	Slight	Moderate	Severe
Headache	None	Slight	Moderate	Severe
Eye strain	None	Slight	Moderate	Severe
Difficulty focusing	None	Slight	Moderate	Severe
Increased salivation	None	Slight	Moderate	Severe
Sweating	None	Slight	Moderate	Severe
Nausea	None	Slight	Moderate	Severe
Difficulty concentrating	None	Slight	Moderate	Severe
“Fullness of the head”	None	Slight	Moderate	Severe
Blurred vision	None	Slight	Moderate	Severe
Dizzy (eyes open)	None	Slight	Moderate	Severe
Dizzy (eyes closed)	None	Slight	Moderate	Severe
Vertigo (Giddiness)	None	Slight	Moderate	Severe
Stomach awareness	None	Slight	Moderate	Severe
Burping	None	Slight	Moderate	Severe

Figure 28: Data collection and SSQ form

References

[1] Adrien Maglo, Guillaume Lavoué, Florent Dupont, and Céline Hudelot. 3d mesh compression: Survey, comparisons, and emerging trends. *ACM Computing Surveys (CSUR)*, 47(3):1–41, 2015.

[2] Alireza Javaheri, Catarina Brites, Fernando Pereira, and Joao Ascenso. Point cloud rendering after coding: Impacts on subjective and objective quality. *IEEE Transactions on Multimedia*, 23:4049–4064, 2020.

[3] Mohammad Hosseini and Christian Timmerer. Dynamic adaptive point cloud streaming. In *Proceedings of the 23rd Packet Video Workshop*, pages 25–30, 2018.

[4] Jeroen van der Hooft, Maria Torres Vega, Christian Timmerer, Ali C Begen, Filip De Turck, and Raimund Schatz. Objective and subjective qoe evaluation for adaptive point cloud streaming. In *2020 twelfth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE, 2020.

[5] Andrew Perkis, Christian Timmerer, Sabina Baraković, Jasmina Baraković Husić, Søren Bech, Sebastian Bosse, Jean Botev, Kjell Brunnström, Luis Cruz, Katrien De Moor, et al. Qualinet white paper on definitions of immersive media experience (imex). *arXiv preprint arXiv:2007.07032*, 2020.

[6] Martín Varela, Lea Skorin-Kapov, and Touradj Ebrahimi. Quality of service versus quality of experience. In *Quality of experience*, pages 85–96. Springer, 2014.

[7] Satu Jumisko-Pyykkö, Jukka Häkkinen, and Göte Nyman. Experienced quality factors: qualitative evaluation approach to audiovisual quality. In *Multimedia on Mobile Devices 2007*, volume 6507, page 65070M. International Society for Optics and Photonics, 2007.

[8] Satu Jumisko-Pyykkö, Tomi Haustola, Atanas Boev, and Atanas Gotchev. Subjective evaluation of mobile 3d video content: depth range versus compression artifacts. In *Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V*, volume 7881, pages 126–137. SPIE, 2011.

[9] Emil Dumic, Federica Battisti, Marco Carli, and Luis A da Silva Cruz. Point cloud visualization methods: A study on subjective preferences. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 595–599. IEEE, 2021.

[10] ITURBT Recommendation. 500-13: Methodology for the subjective assessment of the quality of television pictures. *Geneva: International Telecommunication Union*, 2012.

[11] Evangelos Alexiou, Irene Viola, Tomás M Borges, Tiago A Fonseca, Ricardo L De Queiroz, and Touradj Ebrahimi. A comprehensive study of the rate-distortion performance in mpeg point cloud compression. *APSIPA Transactions on Signal and Information Processing*, 8, 2019.

[12] Yan Huang, Jingliang Peng, C-C Jay Kuo, and M Gopi. A generic scheme for progressive point cloud coding. *IEEE Transactions on Visualization and Computer Graphics*, 14(2):440–453, 2008.

[13] Luis A da Silva Cruz, Emil Dumić, Evangelos Alexiou, Joao Prazeres, Rafael Duarte, Manuela Pereira, Antonio Pinheiro, and Touradj Ebrahimi. Point cloud quality evaluation: Towards a definition for test conditions. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2019.

[14] Chao Cao, Marius Preda, and Titus Zaharia. 3d point cloud compression: A survey. In *The 24th International Conference on 3D Web Technology*, pages 1–9, 2019.

[15] D Graziosi, O Nakagami, S Kuma, A Zaghetto, T Suzuki, and A Tabatabai. An overview of ongoing point cloud compression standardization activities: Video-based (v-pcc) and geometry-based (g-pcc). *APSIPA Transactions on Signal and Information Processing*, 9, 2020.

[16] Euee S Jang, Marius Preda, Khaled Mammou, Alexis M Tourapis, Jungsun Kim, Danillo B Graziosi, Sungryeul Rhyu, and Madhukar Budagavi. Video-based point-cloud-compression standard in mpeg: From evidence collection to committee draft [standards in a nutshell]. *IEEE Signal Processing Magazine*, 36(3):118–123, 2019.

[17] Sebastian Schwarz, Marius Preda, Vittorio Baroncini, Madhukar Budagavi, Pablo Cesar, Philip A Chou, Robert A Cohen, Maja Krivokuća, Sébastien Lasserre, Zhu Li, et al. Emerging mpeg standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):133–148, 2018.

[18] Ricardo L De Queiroz and Philip A Chou. Compression of 3d point clouds using a region-adaptive hierarchical transform. *IEEE Transactions on Image Processing*, 25(8):3947–3956, 2016.

[19] K Mammou, A Tourapis, J Kim, F Robinet, V Valentin, and Y Su. Proposal for improved lossy compression in tmc1. *ISO/IEC JTC1/SC29/WG11 m42640, San Diego, US*, 2018.

[20] Jae Shin Yoon, Zhixuan Yu, Jaesik Park, and Hyun Soo Park. Humbi: A large multi-view dataset of human body expressions and benchmark challenge. *arXiv preprint arXiv:2110.00119*, 2021.

[21] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J Black. Dynamic faust: Registering human bodies in motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6233–6242, 2017.

[22] Emin Zerman, Pan Gao, Cagri Ozcinar, and Aljosa Smolic. Subjective and objective quality assessment for volumetric video compression. *Electronic Imaging*, 2019 (10):323–1, 2019.

[23] Matthew Trumble, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John Collomosse. Total capture: 3d human pose estimation fusing video and inertial sensors. In *Proceedings of 28th British Machine Vision Conference*, pages 1–13. University of Surrey, 2017.

[24] Ignacio Reimat, Evangelos Alexiou, Jack Jansen, Irene Viola, Shishir Subramanyam, and Pablo Cesar. Cwipc-sxr: Point cloud dynamic human dataset for social xr. In *Proceedings of the 12th ACM Multimedia Systems Conference*, pages 300–306, 2021.

[25] Evangelos Alexiou, Nanyang Yang, and Touradj Ebrahimi. Pointxr: A toolbox for visualization and subjective evaluation of point clouds in virtual reality. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2020.

[26] JBIG JPEG. Iso/iec jtc 1/sc 29/wg 1 (itu-t sg16). 2009.

[27] Sujun Zhang, Wei Zhang, Fuzheng Yang, and Junyan Huo. A 3d haar wavelet transform for point cloud attribute compression based on local surface analysis. In *2019 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2019.

[28] Rafael Pagés, Emin Zerman, Konstantinos Amplianitis, Jan Ondřej, and Aljosa Smolic. Volograms & V-SENSE Volumetric Video Dataset. *ISO/IEC JTC1/SC29/WG07 MPEG2021/m56767*, 2021.

[29] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 303–312, 1996.

[30] Danillo Bracco Graziosi. Video-based dynamic mesh coding. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 3133–3137. IEEE, 2021.

[31] MPEG. mpeg-pcc-tmc2. URL https://github.com/MPEGGroup/mpeg-pcc-tmc2.

[32] MPEG 3DG. Doc. n18474: Common test conditions for point cloud compression. *Tech. Rep., ISO/IEC JTC1/SC 29/WG11*, 2019.

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

[34] ITU. Subjective video quality assessment methods for multimedia applications. 2008.

[35] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[36] ITU-T. Subjective test methodologies for 360 video on head-mounted displays. 2020.

[37] ITU-T. Subjective video quality assessment methods for multimedia applications. 2021.