Università degli Studi di Padova

Dipartimento di Ingegneria dell’Informazione

Master Degree in ICT for Internet and Multimedia

User Pattern Exploration in Immersive Applications

Supervisor:
Prof.ssa Federica Battisti

Co-Supervisor:

Candidate:
Tridrik Dey
2050800

Graduation Date: 27th November, 2023

Academic Year 2022-2023

Abstract

Point clouds have become essential for different industries that require 3D modelling of objects or environments. Thus exploration of point clouds has become imperative and subjective evaluation is often used to understand how humans perceive, interpret, and interact with the point cloud data. This interaction as an outcome naturally involves physical mobility around the object. Hence, human trajectory plays a crucial role in accessing and analysing point cloud models. When exploring a 3D point cloud model, users navigate through the data set and view different parts of it from different angles and perspectives in order to fully understand and interpret the data. This physical movement pattern around a scene results in individualistic camera paths, taking into account the same data-set for all the users. Therefore, by aggregating all camera paths and extrapolating a mean trajectory, it's possible to formulate a collective reference path for generating a more comprehensive 2D video that can then be used for further subjective assessment and analysis. This further analysis highlights precise visual evaluation and gives insights into the completeness of the integral data.

The thesis is structured as follows:

Chapter 1: Introduction Brief explanation of point cloud technologies: Introduce the concept of point clouds. Challenges of point clouds: Discuss the dificulties and complexities associated with handling and processing point cloud data. Objectives of the thesis: Clear elaboration of the goals and aims of the research.

Chapter 2: State-of-the-Art Exploring Related works: This chapter dives deep into details of the contemporary works to provide an in-depth understanding of the latest developments. Data-sets: An exploration of the data-sets available for analysis sets the stage. It showcases the foundation upon which research is built. Compression Techniques: Delving deeper into various techniques used to compress point clouds. User Experience Quality: Examining user experience quality, examining how it factors into the handling and processing of point clouds. Subjective Metrics: Finally, the use of subjective metrics illuminates how researchers evaluate and measure the effectiveness of their approaches through subjective human experiences.

Chapter 3: Subjective Experimentation Collecting the Data set:: This section highlights the pivotal role of assembling a comprehensive data set, which forms the bedrock of the subjective experiment. Acquiring Environment: Ensuring the experiment unfolds in an environment conducive to accurate observations. Methodology: This part guides the entire experiment. It expounds on the systematic approach chosen to carry out the subjective experimentation, elucidating the steps taken to ensure validity, objectivity, and actionable results.

Chapter 4: Experimental Results Outliers Removal: This section delves into the process of identifying and removing outliers from the collected data. Quality of Experience (QoE) Analysis: In this section the focus shifts to the analysis of Quality of Experience (QoE). This facet examines how users perceive and interact with the point clouds, providing valuable insights into the usability and effectiveness of the presented subject matter. Movement Analysis: This component unveils the movement analysis aspect of the experimental results. It investigates how movements within the point clouds influence user engagement and comprehension.

Chapter 5: Conclusion This section binds together the insights garnered from the experimental results. It synthesizes the outcomes of outliers removal, Quality of Experience (QoE) analysis, and movement analysis to paint a comprehensive picture of the experiment.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Federica Battisti, for her guidance and support throughout this research. She has contributed immensely in guiding and teaching me about the subject matter, and her insights have been invaluable for my study. I am also grateful for her encouragement and patience, especially during the challenging times.

I would also like to thank the participants who took part in the subjective experiment. Their feedback was essential for the materialization of this study.

In addition, I would like to thank the following people for their help and support:

My late father, Bhabesh Chandra Dey and my mother Lipi Dey, for their unwavering passion, patronage and protection throughout my career. My brother, Trideb Dey, my elder sister Baralika Dey, my cousin Tridha Deb, my extended family and friends, for their encouragement and advice. The staff at the department of Information Engineering, for their assistance.

Finally, yet importantly, I am wholeheartedly thankful to my spiritual master, Revered Srimat Swami Suhitanandajee Maharaj, Vice President of Ramakrishna Math and Ramakrishna Mission, to the Divine Providence and to all those individual whose contributions I may have inadvertently overlooked in this acknowledgment.

I am grateful for the opportunity to have conducted this research, and I hope that it will make a contribution to the field of point cloud compression.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Need for Point Cloud Compression:

In recent years, The increasing popularity of multimedia, has necessitated the advancement of eficient compression and transmission algorithms to effectively manage the ever increasing demand for data-intensive multimedia content. Point clouds, which represent three-dimensional (3D) geometric data, is of high significance because of its extensive use in multimedia content. Their utility are spread out in diverse fields, such as virtual and augmented reality, topography and autonomous driving. Point clouds can be technically defined as a volumetric visual data that represents 3D scenes and objects, comprising set of points in 3D space. Each point has its geometric coordinates(x,y,z) establishing its spatial position containing characterized attributes like color, reflectance, or normal vectors. These point clouds can be derived from computer-generated 3D models or acquired from real-world environments through multiple cameras or specialized sensors like Light Detection and Ranging (LIDAR)s. In the context of sequential image frames, the terms ”dynamic point clouds” and ”volumetric videos” are pertinent and this multimedia format holds significant prominence in Augmented Reality and Virtual Reality technologies, affording users Six Degrees of Freedom (6DoF) viewing capabilities and immersive encounters that are crucial in applications like telepresence.

The need for subjective Quality Assessment over Objective quality assessment:

The process of quantizing point coordinates to fit specific precision ranges is reffered as voxelization. This involves mapping all points within a voxel, which is a three-dimensional grid's cubic unit, to the voxel's central point. Considering that a specific voxel can incorporate a multitude of points, often in millions, the associated data volume poses considerable challenges in terms of storage, transmission, and processing. To illustrate, an uncompressed dynamic point cloud featuring 1 million points per frame and operating at a frame rate of

30fps would necessitate a bandwidth of 3.6 Gbps. To overcome these challenges, substantial research and development initiatives has been devised to find compression techniques tailored for point clouds. It is imperative to thoroughly assess the effectiveness and eficiency of these compression methods. This ensures that the resulting compressed and transmitted point clouds maintain a satisfactory standard of quality and fidelity. Conventional objective quality metrics have been utilised to measure the effectiveness of compression algorithms. However, these metrics often fall short in fully capturing the comprehensive subjective sense of quality that human observers perceive. Moreover, in the context of emerging technologies such as point clouds and virtual reality headsets, it is imperative to delve into subjective evaluations and behavioral observations. This not only aids in comprehending user interactions with multimedia content but also establishes benchmark that contribute to refining precise objective metrics. Furthermore, subjective quality assessment plays a pivotal role in evaluating visual perception and user experiences related to compressed and transmitted point clouds. Gaining insights into how users perceive the quality of compressed point clouds is of paramount significance in optimizing compression algorithms and developing systems that align with user expectations.

Objectives:

The primary objective of this master's thesis is to investigate the subjective Quality of Experience (QoE) associated with compressed and transmitted point clouds. The study involves a comprehensive analysis of outcomes derived from a subjective testing procedure executed at the University of Padua. Through the implementation of a subjective quality assessment experiment, this work seeks to identify valuable insights into the perceptible effects of compression techniques and transmission processes on the perceived quality of point clouds. Additionally, it will touch upon the potential correlation between various rated compressed point clouds relative to original and the participants' patterns of movement. The central objective of this thesis encompasses: 1. Conducting a comprehensive review of the current state of the art and recent experiments within the field. 2. Developing simulation methodologies for the point clouds, employing one of the available datasets. 3. Devising and executing a structured subjective quality assessment experiment that serves to evaluate the perceived quality of compressed point clouds. 4. Analysing the acquired data and subjective ratings to discern the factors that wield influence over the quality of experience and the resultant movement behaviours of the participants. 5. Scrutinizing the outcomes to ascertain any potential correlations. Through the successful accomplishment of these objectives, this thesis undertakes to produce meaningful contributions to the advancement of point cloud compression methodologies.

Chapter 2

Related Works

In this chapter, we will embark on a comprehensive exploration of various facets pertinent to our subject matter. We will delve into the examination of accessible datasets, elucidate key conceptual foundations, and provide an in-depth analysis of state-of-the-art compression algorithms. Additionally, we will scrutinize the methodologies employed in the subjective evaluation of the quality of point clouds.

2.1 Point cloud datasets

Point clouds are versatile and diverse representations of 3D data, encompassing a wide range of characteristics and applications. They can vary significantly in terms of detail, scale, acquisition method, and use cases, making them a rich area of study within computer vision and related fields.

Diversity in Point Clouds:

Point clouds can encompass a vast spectrum, from highly detailed heritage models to extensive topographical scans. These variations cater to different applications, such as cultural preservation and geographic mapping. Some point clouds are captured dynamically, often with sparse data, to support applications like autonomous vehicles. In contrast, others are meticulously acquired in controlled environments, yielding high-quality dynamic point clouds. The technology used for point cloud acquisition can be categorized into direct and indirect methodologies. Direct methods involve specialized hardware like time-of-flight cameras and LIDAR scanners, purpose-built for 3D data capture. Indirect methods rely on post-processing algorithms to extract 3D information from 2D data, such as images from multiple cameras.

Academic Contribution:

The academic community has made significant contributions to the development of point cloud datasets. These datasets are instrumental in advancing research in various domains, including computer vision and robotics.Due to the diversity of use cases and technologies, point cloud datasets exhibit variations in content, capture settings, post-processing techniques, data quality, and the size of objects.

Static Point Clouds: Static point clouds are 3D representations of immobile objects or scenes, serving as digital snapshots of the physical environment at a specific point in time. These point clouds are typically created using specialized equipment like LIDAR scanners or structured light systems, which emit laser or light pulses to measure the distance between the sensor and objects in the environment. Static point clouds are extensively utilized in various fields, including archaeology, cultural heritage preservation, and architectural documentation.

One illustrative example of a static point cloud dataset is the PointExtended Reality (XR) dataset [7] as is Figure 2.1, made available by the École polytechnique fédérale de Lausanne (EPFL). This dataset is especially noteworthy for its focus on high-quality cultural heritage models. It contains 20 meticulously captured point cloud models of cultural artifacts, architectural marvels, and heritage sites. These static point clouds are remarkable for their ability to faithfully represent the intricate details, textures, and geometric intricacies of these objects and structures. Researchers and professionals in the fields of cultural preservation, 3D modeling, and historical documentation rely on datasets like PointXR to advance their work, ensuring the accurate preservation and analysis of culturally significant assets for future generations.

(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 [7]

Dynamic Point Clouds: Dynamic point clouds represent a dynamic, real-time capture of three-dimensional (3D) data, providing a detailed and time-varying representation of the surrounding environment. Unlike static point clouds, which capture stationary scenes or objects, dynamic point clouds are capable of capturing moving subjects, scenes, or objects in motion. They are commonly used in various applications, including robotics, augmented reality, virtual reality, and computer vision. Capturing dynamic point clouds often requires sophisticated setups with synchronized cameras or sensors to track and record the movement of objects or scenes accurately. These dynamic representations play a crucial role in applications such as motion analysis, object tracking, and immersive experiences, where real-time interaction and responsiveness are essential.

Challenges in Dynamic Point Cloud Datasets: - Dynamic point cloud datasets are less common due to their complexity. They typically require intricate setups with numerous synchronized cameras in controlled environments, often referred to as ”green rooms.” - Capturing dynamic scenes with high-quality point clouds is challenging, and not many datasets focus on this aspect.

Some of the examples of different databases for dynamic point cloud are as following:

V-SENSE Volumetric Video Quality Database 2 [11]:

The ”V-SENSE Volumetric Video Quality Database 2” is a noteworthy dataset within the domain of dynamic point clouds. It was meticulously captured by the V-SENSE team at Trinity College, and its primary focus is on providing data for volumetric video quality assessment and research. Dataset Contents: This dataset comprises four distinct point clouds that depict moving human figures engaged in various activities, including interactions with objects. These dynamic scenes are captured with the intention of simulating real-world scenarios, making them valuable for a wide range of applications, including computer graphics, virtual reality, and immersive communication. Limitations: Despite its significance, the ”V-SENSE Volumetric Video Quality Database 2” is not without its limitations. Notably, the dataset exhibits noticeable artifacts within the point clouds. These artifacts can manifest as irregularities, distortions, or inaccuracies in the 3D data, potentially affecting the overall quality and realism of the captured scenes. Another limitation of the dataset is its relatively low point density. Point density refers to the number of data points (or points per unit of space) used to represent a 3D object or scene. In the case of this dataset, the lower point density can result in a less detailed representation of the captured scenes, potentially making them appear blurry or less realistic when compared to datasets with higher point densities. Research Implications: While the ”V-SENSE Volumetric Video Quality Database 2” has these limitations, it remains a valuable resource for researchers and practitioners in the field. The dataset’s real-world scenarios and dynamic content provide opportunities for studying volumetric video quality and developing algorithms to enhance the fidelity of captured scenes. Researchers using this dataset should be mindful of its limitations and consider them when designing experiments or drawing conclusions. Despite these challenges, the dataset contributes to the ongoing efforts to advance the field of volumetric video and dynamic point cloud research. It underscores the importance of addressing artifacts and improving point density to further enhance the realism and quality of dynamic point clouds in future datasets and applications.

CWI Point Cloud Social XR Dateset [13]: The CWI Point Cloud Social XR Dataset as in Figure 2.2 is a valuable addition to dynamic point cloud datasets. It was created by capturing dynamic point clouds using seven synchronized Azure Kinect DK devices. This dataset comprises 45 sequences, each depicting individuals engaged in common social activities within real-time communication scenarios. These scenarios cover four distinct social experiences, including "Education and Training," "Healthcare," "Communication and Social Interactions," and "Performance and Sports." Notably, the dataset includes synchronized audio tracks, enhancing the immersive experience of volumetric videos.

Figure 2.2: Examples of sequences from the CWI dataset [13]

However, it is important to note that the use of a relatively low number of cameras during data capture has led to certain limitations in the dataset as shown in Figure 2.2. These limitations manifest as a significant number of holes and artifacts in the output point clouds. While these flaws may make the dataset less suitable for certain quality of experience experiments, they contribute to the dataset's authenticity and realism, particularly in the context of real-time communications. As a result, the CWI Point Cloud Social XR Dataset holds significant importance within the research community.

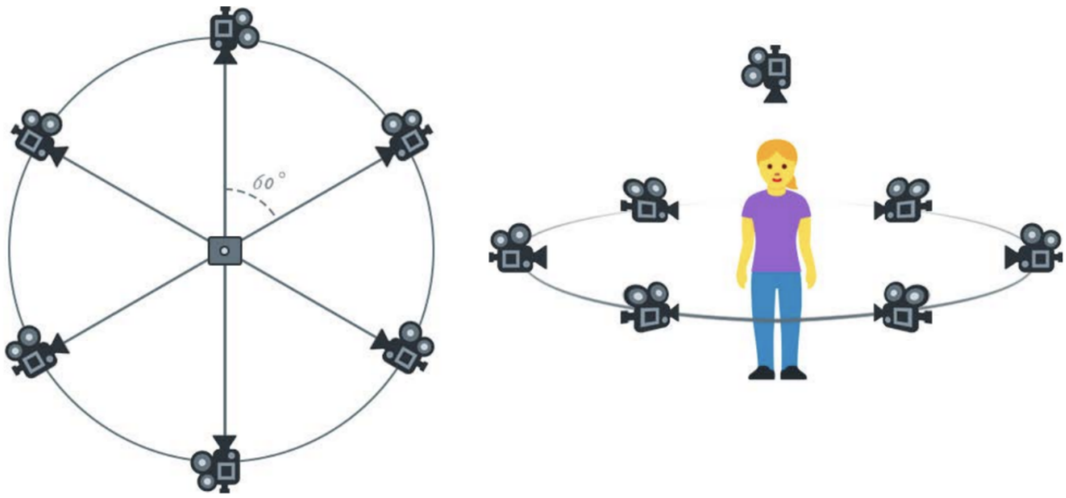

Figure 2.3: Example of a camera arrangement used for the CWI dataset [13]

JPEG Pleno Datasets - 8i Voxelized Full Bodies (8iVFB v2) [8]: One of the most renowned and widely used datasets in the research community is the JPEG Pleno dataset, specifically the ”8i Voxelized Full Bodies (8iVFB v2).” This dataset consists of four sequences capturing full human body movements. The data was collected using a set of 42 RGB cameras organized into 14 clusters, with each sequence recording at 30 frames per second (fps) for a duration of 10 seconds. The resulting data has an impressive space resolution of 1024x1024x1024 voxels, resulting in 10 bits of depth information. In practical terms, for a subject with a typical height of 1.8 meters, a voxel at depth 10 corresponds to approximately 1.75mm on one side. Figure 2.3 shows real setup used in 8i labs.

Figure 2.4: Real set up used in the 8i labs

8i Voxelized Surface Light Field (8iVSLF) Datasets [22]: A complementary dataset, also provided by MPEG, is the ”8i Voxelized Surface Light Field (8iVSLF) Dataset.” This dataset offers a single-frame version of the sequences available in 8iVFB v2, along with a high-quality voxelized volumetric video. The volumetric video sequence was captured using a set of 39 RGB cameras organized into 13 clusters, each functioning as a logical RGBD camera. The capture rate is maintained at 30 fps for a total duration of 10 seconds. After merging the data, a 3D representation with 12 bits of depth information is obtained, resulting in an impressive space resolution of 4096x4096x4096 voxels. Notably, in the volumetric video, each voxel represents approximately 1x1x1 mm of the physical capture space. Given that the human figure typically occupies less than half of the height of the voxelized space, the height of the captured human subjects is under 2 meters.

In conclusion, these dynamic point cloud datasets, including the CWI Point Cloud Social XR Dataset, 8i Voxelized Full Bodies (8iVFB v2), and 8i Voxelized Surface Light Field (8iVSLF) Dataset, serve as invaluable resources for researchers in various fields. They enable the study of complex dynamic scenes and the development of cutting-edge technologies in the realms of computer vision, augmented reality, and immersive communication. While these datasets may have certain limitations, they significantly contribute to the advancement of 3D data processing and analysis.

2.2 Point Cloud Data Compression

The points provided below discusses the the challenges associated with point cloud data and the efforts made by the Moving Picture Experts Group (MPEG) to standardize compression techniques for different types of point cloud data:

Storage and Transmission Challenges: Point clouds are data representations used in 3D modeling, where objects are represented as a large number of points in space. These point clouds can be very data-intensive because, to accurately represent an object, a significant number of points are required. Additionally, each point may have associated attributes like color, surface normal, and reflectance. This leads to significant storage requirements and high transmission costs, making it challenging to handle large point cloud datasets eficiently.

Unstructured Nature of Point Clouds: Unlike regular grids used in 2D images or 3D volumetric data, point clouds are unstructured. Points can be located anywhere in 3D space, and their distribution is not fixed or regular. This unstructured nature makes it dificult to adapt traditional 2D compression algorithms or develop new ones tailored to point cloud data.

Variability in Point Cloud Data: Point clouds are dynamic and vary over time. In some applications, such as capturing real-world scenes using LiDAR or 3D scanning, the number of points in a point cloud may change over time. This dynamic nature adds complexity to compression algorithms as the data is not fixed.

MPEG Point Cloud Compression Standardization: To address these challenges, the MPEG group initiated a standardization effort called MPEG Point Cloud Compression (PCC). This effort aimed to develop compression standards specifically tailored to point cloud data. To cover the wide range of applications, three categories of test data were defined:

Static Point Clouds: These involve high-quality, static point clouds with millions to billions of points and optional additional attributes like color. This category represents scenarios where detailed and high-quality 3D data is essential.

Dynamic Point Clouds: This category deals with scenarios where point clouds are less dense, may include color information, and involve temporal changes. Dynamic point clouds are more common in applications like real-time 3D tracking.

Dynamically Acquired Point Clouds: This category represents point clouds captured dynamically over time, typically with millions to billions of points, color information, optional attributes, and sequences of static point clouds captured on the fly.

Compression Technologies Chosen:

Surface Point Cloud Compression (S-PCC): This technique is suitable for static point clouds and focuses on compressing the spatial positions of points, often suitable for sparse point distributions.

Video-Based Point Cloud Compression (V-PCC): Designed for dynamic point clouds, V-PCC involves projecting 3D models into 2D frames, which are then merged into images. Existing video coding standards are leveraged to compress these sequences of images.

LiDAR Point Cloud Compression (L-PCC): Tailored for dynamically acquired point clouds, L-PCC focuses on compressing LiDAR data, which is commonly used in applications like autonomous vehicles and remote sensing.

Geometry-Based Point Cloud Compression (G-PCC): Recognizing the similarity between S-PCC and L-PCC, these two techniques were merged into G-PCC. G-PCC is used for compressing point clouds with sparse distributions and focuses on coding spatial positions.

In summary, the development of compression standards for point cloud data by MPEG addresses the unique challenges posed by the large, unstructured, and dynamic nature of point cloud datasets. Different compression techniques are tailored to specific scenarios, allowing for eficient storage and transmission of point cloud data across various applications.

2.3 V-PCC standard

The V-PCC (Video Point Cloud Compression) framework is centered around the optimization of existing video codecs to eficiently compress volumetric video data. To accomplish this, the point cloud sequence must undergo a transformation into a 2D video format, making it compatible with standard video codecs. However, this projection from 3D to 2D inherently results in some loss of information. Consequently, the algorithm generates two separate video sequences: one containing geometry information and the other containing texture information.

Additionally, the algorithm computes supplementary metadata, specifically an occupancy map and auxiliary patch information. These metadata elements are crucial for enabling the decoder to properly interpret the two video sequences. It's important to note that the metadata information is compressed separately and then integrated with the compressed video bitstreams. It's worth mentioning that the metadata typically occupies a modest portion of the overall bitstream, ranging from 5 percent to 20 percent [25]. This underscores the fact that the majority of the essential information is encapsulated within the two video sequences, which are already eficiently compressed using established techniques.

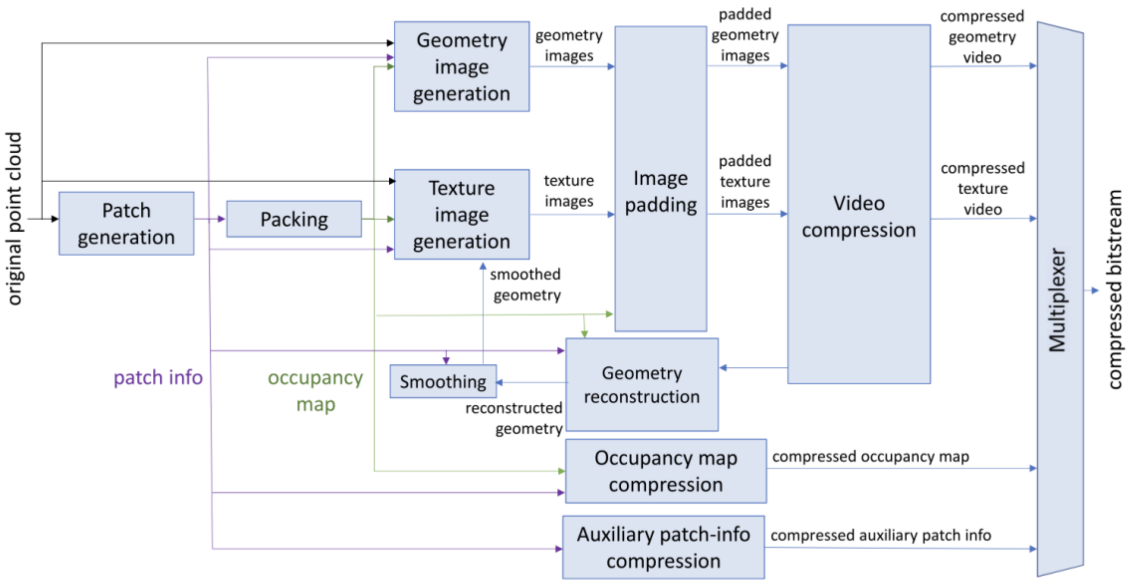

Figure 2.5: V-PCC coding scheme [25]

Patch Generation and Packing:

In order to maximize the eficiency of video codecs, the V-PCC compression process prioritizes the creation of video sequences characterized by strong spatio-temporal correlation. A fundamental goal in this context is the establishment of a temporally coherent, low-distortion, and injective mapping, whereby each point within the 3D point cloud is assigned to a cell within the 2D grid [9]. A simplistic approach, such as straightforwardly projecting the 3D model onto a cube or sphere, fails to ensure lossless reconstruction due to its disregard for common issues like auto-occlusion of points and hidden surface problems inherent in point clouds. These issues often result in significant distortions [5] To address these challenges, V-PCC employs a segmentation strategy, dividing the point cloud into numerous 3D patches, which are essentially contiguous regions. Subsequently, each patch is independently projected onto the 2D plane. Given that multiple points within the point cloud can map to the same 2D pixel, this approach allows for the creation of multiple layer maps, each associated with a specific depth range. An illustrative example of a 3D patch and its projections can be found in Figure 2.6.

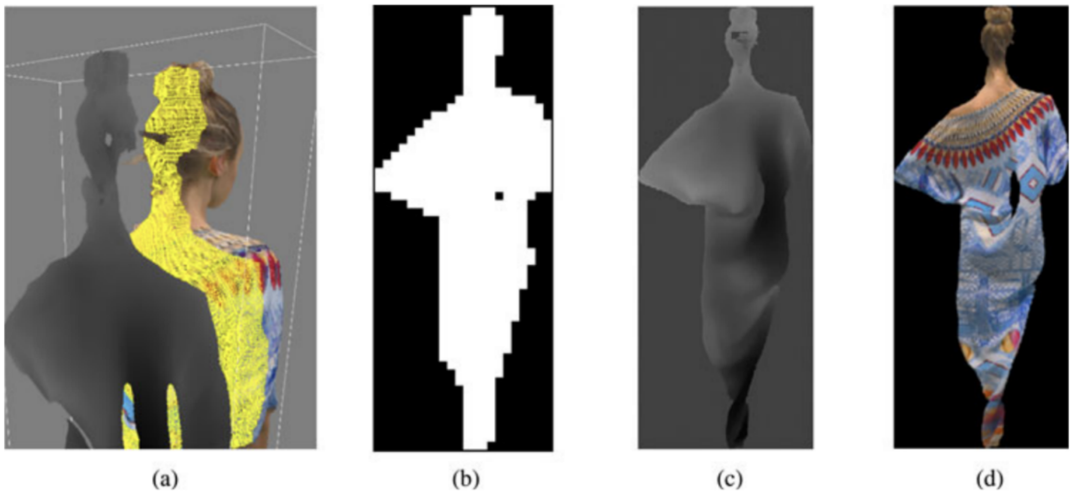

Figure 2.6: Example of patch projection: (a) 3D patch, (b) 3D Patch Occupancy Map, (c) 3D Patch Geometry Image, (d) 3D Patch Texture Image [5]

The process of packing each 2D patch into a 2D image involves several steps. For the initial frame, a blank 2D image of dimensions WidthxHeight pixels is created. Next, all the 2D patches are sorted by size, and, starting with the largest, a straightforward scan search algorithm seeks the first available location that guarantees insertion without overlap. To enhance the chances of finding suitable locations, four orientations are explored in conjunction with mirroring. The occupancy map is then populated, taking into account pixels containing valid depth values. Should no available location be found, the Height of the 2D image is doubled, and the process continues. Once all the patches are accommodated, the image is trimmed down to the minimal achievable Height.

As previously mentioned, establishing spatio-temporal correlation between frames is of paramount importance. To achieve this objective, the algorithm systematically seeks matches between 2D patches across different frames. When a match is identified, V-PCC endeavors to insert it into a location within the image that aligns closely with its position in the previous frame.

Geometry and Occupancy Map:

In the context of the geometry map, only the luminance channel is utilized for encoding the depth information of the points. Given the inherent arbitrariness in the shape of 2D patches, an additional map, known as the occupancy map, becomes essential. The occupancy map is essentially a binary image where a pixel with a value of 1 signifies a pixel in the geometry map that contains depth data. Conversely, pixels assigned a value of 0 correspond to locations within the 2D image that are not utilized to convey depth information. In Figure 2.7 an example of packed patches is depicted below.

Figure 2.7: Example of packed patches, respectively: Occupancy map, Geometry map and Texture map [5]

Additional Steps:

Within the V-PCC algorithm, a multitude of supplementary steps are undertaken to enhance compression eficiency, output quality, and enable seamless data decoding. Notable among these steps are:

Image Padding: This involves assigning values to the vacant pixels within the texture and geometry maps. The aim is to create smoother transitions between patches, thereby improving the performance of video codecs.

Geometry and Color Smoothing: This process may be applied to mitigate artifacts that can arise at the boundaries of patches during their reconstruction as a 3D point cloud. It contributes to a more visually appealing result.

Atlas Metadata: The decoder necessitates additional information known as atlas metadata. This includes details like patch positions, rotations, or other pertinent data that must be transmitted to the decoder for accurate reconstruction.

Decoder:

The decoding process unfolds in two distinct phases. Initially, it focuses on decoding the bitstream and retrieving critical components such as the occupancy and geometry maps, 2D video frames with attributes, and patch-related information for each frame. Once all data is available, the algorithm embarks on geometry and attribute reconstruction. In this phase, patches are re-projected to their original positions in the 3D space, resulting in a point for every occupied pixel within the occupancy map.

However, it's important to note that the video encoding and processing applied during encoding may introduce artifacts and discontinuities in the reconstructed point cloud. To address this, a smoothing process is often applied to alleviate the severity of these artifacts, ultimately contributing to a more refined output [9]. In Fig. 2.8 the decoding pipeline followed by the V-PCC standard is shown.

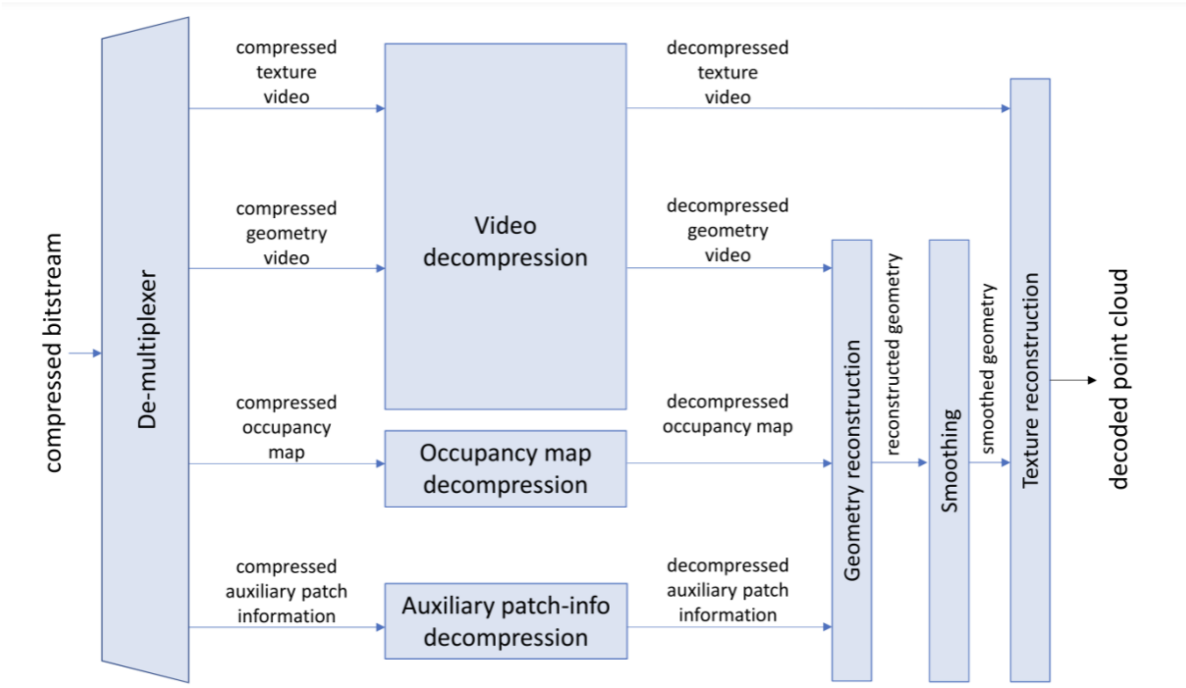

Figure 2.8: V-PCC decoding scheme [25]

2.4 Transmission

Point clouds have garnered significant attention in recent years due to their unique characteristics and potential applications. Many envisioned use cases involving point clouds, particularly those exploiting volumetric videos, demand real-time transmission across unreliable networks. A prime example is telepresence applications, where point clouds can provide 6 Degrees of Freedom (6DoF), transforming a standard video call into an immersive experience, granting users the freedom to explore and move as they would in the physical world. However, transmitting and delivering point clouds over networks present substantial challenges. These challenges include the need for ultra-high-bandwidth transmission, which is currently not readily available [21]. Furthermore, the unreliability of typical networks hinders real-time interactions and applications that rely on accurately timed point clouds.

While eficient compression techniques are being explored to address the bandwidth requirements of point clouds, limited attention has been given to achieving eficient and reliable transmission. Moreover, few studies have delved into the impact of unreliable networks on the quality of the user experience.

One notable advancement in the realm of point cloud transmission over networks is the work by Hosseini et al. [12], which integrates Dynamic Adaptive HTTP Streaming (DASH) with an algorithm that spatially sub-samples dynamic point clouds to create representations at multiple quality levels. Subsequent research extended DASH, such as [21], where a rate adaptation heuristic considers factors like location, focus, and available bandwidth to set the quality of point cloud objects. In [20], Li et al. aimed to maximize Quality of Experience (QoE) by selecting optimal quality levels for partitioned point cloud video tiles while considering computational and communication constraints.

These studies have yielded valuable insights and advancements, but they all rely on DASH, an algorithm inherently designed for reliable network transport protocols. This contrasts with the concept of real-time applications, where unreliable transport protocols (e.g., UDP) are preferred due to their ability to minimize latency. Few studies have explored the challenges of streaming point clouds over error-prone networks. In [4], the Network Abstraction Layer Units (NALUs) of the V-PCC bitstream are examined, and to simulate packet losses, some are altered to sequences of zeros. The same paper proposes a technique to minimize losses and retrieve as much information as possible. Therefore, it is imperative to expand research in this area to gain a deeper understanding of QoE thresholds and explore potential solutions that will facilitate the widespread adoption of multimedia applications involving point clouds.

2.5 Quality of Experience

Throughout the preceding sections, the term ”Quality of Experience” (QoE) has been frequently invoked without a formal description. In [2], QoE is defined as ”The degree of delight or annoyance experienced by a user while using an application or service.” It’s worth noting that, as stipulated in the same recommendation document, the definition of QoE is expected to evolve over time due to ongoing research in this field.

Several factors influence QoE, including the type and characteristics of the application or service, the context in which it is used, user expectations and their fulfillment, cultural background, socioeconomic considerations, psychological profiles, the emotional state of the user, and other factors whose number is likely to expand as research progresses [2].

There are two primary methods for evaluating QoE: subjective metrics and objective metrics. Subjective metrics rely on human perception and involve conducting subjective assessment methods. While these methods yield results that represent ground truth, they are often time and resource-intensive.

On the other hand, objective metrics are algorithms designed to assess the degradation in users' QoE. They are fast, eficient, produce reproducible outcomes, and yield numerical results for easy comparison. However, it is possible that their results may not always align with end-user perceptions; for instance, they may not correlate well with the outcomes of subjective experiments. To gain a comprehensive understanding of how humans perceive multimedia, both subjective and objective evaluation methods should be used in tandem. The study of QoE is crucial as it offers valuable insights into user perception and may uncover behavioral patterns that other metrics cannot predict. Additionally, precise objective metrics expedite the study and evaluation of the effects of processes like compression, transmission, or any operation involving the introduction of noise or artifacts.

In the realm of 2D images and videos, subjective evaluations remain the most accurate quality assessment tool. However, years of research have produced robust objective metrics that are now widely employed.

Conversely, when dealing with point clouds, particularly in the context of volumetric videos, research in this field is still in its early stages, and there are no established or de facto standards in place as of yet.

Objective Metrices:

Objective metrics are computational algorithms or models designed to measure specific perceptual or technical aspects automatically. They aim to evaluate particular characteristics objectively, without the reliance on human assessments.

While the existing metrics for 2D multimedia do not fully capture the intricacies of 3D models, some research has proposed extensions. For instance, in [19], an adaptation of Peak Signal-to-Noise Ratio (PSNR) (Peak Signal-to-Noise Ratio), a common measure used for 2D images and videos, was suggested. Other metrics, such as point-to-point or point-to-plane metrics [19], measure point-wise distortion to derive a quality assessment. However, [1] concluded that the objective metrics available at that time were not consistently able to measure quality as perceived by humans. They also noted that the choice of dataset could significantly impact the QoE results when using V-PCC as a compression algorithm. For example, the voxelization of the 8iVFB dataset resulted in a grid-like disposition of points, while other datasets featured a raw point disposition that was more susceptible to artifacts after the V-PCC process.

Subjective Tests:

Subjective metrics are based on human assessments, involving the collection of user feedback through subjective assessment methods like questionnaires and user ratings. These metrics aim to capture users' opinions, perceptions, and preferences. While there have been numerous experiments on the quality of static point clouds, studies on the quality of dynamic point clouds are relatively limited. This scarcity is particularly noticeable in research conducted within immersive environments, such as virtual reality, where only a few studies are available.

In a study evaluating the quality of compressed volumetric videos under various conditions, it was observed that double stimulus methodologies can lead to evaluations of differences in quality rather than assessments of the quality of the test sequence itself. Consequently, a single stimulus method can offer deeper insights [26].

A typical testing methodology for evaluating dynamic point clouds involves using a single stimulus and employing an absolute category rating (Absolute Category Rating (ACR)) for evaluation. ACR is a quality evaluation approach where participants use a 5-point scale with defined categories such as Excellent, Good, Fair, Poor, and Bad.

Following the guidelines in [14], experiments involving 3D data should adhere to specific conditions: Test sequences should be presented on a mid-gray background to minimize distractions or potential influences on the displayed content. There should be a mid-gray field before and after the test sequence. The gray screen before the sequence should last a maximum of 3 seconds and may contain information, while the voting screen should be displayed for a maximum of 10 seconds. The duration of the test sequence itself should be approximately 10 seconds.

Chapter 3

Subjective Experiment

In this chapter, the experimental procedure is meticulously delineated, and the decision-making process underlying the development of the test is thoroughly demonstrated and justified.

3.1 Dataset - Introduction

In the context of this specific thesis research, I have utilized point cloud datasets curated within the comprehensive BASICS Database [18]. This meticulously curated repository serves as a valuable resource, offering a wide array of high-quality point clouds for various research purposes. Leveraging datasets from the BASICS Database [18] has been instrumental in ensuring the robustness and credibility of the data employed in this thesis work. This scholarly repository has played a pivotal role in enhancing the depth and scope of my research endeavors, providing access to diverse and pertinent point cloud data that aligns seamlessly with the objectives of the thesis.

<u>Rationale for the utilisation of the dataset from the BASICS Database</u>

In the pursuit for the thesis work of point cloud research, particularly within the scope of telepresence applications, the selection of an appropriate dataset is pivotal. The refinement of this research in my Master's thesis is significantly underpinned by the utilization of the BASICS database[18] as said above. This choice is fortified by the database's robust architecture, which systematically addresses the multifaceted challenges and deficiencies evident in the previously existing datasets. The BASICS database thus emerges as an indispensable asset, driving the precision and relevance of the research outcomes in point cloud compression and quality assessment methodologies.

Following are few of the rationale behind the utilisation of this database:

Mitigation of Existing Dataset Deficiencies:

The adoption of the BASICS database[18] reflects a strategic approach to overcome the prevalent deficiencies in existing datasets. These encompass a range of challenges, including limited diversity, inadequate scaling, and inconsistencies such as normalization issues. The BASICS database[18], through its innovative design, eradicates these shortcomings, thereby augmenting the reliability and integrity of the research findings.

Diversity, Scale, and Fundamental Building Blocks:

In the realm of telepresence applications, the selection of a versatile and comprehensive database is crucial for facilitating nuanced and thorough research. The BASICS database[18] emerges as a fundamental asset in this context, showcasing a commendable amalgamation of diversity and scale, meticulously aligned with the core requirements of telepresence paradigms.

This database is enriched with a diversity of data, which is quintessential for cultivating a well-rounded understanding of point cloud compression methodologies. Specifically, it encompasses a curated collection of human subjects, animals, inanimate objects, and architectural landscapes. Such diversity not only fosters a comprehensive exploration of point cloud compression strategies but also nurtures a nuanced appreciation of their performance and applicability across a varied spectrum of subjects and scenarios.

Intricately intertwined with this diversity is the thoughtful assembly of fundamental building blocks within the BASICS database[18]. A meticulous approach has been employed in material selection, ensuring that the database encapsulates essential elements intrinsic to telepresence applications. The dataset is structured around three pivotal categories, namely, humans and animals, inanimate objects, and buildings and landscapes. This strategic categorization underpins the database's robustness, ensuring its profound relevance and applicability in pioneering advancements in the domain of point cloud quality assessment for telepresence applications.

Robust Pre-processing of the BASICS Database:

The authors of the BASICS database[18] meticulously curated and pre-processed the included 3D models to establish a dataset marked by precision and resilience against potential distortions. They embarked on a thorough conversion process, transforming the diverse array of models into point clouds to foster consistency and reliability. Models initially in PLY point cloud format were selectively refined, with a principal focus on voxelization. On the other hand, 3D meshes required a comprehensive suite of modifications. The authors harmonized the format disparities by converting all meshes into a universal Object File (OBJ) format, leveraging tools such as Blender and Meshlab to execute this transformation seamlessly.

In their commitment to dataset integrity, they conducted an exhaustive cleaning process on the 3D meshes. Elements within the meshes exhibiting transparent or reflective attributes, alongside components depicting incomplete reconstructions, were methodically rectified. This rigorous approach was instrumental in obliterating potential distractions and distortions, thereby enhancing the overall user experience. Further solidifying the dataset's robustness, the authors unified the improved meshes into single OBJ files. This strategic move streamlined subsequent sampling processes and facilitated a smoother operational workflow. A critical assessment and fine-tuning of material properties were also undertaken, eliminating inconsistencies, particularly reflective characteristics incapable of authentic replication in point cloud format. CloudCompare was employed in a sophisticated sampling process, extracting point clouds from the surfaces of the meticulously refined 3D meshes. This operation was pivotal in capturing a wealth of attributes, such as location, color, and normals for each constituent point in the point clouds.

Concluding their pre-processing, the authors implemented voxelization on the point clouds, utilizing 10-bit quantization. This nuanced strategy ushered in many benefits, including establishing a conducive range for eficacious point cloud processing and rendering. Additionally, it enhanced stability, predictability, and rendering quality by integrating voxelized coordinates with cube-based rendering techniques.

Utilizing BASICS Database's Advanced Compression Techniques:

In their formidable effort to forge processed point clouds (PPC), the BASICS authors have strategically embraced an array of compression methodologies. Noteworthy among these are the octree-based compression method Moving Picture Experts Group (MPEG) GPCC, the video-based compression method MPEG VPCC, and the learning-based compression method known as GeoCNN. These selections were not arbitrary but were honed based on their pivotal roles and specialized functionalities in the MPEG standardization universe. Their summarised explanations go as below:

MPEG GPCC[8], a geometry-based point cloud compression method, is characterized by its proficiency in encoding point clouds directly within the 3D space. Its methodology encapsulates advanced techniques such as octree or trisoup methods, and it exhibits a remarkable capability in encoding attributes, such as color, by deploying state-of-the-art methods like Random Access and Hierarchical Tree Structure (RAHT) and PredLift transforms. In this database only octree method was used due to trisoup generated uneven structure or holes on reconstructed point clouds.

The Video-based Point Cloud Compression (VPCC) technique [8] represents an innovative approach for point cloud compression. In this method, the point cloud is projected onto the facets of a cube. These projections, encompassing both color and depth information (which conveys the intrinsic three-dimensional structure of the points), are then encoded using established video compression techniques, notably High Eficiency Video Coding (HEVC)/H.265 or VVC/H.266. A key advantage of VPCC lies in its ability to leverage the innate temporal coding capabilities of these traditional video codecs, thereby enabling eficient compression of dynamic point cloud sequences.

GeoCNN[3], leads the charge in compressing voxelized point clouds by first performing block partitioning. It navigates from block partitioning to transformation through a variational autoencoder, transitioning from binary occupancy voxel grids to latent spaces. This intricate process, infused with learned entropy models, specialized quantization, and entropy coding techniques, culminates in the creation of decoded blocks and the holistic decompression of point clouds.

The decision to employ the BASICS database in my Master's thesis is underscored by its exemplary alignment with the research objectives, particularly in the spheres of point cloud compression and quality assessment for telepresence applications. The database's sophisticated architecture, characterized by extensive diversity, scale, and meticulous normalization processes, empowers the finding with a solid foundation for achieving innovative and reliable outcomes. Through this strategic selection, the thesis is poised to contribute meaningfully to the advancement of knowledge and practices within the evolving field of point cloud research for telepresence applications.

3.2 Dataset Utilised

Table 3.1: Dataset for Testing

Source Point CLoud (PC)	GPCC-Octree-Predlift	GPCC-Octree-RAHT	VPCC
p03.ply (Humans)	p03_gpcc-oct-predlift_r02.ply	p03_gpcc-octree-raht_r02.ply	p03_geocnn_r01.ply
Continued on next page			

Table 3.1 – continued from previous page

Source PC	GPCC-Octree-Predlift	GPCC-Octree-RAHT	VPCC
	p03_gpcc-oct-predlift_r03.ply	p03_gpcc-octree-raht_r03.ply	p03_geocnn_r02.ply
	p03_gpcc-oct-predlift_r05.ply	p03_gpcc-octree-raht_r05.ply	p03_geocnn_r03.ply
	p03_gpcc-oct-predlift_r06.ply	p03_gpcc-octree-raht_r06.ply	p03_geocnn_r04.ply
p21.ply (Elephant)	p21_gpcc-oct-predlift_r02.ply	p21_gpcc-octree-raht_r02.ply	p21_geocnn_r01.ply
	p21_gpcc-oct-predlift_r03.ply	p21_gpcc-octree-raht_r03.ply	p21_geocnn_r02.ply
	p21_gpcc-oct-predlift_r05.ply	p21_gpcc-octree-raht_r05.ply	p21_geocnn_r03.ply
	p21_gpcc-oct-predlift_r06.ply	p21_gpcc-octree-raht_r06.ply	p21_geocnn_r04.ply
p40.ply (Monument)	p40_gpcc-oct-predlift_r02.ply	p40_gpcc-octree-raht_r02.ply	p40_geocnn_r01.ply
	p40_gpcc-oct-predlift_r03.ply	p40_gpcc-octree-raht_r03.ply	p40_geocnn_r02.ply
	p40_gpcc-oct-predlift_r05.ply	p40_gpcc-octree-raht_r05.ply	p40_geocnn_r03.ply
	p40_gpcc-oct-predlift_r06.ply	p40_gpcc-octree-raht_r06.ply	p40_geocnn_r04.ply
p44.ply (Trumpet)	p44_gpcc-oct-predlift_r02.ply	p44_gpcc-octree-raht_r02.ply	p44_geocnn_r01.ply
	p44_gpcc-oct-predlift_r03.ply	p44_gpcc-octree-raht_r03.ply	p44_geocnn_r02.ply
	p44_gpcc-oct-predlift_r05.ply	p44_gpcc-octree-raht_r05.ply	p44_geocnn_r03.ply

Continued on next page

Table 3.1 – continued from previous page

Source PC	GPCC-Octree-Predlift	GPCC-Octree-RAHT	VPCC
	p44_gpcc-oct-predlift_r06.ply	p44_gpcc-octree-raht_r06.ply	p44_geocnn_r04.ply
p56.ply (Hut)	p56_gpcc-oct-predlift_r02.ply p56_gpcc-oct-predlift_r03.ply p56_gpcc-oct-predlift_r05.ply p56_gpcc-oct-predlift_r06.ply	p56_gpcc-octree-raht_r02.ply p56_gpcc-octree-raht_r03.ply p56_gpcc-octree-raht_r05.ply p56_gpcc-octree-raht_r06.ply	p56_geocnn_r01.ply p56_geocnn_r02.ply p56_geocnn_r03.ply p56_geocnn_r04.ply

Table 3.2: Dataset for Training

Source PC	GPCC-Octree-Predlift	GPCC-Octree-RAHT	VPCC
p02.ply (Humans)			
		p42_gpcc-octree-raht_r02.ply p42_gpcc-octree-raht_r03.ply p42_gpcc-octree-raht_r04.ply	

3.3 Compression

In the exploration of point cloud compression methodologies, the research work from which my thesis draws inspiration leverages a spectrum of sophisticated techniques, each contributing uniquely to the enhancement of point cloud processing. The authors have meticulously prepared the processed point clouds (PPC) using an array of carefully selected compression methods, thereby cultivating a multifaceted approach to point cloud compression.

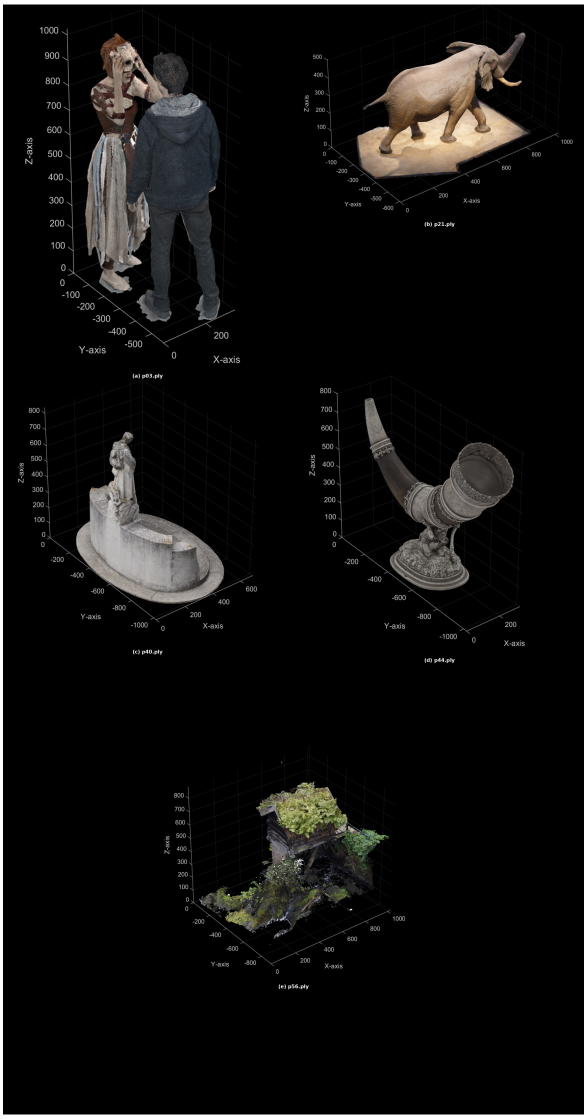

Figure 3.1: Dataset for Testing

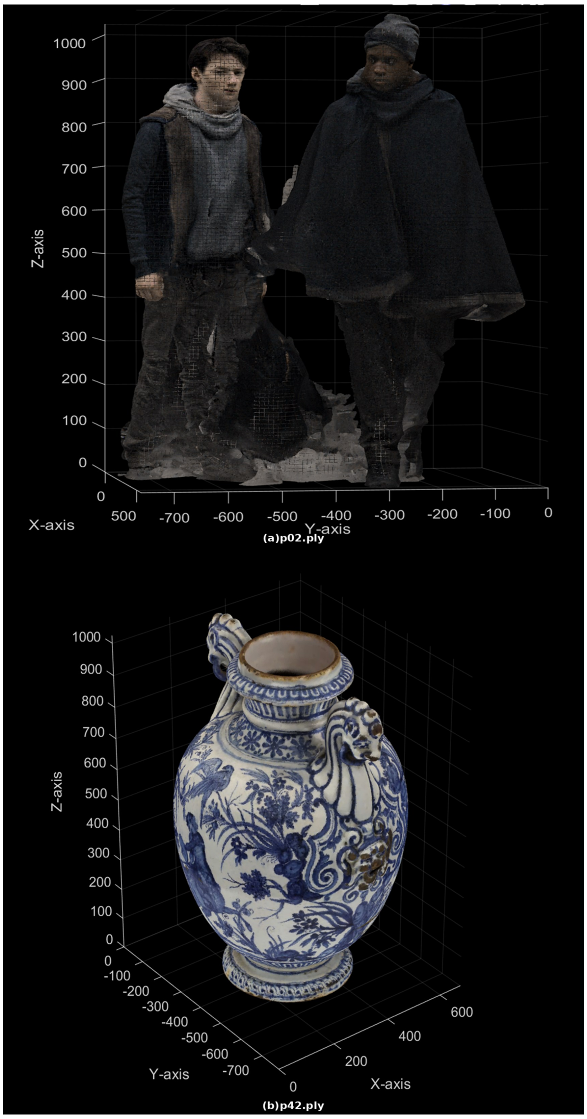

Figure 3.2: Dataset for Training

The ingenuity of the chosen methods unfolds through four distinctive compression techniques, namely GeoCNN[3], MPEG Geometry-based Point Cloud Compression (GPCC) (GPCC-Octree-RAHT, and GPCC-Octree-Predlift)[8], as well as the MPEG Video-based Point Cloud Compression (VPCC)[8]. Each method brings forth a nuanced methodology and a technical advantage aimed at achieving optimal compression of the point clouds.

The GPCC[8] method stands out with its concentrated focus on encoding the point clouds directly within the 3D space. It employs techniques such as octree or trisoup (triangle soup) methods. This method is further diversified into two sophisticated types: GPCC-Octree-Predlift and GPCC-Octree-RAHT, each manifesting unique attributes and methodologies.In a strategic alignment with dynamic point clouds, the MPEG VPCC method carves its niche. Its approach is characterized by a vivid projection of point cloud content onto depth and texture maps. Utilizing state-of-the-art video encoders such as HEVC, VPCC[8] navigates the complexities of compressing point clouds that exhibit temporal changes, enhancing the adaptability and relevance of the compression processes.GeoCNN[3], a learning-based algorithm, enhances the compression landscape with its approach. It navigates the compression pathway with a sequence of processes involving block partitioning and variational autoencoders, culminating in transformation and reconstruction of the point clouds.

Furthermore, it's imperative to underscore that each of these formidable compression methods is manifested across four compression levels(as illustrated in the chart above), illustrating a comprehensive and layered approach to point cloud compression. This multifaceted methodology underscores the robustness and adaptability of the compression strategies employed, each finely tuned to meet the intricate demands of point cloud processing and enhancement in telepresence applications.

3.4 Mouse Coordinates

In the aim of enhancing our understanding and proficiency in handling point clouds, this thesis aims to delve into the intricate interactions between users and point cloud representations. One of the pivotal objectives of this endeavor is to meticulously record the mouse movement coordinates, specifically the x and y coordinates, during user interaction with point clouds within a MATLAB environment. This endeavor is instrumental in unraveling the nuances of user engagement, navigation, and interaction dynamics, providing profound insights that can be pivotal for optimizing point cloud representations and interfaces for enhanced usability and user experience.

For this purpose, a curated selection of point clouds has been utilized, namely P03.ply, P21.ply, P40.ply, p44.ply, and p56.ply as stated above in the chart taken from the BASICS Database[18]. These point clouds are not only instrumental due to their intrinsic characteristics but also due to their processed variations, which have been subjected to an array of sophisticated compression techniques. Each point cloud has been meticulously compressed using four distinctive compression methods, namely GeoCNN[18], MPEG Geometry-based Point Cloud Compression (GPCC)[8] with subtypes GPCC-Octree-RAHT and GPCC-Octree-Predlift, and MPEG Video-based Point Cloud Compression (VPCC)[8]. These compression methods, each with their unique approaches and algorithms, impart a spectrum of qualities and characteristics to the point clouds, thus enriching the landscape of interaction scenarios and possibilities. Furthermore, each compression method has been applied at four distinct levels of compression, thereby creating a multifaceted array of point clouds, each with varied characteristics, complexities, and details. This structured approach allows for a comprehensive exploration of the impact of compression on the interaction dynamics and user engagement with the point clouds, thus broadening the horizons of understanding and analysis.

In the MATLAB environment, where the interaction takes place, the systematic recording of mouse movement coordinates is meticulously executed and diligently stored in an Excel file for further usage. This process of capturing the coordinates is paramount to understanding the user navigation pathways, interaction preferences, and the overall engagement blueprint. It opens doors to insightful data that is pivotal for analyzing and optimizing the interfaces and representation formats of point clouds for heightened usability and enriched user experiences.

The data gathered from the mouse movements, in coordination with the specified point clouds and compression methodologies, forms a robust foundation for in-depth analysis. This allows for the exploration of patterns, tendencies, and preferences in user interactions, and how these dynamics are influenced by the various compression methods and levels. The overarching aim is to cultivate a nuanced understanding that could be instrumental in steering future developments, enhancements, and optimizations in point cloud interactions for telepresence applications and beyond.

3.5 Equipment and Environment

The experimental assessment of 3D point cloud quality was conducted within a controlled setting to ensure the reliability and validity of the Absolute Category Rating (ACR)[23] process. This section delineates the hardware, software, and environmental conditions underpinning the experiment.

Hardware

The central hardware component employed in the experiment was a Lenovo ThinkPad X1 Carbon 7th Generation. This system features an Intel(R) Core(TM) i7-8565U CPU with a base clock speed of 1.80GHz and turbo boost capabilities up to 4.60 GHz. With four cores and eight logical processors, the laptop is well-suited for handling the intensive computational demands of 3D point cloud manipulation and rendering. The system operates on a 64-bit architecture, an essential specification for leveraging the capabilities of the MATLAB environment and Computer Vision Toolbox utilized in the study.

System stability and reliability are assured by the laptop's UEFI Basic Input/Output System (BIOS), version N2HET44W (1.27), dated January 15, 2020. The BIOS, in conjunction with the System Management BIOS (SMBIOS) version 3.1, ensures optimal hardware-software integration. An embedded controller version 1.14 further maintains the integrity of system processes during the experimental operations.

Software

The experimental procedure was facilitated through the MATLAB environment, a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numerical computation. Specifically, the MATLAB version (R2023b) in use adhered to the software's licensing agreements at the time of the experiment, with the exact version number R2023b, it's expected to be contemporaneous to the hardware's last BIOS update.

The Computer Vision Toolbox was utilized for the presentation and manipulation of 3D point cloud images. This toolbox provides algorithms, functions, and apps for designing and testing computer vision, 3D vision, and video processing systems. The exact toolbox version is similarly aligned with the MATLAB environment to ensure compatibility and function reliability.

Experimental Environment

The experimental environment was standardized to mitigate variables that could potentially influence the subjective assessment of 3D point cloud quality. Before initiating the ACR process, the Lenovo ThinkPad X1 Carbon 7th Generation's display settings were carefully calibrated to ensure accurate color representation and consistency throughout the study. The laptop's integrated Wide Viewing Angle and High-Density FlexView Display offered a desktop resolution of 2560x1440, ensuring that fine details within the 3D point clouds were clearly visible to participants. The display operated at a 60 Hz refresh rate to maintain a smooth visual flow and was set to an 8-bit color depth in the Red Green Blue (RGB) color format within the Standard Dynamic Range (SDR), to present the point clouds with true-to-life colors and contrasts. This setup, powered by the Intel(R) UHD Graphics 620 adapter, was crucial in providing a consistent and high-fidelity viewing experience, thereby supporting the integrity of the subjective evaluations. The seating arrangements and viewing distance were also standardized based on ergonomic principles to provide a comfortable and consistent viewing experience for all participants, thereby limiting physical strain that could influence subjective quality judgments.

3.6 Participants

In the study, a group of 32 university students participated, comprising 14 women and 18 men, showcasing a diverse international representation. These individuals hailed from a variety of countries, enriching the study with their distinct cultural and academic backgrounds. Predominantly, these participants were from scientific fields, although a notable minority came from non-scientific disciplines. Prior to the commencement of the tests, all participants were duly informed about the nature and purpose of the study. Explicit consent was obtained both for participating in the tests and for the subsequent use of their data in the research analysis. This careful consideration of consent and diverse participant makeup ensured an ethical and inclusive approach to the study.

3.7 Subjective Experiment Methodology

Subjective quality assessment of point cloud content is generally divided into interactive and passive methodologies[16]. The interactive method grants observers the autonomy to examine the point cloud from any desired perspective, often utilizing tools within augmented or virtual reality environments, thus offering a more immersive evaluation experience. Contrarily, the passive technique involves viewers observing the point cloud via a fixed camera path, akin to watching a conventional video. Although both methods present unique pros and cons, research indicates that the differences in subjective opinions elicited by these approaches are not statistically much different[6]. However, for the purposes of our study, we have chosen to employ the interactive paradigm. This decision is predicated on our objective to closely mimic real-world usage of point cloud content, thereby enhancing the validity of our findings and facilitating a more engaging and natural interaction for participants within our experimental framework.

Methodology:

Several methodologies can be found in the literature and recommendations for subjective quality assessment of traditional image and video sequences [15].The selection of the Absolute Category Rating (ACR) method for the subjective quality assessment of point cloud degradation in our research is grounded in several critical considerations informed by current literature and practical constraints. ACR is distinguished by its straightforward approach where each stimulus is evaluated independently without a direct comparison to a reference or an impaired sample. This singularity in presentation inherently simplifies the task for participants, facilitating quicker and potentially less biased responses. Mantiuk et al.[24] have highlighted the comparative effectiveness of methodologies like Pairwise Comparison (PC) for image and video content due to its simple experimental procedure; however, they acknowledge that there is no significant disparity in accuracy between ACR and Double Stimulus Impairment Scale (DSIS) methods [10]. While PC offers accuracy, it becomes impractical with the escalation in the number of test conditions due to the exponential increase in required comparisons [10]. While Nehme et al.[27] suggest that DSIS[17] might yield more accurate discrimination in 3D content evaluations [27], such accuracy must be weighed against practicality and the scope of the study. The ACR method[15], despite its noted limitations in cases where participants may not be familiar with the pristine models, offers a viable balance between eficiency and reliability for a study with other conditions to be tested. In the case of our research, involving five point clouds each with 15 degradations, the use of DSIS[24] could result in participant fatigue and extended duration of the assessment process due to the sheer volume of side-by-side comparisons required.

Furthermore, the ACR method[15] aligns with our resource constraints and our objective to conduct a evaluation across multiple point clouds and degradations. It allows for a streamlined and scalable approach suitable for our in-house environment, where ease of understanding and brevity of task are crucial. Moreover, in adopting the ACR method[15], we are also able to circumvent the cognitive load and complexity introduced by side-by-side comparative evaluations, thus reducing the risk of inconsistent ratings due to fatigue or confusion. By employing ACR[15], we aim to gather a general consensus on the perceived quality of each degraded point cloud, which is anticipated to offer substantial insights into the threshold levels of degradation detectable by typical observers.

Test Procedure:

In the subjective evaluation study, we conducted a comprehensive assessment of three-dimensional (3D) point cloud visualizations to gather user feedback on their quality. The procedure for this study was carefully designed and executed in several phases as follows:

1. Briefing Phase:

Initially, participants were briefed about the nature of 3D point cloud images. Point clouds represent 3D spaces by using large numbers of points, which can create detailed and immersive visual representations. It was explained to them that these point clouds would be displayed on the screen one at a time, each for a duration of 10 seconds. Before interacting with the Point Cloud the participants name will be registered using the following dialog box:

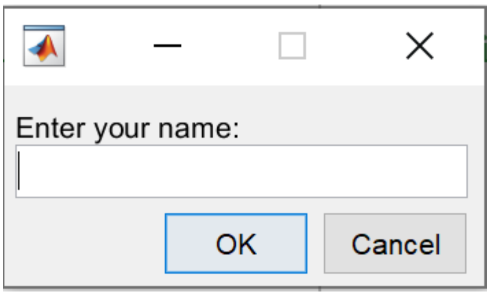

Figure 3.3: User Name input Dialog Box

Following the viewing and interacting of each point cloud, the participants were instructed to rate their visual experience on a scale from 1 to 5, where 1 represented the worst quality and 5 represented the best quality. The classification goes as : 1 (Poor/Lowest), 2 (Fair/Below Average),3 (Average/Moderate),4 (Good/Above Average),5 (Excellent/Highest).

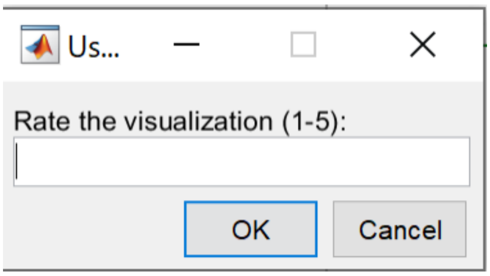

Figure 3.4: User Rating from 1 to 5

Prior to commencing the test, I requested participants to provide their names and obtained their consent for the experiment, ensuring ethical research practices.

2. Training Phase

To familiarize the participants with the assessment process and to set a baseline for their expectations, it was introduced a training session of few point clouds using a set of training datasets recruited from the BASICS database [18].

The datasets included:

- ‘p02.ply‘
- ‘p42_gpcc-octree-raht_r02.ply‘
- ‘p42_gpcc-octree-raht_r03.ply‘
- ‘p42_gpcc-octree-raht_r04.ply‘

During this phase, the participants were allowed to view and interact with these point clouds to understand the type of content they would be evaluating. This step ensured that the participants were comfortable with the evaluation process and reduced the potential for variance due to unfamiliarity with the format of the point clouds.

3. Testing Phase

In the main testing phase, the actual evaluation dataset was introduced. It was implemented a MATLAB script to display each ‘.ply‘ file from the dataset in a controlled environment. Participants interacted with the visualizations using the mouse, and the MATLAB program recorded the mouse coordinates over a 10-second interval to capture their interaction patterns. This interactivity was crucial as it provided a more engaging experience and allowed for a better assessment of the visual quality.

The script was designed to automate the transition between point clouds, display each for the specified interval, and prompt for user ratings immediately following each viewing. In total, the participants evaluated 5 different 3D images with 15 compression and degradation levels each resulting in 85 images as total . After viewing each image, a dialog box was presented where they entered their rating. These ratings were then automatically recorded and associated with the corresponding point cloud file name for later analysis.

To ensure accuracy and organization, each participant’s ratings and mouse coordinates were saved into separate Excel sheets, labeled with their provided username. This organization of data facilitated a smooth and eficient process for both participants and analyser, resulting in a comprehensive dataset ready for further statistical analysis.

The data collected through this methodological approach provided invaluable insights into the subjective perception of point cloud quality and allowed for a detailed analysis of user experience and interaction with 3D point cloud visualizations.

Chapter 4

Experimental Results and Analysis

4.1 Introduction

In this chapter, we delve into the comprehensive analysis of the user interactions within the context of visualizing 3D Polygon File Format (PLY) files using MATLAB. The primary objective of this chapter is to systematically assess and quantify user behavior during a 10-second interaction period with the visualized PLY files. The analyzed parameters encompass four critical aspects: the range of X and Y-axis movement, total distance traveled by the user's cursor, inactive time (periods of mouse inactivity), and the mean angle of movement which corresponds to the mean direction. The details of the framework which examines this four key parameters are written below:

1. Range of X and Y-Axis Movement: This parameter measures the extent of horizontal and vertical cursor movements, providing insights into the spatial dynamics of user interaction with the interface.

2. Total Distance Traveled by the User's Cursor: By quantifying the cumulative distance covered by the cursor, this metric offers an aggregate view of user activity, indicating the level of engagement and the extent of exploration within the digital environment.

3. Inactive Time (Periods of Mouse Inactivity): This factor captures the durations of non-interaction, reflecting pauses in user activity. It serves as an indicator of user attention span on the 3d file visualised.

4. Mean Angle of Movement: Representing the average direction of cursor movement, this parameter reveals dominant interaction patterns, shedding light on user navigational preferences and tendencies within the application.

Each of these parameters is meticulously computed using a dedicated MATLAB code, ensuring precision and reliability in data collection. The results are systematically organized and stored in separate Excel files for each parameter.

In addition to the aforementioned parameters, a separate Excel sheet was meticulously created to capture the Mean Opinion Score (MOS) values assigned by each user to individual 3D files. The MOS, reflects subjective user evaluations, providing a direct measure of user perception and satisfaction.

This dedicated MOS dataset plays an important role in our comprehensive analysis. It is carefully mapped against the other Excel files, where the parameters of X and Y-axis movement range, total cursor distance traveled, inactive time, and mean angle of movement are stored. This strategic alignment of datasets enables a multi-dimensional analysis, allowing for the exploration of potential correlations between objective interaction metrics and subjective user ratings.

By integrating the MOS values with the quantifiable interaction data, the approach facilitates a holistic understanding of user engagement. It allows for the identification of patterns and trends linking user behavior (as captured by the interaction metrics) with user satisfaction (as indicated by MOS). This alignment is instrumental in drawing meaningful insights, enabling us to decipher how different aspects of user interaction influence their overall experience and satisfaction.

This structured approach not only facilitates detailed analysis but also enables easy comparison and correlation of different interaction metrics. The data thus amassed forms a robust foundation for analysis, providing valuable insights into user behavior.

4.2 Outliers removal

The incorporation of a 95% Confidence Interval (Confindence Interval (CI)) for outlier detection in the scatter plots is a crucial aspect of the analysis, serving to enhance the reliability of the interpretations drawn from the data.

Importance of 95% CI in Outlier Detection:

1. Statistical Significance: The 95% CI provides a statistically significant range for each data point. It indicates that, under the assumption of normal distribution, there's a 95% probability that the true value lies within this interval. This is critical for understanding the precision of the data and the extent to which it can be trusted.

2. Outlier Identification: By visualizing the 95% CI, outliers can be effectively identified. An outlier is a data point that significantly deviates from others in a dataset. If a data point

lies outside the 95% CI range, it can be considered an outlier. This helps in assessing the homogeneity of the data and ensuring that the analysis is not skewed by anomalous values.

3. Data Quality Assessment: The CI error bars also serve as a tool for assessing the quality of the data. Large error bars might indicate high variability in the data or a small sample size, prompting further investigation or data collection.

In summary, the use of a 95% CI for outlier detection and removal in the scatter plots provides a clear visual representation of the data's reliability and helps in making informed decisions about the inclusion or exclusion of certain data points in the analysis.

4.3 Graph Utilised

Here's a general breakdown applicable to the graphs depicting Mean Opinion Score (MOS) versus various interaction parameters:

X-Axis Representation: In each graph set, the x-axis represents a specific user interaction parameter: 1. Range of Movement (units) – indicating the extent of cursor movement in a speCIfic direction. 2. Total Distance Traveled by Cursor (units) – reflecting the cumulative distance the cursor has covered. 3. Periods of Mouse Inactivity (time units) – showing the duration of times when there was no cursor movement. 4. Mean Angle of Cursor Movement (degrees/radians) – representing the average direction of cursor movement.

Y-Axis Representation: The y-axis consistently represents the 'Mean Opinion Score (MOS)' across all graph sets. MOS is a metric used to evaluate the quality of user experience post-compression. A higher MOS denotes a superior perceived quality.

Scatter Plot Points: Each point on the scatter plots corresponds to a specific observation or test instance. Its position is determined by the value of the user interaction parameter on the x-axis and the corresponding MOS on the y-axis.

Error Bars: The error bars on each point represent the 95% Confidence Interval (CI) for the MOS. This indicates the range within which the true mean of the MOS is likely to fall with 95% confidence.

Series Comparison: Different series in the graphs compare the Source (Source (SRC)) or the original uncompressed data against different types of compression techniques at various levels.

Correlation Coeficient and P-Value: The graphs include the correlation coeficient (r) and p-value (p) in the legend for each series. The correlation coeficient measures the strength and direction of the linear relationship between the MOS and the interaction parameter. A value of r close to 1 or -1 implies a strong linear relationship, while a value near 0 suggests little to no linear relationship. The p-value is used to assess the statistical significance of the correlation. A low p-value (typically less then 0.05) implies strong evidence against the null hypothesis, suggesting a significant relationship between the phenomena.

This format provides a comprehensive view of the relationships between various user interaction parameters and the perceived quality of 3D models, as reflected in the Mean Opinion Score.

4.4 Correlation Analysis

In this section, correlation analysis is conducted between the Mean Opinion Score (MOS) and a set of key user interaction parameters. These parameters include the range of movement along the X and Y axes, the total distance traversed by the user's cursor, periods of inactivity as indicated by mouse inactivity, and the mean angle of cursor movement, which represents the predominant direction of user interaction. This correlation analysis is imperative in understanding the intricate relationships between subjective user ratings (MOS) and objective measures of user engagement. For this analysis a two-level analysis process is outlined, focused on evaluating the performance and effects of different compression techniques on 3D files. The expanded explanation of this two-level analysis is given below and the names of the two level will be used in further sections.

Intra-Compression Technique Analysis:

At this level, the primary focus is on individual compression techniques (GeoCNN, GPCC-Predlift, GPCC-Raht, and Video BAsed Point Cloud Compression (VPCC)) applied to 3D models (p03, p21, p40, p44, p56).For each compression technique, a series of graphs is generated that illustrate the trend or relationship between four distinct levels of compression and the original source (SRC) file.This analysis aims to understand how varying degrees of compression within a single method affect the 3D model's quality, user experience, or other metrics being measured. Then the analysis proceeds by examining whether the observed trend in intra-compression of the specified type is consistent across other PLY models.

Inter-Compression Technique Comparison:

This level of analysis involves comparing the trends observed across the different compression techniques using the same set of graphs.The objective here is to discern how each compression method (GeoCNN, Geometry Based Point Cloud Compression (GPCC)-Predlift, GPCC-Raht, VPCC) performs relative to the others in terms of the same metrics.

4.4.1 Range of Movement of Cursor in both X and Y axis

To conduct a comprehensive correlation analysis of range of cursor movement (spanning both the x and y axes), graphs were plotted with the Mean Opinion Score (MOS) against the range of cursor movement. This analysis is applied individually and seperately to the x and y axes for enhanced clarity and precision. The data set encompasses a variety of 3D files, specifically models p03, p21, p40, p44, and p56 as mentioned in earlier chapters. These models have been subjected to various state-of-the-art compression techniques, namely GeoCNN, GPCC-Predlift, GPCC-Raht, and VPCC. For each 3D file, we have produced a series of graphs that intricately showcase the impact of each compression technique at four distinct levels of compression.

Range of the movement of cursor of X-axis(horizonataly):

Figure 4.1: Scatter Plot of MOS vs Range Of Movement (X-axis) of p03,p21,p40,p44,p56 (each row)

Range of the movement of cursor of Y-axis(verticaly):

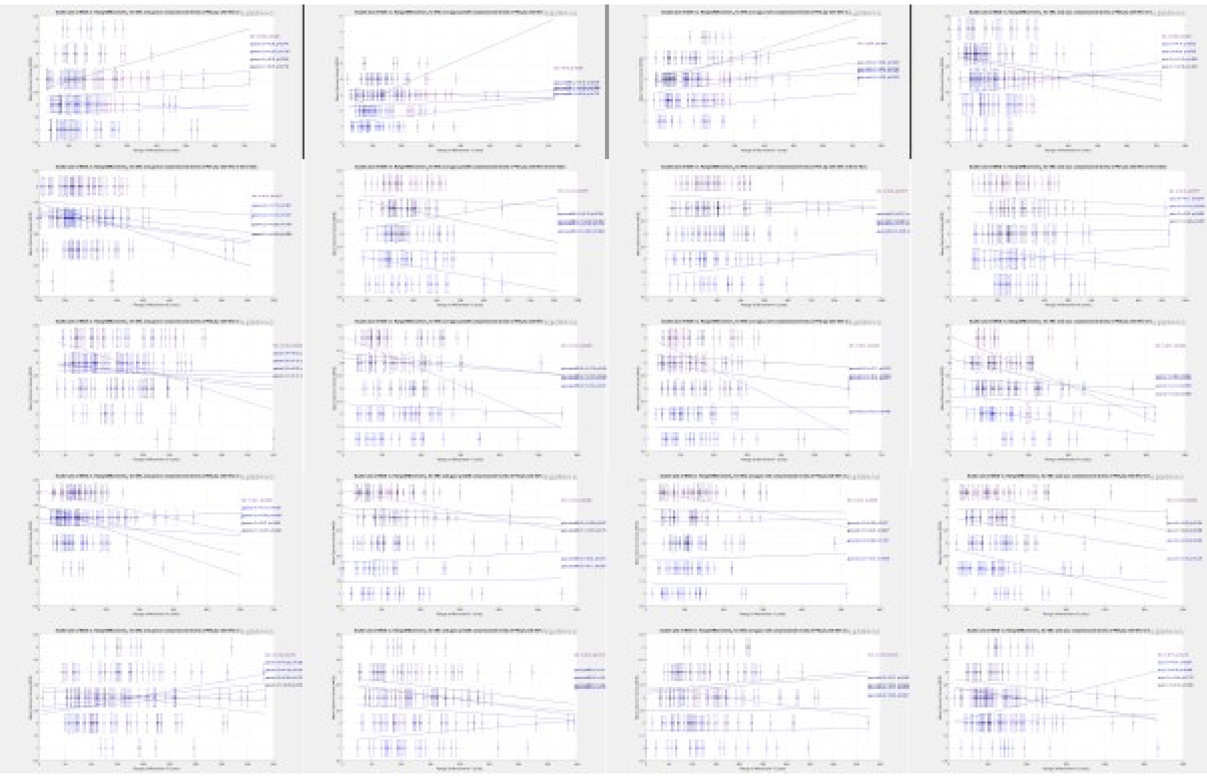

Figure 4.2: Scatter Plot of MOS vs Range Of Movement (Y-axis) of p03,p21,p40,p44,p56 (each row)

Inter-Compression Technique Comparison:

1.GeoCNN Compression

Visually, we can observe by combining the analysis of both the X and Y-axis scatter plots for the GeoCNN compression technique across different PLY models the following:

Correlation Trends: The correlation coeficients (r-values) for both X and Y-axis movements are mostly close to zero for all levels of compression across the different PLY models, suggesting a generally weak linear relationship with the Mean Opinion Score (MOS).

Statistical Significance: The majority of p-values across both axes are above the 0.05 threshold, indicating that the correlations are not statistically significant. This suggests that there is no strong evidence to support a consistent impact of the GeoCNN compression levels on the range of cursor movement in relation to the perceived quality.

Exceptions: Notable exceptions where a significant correlation is observed include: - For the X-axis, P56 shows a significant correlation at the L4 compression level. - For the Y-axis, P40 and P44 demonstrate significant correlations for the SRC and L3 compression levels, respectively.

Overall Trend: Across both the X and Y-axis movements, the data does not show a strong or consistent trend across the different PLY models. This suggests that the GeoCNN compression technique does not have a uniform impact on the MOS related to the cursor movement range.

Significance of Exceptions: The exceptions observed are important as they indicate instances where the GeoCNN compression technique does impact the perceived quality of cursor movement. These exceptions, such as the significant negative correlation for SRC data in P40 on the Y-axis and the significant correlation for L4 compression in P56 on the X-axis, suggest that there may be model-specific factors or certain levels of compression that affect user perception more noticeably. These exceptions warrant further investigation to understand the underlying factors that contribute to these significant correlations.

2.GPCC-Predlift Compression

To provide a combined analysis of the observations for the GPCC-predlift compression technique on both X and Y axes from the provided data, here is the summary based on the scatter plots:

Correlation Trends: The correlation coeficient (r) values for GPCC-predlift compression levels across various participants are generally low, indicating a weak linear relationship between MOS scores and Range of Movement (both X and Y axes).

Statistical Significance: Most p-values are above the typical alpha level of 0.05, indicating that the correlations are not statistically significant.

Exceptions: A notable exception is the stronger negative correlation observed for P03 on the X-axis at level L2, suggesting a possible trend where higher compression may correlate with lower MOS scores. Here the p-value is below 0.05, suggesting a statistically significant relationship. Certain data points, such as P44 on the X-axis, show a moderate negative correlation with statistical significance, which stands out against the overall weak correlation trend. - These exceptions suggest that there may be participant-specific factors or content-specific features that affect the perceived quality and its relationship with compression, which is not consistent across all samples.

Overall Trend: The overall trend suggests that for the GPCC-predlift compression technique, there is no consistent strong relationship between the compression levels and the perceived quality across different participants.

Significance of Exceptions: - The exceptions where a statistically significant correlation is observed suggest that under certain conditions like participant-specific factors or content-specific features may affect the perceived quality and its relationship with compression.

3.GPCC-Raht Compression

We can observe by combining the analysis of both the X and Y-axis scatter plots for the GPCC-Raht compression technique across different PLY models the following

Correlation Trends: For the X-axis movement, correlation coeficients are consistently low across all subjects (P03, P21, P40, P44, P56) with no clear trend suggesting a relationship between MOS and the range of movement. For the Y-axis movement, only one instance shows a moderate correlation, while the rest remain low, similar to the X-axis findings.

Statistical Significance: Almost all correlation coeficients are not statistically significant with p-values exceeding the typical alpha level of 0.05. The statistical significance is generally absent across both axes and all compression levels.

Exceptions: A notable exception is found in subject P03 for the Y-axis movement at compression level L2, where the correlation is moderate and statistically significant ($r = 0.50$, $p = 0.003$).

Overall Trend: The overarching trend suggests no significant relationship between the range of movement and the MOS for the ‘GPCC-raht‘ compression method across the majority of the dataset.

Significance of Exceptions: The exception for P03 on the Y-axis at L2 compression level stands out against the overall trend. It suggests that for this specific subject and axis, the range of movement might have a more pronounced effect on the MOS. This exception could be due to unique characteristics of the subject's data or the specific nature of the compression at L2 for the Y-axis movement, warranting further investigation to understand the discrepancy from the overall trend.

4.VPCC Compression

The scatter plots for the vpcc compression levels in relation to the Mean Opinion Score (MOS) against the Range of Movement in both X and Y axes for different P values have been plotted above. Following points can be observed:

Correlation Trends: The correlation between MOS and the Range of Movement varies across different compression levels (L1-L4) and the SRC group. Some show a slight positive trend, others negative, and some display no clear trend at all.

Statistical Significance: The p-values indicate the statistical significance of the correlation coeficients (r). Most of the p-values are above the typical alpha level of 0.05, suggesting that many of the observed correlations are not statistically significant.

Exceptions: There are exceptions where some compression levels show statistically significant correlations. For example, a few plots show a p-value less than 0.05 for certain compression levels, indicating a statistically significant correlation in those instances.

Overall Trend: The overall trend is not consistent across all plots. While some plots indicate a slight negative correlation between MOS and the Range of Movement, others suggest a slight positive correlation or no correlation at all.

Significance of Exceptions: The significance of the exceptions where p-values are below 0.05 suggests that there may be specific instances where the compression level has a more pronounced impact on the perceived quality (as expressed by MOS) related to the range of movement. However, given the variability across different plots and P values, these exceptions do not establish a clear general pattern.

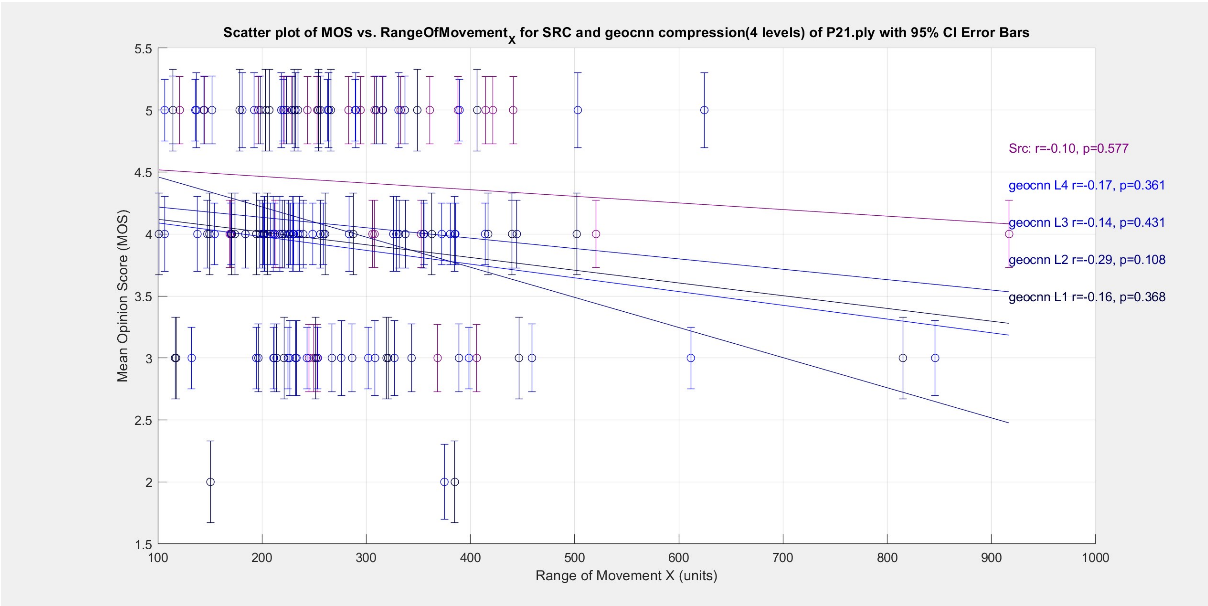

Figure 4.3: For Demonstration: Pearson coeficient shows that it is very weak-negatively correlated and the p-value shows it's not statistically significant as it's value is quite more then 0.05 threshold

Inter-Compression Technique Comparison:

When comparing the GeoCNN, GPCC-Predlift, GPCC-Raht, and VPCC compression techniques, the collective data presents a landscape where correlations between the Mean Opinion Score (MOS) and the range of cursor movements are generally weak across both X and Y axes.

Correlation Trends: For all compression techniques, r-values are predominantly near zero, suggesting a negligible linear relationship with MOS. This indicates that, irrespective of the compression method applied, there is no strong inherent connection between the extent of cursor movement and the perceived quality of the models.

Statistical Significance: In the cross-technique picture, p-values often exceed the 0.05

significance threshold, hinting that the observed correlations are unlikely to reflect true effects within the data. This holds across different PLY models and compression levels, underscoring a widespread lack of statistical significance in the impact of compression techniques on user perception.

Exceptions: However, exceptions to this trend exist and are critical to the analysis. For instance, the GeoCNN technique displayed significant correlations in specific cases, such as P56 on the X-axis and P40 on the Y-axis. GPCC-Predlift also had its standout moments, notably with P03 on the X-axis at the L2 level. GPCC-Raht's exception came from P03 on the Y-axis at L2. These notable deviations highlight instances where compression levels do indeed align with changes in perceived quality, albeit in a non-uniform and participant-specific manner.

Overall Trend: The overarching pattern suggests that the relationship between compression levels and perceived quality is not only weak but also inconsistent. No single compression technique emerges as markedly superior or inferior in maintaining quality perception across the range of cursor movements.

Significance of Exceptions: These anomalies may signify underlying variables that influence perception differently in certain contexts or subjects. For instance, the significant negative correlations for specific PLY models at certain compression levels may point towards a more complex interplay of compression artifacts, model complexity, and subjective quality assessment criteria used by participants.

4.4.2 Total Distance travelled by the cursor

To conduct a thorough analysis of the correlation between the perceived quality of 3D models and the total distance travelled by the cursor, we plotted the Mean Opinion Score (MOS) against this distance. The data set encompasses a variety of 3D files, specifically models p03, p21, p40, p44, and p56 as mentioned in earlier chapters. These models have been subjected to various state-of-the-art compression techniques, namely GeoCNN, GPCC-Predlift, GPCC-Raht, and VPCC. For each 3D file, we have produced a series of graphs that intricately showcase the impact of each compression technique at four distinct levels of compression.

The set of Graphs demonstrates Total Distance Travelled by the Cursor:

Intra Compression Technique Analysis

1. GeoCNN Compression

Analysis of GeoCNN Compression individually and across PLY.

Correlation Analysis: The correlation coeficients across the five GeoCNN graphs are

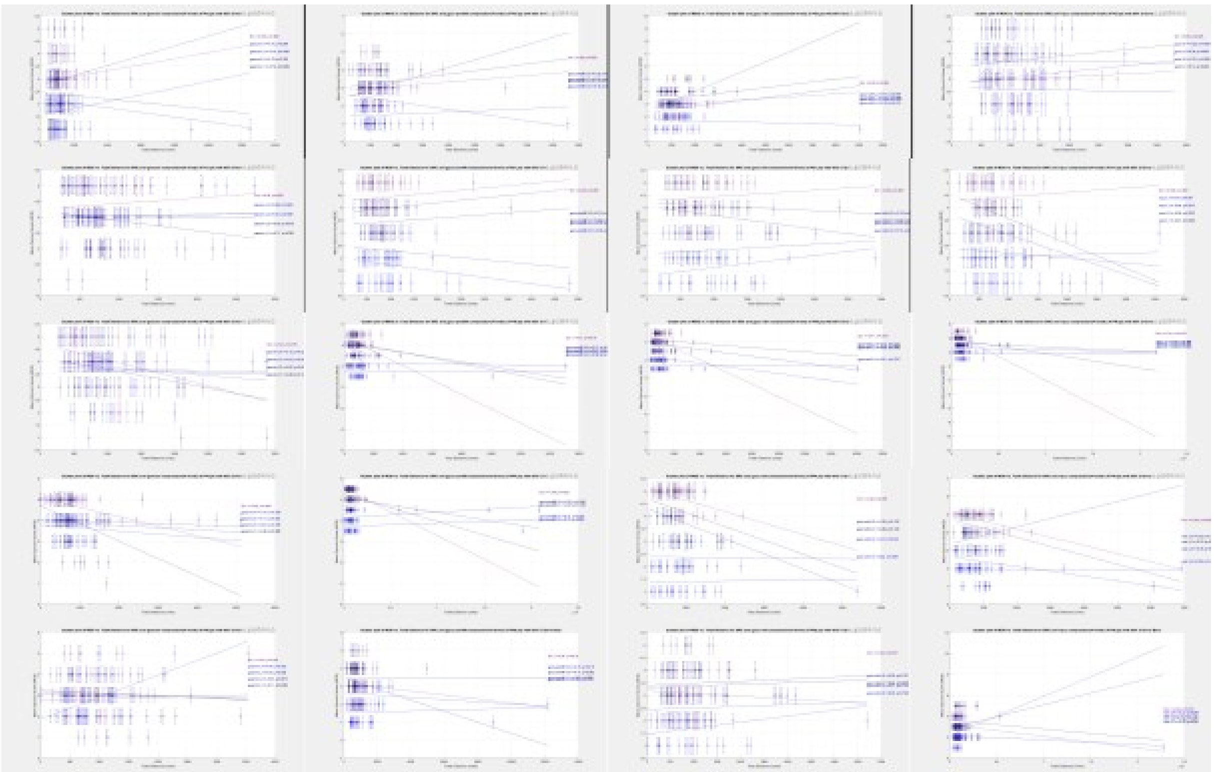

Figure 4.4: Scatter Plot of MOS vs Total Distance of p03,p21,p40,p44,p56 (each row)

consistently low, with most r values indicating very weak linear relationships between the Mean Opinion Score (MOS) and Total Distance. There are occasional moderate correlations, but these are exceptions rather than the norm.

Statistical Significance: Statistical significance is generally lacking in the GeoCNN compression data set, with most p-values well above the 0.05 threshold. This suggests that any observed correlations are not strong enough to confirm a reliable pattern across the data.

Overall Trend: Despite the low correlation coeficients, there is a slight negative trend observable across the graphs. This suggests that as the Total Distance increases, the MOS tends to decrease, albeit not strongly.

Exceptions: Notable exceptions are found in certain graphs where the SRC or a specific level of GeoCNN compression, like L4, shows a moderate correlation that is statistically significant. For instance, in the P40 graph, the SRC had a moderate negative correlation with statistical significance, and in the P44 graph, the SRC showed a moderate positive correlation that was statistically significant.

Significance of Exception: The exceptions are significant in that they break the pattern of weak correlations. The statistically significant correlations for the SRC in P40 and P44 suggest that there may be specific cases where the relationship between MOS and Total Distance becomes more pronounced and potentially important for understanding the perceived quality.

2. GPCC Predlift Compression

Analysis for GPCC Predlift Compression

Correlation Analysis: Across the five GPCC Predlift compression graphs, the correlation between MOS and Total Distance is generally weak, with r values suggesting a minimal linear relationship.

Statistical Significance: In the analysis of the GPCC Predlift compression graphs, there are no instances where the correlation reaches statistical significance. All p-values are above the conventional 0.05 threshold, indicating that the observed correlations could be due to chance.

Overall Trend: There is a consistently weak negative trend across all graphs for GPCC Predlift compression. This suggests that an increase in Total Distance might be associated with a slight decrease in MOS, but the relationship is not strong.

Exception: There are no notable exceptions in the GPCC Predlift compression analysis. All levels of compression follow the same general weak trend without any statistically significant outliers.

Significance of Exception: Since there are no exceptions in the data, there is no significant deviation to explore. The GPCC Predlift compression appears to maintain a consistent pattern of weak correlations across different data sets.

3. GPCC Raht Compression Analysis for GPCC Predlift Compression

Correlation Analysis: The GPCC Raht compression graphs show weak correlationst. However, there is a mix of weak negative and occasional weak positive correlations.

Statistical Significance: Statistical significance is largely absent in the GPCC Raht compression graphs, with the exception of one level in Graph 1 (P03), where L2 demonstrates a significant positive correlation.

Overall Trend: The overall trend for GPCC Raht compression is mixed. While most levels show a weak negative correlation with MOS as Total Distance increases, one level in one graph shows a positive correlation, indicating that not all compression levels affect perceived quality in the same way.

Exception: The significant positive correlation for L2 in Graph 1 (P03) of GPCC Raht compression stands out as an exception. This suggests that, in some cases, a specific level of Raht compression may lead to an increase in MOS with an increase in Total Distance.

Significance of Exception: The significance of the exception in GPCC Raht compression suggests that there may be specific conditions under which a certain level of compression could positively influence the perceived quality. However, this exception is not consistent across other graphs, which implies that it may be an isolated case rather than a trend.

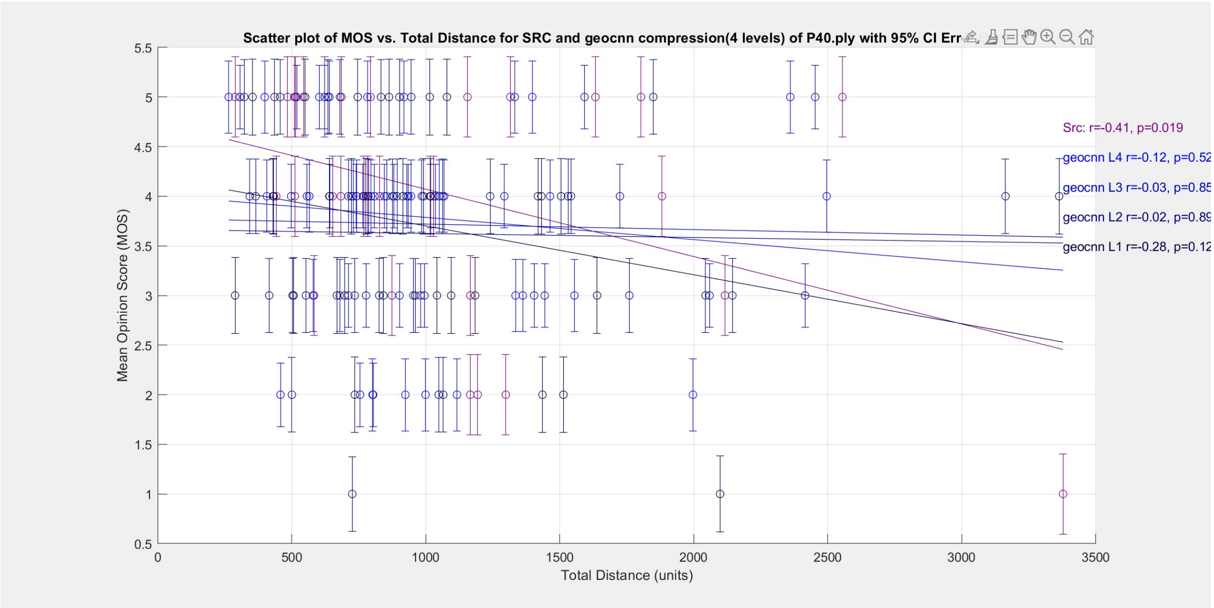

Figure 4.5: For Demonstration: Pearson coeficient shows that it is very weak-negatively correlated and the p-value shows it's not statistically significant as it's value is quite more then 0.05 threshold

4. VPCC Compression

Correlation Analysis: The VPCC compression graphs exhibit very weak correlations throughout, with r values suggesting negligible linear relationships between MOS and Total Distance.

Statistical Significance: The VPCC compression analysis also shows a lack of statistically significant correlations. The p-values across the graphs do not indicate any reliable patterns that would be considered statistically significant.

Overall Trend: The trend for VPCC compression is even weaker than for GeoCNN, with no consistent direction indicated across the graphs. This suggests that there is no strong relationship between MOS and Total Distance for VPCC compression.

Exceptions: There are very few exceptions in the VPCC graphs, and even these do not reach a level of statistical significance. All levels of compression tend to follow the same weak or non-existent trend.

Significance of Exception Given the lack of statistically significant exceptions, there is little to indicate that VPCC compression's relationship between MOS and Total Distance deviates from a general trend of weak correlation. This suggests that other factors may be more critical in determining the perceived quality as measured by MOS for VPCC compression.

Inter-Compressional Technique Analysis

When cross-comparing the four compression techniques, GeoCNN occasionally presents moderate correlations which are significant, potentially indicating a more nuanced relationship between Total Distance and MOS. GPCC Predlift and VPCC compressions show consistently weak or negligible correlations without significant exceptions, suggesting that Total Distance may not be a strong predictor of MOS. GPCC Raht compression stands out for one instance of significant positive correlation, hinting at the possibility of certain levels of Raht compression impacting perceived quality differently. However, this was not a consistent finding across all Raht graphs.

Overall, these findings underscore that while there are overarching trends within each compression technique, the strength and significance of the relationship between Total Distance and MOS can vary. GeoCNN and GPCC Raht occasionally show significant correlations, but these are more the exception than the rule, with the majority of the data suggesting a weak influence of Total Distance on perceived quality.

4.4.3 Average angle or Direction of the cursor

To provide a detailed examination of the connection between the perceived quality of 3D models and the average angle of cursor movement, the Mean Opinion Score (MOS) against this directional metric was plotted.

The set of Graphs demonstrates Average Angle(Direction) by the cursor

Figure 4.6: Scatter Plot of MOS vs Average Angle of p03,p21,p40,p44,p56 (each row)

Intra-Compression Technique Analysis

1. GeoCNN Compression

Correlation Analysis: Across all the graphs for GeoCNN compression, there is a consistently weak negative correlation between the Mean Opinion Score (MOS) and the average angles. This suggests a general but slight tendency for the MOS to decrease as the average angles increase.

Statistical Significance: In none of the graphs is the correlation statistically significant, with all p-values remaining above the 0.05 threshold. This indicates that the weak correlations observed may not be reliable and could be due to random variation in the data.

Overall Trend: The trend is uniformly weak across all datasets, suggesting a very subtle decrease in MOS with increasing average angles. However, due to the lack of statistical significance, this trend cannot be confirmed as a consistent pattern across the board.

Exception: Specific exceptions in the GeoCNN dataset include:

In the P03 graph, the highest level of compression (L4) has a weaker correlation (r = -0.16) compared to SRC, but with no statistical significance (p = 0.369). The P21 graph shows that the L3 level has a slightly better correlation (r = 0.17) than others, but again without statistical significance (p = 0.352). For P40, the L1 level displays almost no correlation (r = -0.15), indicating a negligible relationship.

Significance of Exception: These exceptions across different levels of compression (L3 and L4 in particular) show slight variations in the strength of correlation, but none are statistically significant. Therefore, these exceptions do not provide meaningful insights into the relationship between MOS and average angles in GeoCNN compressed data.

2. GPCC Predlift Compression

Correlation Analysis: Across the GPCC Predlift compression graphs, there is a generally weak negative correlation between MOS and average angles. This suggests that higher compression may be associated with a slight decrease in perceived quality.

Statistical Significance: In all cases, the correlations are not statistically significant, as indicated by p-values greater than 0.05. This indicates that the data does not provide strong evidence for a consistent relationship between the variables.

Overall Trend: The overall trend across different content (P03, P21, P40, P44, P56) shows that the impact of compression on MOS is not significant, and any observed trends in MOS changes with respect to average angles are weak.

Exceptions: P21 and P56 show a stronger negative correlation at the highest level of compression (L4), suggesting that at these specific instances, the MOS might be more sensitive to changes in average angles.

Significance of Exception: Despite these exceptions, the lack of statistical significance implies that they may not be representative of a true underlying effect but could be due to random variation within the dataset.

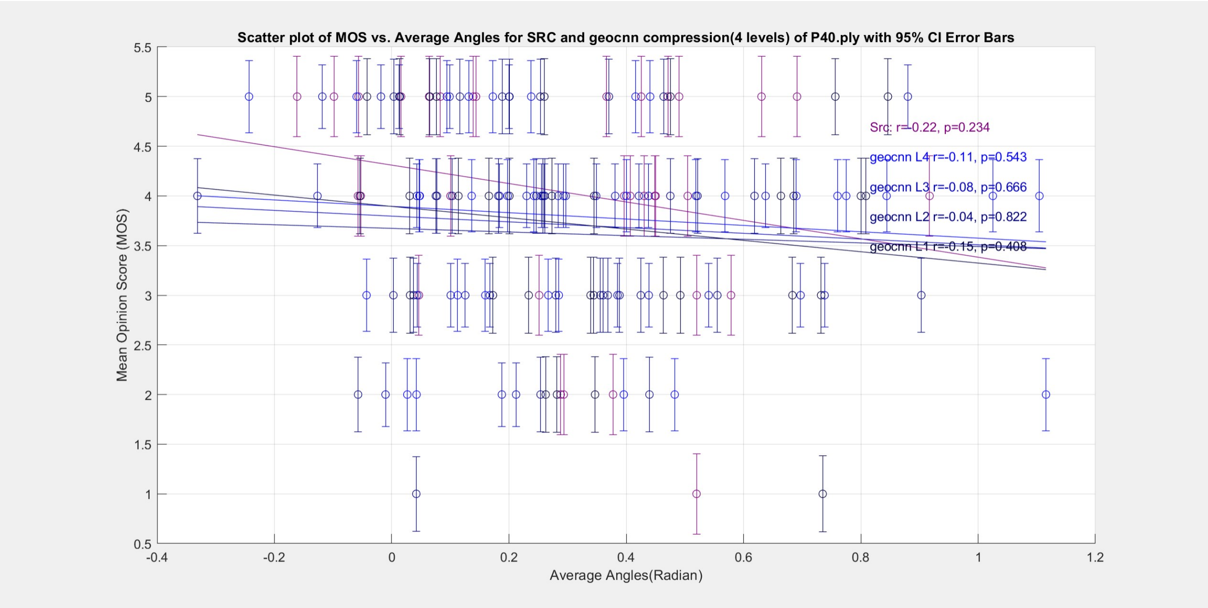

Figure 4.7: For Demonstration: Pearson coeficient shows that it is very weak-negatively correlated and the p-value shows it's not statistically significant as it's value is quite more then 0.05 threshold

3. GPCC Raht Compression

Correlation Analysis: The correlation between MOS and average angles for GPCC Raht compression is also weak and varies in direction, with some levels showing a weak positive correlation.

Statistical Significance: The correlations for Raht compression are not statistically significant, with all p-values exceeding the conventional threshold for significance.

Overall Trend: There is no consistent trend across the datasets; the relationship between MOS and average angles is not strong, and the direction of the correlation varies.

Exceptions: Notable exceptions are observed in P21 and P44, where L3 shows a weak positive correlation, contrary to the general negative correlation trend seen in other datasets and compression levels.

Significance of Exception: Although these exceptions are interesting, they do not hold statistical significance. This suggests that while there may be instances where the relationship between MOS and average angles appears different, these instances are not enough to conclude a reliable pattern across the GPCC Raht compression technique.

4. VPCC Compression

Correlation Analysis:

The VPCC compression graphs display weak negative correlations between MOS and average angles, with a general but not strong suggestion that MOS could decrease with increasing angles.

Statistical Significance: All observed correlations for VPCC compression are not statistically significant, with p-values remaining above the 0.05 threshold across all graphs. This consistently indicates that any correlation present is weak and potentially non-existent.

Overall Trend:

The trend for VPCC is also weakly negative, consistent with the GeoCNN results, but without significant statistical backing, the relationship between MOS and average angles is not established.

Exceptions: Specific exceptions in the VPCC dataset include:

The P03 graph's L2 level shows a somewhat stronger negative correlation (r = -0.26) compared to other levels, yet without statistical significance (p = 0.158). In P21, the L2 level again shows a correlation (r = -0.21) that mirrors the SRC, but it lacks statistical significance (p = 0.242). The P40 graph shows a slightly stronger correlation at L4 (r = -0.11), but it is not statistically significant (p = 0.543).

Significance of Exception: The exceptions noted, such as at L2 and L4 levels, do not present a significant deviation from the overall weak correlation trend. The lack of statistical significance suggests these exceptions do not represent a meaningful relationship between MOS and average angles in the VPCC compressed data.

Inter-Compressional Technique Analysis

General Trends Across Techniques:

Weak Negative Correlation: A common thread in the analysis of GeoCNN, GPCC Predlift, GPCC Raht, and VPCC compression techniques is the presence of a generally weak negative correlation between Mean Opinion Score (MOS) and average angles. This suggests a subtle trend where higher compression may lead to a slight decrease in perceived quality.

Lack of Statistical Significance: In all techniques, the observed correlations fail to reach statistical significance, with p-values consistently above the 0.05 threshold. This indicates that these weak correlations might be due to random variations rather than indicative of a real underlying trend.

Inconsistency in Correlation Direction: Particularly for GPCC Raht, the direction of the correlation is not consistent, occasionally showing weak positive correlations. This contrasts with the generally negative trends observed in other techniques.

Notable Exceptions and Their Implications: Across the techniques, certain levels of compression (like L2 and L4 in VPCC and L3 in GPCC Raht) show slightly different correlation strengths. However, these deviations are not statistically significant. In some cases, such as with P21 and P56 in GPCC Predlift, there is a stronger negative correlation at the highest level of compression. While intriguing, the lack of statistical significance means these observations don't confirm a broader pattern.

4.4.4 Inactive Time (Periods of Mouse Inactivity)

In a detailed correlation analysis, graphs were plotted to examine the relationship between Mean Opinion Score (MOS) and periods of mouse inactivity.

The sets of Graphs for Period of Inactive Time:

Figure 4.8: Scatter Plot of MOS vs Inactive Time of p03,p21,p40,p44,p56 (each row)

Intra-Compression Technique Analysis

1. GeoCNN Compression

Correlation Analysis: The correlation coeficients across the five graphs for GeoCNN compression are generally low, with values typically ranging from -0.10 to 0.53. These suggest weak to moderate linear relationships, indicating that the average inactive time may not be a strong predictor of the Mean Opinion Score (MOS) for GeoCNN compression across different 3D models.

Statistical Significance: Statistical significance is generally lacking in the correlation between average inactive time and MOS, as most p-values are above the 0.05 threshold. This indicates that the correlations observed are not reliable enough to infer a consistent effect of inactive time on perceived quality across the different levels of GeoCNN compression.

Exception: The notable exception is the graph (P44 GeoCNN), where the Level 4 (L4) compression shows a moderate positive correlation (r=0.53) with a statistically significant p-value (0.002). This is an outlier when compared to the other models and compression levels, which do not exhibit such a strong relationship.

Overall Trend: Overall, there is no consistent trend across the different 3D models when considering the impact of GeoCNN compression on the MOS relative to average inactive time. This suggests that other factors may be more influential in determining the perceived quality post-compression.

Significance of Exception: The significance of the exception observed in P44's L4 compression suggests that for this specific model at this particular level of compression, there might be a more pronounced and measurable impact of average inactive time on perceived quality. This could be due to unique characteristics of the model or the compression technique's performance at this level, warranting further investigation.

2. GPCC Predlift Compression

Correlation Analysis: The correlation between Mean Opinion Score (MOS) and Average Inactive Time across the five datasets for the GPCC Predlift compression is generally weak. Most correlation coeficients are close to zero, indicating no strong linear relationship.

Statistical Significance: The majority of the p-values are above the 0.05 threshold, suggesting that the correlations are not statistically significant, with the notable exception of Level 3 compression in P03 and Level 4 in P56, where the p-values indicate significant correlations.

Overall Trend: There is no consistent overall trend across all datasets; however, a slight negative trend in SRC is observed in some cases, while certain levels of GPCC Predlift compression show a positive trend. This indicates a potential variance in how different levels of compression affect the MOS depending on the dataset.

Exception: The exceptions to the general lack of correlation are found in P03 and P56. In P03, Level 3 compression displays a significant positive correlation, whereas in P56, Level 4 compression shows a significant positive correlation.

Significance of Exception: The exceptions are significant in the context of their respective datasets. For P03, the significance (p = 0.036) of the exception suggests that at Level 3 compression, there might be a perceptible relationship between compression and MOS. In P56, the positive trend at Level 4 (p = 0.041) could indicate that as the Average Inactive Time increases, the MOS tends to be higher, potentially implying that less aggressive compression (or better quality at Level 4) is positively received.

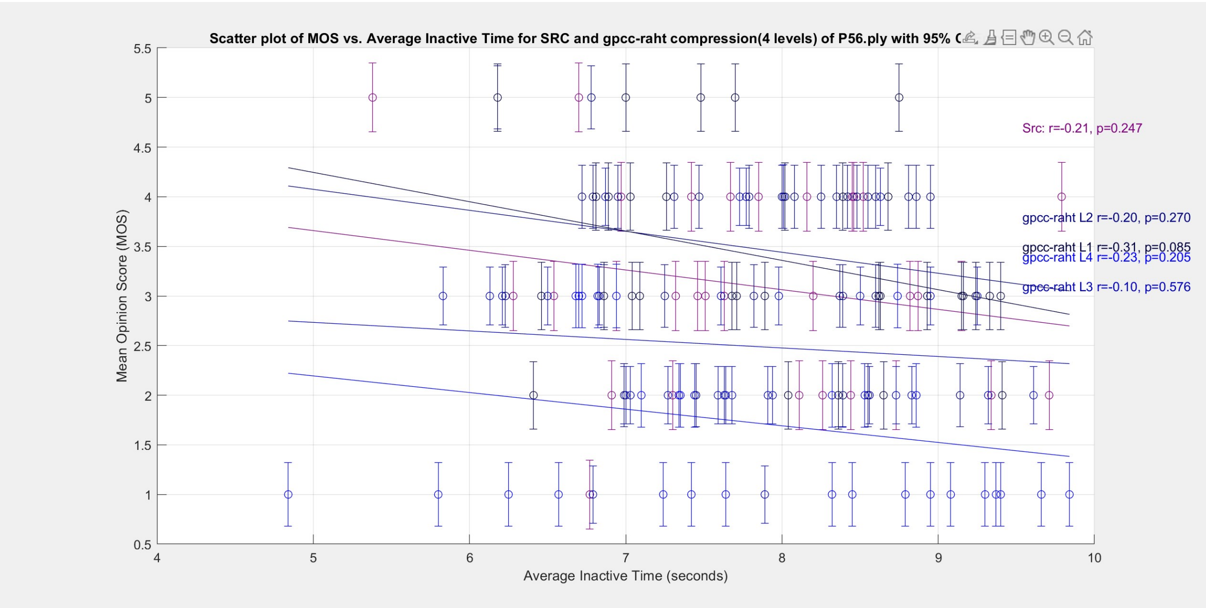

Figure 4.9: For Demonstration: Pearson coeficient shows that it is very weak-negatively correlated and the p-value shows it's not statistically significant as it's value is quite more then 0.05 threshold

3. GPCC Raht Compression:

Correlation Analysis: The Raht compression analysis across the five datasets shows weak to no correlation between MOS and Average Inactive Time. The correlation coeficients vary slightly but remain low.

Statistical Significance: The p-values for the SRC and Raht compression levels are predominantly non-significant. Notably, P21 with Raht compression at Level 1 and Level 2 shows statistically significant correlations.

Overall Trend: The overall trend for SRC and Raht compression levels does not suggest a clear pattern. Any discernible trends are weak and do not establish a clear direction across datasets.

Exception: The exceptions in the Raht compression are observed in P21, where Level 1 and Level 2 show statistically significant positive correlations, unlike other datasets where no significant trends are found.

Significance of Exception: The significance of these exceptions (p-values of 0.017 and 0.014 for Levels 2 and 1 respectively in P21) suggests that for this particular dataset, there is a meaningful relationship between compression and MOS. This could be due to dataset-specific characteristics or viewer preferences that are more apparent under certain compression levels.

4. VPCC Compression

Correlation Analysis: The correlation coeficients for VPCC compression are also generally low, ranging from -0.29 to 0.48 across the five graphs. This suggests that the relationship between average inactive time and MOS is weak to moderate at best.

Statistical Significance: The p-values for most of the VPCC compression levels are above 0.05, indicating that the correlations are not statistically significant, with a couple of exceptions. This implies that the relationship between inactive time and MOS is not consistent across different levels of VPCC compression.

Exception: The exceptions in the VPCC compression are found in P21's Level 1 (L1) compression with a p-value of 0.005, indicating statistical significance, and P44's Level 4 (L4) compression with a p-value of 0.040. These exceptions stand out from the overall pattern of non-significant results.

Overall Trend: There is no clear overall trend across the VPCC compression graphs that would indicate a strong influence of average inactive time on MOS. The data suggests that average inactive time is not a strong predictor of perceived quality for VPCC compression.

Significance of Exception: The exceptions in VPCC compression (P21 L1 and P44 L4) suggest that there may be instances where the average inactive time can have a more substantial effect on the perceived quality, at least for certain models at specific compression levels. These exceptions are significant because they defy the general trend of non-significance and indicate potential areas where VPCC compression could be optimized or requires further scrutiny.

Intra-Compression Technique Analysis

When comparing across the four different compression techniques—GeoCNN, GPCC Predlift, GPCC Raht, and VPCC—the analyses reveal a few overarching patterns and distinct exceptions.

Correlation Analysis: Across all techniques, the correlation coeficients between Mean Opinion Score (MOS) and Average Inactive Time are generally low. This suggests a weak to moderate relationship across various 3D models and compression levels, indicating that Average Inactive Time is not a consistent predictor of MOS regardless of the compression technique used.

Statistical Significance: For most of the compression techniques and their levels, statistical significance is absent, with p-values largely exceeding the 0.05 threshold. This lack of significance is consistent across GeoCNN, GPCC Predlift, GPCC Raht and VPCC compressions, with sporadic exceptions that do not form a pattern indicative of a strong relationship between inactive time and perceived quality.

Overall Trend: No consistent trend is observable across the different datasets for any of the compression techniques. The slight negative trend observed in some SRC cases does not hold across all datasets or compression levels, indicating that the impact of compression on MOS is not uniform.

Exceptions: Notable exceptions occur within certain levels of compression for specific datasets, suggesting that while the general trend is weak, particular instances stand out. For example, P21 shows significant positive correlations at certain levels in both Raht and VPCC compressions. Similarly, P44's GeoCNN Level 4 compression and P03's GPCC Predlift Level 3 compression present significant correlations, indicating a unique relationship between inactive time and perceived quality for those specific cases.

Significance of Exception: The exceptions observed are significant in their contexts, implying that for certain models and compression levels, there is a noticeable impact on perceived quality. This is particularly true for P21 in Raht and VPCC compressions, where the significant positive correlation suggests that as inactive time increases, so does the MOS, albeit in isolated instances. These exceptions highlight potential unique interactions between compression techniques and models that warrant further investigation to understand and optimize perceived quality.

4.5 Directional Graph Analysis

For Demonstration out of the 85 directional graphs generated only one of it is presented here for explanation.

The image on the next page is a visual representation of the study of mouse movement . The key points of the graph are:

1. Directional Graphs: The lines and curves represent the paths taken by the mouse cursors of the participants as they interacted with the images. Each participant's mouse movement is plotted with a different color.

2. Data Aggregation: This graph is an aggregate of all mouse movements from 32 participants who interacted with 85 images. It is of one of the 85 images. This suggests that the experiment may have been focused on how users interact with these images, possibly to study their attention, interest, or how they process visual information.

3. Metrics Recorded: The graph provides an average of various metrics across all the

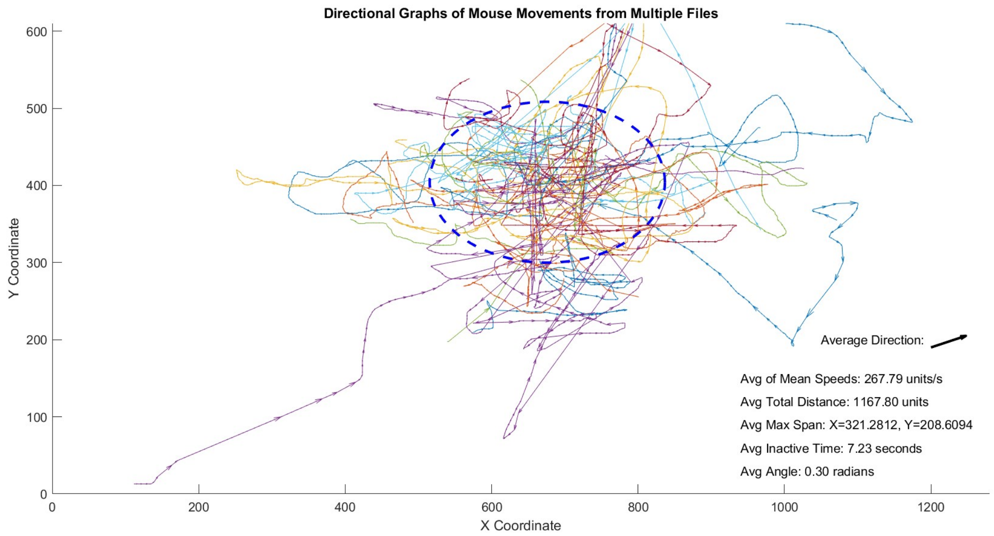

Figure 4.10: One of the 85 Directional Graph

participants and images, such as: - Mean Speeds: The average speed of the mouse movement, which can indicate how quickly participants moved the mouse between points. - Total Distance: The average total distance covered by the mouse, which can reflect how much participants explored the images. - Max Span: The maximum extent of mouse movement in both X (horizontal) and Y (vertical) directions, indicating the area within which most mouse activity was concentrated. - Inactive Time: The average time during which the mouse was not moving, which might indicate moments of pause for thought or decision-making. - Angle: The average angle of mouse movement, possibly indicating the general direction of mouse paths or the orientation of the images.

4. Central Ellipse: The ellipse in the center of the graph represents the area where mouse activity was most dense or concentrated. The height and width of the ellipse correspond to the average maximum span in the X and Y coordinates, showing where participants focused their attention the most.

5. Average Direction: The arrow marked as 'Average Direction' shows the overall trend or the dominant direction of mouse movement across all the data. This could indicate a bias in how the participants tended to move the mouse or interact with the images.

The significance of each point and path on the graph is that they collectively represent the behavior patterns of users interacting with visual stimuli. The graph provides a visual representation that mirrors the user interface layout employed in the visualization study. The densely populated regions, where the mouse movements are highly concentrated, appear to align with the central focal point of the displayed ply images. This suggests that the participants' interactions predominantly revolved around a central point of interest within the images, which could be inferred as the 'center of gravity' in the visual context. This concentration of activity may indicate that the design of the PLY images naturally draws the viewer's attention towards the center region, guiding the mouse movements of the participants accordingly.

Chapter 5

Conclusions

The correlation analysis explored the relationship between the Mean Opinion Score (MOS) and several user interaction parameters: range of X and Y-axis movement, total distance traveled by the cursor, periods of mouse inactivity, and mean angle of movement. This multifaceted investigation was crucial for understanding how different compression techniques affect user engagement with 3D models.

Range of Movement: Both X and Y-axis movements showed mostly weak correlations with MOS across various compression techniques, with the majority of p-values above the 0.05 significance threshold, indicating no strong evidence of a consistent impact. Exceptions like the P56 model on the X-axis and P40 and P44 on the Y-axis demonstrated significant correlations in isolated instances, suggesting model-specific factors could influence user perception.

Total Distance Traveled: Across the GeoCNN, GPCC-Predlift, GPCC-Raht, and VPCC compression techniques, the correlation between the total distance traveled by the cursor and MOS was generally weak. Notable exceptions were observed in the GeoCNN dataset, where moderate correlations occurred, such as with the SRC data in the P40 and P44 models, pointing towards specific instances where the relationship becomes more pronounced.

Inactive Time: For periods of mouse inactivity, the correlation with MOS was weak to moderate, indicating that inactive time might not be a strong predictor of MOS for GeoCNN compression across different 3D models. However, a significant positive correlation was found in P44's L4 compression in the GeoCNN dataset, suggesting a potential impact of inactive time on perceived quality in specific contexts.

Mean Angle of Movement: The average angle of cursor movement also showed a weak negative correlation with MOS across all techniques, with no consistent or statistically significant trend established. Variations in correlation strength were noted at different levels of compression but lacked statistical significance, implying that these did not provide meaningful insights.

5.1 Limitations and Opportunities for Improvements

1. Challenges in Discriminating Mouse Interactions: The current methodology for capturing mouse movement data faces a critical limitation in differentiating between dragged and non-dragged interactions. The primary function utilized, 'onMouseMove', records cursor movement without discriminating between these interaction states. Efforts to integrate auxiliary functions that could flag and categorize movements as either dragged or non-dragged have led to resource conflicts. These conflicts adversely affect the granularity and accuracy of the data, undermining the integrity of the recorded coordinates. Which is why usage of this auxillary functions has been bypassed for current conditions. But the precision of data capture is paramount; without high accuracy, the utility of the mouse movement data in understanding user behavior is significantly compromised. Consequently, this presents a need for improvement in the data collection mechanism to ensure a comprehensive and precise analysis of mouse dynamics.

2. Participant numbers Considerations and Digital Proficiency Classification: The experimental design incorporated a relatively modest sample of 32 participants, which may yield trends that could be statistically insignificant. The size and diversity of the participant pool are crucial factors that can substantially influence the reliability and generalizability of the findings. Furthermore, the experiment did not implement a mechanism to classify participants based on their digital adeptness. The absence of this differentiation presents a notable challenge, as the variance in digital proficiency can profoundly impact user behavior and interaction patterns with digital interfaces. Addressing this gap by expanding the participant base and establishing criteria for digital skill levels stands as a significant opportunity for enhancement. Such stratification would enable a more in-depth analysis, providing deeper insights into how digital fluency influences interaction trends and strengthen the overall finding.

3. Refinement of Visual Space Classification and Expansion of Data Sets: The experiment utilized a three-category classification system for .ply files—Humans & Animals (HA), Inanimate Objects (IO), and Buildings and Landscapes (BL). However, this classification overlooks the varying spatial occupation of different 3D images within the visual interface. For instance, a 3D representation of a human figure versus a building does not uniformly fill the viewer's interface, potentially affecting the user's interaction with the content. A classification

system based on the proportion of visual space occupied may offer a more refined approach to understanding user engagement with the interface. Additionally, incorporating a broader array of .ply files would likely yield a more comprehensive analysis. It is also advisable to decentralize the focus on compression techniques across multiple experiments to assess their impacts more granularly. Such methodological enhancements are anticipated to yield richer insights and facilitate a more nuanced understanding of user interactions with 3D visual content.

4. Advancing Statistical Techniques for Pattern Identification: In the current experiment, traditional statistical methods like the Pearson coeficient and p-value calculations were employed to analyze the data. While these are foundational tools in statistical analysis, they may not be suficient to uncover complex patterns, especially in datasets with a considerable amount of noise or variability. The limitation here lies in the potential for in-depth trends to remain hidden beneath the surface of apparent randomness. To overcome this, future enhancements could include the integration of advanced statistical approaches.

In conclusion, when analyzing the data using the four key parameters—1) Range of X and Y-Axis Movement, 2) Total Distance Traveled by the User's Cursor, 3) Inactive Time (Periods of Mouse Inactivity), and 4) Mean Angle of Movement—it becomes evident that the perceived quality of 3D models by users is not predominantly influenced by these aspects of cursor dynamics as of now. Presently, this observation suggests that factors such as model complexity, visual fidelity, and individual user preferences might play a more significant role in determining the Mean Opinion Score (MOS). The exceptions observed in specific cases highlight the importance of further study to delve into how compression techniques impact user engagement and perception under particular conditions.

Implementing the suggested improvements, such as refining data collection methods to distinguish between dragged and non-dragged interactions, expanding the participant base, adopting a more precise classification of visual space in the interface, and utilizing advanced statistical techniques, could lead to more meaningful and reliable results. Such enhancements in the methodology are anticipated to provide a deeper understanding of the relationship between user interaction patterns and their perception of 3D model quality.

Acronym

ACR Absolute Category Rating

BIOS Basic Input/Output System

CI Confindence Interval

CWI Centrum Wiskunde Informatica

DSIS Double Stimulus Impairment Scale

GPCC Geometry Based Point Cloud Compression

HEVC High Eficiency Video Coding

LIDAR Light Detection and Ranging

MOS Mean Opinion Score

MPEG Moving Picture Experts Group

OBJ Object File

PC Point CLoud

PLY Polygon File Format

PSNR Peak Signal-to-Noise Ratio

RAHT Random Access and Hierarchical Tree Structure

RGB Red Green Blue

SRC Source

VPCC Video BAsed Point Cloud Compression

XR Extended Reality

Bibliography

[1] Evaluation criteria for PCC (Point Cloud Compression), volume Recommendation ISO/IECJTC1/SC29/WG11 MPEG2016/n16332. ITU-T.

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

[3] Zerman E. Quach M. Chetouani A. Smolic A. Valenzise G. Le Callet Ak, A. Basics: Broad quality assessment of static point clouds in a compression scenario. P.(2023).

[4] R. Rajesh W-T. Ooi C-H. Wu, X. Li 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, pages 98–105, 2021.

[5] S. Kuma A. Zaghetto-T. Suzuki D. Graziosi, O. Nakagami 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.

[6] E. Zerman I. Viola-G. Lavou ´ e A. Ak A. Smolic ´ P. Le Callet E. Alexiou, Y. Nehme and P. Cesar. "subjective and objective quality assessment for volumetric video," in immersive video technologies. Elsevier, 2023.

[7] N. Yang E. Alexiou 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.

[8] T. Myers E. d'Eon, B. Harrison 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.

[9] K. Mammou A.M. Tourapis J. Kim D.B. Graziosi S. Rhyu E. S. Jang, M. Preda 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), pages 118–123, 2019.

[10] G. Valenzise R. Mantiuk E. Zerman, V. Hulusic and F. Dufaux. ”the relation between mos and pairwise comparisons and the importance of cross-content comparisons,”in human vision and electronic imaging conference. IST International Symposium on Electronic Imaging (EI 2018), Burlingame, United States,, January, 2018.

[11] P. Gao E. Zerman, C. Ozcinar 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.

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

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

[14] J. Li I. Viola, S. Subramanyam 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.

[15] ITU-R. Methodology for the subjective assessment of the quality of television pictures. ITU-R Recommendation BT.500-14, 2019.

[16] ITU-T. Subjective video quality assessment methods for multimedia applications (Recommendation P.910), volume Recommendation ISO/IECJTC1/SC29/WG11 MPEG2016/n16332. International Telecommunication Union. from https://www.itu.int/rec/T-REC-P.910.

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

[18] ITU-T. Subjective methods for the assessment of stereoscopic 3dtv systems. recommendation itu-r bt.2021. 2021.

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

[20] Z. Liu W. Sun J. Li, C. Zhang 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.

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

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

[23] G. Valenzise M. Quach and F. Dufaux. Improved deep point cloud geometry compression. IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP), 2020.

[24] A. Tomaszewska R. K. Mantiuk and R. Mantiuk. Comparison of four subjective methods for image quality assessment. Comput. Graph. Forum, vol. 31, no. 8, p. 2478–2491,, December, 2012.

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

[26] Y. Xu Y. Liu, Q. Yang and Z. Ma. Assessing objective metrics for point cloud compression. 2021.

[27] F. Dupont P. LeCallet Y. Nehme, J.-P. Farrugia and G. Lavou ´ e. “comparison of subjective methods, with and without explicit reference, for quality assessment of 3d graphics,”. ACM Symposium on Applied Perception 2019, ser. SAP ’19. New York, NY, USA: Association for Computing Machinery,, 2019.