Concepts for operating ground based rescue robots using virtual reality

Bachelor thesis by Jingyi Jia
Date of submission: September 6, 2021

1. Review: Prof. Dr. Max Mühlhäuser
2. Review: Julius von Willich
Darmstadt

TECHNISCHE UNIVERSITÄT DARMSTADT

Computer Science
Department
Telekooperation
Prof. Dr. Max Mühlhäuser

Erklärung zur Abschlussarbeit
gemäß §22 Abs. 7 und §23 Abs. 7 APB der TU Darmstadt

Hiermit versichere ich, Jingyi Jia, die vorliegende Bachelorarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Mir ist bekannt, dass im Fall eines Plagiats (§38 Abs. 2 APB) ein Täuschungsversuch vorliegt, der dazu führt, dass die Arbeit mit 5,0 bewertet und damit ein Prüfungsversuch verbraucht wird. Abschlussarbeiten dürfen nur einmal wiederholt werden.

Bei der abgegebenen Thesis stimmen die schriftliche und die zur Archivierung eingereichte elektronische Fassung gemäß §23 Abs. 7 APB überein.

Bei einer Thesis des Fachbereichs Architektur entspricht die eingereichte elektronische Fassung dem vorgestellten Modell und den vorgelegten Plänen.

Darmstadt, 6. September 2021

Jingyi Jia

Jingyi Jia

Abstract

Rescue robotics have been increasingly used in crisis situations, such as exploring areas that are too dangerous for humans, and have become a key research area. A number of studies use Virtual Reality (VR) to improve the degree of immersion and situation awareness. However, there remains a need to explore an intuitive, easy-to-use interaction pattern, which increases the efficiency of search and rescue and allows the user to explore the environment intentionally. This thesis presents a preliminary VR-based Human-Robot Interaction system in terms of ground based rescue robots, with the aim to find an ideal interaction pattern. The proposed system offers four different operation modes and corresponding test scenes imitating a post-disaster city. This thesis includes a user study in which four different operation modes are tested in turn. The user study reveals that the ideal interaction pattern should reduce the complexity of the operation as much as possible. Instead of allowing the user to adjust the robot's direction and speed themselves, it is recommended to set a target point and let the robot navigate to the target point automatically.

Zusammenfassung

Rettungsroboter werden zunehmend zur Bewältigung von Krisensituationen eingesetzt z.B. dort, wo es für den Menschen zu gefährlich ist. Eine Reihe von Studien nutzt VR als Plattform für die Mensch-Roboter-Interaktion, da dies die Immersion erhöhen und das Situationsbewusstsein verbessern kann. Es besteht jedoch nach wie vor Bedarf für eine einfach zu bedienende Interaktionsmethode, um das Personal bei der Steuerung des Roboters zu entlasten und die Effizienz und Effektivität von Such- und Rettungsarbeiten zu erhöhen. In dieser Abschlussarbeit wird ein vorläufiges VR-basiertes Mensch-Roboter-Interaktionssystem in Bezug auf bodenbasierte Rettungsroboter vorgestellt. Ziel ist, ein möglichst intuitive Steuerungsmethode mit geringer mentaler Ermüdung zu finden. Das vorgeschlagene System bietet vier verschiedene Betriebsmodi und entsprechende Testszenen. Diese Arbeit beinhaltet eine Nutzerstudie (die vier Betriebsmodi werden nacheinander getestet und verglichen). Das Fazit ist, dass die ideale Interaktionsmethode die Komplexität der Steuerung so weit wie möglich zu reduzieren.

Contents

1 Introduction

In recent years, natural disasters such as earthquakes, tsunamis and potentially nuclear explosives have seriously threatened the safety of human life and property. The number of disasters, as well as their severity and complexity, have gradually increased. The first 72h after a disaster is the golden rescue time, but the unstructured environment of disaster sites makes it difficult for rescuers to work quickly, efficiently and safely.

Rescue robots have the advantages of high mobility. They can work continuously to improve the efficiency of search and rescue. In addition, those robots can map the terrain and detect sound, gas and temperature within the ruins by carrying a variety of sensors. Rescue robots can assist or replace the rescuers, thus avoiding injuries caused by the secondary collapse and reducing risks for rescuers.

In fact, rescue robots have already been used in a number of disaster scenarios. The Center for Robot-Assisted Search and Rescue (CRASAR) used rescue robots for Urban Search and Rescue task during the World Trade Center collapse in 2001 [1] and has employed rescue robots at multiple disaster sites to assist in finding survivors, inspecting buildings and scouting the site environment etc. [2]. Anchor Diver III was utilized as underwater support to search for bodies drowned at sea after the 2011 Tohoku Earthquake and Tsunami [3].

Considering the training time and space constraints for rescuers [4], and the goal of efficiency and fluency collaboration [5], the appropriate Human-Robot Interaction approach deserves to be investigated. Some of the existing Human-Robot Interaction solutions include Android software [6] [7], gesture recognition[8] [9] [10], facial voice recognition [11], adopting eye movements [12], Augmented Reality (AR) [13] and VR.

Among them, VR has gained much attention due to its immersion and the interaction method that can be changed virtually. VR is no longer a new interaction technique. With the development of technology in recent years, VR devices are gradually becoming more accessible to users. With the improvement of hardware devices, the new generation of VR headsets has higher resolution and a wider field of view. While VR headsets are often considered as entertainment devices, VR brings more than that. It plays an important role in many fields such as training, education and medical care.

The use of VR in Human–Robot Collaboration also has the potential. In terms of reliability, VR is reliable as a novel alternative to Human-Robot Interaction. The interaction tasks that users can accomplish with VR do not differ significantly from those using devices in real environments [14]. In terms of user experience and operational efficiency, VR headsets can provide users with stereo viewing cues, which makes collaborative Human-Robot Interaction tasks in certain situations more

efficient and performance better [15]. A novel VR-based practical system for immersive robot teleoperation and scene exploration can improve the degree of immersion and situation awareness, thus increasing the accuracy of the robot's navigation and the awareness of objects in the scene. In contrast, this level of immersion and interaction cannot be reached with video-only systems [16].

However, there remains a need to explore Human-Robot Interaction patterns and improve the level of Human-Robot Integration [17]. Intuitive and easy-to-use interactive patterns can enable the user to explore the environment intentionally and improve the efficiency of search and rescue as much as possible. The appropriate interaction method should cause less mental and physical exhaustion, reducing the need for the user to frequently exit the VR environment for rest.

For this purpose, this thesis presents a preliminary VR-based system that simulates the cooperation between ground based rescue robots and humans with four different operation modes and corresponding test scenes, which imitate a post-disaster city. The test scene simulates a robot leveraging Unity to construct a virtual 3D scene. The proposed system simulates the process that the robot uses Light Detection and Ranging (LiDAR) to detect the surrounding environment. In order to find an interactive approach that is as intuitive and easy-to-use as possible, a user study was performed after the development was completed.

In Chapter 2, related work involving the integration of VR and Human-Robot Interaction is presented. Chapter 3 provides details of the proposed system, including the techniques used for the different interaction modes and the setup for test scenes. Chapter 4 explains the design and procedure of the user study. Chapter 5 and 6 present the results of the user study and analyze the advantages and disadvantages of the different operation modes and the directions for future work. Finally, in Chapter 7, the thesis is concluded.

2 Related Work

In this chapter, some research on the integration of VR and Human-Robot Interaction will be discussed. The relevant literature and its contributions will be briefly presented.

The topic of VR and Human-Robot Interaction is an open research topic with many kinds of focus perspectives.

Human-Robot Interaction platforms combined with virtual worlds have many applications. It can be used, for example, to train machine operators in factories. Elias Matsas et al. [18] provided a VR-based training system using hand recognition. Kinect cameras are used to capture the user's positions and motions, and virtual user models are constructed in the VR environment based on the collected data. Users will operate robots and virtual objects in the VR environment, and in this way, learn how to operate the real robot. The framework proposed by Luis Pérez et al. [19] is applied to train operators to learn to control the robot. Since the environment does not need to change in real time, but rather needs to recreate the factory scene realistically, a highly accurate 3D environment was constructed in advance using Blender in combination with a 3D scanner.

Building 3D scenes in virtual worlds based on information collected by robots is also a research highlight. Wang, et al. [17] were concerned with the visualization of the rescue robot and its surroundings in a virtual environment. The proposed Human-Robot Interaction system uses the incremental 3D-NDT map to render the robot's surroundings in real time. The user can view the robot's surroundings in a first-person view through the HTC VIVE and send control commands through arrow keys on the motion controllers. A novel VR-based practical system is presented in [16] consisting of distributed systems to reconstruct the 3D scene. The data collected by the robot is first transmitted to the client responsible for reconstructing the scene. After the client has constructed the 3D scene, the set of actively reconstructed visible voxel blocks is sent to the server responsible for communication, which has a robot-based live telepresence and teleoperation system. This server will then broadcast the data back to the client used by the operator, thus enabling an immersive visualization of the robot within the scene.

Others are more concerned about the manipulation of the robotic arm mounted on the robot. Moniri et al. [20] provided a VR-based operating model for the robotic arm. The user wearing a headset can see a simulated 3D scene at the robot's end and send pickup commands to the remote robot by clicking on the target object with the mouse. The system proposed by Ostanin et al. [21] is also worth mentioning. Although their proposed system for operating a robotic arm is based on Mixed Reality (MR), the article is highly relevant to this thesis, considering the correlation of MR and VR and the proposed system detailing the combination of Robot Operating System (ROS) and robotics.

In their system, ROS Kinect is used as middleware and responsible for communicating with the robot and the Unity side. The user can control the movement of the robot arm by selecting predefined options in the menu. In addition, the orbit and target points of the robot arm can be set by clicking on a hologram with a series of control points.

To summarize, previous work has studied methods and tools for VR-based Human-Robot Interaction and teleoperation. However, only few studies focus on the different interactive approaches for Human-Robot Interaction.

3 Implementation

In this chapter, the tools and techniques used in building this VR-based Human-Robot Interaction system are described. Particular emphasis is given to the interaction techniques for different operation modes. In addition, the setup of the robot and the construction of test scenes will also be covered in this chapter.

3.1 Overview

One of the aims of this thesis is to design and implement a VR-based Human-Robot Interaction system with different methods of operating the robot in order to find out which operation method is more suitable to control the rescue robot. Further, this thesis aims to provide some fundamental insights for future development directions and to provide a general direction for finding an intuitive, easy-to-use and efficient interaction approach for Human-Robot Interaction. Therefore, the proposed system was developed using Unity, including four operation modes and corresponding test scenes for simulating post-disaster scenarios. In each operation mode, the user has a different method to control the robot. In addition, the system should simulate the process by which the robot scans its surroundings and the computer side cumulatively gets a reconstructed 3D virtual scene. The test environment was thus implemented in such a way that the scene seen by the user depends on the robot's movement and the trajectory it travels through.

3.2 System Architecture

The proposed system uses HTC VIVE as VR device. It has a resolution of 1080 × 1200 per eye, resulting in a total resolution of 2160 × 1200 pixels, a refresh rate of 90 Hz, and a field of view of 110 degrees. It includes two motion controllers and uses two Lighthouses to help track the headset's position and the motion controllers.

Unity was chosen as the platform to develop the system. Unity is a widely used game engine with Steam VR plugin [1], which allows developers to focus on the VR environment and interactive behaviors in programming, rather than specific controller buttons and headset positioning, making

---

[1]https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647

VR development much simpler. Another reason why Unity was chosen as a development platform was its potential for collaboration with ROS, a frequently used operating system for robot simulation and manipulation, which is flexible, low-coupling, distributed, open source, and has a powerful and rich third-party feature set. In terms of collaboration between Unity and ROS, Siemens provides open-source software libraries and tools in C# for communicating with ROS from .NET applications [2]. Combining ROS and Unity to develop a collaborative Human-Robot Interaction platform proved to be feasible [22]. Since the focus of this work is on Human-Robot Interaction, collaboration and synchronization of ROS will not be explored in detail here.

3.3 Robot

The proposed system needs to simulate the process that a robot uses a LiDAR remote sensor to detect the real environment and synchronize it to Unity. Thus, a sphere collision body was set up on the robot as seen in 3.1. The collider is an invisible Unity component that can interact with other colliders by collision detection and a trigger event (onTriggerEnter function). In this system, the sphere collision body defines the range of the robot's LiDAR detection. It will interact with invisible objects in the scene, such as buildings, during the robot's movement, and transform the Layers of the objects into visible Layers. The robot's driving performance, such as the number of collisions, average speed, total distance, will be recorded in each test. The detailed recorded information can be seen in Chapter 5. The movement of the robot depends on the value of the signal that is updated in each mode. In addition, the robot's Gameobject has the NavMeshAgent [3] component, which supports the robot's navigation to the specified destination with automatic obstacle avoidance in the test scene. The virtual robot has three cameras. One of the cameras is a simulation of a surveillance camera mounted on the robot, which can see all the items in the scene, although the distant items are not yet detected by LiDAR. Two of these cameras are set up in such a way that they can only see the area detected by LiDAR. Each camera captures what it sees and modifies the bound image in real time. The four operation modes described later all use the camera viewport as a monitoring screen by rendering the camera viewport on UI canvas.

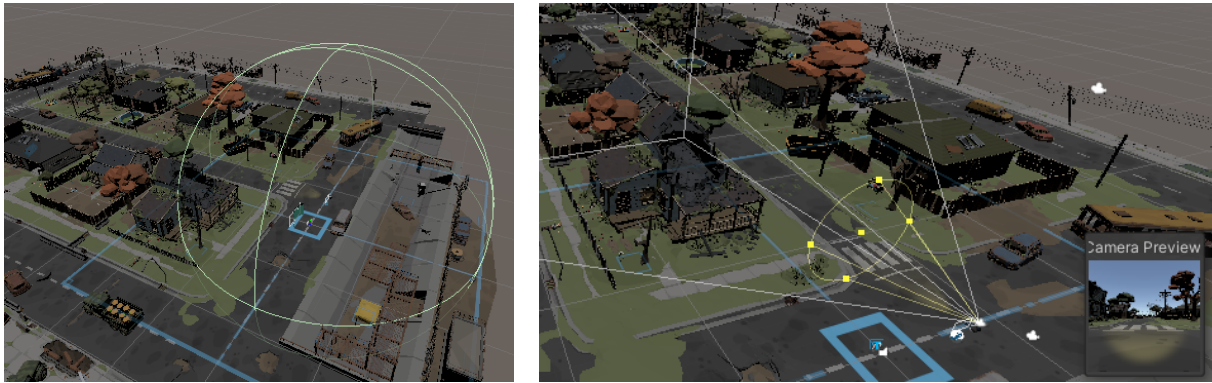## 3.4 Interaction techniques

This system has four different approaches to control the robot. Each mode has its own distinctive features:

1. In Handle Mode, the user will send control commands directly using the motion controller.

2. In Lab Mode, a simulated lab is constructed in the VR environment and the user will use virtual buttons in the lab to control the rescue robot.

---

[2]https://github.com/siemens/ros-sharp
[3]https://docs.unity3d.com/ScriptReference/AI.NavMeshAgent.html

(a) LiDAR Collider

(b) Surveillance camera

Figure 3.1: Robot

3. In Remote Mode, the user can set the driving destination directly.

4. In UI Mode, the user has a virtual menu and sends commands via rays from the motion controller.

In order to improve the reusability of the code and facilitate the management of subsequent development, the classes that manage the interaction actions of each mode implement the same interface. A graphical representation of the system structure is given in the UML activity diagram in Figure 3.2.

3.4.1 Handle Mode

In this mode, the user controls the robot's movement directly through the motion controller in the right hand. The touchpad of the motion controller determines the direction of rotation of the robot. The user can control the robot's driving speed by pulling the Trigger button. Figure 3.4 shows the HTC VIVE motion controller. The robot rotation direction will read the value of the touchpad X-axis. The range of values is $[-1, 1]$. Forward speed reads the Trigger button passed in as a variable of type SteamVR_Action_Single, and the range of the variable is $[0, 1]$. With the right-hand menu button, the surveillance screen can be turned on or off. The monitor window can be adjusted to a suitable position by dragging and rotating it. In the literature dealing with VR and Human-Robot Interaction, many researchers have used a similar operational approach. Therefore, as a widely used, and in a sense default operation approach, this mode was designed as one of the operation modes.

3.4.2 Lab Mode

This pattern was designed with reference to the system proposed by [18][19]. Their frameworks are used to train operators to work with the robot, avoiding risks and saving learning costs. In addition, they also mentioned that being in a simulated factory or laboratory can improve immersion. Therefore, in this mode, a virtual laboratory environment is constructed, in which simulated buttons,
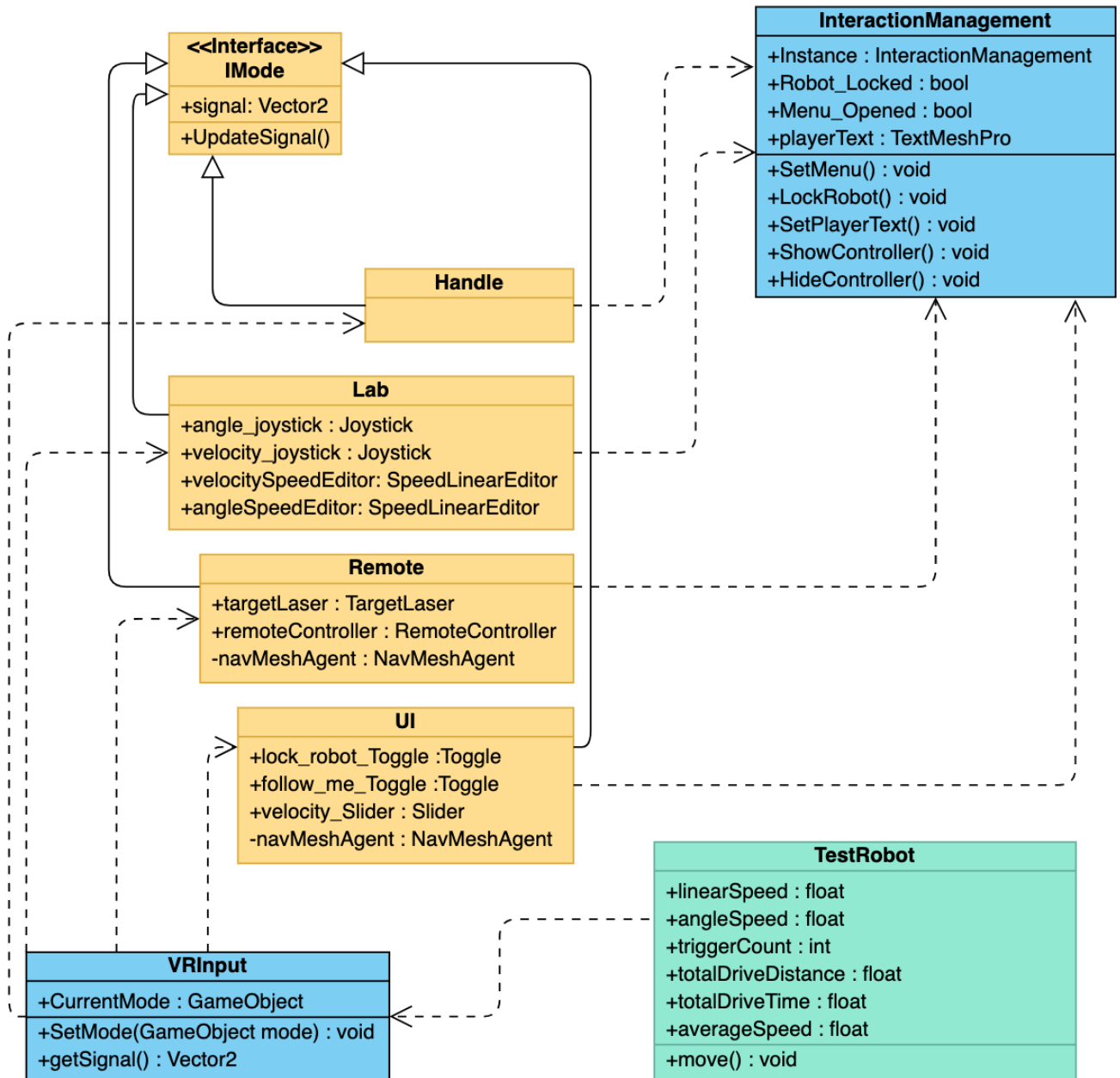
Figure 3.2: UML Class diagram for the main structure of the system
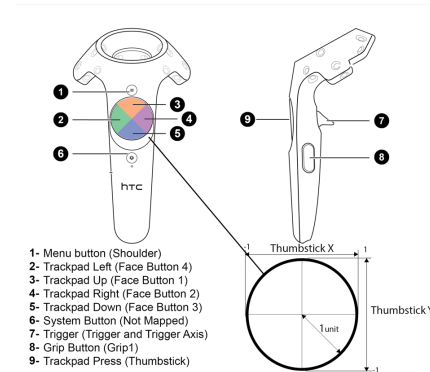
Figure 3.3: Handle Mode

Figure 3.4: HTC handle illustration

1- Menu button (Shoulder)
2- Trackpad Left (Face Button 4)
3- Trackpad Up (Face Button 1)
4- Trackpad Right (Face Button 2)
5- Trackpad Down (Face Button 3)
6- System Button (Not Mapped)
7- Trigger (Trigger and Trigger Axis)
8- Grip Button (Grip1)
9- Trackpad Press (Thumbstick)

controllers, and monitoring equipment are placed. The laboratory consists of two parts. The first part is the monitoring equipment: the monitoring screen is enlarged and placed at the front of the lab as a huge display. The second part is the operating console in the center of the laboratory, which can be moved by the user as desired. This is due to the fact that users have different heights and may wish to operate the robot in a standing or sitting position. The user can use the buttons on the table to lock the robot or let it move forward automatically. In the middle of the console are two operating joysticks that determine the robot's forward motion and rotation respectively. The part that involves virtual joystick movement and button effects uses an open-source GitHub project VRtwix[4]. With the sliding stick on the left, the user can edit the speed of the robot's forward movement and rotation.
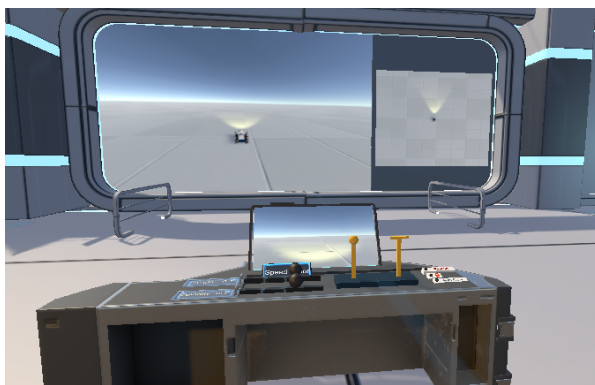
(a) Overview
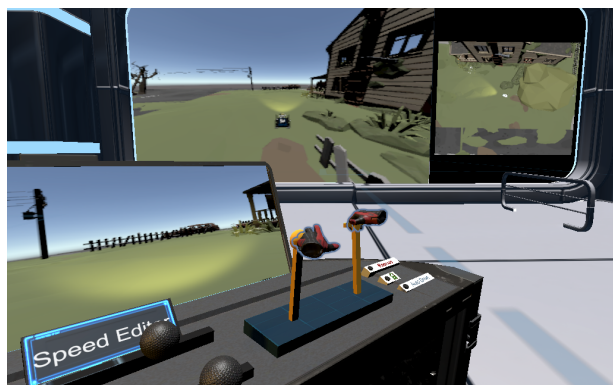
(b) Operating console

Figure 3.5: Lab Mode

3.4.3 Remote Mode

In this mode, the user can set the driving target point directly or control the robot by picking up the remote control placed on the toolbar. The target point is set by the ray emitted by the right

---

[4]https://github.com/rav3dev/vrtwix

motion controller. This process is similar to setting a teleportation point. After the target point is set, a square representing the destination is shown in the scene, and the robot will automatically travel to the set destination. The entire driving process uses the NavMeshAgent component and is therefore capable of automatic obstacle avoidance. A movable toolbar with remote control and a monitoring device can be opened by clicking on the menu button. The remote control is a backup precaution if the automatic navigation fails to navigate the target point properly. The user can adjust the direction of the robot's travel by using the remote control. The pickup and auto-release parts use the ItemPackage component available in the Steam VR plugin.

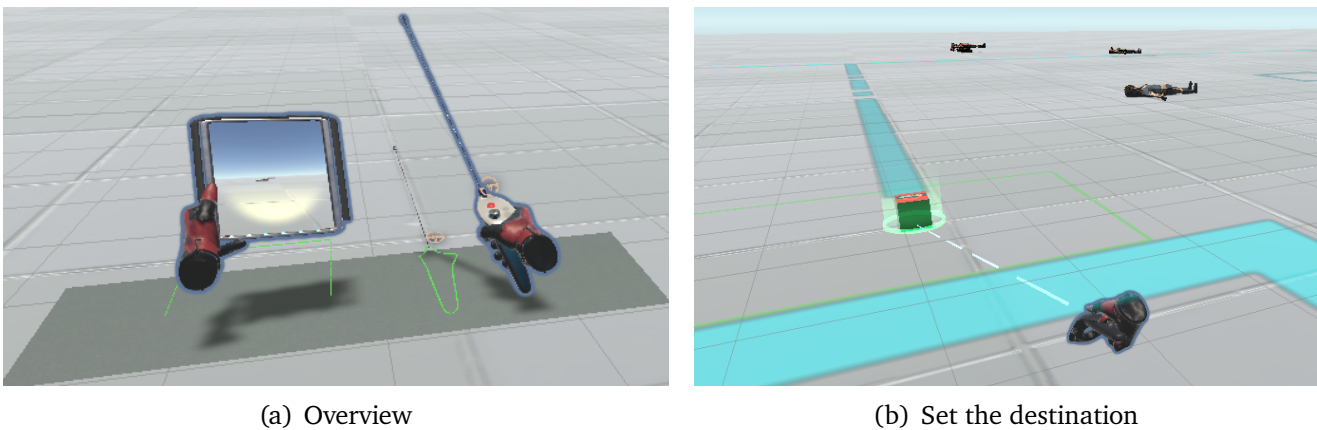(a) Overview   (b) Set the destination

Figure 3.6: Remote Mode

3.4.4 UI Mode

The virtual menu is also an interaction method that is often used in VR, so this mode is proposed. In this mode, the user interacts with the virtual menu using the ray emitted by the right motion controller. The virtual menu is set up with buttons for the direction of movement, a speed controller, and buttons to open and close the monitor screen. In addition to this, an additional follow function is added to the menu, allowing the robot to follow the user's position in the virtual world. This is intended to let the user concentrate on observing the rendered VR environment. Also, having a real robot following the user's location in the virtual world is a novel, unique Human-Robot Interaction approach in VR. The robot's automatic navigation uses the NavMeshAgent.
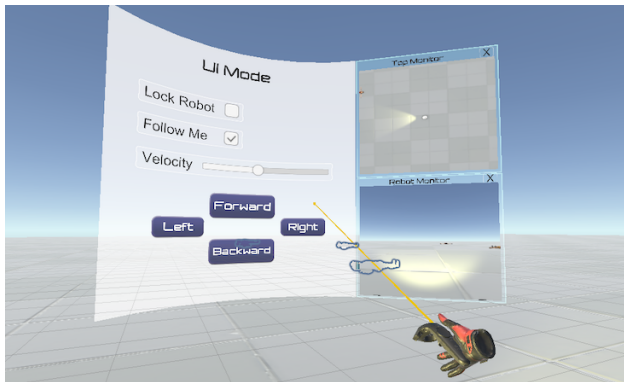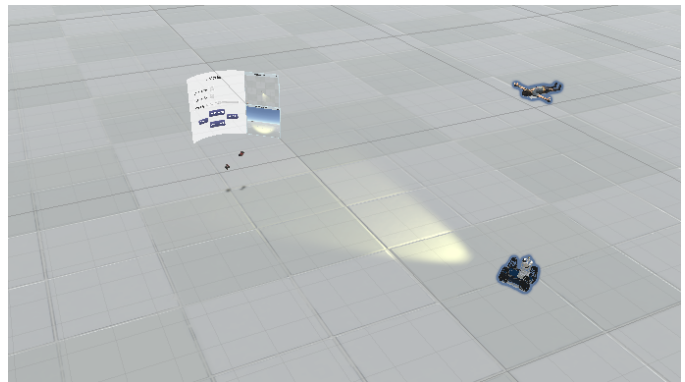
3.5 Test Scene

In order to simulate the use of rescue robots in disaster scenarios, the test scenes were built to mimic the post-disaster urban environment as much as possible. The POLYGON Apocalypse [5], available on

---

[5]https://assetstore.unity.com/packages/3d/environments/urban/polygon-apocalypse-low-poly-3d-art-by-synty-154193

(a) Overview

(b) Follow Function

Figure 3.7: UI Mode

the Unity Asset Store, is a low poly asset pack with a large number of models of buildings, streets, vehicles, etc. This resource pack was used as a base. Additional collision bodies of the appropriate size were manually added to each building and obstacle after the resource pack was imported, which was needed to help track the robot's driving crash in subsequent tests.

Considering that four operation modes need to be tested, four scenes with similar complexity and composition but different road layouts and placement of buildings were constructed. The similarity in complexity of the scenes ensures that the difficulty of the four tests is basically identical. The different scene setups ensure that the scene information learned by the user after one test will not make him understand the next test scene and thus affect the accuracy of the test data.

The entire scene is initially invisible, and the visibility of each object in the test scene is gradually updated as the robot drives along. Ten interactable victim characters were placed in each test scene. The placement can be next to the car, the house side and some other reasonable places.

(a) Obstacle
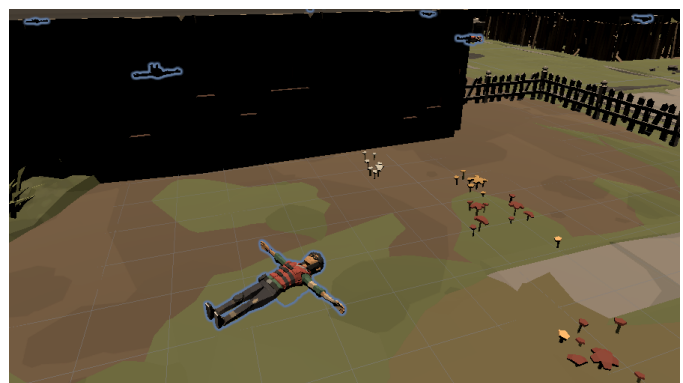
(b) Victims

Figure 3.8: Test Scene

4 Evaluation of User Experience

This chapter describes the design and detailed process of the user evaluation. The purpose of the performed user study is to measure the impact of four different operation modes on rescue efficiency, robot driving performance, and psychological and physiological stress and fatigue. For this purpose, participants are asked to find victims in a test scene using different operation modes and answer questionnaires after the test corresponding to each mode of operation.

4.1 Study Design

The evaluation for each operation mode consists of two main parts. The first part is the data recorded during the process of the participant driving the robot in the VR environment to find the victims. The recorded data includes information about the robot's collision and the speed of driving etc. The rescue of the victims was also recorded. The second part is the questionnaires that participants filled out after each test. The Official NASA Task Load Index (TLX) was used to measure the participant's subjective workload assessments. Additionally, participants were asked specific questions for each mode and were asked to select their favorite and least favorite operation mode. In order to reduce the influence of order effects on the test results, the Balanced Latin Square was used when arranging the test order for the four operation modes.

4.2 Procedure

4.2.1 Demographics and Introduction

Before the beginning of the actual testing process, participants were informed of the purpose of this work, the broad process and the content of data that would be collected. After filling in the basic demographics and signing a consent form, the features of each of the four modes of operation and their rough usage were introduced verbally with a display of the buttons on the motion controllers.

4.2.2 Entering the world of VR

After the essential introduction part, participants would put on the VR headset and enter the VR environment to complete the rest of the tutorial. Considering that participants might not have experience with VR and that it would take time to learn how to operate the four different modes, the proposed system additionally sets up a practice pattern and places some models of victims in the practice scene. After entering the VR world, participants first needed to familiarize themselves with the opening and closing menu, as well as using the motion controllers to try to teleport themselves, or raise themselves into mid-air. Finally, participants were asked to interact with the victim model through virtual hands. After this series of tutorials, participants were already generally familiar with the use of VR and how to move around in the VR world.

4.2.3 Practice and evaluation of modes

Given the different manipulation approaches for each mode and possible confusion between the different modes, participants would take turns practicing and directly evaluating each mode immediately afterward.

The sequence of modes to be tested is predetermined. The order effect is an important factor affecting the test results. If the order of the operation modes to be tested would be the same for each participant, the psychological and physical exhaustion caused by the last operation mode would inevitably be more. In order to minimize the influence of the order effect on the results, the Balanced Latin Square with a size of four was used to arrange the test order of the four operation modes.

Participants automatically entered the practice scene corresponding to the relevant operation mode in the predefined order. After attempting to rescue 1-2 victims and when participants indicated that they were familiar enough with this operation mode, they would enter the test scene. In the test scene, participants had to save as many victims as possible in a given time limit. Participants were required to move the robot around the test scene to explore the post-disaster city and rescue victims. During this process, if the robot crashes with buildings or obstacles, besides the collision information being recorded as test data, participants would also receive sound and vibration feedback. The test would automatically end when time ran out or when all the victims on the scene have been rescued. Participants were required to complete the evaluation questionnaire and TLX form at the end of each test. This process was repeated in each mode of operation.

After all the tests were completed, participants were asked to compare the four operation modes and select the one they liked the most and the one they liked the least. In addition, participants could give their reasons for the choice and express their opinions as much as they wanted, such as suggestions for improvement or problems found during the operation.

5 Results

In this chapter, the results of the user study, obtained by the method described above, will be presented. The results are divided into three parts as follows:

- Participants: Participants' basic information will be counted.

- Quantitative Results: The collected data about the robot performance and participants' ratings will be displayed.

- Qualitative Results: The comments given by participants and the phenomena observed in the tests are analyzed.

5.1 Participants

A total of 8 volunteers participated in the user study (3 females and 5 males between 22 and 32 years, mean age 25.75 years). Five of them were computer science students at the university. Four participants had no previous experience with VR, and the other four participants had only limited experience.

5.2 Quantitative Results

Part of the data for the quantitative analysis comes from the robot's performance and testing results, which were automatically recorded by the proposed system during the tests. The other part of the data comes from the questionnaires that the participants filled out after the test.

5.2.1 Robot Performance

The overall robot performance is reported in Figure 5.1.

The number of collisions between the robot and objects reflects the probability of the robot being damaged. Lab Mode got the worst results with an average collision times of 26.75 with a standard error of 18.07. Handle Mode had the second-worst result with an average collision times of 21.5 with a standard error of 11.45. Remote Mode and UI Mode performed similarly, they both had a few collision times and a low standard error ($M_{Remote} = 2.25$, $SD_{Remote} = 1.79$, and $M_{UI} = 3.875$, $SD_{UI} = 1.96$).

In Lab Mode, the robot traveled the most distance and the most time. During the five-minute test period, the robot drove for an average of 243 seconds and covered a total of 753 meters. The average speed of the four modes did not differ significantly, but the standard error of the samples in Handle and Lab modes was significant. In both modes, it was found that some participants drove the robot very slowly and cautiously, while some participants tended to drive at maximum speed on the road and braked as soon as they noticed a distressed person. In Remote and UI modes, the robot's driving route was mainly controlled by the computer, so the average speed was basically the same and the deviation values were very small.
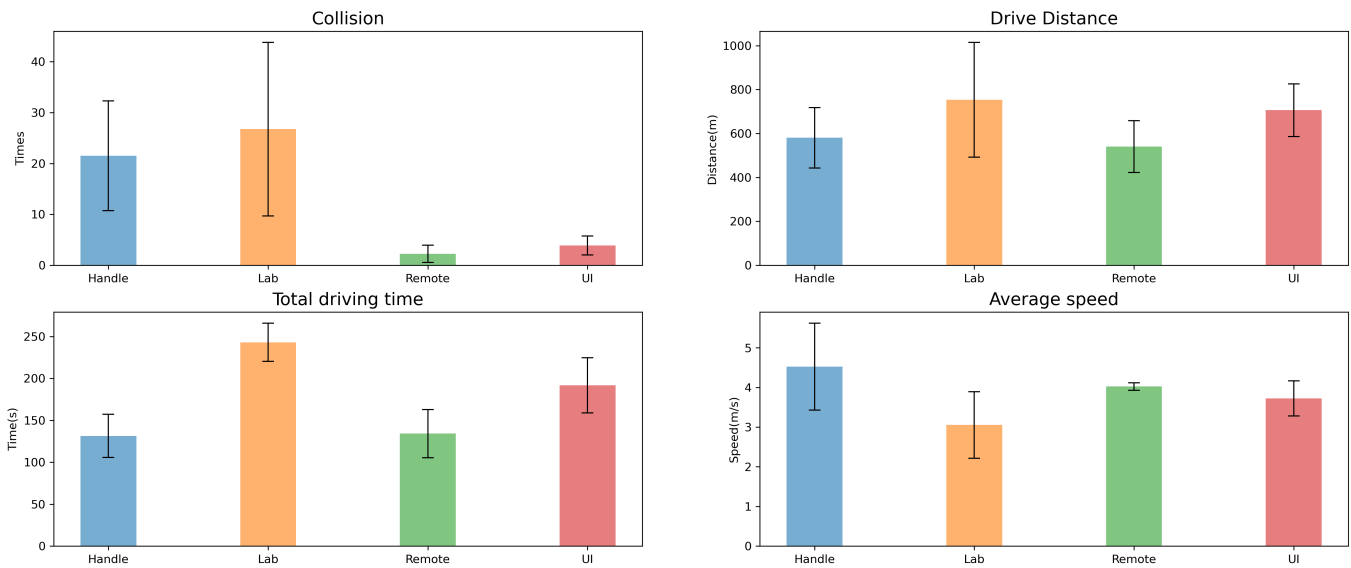

Figure 5.1: Robot Performance. (All error bars indicate the standard error.)

5.2.2 Rescue situation

The results of rescuing victims are shown in Figure 5.2. In general, the average number of rescued victims was the highest in Remote and UI modes. In both modes, some participants rescued all the victims within the time limit, and even completed the rescue task from half a minute to one minute
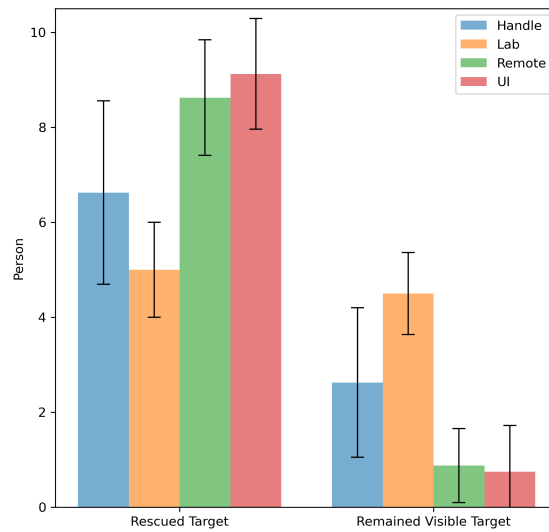
Figure 5.2: Rescue situation. (All error bars indicate the standard error.)

earlier. In the Lab Mode, the remaining visible victims were the most. This means that participants were more likely to overlook details in the scene, or even pass by the victim without noticing them. This could be attributed to the poor display quality of the screen, or it could be due to the complexity of the operation which makes the participant not have time to take into account every detail in the scene.

5.2.3 TLX Score

The result of TLX reveals the perceived cognitive workload. The smaller the score is, the less workload the operation mode brings to the participant. Figure 5.3 contains the mean and standard error of each scale. The standard error was significant for each scale because participants could only evaluate the workload of each mode of operation relatively, and they had different standard values in mind.

In this subsection, the total score will be analyzed. Three of the subjective subscales are selected additionally for detailed explanation.

As can be seen, similar score could be found for each scale in Lab Mode and Handle Mode, and their values were significantly worse than the other two modes. In the majority of cases, the average load in these two modes was almost twice as high as in Remote and UI modes. ($total_{Handle} = 57.29$, $total_{Lab} = 61.67$, $total_{Remote} = 30.24$, $total_{UI} = 28.23$)

The difference in scores was huge in terms of mental demand. ($mental_{Handle} = 73.125$, $mental_{Lab} = 61.25$, $mental_{Remote} = 21.43$, $mental_{UI} = 15.63$) The Handle Mode required the most mental activity. Participants said they needed to recall what buttons to press to achieve a certain effect or consider how to turn the robot to get it moving towards the desired position. They also had to pay attention to the distance between themselves and the robots so that the robots would not leave their sight.
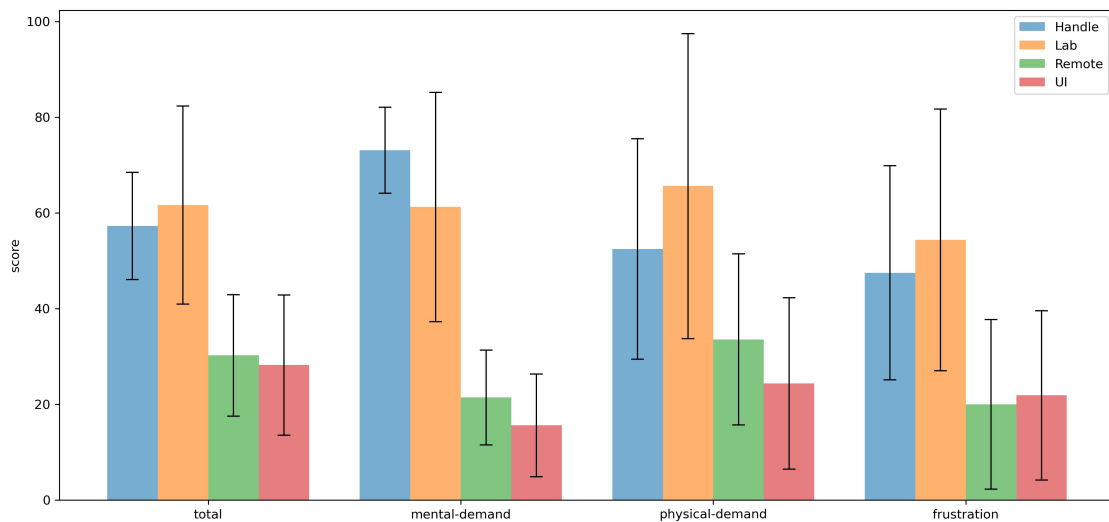
Figure 5.3: The average score of TLX. (All error bars indicate the standard error.)

Physical demand in Lab Mode was highest. It was observed that they frequently looked up and down to view different screens. In addition, some participants maintained their arms in a flat position while operating the virtual joystick, which also caused some fatigue on the arms.

In Lab Mode and Handle Mode, they felt more negative emotions, such as irritability, stress, etc. ($frustration_{Handle} = 47.5$, $frustration_{Lab} = 54.38$, $frustration_{Remote} = 20$, $frustration_{UI} = 21.88$.)

Overall, the Remote and UI modes received better scores on TLX.

5.2.4 Likert Questionnaire Results

After each test was completed, participants were also asked to fill out a questionnaire. The results can be seen in Figure 5.4. The questionnaire uses a 5-point Likert scale. 1 means that participants considered that their situations did not fit the question description at all. 5 indicates that their situations fit perfectly.

"I found it easy to move robot in desired position."

The Remote Mode received a highly uniform rating. All eight participants gave it the highest score of 5. The UI Mode also showed good ratings, but two participants scored out of the majority. They explained that sometimes they forgot to cancel the follow function, causing the robot to drive to undesired locations. In this regard, Lab Mode performed the worst, with an average score of 3. And three participants gave a low score of 2 to Lab Mode.
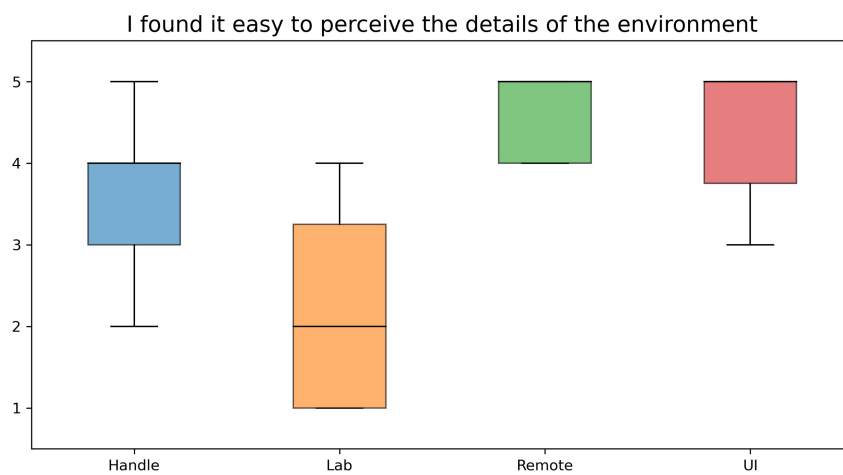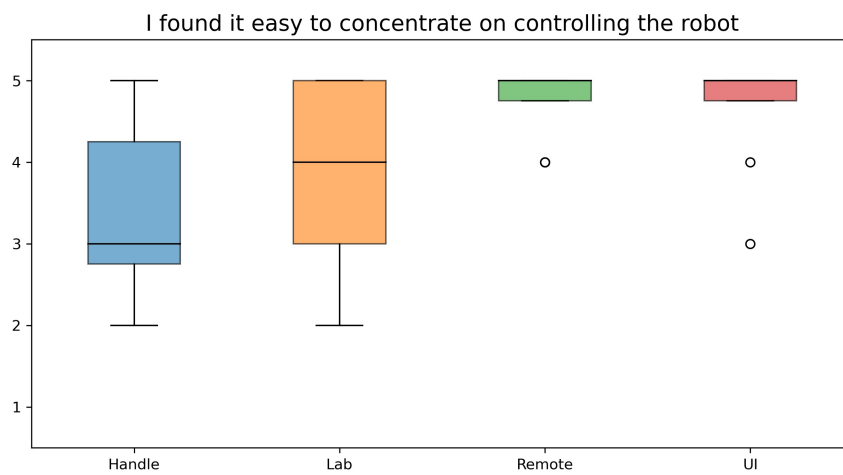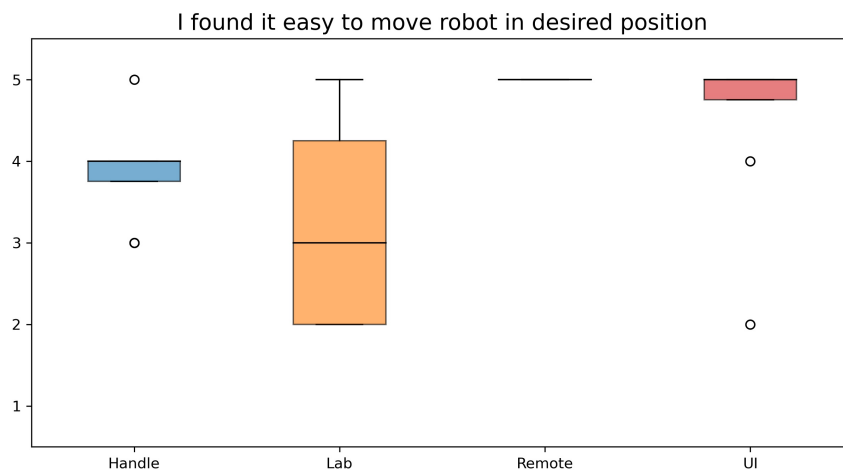
Figure 5.4: Likert Questionnaire Results.(1:strongly disagree, 5:strongly agree)

"I found it easy to concentrate on controlling the robot."

In Handle Mode, participants seemed to have more difficulty concentrating on controlling the robot. This could be partly attributed to the fact that the participants had to adjust their distance from the robot when it was about to leave their field of view. On the other hand, participants had to control the direction and movement of the cart, which was not as easy in the Handle Mode as in the Remote and UI modes. Therefore, although participants needed to adjust their position while controlling the robot in all three modes, the other two modes scored better than Handle Mode. Some participants thought that they could concentrate well in Lab Mode because they did not need to turn their heads and adjust their positions frequently as in the other 3 modes. However, some participants held the opposite opinion. They stated that they needed to recall how the lab operated, and that multiple screens would also cause distractions.

"I found it easy to perceive the details of the environment."

The Lab Mode had the worst rating. This could partially be explained by the use of a screen to show the scene, rather than through immersion. Another reason was the poor display quality of the screen, which some participants felt was very blurred, making it impossible for them to observe the details of the scene. The results of Handle Mode differed greatly. Five participants gave a rating higher than 4. However, three participants gave a score of 2 or 3. The reasons were similar to what was mentioned in last question. The difficulty of operation and the need to alternate between controlling the robot and their own position made it difficult for them to focus on the scene.

5.3 Qualitative Results

This section will discuss the feedback from participants. Overall, every participant gave positive comments about operating the robot in a VR platform. They thought the proposed system was exciting and did allow them to perceive more details in the post-disaster environment than the traditional video-based manipulation. The feedbacks obtained from each mode will be listed next.

70% of participants ranked Lab Mode as the least preferred mode. Some participants were very unaccustomed to using VR handles to grasp objects, which made it difficult for them to operate the robot with virtual joysticks smoothly. For those who have VR experience, even without any hints and learning, they subconsciously understood what each button and joystick represented and were able to operate the robot directly. Nevertheless, for the actual rescue experience in the test focus, both kinds of participants responded that the robot's operation was more complex and difficult than the other modes. Participants attributed the reasons to obstacles in the environment. One of the participants said: "*There is no physical access to the joystick. So it is slightly tough for me to control the robot.*" In some cases, when the robot was stuck in a corner, it took them much effort to get the robot out of this situation. Also, since the Lab Mode uses a simulated screen, the Lab Mode is not as good

as the other three in terms of observing the details of the scene. Participants felt that the simulated screen was blurred, and the frequent switching between multiple screens made them very tired.

Handle Mode directly uses motion controllers for moving the robot. The evaluation of this operation mode depends in large part on the design of the motion controllers. The proposed system uses the touchpad on the HTC VIVE motion controllers to control the direction of the robot. More than half of the users thought that the touchpad made them less flexible when operating the robot's steering. Participants were often unable to accurately touch the correct position of the touchpad when using it, and it was very likely to be touched by mistake. At the end of the experiment, these participants were additionally invited to re-operate the robot using the VR controller with thumb sticks, and said that using thumb sticks was easier for them to control the direction. Some participants said that they did not like the two monitoring screens provided by this mode. The additional surveillance screens made them subconsciously distracted to observe them, preventing them from concentrating on the rescue mission. Others, however, thought that the monitor was particularly helpful. As it was very difficult to control the robot while teleporting themselves, they first relied on the monitor screen to drive the robot to a place, and then teleported themselves to the location of the robot. The experiment also found that participants tended to forget that the two monitor screens could be closed, and they usually tried to drag the screens to places where they did not affect their view and dragged them back when they wanted to use them.

Remote Mode and UI Mode that use intelligent obstacle avoidance driving algorithm were most well-received. Participants felt that in both modes they did not need to worry about how to control the robot's steering and forward speed, but that the computer was responsible for everything, allowing them to focus on virtual world exploration.

For the UI Mode, one of the participants remarked: "*I can just let the robot follow me. I don't need to think about how to operate the robot. This way I can concentrate on the rescue.*" In the user study, all participants could easily learn how to operate the UI menu. This may be explained by the fact that the menu interface was very familiar to them. It was observed that all participants did not use the direction buttons and monitoring screens in the virtual menu. At the beginning of the test, they all turned on the follow function directly and adjusted the robot's driving speed to the maximum. After that, the robot was more like a moveable LiDAR sensor. This therefore led to the fact that these participants could completely disregard the location of the robot and just explore the VR world on their own. One participant in the experiment teleported so fast that when he reached a location and had been waiting for a while, the robot was still on its way. In fact, the problem of not being able to find the robot happened in Handle Mode as well.

In contrast, Remote Mode solves this problem of the robot not being in view. One participant stated that "*The robot is always in sight, so I don't have to waste extra time looking for the robot. Control of the robot is also very easy.*" Another participant reflected that after setting the destination of the trolley operation, he would subconsciously observe the movement of the robots, thus making him always know where the robot was. They also thought it was very easy in this mode to operate the robot. Many participants alternated between using the right- and left-hand rays, first setting the robot's moving target point with the right-hand ray, and then teleporting themselves there with the left-hand ray. The backup measures (remote controller) were almost not used in the actual test.

When it came to the robot's inability to navigate automatically to the destination, the participants preferred to move the robot by resetting the destination point.

In addition to this, participants were found lost in each of the operation modes. They could forget whether the place was already visited by themselves. Similar behavior was observed in all participants. They passed by the same place over and over again, and sometimes simply stayed within the confines of a known scene and did not explore further.

6 Discussion

In general, an ideal VR-based robotics operation method should eliminate as much complexity as possible for the user.

For the Lab Mode, the least favorite mode among the participants, it can be concluded that unless the VR operating system is developed for training operators to learn to operate the robot in a real environment, a lab-like operation mode is not desirable. Suppose one wants to develop an interaction approach like Handle Mode, where the robot is operated directly using the controller. In that case, it should be taken into account whether the user needs to move their position frequently. As mentioned before, if the user needs to control the exact movement of the robot themselves and at the same time change their position, the user may have a higher workload, more difficulty in observing the details of the scene and concentrating on controlling the robot. The choice of VR handle is also important. VR Motion controllers with thumb sticks are recommended for better directional control.

Interaction approaches like Remote Mode and UI Mode should be the future direction to focus on. Both of these operation modes significantly simplify the process of controlling the robot by using intelligent obstacle avoidance algorithms to allow the robot to automatically navigate to the desired destination. The proposed system currently uses the NavMeshAgent component and simulates a post-disaster scene instead of reconstructing it by scanning the real site through LiDAR. Therefore, there remains a need for an intelligent obstacle avoidance algorithm when using a real robot. Considering that the rescue efficiency and the possibility of the robot being damaged rely strongly on the algorithm, this algorithm should be accurate enough so that the user can entirely depend on the computer to control the robot.

In some tests, the monitoring screens were not used. The additional features, which are provided by Remote and UI modes that allow the user to control the robot themselves, were also rated as redundant by participants. All the items mentioned above might somewhat complicate the interaction process. When too many features are available simultaneously, it seems difficult for the user to decide which feature to use in certain situations. Sometimes users may even forget functions that are already provided, such as monitoring screens can be switched off if they block the view. It should be noted that there were only eight participants in this user study and that all of them were unfamiliar with VR. It is thus uncertain whether different results would be obtained if the number of participants is expanded. Even so, the presence of the monitor screen is still necessary, considering that some participants responded it could provide valuable information. Further work is needed to optimize the monitoring screens so that they do not obscure the view, and adjusting the monitor does not complicate the whole interaction.

Apart from that, extra maps should be provided. It was mentioned in the results part that participants often got lost in the scenes. They did not know if a location had been visited before and repeatedly searched for the same location. This may lead to a decrease in search efficiency and even, to a certain extent, discourage users from exploring unknown areas that have not yet been scanned by the LiDAR. The map should hence indicate the fields that have been scanned by the LiDAR and the user's own current location. The user can see the overall outline of the detected area and the trajectory he or she has taken from the map provided, thus giving the user a clear overview of the scene. In the user study, it was found that some participants often forget the location of the robot. Therefore, the relative coordinates of the user and the robot should also be provided on the map. How the map is presented is also worth considering. As noted before, the whole interaction pattern should be as simple as possible.

7 Conclusion

This thesis presents a preliminary VR-based Human-Robot Interaction system in terms of ground based rescue robots. This work aims to find an ideal interaction method or provide a general direction for future development. For this purpose, the proposed system offers four different operation modes and corresponding test scenes imitating a post-disaster city. This thesis shows an overview of the simulated robot, interaction techniques and the construction of the test environment. Eight participants were invited to conduct a user study. Based on the obtained results, it can be concluded that an ideal VR-based robotics operation method should eliminate as much complexity as possible. An intelligent obstacle avoidance algorithm is recommended instead of the user operating the robot themselves to steer and move forward. Additional functions, such as monitoring screens, need to be optimized so that they do not complicate the whole interaction process. The system also requires maps to show the user which areas have been detected and where the robot is located.

Future work should focus on the intelligent obstacle avoidance algorithm when using a real robot. The next stage is to develop a live telepresence and teleoperation system with the real robot. Considering that the system proposed in this thesis only simulates the disaster rescue process, the conclusions obtained may be limited. Additional testing and user surveys should be carried out in the future after building a collaborative VR-based system with real robots.

List of Figures

Glossary

Augmented Reality Using computer technology to superimpose virtual information onto the real world and display it through mobile phones, tablets and other devices, thus achieving a great fusion of the real and the virtual.

HTC VIVE A virtual reality headset developed by HTC and Valve. It has a resolution of 1080×1200 per eye, resulting in a total resolution of 2160×1200 pixels, a refresh rate of 90 Hz, and a field of view of 110 degrees. It includes two motion controllers and uses two Lighthouses to help track the headset position and the motion controllers.

Human-Robot Interaction A study of the interaction between humans and robots. It is a multi-disciplinary field encompassing human–computer interaction, artificial intelligence, robotics, natural language understanding, design, and psychology.

Human–Robot Collaboration A study of collaborative processes in human and robot agents work together. It includes classical robotics, human-computer interaction, artificial intelligence, design, cognitive sciences and psychology.

Light Detection and Ranging A laser-based remote sensing technology. It uses the laser as the light source and employs photoelectric detection technology. It can produce a map of the surveyed area known as a 3D point cloud.

Mixed Reality Mixed Reality (MR) enhances the realism of the user experience by introducing realistic scene information into the virtual environment, bridging the gap between the virtual world, the real world, and the user with interactive feedback information.

NavMeshAgent Navigation mesh agent provided by Unity. This component can be attached to game objects to intelligently navigate to a destination and avoid obstacles. Terrain and objects within the scene must be set to static and marked with a moveable range before being used.

Robot Operating System A set of software libraries and tools for robotics. Robot Operating System (ROS) provides various services for operating system applications (e.g. hardware abstraction, underlying device control, common function implementation, inter-process messaging, package management, etc.), as well as tools and functions for acquiring, compiling, and running code across platforms. ROS mainly uses loosely coupled peer-to-peer process network communication and currently mainly supports Linux systems.

Steam VR plugin A Unity plugin from Valve Corporation. Applications developed using the SteamVR plugin can be adapted to mainstream PC VR software. The plugin simplifies these parts for developers: loading 3D models for VR controllers, handling input from these controllers, and estimating hand posture when using these controllers.

The Center for Robot-Assisted Search and Rescue A nonprofit corporation dedicated to unmanned systems technology for public safety. It aims to promote the effective use of unmanned systems by formal emergency management agencies.

The Official NASA Task Load Index The NASA Task Load Questionnaire (NASA-TLX, NASA Task Load Index) is a subjective workload assessment tool whose primary purpose is to provide a subjective workload assessment of operators of various user interface systems. Using a multidimensional rating process, the total workload is divided into six subjective subscales. The user should choose the weighted score for each measurement. The simplified version, also called Raw TLX, has the same weighted score for all subscales and some of the subscales could be excluded. This simplified version is also used in this thesis.

Unity Unity is a cross-platform game engine developed by Unity Technologies, allowing users to easily create interactive content such as 3D video games, architectural visualizations, real-time 3D animation and other types of interactive content.

Virtual Reality Featuring immersion, interactivity and conceptualization. By simulating the functions of human sensory organs such as sight, hearing and touch, it allows people to immerse themselves in a computer-generated virtual world and to communicate in real time through language and gestures, enhancing the sense of entry and immersion.

Acronyms

AR Augmented Reality

CRASAR The Center for Robot-Assisted Search and Rescue

LiDAR Light Detection and Ranging

MR Mixed Reality

ROS Robot Operating System

TLX The Official NASA Task Load Index

VR Virtual Reality

References

[1] J. Casper and R. R. Murphy. "Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33.3 (2003), pp. 367–385. DOI: `10.1109/TSMCB.2003.811794`.

[2] R. R. Murphy. "A decade of rescue robots". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2012, pp. 5448–5449. ISBN: 2153-0866. DOI: `10.1109/IROS.2012.6386301`.

[3] Ya-Wen Huang et al. "Operation of underwater rescue robot anchor diver III during the 2011 Tohoku Earthquake and Tsunami". In: *OCEANS'11 MTS/IEEE KONA*. 2011, pp. 1–6. ISBN: 0197-7385. DOI: `10.23919/OCEANS.2011.6107198`.

[4] R. R. Murphy. "Human-robot interaction in rescue robotics". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2 (2004), pp. 138–153. DOI: `10.1109/TSMCC.2004.826267`.

[5] Guy Hoffman and Cynthia Breazeal. "Effects of Anticipatory Action on Human-Robot Teamwork Efficiency, Fluency, and Perception of Team". In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. HRI '07. Arlington, Virginia, USA: Association for Computing Machinery, 2007, pp. 1–8. ISBN: 9781595936172. DOI: `10.1145/1228716.1228718`. URL: `https://doi.org/10.1145/1228716.1228718`.

[6] S. Sarkar et al. "Earthquake rescue robot: A purview to life". In: *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. 2017, pp. 1–7. DOI: `10.1109/ICECCT.2017.8118044`.

[7] F. Faisal and S. A. Hossain. "DOORMOR: A Functional Prototype of a Manual Search and Rescue Robot". In: *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. 2019, pp. 1–6. DOI: `10.1109/ICASERT.2019.8934515`.

[8] Patrick Sousa et al. *Human-Robot Interaction Based on Gestures for Service Robots*. Vol. 27. Oct. 2017, pp. 700–709. ISBN: 978-3-319-68194-8. DOI: `10.1007/978-3-319-68195-5{\_}76`.

[9] Alessandro Giusti et al. "Human-Swarm Interaction through Distributed Cooperative Gesture Recognition". In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '12. Boston, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 401–402. ISBN: 9781450310635. DOI: `10.1145/2157689.2157818`. URL: `https://doi.org/10.1145/2157689.2157818`.

[10] J. Nagi et al. "Human-swarm interaction using spatial gestures". In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014, pp. 3834–3841. ISBN: 2153-0866. DOI: `10.1109/IROS.2014.6943101`.

[11] S. Pourmehr et al. ""You two! Take off!": Creating, modifying and commanding groups of robots using face engagement and indirect speech in voice commands". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013, pp. 137–142. ISBN: 2153-0866. DOI: `10.1109/IROS.2013.6696344`.

[12] J. Ma et al. "A Novel EOG/EEG Hybrid Human–Machine Interface Adopting Eye Movements and ERPs: Application to Robot Control". In: *IEEE Transactions on Biomedical Engineering* 62.3 (2015), pp. 876–889. DOI: `10.1109/TBME.2014.2369483`.

[13] João Soares, Alberto Vale, and Rodrigo Ventura. "A Multi-purpose Rescue Vehicle and a human–robot interface architecture for remote assistance in ITER". In: *Fusion Engineering and Design* 98-99 (2015). Proceedings of the 28th Symposium On Fusion Technology (SOFT-28), pp. 1656–1659. ISSN: 0920-3796. DOI: `https://doi.org/10.1016/j.fusengdes.2015.06.148`. URL: `https://www.sciencedirect.com/science/article/pii/S092037961530199X`.

[14] V. Villani, B. Capelli, and L. Sabattini. "Use of Virtual Reality for the Evaluation of Human-Robot Interaction Systems in Complex Scenarios". In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2018, pp. 422–427. ISBN: 1944-9437. DOI: `10.1109/ROMAN.2018.8525738`.

[15] O. Liu et al. "Understanding human-robot interaction in virtual reality". In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2017, pp. 751–757. ISBN: 1944-9437. DOI: `10.1109/ROMAN.2017.8172387`.

[16] Patrick Stotko et al. *A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot*. Nov. 2019. DOI: `10.1109/IROS40897.2019.8968598`.

[17] P. Wang et al. "A novel human-robot interaction system based on 3D mapping and virtual reality". In: *2017 Chinese Automation Congress (CAC)*. 2017, pp. 5888–5894. DOI: `10.1109/CAC.2017.8243836`.

[18] Elias Matsas and George-Christopher Vosniakos. "Design of a virtual reality training system for human–robot collaboration in manufacturing tasks". In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11.2 (2017), pp. 139–153. DOI: `10.1007/s12008-015-0259-2`. URL: `https://doi.org/10.1007/s12008-015-0259-2`.

[19] Luis Pérez et al. "Industrial robot control and operator training using virtual reality interfaces". In: *Computers in Industry* 109 (2019), pp. 114–120. DOI: `https://doi.org/10.1016/j.compind.2019.05.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0166361518308546`.

[20] M. M. Moniri et al. "Human Gaze and Focus-of-Attention in Dual Reality Human-Robot Collaboration". In: *2016 12th International Conference on Intelligent Environments (IE)*. 2016, pp. 238–241. ISBN: 2472-7571. DOI: `10.1109/IE.2016.54`.

[21]   M. Ostanin et al. "Human-robot interaction for robotic manipulator programming in Mixed Reality". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. 2020, pp. 2805–2811. ISBN: 2577-087X. DOI: 10.1109/ICRA40945.2020.9196965.

[22]   D. Whitney et al. "ROS Reality: A Virtual Reality Framework Using Consumer-Grade Hardware for ROS-Enabled Robots". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2018, pp. 1–9. ISBN: 2153-0866. DOI: 10.1109/IROS.2018.8593513.