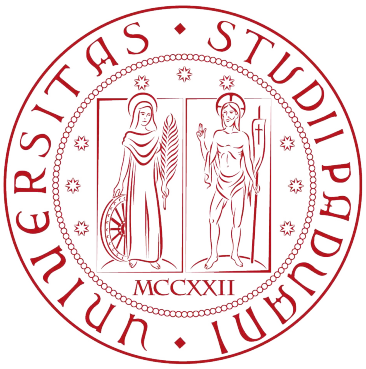

Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Master Degree in ICT for Internet and Multimedia

Impact of Point Clouds transmission and compression errors on the user experience

Supervisor:
Prof.ssa Federica Battisti

Co-Supervisor:
Prof. Jesús Gutiérrez

Candidate:
Matteo Dal Magro
2019281

Academic Year 2022-2023

Abstract

Point Clouds (PC) have gained significant attention in recent years as a means of representing visual data in immersive applications, such as virtual and augmented reality. A point cloud is a collection of points in three-dimensional space, each defined by its spatial coordinates, potentially along additional attributes such as color, opacity and surface normal.

It is crucial to evaluate the impact of compression and transmission on the Quality of Experience (QoE) of point cloud data. QoE refers to the subjective perception of users when interacting with compressed point clouds, taking into account factors such as visual fidelity, interactivity, and overall user satisfaction.

Understanding the implications of point cloud compression and transmission on the QoE is of utmost importance as it directly affects the user's perception and acceptance of compressed point cloud content. This thesis aims at evaluating the trade off between compression, transmission and QoE.

The thesis is structured as follows.

In the introduction there will be a brief explanation of point clouds technologies, of the challenges they impose and of the objectives of the thesis. The second chapter will dive deeper into the state-of-the-art and the related works in the research community, exploring available datasets, point cloud compression technologies, Quality of Experience, and subjective metrics. The third chapter focuses on the design choices taken during the development of the subjective experiment. The fourth chapter will discuss the obtained results. Lastly, in the fifth chapter the conclusion will be drawn.

Acknowledgements

I would like to express my deep gratefulness to my supervisors Federica Battisti and Jésus Gutiérrez, first of all for giving me the opportunity to develop my thesis abroad, then for the guidance, support and friendliness they always gave me. Besides my supervisor I would like to thank the whole GTI group for being so welcoming, nice and for making me feel part of the group since the first day. In particular I would like to thank Carlos for his fundamental help in the development of the thesis and Pedro and Adriana for always being there to have a break and to teach me spanish.

I am deeply indebted to all the friends I made while in Madrid, you made this last 9 months the happiest period in my life, it wouldn't have been even remotely as good without you, if I fell in love with Madrid it's because of you. I would like to thank my flatmates, especially Inès, Val and Mathias, you helped in creating a special family type relationship in our home, thank you for always being there to talk until 4 in the morning. I extend my thanks to Cecilia and Marzia, you were the first two persons I have met in Madrid and I couldn't have been luckier, thank you for sharing this whole experience with me, you were fundamental. I would like to thank Lorenzo, I am so glad to have had a friend like you in Madrid, I hope we will share more experiences together. I then need to thank Sofia, Lidia and Jacqueline, hanging out with you is always so easy and so natural, I hope one day you will be part of my daily life again. Thanks Lidia for being so honest and direct, thanks Jacqueline for always having an interesting view on the most random topics. Lastly, thank you Sofia for being your weird self, I still cannot believe how we clicked so easily and naturally, I will always hold you close to my heart. I can't forget to thank Nathan, Louis, Lisa, Sarah and all the feel free to enjoy group, you were all part of this journey and you made it perfect. I will never find such a messy group of random people again, everyone is so different yet so similar, I will be forever grateful to all the moments we spent together.

Words cannot express my gratitude to Vale, you are the best sister I could have ever asked for, knowing that you are always there if I need it makes me feel safe, thank you for your wise

words, I have a lot to learn from you. I then need to thank my childhood friends, we know each other since so long and you saw me grow and change during the years, yet we still are sharing moments and memories, thanks for being there when i need it the most.

Lastly, I must thank my family, in particular my parents. You were always there to support and to help me in my choices, even when you didn't agree with them. Even though we are growing to be different persons you still manage to teach me so much. If I am here now it is thanks to you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In recent years, the exponential growth of multimedia usage has requested the development of efficient compression and transmission algorithm to accommodate the increased demand of data intensive multimedia.

Point clouds, which represent three-dimensional (3D) geometric data, have gained significant attention due to their wide range of applications, as a matter of fact they can be found in fields such as autonomous driving, cultural heritage, topography, virtual and augmented reality. Point clouds are volumetric visual data that represent 3D scenes and objects. As the name suggests, they are composed by set of points in the 3D space. Each point is characterized by three geometric coordinates (x,y,z), defining its position in the space, as well as optional additional attributes such as color, reflectance, or normal. Point clouds can be generated from computer-based 3D models or captured from real-world environments using multiple cameras or specialized devices like LIDARs. When there is a sequence of frames we talk about dynamic point clouds or volumetric videos. This form of multimedia plays a significant role in Augmented Reality and Virtual Reality technologies, allowing users to have Six Degrees of Freedom (6DoF) viewing and immersive experiences fundamental to applications such as tele-presence. The coordinates of a point can be of any value, but they are often quantized to fit a certain precision range, this process is called voxelization, as all the points inside a voxel (a cube of the three-dimensional grid) will be mapped to the center of the voxel. A single point cloud can easily reach millions or billions of points, taking up a huge volume of data and posing significant challenges in the storage, transmission and processing aspects. As an example if left uncompressed, a typical dynamic point cloud of 1 million points per frame and at 30fps would require a bandwidth of 3.6 Gbps [4].

To address these challenges, in the last years, compression techniques for point clouds

have been researched and developed. Evaluating the effectiveness and the efficiency of these compression methods is crucial to ensure that the compressed and transmitted point clouds maintain an acceptable level of quality and fidelity.

While objective quality metrics have traditionally been employed to measure the performance of compression algorithms, they often fail to fully capture the subjective perception of quality experienced by human observers. Moreover, when dealing with recent technologies, such as point clouds and virtual reality headsets, it is fundamental to study subjective evaluations and behaviours, both to understand how the user experiences the multimedia and to have ground truths scores to help develop accurate objective metrics.

Secondly, subjective quality assessment plays a vital role in evaluating the visual perception and user experience of compressed and transmitted point clouds. Understanding how users perceive the quality of compressed point clouds is essential for optimizing compression algorithms and designing systems that cater to user expectations.

1.1 Objectives

This master thesis aims to investigate the subjective Quality of Experience (QoE) of compressed and transmitted point clouds. The study will analyze the results obtained from a subjective test carried out at the Universidad Politécnica de Madrid. By conducting subjective quality assessment experiment, the thesis aims to provide valuable insights into the impact of compression techniques and transmission on the perceived quality of point clouds, however it will also take a brief look into the correlation of different sickness questionnaires and into the movement behaviours of the participants.

The objectives of this thesis include:

- Studying the current state of the art and the recent experiments conducted
- Coding and simulating transmission on one of the available datasets
- Designing and conducting a subjective quality assessment experiment to evaluate the perceived quality of compressed and transmitted point clouds
- Analyzing the obtained subjective ratings to identify factors influencing the subjective quality of experience and movement behaviours
- Analyzing the results from three different sickness questionnaires to evaluate their correlation

By addressing these objectives, this thesis aims to contribute to the advancement of point cloud compression and transmission techniques, with a particular focus on enhancing the subjective quality of experience for end-users. The findings of this research will provide valuable insights for researchers and developers involved in the design and implementation of efficient and visually pleasing point cloud systems.

Chapter 2

Related Works

In this chapter we are going to explore the available datasets, the key concepts and the state-of-the-art on the compression algorithms and on the subjective evaluation of quality of point clouds.

2.1 Point cloud datasets

Point clouds can differ a lot between themselves. They can range from being extremely detailed heritage models to huge topographical scans, from dynamically taken, scarcely populated, scans for autonomous vehicles to high quality dynamic point clouds grabbed in a controlled environment. They can also differ in the technology used to capture them. There are direct methodologies, i.e. using technology explicitly developed to output 3D data, such as time-of-flight cameras or LIDAR scanner. There are also indirect methodologies, in which algorithm can be used in the post processing to extract 3D information from data that does not represent 3D data, for example fusing the data from various 2D cameras to create a 3D model. Many point cloud datasets were developed by the academic community to push forward the research process. Due to wide variety of use cases and technologies, datasets can vary a lot in the type of content, the capturing setting, the post processing of the data, the quality and the size of the objects [1], [29], [22].

Example of static point clouds can be found in the PointXR dataset [1] provided by the EPFL that consists of 20 high-quality cultural heritage models, some examples can be seen in Fig.2.1.

Not many datasets focus on dynamic point clouds, as they usually require a more complex set up, with up to hundreds of synced cameras in a controlled environment, often a green room. In Fig.2.2 is shown an example of a basic camera arrangement, while in Fig.2.3 the

(a) Guanyin (2286975) (b) Muse (2092701) (c) Roy (2272748) (d) Shield (2380435) (e) Tiki (1540246)

Figure 2.1: Examples of static point clouds from the pointXR dataset [1]

actual setting used by the 8i labs to capture holograms and point clouds.

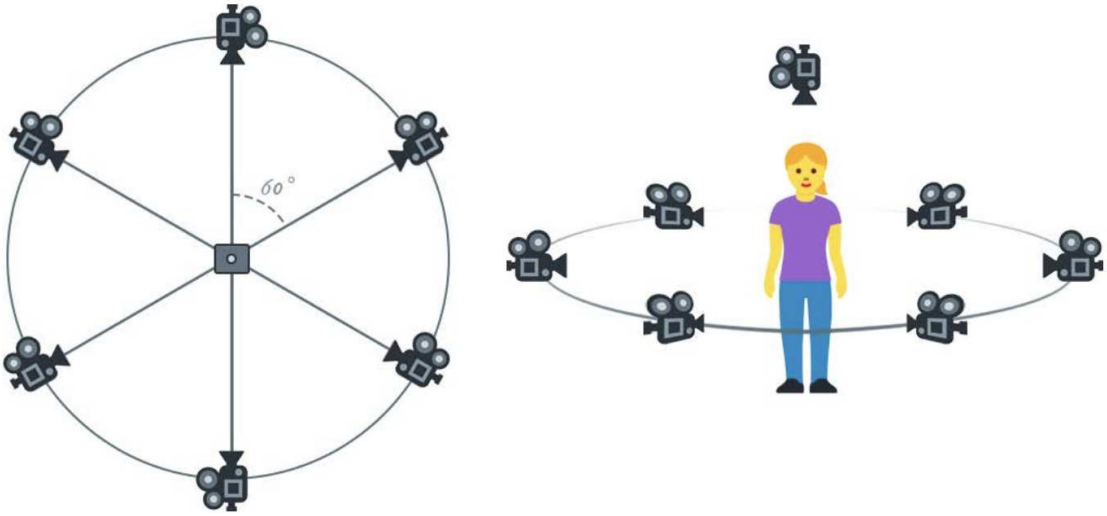

Figure 2.2: Example of a camera arrangement used for the CWI dataset [22]

An important dataset is the "V-SENSE Volumetric Video Quality Database 2" [29], captured by the V-SENSE team at the Trinity College. It provides four point clouds of moving human figures, one of which is interacting with an object (i.e. a football). The point clouds were crafted out of the raw data using the method of Pagés et al. [20], however they do not provide information about the capture system. Overall the dataset presented some flaws, the point clouds had some noticeable artifacts and did not have a high number of points, making them more blurry and less realistic.

Another recent dataset is the CWI Point Cloud Social XR Dataset [22], in which dynamic point clouds were captured using seven synchronized Azure Kinect DK devices, as shown in Fig. 2.2, the dataset is composed of 45 sequences. The sequences focus on individuals engaged in common social activities within real-time communication cases. The dataset counted four distinct social experience scenarios, namely "Education and Training," "Healthcare," "Commu-

Figure 2.3: Real set up used in the 8i labs

nication and Social Interactions," and "Performance and Sports" (Fig.2.4).

Figure 2.4: Examples of sequences from the CWI dataset [22]

This datasets also provided synchronized audio tracks to play with the volumetric videos. It is to be noted that the low number of cameras used to create this dataset influenced the quality of the output point clouds, which show a huge number of holes and artifacts. This flaws make the dataset ineffective for the quality of experience experiment that was the objective of this project, however they show more realistic point clouds for real time communications, thus making it a really important dataset for the research community.

The most famous and used dataset in the research community is the JPEG Pleno dataset "8i Voxelized Full Bodies (8iVFB v2) - A Dynamic Voxelized Point Cloud Dataset" [3] in which four sequences of full human bodies were grabbed using a set of 42 RGB cameras, arranged in 14 clusters, at 30fps for a total of 10s each. The output data has a space resolution of 1024x1024x1024 voxels, which results in 10 bits in the depth channel. Since the subject height

is typically the larger dimension, for a 1.8m tall subject, a voxel at depth 10 would be about 1.8m/1024voxel $\approx$ 1.75mm per voxel on one side.

A similar dataset, provided by the MPEG, is the 8i Voxelized Surface Light Field (8iVSLF) Dataset [15] which provides a one frame version of the sequences available in 8iVFB v2 plus a high quality voxelized volumetric video. This last sequence was captured with a set of 39 RGB cameras organized in 13 clusters, each one acting as a logical RGBD camera, at a 30 fps rate for a total of 10s. After merging the data a 3D representation with 12 bits on the depth channel is obtained, for a space resolution of 4096x4096x4096 voxels. In the volumetric video each voxel represents approximately 1x1x1 mm of the physical capture space, and since the human figure takes up less than half of the height of the voxelized space, making its height under 2 m tall.

2.2 Point Cloud Compression

One of the primary challenges associated with point clouds is to deal with the huge storage requirements and the high transmission costs they impose. This is due to the fact that, to represent an object with an high quality point cloud, a large number of points is required and to each point a multitude of additional attributes, such as color, surface normal, and reflectance can or have to be attached. The intrinsic unstructured characteristics of point clouds makes it hard to develop compression algorithms. For example, raw point clouds are not located on a grid, and points can lay anywhere in the space, making it complex to adapt available 2D compression algorithms or to develop efficient new ones. Moreover, point clouds differ from other multimedia formats, like images or videos, in which there is a fixed amount of pixels always in the same position and each one with an associated value. In fact, in a volumetric video, frames usually do not have a fixed number of points, meaning some of them may appear or disappear during time, and even if voxelized not all voxels will contain information.

To address these challenges, the Moving Picture Experts Group (MPEG) published a call for proposals (CfP) in 2017. Due to the wide range of applications, the MPEG Point Cloud Compression (PCC) standardization activity opted to have three categories of test data:

- Static point clouds: Millions to billions of points, high quality, colors, optional additional attributes

- Dynamic point clouds: Less points, color, temporal information

- Dynamically acquired point clouds: Millions to billions of points, colors, optional additional attributes, sequences of static point clouds grabbed dynamically

In the last years, several approaches have been proposed, and three technologies were chosen as test models for the three categories:

- Surface point cloud compression for (S-PCC) for Static point clouds
- Video-based point cloud compression (V-PCC) for Dynamic point clouds
- LIDAR point cloud compression (L-PCC) for Dynamically acquired point clouds

Due to the similarity between S-PCC and L-PCC the two have later been merged in a Geometry-based point cloud compression (G-PCC) standard, equivalent to the combination of L-PCC and S-PCC.

Geometry-based techniques, are appropriate for sparse distributions, they revolve around the coding of the spatial position, G-PCC is derived from these methods.

Video-based, equivalent to V-PCC, appropriate for point sets with a relatively uniform distribution of points. The idea is to make projection of each frame of the 3D model to a 2D space, the projections are then merged into images, one per each frame, the pivotal point is to then take advantage of existing video coding standards to code the sequence of images.

2.2.1 V-PCC standard

As mentioned above, V-PCC revolves around exploiting existing video codecs to achieve an efficient compression of volumetric videos information. To achieve this, the point cloud sequence needs to be processed to become a 2D video, it will then be possible to apply the standard video codecs available. Projecting a 3D model to a 2D space creates loss of information, for this reason the algorithm actually creates two video sequences, one containing the geometry information and a second one containing the texture information. Moreover additional metadata, specifically an occupancy map and auxiliary patch information, are computed in order for the decoder to have a way to interpret the two video sequences. The additional metadata is compressed separately and then multiplexed with the compressed videos bitstreams. It should be noted that the metadata information takes up 5 to 20 % of the bitstream [24], meaning most of the information resides in the two video sequences, and is already efficiently compressed with known techniques.

In Fig. 2.5 the encoding pipeline followed by the V-PCC standard.

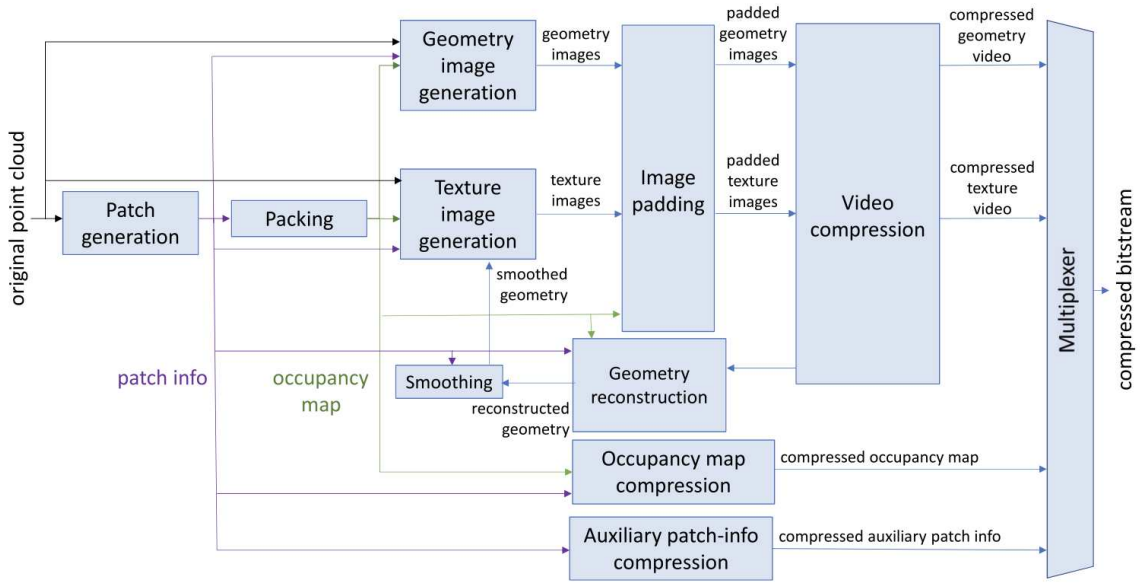

Figure 2.5: V-PCC coding scheme [24]

2.2.1.1 Patch generation and packing

To allow the video codecs to be as efficient as possible it is necessary to generate video sequences with a strong spatio-temporal correlation. For this reason one of the main objectives of the V-PCC compression process is to find a temporally coherent, low-distortion, injective mapping, which assigns each point of the 3D point cloud to a cell of the 2D grid [12].

A naive approach, such as projecting the 3D model to a cube, or a sphere, would not guarantee lossless reconstruction, as it does not take into account the frequent auto-occlusion of points and hidden surface problems typical of point clouds, leading to considerable distortions [4]. To solve this problems, V-PCC segments the point cloud in many 3D patches, which are just connected regions. Each patch is then independently projected to the 2D plane. Since a point cloud may have multiple points mapped to the same 2D pixel, it allows to have multiple layer maps, each one associated with a certain depth range. An example of 3D patch and its projections can be seen in Fig. 2.6.

Each 2D patch needs to be packed in a 2D image. For the first frame the process is simple, a blank 2D image of size WidthxHeight pixels is created, then all the 2D patches are ranked in size, and, starting from the biggest a simple scan search algorithm looks for the first available location that guarantees an overlap-free insertion. To optimize the chances of finding good locations, four orientations combined with mirroring are tried. The occupancy map is then filled in by taking into account pixels that contain valid depth values. If there is no available location, the 2D image is doubled in Height, and the process goes on. Once all the patches are placed, the image is trimmed down to have the minimal Height possible.

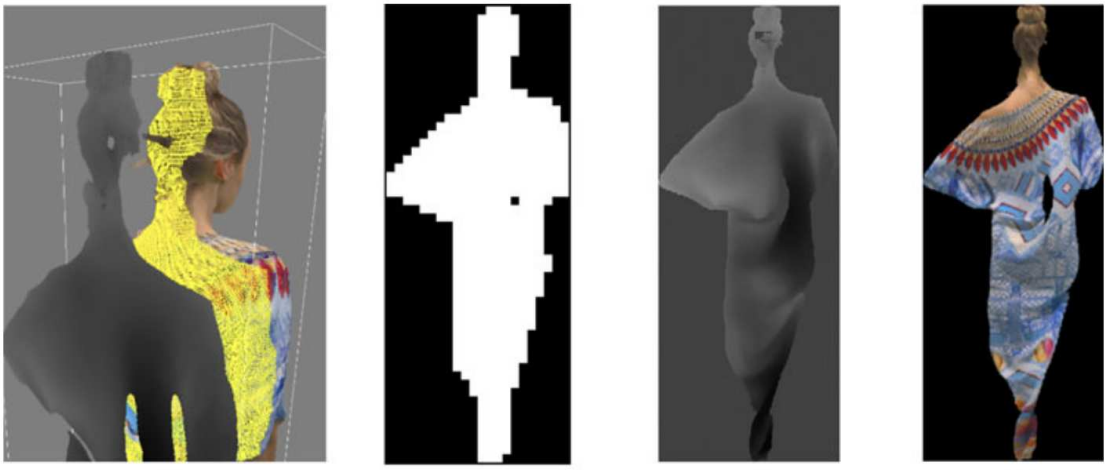

Figure 2.6: Example of patch projection: (a) 3D patch, (b) 3D Patch Occupancy Map, (c) 3D Patch Geometry Image, (d) 3D Patch Texture Image [4]

As said above, it is fundamental to have spatio-temporal correlation between frames, to achieve this the algorithm searches for matches between 2D patches in different frames, when a match is found V-PCC tries to insert it in a similar location in the image as the previous one.

2.2.1.2 Geometry and occupancy map

The geometry map uses only the luminance channel to code the depth of the points. Due to the nature of the 2D patches, that makes their shape arbitrary, it is necessary to have a second map, called occupancy map. The occupancy map is a binary image in which a pixel with a value of 1 represents a pixel in the geometry map that carries depth data, on the other hand pixels with a value of 0 are associated with locations in the 2D image that are not used to carry depth data. In Fig.2.7 an example of packed patches.

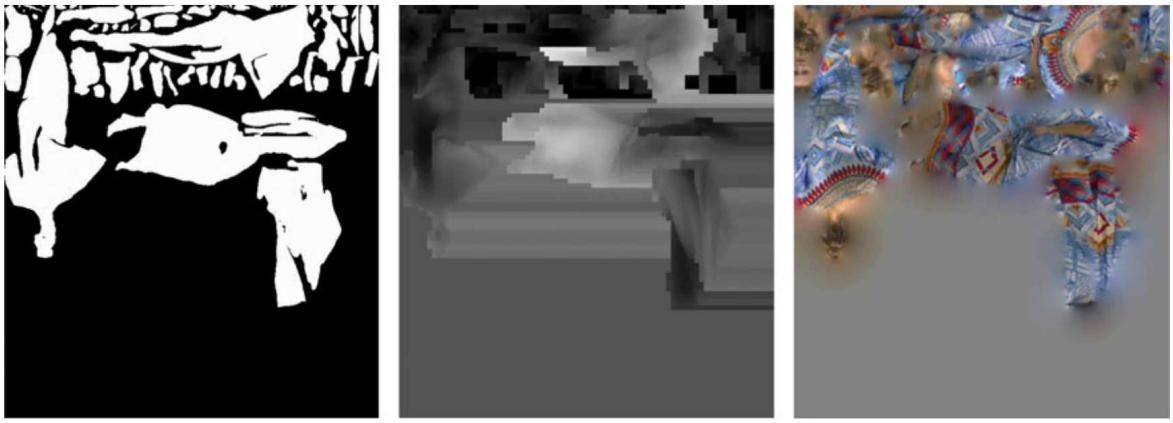

Figure 2.7: Example of packed patches, respectively: Occupancy map, Geometry map and Texture map [4]

2.2.1.3 Other steps

The V-PCC algorithm involves many other steps to improve the efficiency of the compression, the quality of the output and to allow the decoding of the data.

A couple of steps worth mentioning are:

- Image padding: In which the empty pixels of the texture and the geometry map are assigned a value to make the transition between patches smoother and improve the efficiency of the video codecs
- Geometry and color smoothing: A process that may be applied to reduce artifacts that may arise at the borders of patches when reconstructing them as a 3D point cloud
- Atlas metadata: The decoder needs some additional information, called atlas metadata, such as the position and rotation of patches or other information that has to be sent to the decoder

2.2.1.4 Decoder

The decoding process is split in two phases. In the first one it takes care of decoding the bitstream, recovering the occupancy and geometry map, the attribute 2D video frames and the patch information associated with very frame. Once all the data is available the algorithm starts with the geometry and attribute reconstruction, in which it re-projects the patches to their original place in the 3D space creating a point per each occupied pixel in the occupancy map. The video encoding and the processing applied at the encoder will introduce artifacts and discontinuities in the reconstructed point cloud, to solve this, usually a smoothing process is applied in order to alleviate the severity of such artifacts [12].

In Fig. 2.8 the decoding pipeline followed by the V-PCC standard is shown.

2.3 Transmission

As mentioned above, due to the unique characteristics they have and the possible applications they allow, point clouds gained a lot of attention in the last years. Some of the promising use cases involving point clouds, and many of the foreseen applications exploiting volumetric videos, require real-time transmission over unreliable networks. The typical example is tele-presence applications, point clouds would allow 6 Degrees of Freedom (6DoF) turning a normal video call into a truly immersive experience, where the user is free to explore and

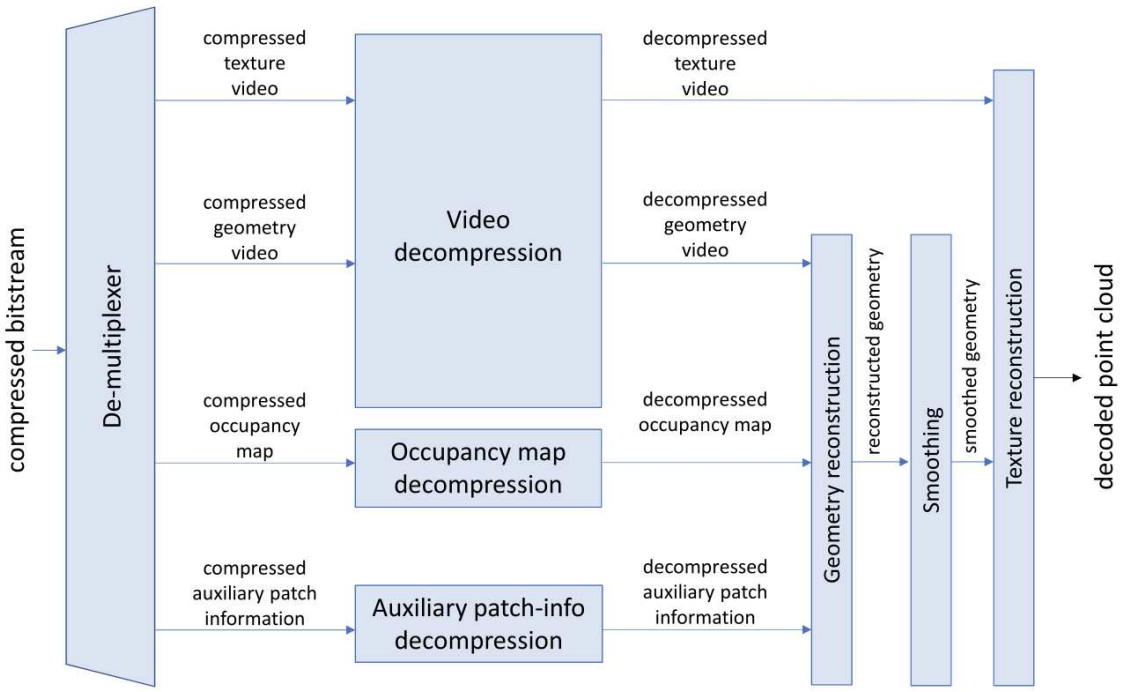

Figure 2.8: V-PCC decoding scheme [24]

move like he/she would in the real world. However, transmitting and delivering point clouds over networks poses big challenges, as they would need ultra high bandwith transmission, that nowadays are not available [25], even more challenges arise when considering the typical unreliable networks, hindering real-time interactions and applications that rely on accurate and correctly timed point clouds.

While the first challenge, i.e. dealing with the huge bandwith requirements of point clouds, is being tackled by studying efficient compression techniques, not many studies focused on how to achieve an efficient and reliable transmission. Moreover, not many studies researched the effects of unreliable networks on the quality of experience.

A notable advancement, regarding solution for point cloud transmission over networks, was developed by Hosseini et al. [7] in which a Dynamic Adaptive HTTP Streaming (DASH) is merged together with an algorithm that spatially sub-samples dynamic point clouds in order to create representations at multiple quality levels. Other works then extended the DASH, in [26] it was proposed to set the quality of the point cloud objects based on a rate adaptation heuristics that considers location, focus and available bandwidth. In [17], Li et al. aimed at maximizing the QoE by choosing the best possible quality levels, considering computational and communication constraints, for each partitioned point cloud video tile.

The studies above provided important insights and advancements, however, all of them, built up from the DASH, which is an algorithm that by nature uses reliable network transport

protocols. This clashes with the whole concept of real time applications, as a reliable network values the delivery of the data more than the low latency. In typical real time applications, such as video calls or online gaming, only unreliable transport protocols are used (e.g. UDP) as they guarantee the lowest latency possible.

Only a few studies focused on the effects of the streaming of point clouds over error prone networks. In [28] the Network Abstraction Layer Units (NALUs) of the V-PCC bitstream are located and, to simulate packet losses, some of them are changed to sequences of zeros. A technique to minimize the losses retrieving as much information as possible is also proposed in the same paper.

It is thus fundamental to expand the research in this area, to allow a better understanding on the QoE thresholds and on possible solutions that will help in making this multimedia widely used.

2.4 Quality of Experience

In the previous sections, the concept of Quality of Experience (QoE) was used more than once, and often the discussion revolved around the effects on it, however the concept was never described.

In [11], QoE is defined as "The degree of delight or annoyance of the user of an application or service".

It is worth mentioning that, as it is said in the same recommendation document, since there is on-going research on this topic, the QoE definition is expected to evolve over time.

QoE influencing factors include the type and characteristics of the application or service, context of use, the user's expectations with respect to the application or service and their fulfilment, the user's cultural background, socioeconomic issues, psychological profiles, emotional state of the user, and other factors whose number will likely expand with further research [11].

There exist two methods to evaluate QoE, the first one is through subjective metrics, the second one through objective ones.

Subjective metrics rely on human perception and involve carrying out subjective assessment methods, while these methods provide ground truth results they are time and resource consuming.

Objective metrics on the other hand are algorithms that aim to evaluate the level of degradation in the QoE of users, they are fast, efficient, their outcome can be reproduced and they

give out numerical result that can be easily compared. However, it may happen that their results will not provide valuable information regarding the end user perception, for example their outcomes may not correlate well with the ones of subjective experiments.

The two QoE evaluation methods should be used together to provide a deeper understanding on how humans perceive multimedia. The study of QoE is fundamental as it provides useful insights on users perception, not only that, it may lead to the discovery of behavioral patterns that other metrics cannot predict, helping in the development of more accurate objective metrics. On the other hand accurate objective metrics make it faster and more efficient to study and evaluate effects of compression, transmission or any other process that involves the creation of noise or artifacts.

Regarding 2D images and videos, while subjective evaluations are still the most accurate quality assessment tool, years of research provided powerful objective metrics that are now widely used.

When dealing with point clouds, and even more with volumetric videos, the research process just started and there are still no defined or de-facto standards.

2.4.1 Objective metrics

Objective metrics are algorithms or models that measure specific perceptual or technical aspects automatically. They aim to evaluate certain characteristics in an objective way, without relying on humans.

While the metrics already available for 2D multimedia do not fully grasp the complexity of 3D models, in the research some extensions have been proposed, for example in [16] they proposed an adaptation of PSNR, a common measure used for 2D images and videos.

Other metrics have been proposed, such as point-to-point or point-to-plane [8], which measure a point-wise distortion to obtain a quality measure. However in [18] concluded that the objective metrics available at that time were not able to measure the quality constistently with what humans perceived. They also noted that, using V-PCC as a compression algorithm, very different results in terms of QoE could be obtained depending on the dataset used, for example the 8iVFB has been voxelized, and thus shows a grid like disposition of points, while other datasets have a raw disposition of points which is more prone to give artifacts after the V-PCC process.

2.4.2 Subjective tests

Subjective metrics are human based, they involve gathering feedback from users through subjective assessment methods such as questionnaires, and user ratings. These metrics try to capture the users' opinions, perception and preferences.

Static point cloud quality has been researched in many experiments, however the number of studies on the quality of dynamic point clouds is still limited, and if considering studies conducted in immersive environments (e.g. using virtual reality) only a few are available.

In a study over quality of compressed volumetric videos under various conditions it was concluded that double stimulus methodologies can lead to an evaluation of a difference of quality, instead of an evaluation of the quality of the test sequence, thus a single stimulus method can give more insights [27].

A typical test methodology to evaluate dynamic point clouds would be to have a single stimulus and to use a absolute category rating for the evaluation.

Absolute category rating is a way of evaluating quality in which participants have a 5 scores scale, defined as: Excellent, Good, Fair, Poor and Bad.

Following the recommendations in [9], an experiment over 3D data should be conducted under this conditions: Test sequences have to be in a mid-gray field to avoid distractions or effects on the shown media, there should be a mid-grey field before and after the test sequence. The gray screen before the sequence should last at maximum 3 seconds and can contain information, the voting screen should last a maximum of 10 seconds. The duration of the test sequence should be at around 10 seconds.

Chapter 3

Subjective experiment

In this chapter the experiment is described and the choices made during the test development are described and justified.

3.1 Dataset

As test sequences it was decided to use the 8iVFB dataset [3], which, as described in the previous chapters, provides 4 voxelized dynamic point clouds. The sequences are called Loot (or S23), Redandblack (or S24), Soldier (or S25) and Longdress (or S26) and can be seen in Figs. 3.1, 3.2, 3.3 and 3.4.

These point clouds only have geometrical and color attributes, they are 300 frames long at 30fps, making them 10s long.

Table 3.1 reports a summary of their characteristics.

Sequence	Frames	fps	Avg # points	Format	Attributes
Loot (S23)	300	30	700K	ply	RGB
Redandblack (S24)	300	30	700K	ply	RGB
Soldier (S25)	300	30	1.2M	ply	RGB
Longdress (S26)	300	30	800K	ply	RGB

Table 3.1: 8iVFB dataset basic informations

There are three main reasons this dataset was chosen between the few available ones. First of all it is the highest quality and more consistent dataset, the point clouds do not show holes and only have few artifacts, this is consistent between all four sequences. Secondly, it is the most used dataset in research, while it is necessary to expand the research to different sequences in order to avoid biases due to the shown figure, since there are not many other available datasets, it is interesting to develop a thorough investigation over the same dataset

Figure 3.1: Loot (S23)

Figure 3.2: Redandblack (S24)

trying to connect the dots with the findings of previous studies. Then, as one of the main use cases of volumetric videos could be tele-presence or social applications, it is fundamental to study point clouds of human figures as they will be the main focus in these applications.

In [8], it is recommended to show training sequences to the participants, in order to make them comfortable with the test process and to give them references of what can be best or worst quality, this is especially necessary when dealing with multimedia that are not yet common in the life of people, such as volumetric videos. To avoid biases, the training sequences should be different from the testing ones.

For this reasons, the MPEG dataset 8iVSLF [15] was chosen for the training process. The main reason that made it the right choice for this experiment is that it provided one high-resolution voxelized dynamic point cloud, called Thaidancer, which is shown in Fig. 3.5. Secondly, as discussed in the previous chapter, this point cloud has been grabbed by the same lab as the ones of the test using a similar set up, making it the perfect training.

It is worth noting that this volumetric video had too many points and attributes, making

Figure 3.3: Soldier (S25)

Figure 3.4: Longdress (S26)

it too detailed with respect to the test sequences and also not practical for the experiment. To solve this, all the unnecessary attributes, such as the point normal vector and the RGBs of the points as seen by each camera rig, were removed. Then a down-sampling process was applied to reach an average of 800k points per frame, this was achieved using PyMeshLab, a python library that implements methods from the open source 3D processing program MeshLab.

3.1.1 Compression

For this experiment it was chosen to use the V-PCC compression algorithm. Both the training sequence and the testing sequence were compressed. The implementation provided by MPEG through GitHub [19] was used as software to compress and decompress point clouds, more specifically the last available release, 18.0, was used.

Out of the standard rates of compression, only R1, R2 and R5 were chosen, their characteristics are shown in Tab. 3.2.

Figure 3.5: Thaidancer

R1 is the highest compression factor which yields low quality decompressed point clouds. R2 is still a quite high compression factor, but it manages to preserve more details after decompression. Lastly, R5 is the lowest compression factor that gives out the most detailed point clouds.

Only 3 rates out of 5 were selected, because, having more would have meant longer test sessions, which in turn would have meant less participants and more distress on them. The reason behind the choice of the rates stems from the results of a preliminary study, conducted last year [5]. Following the conclusions of that study, the rates were chosen to have an uniform distribution of scores. In fact, in [5], the results shown that out of the 5 standard rates, statistically, some had the same Mean Opinion Score (MOS) as others, thus making them less interesting to study. In different words, the objective was to have rates that provided statistically different MOS.

Rate	Geometry QP	Texture QP	Occupancy Map Precision
R1	32	42	4
R2	28	37	4
R5	16	22	2

Table 3.2: V-PCC Rate settings for the test stimuli

3.1.2 Transmission

As mentioned in the previous chapter, not many studies have yet explored the effects of transmission of point clouds over lossy networks.

One of the objectives of this thesis is to research on the effects of losses in the V-PCC bitstream. The work done in [2] was used as a starting point for the research and from there it was expanded to consider transmission errors.

To simulate errors, Matlab was used. In brief, the adopted scripts, read the V-PCC compressed point cloud, which corresponds with the bitstream, they parse it in order to read the NALUs, then, depending on the chosen loss ratio, a number of bits are replace in the stream in order to create losses. The V-PCC decoder is fragile with regards to losses, the bitstream carries some fundamental information, that, if lost, does not let the algorithm work, bringing to an abrupt stop the process. That is why the method used to simulate transmission errors attacks only specific parts of the bitstream. Taking this measure however does not always ensure a correct decodification of the corrupted bitstreams.

The provided scripts have been slightly modified to loop the transmission simulation until all the corrupted bitstreams could be decoded.

The simulation of the transmission was done with the following loss rates:

- L1: 0.01%
- L2: 0.02%
- L5: 0.05%
- L10: 0.1%
- L20: 0.2%
- L50: 0.5%.

The simulation process ran for 4 days on 20 point clouds (the 4 selected point clouds compressed on all 5 standard compression rates) thus leading to 120 transmissions simulated.

Following the same reasoning done for the compression rates, due to constraints on the experiments length, only three loss rates were chosen per each point cloud. Due to the random nature of the algorithm used for the transmission, degradation can vary between different simulations with the same loss rate. Considering the above factors, the rates were chosen after a preliminary evaluation of the results, the objective was to have three strongly different degradation levels.

The loss rates assigned are shown in Tab. 3.3.

Sequence	Low Loss Rate	Mid Loss Rate	High Loss Rate
S23	L1	L5	L50
S24	L10	L20	L50
S25	L5	L10	L50
S26	L5	L10	L50

Table 3.3: Loss Rates used per each point cloud sequence

From now on specific sequences will be referred to with an alphanumeric string, the first three symbols will identify the figure being seen, as explained in the previous section, for example ”S23” is the loot sequence. They will be followed by ”C2AI” and then the compression rate used. Lastly there will be an underscore followed by the loss rate.

As an example, the loot sequence, compressed with rate R5 and with loss L50 will be called ”S23C2AIR5_L50”, a frame of this point cloud can is shown in Fig. 3.6.

3.2 Eye tracking

One of the objectives of the experiment was to analyze where users focused when looking at the point clouds. To get a rough idea about this, [5] and [23], looked at the angle and position of the HMD, however this gives insights on what the user was not looking, while it can only give a loose estimate, often based on assumptions, of what the user was focusing on.

To solve this, eye tracking functions in the HMD were enabled in order to collect and save precise data of the users gaze. This data, merged with the position and rotation of the HMD, allows to completely reproduce the behaviour of a user and to reconstruct what was being stared at any given point in time. This could provide huge benefits to the research, allowing to tailor algorithms to adapt to typical human behaviours.

While this data has been collected, its analysis is out of the scope of this thesis and will not therefore be covered here.

Figure 3.6: S23C2AIR5_L50

3.3 Equipment and Environment

In this section the equipment, the software and the virtual environment used in the experiment will be briefly described.

3.3.1 Virtual environment

The experiment had to be conducted in a virtual reality environment, allowing 6DoF to the participants. A room roughly the size of 3x4 m has been used to create the conditions for freedom of movement of the users.

An application was developed to show volumetric videos in a mid-grey [128,128,128] empty environment. Point clouds were rescaled to be displayed at human size (approximately between 1.60 and 1.80m of height) and they were placed at location (X=-0.3; Y=0.1; Z=1.25) in order to be in a central place with respect to the real room. Point clouds were rendered so that, at a reasonable distance, no discontinuities were visible. In Fig. 3.7 it is shown an example of what an observer could see when visualizing the sequence S23C2AIR5 during the experiment.

Figure 3.7: S23C2AIR5

The evaluation interface used to score the point clouds is shown in Fig. 3.8.

3.3.2 Head Mounted Display

The HTC vive pro eye 2 HMD was used to conduct the experiment. The choice to use this specific headset was mainly due to its eye tracking capabilities. Secondly it is relatively widely used in the VR community, making it a realistic choice for a real user. Finally, a specific external module can be attached to this headset to make it work in wireless mode.

The headset features a resolution of 2448 × 2448 pixels per eye (4896 x 2448 pixels combined), FOV of 120 °, 90 Hz refresh rate and eye tracking capabilities at 120Hz.

To track the headset and the controllers the base stations need to be placed in the room, to ensure a correct and precise tracking a total of three base stations were used.

At first the idea was to use the visor in wireless mode, however, after some tests, the connection proved to be unstable, therefore the experiment was conducted with a cable connection.

3.3.3 Unity

To experiment has been designed in Unity, a game-engine that can also be used to build other interactive content ans that is frequently used in research.

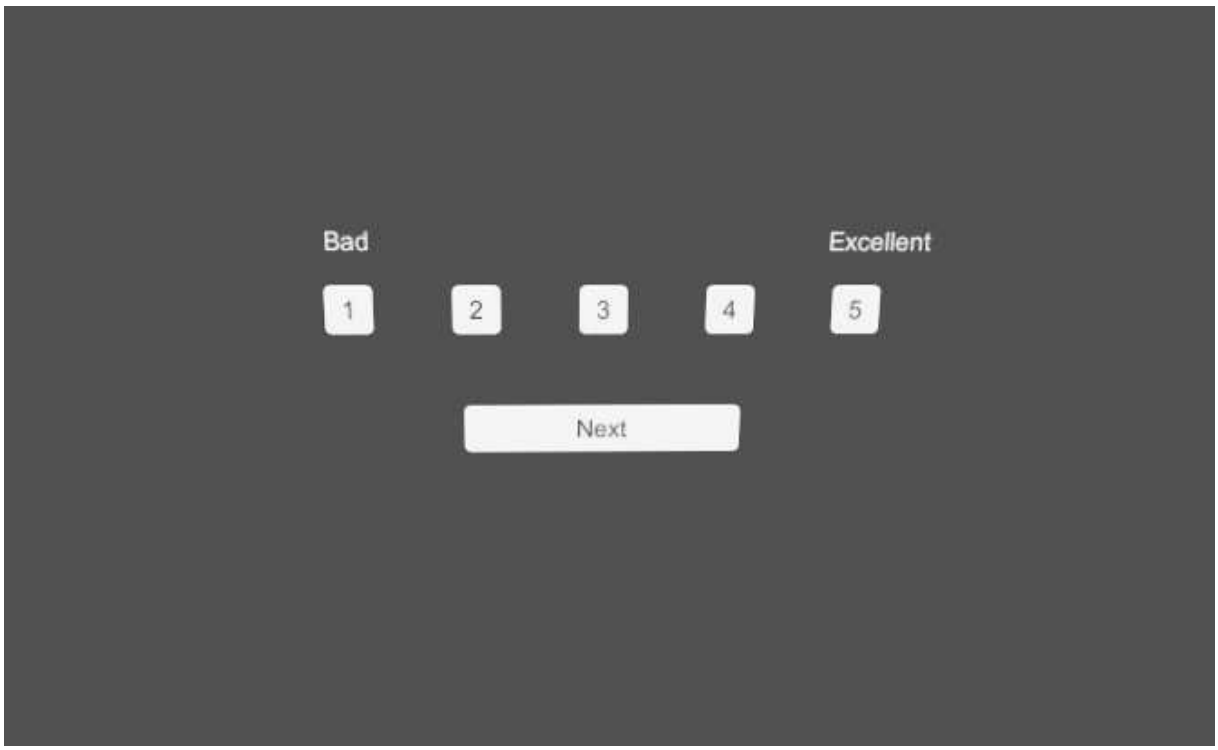

Figure 3.8: Voting interface

The developed project is an extension of the project used by [2]. The original project features a similar structure, it was however tailored to use with another HMD, it did not feature eye tracking capabilities it used Unity 2017 and it was not efficient at all.

The current project has gone through various major changes that posed multiple challenges.

First of all, the original project was developed with an old version of Unity, to use its full potential and allow the use of new standards and technologies the project was moved to a newer version of Unity, specifically the 2020 LTS. Doing this caused many conflicts that had to be resolved one by one.

The whole project was then updated to use OpenXR. Before OpenXR to target different visors it was necessary to develop many versions of the project, even with different code. The use of OpenXR allows to use only one implementation with a wide range of devices that will work in a plug and play fashion.

Another major change was the development of the eye tracking scripts that enabled its deployment in real time. While the VR community is still quite small, the eye tracking community can be properly called a niche, it was in fact hard to encounter information and descriptions that would explain how to make it work. This process took longer than expected, not because of it being hard, but because of the confused and often conflicting documenta-

Figure 3.9: Vive pro eye 2 headset with controllers and base stations

tion that can be found online. Nevertheless, the eye tracking functionalities were correctly implemented and proved to work really well in real time.

Additionally, the project underwent some changes to improve its efficiency and adaptability. The first change that was done was the implementation of coroutines, these made it possible to start the loading of frames during dead times (e.g. when the user had to vote the quality of the previous volumetric video). In this way most of the frames were loaded before it was needed to show the point cloud, the remaining ones were loaded at the same time as the observers were watching the point cloud, all of this without blocking or causing any lag.

The project was then cleaned with the creation of standard scenes that would be called back with specific scripts, making it easy to change and adapt the experiment to new advancements. This resulted really helpful because, during the months, the experiment process was adapted and changed many times.

Lastly the code was revised and optimized, removing unnecessary lines, moving all hard-coded paths to a single file and extending most of the functions and snippets of code to be able to adapt to reasonable changes in the project.

3.4 Methodology

The experiment followed the recommendations in [8], [11] and [9]. In the next subsections a description of the experiment and the methodologies used.

3.4.1 Experiment process

The experiment process is as follows:

1. Welcome (5 min): Briefing and informed consent
2. Setup (5 min): Visual acuity test, demographic data and first sickness questionnaires
3. Training (3 min): Example scene and evaluation of 4 point clouds
4. Session 1 (10 min): Evaluation of 24 point clouds
5. Break (5 min): Second sickness questionnaires and rest
6. Session 2 (10 min): Evaluation of 24 point clouds
7. Debriefing (2 min): Third sickness questionnaires

The experiment started with a visual acuity test, to check for eventual participants that did not have a sight good enough to be part of the test.

If the visual acuity test was successful the experiment continued with the first questionnaire, this featured a first page to be filled with personal information (such as country, age, gender) followed by three different sickness questionnaires.

The users could then put on the visor and to start the virtual reality experience by seeing a 45 s example scene. In this scene a continuous loop of the training point cloud was shown. This allowed users to get comfortable with the virtual environment, its boundaries and with the concept of point cloud and its characteristics.

The experiment continued with the calibration process, necessary to obtain accurate eye tracking data. In case the calibration failed it was repeated until successful.

After the calibration the experiment moved to the training scenes, in which, the users saw four training point clouds in the following order: ThaiR1, ThaiR2, ThaiR5 and ThaiR5_L50. After each point cloud the user had to vote the quality of the sequence just seen. Participants were informed about the sequences they were seeing to make them aware of what was supposed to be bad or good quality.

After the training, when they felt ready, subjects could move to the actual testing. In this first part participants saw 24 point clouds, and, as in the training, they had to evaluate the quality of each one of them.

After this first block of evaluations, subjects were requested to take a small break. The break served a dual purpose, on one hand it gave some rest to the participants, on the other

hand it gave the opportunity to run a second form, only composed by the three sickness questionnaires.

When participants felt well rested they could follow with the second block of evaluations, which was still composed of 24 sequences to be evaluated.

Finally, subjects were asked to fill in a last form that was still composed by the three sickness questionnaires and lastly they were remunerated.

Overall the experiment took an average of 40 minutes to be completed.

3.4.2 Quality evaluation

To evaluate the quality of point clouds the single stimulus approach paired with the absolute category rating were chosen. The order of the sequences to be tested was randomized to avoid biases or learning effects.

3.4.3 Sickness questionnaires

Three different sickness questionnaires were proposed to the participants.

The three questionnaires are:

- Simulator Sickness Questionnaire (SSQ) [13]
- Virtual reality sickness questionnaire (VRSQ) [14]
- Vertigo score [21].

SSQ was developed to be used by the military to evaluate the effects of simulator training [13]. Since it was available it quickly became the standard in VR sickness questionnaires. However it is a long questionnaire, that focuses on symptoms that are not of interest in the case of VR.

For this reasons a shorter version of the form, called VRSQ was proposed in [14]. While this questionnaire reduces considerably the number of questions, it still asks many.

The vertigo score [21] is instead a one question form that aims at estimating the overall well-being of subjects in a quick way. The vertigo questionnaire asks the question ”Are you feeling any sickness or discomfort now?” and offers a 5 levels scale that ranges from ”no problem” to ”unbearable”.

This three questionnaires are the ones recommended by the ITU in [10]. It was chosen to have the subjects undergo all of the questionnaires to study the correlation between them.

3.4.4 Participants

42 subjects (21 men, 20 women and 1 that preferred not to answer) aged between 20 and 30 (mean of 23 and standard deviation 1.8) participated in the test. It is worth noting that 34% of the subjects were international students. Regarding the study background, 64% of the observers studied telecommunication engineering, another 12% studied in other areas of engineering and 5% in scientific subjects. Participants also answered a question about their experience using VR headsets. Results shown that most of the observers, 48%, used HMDs less than 5 times, for 26% of them it was the first time, 14% used it between 5 and 20 times and 12% used it more than 20 times. Participants were underwent a visual acuity test to assess the (corrected-to-) normal vision. To check on the color perception the Ishihara test was conducted, to which an observer resulted to be colorblind.

After the tests were conducted, some data had to be discarded. Specifically, two users' movement and eye tracking data were discarded due to a calibration problems during the test. Lastly the two users' sickness questionnaire data were discarded since the online forms did not register some of their answers.

Chapter 4

Experimental Results

The data collected during the experiment can be divided into 4 main categories:

- Scores on the quality of point cloud sequences
- Movement and eye tracking data
- Sickness questionnaire results
- Demographic data

This large amount of data data makes it possible to conduct many different types of analyses, to look for insights in many areas or to try to answer to specific questions.

In this work we will focus on:

- QoE analysis: Comparing scores given to different combinations of compression and loss rates.
- Movement analysis: Searching insights in pattern or interesting behaviours
- Sickness questionnaire analysis: Investigating the impact of VR on participants and the correlation between different questionnaires

4.1 Outliers removal

As mentioned previously, some participants data was removed for the analysis. The data removed was about the MOS and the movements for two observers who witnessed calibration problems during the experiment, however their sickness questionnaires were kept as they

could still provide useful information. Then, also the sicknesses questionnaire data of two different participants was took off from the analysis, this because for both of them, one of the form was not registered correctly.

Apart from this, no outlier removal procedure was applied. This choice was stems from the fact that, due to the novelty of this technology and the overall inexperience of participants, it is hard to define what is an outlier and what is normal.

4.2 QoE analysis

In this section the results on the mean opinion score obtained by different combinations of loss and compression rates will be discussed.

In Fig. 4.1 the MOS for the compressed sequences without transmission errors and 95% confidence interval are shown. As it can be easily seen, the scores of different compression rates are diverse and statistically independent. This is in accordance with was found in [5], which was the study used to choose the rates in the first place. It is worth noting that this experiment, compared to the one in the paper, was conducted with more participants.

In Figs. 4.2 and 4.3 are shown the MOS of all the sequences, compression and loss rates combinations. The graph is color coded to highlight different human figures (i.e. S23, S24, S25 and S26) and different compression rates between each figure. The 95% confidence intervals are also present.

First of all, as expected, for any pair of sequence and compression rate (e.g. S23 and R1, S23 and R2..), the influence of high loss rates is strong. In fact over a certain threshold it always creates statistically significant changes in the MOS.

Secondly, something that was also expected to be seen, and that is proved by the graph, is that generally low loss rates do not impact the score. In fact, most of the times, the MOS of point clouds without losses and the ones with low (or even mid) loss rates were statistically the same.

From the picture it can also be noted that, generally, compression rates play a big role even for the worst lost rates, in fact the scores obtained for compression R5 and the worst loss rate L50 are usually comparable with the scores obtained by the compression R1 without any losses.

On the other hand, the only noticeable difference between the general scores of different figures is that, sequence S26 (longdress) resulted in lower MOS with low compression rate and

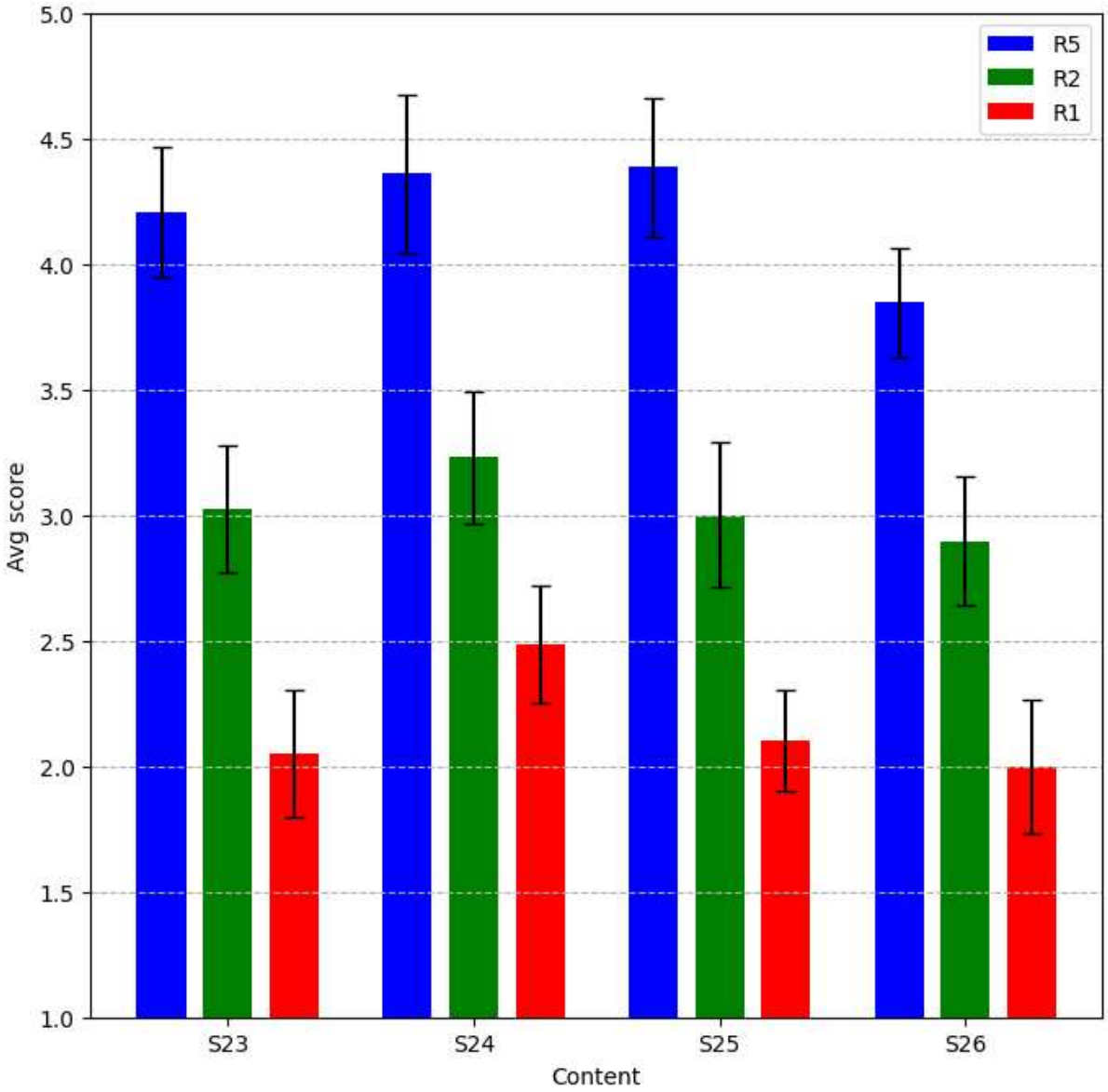

Figure 4.1: MOS aggregated on compression rates only

low or none loss rates compared to the other sequences. A similar trend was observed in [2], which hypothesized that the many different colors, patterns and movements that characterize the figure could have had an impact. Given the contradictory findings it is hard to draw any conclusion, more studies need to be conducted.

Fig. 4.4 shows the MOS aggregated on the loss rates, meaning it is averaged for the various compression rates. Each sequence has 4 MOS bars, the blue one regards the no losses, the green bar regards the lowest loss rate (for that sequence), the red one the middle loss rate and the yellow one the highest loss rate. Each MOS bar has the 95% confidence interval and, if relevant, the loss rate.

As it can be seen in the picture, the sequences S24 and S25 highlight really well the strong effects of different loss rates. As for the sequences S23 and S26, the effects is still shown, but

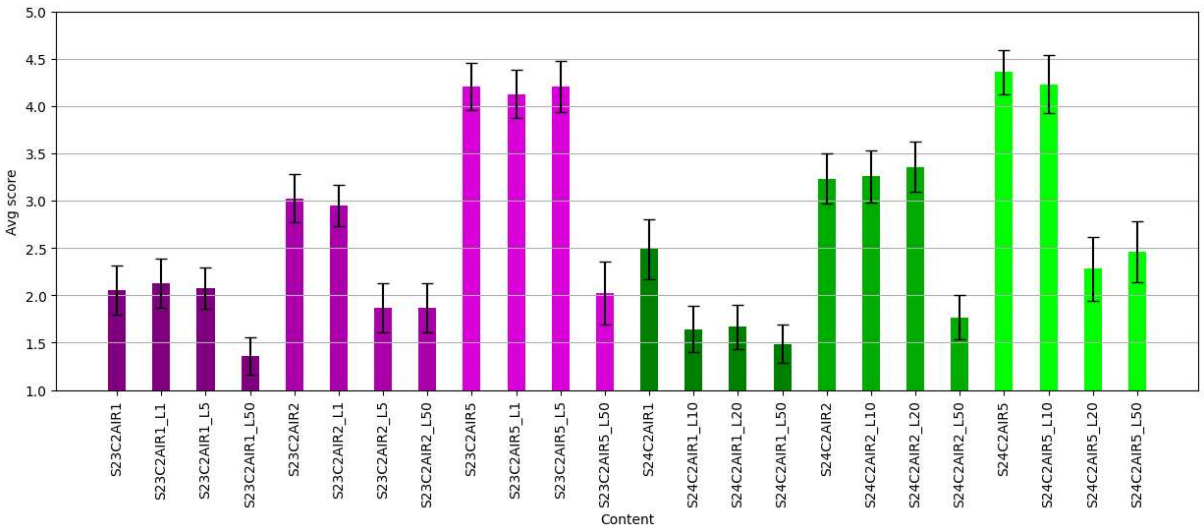

Figure 4.2: MOS for all combinations of loss and compression rates of S23 and S24

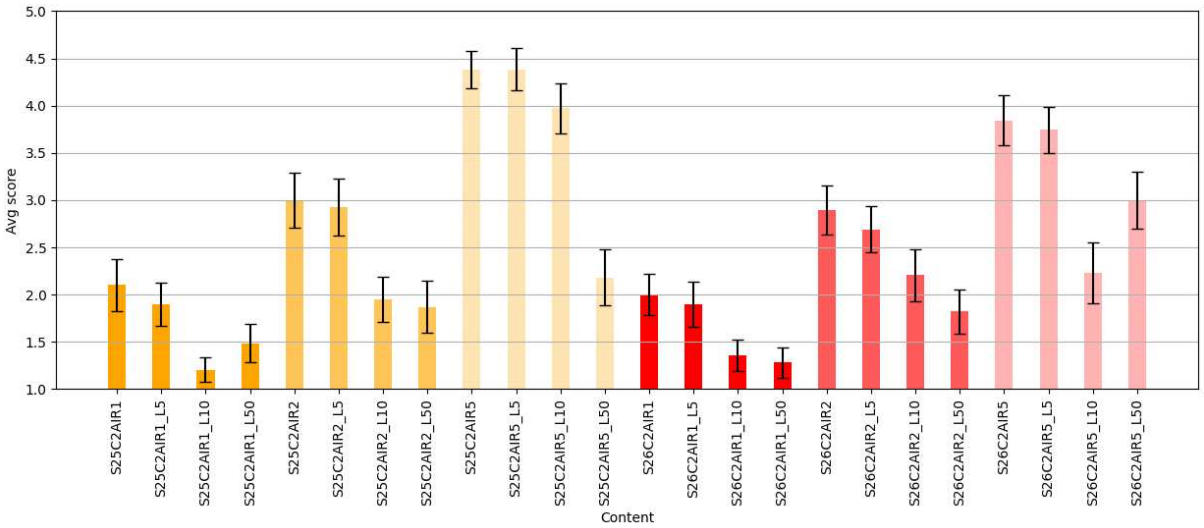

Figure 4.3: MOS for all combinations of loss and compression rates of S25 and S26

in their case it is more evident how the lowest loss rate does not almost impact the MOS.

It needs to be mentioned again that, the algorithm that simulates the transmission of errors is randomized, just like it would be random in a real case scenario. So it is expected that repeating the experiment would not give the same results, but the trend would be supposed to be the same. It is important to mention this, as it emphasize that it is normal to notice results that don't follow the pattern and that thus it is not possible to draw conclusions on single sequences or specifics loss rates threshold.

Lastly in Fig. 4.5 there are the MOS of point clouds aggregated on both compression and loss rates, thus loosing the information about the sequences. This, once again, shows the same trends found before, hence underlining how the findings were not dependent on the specific figures being seen.

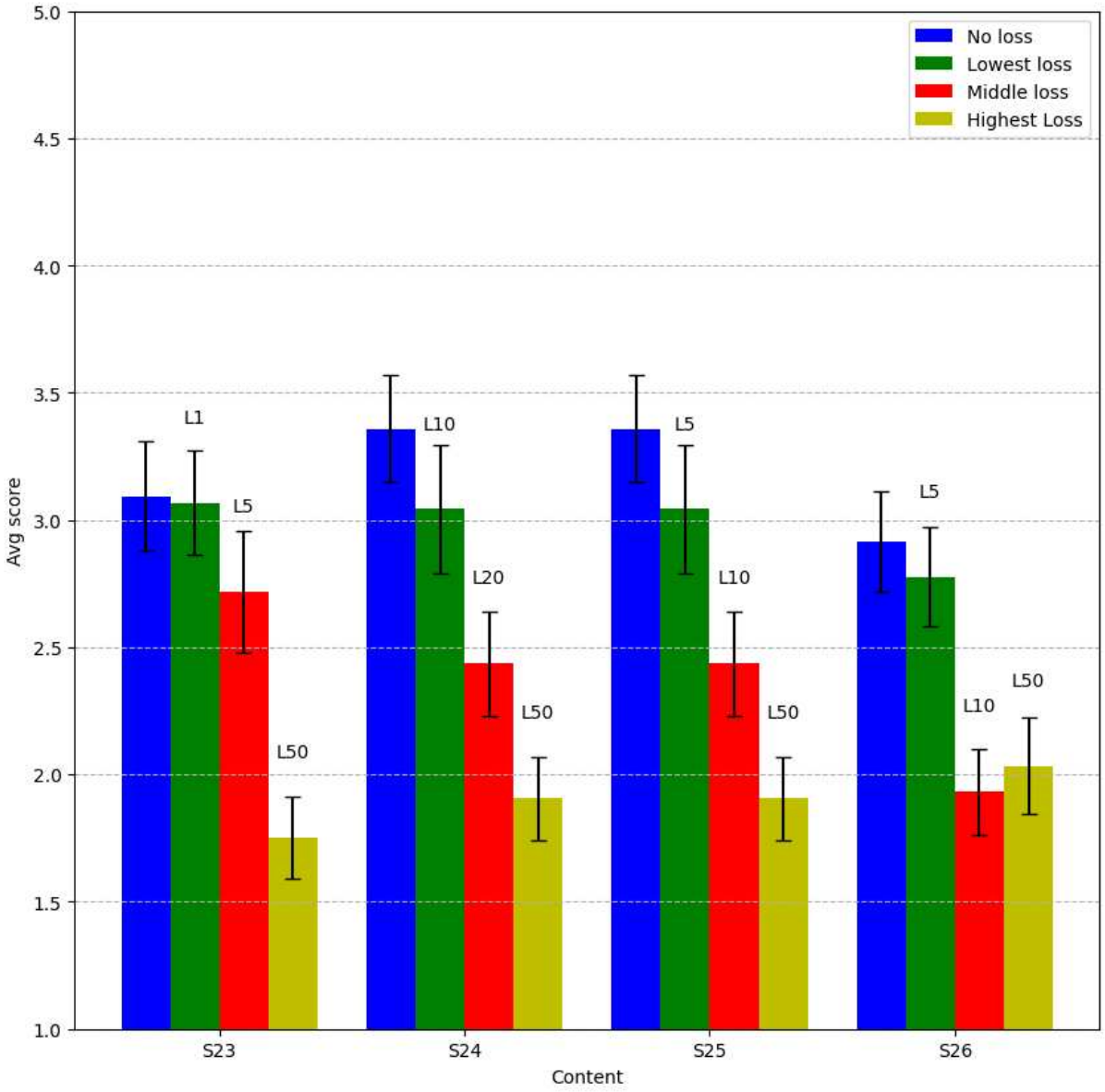

Figure 4.4: MOS aggregated on loss rates

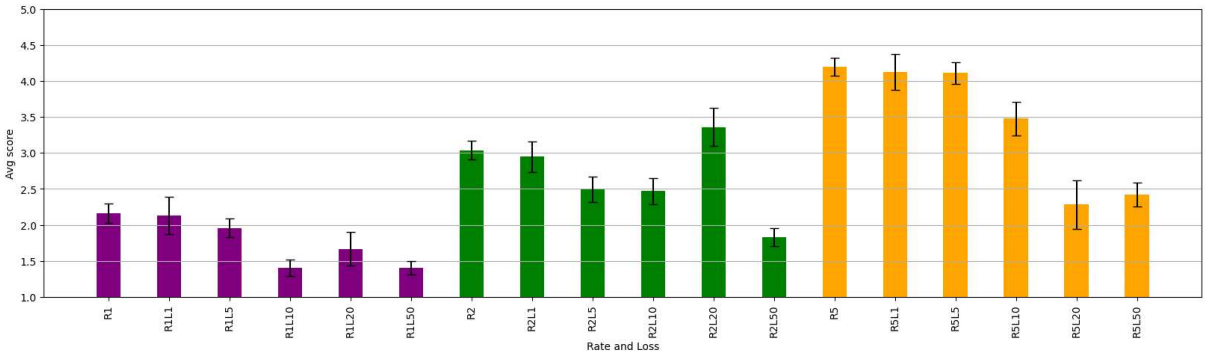

Figure 4.5: MOS aggregated on loss and compression rates

4.3 Movement analysis

Regarding the movement analysis, the idea was to investigate the behaviour and preferences of users while experiencing volumetric videos.

4.3.1 Movement heatmaps

It was decided to use heatmaps as the method to represent the movement data as they are easy and can provide visual insights. In all the subsequent heatmaps the red points represent the approximate location (unity world coordinates) of the point clouds, which are always turned to face the top of the graph.

In Fig. 4.6 the heatmaps of users behaviour during different session is shown. There are no visible differences between the two test sessions. However it seems that during the training participants were more spread around the space. This could be explained by the fact that the observers were inexperienced with point cloud, and that they learnt during the training what was best.

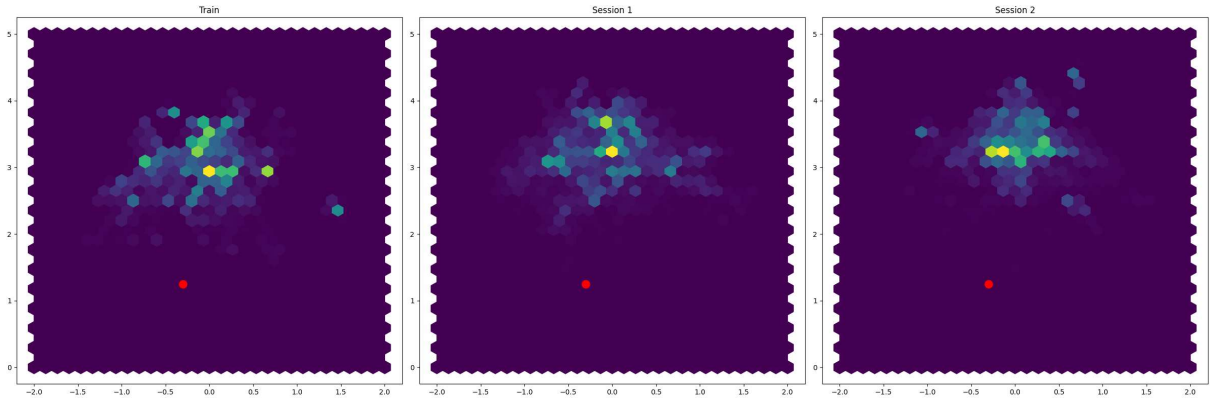

Figure 4.6: Movement heatmap of training, first session and second session

In the Figs. 4.7, 4.8, 4.9 are shown the heatmaps aggregated on sequence, compression rates and lastly loss rates. Overall there are no substantial differences between any of the plotted heatmaps, which is in accordance with [5]. The only noticeable dissimilarity can be seen in the loss rate heatmaps, in which the ones of loss rates L1 and L20 are less uniformly distributed, however, this is most probably due to the fact that L1 was used only for S23 and L20 only for S24, hence less data was collected for them with respect to the other rates.

All the heatmaps suggest that users prefer to experience point clouds from the front, however they often tend to move slightly to be able to also see the sides, this behaviour confirms what was found in [5].

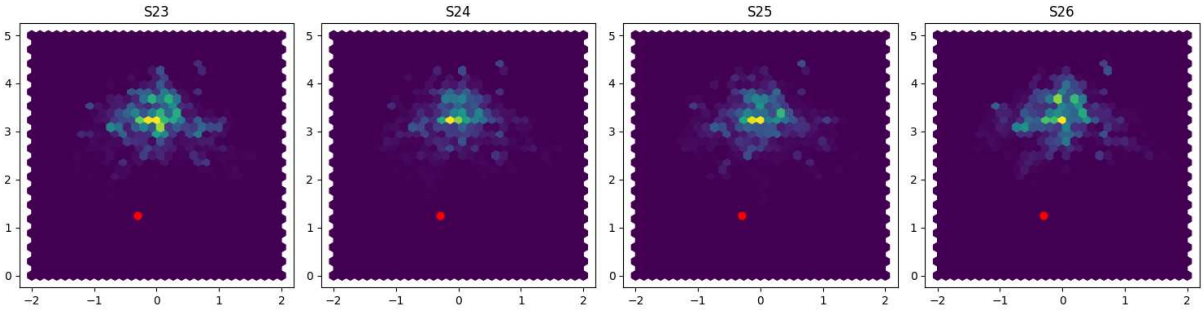

Figure 4.7: Movement heatmap aggregated on sequences

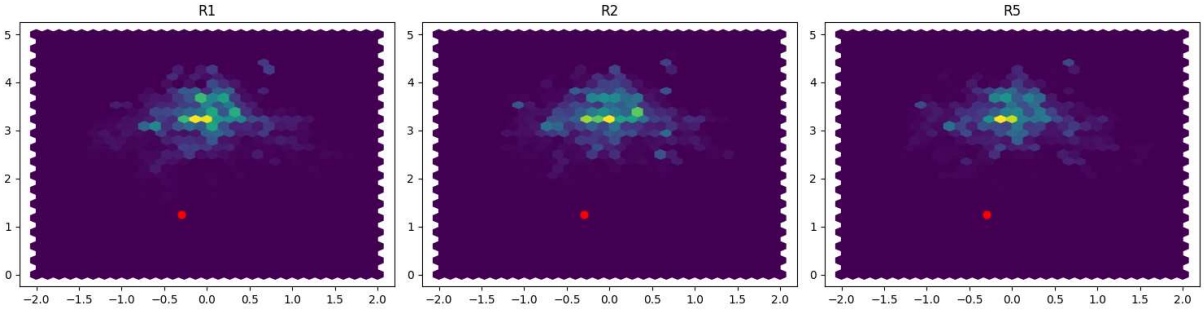

Figure 4.8: Movement heatmap aggregated on compression rates

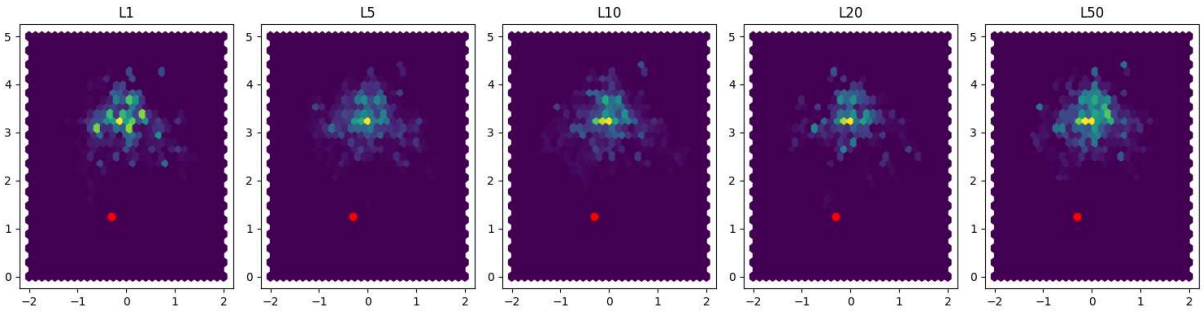

Figure 4.9: Movement heatmap aggregated on loss rates

4.3.2 Viewing direction

Studying the headset rotation can give an idea of what was in the field of view of observers during the experiment, hence can give some insights to what the participants might have been looking to. Since observers mostly moved their head along the pitch axis, this was the only one studied.

In Fig. 4.10 the viewing angle during the different sessions are shown, in Fig. 4.11 are shown the viewing angles for different sequences, lastly in Fig. 4.12 and 4.13 the viewing angle of different compression rates and different loss rates can be seen.

Once more, no dissimilarities worth noting are found between the various conditions evaluated, which is still in accordance with [5].

It can be noted that participants tend to look straightforward, or slightly upwards, this however is in contradiction with what was found in [5]. This contradiction proves that more studies are needed on the subject and that to really understand where participants focus when viewing volumetric video it is necessary to resort to the use of eye tracking data.

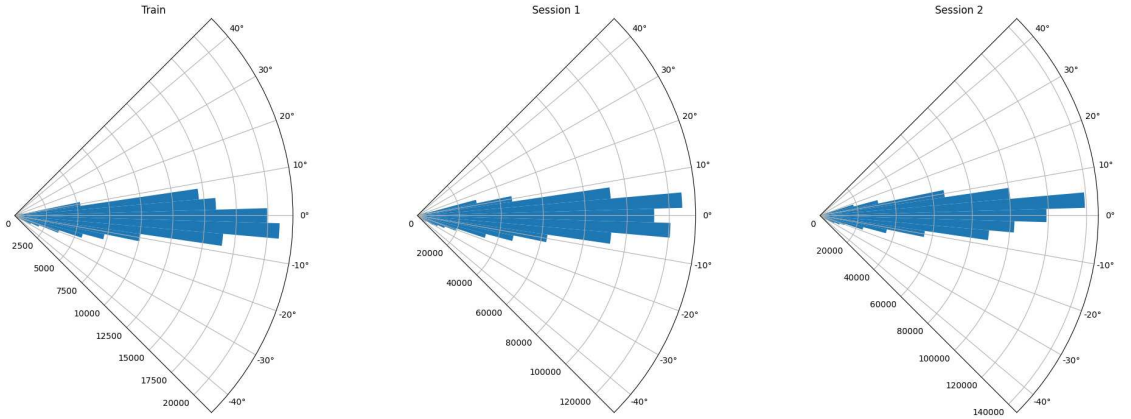

Figure 4.10: Viewing direction of training, first session and second session

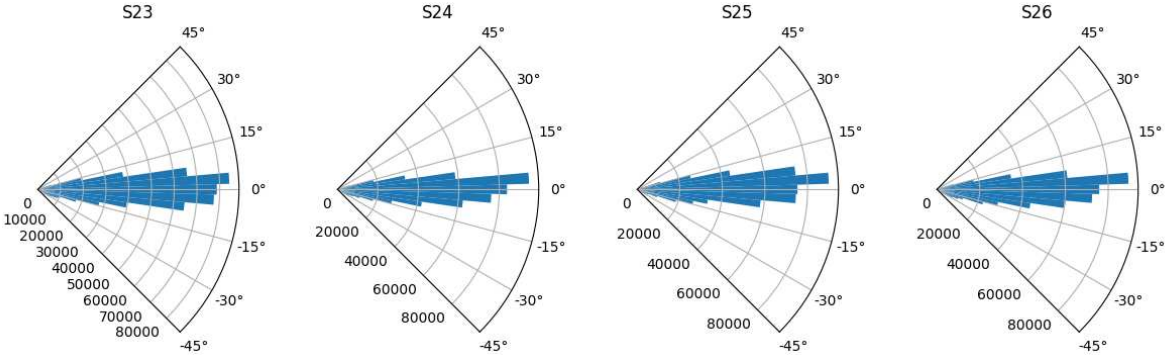

Figure 4.11: Viewing direction aggregated on sequences

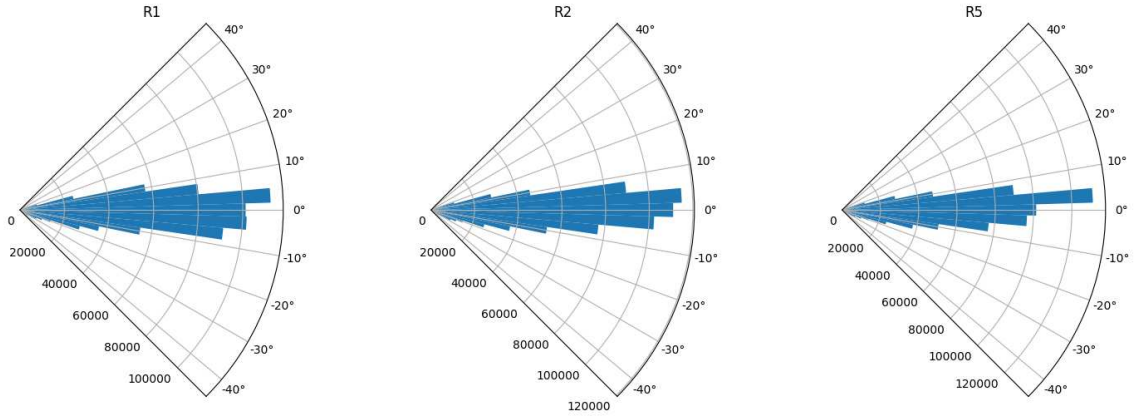

Figure 4.12: Viewing direction aggregated on compression rates

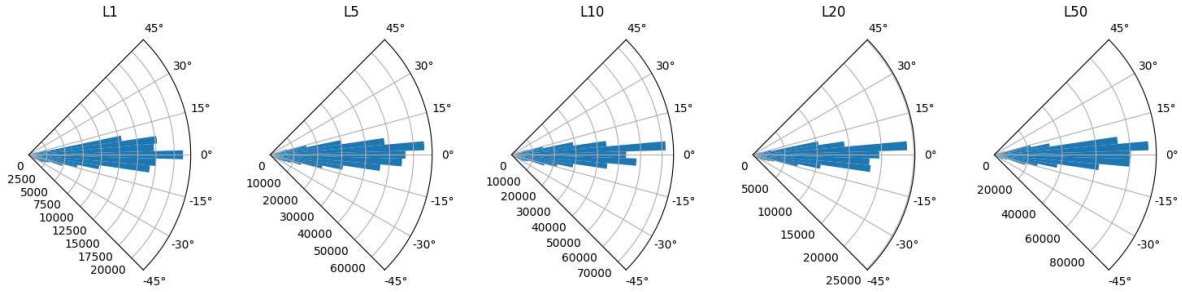

Figure 4.13: Viewing direction aggregated on loss rates

4.4 Sickness questionnaire analysis

As previously stated 3 different sickness questionnaires were filled by each participant, each one of them has been submitted 3 times, firstly before starting with the VR experience, the second in the short break, and the last at the end of the experiment. The 3 sickness questionnaires are:

- Simulator Sickness Questionnaire (SSQ): the de-facto industry standard, that will be used as reference
- Virtual Reality Sickness Questionnaire (VRSQ): shorter version of the SSQ
- Vertigo Score: One question questionnaire.

In Fig. 4.14 the overall distribution of the SSQ scores, as it can be seen the vast majority lies under a total score of 40, meaning the test did not create much discomfort to the participants.

The distribution of the severity of symptoms also confirm this. In fact, as it can be seen in Fig. 4.15, more than 95% of the symptoms were rated as none or slight. It is worth noting that sweating was rated as severe one time, however this was due to external factors, such as the abnormal warmth of the room the day of the test and the specific characteristics of the subject (which reported to suffer from excessive sweating often).

Overall this results are in accordance with [6].

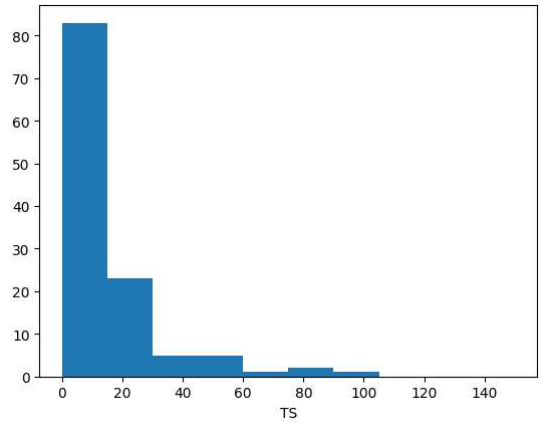

Figure 4.14: SSQs total scores distribution

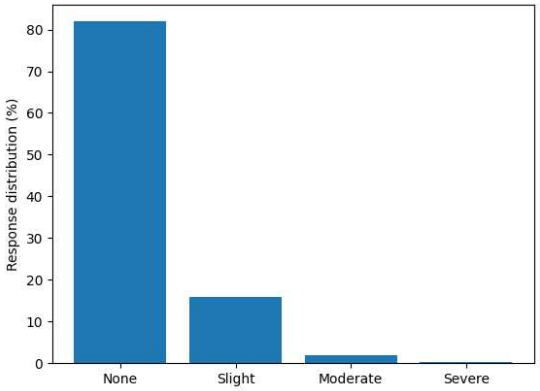

Figure 4.15: SSQs symptoms severity distribution

In Fig. 4.16 it is shown the trend of the average total scores, with a 95% confidence interval, during the various sessions. SSQ is plotted over the total score, while VRSQ and vertigo total scores are normalized to 1.

It can be seen that, overall, the three measures show some level of correlation. The SSQ is quite stable over the sessions, possibly due to the fact that it considers many other symptoms that are not really relevant for virtual reality applications. VRSQ, instead, slightly increases during sessions, which is reasonable. The vertigo score has the highest increase in the three sessions, and this can be explained by the fact that it only offers a 5 levels scale to evaluate discomfort.

Fig. 4.17 shows the correlation of the three questionnaires, comparing the total score obtain in the SSQ to the normalized one obtained by VRSQ and vertigo in the various sessions. Showing that VRSQ is highly correlated with the SSQ, while for vertigo there is a correlation with SSQ but it is not that strong.

Table 4.1 shows the Pearson correlation of the various questionnaire scores, which confirms what can be seen in Fig. 4.17.

This last result is partly in accordance and partly not with the findings of [6]. In fact, in that paper the authors found that SSQ and VRSQ are highly correlated, which is confirmed here, however also an high correlation between SSQ and vertigo (and hence between VRSQ and vertigo) was found, which is not the case in this study. It is to be noted that the research of [6] was conducted with over 300 participants in various labs, thus making it statistically stronger. For this reason, it can be assumed that repeating the research with more participants would yield results more similar to the one of the paper, however, more research is still needed to confirm this.

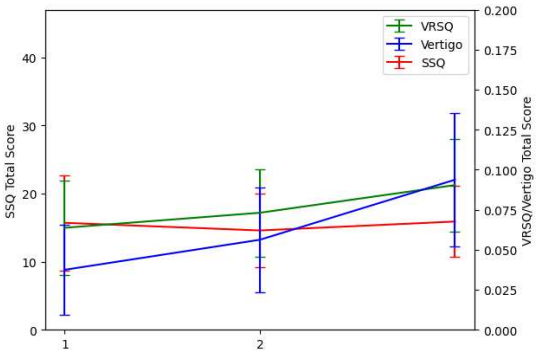

Figure 4.16: Viewing direction aggregated on compression rates

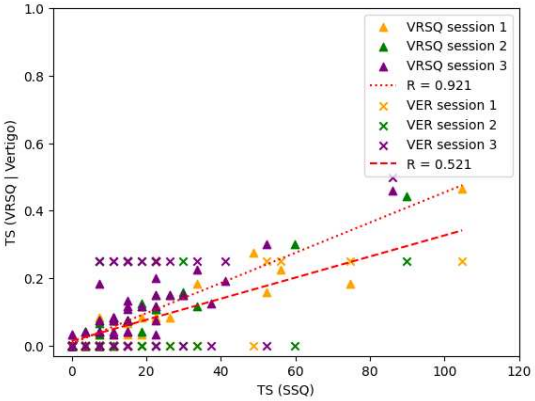

Figure 4.17: Viewing direction aggregated on loss rates

Questionnaire	SSQ	VRSQ	Vertigo
SSQ	1.000	0.921	0.522
VRSQ	0.921	1.000	0.558
Vertigo	0.522	0.558	1.000

Table 4.1: Sickness questionnaires pearson correlation

Chapter 5

Conclusions

In this research work the effects of compression and transmission on the perceived quality of dynamic point clouds were studied. The work consisted in a first part of research on the topic, this was followed by the development of the experiment, the actual experiment, which involved 42 participants and lastly the data analysis.

For the experiment, the single stimulus assessment was chosen and an absolute category rating with 5 levels was used as evaluation method. During the experiment participants had to fill in three forms, one at the start, one at the first break and the last one at the end. Each form contained three different sickness questionnaires.

The data collected was composed of demographics about the observers, sickness questionnaires results, quality evaluation scores of the point clouds, movement and eye tracking data.

This huge amount of data was used to get insights on many aspects.

First of all, in this work, thresholds on compression rates and perceived quality were compared with previous works confirming that, out of the standard V-PCC rates, R1, R2 and R5 consistently provide three very different quality levels. Then the focus was put on the effects of transmission losses and both compression and transmission losses. Here it was found that at low loss rates effects are not strong or not even noticeable. Moreover data made it evident that compression quality plays a big role even in presence of high transmission losses. Data made it also evident that quality impairments were not dependent on the human figure seen. About this last point, more studies, with different and varied point clouds, are needed.

Switching to the movement of participants, the data was analyzed in search of patterns and interesting behaviours. Not many insight were found, overall it was confirmed what was already discovered in previous studies, users tend to move very little, standing in front of the figure or at most moving slightly to the sides. Not many differences were visible between all

the various conditions. The only possible pattern is the evolution of the preferred places of users during the various sessions, in this regard it seems that users spread around the available space during the training, then, during the first session, they learnt their preferred location, which was then kept during the second session. However this needs to be checked with more studies.

Regarding users behaviour, also the viewing direction was studied briefly. About this topic, the data confirmed other studies on the fact that there are no big differences neither during the various sessions, nor between the various combinations of sequences, transmission and loss rates. However in this work it was found that users mainly looked straightforward or slightly upwards, while in a previous study it was found the opposite. This makes evident the need of studying eye tracking data to understand where users focus.

Lastly, an analysis on the discomfort given by virtual reality and the correlation of various sickness questionnaires was conducted. In this context it was seen that the experiment did not give much discomfort to the participants, generally they did not suffer of any of the symptoms of the SSQ. Regarding the comparison of different sickness questionnaires it was found that SSQ and VRSQ show a high correlation, hence making them interchangeable. On the other hand, SSQ and vertigo only showed a moderate correlation. Based on previous studies it was expected to find both questionnaires highly correlated with the SSQ.

Overall this research work provided some insights and confirmed some findings from previous works, however it was still inconclusive in some aspects or it provided results in disagreement with other studies. This makes it evident how more studies are needed in this subject.

In this work, a huge amount of data was collected, however only a part of it was used. Future works will be conducted on this data studying eye gazes to have insights on where participants focused during the experiment. Moreover average movement trajectories will be used together with a replay tool already developed, to create test sequences for objective metrics.

Bibliography

[1] E. Alexiou, N. Yang, and T. 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, 2020.

[2] G. Dandyyeva. Design and development of a subjective test methodology for quality of experience evaluation of point clouds. Master's thesis, Università degli Studi di Padova, 2022.

[3] E. d'Eon, B. Harrison, T. Myers, and P.A. Chou. "8i voxelized full bodies - a voxelized point cloud dataset". *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006*, 2017.

[4] 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.

[5] J. Gutierrez, G. Dandyyeva, M. Dal Magro, C. Cortes, M. Brizzi, M. Carli, and F. Battisti. Subjective evaluation of dynamic point clouds: Impact of compression and exploration behavior. Accepted at the 2023 European Signal Processing Conference (EUSIPCO).

[6] J. Gutiérrez, P. Pérez, M. Orduna, A. Singla, C. Cortés, P. Mazumdar, I. Viola, K. Brunnström, F. Battisti, N. Cieplińska, D. Juszka, L. Janowski, M. Leszczuk, A. Adeyemi-Ejeye, Y. Hu, Z. Chen, G. V. Wallendael, P. Lambert, C. Díaz, J. Hedlund, O. Hamsis, S. Fremerey, F. Hofmeyer, A. Raake, P. César, M. Carli, and N. García. Subjective evaluation of visual quality and simulator sickness of short 360° videos: Itu-t rec. p.919. *IEEE Transactions on Multimedia*, 24:3087–3100, 2022.

[7] M. Hosseini and C. Timmerer. Dynamic adaptive point cloud streaming. In *Proceedings of the 23rd Packet Video Workshop*. ACM, jun 2018.

[8] ITU-T. Evaluation criteria for PCC (Point Cloud Compression. Recommendation ISO/IEC JTC1/SC29/WG11 MPEG2016/n16332.

[9] ITU-T. Subjective methods for the assessment of stereoscopic 3DTV systems. Recommendation ITU-R BT.2021.

[10] ITU-T. Subjective test methodologies for 360o video on head-mounted displays. Recommendation ITU-T P.919.

[11] ITU-T. Vocabulary for performance, quality of service and quality of experience. Recommendation ITU-T P.10/G.100 (2017) – Amendment 1.

[12] E. S. Jang, M. Preda, K. Mammou, A.M. Tourapis, J. Kim, D.B. Graziosi, S. Rhyu, and M. 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.

[13] 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, November 1993.

[14] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69:66–73, 2018.

[15] M. Krivokuća, P.A. Chou, and P. Savill. "8i voxelized surface light field (8ivslf) dataset". *ISO/IEC JTC1/SC29 WG11 (MPEG) input document m42914*, 2018.

[16] J. Li, X. Wang, Z. Liu, and Q. Li. A qoe model in point cloud video streaming. *ArXiv*, abs/2111.02985, 2021.

[17] J. Li, C. Zhang, Z. Liu, W. Sun, and Q. Li. Joint communication and computational resource allocation for qoe-driven point cloud video streaming. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pages 1–6, 2020.

[18] Y. Liu, Q. Yang, Y. Xu, and Z. Ma. Which one is better: Assessing objective metrics for point cloud compression. 2021.

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

[20] R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, and A. Smolić. Affordable content creation for free-viewpoint video and vr/ar applications. *Journal of Visual Communication and Image Representation*, 53:192–201, 2018.

[21] P. Perez, N. Oyaga, J. J. Ruiz, and A. Villegas. Towards systematic analysis of cybersickness in high motion omnidirectional video. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3, 2018.

[22] I. Reimat, E. Alexiou, J. Jansen, I. Viola, S. Subramanyam, and P. Cesar. Cwipc-sxr: Point cloud dynamic human dataset for social xr. In *Proceedings of the 12th ACM Multimedia Systems Conference*, MMSys '21, page 300–306, New York, NY, USA, 2021. Association for Computing Machinery.

[23] S. Rossi, I. Viola, and P. Cesar. Behavioural analysis in a 6-dof vr system: Influence of content, quality and user disposition. In *Proceedings of the 1st Workshop on Interactive EXtended Reality*, IXR '22, page 3–10, New York, NY, USA, 2022. Association for Computing Machinery.

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

[25] J. van der Hooft, M. Torres Vega, C. Timmerer, A. Begen, F. De Turck, and R. Schatz. Objective and subjective qoe evaluation for adaptive point cloud streaming. 05 2020.

[26] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner. Towards 6dof http adaptive streaming through point cloud compression. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, page 2405–2413. Association for Computing Machinery, 2019.

[27] I. Viola, S. Subramanyam, J. Li, and P. Cesar. On the impact of vr assessment on the quality of experience of highly realistic digital humans. *Quality and User Experience*, 7:1–32, 2022.

[28] C-H. Wu, X. Li, R. Rajesh, W-T. Ooi, and C-H. Hsu. Dynamic 3d point cloud streaming: Distortion and concealment. In *Proceedings of the 31st ACM Workshop on Network*

and Operating Systems Support for Digital Audio and Video, NOSSDAV '21, page 98–105. Association for Computing Machinery, 2021.

[29] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic. Textured mesh vs coloured point cloud: A subjective study for volumetric video compression. In *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.