



Final Thesis Report

MSc. GIMA Thesis

Labeling and segmentation of treepoint cloud in immersive virtual reality.

Filippos Sakellariopoulos

2078015

f.sakellariopoulos@students.uu.nl

Supervisors Alexander Klippel, Jiayan Zhao
Responsible Professor: Ron van Lammeren



Abstract

Effective data annotation is crucial for the advancement of machine learning techniques for autonomous labeling of point cloud datasets for environmental studies and other fields. Traditional point cloud annotation workflows utilize 3D datasets on a 2D screen, which degrades the user experience and negatively impacts performance. Concepts like embodied digital twins and virtual reality may offer an alternative approach for annotating and exploring datasets in 3D. However, there is a lack of research on determining which annotation approaches offer the most optimal solution for labeling and segmenting components of individual tree point clouds. Additionally, only a few studies refer to users' evaluations of usability and workload and compare different annotation approaches, as well as examine the influence of participants individual differences. To address this gap, game engine technologies were used to recreate a virtual annotating environment and an interactive experience to capture the users' experience and performance on two commonly used annotation approaches and their combinations.

The first annotating method, namely the Sphere Selection approach, offered a 3D pointer with adapting shapes for free-style detailed point selection. The second method, namely the Container Creation approach, offered a solution based on control points and nodes for polygon creation, offering a quick approach for massive point selection and annotation. The combination of the two methods formed the third approach, namely, the combination of the two approaches. The results indicate that the sphere selection approach is optimal in terms of time and accuracy for annotating components of individual trees. Participants reported that this approach was perceived as more effective, enjoyable, and the least physically and mentally demanding annotation technique compared to the other two. The findings of this research highlighted the potential of each approach for future environmental applications for labeling 3D point cloud data in virtual reality.

The findings of this research contribute to field data annotation and highlight the potential for immersive technologies in environmental studies. Future researchers can focus on further investigation of the potential of each annotation approach for labeling and segmenting individual tree point clouds and further investigation of participant's individual differences on performance and usability. Finally, the findings of this thesis could be a starting point for the design criteria for 3D VR tree annotation tools and applications and the participant application usability study.

Keywords: Embodied digital twins, virtual reality, 3D data annotation, point cloud, Unity, user evaluation

Table of Contents

1. Introduction.....	5
1.1 Background.....	5
1.2 Scope and limitations.....	6
1.3 Research objectives and research questions	7
2. Theoretical background and related work.....	8
2.2 Digital Twins in Environmental Sciences.....	9
2.3 Point clouds and annotation challenges	9
2.4 Existing tools and approaches.....	11
2.5 Evaluation of accuracy, time, usability, and workload.....	13
2.6 Importance of gender reporting in virtual reality studies.....	14
2.7 Literature Review summary and discussion	15
3. Methodology.....	17
3.3 3D Point cloud data	17
3.4 Software	19
3.4.1 Cloud Compare	19
3.4.2 Unity	19
3.4.3 SPSS.....	19
3.5 Development and annotation environment	20
3.6 WANDER Lab's annotation tools	20
3.6.1 Meta Oculus 2 controllers features	21
3.6.2 Annotating with the Sphere Selection Approach.....	23
3.6.3 Annotating with the Container Creation Approach	24
3.6.4 Annotating with the combination of the two approaches	25
3.7 Experiment structure and users' evaluation process	25
3.7.1 Introductory Phase	25
3.7.2 Annotation Phase	26
3.7.3 Data Collection and Users' Evaluation.....	28
3.7.4 System Usability Scale and Nasa Task Load Index calculation	29
3.7.5 Measuring time and accuracy calculation.....	30
3.8 Data Analysis	30
4. Results.....	31
4.1 Participants.....	31
4.3 System Usability Scale	31
4.4 NASA Task Load Index.....	32
4.5 Time	33
4.6 Accuracy	33
4.7 Effect of Approach and Gender on the SUS scores	34
4.8 Effect of Approach and the Gender on the TLX scores.....	34
4.9 Effect of Approach and the Gender on the Time variable	34
4.10 Effect of Approach and the Gender on the Accuracy scores	35
4.11 User's feedback and comments	35
4.9.1 Sphere Selection Approach feedback	35
4.9.2 Container Creation Approach feedback.....	36
4.9.3 Combination of two approaches feedback.....	36

5.	Discussion.....	37
5.1	Time and Accuracy	37
5.2	Usability, workload, and overall satisfaction assessment.....	38
5.3	Environment and Data discussion.....	40
5.4	Navigational features discussion.....	40
5.4	Limitations	41
5.4.1	Participants and experiment setup	41
5.4.2	Accuracy evaluation	41
5.4.3	Other challenges.....	42
6.	Conclusion	43
7.	References	44
Appendix 1:.....		48
A.1.1	Unity Project and Annotation Tool Setup	48
A.1.1.1	Download and Setup of UnityHub	48
A.1.1.2	Set up Gitlab	48
A.1.1.3	Gaining access and installation the Unity Packages.....	49
A.1.1.4	Oculus Setup	49
A.1.1.5	Developer Mode Enabling	49
A.1.1.6	Developer Organization Establishment.....	50
A.1.1.7	Connect Oculus to Unity	51
A.1.1.8	Troubleshooting during the Unity project setup	51
A.1.2	Unity Project package settings	52
A.1.2.1	LAZ Loader	52
A.1.2.2	LAZ Renderer	54
A.1.2.3	Creating a new room/ Join an existing room	56
A.1.2.4	Further settings during the experiment.....	57
Appendix 2: Questionnaires.....		59
A.2.1	Demographics Form:.....	59
A.2.2:	System Usability Scale (SUS) Index Questionary	60
A2.3:	NASA Task Load Index (NASA- TLX) Questionary	60
A.2.3:	Final Questionary- Comments	61

1. Introduction

1.1 Background

The term remote sensing refers to the study and practice of collecting and interpreting information about the environment from a distance, primarily by sensing radiation that is naturally emitted or reflected by the earth's surface or from the atmosphere, or by sensing signals transmitted from a device and reflected back to it (Lillesand et al., 2015). In many cases, it could be utilized to improve detection, mapping, and monitoring due to its numerous useful characteristics (Kumar et al., 2013). Several remote sensing solutions are presently accessible to the environmental sciences community due to recent breakthroughs in sensor technology, data management, and data analytics (Khanal et al., 2020). Because of their outstanding measurement precision, low acquisition time, and high degrees of detail, remote sensing techniques have tremendous potential for gathering spatial information on forests (Trochta et al., 2017).

The employment of high-performance Light Detection and Ranging (LiDAR) remote sensing technology in environmental science applications has lately become a feasible method and important environmental sensing technique for feature extraction and 3D mapping. High-quality data on three-dimensional characterizations of forest structures is often obtained by LiDAR measurements of forest regions (Akay et al., 2009). Thus, the combination of LiDAR and current developments in computer 3D graphics makes it possible to create accurate digital representations of natural features (e.g., plants, individual trees, and forests). Specifically in the field of forestry, the employment of the technology of terrestrial laser scanning (TLS) offers more accurate 3D representations that lead to a better understanding of the forest ecology and dynamics of forest systems (Trochta et al., 2017).

In the fields of environmental sciences, remote sensing techniques such as terrestrial and drone-based LiDAR allow for a tremendous amount of accurate 3D data to be collected with a quick turnaround. The process of classification in 3D point clouds has become more relevant for ecologists and foresters due to the widespread use of LiDAR sensors in forest monitoring and management (Henry et al., 2015). Devices like laser scanners and depth cameras can generate 3D point clouds of trees, but the produced data are disorganized, unlabeled masses. In their current form, these data don't represent different components of the individual tree's point clouds (tree trunk, branches, and leaves). Trees are essential to the ecosystem's healthy functioning and provide a variety of environmental benefits, including regulation of water systems, maintenance of air quality, carbon sequestration, and promotion of biodiversity. Consequently, detailed and updated tree inventories are crucial for monitoring and preserving ecological environments (Gupta et al., 2018).

In this regard, approaches such as computer vision have made remarkable strides in their ability to identify and recognize a variety of features from point cloud data. Yet, these approaches are hindered from scaling to a high number of item categories by the size and nature of existing datasets. Additionally, existing deep learning approaches may offer individual tree detection and species classification from these data (Persson et al., 2022), but do not differentiate the different components of individual trees. Moreover, a significant issue with these techniques for automatic labeling is their low accuracy (Russel et al., 2009; Goel et al., 2020). Furthermore,

supervised machine learning techniques have been used to automatically segment point cloud data with promising accuracy (Mirzaei et al., 2022). However, the laborious nature of producing enough training data hinders the scalability and quality of the outcome. In recent years, the study of efficient data annotation has emerged as its own discipline.

In many cases, users will need to manually annotate the point cloud to pinpoint certain features (Bacim et al., 2014). Consequently, segmentation of complicated 3D objects relies mostly on manual work. As described by Doula et al. (2022), annotation is a time-consuming activity that must be performed by humans for deep learning models to be trained. A main issue that is prevalent in the literature is the representation of 3D datasets on a 2D desktop screen, which degrades the user experience and negatively impacts performance (Zimmer et al., 2019; Doula et al., 2022). In the traditional workflow of point cloud annotation, users need to draw a hull around a subset of points displayed on the flat screen. To appreciate the spatial layout of a 3D point cloud on a 2D screen, the lack of binocular depth cues from a single perspective necessitates that users alter their views using the mouse or keyboard, which often ends up with important geometric structures missing (Simpson, 2020).

Recent improvements in digital twins and virtual reality (VR) technology have made it possible to explore the full potential of this technology on 3D environmental datasets, especially LiDAR point cloud data collected from forests (Toma et al., 2012). The VR community is constantly expanding due to the availability of several low-cost or free data sources, free software tools, and applications for various VR devices, such as the developing platform game-engine Unity. Although there are limited publications on these concepts, the use of virtual reality applications and 3D-VR environments have shown great potential in handling 3D point clouds and annotation processes (Monica et al., 2017; Wirth et al., 2019). Thus, it is believed that the creation of the required training datasets for machine learning algorithms for forestry is promising, based on these concepts.

Despite these technological strides, there is a lack of research exploring and comparing different annotating techniques in 3D-VR and their performance. Additionally, annotating techniques for labeling individual tree components have now been examined. Furthermore, as pointed out by Wirth et al. (2019), the user evaluation of 3D data annotation methods is not present in most of the studies. Although researchers such as Monica et al. (2017) evaluated labeling processes based on the user's experience as a part of their research, their approach was descriptive and limited. As a result, the benefits of different annotation approaches for segmenting and labeling components of tree point cloud are still unclear. Moreover, publications (Shafer et al., 2017; MacArthur et al., 2021; Lorenz et al., 2023) discuss the influence of individual characteristics, but the possible relation between them, performance and users' experience while annotating in 3D is unknown.

1.2 Scope and limitations

The scope of this research was to investigate and to evaluate different data annotation approaches by a group of 15 participants and to provide better understanding to annotate point cloud data more effectively. The goal was to provide insight into the field point cloud data annotation to researchers, educators, and practitioners. This evaluation was performed while annotating components of individual trees. The context of this study was related to forestry and virtual reality, to align with the current of future research interests of the WANDER Lab of Wageningen University, which provided the necessary tools for completing this thesis.

The vision of this research was to set a baseline for future studies and applications, since there

are no existing publications relating the concepts of data labeling in virtual reality and user evaluation in environmental applications. For this reason, this thesis could be a starting point for understanding the capabilities of different annotation approaches for these concepts. Finally, since the creation of efficient training datasets for environmental applications is a rising topic, this research could provide inspiration for future projects and research.

For this effort, innovative approaches were followed by employing the concepts of digital twins and virtual reality for exploring and annotating point cloud of trees. The outstanding and unique characteristics of these applications offered advanced capabilities for exploring and interacting with the point cloud datasets while performing labeling and segmentation processes in 3D. This experience was followed by an overall evaluation of the approaches, in terms of time, accuracy, usability, overall user satisfaction and an attempt to study their relation to the participant's individual differences.

The primary limitation of this research was the accuracy measure approach, since it was based on a visual comparison between the annotated point clouds of participants and point clouds (ground truth) annotated with Cloud Compare software (verified by an expert annotator). This could result in subjective results. Furthermore, the user's evaluation process setup came with weaknesses, since one of the examined annotation approaches was the combination of two other approaches and it was necessary to be evaluated always at the last. This design choice was not able ensure the comparability on performance of participants among the three conditions. Moreover, the effects of the known limitations of VR sickness and discomfort are collected, reported but not examined in depth, nor related specifically with each annotating approach. Finally,

1.3 Research objectives and research questions

The main objective of this thesis was to evaluate three 3D - VR approaches, designed for point cloud data to annotate individual tree components. Therefore, the main research questions were:

RQ1.1 Which annotating technique is considered the most optimal, in terms of the time and accuracy for annotating different components of the individual tree's point clouds in VR?

RQ1.2 Is there a relationship between the performance of different annotating approaches in VR and participants individual differences?

RQ2.1 What is the user experience (including usability and workload) with the different annotating approaches in VR?

RQ2.2 Is the usability and workload pervence of different annotating approaches in VR related to the participant's individual differences?

RQ2.3: Which recognized elements are present in annotation approaches to enhance usability, particularly in terms of efficiency?

2. Theoretical background and related work

The following chapter summarizes several scientific topics that were considered essential elements of this research. First, the concept of Digital Twins and the relative terms of immersive technologies and virtual reality are presented. The following chapters include crucial information related to LiDAR technology and point cloud data in environmental sciences. Afterwards, the principles of data annotation, relevant existing applications and annotation tools are explored. The last section includes information about existing evaluation methods and processes. This chapter provides necessary information which are linked to answering RQ1.1 and provides the necessary information for the methodologies used to answer RQ2.1.

2.1. The concept of Digital Twins and immersive technologies

As it described in the literature by Panetto et al. (2019), the current digital transformation of businesses necessitates the creation and use of digital models, known as digital twins, that represent an asset of knowledge about the actual processes involved. The term Digital twins refers to simulations of real-world objects of high accuracy, which may be used for assessment, optimization, and prediction (Graessler and Poehler, 2017; Semeraro et al., 2021).

As described by Klippel et al. (2021) the concept of Digital Twins can be combined with the concept of Embodiment. Through the usage of Embodied Digital Twins, users may be placed in digital environments that become an integrated part of the user's knowledge base, and unique, in-depth experiences can be built around the user's physical presence. Through these experiences, users can perform cognitive assignments, interact with aspects of the virtual environment, and enhance their knowledge. Given the location-specific character of the fields they intersect with, embodied digital twins hold special significance for environmental applications (Klippel et al., 2021).

The use of Digital Twins in Environmental Sciences can be traced back to the early 2000s, with the development of virtual reality (VR) and augmented reality (AR) technologies. The phrase "immersive technologies" is used to describe technologies that offer a fully submerged experience by engaging all users' senses (Jerald, 2016; Klippel et al., 2021). As it is known, all aspects of human perception, including education, research, communication, and choice-making, are founded on the individual's past experiences. Klippel et al. (2019, 2021) through numerous publications, point out the way that immersive technologies have differentiated the way of understanding environmental processes.

These digital environments can be explored with equipment, such as head-mounted displays (HMDs). Professionals and decision-makers can enhance their knowledge base through embodied experiences apart from the use of data and traditional methodologies. This is achieved through a great deal of freedom of choice, which lets them focus on their own perspectives while interacting with digital content. HMDs' tracking capabilities provide the user agency on par with that in the physical world. To have a truly embodied experience, this degree of agency is required (Klippel et al., 2021).

One potential limitation of these types of equipment and methodologies, which is discussed in the literature, is simulation sickness and general discomfort during virtual reality experiences and especially the longer a user is exposed to them. These effects can be a frequent issue

problem in virtual reality applications, and it is distinguished by symptoms such as nausea, confusion, and general discomfort and presents a significant obstacle to collecting accurate data from individuals (Ihmedy-Steinke et al., 2017; Wallgrun et al., 2019).

2.2 Digital Twins in Environmental Sciences

The concept of the Digital Twin in the case of environmental sciences has been referred to in the past as virtual geographic environments, capturing similar concepts and techniques (Lin et al., 2015; Klippel et al., 2021). In the work of Klippel et al. (2021), the concept of embodied digital twins for environmental applications was presented for cases such as research, education, communication, and decision-making. As they describe in their work, based on different examples, these applications represent the state-of-the-art, by incorporating photogrammetry and remote sensing methods and can be used to study inaccessible or remote locations Klippel et al. (2021).

Applications of Digital Twins that are worth mentioning are those of Thrihnukar volcano in Iceland (Zhao and Klippel, 2019a) and the Critical Zone Investigator (Sajjadi et al., 2014). The Thrihnukar volcano study utilized satellite imagery data and a Digital Elevation Model (DEM) to create a simple recreation of the study area. The volcanic system's interior structure was then constructed with Structure-from-Motion photogrammetry and terrestrial laser scanning (Zhao and Klippel, 2019a). This offered geoscientists a high-resolution point cloud of the volcano's interior and a detailed terrain model portraying the geographical position of the volcanic complex. Additionally, tools such as scaling, rotation, and translation, as well as a set of geometric measurement tools such as length, area, and volume are available on the workbench enabling direct manipulation of georeferenced data sets (Zhao and Klippel, 2019a).

Similarly, Digital Twins were utilized to explore the Critical Zone in the second case provided by Klippel et al. (2021) for the Critical Zone (the layer of the Earth that is vital to the food, water, and energy supply) observatories project. The project's objective was to establish facilities for the research of chemical, physical, and biochemical functions in the Critical Zone. Based on real time data, CZ Investigator, a meaningful adventure game with the overarching goal of introducing people to the notion of the Critical Zone. The objective of this project is to help users learn more about the Critical Zone, the interdependence of the Critical zone's food, energy, and water systems, and the effects of both natural and human-caused processes (Sajjadi et al., 2014; Klippel et al. 2021).

This subsection explored different environmental applications and pointed out some key features of Embodied Digital Twins that were relevant to this thesis. Secondly, these publications provided concrete evidence that these concepts could be successfully applied in environmental studies. More precisely, the Critical Zone project (Sajjadi et al., 2014; Klippel et al. 2021) investigated the idea of game-like experience, with exploring and achieving tasks while interacting with features of the virtual environment. A similar idea was followed for the annotation the evaluation of the annotation's methods of this thesis, as different annotations scenes, annotation features and tasks were given to the participants in the form of an interactive experience.

2.3 Point clouds and annotation challenges

Forestry studies have extensively explored the use of point cloud data. Airborne LiDAR, a technique for acquiring point cloud data, has demonstrated its reliability and effectiveness in forest inventory and analysis, from individual trees to large and complex landscapes (Sheridan

et al., 2014). Point cloud data, a form of 3D data that captures the surface of objects or environments, has gained significant attention in various fields, including computer vision, robotics, and virtual reality (Kork et al., 2020). In the specific context of forestry, point cloud data has proven valuable for applications such as plant identification, biomass calculation, and forest monitoring (Lechner et al., 2020).

Generally, according to O'Mahony et al. (2018) in terms of size, cost, spectral sensitivity, and depth sensing capabilities, vision sensor technology has come a long way in recent years and lead to an ascent in the ability of object identification (O'Mahony et al., 2018). These advancements have led to widespread use in a variety of industries, including autonomous vehicles, healthcare -but most importantly for this research- in the fields of forestry and GIS. In continuance to these advancements, as the price of 3D sensors drops and their perceptual capabilities increase, the field of 3D deep learning expands at a rapid pace. Furthermore, the improvement of Convolutional Neural Networks (CNN) has greatly impacted machine learning and cross validation (techniques for using all training data to inform models and forecast performance on new data) (O'Mahony et al., 2018).

However, the primary obstacle in this industry is the time and effort required for creating and annotating datasets, especially for 3D segmentation tasks where every point in 3D space must be labeled correctly as they are essential components in the process of data mining for autonomous annotating processes (O'Mahony et al., 2018, 2019). These challenges that come with these data, are that significant, to the point that several authors describe them as 'dreadful' and 'extremely time consuming'. On the other hand, their implementation is recommended, as annotation tools that offer a 2D visual interface fail to accurately capture the boundaries belonging to different classes or categories (O'Mahony et al., 2019). The necessity for annotated data from users for the purpose of training such a system becomes increasingly challenging as the need for massive amounts of data increases (O'Mahony et al., 2018, 2019; Doula et al., 2022).

In their paper, O'Mahony et al. (2018) pointed out that point cloud data come with a high number of points with significantly variable density and without an evident structure are difficult to comprehend, navigate, and choose on a 2D monitor. Additionally, it is reported that running high-quality, universal classifiers is challenging, and often requires more data in 3D than 2D. This happens because of the third dimension in 3D space causes the parameters of classifiers to significantly increase while the peculiar phenomena of 3D space, such as occlusion or changes in point density, produce a multitude of patterns for the same output classes (O'Mahony et al., 2019).

To annotate a 3D point cloud typically with the use of a monitor and software like Cloud Compare, the user is required to rotate the cloud, lock the view, and create a closed polygon that would divide the cloud in half, inside and outside the polygon. This procedure must be repeated until every point cloud was classified under a specific category and points are stacked according to their classification (O'Mahony et al., 2019). This approach is not only considered challenging and time-consuming, but also reduces the perception of visualized data and sacrificing a whole dimension due to hardware (Wirth et al., 2019).

This subsection explored the various challenges of data annotation, particularly in the context of 2D monitor annotation restrictions for 3D point cloud datasets, are substantial and have been recognized as time-consuming and difficult by various authors (O'Mahony et al., 2018, 2019; Doula et al., 2022). Therefore, these addressed inefficiencies inspired the content of this research, to examine alternative concepts over the 2D monitor approaches and find the less

time demanding and accurate annotation approaches.

2.4 Existing tools and approaches

The field of 3D dataset annotation has seen considerable progress in recent years, driven by the need to perform accurate annotations for various applications (Kaur & Singh, 2022). This chapter explores different tools and approaches used for annotating 3D point cloud datasets, from early references like Antsy 3D to modern innovations such as VR-based systems (Veit et al., 2013; Bacim et al., 2014; Lubos et al., 2014; Wirth et al., 2019). Although some approaches of this subsection used 2D monitor technologies, it was considered essential to mention them, as they were the basis for future applications and some of their characteristics were used in this research (Forsyth & Ponce, 2002; Russel et al., 2009; Wirth et al., 2019). By examining these diverse approaches, this chapter aims to provide valuable insights into the current state and potential future directions of 3D data annotation approaches.

The study and practice of 3D dataset annotation have been of recent development. As Wirth et al. (2019) suggested, one of the first tools for point annotation is the Antsy 3D tool released in 2007, which was a point-and-click annotator for web browsers. This tool can be considered relatively outdated, since it did not offer annotating options for 3D data, but it was considered a reference point and source of inspiration for contemporary labeling methods and approaches, which lead to the state of the art in annotating 3D data.

Furthermore, in the work of Russell et al. (2009), based on the previous concept of the Antsy 3D tool, the LabelMe tool offered online annotation capabilities as a toolbox in the programming software of MATLAB. This tool offered a database and web-based annotation tool for collaborative two-dimensional picture annotation. The web application included features such as polygon drawing, picture queries, and database exploration (Russel et al., 2009). This is considered to have significantly contributed to the science of computer vision, which is related to computers' high-level interpretations of digital photos and movies- a topic of research that cuts across many different fields (Forsyth & Ponce, 2002).

A significant addition to this matter was made by Monica et al. (2017) as they created two scenes with different objects and settings to be labelled by users in VR. In their approach, two methods for measuring the amount of time required for scene annotation were presented. The first is their proposed technique based on control points. The second approach involved rectangle selection. The annotators selected a rectangle containing every point in the current direction of view. (Monica et al., 2017; Wirth et al., 2019).

In the work of Veit et al., (2013) a different approach was presented, developed on a Unity3D-based system for marking up 3D points with annotations (Wirth et al., 2019). The developed platform namely Go'Then'Tag gave the ability to the users to annotate 3D point clouds. The two main concepts on which this tool was built were the use of a tracked multi-touch input device and the gradual refining of pre-selections to make the selection process more efficient. The offered capabilities of this tool included a set of features for manipulating and labeling multi-layered 3D data sets, tools for easy modification of subsets of points in a 3-D point cloud, tools for managing and editing tag trees, including support for nested and intersecting tag hierarchies (Veit et al., 2013).

Although the tool developed by Veit et al. (2013) was designed for 3D object manipulation and selection, it is important to point that the annotating process of the data, including the tag and text modification was handled in a 2D environment of a smartphone. This was done, despite studies (Wilkes et al., 2012) indicating that mobile devices with tracked, multi-touch interfaces

offered no functional benefits. It is important to mention that this was considered the state of the art of this study, since through their work, Veit et al., (2013) introduced the concept of performing activities with datasets while using the monitored device as a 3D pointer, and text and tag hierarchy revisions with device's touchscreen. When it comes to the procedure of point selection, both individual point and point group selections were offered to the user. The former selection was made with the use of a pencil tool for assisting accurate selection, while for the latter case geometric shapes such as a cube or a sphere can be employed. Additional adaptations can be made, since the user may adjust both the shape's size and its depth location along the ray, with the use of multi-touch controls (Veit et al., 2013).

Furthermore, in the work of Wirth et al. (2019) a tool for marking cuboid-like entities in point clouds was presented and made accessible to the public as open-source software and in the Unity Asset Store. This platform was made for annotating 3D data in the surroundings autonomous vehicles. Their approach introduced the concept of gamifying data annotation, to solve the issue of low motivation from student assistants to perform data annotation. An evaluation process was conducted on their VR labelling technique in terms of annotation cost and accuracy. Additionally, it is highlighted that there was little empirical research performing thorough evaluations of data annotation in VR (Wirth et al., 2019).

Another crucial factor that was worth investigating was the ability of the user to navigate the VR scene to successfully annotate 3D datasets. To facilitate movements inside the virtual environment, Coffey et al. (2012) introduced a method that employs the utilization of two screens. The user only observes the 3D item in VR, but they may interact with it using many movements on the touch screen. On the second, larger screen, viewers may watch as the user gives their presentation (Coffey et al., 2012 & Wirth et al., 2019). Bacim et al. (2014) and Lubos et al. (2014) developed techniques for adding free-form gesture annotations to point clouds. They concluded that in the context of segmenting more complicated 3D point clouds, users' precision in VR may be greater (Bacim et al., 2014; Lubos et al., 2014). Additionally, in the work of Veit et al., (2013), the use of the smartphones screen has a predominant role in the data handling process and assistance of the point annotation. The annotation of the 3D datasets was assisted by orientating capabilities offered by their tool; that is, sliding gestures rotate the selected node around the x and y axes to change its orientation.

In the work of Hmida et al. (2012), the WiDOP platform was presented, which is able to construct an indexed scene from disorganized 3D point clouds so that the results may be viewed using VR technology. On the other hand, this application is related to semantic web applications and ontology (Hmida et al., 2012; O'Mahony et al., 2019) In addition, Stets et al. 2017 describe a virtual reality application for labeling and managing point cloud datasets, which allows the user to roam around and explore a succession of room-scale point clouds while the application is playing or paused. A tracked hand-held controller allows the user to pick and choose specific regions of a point cloud for tagging. When the points are labeled, they are highlighted with a different color. The labeled points can be followed from one frame to the next with the help of a tracking algorithm (Stets et al., 2017; O'Mahony et al., 2019).

This subsection explored various tools and approaches for data labelling and annotating 3D point cloud datasets, highlighting key influences on the present research. The work of Russel et al. (2009) introduced an application with features like polygon drawing. In combination with Monica et al. (2017) who proposed an approach based on control points-nodes, the creation of a container-based annotation approach was conceptualised in this research. Furthermore, Veit et al. (2013) presented a Unity3D-based system, introducing the idea of using Unity3D and VR for annotation environments. Additionally, their use of a multi-touch input device and efficient

pre-selections further inspired the adoption of the Meta Oculus Quest 2 in this research. Although Veit et al.'s (2013) tool handled annotation in a 2D smartphone environment, their concept of using the device as a 3D pointer and adapting shapes for point selection influenced the development of the sphere selection approach. Additionally, ideas from Hmida et al. (2012) and Stets et al. (2017) impacted the use of VR navigation controls for improved annotation assistance.

2.5 Evaluation of accuracy, time, usability, and workload

There was a limited number of publications discussing users' evaluation processes of 3D point cloud annotation techniques in VR. As it was pointed out by Wirth et al., (2019) that there were no standards for data annotation to compare their evaluation. Additionally, it was highlighted that there was a limited number of publications, referring to evaluation of data annotation in VR. They pointed out that there were no standards for data annotation against which compare their evaluation (Wirth et al., 2019). Furthermore, one of the few references regarding the usability testing of virtual reality applications from Kamińska et al. (2022) pointed out that understanding the concept of usability within the context of VR presents significant challenges.

In the work of Wirth et al. (2019), after the completion of tasks by using their labelling technique in VR, an evaluation process was conducted in terms of annotation time and accuracy. Following the steps of the previous studies (Monica et al., 2017), Wirth et al. (2019) made time measurements of simple annotating tasks from inexperienced students. The users were given a brief amount of time (around 30-45 minutes) to get familiar with the annotating equipment, and afterwards they were introduced to four scenarios of 3D object detection. Additionally, a few hints for efficient data annotation were given.

Furthermore, Wirth et al. (2019) concluded that the assessment of accuracy was rather a challenging process. For assessing this factor, the researchers generated their own Ground Truth (GT), since they had a richer experience as annotators, comparing to the student annotators. Furthermore, they established several requirements to measure accuracy. In this specific case the annotators were asked to contain all points belonging to the current object in a bounding box, exclude all points that do not belong to the current item, and have a form that is as close as feasible to the shape of the physical thing itself. Finally, label precision was evaluated by counting the proportion of objects with correct labels vs those with incorrect ones. In the study of Wirth et al. (2019) the only reported feedback of the users was related to fatigue while annotating data- since the annotation process was made more engaging in virtual reality- and to VR sickness (around one-third of the students) following the experiments (Wirth et al., 2019). As a result, there were no reports regarding the usability of these applications.

Furthermore, according to the ISO 9241 standard from 1998, with further information from ISO/IEC 25010, usability is defined as the effectiveness, efficiency, and satisfaction of specified users in achieving specified goals in a specific context of usage. Learnability, efficiency, memorability, errors, and satisfaction are all important factors in determining a design's usability. Learnability refers to a user's ease in performing specific tasks when dealing with a concept for the first time. Once individuals feel comfortable with the design, efficiency assesses how quickly they can execute activities. Memorability is concerned with the user's ability to regain expertise after a time of inactivity with the design. The evaluation of errors considers the frequency of user errors and the design's ability to assist error recovery. Finally, satisfaction assesses the level of satisfaction and comfort felt by participants while interacting

with the system (Kamińska et al., 2022).

As Kamińska et al. (2022) noted, there are several methods for enhancing usability, but user testing is the most fundamental and beneficial. The source of data for subjective usability testing is the perceived emotions among those participating in a virtual reality experience. Due to the straightforward data acquisition methodology that does not require a laboratory, monitoring devices, or complex biomedical data analysis following measurement, this method is widely utilized by usability researchers. Preparing an effective questionnaire that captures valuable usability data, including follow-up questions, without being excessively lengthy and repetitive is a complex process. Thus, ready-made questionnaires and measures are utilized frequently. A commonly used approach, proposed by both IJsselstein (2013) and Kamińska et al. (2022) is the System Usability Scale (SUS; Brooke, 1996) and in the work of the former author it was utilized for evaluating game-related applications- which is considered relevant with the concept of this research.

In the work of Xi et al. (2023) the concepts of VR and workload estimation are examined. Workload is typically defined as the amount of effort or investment (e.g., physical, mental, emotional) an individual expends to complete a task (Hart, 2006). Factors that can be influential are experience, and skill, as well as external factors such as the nature, novelty, difficulty, and number of tasks a person completes (Hart, 2006; Xi et al., 2023). As Hart (2006) pointed out, workload is an important evaluation factor for the technologies that support virtual and augmented vision. As Xi et al. (2023) explained, numerous evaluation methods, such as objective measures based on indicators of performance and psychophysiological information, as well as metrics focused on subjective experiences, have been developed because of the need to evaluate the workload associated with human-computer interaction. A methodology that can cover the multidimensional nature of workload is the NASA- Task Load Index (NASA-TLX; Hart, 2006).

This subsection presented the evaluation of time and accuracy measurement approaches found in the literature, specifically for 3D point cloud data annotation techniques in VR. Although the sources of these findings were limited (Monica et al., 2017; Wirth et al., 2019; Kamińska et al., 2022; Xi et al., 2023), their concepts influenced greatly this research. The setup of the experiment of this research was like the one of Wirth et al. (2019), with the creation of different stages (introductory phase and annotation phase) and the use of different scenarios for the student annotators to explore. Furthermore, the concept of generating data for establishing standards for meaning accuracy as Ground Truth, was considered an ideal approach for this research. Capturing the user's feedback on VR sickness while annotating was also included in this research, as it reported to be an important limitation for VR applications (Wirth et al., 2019). It is important to note that none of the studies evaluated the labelling process based on the users' experience and workload through qualitative methods. For this reason, further literature investigation on usability and workload on in the broader concept of VR showed that System Usability Scale (SUS; Brooke, 1996) and NASA- Task Load Index (NASA-TLX; Hart, 2006) could be employed.

2.6 Importance of gender reporting in virtual reality studies

MacArthur et al., (2021) through their systematic literature review on virtual reality research, gender, and VR discomfort, pointed the lack of reporting on what is gender assumed to be (e.g., biologically, socially). The findings of their review suggested that none of the found studies addressed gender beyond the traditional binary framework. Furthermore, these studies did not acknowledge the evolving understanding of gender within a wider social and cultural

context (MacArthur et al., 2021). It is unknown the studies offered options for self-identification, in accordance with feminist, queer, and trans studies.

Despite this, gender and other demographic information should be reported to better understand how each person is affected by VR and it is important for at every stage of VR study design (MacArthur et al., 2021). As Peck et al. (2020) suggested, “Demographic information must be included when reporting characteristics of participants, including age, gender, and race/ethnicity, so that readers can accurately interpret the studied population and future meta-analyses of participant demographics can be performed”. Additionally, the significance of broad usability for VR to be relevant in educational settings, science, industry, and more is emphasized by usability research. Despite academics' attempts to improve the VR user experience and usability with design suggestions, few articles examined the potential gender differences in this field (MacArthur et al., 2021).

Since MacArthur et al., (2021) investigated VR discomfort, it is important to mention that when it comes to gender and symptoms of VR discomfort, there is a plethora of articles (Lo & So, 2001; Kim et al., 2005; Shafer et al., 2019) relating these two factors. On the other hand, other studies (Knight & Arns, 2006; MacArthur et al., 2021) reported that male and female participants in VR applications experiences similar level of VR discomfort. Still, there is evidence that gender might influence the VR experience of users (MacArthur et al., 2021). As a result, defining potential relations of gender with other factors of VR experiences such as usability and performance was considered interesting for this thesis.

This subsection highlighted the importance of incorporating gender in virtual reality research and the possible relations between gender and VR sickness were mentioned. As discussed in the previous subsection, feedback on VR sickness was included in this research. Although VR sickness is a known limitation (Wirth et al., 2019), the scope of this research does not aim to explore in detail its effects on performance and usability, nor to relate this factor to the gender of participants. This is one of the future research recommendations. On the other hand, since there is known relations between the factor of gender and VR and existing literature gaps on usability research, it was considered important to investigate relations between gender (among other factors), usability, and performance.

2.7 Literature Review summary and discussion

This subsection presents the findings and conclusions from the conducted literature review and existing literature gaps. Furthermore, it summarizes the key concepts and approaches that influenced the different components and stages of this research and the research gaps that this study aims to address.

To begin with, the literature review explored different environmental applications and identified key features of Embodied Digital Twins that are relevant to this research. These applications demonstrated the potential for successful applications in environmental studies, as they utilized a game-like experience to engage users in tasks while interacting with components of virtual environmental (Sajjadi et al., 2014; Klippel et al. 2021). This design influenced the concept of this research, by incorporating interactive experiences in the annotation evaluation process, giving the freedom to the participants to navigate through different annotation scenes, features, and tasks.

One of the most significant challenges highlighted in the literature revolved around data annotation, particularly within the context of 3D point cloud datasets constrained by 2D

monitor limitations. Numerous authors (O'Mahony et al., 2018, 2019; Doula et al., 2022) acknowledged the time-consuming and challenging nature of the traditional 2D monitor annotation approaches. These known limitations were proven as motivation to explore alternative annotation approaches that had the potential to influence this research.

Further investigation into alternative approaches and existing approaches had a great impact on the conceptualization of this research's approach. The Sphere Selection Approach was inspired from the work of Veit et al. (2013) and incorporated the concept of using gaming engines and 3D point cloud data annotation techniques. Veit et al. (2013) presented a Unity3D-based system and explored the potential of using Unity3D and VR for efficient and interactive annotation environments. While the tool of Veit et al. was designed for annotation in a 2D smartphone environment, their idea of using the device as a 3D pointer and adapting shapes for point selection influenced the conceptualization of the Sphere Selection Approach. Additionally, the integration of Meta Oculus Quest 2, a VR device with multi-touch input capabilities, was influenced by their efficient pre-selections and 3D pointer concept.

The Container Creation Approach was influenced by the work of Russel et al. (2009) and Monica et al. (2017). Both publications introduced concepts related to control point-nodes for annotation. Russel et al. (2009) presented an application with polygon drawing features, while Monica et al. (2017) proposed an approach based on control point-nodes for annotation tasks. The combination of these concepts influenced the conceptualization of the Container Creation Approach.

Overall, the findings from the reviewed literature indicated three major literature gaps. Firstly, although the discipline of data annotation is emerging, there are few publications related to 3D data annotation in environmental studies, and more specifically in forestry applications. Secondly, there was a lack of publication exploring and comparing different annotating techniques in 3D and performance. Finally, as pointed out by Wirth et al. (2019) and based on the literature review, the evaluation of 3D data annotation methods and user's evaluation are not present in most of the studies. It is also important to point out that studies such as the one of Monica et al. (2017), who evaluated the labelling process based on the user's experience, were based solely on qualitative methods.

In the study of Wirth et al. (2019) the only reported feedback of the users was related to fatigue while annotating data since the annotation process was made more physically engaging in virtual reality. Furthermore, the possible relation of performance and usability with individual differences (gender) for 3D data annotation has not been investigated. Therefore, the objective of this research is to address the gaps related to performance comparison of annotating techniques in VR for environmental 3D point clouds of and capturing users' experience thought evaluating processes. This knowledge is considered essential for the design of future time and accuracy efficient, user- friendly tools for annotating in VR. Finally, these outcomes could lead to design criteria of 3D VR tree annotation tools and application and the participant application usability study.

3. Methodology

Following the literature review discussion, where the literature gaps and the key concepts which influenced this study were highlighted, in this chapter the methodology is presented. To hint all the details of this chapter, it was considered efficient to divide the methodology of this research into different sections, including the annotation environment and datasets, the different software utilized for specific tasks, the annotation tools of WANDER Lab, the structure of the experiment and users' evaluation process, the data collection, the calculation of the SUS and NASA-TLX scores, and finally the calculation of the time and accuracy factors.

3.3 3D Point cloud data

The components of these environments belonged to a collection of tropical trees from three different study sites, namely Indonesia (peat swamp forest in Central Kalimantan, Borneo), Peru (lowland tropical moist forest in Madre de Dios), and Guyana (lowland tropical moist forest in Cayuni-Mazaruni) (Gonzalez de Tanago et al., 2018). One important characteristic of this dataset is that the point-cloud did not include any RGB color information. These datasets will be used for performing annotations, leading to answer of RQ1.1.

This selected dataset from Gonzalez de Tanago et al. (2018) was considered optimal for this study since it offered a wide selection of tropical trees with a relatively simplified structure. This selection was expected to make the annotation process of the different parts of the trees straightforward for the users (Figure 1). To be more specific, other types of trees, such as coniferous species, would be challenging for the users to annotate, as the dense canopy would not allow the distinguishing of the different parts of the trees just from a point cloud without RGB information. More complex trees could be a point of interest for future research. Additionally, the simplified structure of these tropical trees was an asset for this research, not only for the annotation process but also for the user's navigation and inspection of the dataset. Complex tree structures could lead to confusion in navigation, as the annotators would have to navigate constantly through the point cloud, possibly leading to confusion, downgrading the user's experience, and compromising the flow of the annotation process. It is important to mention that the different trees will present differences in their structure (e.g., with long vs. short tree trunks, or with leaves vs. without leaves).

From the collection of Gonzalez de Tanago et al. (2018), four tree point clouds were selected to be included in the annotation environment. This selection was performed after close inspection of the different trees in this dataset, as the goal was to include trees of different sizes, geometrical features, number of branches, canopy density, and overall complexity. This number was considered optimal, as the users will be given an adequate number of cases to perform the annotation tasks and form a complete picture of the capabilities of the provided methods. Additionally, a higher number of trees the point clouds was not considered necessary, as the possibility of discomfort and motion sickness is expected the longer the user wears the spends time in VR (Kim et al., 2018). This factor is discussed in the discussion as a limitation of annotating in VR.

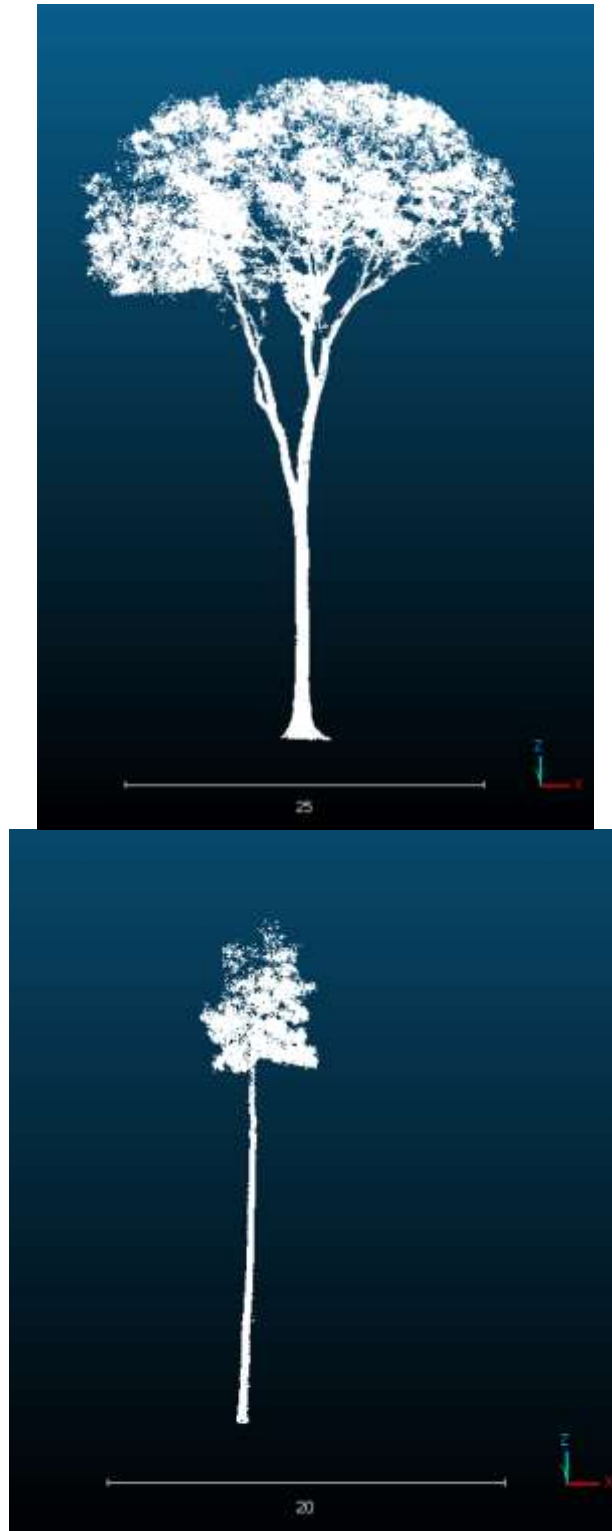


Figure 1: Examples of two-point cloud trees used from the dataset of Gonzalez de Tanago et al. (2018).

3.4 Software

This subsection provides a summary of the software utilized for the setup of the annotation environment, the manipulation of the datasets, and the result analysis of the user's evaluation. The implementation of Cloud Compare allowed the visualization, creation of ground truth point clouds, and extraction in a usable format for Unity (for measuring accuracy and answering RQ1.1). Unity played the most significant role in the creation of the annotation environments and the VR experience, with the point clouds and the necessary packages and tools from the WANDER Lab of Wageningen University (Knuiman, 2022). Services provided by Google Forms are utilized to capture the user's experience through several questions. A detailed description of the components of these questionnaires (for capturing the user's feedback and experience and answering RQ2.1 and RQ2.3) and their methodology is described in Appendix II and the respective subsection of the methodology. Finally, the statistical analysis of the results was completed with SPSS.

3.4.1 Cloud Compare

Cloud Compare is an open-source 3D point cloud and mesh processing software designed for geospatial analysis, 3D modeling, and visualization applications. It offers an extensible set of tools for manipulating, comparing, and registering point clouds produced by 3D scanning devices, LiDAR data, or photogrammetry techniques. In general software enables users to perform tasks including point cloud alignment, filtering, segmentation, and surface reconstruction due to its intuitive interface and extensive functionality (Duarte et al., 2017). Specifically for this research, it was used to create ground truth point cloud data, to answer RQ1.1.

3.4.2 Unity

Unity is a robust and widely utilized cross-platform game development engine that has applications outside of the gaming industry. Unity is best known for its ability to create interactive video games, but it also has significant potential for applications in science (Solomon, 2018). Its real-time rendering capabilities, intuitive interface, and support for multiple platforms make it an appealing tool for researchers and scientists (Gregory, 2018). The incorporation of VR technology into Unity increases its potential for scientific applications. Unity's robust support for generating immersive VR experiences makes it an indispensable instrument for scientific research and visualization in VR environments (Soto-Martin et al., 2020). Overall, the use of Unity and the packages of the WANDER Lab of Wageningen University (Knuiman, 2022) offered a range of benefits for this research, such as increased navigability and advanced interactive tools with the use of the Meta Oculus Quest 2 headset. Specifically, for this research this software was used to answer RQ1.1

3.4.3 SPSS

SPSS (Statistical Package for the Social Sciences) is a comprehensive and widely used statistical software developed to analyze and interpret data in many fields of research. It provides various statistical techniques for data manipulation, descriptive analysis, hypothesis testing, and advanced modeling, all accessible through an intuitive interface (Rahman, 2021). In this research, SPSS offered an extensive analysis of the users' evaluation data of the VR experience and annotation tool and unveiled the relationship between distinct factors. Specifically, it was used to perform ANOVA tests to answer RQ1.2 and RQ2.2.

3.5 Development and annotation environment

This research's scope was limited to the study of the labelling approaches' experience. Although this research included the generation and set-up of a simplistic 3D environment in Unity, the development of the annotation tool packages was done by Knuiman (2022), a developer of the WANDER Lab of Wageningen University. Several meetings were conducted in order to finalize the annotation tools. The scope of this project did not include the development of a fully functional toolbox, but rather a simplistic and limited applied perspective as a means of capturing and valuing differences in performance and user experience.

For the generation of the 3D environment in Unity, a few simple steps were followed, including the creation of a new Unity project, and simply importing the necessary packages provided by Knuiman (2022) and the selected tree point clouds (Gonzalez de Tanago et al., 2018) (Figure X). More information and detailed guidelines related to the functionalities and settings of the provided packages can be found in Appendix I.

It is important to mention that the case study environment, in which users performed the annotation of a tree point cloud, did not depict an actual representation of a physical environment but rather an artificial one. The advantage of this design choice is that it provides a highly controlled representation where participants solely focus on annotation activities. The environment was simplified without any other environmental details or processes, except for the tree point clouds. This representation allowed the users to perform the annotation process for the different tree cases and capture their performance during the annotation process. Additionally, it was considered important to exclude any objects and artificial representations that were not related to the tree point clouds, as the scope of this research is limited to the implementation of the annotation methods, their evaluation and not the reconstruction of an accurate environmental model.

3.6 WANDER Lab's annotation tools

This chapter provides a complete overview of WANDER Lab's annotation tool features (Knuiman, 2022). Information regarding the setup of the Unity project and the necessary packages can be found in Appendix I, as they are not related to the scope and the methodology of this project. Firstly, the general functionalities of each Meta Oculus Quest 2 controller are described, followed by the navigation feature provided in this VR experience. These subsections are followed by a description of each annotation methodology utilized in this research. Finally, the experiment structure and users' evaluation process, as well as the data analysis methodology.

The first step of this research was using a simplified annotating tool for 3D environments in the Unity game engine, owned by the WANDER Lab of Wageningen University, The Netherlands. This tool offered two annotating solutions, namely the Sphere Selection Approach and the Container Creation Approach, respectively. As briefly described in the Research goals and limitations, the former allowed a free-style selection of parts of the point cloud surface to define an area of interest, while the latter one—inspired by the annotating approach of Doula et al. (2022)—allowed the user to pick the point cloud points within a container such as a box. This method was developed specifically for this study. In the second method, like the approach of Doula et al. (2022), to annotate an object of interest, the user surrounds it with multiple points to determine the perimeter of the bounding box. Both approaches involve selecting a semantic label prior to selecting the points.

This tool is available in the form of a Unity package in the repository of the WANDER Lab (Knuiman, 2022). Although allowing participants to choose the number of categories based on their understanding could be an option, the disadvantage of this approach is that it could result in highly confusing data from the participants, thereby increasing the difficulty of data analysis. In this research, three categories will be available, namely, tree trunk, tree branches, and leaves.

3.6.1 Meta Oculus 2 controllers features

In this subsection WANDER Lab's annotation tools packages for Unity and Meta Oculus Quest 2 came with two types of features, for point cloud annotation and for navigation. Each Meta Oculus Quest 2 controller had assigned features to each button. The ones on the left controller offered controls related to the annotation process, while the ones on the right assisted the navigation of the user around the VR scene. Both the Sphere Selection and Container Creation approaches came with some unique features assigned to the left controller. The specific interaction features are presented in the following subsections.

For each annotation method below were the common features assigned to the left controller (Figure 2):

- "Minimizing and Maximizing" the radius of the annotation laser pointer assigned to the thumb stick. While using the controller during the VR experience, the users could move around the laser beamer, which helped the annotation. This pointer was an important assisting feature for both the Sphere Selection Approach and Container Creation Approach. Changing the radius helped adjust the length of the laser pointer, to accurately perform the annotation. Details on each annotation technique are presented in the respective chapter.
- "Switch annotation method" feature assigned to the X button. By clicking this button, the user could instantly switch between the two annotation methods during the experiment. This feature was necessary during the combination of the two approaches.
- "Annotate" by holding the front trigger. While annotating with Sphere Selection Approach, the user had to press the trigger continuously and move the controller for the laser pointer to capture all the desired points of the point cloud. On the other hand, while annotating with the Container Creation Approach, by pressing the front trigger, nodes were assigned that corresponded to the edges of the desired container.
- "Classifier" by holding the side trigger. This button allowed the user to instantly shift between the three annotating classes, namely tree trunk, tree branches, and leaves.

The features of the right controller were (Figure 2):

- "Fly/ Teleport distance and rotation" assigned to the thumb stick. The user could either fly freely around the scene by aiming the beam in the desired direction and accelerating, or teleport and rotate while standing still at a specific spot of the VR environment, depending on the enabled navigation mode.
- "Teleport" feature assigned to the Front Trigger. This button allowed the users to instantly teleport forwards, offering the user a faster navigation method. By using the teleport feature in combination with the rotation feature, the user can perform time efficient navigation, which offers the potential of detailed annotation.
- Reverse feature by using the side trigger. This offers the possibility of moving backward

- a feature like the teleport
- Change the Navigation feature assigned to button A. Through this action the user can shift between the Fly and Teleport navigation style.

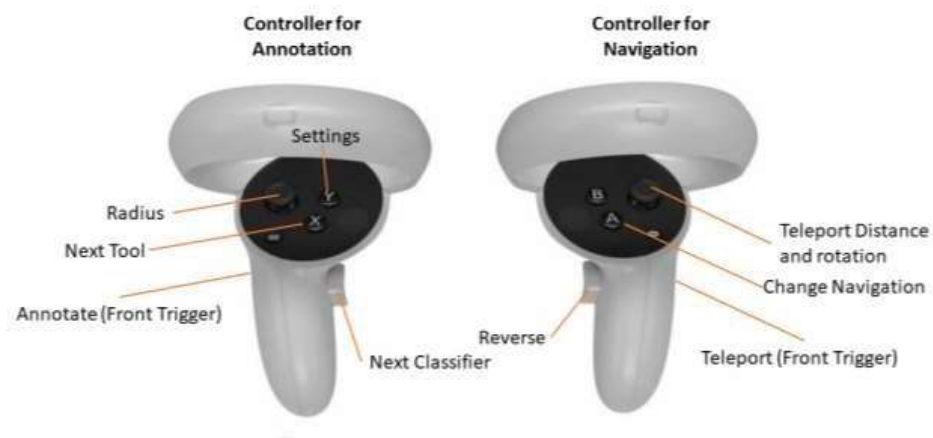


Figure 2: Different features assigned to the Oculus controllers.

3.6.2 Annotating with the Sphere Selection Approach

This subsection describes the annotation process and characteristics of the Sphere Selection approach. This annotation approach was based on manually choosing the desired points of the tree point cloud with the use of an annotating sphere-shaped beamer on the left controller. The first step of the annotation process was to identify the class- label that the user wanted to annotate. As mentioned in the previous subsection for the left controller features, this could be done by pressing the side trigger button. Once the participants pressed the side trigger button, the color of the beamer on the left controller changed, indicating that the controller was annotating a specific class. The same classes applied for the Sphere Selection Approach, the Container Creation Approach, and their combinations. The assigned color indicated the following classes:

- Brown color: tree trunk category
- Light Brown color: tree branches category
- Green color: leaves category

Once the participants decided the class that they wanted to annotate with, they could navigate towards the tree point cloud for the annotating beamer of the left controller to touch the points of the point cloud (Figure 3). Another way of making the beamer contact the points of the point cloud was to expand or minimize the radius to the beamer extension by using the Minimizing and Maximizing feature of the thumb stick on the left controller.

This approach's most distinguishing characteristic was the ability to change the sphere-shaped beamer's size at any time during the annotation process. For this reason, the approach was given the name Sphere Selection. This could be done by moving the joystick of the left controller left to reduce the size of the sphere and right to increase it. Once the size of the sphere-shaped beamer and the radius were set, the participant could hold the left controller's front trigger to annotate. By continuously pressing the front trigger, extended parts of the tree point cloud were labeled. On the other hand, by pressing the trigger instantly, smaller parts of the point cloud were annotated with greater accuracy and detail. The combination of the sphere shape, beamer radius, and annotation were highly dependent on the approach and personal strategy of each participant.

It is important to mention that the annotation on the point cloud could only be achieved while the left controller beamer contacted the points of the tree point cloud. As a result, this process needed to be precise and detailed. In case the participant wished to remove the given label from an already annotated part, by pressing the side trigger of the left controller, a fourth color category was available. While using this category, depicted by the white color, the sphere-shaped beamer was converted into an eraser. By following the exact same annotation style and pressing the front trigger, the parts of the annotated point cloud that encountered the beamer became white (unlabeled).

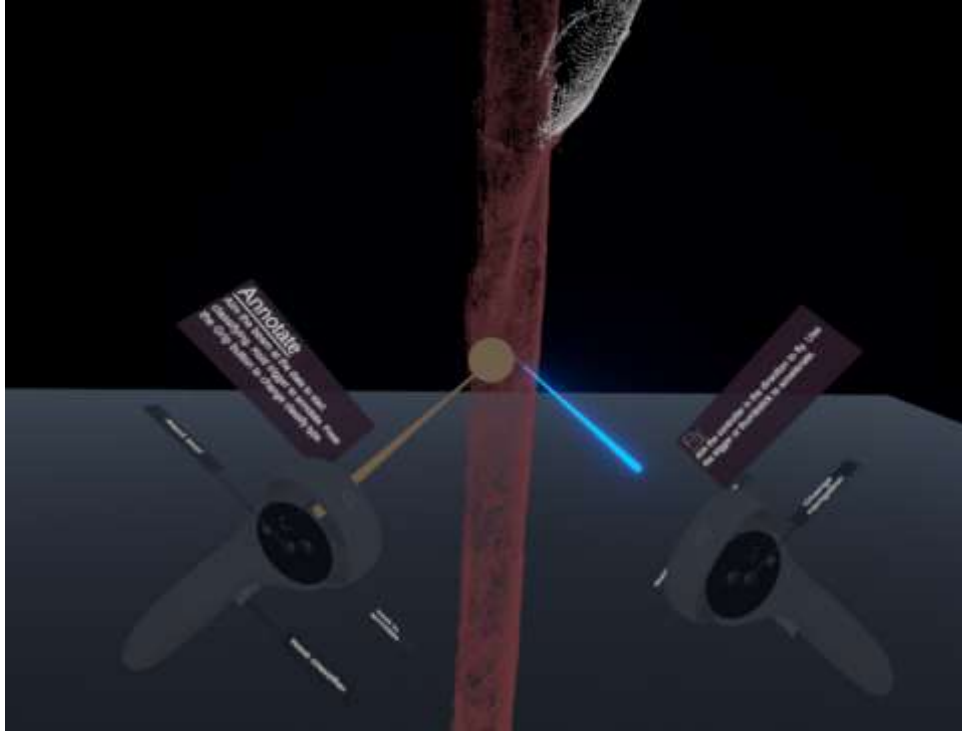


Figure3: Annotating with Sphere Selection Approach. The point of the point cloud represents the annotated area.

3.6.3 Annotating with the Container Creation Approach

The second method involved the creation of containers of different shapes and sizes by placing several nodes inside or around the tree point cloud. For this reason, this methodology was referred to under the name container Creation. Once a container was created, the points of the point cloud included in the container were assigned a specific class. Like the Sphere Selection Approach, the user had the option to choose between the different classes before starting, during the creation, or before placing the last node and completing the desired shape of the container. A major difference was that the fourth class used as an eraser in the Sphere Selection Approach is not present in this methodology. The users needed to add at least three nodes to create a container.

Once the last one is placed, the annotators can finalize the container by pressing the B button on the right controller. The assignment of each node could be done by aiming the beamer of the left controller inside or around the desired part of the tree point cloud and pressing the front trigger once for the assignment of each point (Figure 4). Additionally, the tools allowed the annotators to place the nodes anywhere they desire around the scene. Once the participants finalized the container's creation, all the tree point clouds were selected and labeled with the respective category.

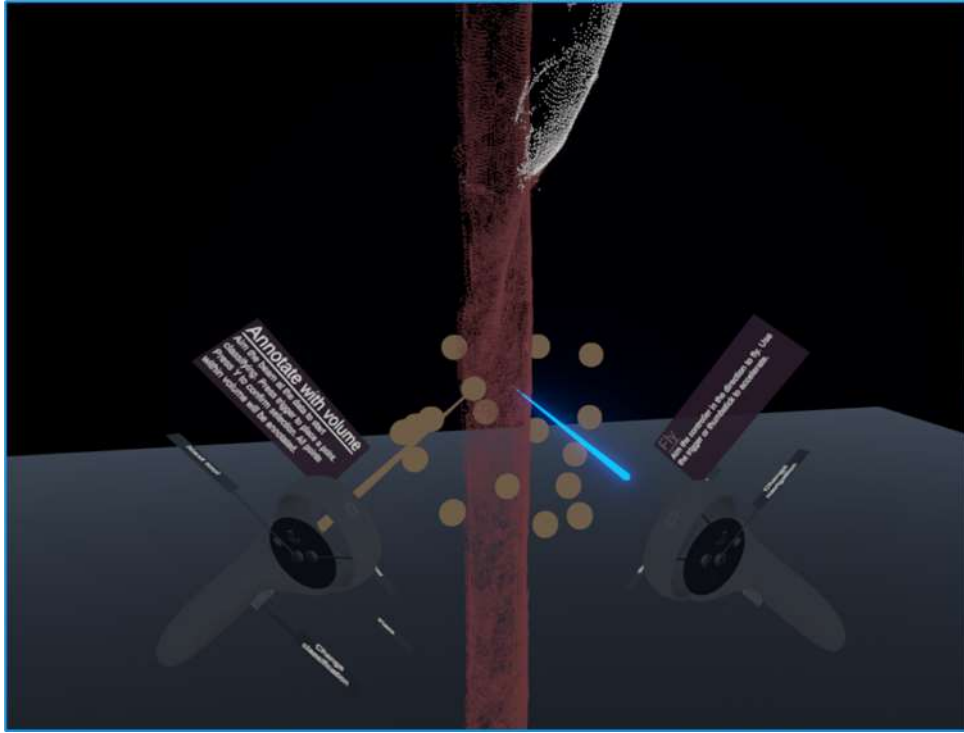


Figure 4: Annotating with the Container Selection approach. The point of the point cloud represents the annotated area.

3.6.4 Annotating with the combination of the two approaches

The annotation by using the combination of the two approaches shared the same characteristics as the Sphere Selection Methodology and the Container Creation Approach. At any moment during the annotation process the participants were able to switch between the functionalities of each approach. The only restriction was that to switch, the users had to finalize the creation of the containers in case they were using the Container Creation Approach.

3.7 Experiment structure and users' evaluation process

This subsection provides information on the different phases of the experiment, the overall evaluation processes utilized for capturing the users' experience and measuring time and accuracy. The group of participants comprised of 15 individuals with diverse scientific backgrounds and positions (bachelor students, master students, researchers, and professors). The requirement for participating in this VR experience was a basic understanding of the different tree components and the willingness to explore the capabilities of VR annotation.

3.7.1 Introductory Phase

A brief introduction was provided to the participants with a brief overview of the concept of this research, the goals and the tasks they had to complete. More specifically, during the experiment's introductory phase, it was considered important not to expose any information related to the tree point clouds data, as the possibility of bias towards the annotation tasks had to be eliminated. On the other hand, it is considered necessary that the user get familiar with and gain some hands-on experience on the Meta Oculus Quest 2 equipment and the different functionalities before the actual conduction of this experiment. It was believed that after an introduction, the users had to spend some time exploring the functionalities of annotation and navigation.

For this reason, after the brief introduction, the users had to explore a sample scene and annotate a point cloud irrelevant to the actual tree point clouds. The point cloud provided during the testing phase was not a file related to forestry, but specifically a representation of a car. Exposing users to a similar point cloud used in actual phases of the experiment could bias the users and compromise the integrity of the experiment. Their performance (time and accuracy) was measured at that stage of the experiment. The introduction stage was brief (a maximum of 5 minutes) as the purpose of this stage was not an in-depth exploration of the capabilities but to gain a general perspective to be able to perform the actual tasks of the experiment. This was since it was expected that some users might feel discomfort towards the final annotating scenes, as at that point the users will have already been annotating in VR for a considerable amount of time.

3.7.2 Annotation Phase

Once the introductory part of the experiment ended, the participants were exposed to the actual tree point cloud datasets and the process of evaluating the given approaches started. Each participant had to annotate three different point clouds, each with one of the employed approaches and their combination as the third approach.

As described above the selected datasets from the collection of Gonzalez de Tanago et al (2018) were given to the users to annotate with gradual complexity as it is expected that the differences in the complexity of the given point cloud, might oppose different challenges during the annotation process. Differentiations in complexity offered the users a plethora of scenarios in which they could gain a complete perspective of the advantages and disadvantages of each individual approach, as well as the combination of them. Each participant's time to fully annotate each point cloud was recorded for each scene.

It is worth noting that during the conduction of the experiment, each participant had to annotate the given point cloud by using a different annotating technique, compared to the previous participant. For instance, Participant 1 annotated with the Sphere Selection Approach in the first scene, and the Container Creation Method for the second scene. The following participant annotated the first scene by employing the Container Creation Method, and then with Sphere Selection Method. The third scene for every participant was the combination of the two methods. It was important not a specific method in the same sequence for every participant (Figure 5).

	Scene 1	Scene 2	Scene 3
Participant 1	<div>Sphere Selection</div> <div>Tree A</div>	<div>Container Creation</div> <div>Tree B</div>	<div>Combination</div> <div>Tree C</div>
Participant 2	<div>Container Creation</div> <div>Tree B</div>	<div>Sphere Selection</div> <div>Tree A</div>	<div>Combination</div> <div>Tree D</div>
Participant 3	<div>Sphere Selection</div> <div>Tree B</div>	<div>Container Creation</div> <div>Tree A</div>	<div>Combination</div> <div>Tree C</div>
Participant 4	<div>Container Creation</div> <div>Tree A</div>	<div>Sphere Selection</div> <div>Tree B</div>	<div>Combination</div> <div>Tree D</div>
...			

Figure 5: Experiment log.

A similar approach was followed for the point cloud files to each of the three stages of the experiment (Figure 5). The collection of trees used for this research presented distinctive characteristics but similar difficulty during the experiment's first two stages. For the first two scenes of each participant, two tree point clouds were selected. Similarly, to the sequence of the annotating approaches, different point clouds were annotated in each scene for each participant. In the case of the approach of the experiment's final stage, meaning the combination of Sphere Selection and Container Creation methods, another two trees were selected for the annotation process (Figure 5). These trees present higher complexity compared to the point clouds of the first two scenes (the potential effect of this design choice is discussed in the Discussion chapter). It was expected that during the experiment's final scene, the users would be more experienced. Thus, the most demanding tasks in terms of approach combined with the tree point clouds, were given in the end. These combinations of the sequences of approaches and different tree point clouds created, to eliminate the creation any patterns through the results of the experiment.

3.7.3 Data Collection and Users' Evaluation

In this subsection the type of data collected before, during and after the annotation phase are presented. These factors were the time needed for the annotation of each tree point cloud, the annotated tree point clouds were collected for estimating the accuracy and finally the users' demographics and experience were recorded through the questionnaires.

During the introductory phase, users' demographic characteristics (age, gender, educational level) were collected to study the effects of individual differences (Figure 6). The responses were gathered via Google Forms. This information was collected to provide an overview of participant's background. The gender factor was used to answer both RQ1.2 and RQ2.2. The questions of each questionnaire can be found in Appendix 2.

For the estimation of the accuracy, the necessary data were the annotated point cloud of each participant captured from each scene (Figure 6). Those were saved separately, to be visually evaluated afterwards. Details over this procedure can be found on the following subsection. Time was measured during the annotation process of each scene. These were collected used for statical analysis to answer the RQ1.1, in combination with the feedback from the final questionnaire.

When it comes to evaluating user performance and experience, different approaches could be followed based on the number of study participants as it is already discussed in the literature review (Drew et al., 2018). In this study, the content of the questionnaires was influenced by related work in immersive virtual reality, such as Kamińska et al. (2022) regarding the usability testing of virtual reality applications. As pointed out in the previous chapters, there were no specific evaluating techniques that could be considered optimal for studies related to VR. As a result, in this study different evaluation methods for the different steps of the evaluation process were employed.

For collecting the data regarding the user's evaluation, after each scene, once the participant completed the annotation with each approach, they had to answer two questionnaires after completing each annotation scene. The first was a collection of questions for calculating the score of the System Usability Scale (SUS; Brooke, 1996) and the second one for the Nasa-Task Load Index (NASA-TLX; Hart, 2006). In total, six questionnaires (two for each approach) were filled by each participant. These questionnaires can be found in the Appendix 2.

Finally, a questionnaire with general questions related to discomfort (in a 1 to 5 scale) during the VR and open questions for commenting each approach were provided to participants. After the calculation of the System Usability Scale (SUS; Brooke, 1996) and Nasa-Task Load Index (NASA-TLX; Hart, 2006) scores, the results were used for statical analysis to answer the RQ2.1, in combination with the feedback from the final questionnaire.

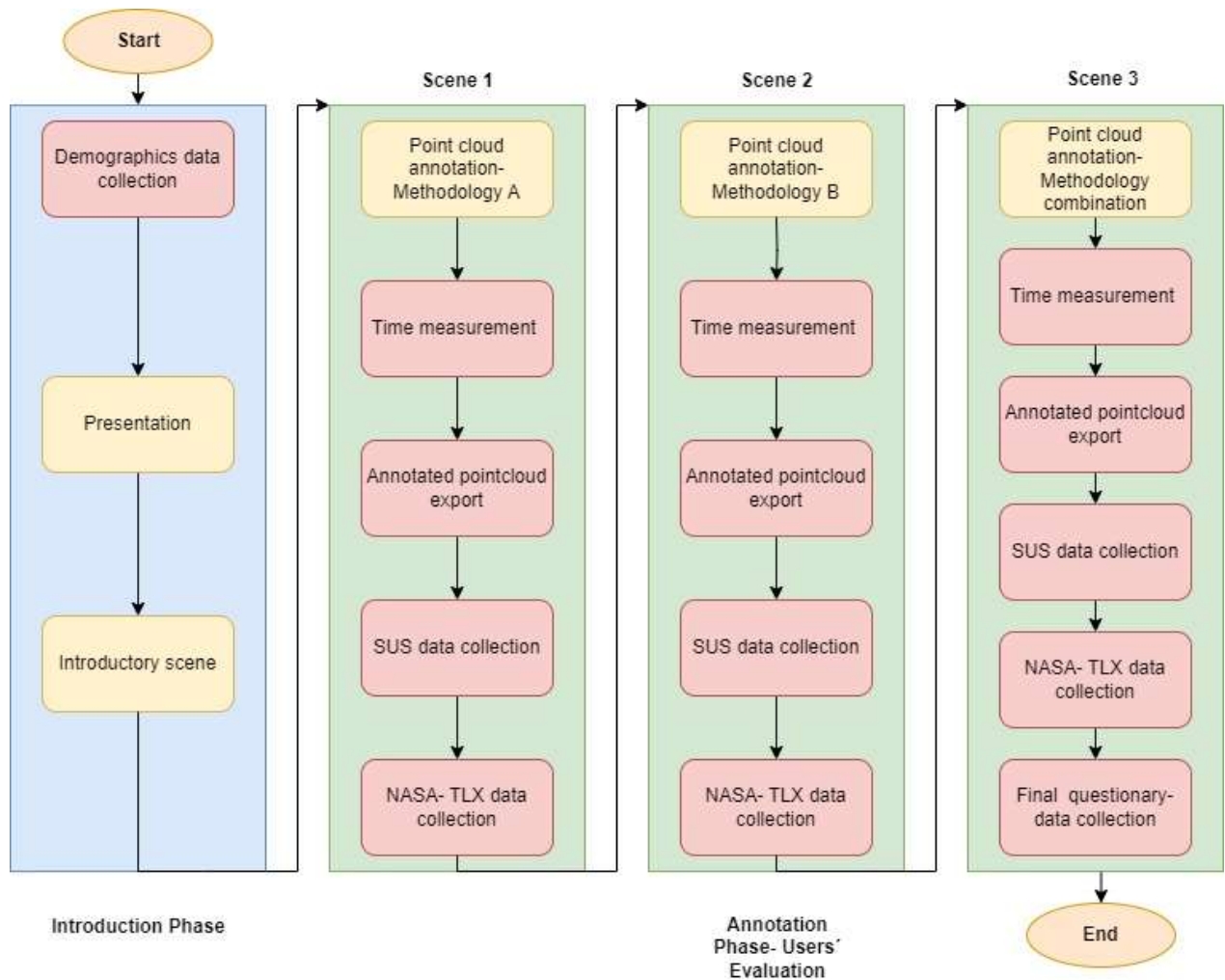


Figure 6: Flowchart of the experiment phases. The yellow boxes represent general stages of the experiment, while red boxes phases where data were collected.

3.7.4 System Usability Scale and Nasa Task Load Index calculation

The System Usability Scale (SUS) by Brooke (1995), was implemented to evaluate the proposed methodology's efficiency of use and user satisfaction. The SUS score provided valuable insight into the produced application's effectiveness, efficiency, and usability. In general, the higher the score, the greater was the workload that the users had to take on during the annotation of the different scenes using the three different approaches.

According to the author, the SUS score was calculated by adding the weighted scores of each question (0–4) for each participant. Subtracting 1 from the scale position provided the answers for questions with odd numbers (1, 3, 5, 7, and 9) 41, whereas subtracting 5 from the scale position yielded the answers for questions with even numbers (2, 4, 6, 8, and 10). The final SUS score calculated by multiplying the total of these individual values by 2.5 (Brooke, 1995). Finally, the efficiency of the examined approaches was calculated by adding together the participants' SUS ratings and dividing them by 15, which is the total number of participants.

For the calculation of the Nasa Task Load Index, after the collection of the data, the values were normalized to a scale from 0 to 100 (NASA-TLX; Hart, 2006). As the methodology suggests the higher the score, the greater is the workload that the user's had to take on during the annotation of the different scenes using the three different approaches.

3.7.5 Measuring time and accuracy calculation

Time measurement for each given task is considered a straightforward process. Once the users were ready to start the annotation process and all the parameters were set, the time needed was captured, until the moment the users declared that they completed the given task. It is important to notice that each participant had to complete every given task without any break in between. After the completion of each annotation round, the time measurements were recorded.

Moving on to the accuracy factor, to assess the annotation results of each user, it was considered necessary to have a ground truth to compare them. For this purpose, the tree-point cloud files, used in the annotation process, were separately annotated in a 2D environment by using Cloud Compare software without time restrictions. To make this comparison meaningful and solid, each part of the tree to be used as ground truth was carefully and methodically annotated. Afterwards, each file was closely examined and approved by a member of the research staff of Wageningen University, considered to be an expert in Virtual Reality and point cloud annotation techniques.

More specifically, for the calculation of the accuracy factor, each annotated point cloud was closely, and side-by-side compared with the ground truth point clouds. The comparison was based on the classes used for annotation, the detail of the annotation and the existence of unlabeled spots. The more similar the results of the users' annotations, compared to the ones of the ground truth files, the higher the given score. All these factors were examined from different angles and perspectives by the author of this thesis and another GIMA student with a major in Forestry. After the inspection phase was conducted, each tree point cloud was rated on a scale from 1 to 5, ranging from poor to excellent results. These results were recorded and imported to SPSS for the statistical analysis of the accuracy factor.

3.8 Data Analysis

In this section the processes followed for statistical analysis with SPSS are presented. These included multiple comparison analysis, like one-way ANOVA to examine differences among the results of the different approaches, and two-way ANOVA analysis to assess the effects of the individual differences.

Specifically, One-Way ANOVA is a statistical test used to determine whether there are statistically significant differences between the means of two or more groups (Ross & Willson, 2017). In this research this test included the results of SUS and NASA- TLX scores, time, and accuracy for each of the of the three approaches. This was used to answer RQ1.1 and RQ2.1.

Finally, two-way ANOVA tests were conducted. Two-Way ANOVA is an extension of one-way ANOVA that can be used to analyze the impact of two independent categorical variables (factors) on a continuous dependent variable. It is possible to determine not only the principal effects of each factor, but also the presence of an interaction effect between the two factors (Corain & Salmaso, 2007). In this research this test was performed to study the effects of:

- The Approach and [individual difference factor] on the SUS scores
- The Approach and the [individual difference factor] on the NASA- TLX scores
- The Approach and the [individual difference factor] on the Time variable
- The Approach and the [individual difference factor] on the Accuracy scores

This test was conducted to answer RQ1.2 and RQ2.2.

4. Results

This chapter includes the results of the statistical analyses and their interpretation for the System Usability Scale (SUS), NASA Task Load Index (NASA-TLX), Time, Accuracy, and multiple comparison tests, for answering RQ1.1, R1.2, RQ2.1, RG2.2. Finally, it includes a summary of the users' feedback from the open questions of the final questionnaire, for answering RQ2.3.

4.1 Participants

There was a total of 15 people that filled who participated in the experiment. There were 9 males and 6 females in the sample, with 60% being male and 40% being female. No one either chose not to disclose their gender or identify as non-binary nor other. The ages of the respondents ranged from 24 to 29, with a mean age of 25.7 years.

Regarding the educational level of the participants, 53.3% declared that they hold a bachelor's degree, while 46.7% stated that they hold a master's degree. None of the participants happened to have education related to the rest of the categories namely less than high school, high school diploma or equivalent, some college or associate degree and doctorate degree.

Regarding the familiarity of the users with virtual reality experiences, the majority declared that either never had a virtual reality experience or consider themselves as beginners.

It is important to mention that the lack of variety of answers regarding the factors of age groups, educational level, and familiarity in VR, lead to the decision to investigate only the individual difference of gender. In future studies, with a higher number of participants, more factors related to individual differences should be investigated.

4.2 Measure of simulation sickness

To capture the possible effects of the virtual experience of the users, once they finished the experiment, they were asked to provide details regarding expected effects related to simulation sickness during their experience and general discomfort after finalizing all the given tasks. The participants shared their experience by answering the two questions, which later were combined into one. In the former case, 40% declaring they felt moderate simulation sickness and discomfort, followed by 33.3% reporting mild effects, 20% reporting significant effects and 6.67% reporting no effects. Finally, none of the participants reported severe simulation sickness and discomfort.

4.3 System Usability Scale

The results of the one-way ANOVA indicated significant differences among the approaches for the SUS scores ($F(2, 42) = 14.748, p < .001$). Post hoc tests revealed that Sphere Selection approach had significantly higher SUS scores compared to the Container Creation approach (mean difference = 21.2333, $SE = 3.3572, p < .001$) and the combination of the two (mean difference = 13.0667, $SE = 3.7840, p = .007$). The Container Creation approach had significantly lower SUS scores compared to Sphere Selection approach (mean difference = -21.2333, $SE = 3.3572, p < .001$) (Figure 7).

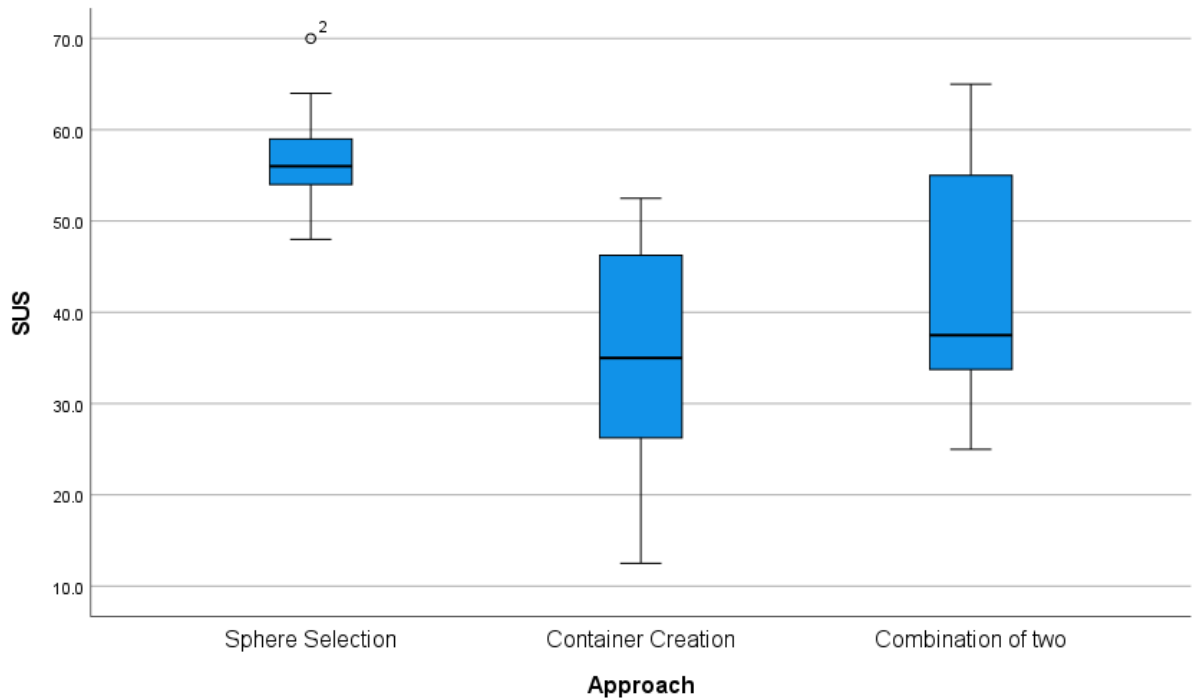


Figure 7: SUS scores for each methodology.

4.4 NASA Task Load Index

Furthermore, for the NASA Task Load Index (NASA- TLX) scores, the one-way ANOVA revealed significant differences among the approaches ($F(2, 42) = 83.638$, $p < .001$). Post hoc tests indicated that Sphere Selection approach had significantly lower NASA- TLX scores compared to Container Creation approach (mean difference = -22.000 , $SE = 1.994$, $p < .001$) and the combination of the two approaches (mean difference = -19.733 , $SE = 1.879$, $p < .001$). Container Creation approach had significantly higher NASA- TLX scores compared to Sphere Selection approach (mean difference = 22.000 , $SE = 1.994$, $p < .001$) (Figure 8).

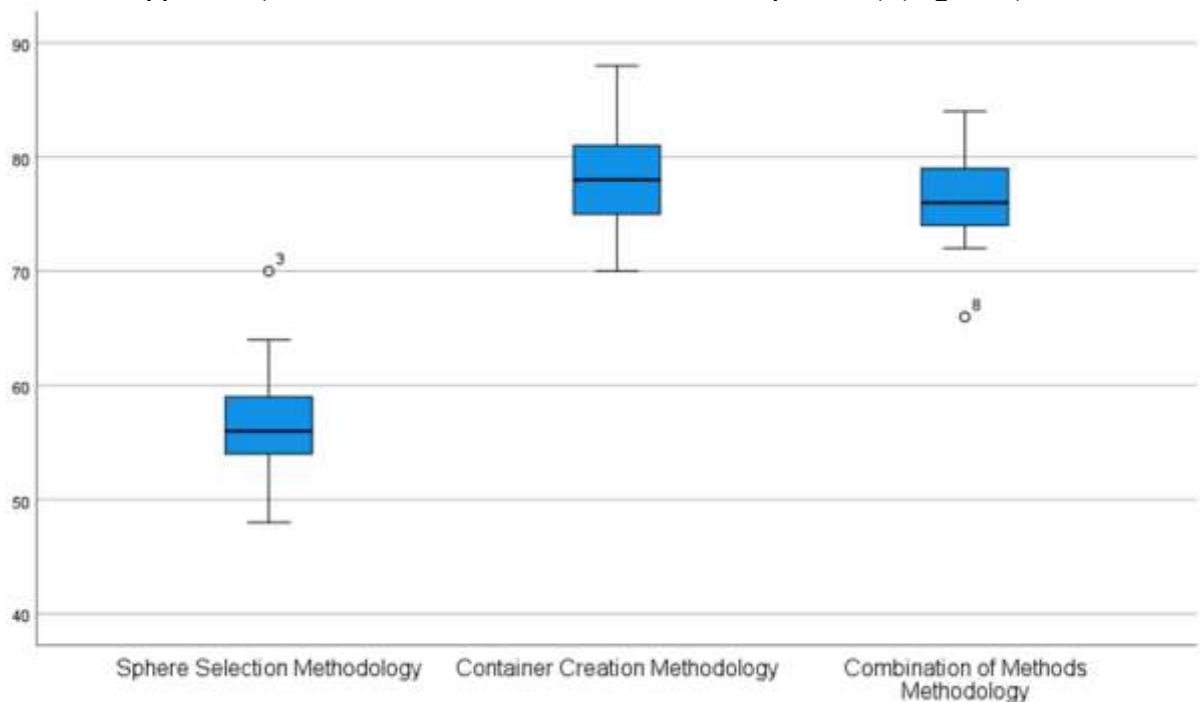


Figure 8: NASA- TLX scores for each approach.

4.5 Time

For the time taken to complete the task, the one-way ANOVA showed significant differences among the approaches ($F(2, 42) = 5.362, p = .008$). Post hoc tests revealed that Combination of the two approaches resulted in significantly lower completion time compared to Sphere Selection approach (mean difference = 4.467, $SE = 1.160, p = .002$). There was no significant difference in completion times between the Container Creation approach and combination of the two approaches (Figure 9).

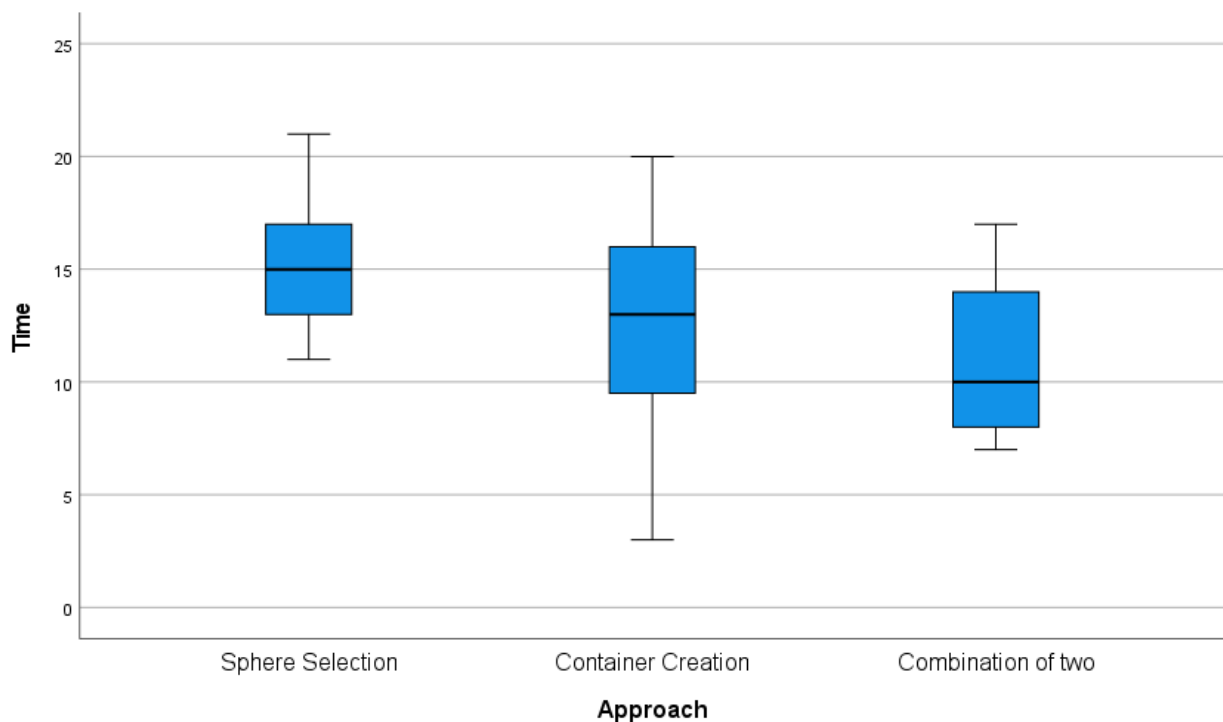


Figure 9: Time scores for each approach.

4.6 Accuracy

The one-way ANOVA results for accuracy indicated differences between the approaches. The between-groups analysis of variance reveals a significant difference ($p = 0.013$) in accuracy scores among the groups. The F-statistic is 4.854, indicating a noticeable difference in accuracy scores, as well. This suggests that the different approaches employed have an impact on the accuracy of the participants' tasks. Based on the post hoc tests, Sphere Selection approach had a significantly higher mean difference compared to Container Creation approach, with a mean difference of 0.933 ($p < 0.05$). Additionally, Sphere Selection approach had no significantly higher mean difference compared to the combination of the two approaches as the mean difference of 0.467 ($p > 0.05$) suggested. The Container Creation approach did not show a significant difference compared to the combination of the two approaches with a mean difference of 0.467 ($p > 0.05$) (Figure 10).

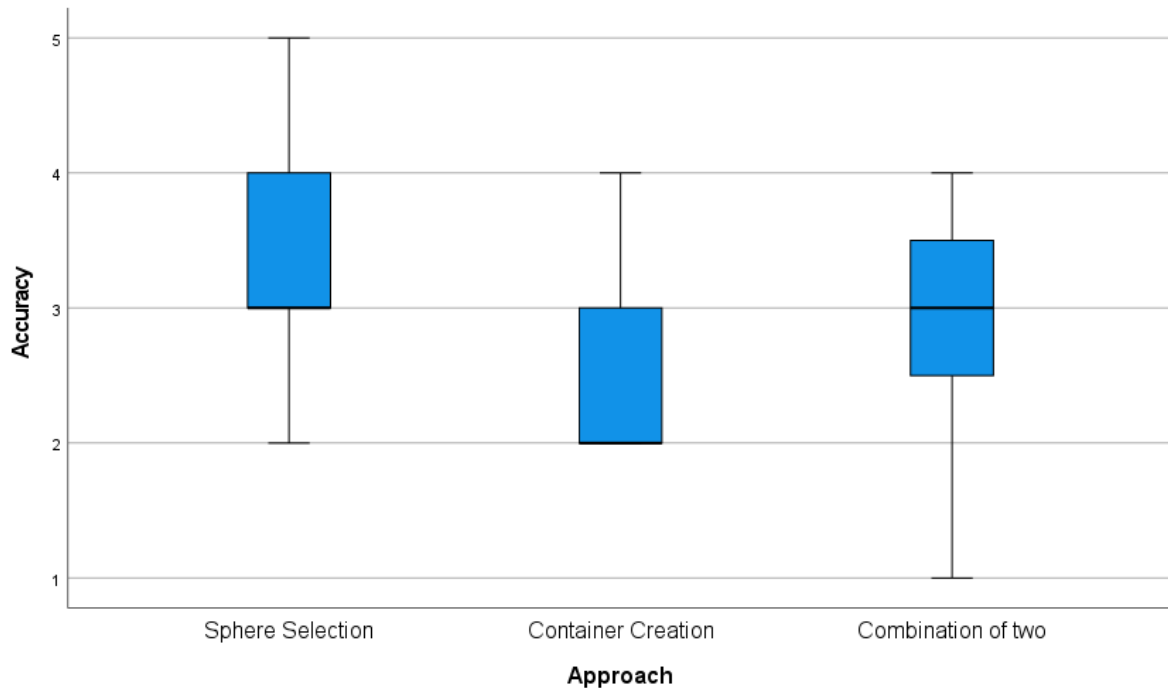


Figure 10: Accuracy scores for each approach.

4.7 Effect of Approach and Gender on the SUS scores

Further into the statistical analysis, a two-way ANOVA analysis was performed, to incorporate and correlate the gender factor with the previous examined variables. By using a two-way ANOVA, the effects of the Approach and Gender on the SUS variable were examined first. The results showed a significant main effect for Approach ($F = 13,112$, $p = 0.001$), indicating that different methodological approaches have a substantial influence on SUS scores. However, there was no significant main effect of Gender ($F = 0.605$, $p = 0.441$), indicating that gender does not influence SUS scores significantly. The interaction between the Approach and Gender was also not statistically significant ($F = 0.060$, $p = 0.942$), showing that the combined effect of Approach and Gender on SUS scores is not statistically significant.

4.8 Effect of Approach and the Gender on the TLX scores

Furthermore, two-way ANOVA was employing for examining the effects of the Approach and the Gender on the TLX score variable. The Approach factor shows a significant effect on TLX scores ($p < 0.001$). This suggests that the different approaches significantly influence participants' TLX scores. Additionally, The Gender factor does not show a significant effect on TLX scores ($p = 0.920$). This means that there is no significant difference in TLX scores between males and females. The interaction between Approach and Gender is statistically significant ($p = 0.018$). This indicates that the effect of Approach on TLX scores differs depending on the gender of the participants. In other words, the relationship between Approach and TLX scores varies based on whether the participant is male or female.

4.9 Effect of Approach and the Gender on the Time variable

The two-way ANOVA was conducted to examine the effects of Approach and Gender on the Time variable. The results indicate significant main effects for both approaches ($F = 7.447$, $p = 0.002$) and Gender ($F = 17.590$, $p < 0.0001$). However, the interaction between Approach and

Gender was not statistically significant ($F = 0.940$, $p = 0.399$). The Approach factor shows that different annotation approaches significantly affect the Time variable. Additionally, the Gender factor demonstrates that there are significant differences in Time between genders. More specifically participants who identified as male presented a mean time of 13:53 (minutes:seconds), with a standard deviation of 2:18, while female participants 17:40 with a standard deviation of 2:48 while annotating with the Sphere Selection approach. For the Container Creation, male participants presented a mean time of 10:20 with a standard deviation of 3:36, while female participants presented a mean time of 16:10 with a standard deviation of 4:10. Finally for the combination of two, male participants presented a mean time of 9:53 with a standard deviation of 3:03 and female participants presented a mean time of 12:30 with a standard deviation of 3:08. However, the interaction effect suggests that the combined influence of Approach and Gender does not significantly impact Time. Overall, these findings highlight the importance of considering both the Approach and Gender factors influential to the Time variable.

4.10 Effect of Approach and the Gender on the Accuracy scores

Additionally, the two-way ANOVA test was conducted to examine the effects of distinct factors on the accuracy scores. The overall model was not statistically significant ($F = 1.875$, $p = 0.121$), indicating that the combination of factors did not have a significant effect on accuracy. Nevertheless, the overall mean accuracy (intersect term) was highly significant ($F = 517.369$, $p < 0.001$), indicating that there were significant differences in accuracy scores overall. The main effect of Approach was statistically significant ($F = 4.469$, $p = 0.018$), indicating that the different employed approaches had a significant impact on accuracy scores. The main effect of Gender was not statistically significant ($F = 0.005$, $p = 0.943$), suggesting that gender did not have a significant effect on accuracy scores. The interaction effect between Approach and Gender was not statistically significant ($F = 0.144$, $p = 0.866$), indicating that the combined effects of Approach and Gender did not significantly impact accuracy scores.

4.11 User's feedback and comments

The final questionnaire included open questions about each of the following approaches, where the users had to indicate the positive and negative aspects, and their comments and recommendations. These results were used to answer RQ2.3.

4.9.1 Sphere Selection Approach feedback

Participants found the Sphere Selection approach to be user-friendly and simple to comprehend, with quick and straightforward application. The labelling was easy to be detailed and manageable, because of the adjustable shape of the sphere, which was considered an important feature by many participants. The delete feature was considered convenient, and overall, the approach was enjoyable. It was characterized by quick mobility, easy rotation, and uncomplicated annotation, providing benefits that were appreciated by the participants.

Furthermore, users pointed that Sphere Selection approach found some limitations that affected their experience. Among them, lacking accuracy and precision - especially in navigating and comprehending sphere depth - was commonly reported. The navigational feature of continuous motion was necessary, according to some. They also expressed the extra effort required in manually drawing every single point as a significant disadvantage. Areas for improvement were identified for Sphere Selection approach based on feedback. While it was deemed easy to use and intuitive, insufficient precision and navigation frequently mentioned.

4.9.2 Container Creation Approach feedback

Compared to Sphere Selection approach, the participants found Container Creation approach to be a challenging and difficult to use. A plethora of users perceived the Container Creation approach to be not as user friendly and the previous approach. Due to the need for extensive navigation, some individuals found the process to be both tiring and time-consuming. Inaccurate and unfinished annotations were common among participants because of the unclear positioning of points. Adding to the challenge, the navigation controlling the deletion feature caused added irritation (as the last feature is assigned the controller related to navigation).

The container creation method was less precise, detailed, and user-friendly than expected, though some did mention the potential for quickly annotating massive quantities of points. However, the majority mentioned that it was difficult to understand and get familiar with. In comparison to Sphere Selection approach, participants found Container Creation approach to be more challenging, less intuitive, and less accurate. Most of the participants preferred Sphere Selection approach due to its ease of use, and better results.

4.9.3 Combination of two approaches feedback

Finally, based on the feedback from the combination of the two approaches, participants had mixed opinions about this approach. Some participants appreciated the combination as it allowed them to use containers for linear parts and the sphere method for leaves. Others felt that the combination was unnecessary and preferred using either one of the methods. Additionally, a considerable number of participants expressed a strong preference for using only the Sphere selection method, since they found it easier to use. Additional comments supported that combination approach and switching between methods to be complex and confusing, as they felt that it required more practice and more training with the individual methods first.

Some additional important notes are that participants noted that the delete feature was assigned to a different controller for each approach in the combination approach, leading to confusion and difficulty in using two methods simultaneously. Other participants on the other hand found Combination of the two approaches to be the best approach, especially when considering the combination of both methods for various parts of the tree. Finally, others felt that the combination was unnecessary, and they preferred using one method consistently.

5. Discussion

The goal of thesis was to explore the potential of point cloud annotation in immersive virtual reality. To achieve this, two research questions (with two sub questions each) were created. The first one was related to which annotating technique is considered the most optimal, in terms of performance (time and accuracy) for annotating different components of the individual tree's point clouds in VR (first sub-question) and the potential influence of gender (second sub-question) the users' experience (including usability, preferences, intentions for future use, and mental workload) with the different annotating approaches (first sub-question) and identically to the first research question, the potential influence of gender (second sub-question).

5.1 Time and Accuracy

To answer the first research question (RQ1.1) "Which annotating technique is considered the most optimal, in terms of the time and accuracy for annotating different components of the individual tree's point clouds in VR?", both the results of the time and accuracy had to be discussed and combined. Hypothetically, the optimal annotation approach would be the one presenting a shorter completion time together with higher accuracy.

By solely considering both accuracy and time, the combination of the two approaches seems to offer the optimal annotation results, as it presented a balance between accuracy and reduced completion time. On the other hand, there are a plethora of factors that could influence these results. For this reason, attempting to answer RQ1.1, only based on the original hypothesis without examining the effects is superficial.

The lower time measurements of the two approaches compared to the Sphere Selection approach may be related to the design of the annotating experiment. As mentioned in the Experiment's structure subsection, the combination of the two approaches was tested by every user during the final scene. As Wirth et al. (2019) suggested, data annotation is a challenging process, and it was reported that non-experienced users (students) were showing low willingness to perform annotation tasks. Although a game-like concept was designed to solve this problem - and for that reason was adapted by the approach of this thesis - many participants of this research expressed lack of motivation during the experiment's last stages. Additionally, participants described the combination of the two approaches as complicated and unnecessary, through open questions. Furthermore, the accuracy scores of the Combination of two approaches did not present a statistically significant difference compared to the Sphere Selection approach. An explanation for this result can be found on participant's feedback (open questions). A considerable number of annotators (8 participants out of 15) declared that the Sphere Selection was their approach of choice, even though they were able to use both approaches, while annotating with the Combination of two approaches.

Moreover, many participants reported moderate simulation sickness and general discomfort (mostly ranging from mild to moderate). Similar complaints were reported by student annotators in the experiment of Wirth et al. (2019). Since the participants had already spent a significant amount of time on that stage annotating in a VR environment, these factors may have influenced the time that the users were willing to spend for the final tasks. It is unknown how much time spent in VR led to those unpleasant effects, as examining this factor was not included in the scope of this research. As Wirth et al. (2019) suggested it is important to examine this factor before hiring someone to annotate in VR for a long time. For this reason, future work could be directed on this issue.

Additionally, since the participants did not experience the individual approaches repeatedly and for an extended period, the combination could affect the willingness of participants to invest as much time as they did in the previous stages of the experiment. As a result, this outcome may not only affect the factor of time, but also the accuracy factor, as the reported complexity of the combination of the two approaches, the lack of motivation during the last stages of the experiment and the reported VR discomfort, could lead to annotations of lower quality. These effects can be described by fatigue caused by the Carry-Over Effect in which the participant becomes less successful at a specific task due to exhaustion from completing earlier experimental procedures (Price et al., 2015; Fan & Smith, 2017). For this reason, future work can be directed on examining different annotation approaches by employing a different experiment design to overcome these limitations.

Another important reason for employing a different experiment design is the sequence of the combination of two approaches in the final scene for each participant and the use of trees with higher complexity. Since participants had to experience first both Sphere Selection and Container Creation approaches separately, it was believed some may have developed a preference over a specific approach. Moreover, it was expected that by the time they had to annotate the final scene, some participants would only utilize their approach of choice. For this reason, annotating a tree of similar complexity would not ensure comparability on performance, as the users would be already more experienced compared to the scenes prior. Although, the incorporation of more complex point clouds was thought to overcome this obstacle, ultimately, this design choice was not able to ensure the comparability on performance of participants among the three conditions.

Based on the above, the combination of the two approaches cannot be safely considered the optimal approach. On the other hand, the answer to RQ1.1 is that the Sphere Selection approach is more viable compared to the other two methods. Still, further investigation and experiments are considered necessary.

Based on the results of this sample of participants, the answer to RQ1.2 is that there was no relationship between the performance (time and accuracy) of different annotating techniques in VR and participants' gender. It is important to mention that although there was no statistical significance between the gender of participants and the effects of discomfort and VR sickness, their relations to different approaches and discomfort and VR sickness were not examined as it was not included in the scope of this project. As VR sickness and discomfort is a known VR related limitation and relations with gender are reported (Ihmedy-Steinke et al., 2017; Wallgrun et al., 2019; MacArthur et al., 2021) further investigation into their relation to performance is necessary.

5.2 Usability, workload, and overall satisfaction assessment

To measure the usability, workload, and overall satisfaction of each approach the System Usability Scale (SUS; Brooke, 1995) and Nasa Task Load Index (NASA- TLX; Hart, 2006) were successfully utilized. The average SUS scores indicated that the Sphere Selection Approach was perceived as more effective and enjoyable than the Container Creation approach and the combination of the two approaches. Furthermore, the average NASA- TLX showed that the Selection Approach was the least physically and mentally demanding annotation technique, compared to the other two methods. The feedback through open questions indicated that the Sphere Selection approach was easy to understand and user-friendly, but some participants expressed concerns about its precision and accuracy. On the other hand, the

Container Creation approach had advantages in terms of speed and innovation, but participants found it challenging to use, hard to understand, and struggled with the node's placement. Combination of the two approaches, received mixed feedback. Some participants appreciated the combination for differentiating between tree parts and allowing faster annotation in specific areas. However, others found it unnecessary, confusing, or struggled with switching between the methods.

Identically to the previous subsections, the same limitations may apply. The experiments design, the possible fatigue caused by the Carry-Over Effect (Price et al., 2015; Fan & Smith, 2017) and VR discomfort and sickness may have influenced the perception of users on usability, workload, and overall satisfaction. For this reason, investigating further users' experience for each methodology individually is essential in future research.

For answering the RQ2.1, it can be concluded that based on the System Usability Scale (SUS) scores and feedback of the 15 participants of this research, the Sphere Selection Approach is considered to a more effective, efficient, and satisfactory annotation experience, compared to the Container Creation approach and the combination of the two approaches (Brooke, 1995). Based on the Nasa Task Load index (NASA- TLX) user's feedback, the Sphere Selection Approach is less physically and mentally demanding annotation technique, compared to the Container Creation approach and the combination of the two approaches, with both being demanding on a similar level (Hart, 2006).

For answering RQ2.2, it can be concluded that usability and workload perveance of different annotating approaches in VR is not related to gender that participants were identified to. Although the importance of examining the relation of gender and usability is highlighted (MacArthur et al., 2021), thoroughly research is necessary to effectively contribute into these literature gaps.

Although the overall design on this experiment had similarities with the related work of Wirth et al. (2019), the usability, workload, and overall satisfaction were not evaluated. Additionally, there are no publications relating 3D data annotation techniques in VR with the examined factors to perform a comparison. This highlights the importance of developing a standardized evaluation processes for 3D for users' evaluation for 3D point cloud annotation techniques in environmental applications and can be described as a baseline.

User feedback made clear which parts of annotation methods are known to help improve usability, especially in terms of efficiency, to answer RQ2.3. The Sphere Selection method has advantages such as being easy to use and understand, and the shape of the sphere can be changed, which many users found useful. The user experience is also made better by the fact that it is easy to move around, rotate, and remove labels. In the same way, the Container Creation method shows how it could be used to efficiently annotate a large number of tree point cloud points. The Combination of Two Approaches is also known for being flexible when it comes to dealing with various parts of annotation tasks.

A significant limitation of this study is that the differences between respondents were mostly examined in terms of gender. This is because the responses were not fairly distributed across age groups, educational levels, or experience with VR. Because there were not many different answers to these questions, it was decided to focus on gender. Future studies with a larger and more diverse group of participants may be able to investigate a wider range of factors that make people different. These could be elements like age, level of education, familiarity with VR, and more. This could provide a greater understanding of the factors that affect how users experience

and performance.

5.3 Environment and Data discussion

It is important to mention that, although point cloud format is considered a prominent type of data for environmental applications, as already discussed in the literature review, it could be argued that other types of data could be also employed for the purpose of this thesis. An example of this data would be meshes, or even point cloud files that come with RGB color information to each point. Although color information could be a useful indication of the different parts of the trees and assist users to differentiate the canopy from the branches, it is considered a compromising factor for this annotating experiment. More specifically it is believed that color information could act as a distracting factor, as the users are expected to simply annotate according to the color, without investigating and understanding the form of the given point clouds. The goal of this research is to enable the critical thinking of participants, to give them the opportunity to fully explore all the offered annotation and navigation features and develop their personal strategy to gain full insight and perform a meaningful evaluation.

Furthermore, the RGB color information is not always present as it occurs from capturing the physical objects by cameras with RGB or infrared sensors, or by post-processing algorithms that analyze the captured data (Mokatren et al, 2022). As it is expected this kind of information is not present in every dataset. As a result, choosing a dataset with color information would automatically exclude a high number of existing and future datasets and make this study case orientated and irrelevant to many cases. As mentioned in the literature review, since there is a lack of research on the topic of point cloud annotation and evaluation, a non-specific case study research is considered essential. Additionally, this study could set a baseline for future research in other dataset types.

Another argument for using non-colored point cloud data is that, for the Unity platform to represent and load complexed colored data files a high computational power is required. This needed power is expected to be even greater, especially while the participants are navigating around the scene. As viewing angle and position changes, the point cloud representation is constantly adapting, with the points resizing to bigger when the user is moving far away, and smaller when they move towards the point cloud. These constant visual adaptations needed from Unity, in the case of color point cloud, the system requirements needed for the smooth operation of this experiment exceed the present equipment. As a result, lower frame rates and possible insufficient loading time are expected which could be another factor to negatively affect the user's experience.

5.4 Navigational features discussion

Furthermore, a crucial factor that was considered an inextricable aspect for the annotation process but was not examined is the different navigation features offered to the users. As mentioned in the methodology, the users could freely choose between the Flying feature, which allowed them to navigate freely and the Teleport feature, accompanied with rotation capabilities. In this research, the effect of the different navigation features on the result of the different annotation approaches is unknown. As navigation was a crucial factor for exploring the tree point cloud datasets in VR, which is the key and the basic concept of this research, there is still an unexplored aspect. Additionally, personal preferences and annotation strategies and combinations between annotation method and navigation method were not examined nor documented. Furthermore, it is unknown if the navigation choice had impact on the accuracy, time, the overall perceiving of the efficiency, user's satisfaction, and workload for each

approach. A potential further step for this research, would be an investigation of these factors through a literature review and ideally the incorporation, examination and correlation with the different annotating approaches accompanied with a more in-depth statistical analysis.

5.4 Limitations

This subsection provides information on other limitations and challenges faced during the conduction of this research.

5.4.1 Participants and experiment setup

It is important to discuss one of the obvious limitations, which is the number of participants in this research. This statistical sample may be considered low, but a substantial challenge of this thesis was the recruitment process of participants. For this kind of experiment, it was important to set up an environment in which participants would be able to use the Oculus equipment safely and without any external factor that could have altered their experience. This is one of the reasons the experiment setup had to be in a specific place, and this made accessibility hard for some potential participants. Additionally, the overall time needed to complete all the distinct phases of the experiment, meaning the introduction presentation, the testing annotation phase, the annotation phase with each approach respectively and the completion of all the questionnaires, required more than one hour per participants. These requirements made the recruitment process challenging, as the overall experiment was considered overall demanding. The same experiment could be performed in the future with a higher number of participants.

5.4.2 Accuracy evaluation

A significant limitation of this evaluation of the accuracy factor is present in this research's approach. Firstly, it is necessary to report that the annotated tree point clouds, made by the participants of this research, and the ground truth files on which the comparison was based on, were produced by different software and different environments. In the case of the former, the annotation environment, was developed in the game engine of Unity and offered three dimensional features and capabilities to the participants, as well as navigation systems which offered two different interaction approaches. The first navigation option was related to free navigation around the annotation scene (Fly feature assigned in the right controller of the Oculus set), while the second to teleport and rotation features (Teleport and rotation features). Both styles offered unique ways of methodically exploring and interacting with the datasets. Also, the annotated trees were produced by the two examined approaches and their combination. As a result, the overall creation of the annotated tree point clouds was based on combination of different navigation and annotation styles, with a given high level of freedom to the users.

On the other hand, the annotated files considered as ground truth for this research were produced in Cloud Compare, which is a software that although it produces 3D visualization of point clouds, the only offered labeling feature is on 2D. To be more specific, the annotation process demanded a constant rotation of the point cloud around the three axes X, Y, Z and clip features, based on drawing 2D polygons and assigning labels on different parts of the trees. As a result, this approach may be lacking when it comes to greater details. Nevertheless, before the comparison of the participant's results, the ground truth point clouds were inspected by a researcher of the Wageningen University, and after corrections, were approved for the comparison.

When it comes to the actual comparison process of the participants products and the ground truth, for the purpose of calculating the accuracy factor, initially a python script was developed. This approach calculated the percentage of the precision, between the given class of each point of the point cloud and the points of the ground truth point cloud. Although this approach was prominent during the testing phase of comparing the classes of point cloud files, it was not implemented for the interpretation of the actual results of this research. This decision was made, since unexpectedly the results of the annotation process were not saved in a usable format. To be more precise, once the annotation of the tree point clouds in the 3D Unity environment was completed, the final product was not an updated version of the existing .las point cloud which included the annotation as an additional field (for example as a scalar field incorporated into the original point cloud), but the annotation was exported as a separate new folder including a collection of .voxel files. The structure and of those files were originally compatible with the existing packages and scripts developed appropriately for Unity. An extensive attempt was made to find an approach to exploit this structure and format and export the labels into scalar fields for each tree point cloud. Nevertheless, this attempt was not proven fruitful, as the lack of information about this topic, the high level of programming knowledge needed were and the time limitation were an obstacle.

An alternative approach which was considered was the comparison of the user annotated point clouds and the ground truth ones, by employing software and approaches based on image analysis. The initial idea was to compare the RGB values of pictures, depicting the annotated trees from different perspectives and angles. This approach was not further explored, neither incorporate is the methodology of this research, since it was considered to compromise the scope of this project from the perspective of time. Additionally, there was uncertainty related to the comparison of visualizations results from different software, approaches, and approaches, making the effectiveness of this approach questionable for this research.

Finally, the employed methodology for the assessment of the accuracy of this research was considered the optimal solution, but as expected there are related limitations. When it comes to optical assessments of multiple images the interpretation of the result and the grading process is solely based on the perspective of an individual. Additionally, since there were no publications found on optical comparisons of labelled point clouds or related concepts which could be useful for this research, the evaluation process was performed empirically, with the assistance of a fellow MSc student with a strong scientific background on Forestry. The evaluation of every annotated tree point cloud was completed after close side-by-side inspection and comparison of each annotated tree with the ground truth file. This approach may come with errors which at this stage of the research cannot be quantified and further explored.

5.4.3 Other challenges

Due to limited knowledge of statistics and statistical analysis, two-way Anova was initially thought to be the ideal test to perform multiple comparisons. After the completion of the result analysis, the supervisor of this thesis pointed that mixed ANOVA test would be more suitable to this research as gender is a between-subject variable, while the annotation approach is a within-subject variable as each participant experienced all three annotation approaches. Unfortunately, due to data loss and hardware issues during the final stages of this thesis, performing another statistical analysis was not possible.

6. Conclusion

In this thesis, the potential of point cloud annotation in immersive virtual reality was successfully investigated. The concept of embodied digital twins, Unity software, WANDER Lab's tools, and environmental datasets of trees were utilized to generate this virtual experience for student annotators. Three annotating approaches were compared in terms of performance, workload, usability, and overall satisfaction, and their relation to participants' individual characteristics was examined. Users' experience evaluation showed that Sphere Selection is considered the optimal approach and has proven to be successful for segmenting and labeling different tree point cloud components. Results did not prove any relation between gender, performance, and users' experiences.

While this approach has proven effective for this thesis, more research should be conducted to fully investigate each annotating approach's potential. To conclude, future work should incorporate the investigation of virtual reality sickness and discomfort's impact on the examination of each annotating approach. Additionally, the effect of users' individual differences on the performance and usability of each approach should be further investigated.

Overall, this research contributes to field data annotation and highlights the potential for immersive technologies in environmental studies. Further advancements and research on designing alternative approaches for generating efficient training datasets can use the findings and concepts of this research as a starting point. As these concepts are not only relevant to environmental applications, other fields could also benefit from the insights of this thesis. Finally, the findings of this thesis could be a starting point for the design criteria for 3D VR tree annotation tools and applications and the participant application usability study.

7. References

- Akay, A. E., Oğuz, H., Karas, I. R., & Aruga, K. (2009). Using LiDAR technology in forestry activities. *Environmental monitoring and assessment*, 151(1), 117-125.
- Bacim, F., Nabiyouni, M., & Bowman, D. A. (2014, March). Slice-n-Swipe: A free- hand gesture user interface for 3D point cloud annotation. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 185-186). IEEE.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Coffey, D., Malbraaten, N., Le, T. B., Borazjani, I., Sotiropoulos, F., Erdman, A. G., & Keefe, D. F. (2011). Interactive slice WIM: Navigating and interrogating volume data sets using a Gonzalezmultisurface, multitouch VR interface. *IEEE Transactions on Visualization and Computer Graphics*, 18(10), 1614-1626.
- Corain, L., & Salmaso, L. (2007). A critical review and a comparative study on conditional permutation tests for two-way ANOVA. *Communications in Statistics—Simulation and Computation*, 36(4), 791-805.
- Doula, A., Güdelhöfer, T., Matviienko, A., Mühlhäuser, M., & Sanchez Guinea, A. (2022). Immersive-Labeler: Immersive Annotation of Large-Scale 3D Point Clouds in Virtual Reality. In *ACM SIGGRAPH 2022 Posters* (pp. 1-2).
- Drew, M. R., Falcone, B., & Baccus, W. L. (2018, July). What does the system usability scale (SUS) measure?. In *International Conference of Design, User Experience, and Usability* (pp. 356-366). Springer, Cham.
- Duarte, L., Teodoro, A. C., Moutinho, O., & Gonçalves, J. A. (2017). Open-source GIS application for UAV photogrammetry based on MicMac. *International Journal of Remote Sensing*, 38(8-10), 3181-3202.
- Fan, J., & Smith, A. P. (2017). The impact of workload and fatigue on performance. In *Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland, June 28-30, 2017, Revised Selected Papers 1* (pp. 90-105). Springer International Publishing.
- Forsyth, D. A., & Ponce, J. (2002). *Computer vision: a modern approach*. prentice hall professional technical reference.
- Goel, A., Tung, C., Lu, Y. H., & Thiruvathukal, G. K. (2020, June). A survey of methods for low-power deep learning and computer vision. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)* (pp. 1-6). IEEE.
- Gonzalez de Tanago, J., Lau, A., Bartholomeus, H., Herold, M., Avitabile, V., Raumonon, P. Martius, C., Goodman, R. C., Disney, M., Manuri, S., Burt, A., Calders, K. (2017). Estimation of above-ground biomass of large tropical trees with Terrestrial LiDAR. *Methods in Ecology and Evolution*. DOI:10.1111/2041-210X.12904
- Graessler, I., & Pöhler, A. (2017, December). Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system. In *2017 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 289-293). IEEE.
- Gregory, J. (2018). *Game engine architecture*. crc Press.
- Grieves, M. (2014). Digital twin: manufacturing excellence through virtual factory replication. *White paper*, 1(2014), 1-7.
- Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. *Transdisciplinary perspectives on complex systems: New findings and approaches*, 85-113
- Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In

Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Henry, M., Réjou-Méchain, M., Jara, M. C., Wayson, C., Piotto, D., Westfall, J., ... & Saint-André, L. (2015). An overview of existing and promising technologies for national forest monitoring. *Annals of Forest Science*, 72(6), 779-788.

Hmida, H. B., Cruz, C., Boochs, F., & Nicolle, C. (2012). From Unstructured 3D Point Clouds to Structured Knowledge-A Semantics Approach. *Semantics-advances in theories and mathematical models*, 213.

Huang, J., Lucash, M. S., Scheller, R. M., Klippel, A., 2020. Walking through the forests of the future: using data-driven virtual reality to visualize forests under climate change. *International Journal of Geographical Information Science*, 1–24.

Ihemedu-Steinke, Q. C., Rangelova, S., Weber, M., Erbach, R., Meixner, G., & Marsden, N. (2017). Simulation sickness related to virtual reality driving simulation. In *Virtual, Augmented and Mixed Reality: 9th International Conference, VAMR 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings 9* (pp. 521-532). Springer International Publishing.

Jerald J., Giokaris P., Woodall D., Hartbolt A., Chandak A., Kuntz S. (2014). Developing virtual reality applications with Unity. *IEEE Virtual Reality (VR)*, 2014, pp. 1-3, doi: 10.1109/VR.2014.6802117.

Kamińska, D., Zwoliński, G., & Laska-Leśniewicz, A. (2022). Usability Testing of Virtual Reality Applications—The Pilot Study. *Sensors*, 22(4), 1342.

Kaur, J., & Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, 81(27), 38297-38351.

Khanal, S., Kc, K., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, 12(22)

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, 69, 66-73.

Klippel, A., Sajjadi, P., Zhao, J., Wallgrün, J. O., Huang, J., & Bagher, M. M. (2021). Embodied digital twins for environmental applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 193-200.

Klippel, A., Zhao, J., Jackson, K. L., La Femina, P., Stubbs, C., Wetzel, R., ... & Oprean, D. (2019). Transforming earth science education through immersive experiences: Delivering on a long held promise. *Journal of Educational Computing Research*, 57(7), 1745- 1771.

Knuiman, B. (2022). Wanderlib. <https://git.wur.nl/wander/wanderlib>

Kumar, S., Arivazhagan, S., & Rangarajan, N. (2013). Remote sensing and GIS applications in Environmental Sciences—A review. *J. Environ. Nanotechnol*, 2(2), 92-101.

Lechner, A. M., Foody, G. M., & Boyd, D. S. (2020). Applications in remote sensing to forest ecology and management. *One Earth*, 2(5), 405-412.

Lubos, P., Beimler, R., Lammers, M., & Steinicke, F. (2014, March). Touching the Cloud: Bimanual annotation of immersive point clouds. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 191-192). IEEE.

MacArthur, C., Grinberg, A., Harley, D., & Hancock, M. (2021, May). You're making me sick: A systematic review of how virtual reality research considers gender & cybersickness. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-15). sense of presence, and gender. *Media Psychology Review*, 11(2), 1.

Mokatren, M., Kuflik, T., & Shimshoni, I. (2022). 3D Gaze Estimation Using RGB-IR Cameras. *Sensors*, 23(1), 381.

Monica, R., Aleotti, J., Zillich, M., & Vincze, M. (2017, October). Multi-label point cloud annotation by selection of sparse control points. In *2017 International Conference on 3D Vision (3DV)* (pp. 301-308). IEEE.

O'Mahony, N., Campbell, S., Krpalkova, L., Carvalho, A., Velasco-Hernández, G. A., Riordan, D., & Walsh, J. (2018, December). Convolutional Neural Networks for 3D Vision System Data: A review. In 2018 12th International Conference on Sensing Technology (ICST) (pp. 160-165). IEEE.

O'Mahony, N., Campbell, S., Carvalho, A., Krpalkova, L., Riordan, D., & Walsh, J. (2019, December). Point cloud annotation methods for 3D deep learning. In 2019 13th International Conference on Sensing Technology (ICST) (pp. 1-6). IEEE.

Panetto, H., Iung, B., Ivanov, D., Weichhart, G., & Wang, X. (2019). Challenges for the cyber-physical manufacturing enterprises of the future. *Annual Reviews in Control*, 47, 200-213.

Price, P. C., Jhangiani, R. S., & Chiang, I. C. A. (2015). Experimental design. *Research Methods in Psychology*.

Ross, A., & Willson, V. L. (2017). One-way anova. In *Basic and advanced statistical tests* (pp. 21-24). Brill.

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *International journal of computer vision*, 77(1), 157-173.

Sajjadi, P., Van Broeckhoven, F., De Troyer, O., 2014. Dynam-ically adaptive educational games: A new perspective. *International Conference on Serious Games*, Springer, 71-76.

Sheridan, R. D., Popescu, S. C., Gatziolis, D., Morgan, C. L., & Ku, N. W. (2014). Modeling forest aboveground biomass and volume using airborne LiDAR metrics and forest inventory and analysis data in the Pacific Northwest. *Remote Sensing*, 7(1), 229-255.

Simpson, M. (2020). Scale and space: Representations in immersive virtual reality (Doctoral dissertation, The Pennsylvania State University).

Solomon, F. (2018). EDUCATIONAL GAME FOR FORESTRY: Creating Unity 3D Game.

Stets, J. D., Sun, Y., Corning, W., & Greenwald, S. W. (2017). Visualization and labeling of point clouds in virtual reality. In *SIGGRAPH Asia 2017 Posters* (pp. 1-2).

Toma, M. I., Gîrbacia, F., & Antonya, C. (2012). A comparative evaluation of human interaction for design and assembly of 3D CAD models in desktop and immersive environments. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 6(3), 179-193.

Trochta, J., Krůček, M., Vrška, T., & Král, K. (2017). 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. *PloS one*, 12(5), e0176871.

Veit, M., & Capobianco, A. (2014, March). Go'then'tag: A 3-d point cloud annotation technique. In 2014 IEEE Symposium on 3D User Interfaces (3DUI) (pp. 193-194). IEEE.

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E. M., Kyrpides, N., ... & Meyer, F. (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC bioinformatics*, 13, 1-5.

Wirth, F., Quehl, J., Ota, J., & Stiller, C. (2019, June). Pointatme: efficient 3d point cloud labeling in virtual reality. In 2019 IEEE Intelligent Vehicles Symposium (IV) (pp. 1693-1698). IEEE.

Zhang, M., You, H., Kadam, P., Liu, S., & Kuo, C. C. J. (2020). Pointhop: An explainable machine learning method for point cloud classification. *IEEE Transactions on Multimedia*, 22(7), 1744-1755.

Zhao, J., Klippel, A., 2019a. Scale - unexplored opportunities for immersive technologies in place-based learning. R. Teather, Y. Itoh, J. Gabbard (eds), *Proceedings of the 26th IEEE Conference on Virtual Reality and 3D User Interfaces*, IEEE, Piscataway, NJ, 155-162.

Zimmer, W., Rangesh, A., & Trivedi, M. (2019, June). 3d bat: A semi-automatic, web-based 3d annotation toolbox for full-surround, multi-modal data streams. In 2019 IEEE Intelligent Vehicles Symposium (IV) (pp. 1816-1821). IEEE.

Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). Remote sensing and image interpretation. John Wiley & Sons.

Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y., & Behnood, A. (2022). 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Advanced Engineering Informatics*, 51, 101501.

Persson, H. J., Olofsson, K., & Holmgren, J. (2022). Two-phase forest inventory using very-high-resolution laser scanning. *Remote Sensing of Environment*, 271, 112909.

Gupta, A., Byrne, J., Moloney, D., Watson, S., & Yin, H. (2018). Automatic Tree Annotation in LiDAR Data. In *GISTAM* (pp. 36-41).

Lorenz, M., Brade, J., Klimant, P., Heyde, C. E., & Hammer, N. (2023). Age and gender effects on presence, user experience and usability in virtual environments—first insights. *PloS one*, 18(3), e0283565.

Appendix 1:

A.1.1 Unity Project and Annotation Tool Setup

A rather challenging part of this thesis was the setup of the Unity Project, the import of the necessary packages, and the setup and connection of the project with the Oculus VR headset. In the following chapters, the different steps of those processes are described. Every step was important in producing the annotation experiment. Nevertheless, the scope of this project, as well as the Research Question are not directly related to unity development but rather focus on the evaluation of the user experience. Therefore, the following chapters incorporate an overview of the different steps needed for the creation of the experimental VR environment. If the reader wishes to gain a more in-depth overview of each step, there are available links for guidelines and useful links. The core of the following chapter was incorporated to this thesis from documentation created by the developer of the annotation tool (Knuiman, 2022) and it was enriched with further information which were considered necessary for the methodology of this research.

A.1.1.1 Download and Setup of UnityHub

UnityHub can be downloaded through the official website <https://unity.com/>. In the case that the user is not already registered, it is mandatory to sign up for a new account. By selecting a personal license, the user can keep using the software at no cost.

It is important and highlighted by the developer and the documentation of the annotation tool that the user must select Unity version 2021.3.11f1 LTS after installing UnityHub. The final digit (11f1) may have changed, as new versions are released frequently, although this is unlikely to be a major concern for the smooth operation of the project (Knuiman, 2022). After the installation is completed, a new Unity Project has to be made, specifically for 3D (UPR).

A.1.1.2 Set up Gitlab

For the purpose of the setup of this annotation experiment, a number of Unity packages have to be downloaded and set up under the same Unity project. These packages are available in the links under each package's description, respectively. Access to these packages can be gained through GitLab. It is necessary to create an account on <https://git.wur.nl/>. Direct access, without registration, is given to the students and the staffs member at WUR.

The next step of the methodology involves the creation of a secure shell (SSH) key and the addition of it to the GitLab profile. The provided guideline be downloaded through the given link for further instructions:

<https://docs.gitlab.com/ee/user/ssh.html#generate-an-ssh-key-pair>

A.1.1.3 Gaining access and installation the Unity Packages

As the author of the documentation suggests (Knuiman, 2022), the following steps should be completed in order to add necessary packages to the Unity Project:

1. Adding the wanderlib package from the URL git@git.wur.nl:wander/wanderlib.git. This can be done through the Package Manager of Unity, as well as for the next packages referred to in this chapter. Alternatively, the packages can also be downloaded by gaining access through the same url and importing the packages manually into the Unity project (Knuiman, 2022).
2. Adding the XR Interaction Toolkit (com.unity.xr.interaction.toolkit@2.0.3) in the project. Further information for this package can be found in the given URL.
3. Running the git bash in the Asset folder. It should be noted the working directory for git bash (cd) is located in the Asset folder of the project. Afterwards, the following should be input into git bash or another terminal by cloning the repository using Git located at git@git.wur.nl:wander/xrframework.git (Knuiman, 2022).
4. The final step is not mandatory, as it includes testing an Example scene provided by the developer of the tool, located in the xrframework folder. Then the demonstration of this example can be experienced by clicking the play button in Unity (Knuiman, 2022).

A.1.1.4 Oculus Setup

Once the headset and controllers are charged, the user presses the power button on the right side of the headset to turn it on. Afterward, by following the on-screen instructions, the user must set up their Oculus Quest, which includes connecting to Wi-Fi, creating an Oculus account with Meta, and pairing the touch controllers. Furthermore, it is important for the user to own a smartphone, or a tablet, or be able to install the Oculus app on their computer to gain further access to the headset's settings, as it is needed for then further steps.

The next step is related to the Oculus and Oculus App setup. It is highly recommended by the developers to keep the hardware and the application up to date. The Oculus application is an important topic for the methodology of this thesis since it enables the connection of Oculus with the computer, and consequently with the Unity Project, via either USB or AirLink. Oculus offers extensive guidelines for configuring the Quest. Further information and guidelines can be found on the official Oculus website.

A.1.1.5 Developer Mode Enabling

The next step of the methodology involves enabling the Developer mode of the Oculus Quest 2, as it is necessary to connect the headset to the Unity project. To enable developer mode, the user needs to activate the Developer mode option in their Oculus app on their smartphone, tablet, or computer. In the case that the user wants to use their smartphone or tablet, it is necessary to download the official Oculus app on their smartphone or tablet via the App Store.

Once the application is downloaded and linked through the user's Meta account, the Oculus Quest 2 must be connected to the app via Bluetooth, or with a USB cable. Once the headset is connected, the user can toggle on the Developer Mode option by navigating through the tabs Settings > Headset > More Settings. It is important to mention that in case the user does not own their personal headset, some features may potentially be restricted or predefined by the

organisation/ administration of the device.

As an example, in the case of this thesis, the device is owned by the WANDER Lab of Wageningen University, and the Developer Mode is already enabled if the user decides to use the headset with the predefined credentials. If the user wants to use their personal account on a device that they do not own, they still have to follow this procedure, even though the administrator's profile has the Developer mode on.

The same procedure can be followed by downloading the Oculus application on the computer. The installation of this software is recommended since the computer version of the application offers an extensive preview of the Oculus settings and can be helpful for the later connection with Unity.

A.1.1.6 Developer Organization Establishment

Furthermore, the next step of the methodology requires getting the Oculus Quest ready for development by establishing a developer organization. Another advantage of creating an organization is the access to additional Oculus development tools and the possibility of realising the developed applications in the Oculus Store. The procedure of forming a developer organization is a one-time activity and is only required once per account.

The developer organizations can be created by using the Oculus Developer Dashboard following the link:

<https://dashboard.oculus.com/>

To do so, it requires logging into an existing Meta Oculus account. Once the user is logged in, to start a new organization, the user must click on the "Create New Organization" link on the main menu. The link will redirect to a blank "Organization Name" section which must be filled up with the company's name. Additionally, the user can choose a "Category" that best describes your company from the list provided. The same applies to choosing an "Organisation Type" from the menu. An optional step is giving to the company some background by filling out the "Description" field. Finally, to start forming the new development group, the user must complete the creating procedure by selecting the "Create" tab. After the establishing of the developer organization, it is possible to distribute virtual reality software via the Oculus Store, have access to Oculus' software development kits (SDKs), APIs, and forums, and network with other developers.

An additional non-mandatory step can be followed for the publication of the new Unity application in the Oculus Developer Dashboard. This can be achieved by clicking on "Create New App". Afterward the user needs to choose a name for the application and select the platform as Oculus Quest. This procedure can be finalised by clicking "Create". Finally, it is important to configure the application for Unity. This can be done by clicking the "Edit App" button on the selected application through Oculus Developer Dashboard and choosing select Unity as the engine under "App Integration." After the completion of the previous steps, it is essential to download and import the Oculus Integration package for Unity in the desired project.

A.1.1.7 Connect Oculus to Unity

Usually, when the user connects the Oculus Quest 2 to Unity, it will automatically generate a new project and import the required assets. As it is mentioned in the respective chapter of the methodology, according to the developer of the annotation tool, making a 3D (UPR) project is necessary for the packages to function properly (Kniuman, 2022). It is recommended to create the Unity Project and the installation of the necessary packages before connecting the Oculus Quest 2 to Unity.

The Unity Project can be set up for connecting with the Oculus Quest 2, by entering the Player option in the Project Settings menu. It is important to make sure that the Virtual Reality Supported option is checked under XR Settings. The following step is the addition of Oculus to the list of compatible hardware by clicking the plus symbol. At this point, the user can link up the Oculus Quest 2 using Unity by connecting the equipment to the computer through USB-C or AirLink. The Oculus button in Unity may be used to initiate a connection. Once the Oculus Quest 2 device is recognized by Unity, the user can begin the development of the Unity Project.

More extensively, after connecting the Oculus Quest to the computer using a USB cable, the user must go to Edit > Preferences > External Tools. In this tab, the Android section should be visible. In this section, by clicking on the browse button next to the SDK field, it is important to select the Android SDK folder on the computer. Once the selection is made, in the unity project, the user must navigate to File > Build Settings and then click on Add Open Scenes button to add their scene to the build settings.

Afterward, by clicking the Player Settings button, the Player Settings window should open. In there, Under XR Settings, select Oculus as the Virtual Reality SDK. Additionally, Under Other Settings, select the Minimum API Level as Android 7.0 (Nougat) or later. Once all these settings are set up, the user can complete this procedure by clicking on the "Build and Run". By doing that, the project will be built and utilized for Oculus Quest 2.

A.1.1.8 Troubleshooting during the Unity project setup

During the setup of the Unity Project there were issues representation of the Oculus Quest 2 controllers during the virtual reality session in Unity. More specifically, the controllers would look like non-moving objects as part of the scene without following the movements of the actual controllers of the annotator during the experimental phase. As a consequence, this had a major impact in the creation of the experiment, as the participants wouldn't be able to see their controllers, the different buttons with their respective information, and the controller beamer (laser). Without the features, the annotation experiment would not be completed.

This issue can be solved by creating, through Visual Studio, a new script with the following code:

```

using System.Collections;
using System.Collections.Generic;
using UnityEngine;

public class ControllerMovement : MonoBehaviour
{
    public GameObject
    LeftcontrollerObject; public GameObject
    RightcontrollerObject;
    // Start is called before the first frame update
    void Start()
    {

    }

    // Update is called once per frame
    void Update()
    {
        if (gameObject.name.Contains("Left"))
        {
            transform.position = LeftcontrollerObject.transform.position;
            transform.rotation = LeftcontrollerObject.transform.rotation;
        }
        else
        {
            transform.position = RightcontrollerObject.transform.position;
            transform.rotation = RightcontrollerObject.transform.rotation;
        }
    }
}

```

A.1.2 Unity Project package settings

According to Knuiman's (2022) guide, at this point the scene must include four important elements for the smooth operation of the project namely, LAZ Loader, LAZ Renderer, XR Wander, and XR Network Manager. Each of the following presented features is based on the guideline provided by Knuiman (2022), with additions and comments related to the settings of this particular project.

A.1.2.1 LAZ Loader

It is important to mention that only one LAZ Loader should be used in each scenario; otherwise the correct function of the scene might be compromised (Knuiman, 2022). For more information on the inspector's various settings, the user can get a tooltip by hovering over them with the mouse. Some of the most significant features and their functions, according to the author, are described below:

Cam

This should include information about the camera since it will be used to find neighbouring octree cells and produce more compact point clouds (Knuiman, 2022). In the project of this thesis, no further changes were made regarding the Cam settings (Figure AI.1).

Target Projection

If several point clouds need to be aligned but use various projections, one must be picked as the standard. We'll use this standard projection to bring all point clouds into the same format. In the case where the origin and destination are the same, nothing takes place (Knuiman, 2022). In this project, the default settings were selected (Figure AI.1).

Target Projection Custom WKT

This option supersedes the Target Projection and enables any feasible projections. The user can discover a projection's WKT value here. The code should be copied and pasted it into this form on <https://epsg.io/> (Knuiman, 2022). In this project, there was not a particular projection selected, as the created environment doesn't indeed to represent a real scene, but rather an artificial one for the sake of measuring the different usability factors during the annotation process (Figure AI.1).

Tree Setup

Maximum tree depth limits the number of points used in the initial tree configuration per cell before branching out. It is noted by the developer of the tool that it is not recommended to select further than 4 depths. It should be taken into account that the number of cells is roughly 8^{depth} (Knuiman, 2022). In this project, 4 depths were selected as it is the maximum recommended value (Figure AI.1).

Number of LODS

This field characterises the number of iterations of a point cloud that exist inside a given volume. Due to possible memory requirements issues, it is recommended by the developer of the tool to use 4 as the maximum value (Knuiman, 2022). In this project, the maximum recommended value was selected (Figure AI.1).

Divide method

The user can specify whether higher LOD levels (fewer points) are created linearly (each time a smaller percentage is subtracted) or quadratically, for instance, by $1/(8^{\text{LODlevel}})$ num point (Knuiman, 2022). The initial value, as the developer states, should be LOD=0. The original value was selected for this project.

Cache to scene

In case the users decide to use scene caching, the octree will not have to be reconstructed from the original file every time. This, however, necessitates storing the information in the actual scene. This causes a substantial increase in the scene file, as Knuiman (2022) suggests. No scene caching was used in this project.

Cached meshes

To aid with debugging, it is helpful to view the contents of the scene file and the final result will not be visible to the user. Furthermore, it is important to note that loading a point cloud file while making changes to the Tree configuration might cause unexpected results (Knuiman, 2022).

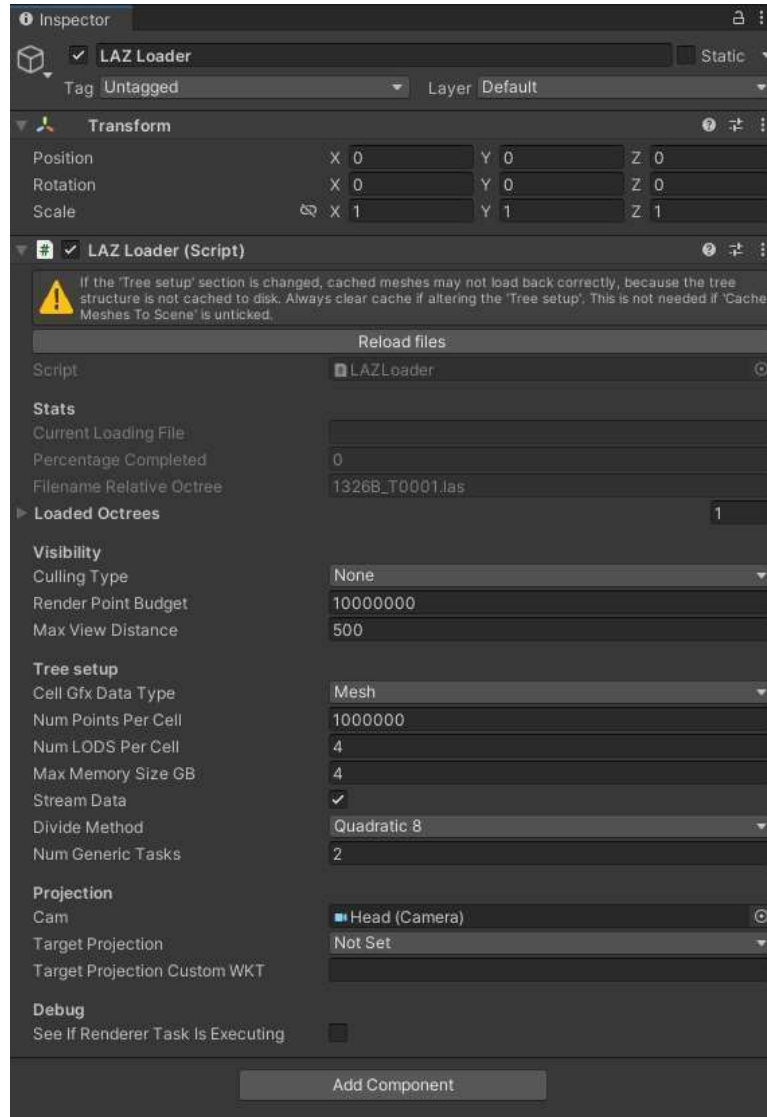


Figure A1.1: Different setting of LAZ Loader in the Unity Project environment.

A.1.2.2 LAZ Renderer

Each renderer is like a Mesh Renderer in that it depicts a cloud of points that must be drawn. It is only loaded once, even if the same point cloud is being used by numerous renderers (Knuiman, 2022). During the creation of the annotation scene of this thesis, each tree was imported individually into Unity by clicking on the Select File feature.

Projection

This method is identical to that used by the LAZ Loader. If the LAZ Renderer and the LAZ Loader are both the same, then no conversion is necessary (Knuiman, 2022). The same

applies to the project of this thesis. No further actions were taken during the setup of this parameter (Figure AI.2).

WKT Projection

As in the setup of the projection parameter of LAZ Loader, the user should look up projection WKT codes to the following link, as this parameter supersedes Projection (Figure AI.2). To copy and paste the proper link, the user should advice <https://epsg.io/> (Knuiman, 2022).

Place type

As the developer of the tool suggests, a signlerenderer can be set as the default. If the user decides to choose the operation "Relative to one Particular," it will be positioned with respect to the one user entered as the origin, as Figure AI.2 suggests. The centre of gravity may be set at will for each LAZ Renderer. As an alternative, the user may choose the world option as the centre. That is not the case in comparison to other rendering methods.

Visualisation

The author suggests that the user should keep the mouse pointer where the parameters are to see the help text for further information. Point clouds can be observed in one of four different ways. Shade, Height, Power, and Grouping. In order for the user to conceptualise the meaning of each of them, LIDAR needs to be performed. Additionally, colours can be associated with their numerical values. It is beyond the scope of the provided guideline, as well as the present thesis, to go through each parameter in detail. As a consequence, it is considered important that the users experiment around with the settings to obtain a thorough grasp. The provided settings can offer significant enhancement to the rendering quality of a certain LAZ Renderer.

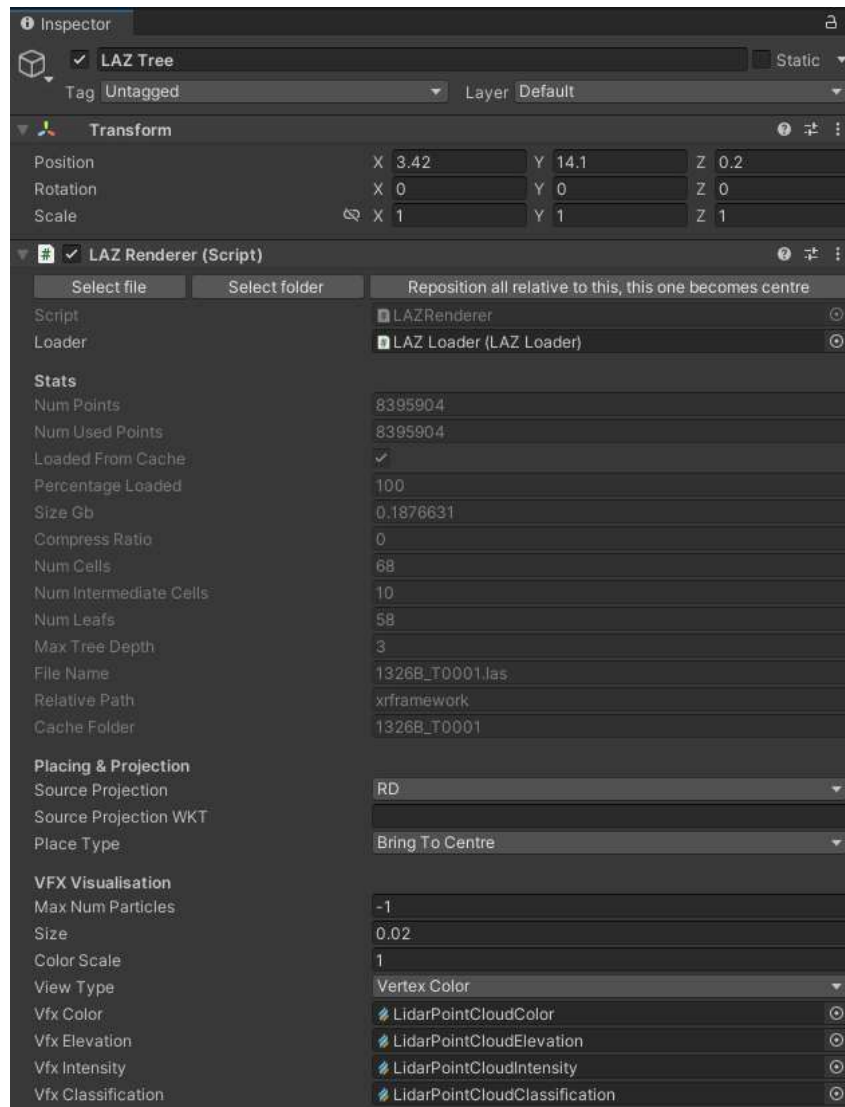


Figure AI.2 Different setting of LAZ Renderer in the Unity Project environment.

A.1.2.3 Creating a new room/ Join an existing room

At the beginning of the experiment, every participant will be briefly introduced to some key information, such as the purpose of the experiment and the variety of activities that they are responsible for carrying out. In addition, before the experiment begins, they will be given information regarding the various capabilities of the controllers and how each action can enhance the process of annotating. The experiment may start as soon as the participants have placed on the Oculus Quest 2 and made the necessary adjustments to it.

The annotating procedure will be performed in a "new room" that each participant is responsible for creating as soon as they have entered the virtual reality environment. This can be done by aiming the beam of any controller in the "Create new room" and pressing the front trigger. Since the experiment has developed into its current shape, it has been decided that each participant would complete the annotation in their "own room". This will allow the experimenter to supervise, capture, and record each participant's activities independently. Also, each activity that is carried out by the participant will be immediately saved. If the participants

wish to pause the experiment for whatever reason, they are free to return at a later time and pick their personal room from the list that is labelled "Available rooms" to pickup where they left off in their annotations (Figure AI.3).



Figure AI.3: Create or join room startup menu.

A.1.2.4 Further settings during the experiment

Once the participant enters their private scene, before and during the annotation process, there is a possibility of changing and adapting specific features related to the tree point cloud, although that is not recommended by the conductors of the experiment, as different settings may alter the conceptualization of the data and lead to non-comparable results. As a result, this may compromise the outcome of the experiment, as the quality and appearance of the trees will not be the same for each participant, leading to unreliable results. The exact settings of the experiment will be decided by the conductors and will be set for each participant beforehand.

As it is mentioned above, the user can enter the Settings menu by pressing the Y button on the left controller. As it is presented in Figure 5, the first characteristic of the point cloud that can be changed is the LAZ colorization. By aiming the beam of any controller and pressing the trigger, the participant can choose between four different colorization options namely Color, Elevation, Classification, and Intensity. The default setting is the Classification option, and it is recommended that should not be changed during the experiment as it is the one that depicts the annotated tree parts with different colors for each class (Figure AI.4).

Furthermore, the next setting option is the Global LOD. In point clouds, the Level of Detail (LOD) is normally set by the point density and accuracy requirements of each application. A larger number of points and greater data about the scanned item or surroundings characterize higher LOD point clouds. Higher LOD, however, necessitates more space and processing power. This value is set by default as medium by the developer of this application (Figure AI.4).

The third feature, namely Quality, refers to the quality of the representation of the tree point cloud. This value will be set to maximum as default for each participant. The last feature is related to Post Processing processes. The term "post-processing" is commonly used to describe a variety of procedures that are applied to point cloud data in order to

enhance its precision, accuracy, and usability. By default, Post Processing should be enabled.



Figure AI 4.: Further settings.

Appendix 2: Questionnaires

A.2.1 Demographics Form:

Please answer the following questions related to basic demographics information. All of the answers you provide will be kept confidential and will only be used for research purposes.

Age: [Fill in your answer]

Gender: [Select one of the following options]

Male

Female

Non-binary

Other

Prefer not to say

Educational Level: [Select one of the following options]

Less than high school

High school diploma or equivalent

Some college or associate's degree

Bachelor's degree

Master's degree

Doctorate degree

How familiar are you with virtual reality experiences? (HTC Vive, Oculus Rift, Samsung GearVR, Google Cardboard, etc?): [Choose one of the answers below]

Never used

Beginner

Intermediate

Advanced

Expert

How familiar are you with virtual reality experiences? (HTC Vive, Oculus Rift, Samsung GearVR, Google Cardboard, etc?): [Choose on of the answers below]:

Never used

Beginner

Intermediate

Advanced

Expert

A.2.2: System Usability Scale (SUS) Index Questionary

1. I found the [approach name] annotating method to be easy to use.
2. I felt confident using the [approach name] annotating method to annotate tree components.
3. I found the [approach name] annotating method to be unnecessarily complex.
4. I thought the [approach name] annotating method was consistent in its performance.
5. I found the various features of the [approach name] annotating method to be well-integrated.
6. I thought there was too much inconsistency in the [approach name] annotating method.
7. I would imagine that most people would learn to use the [approach name] annotating method very quickly.
8. I found the [approach name] annotating method very cumbersome to use.
9. I would need the support of a technical person to be able to use the [approach name] annotating method.
10. I felt very confident using the [approach name] annotating method to annotate tree components.

A2.3: NASA Task Load Index (NASA- TLX) Questionary

1. Mental Demand: How mentally demanding was the task of annotating using [approach name] method?
2. Physical Demand: How physically demanding was the task of annotating using [approach name] method?
3. Temporal Demand: How time-pressured was the task of annotating [approach name] method?
4. Performance: How successful do you feel you were in annotating the data using [approach name] method?
5. Effort: How much mental and physical effort did you need to put in annotating the data [approach name] method?
6. Frustration: How frustrating was the task of annotating using Sphere Selection method compared to Container Selection method?

A.2.3: Final Questionary- Comments

1. To what extent did you experience simulator sickness during the virtual experience?
2. How much general discomfort are you experiencing right now?
3. What do you think are the advantages and disadvantages of the method of Sphere Selection?
4. What do you think are the advantages and disadvantages of the method of Containter Selection?
5. What is your opinion about the combination of these two methods?