Teleoperation of a Fast Omnidirectional Unmanned Ground Vehicle in the Cyber-Physical World via a VR Interface

Yiming Luo
yiming.luo19@xjtlu.edu.cn
Xi'an Jiaotong-Liverpool University
Suzhou, China

Jialin Wang
jialin.wang16@xjtlu.edu.cn
Xi'an Jiaotong-Liverpool University
Suzhou, China

Yushan Pan
yushan.pan@xjtlu.edu.cn
Xi'an Jiaotong-Liverpool University
Suzhou, China

Shan Luo
shan.luo@kcl.ac.uk
King's College London
London, United Kingdom

Pourang Irani
pourang.irani@umanitoba.ca
The University of British Columbia
Kelowna, Canada

Hai-Ning Liang*
haining.liang@xjtlu.edu.cn
Xi'an Jiaotong-Liverpool University
Suzhou, China

ABSTRACT

This paper addresses the relations between the artifacts, tools, and technologies that we make to fulfill user-centered teleoperations in the cyber-physical environment. We explored the use of a virtual reality (VR) interface based on customized concepts of Worlds-in-Miniature (WiM) to teleoperate unmanned ground vehicles (UGVs). Our designed system supports teleoperators in their interaction with and control of a miniature UGV directly on the miniature map. Both moving and rotating can be done via body motions. Our results showed that the miniature maps and UGV represent a promising framework for VR interfaces.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; **User interface design**; *User studies.*

KEYWORDS

Human-robot Interaction, Virtual Reality, Teleoperation, World-in-Miniature, Interface Design

ACM Reference Format:
Yiming Luo, Jialin Wang, Yushan Pan, Shan Luo, Pourang Irani, and Hai-Ning Liang. 2022. Teleoperation of a Fast Omnidirectional Unmanned Ground Vehicle in the Cyber-Physical World via a VR Interface. In *The 18th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (VRCAI '22), December 27–29, 2022, Guangzhou, China.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3574131.3574432

1 INTRODUCTION

Worlds-in-Miniature (WiM) is a technique used as a tool for navigation and object manipulation in virtual reality [Milgram and Kishino 1994]. It is a scaled-down replica of the original environment combining the advantages of an operation space, a car-to-graphic map, and an interface that allows users to observe overview+detail quickly [Danyluk et al. 2021a]. These affordances of WiM can be of benefit for remotely controlling drones via VR. However, the use of WiM in teleoperator–drone manipulation is under-explored. Our review of the literature, particularly from human-robot interaction (HRI), shows that current approaches for teleoperator–drone remote control typically rely on a computer monitor to display information and a keyboard, mouse, or joystick to control the drone [Wonsick and Padır 2021]. In addition, VR is generally used for visualization and to enable operators to interact with 3D environments derived from the 3D physical world. Such applied cognitive engineering thinking to designing HRI interfaces is nothing new [Gorjup et al. 2019], because researchers, to some degree, have already agreed that this type of technology can enhance operators' perception and presence in the virtual environment. Nevertheless, and oddly enough, VR technology has not yet been widely adopted in mainstream HRIs, especially in commercial products.

Usually, in HRI, when the teleoperator controls a vehicle with a bird's-eye view (e.g., derived from the real-time feed from an aerial drone performing photography), the conventional way of manipulating a UGV is with two joysticks to perform rotations and translations (typically, one to control body movements and the other for steering (see details in Figure 1)). However, since the camera's perspective does not move with the direction of the UGV's movements, the teleoperator's control of the UGV could be challenging because it does not reflect a natural mapping [Walker et al. 2019]. When the UGV has the Mecanum wheel [Diegel et al. 2002], the control can be more counter-intuitive and difficult to grasp, especially for non-expert users [Grassini et al. 2020]. In line with Nostad et al. [Nostadt et al. 2020] and Draper et al. [Draper et al. 1998], we consider the teleoperation system has to be designed with full respect to the natural interaction between teleoperators and the system. That means the system has to be designed to fit its teleoperators, rather than the teleoperators [Pan 2021] having to adapt to use it in their daily work practices [Falcone et al. 2022]. Thus, our research question is *how a humanized interactive interface via VR technology can be designed to support teleoperators' in-situ work practice in their tasks.*

To demonstrate that our approach is practical, efficient, and user-friendly, we conducted a user study and compared four conditions

*Corresponding Author

VRCAI '22, December 27–29, 2022, Guangzhou, China

ACM ISBN 979-8-4007-0031-6/22/12…$15.00
https://doi.org/10.1145/3574131.3574432

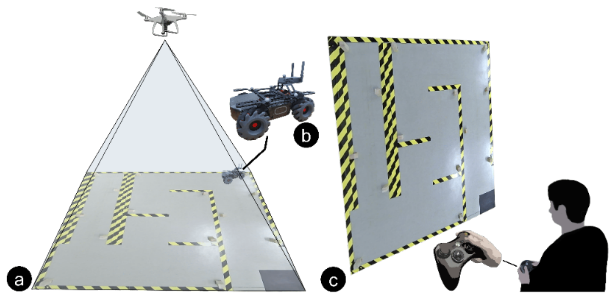

Figure 1: (a) An aerial camera shoots from above to provide a drone's view; (b) A UGV with Mecanum wheels with five tracking spots; and (c) A teleoperator is viewing the image of (a) on a 2D screen and teleoperate the UGV using an Xbox controller.

consisting of two factors, namely (1) map visibility and (2) control methods, to evaluate our method. Our findings show that our added approach to the WiM technique significantly improves the teleoperators' performance, reduces their workload, and enhances their preference for teleoperation tasks. Hence, we offer two potential contributions to the VR community. First, we provide a VR interface based on WiM to aid in the teleoperation of UGVs. Second, our proposed approach opens a door for other researchers who are interested in designing VR interfaces for UGV teleoperation in similar research contexts.

2 RELATED WORK

2.1 Teleoperation via VR Interfaces

With the development of VR applications, more and more researchers want to improve the perception (vision, haptic, and other sensory feedback) of the environment in VR [Luo et al. 2022; Wang et al. 2022]. Some have attempted to provide intuitive interaction approaches to facilitate intuitive real-time remote teleoperation [Naceri et al. 2021]. Ott et al. showed an evaluation study of teleoperation [Ott et al. 2005], Kadavasal and Oliver proposed a multi-model teleoperation approach [Kadavasal and Oliver 2009], and Tran et al. discussed the possibility of using Wizard-of-Oz methods for hands-free control of robots in VR [Tran et al. 2018]. Kazanzides et al. presented remote intervention in space [Kazanzides et al. 2021], and Domingues et al. showed us how to control underwater robots [Domingues et al. 2012]. Li et al. [Li et al. 2022] have focused on how people can collaboratively work in VR environments. They proposed a new collaborative VR system to support two teleoperators working in the same VR environment to control a UGV remotely. In addition, research has shown that VR with an omnidirectional treadmill can be utilized to create a fully immersive teleoperation interface for controlling a humanoid robot. This system is suitable for precise but slow teleoperation [Elobaid et al. 2019]. Besides the slow processing, the cost of equipment and high physical demands for users also represent significant impediments to the system. The control system [Hirschmanner et al. 2019] allows a humanoid robot to be remotely controlled by imitating the user's upper-body posture in a design that mimics the entire human arm pose during teleoperation.

Although there are many commercial immersive VR devices applied to the field of robotics [Hetrick et al. 2020], few research focused on exploring VR interfaces for user-centered remote robot manipulation [Chen et al. 2017]. Some researchers [Luo et al. 2021] in recent years have used a traditional computer interface with an immersive VR one for teleoperation. This work has shown that the VR interface can improve user experience in teleoperation. Another user study [Theofanidis et al. 2017] has compared a VR programming interface with a direct manipulation interface and keyboard, mouse, and monitor manipulation interfaces. They used gesture recognition, rather than VR controllers, to teleoperate a robot. Their results showed that their system could support the training of robot programming. A common characteristic is that work remains the dominant control paradigm for human interaction with robotic systems. Even though it may have merits in various domains, teleoperation can be challenging for novice users in complex environments. Without having a nuanced understanding of users' skills for problem-solving and their goals for task completion [Pan et al. 2021], we will fall into low-level aspects of robot control. In line with that, it is impossible to accurately increase operation effectiveness, support concurrent work, and decrease work stress [Bourdieu 2020]. Moreover, those prior studies focused mainly on studying the mapping of users' gestures to commands for remote robots in first-person view (FPV). In contrast, the third-person view (TPV), such as a bird's-eye view, has not been explored in detail.

2.2 Worlds-in-miniature (WiM) in VR

WiM is a metaphor for a user interface technique that augments an immersive head-tracked display with a hand-held miniature copy of the virtual environment [Stoakley et al. 1995]. WiM offers a second dynamic viewport onto the virtual environment as an addition to the first-person perspective in the VR system. By doing so, objects are directly manipulated either through an immersive viewport or through the three-dimensional viewport offered by the WiM [Stoakley et al. 1995]. WiM interfaces have been used in VR [Bluff and Johnston 2019; Drogemuller et al. 2020] in recent years. WiM was first proposed for virtual environments and meant to provide a small, scaled-down model of an entirely virtual environment that could act as a map and interaction space for users to explore large environments [Stoakley et al. 1995]. Since then, researchers have successively made new additions and improvements. For example, Wingrave et al. [Wingrave et al. 2006] introduced the concept of a scaled scrolling WiM that addressed the issue of users being unable to perform tasks of different scale levels using the original idea of WiM. Trueba et al. [Trueba et al. 2009] utilized an algorithm to automatically analyze a model's 3D structure and select the best WiM views to minimize occlusion issues. Danyluk et al. [Danyluk et al. 2021b] proposed eight dimensions (size, scope, abstraction, geometry, reference frame, links, multiples, and virtuality) to define the design of WiM. Although the previous work lacks user-centered focus, we still get inspiration from WiM to explore the design of VR interfaces that enhance robot teleoperation (especially for land-based robots). By doing so, we enable the user to manipulate immersive virtual robot surrogates that foreshadow the physical robots' actions. Thus, in the present work, we address more users' perception that teleoperation must see through their own eyes.

3 SYSTEM OVERVIEW

3.1 Field Study

Suppose the system can enable the teleoperator to operate with a small range of hand movements in front of their body. In that case, they can perform natural movements similar to actions that operate the steering wheel of a car. Therefore, we are committed to combining hand movements and vision so that operators can grab a virtual surrogate of UGV to complete the teleoperation of an actual UGV in an interface that conforms to traditional driving practices.

3.1.1 Interface Considerations. Considering that the interface is in a flat-like form, we took inspiration from the car's steering wheel (see Figure 2a) and proposed that the angled interface would be more user-friendly. The size of the operation interface also needs to be considered, which is related to the scaling ratio between the size of the field under the aerial view of the drone and the size of the miniature map in the virtual world. To explore the above considerations, we conducted a pilot study.

3.1.2 Pilot Study. We had eight participants in our pilot study, which provided us with some constructive results. The participants were required to adjust the placement of the 2D screen and the miniature map in VR and fill out a questionnaire to collect their preferences on the angle of placement (0, 45, and 90 degrees) and the ratio of the miniature map (1:5, 1:10, and 1:15) to the actual experimental site (2m×2m) (see Figure 2b and c). The questionnaire results showed that 7/8 of the participants preferred 45 degrees because it was comfortable and suitable during the teleoperation of the UGV. They pointed out that lower than 45 degrees or higher than 45 degrees would cause neck discomfort, visual discomfort, and operational challenges. The ratio of our miniature map to the actual site was set to 1:10, limiting the range that the user's hands could move to a 20cm×20cm area. This was found to help avoid fatigue caused by extensive large hand movements.

3.2 System Design

To better support the performance of teleoperators-UGV interaction, our system was not only set in an experimental but also partly actual field. On the site, wooden blocks were placed as targets, and black and yellow warning tapes were used as barriers to guide the movement of the UGV (see Figure 2d and e). The teleoperator would monitor the movement of the UGV and the environment from images in the VR captured from the camera to control the UGV under different experimental conditions to hit the targets but avoid touching the tapes.

3.2.1 Hardware Overview. [Teleoperator side] An HTC Vive Pro set driven by a Windows 10 desktop computer (Intel Core i9- 11900K at 3.5 GHz, 32 GB RAM, and NVIDIA GeForce GTX 3090), one Xbox controller, and one USB camera attached to the ceiling to simulate the drone's view. [UGV side] A RoboMaster EP without the robotic arm, five tracking spots attached on the UGV for the VICON tracking system. [Networking] The two sides are connected by an Orbi router (RBK853 AX6000 WiFi 6 System). The USB cameras and VICON system are connected to the desktop computer by wire.

3.2.2 Software Overview. [Development Tools] The Unity3D game engine (version 2019.3.7f3) with SteamVR for Unity. [Virtual Representation] The surrogate of UGV in our interface was built by a red hemisphere representing the direction of the front of the UGV, and a white cuboid representing the body. We used virtual yellow lines to represent the tapes as warning lines and virtual blocks to represent the actual blocks as targets, which are scaled down by a specific ratio (1:10).

3.2.3 System Workflow. In this section, we demonstrate system workflow see Figure 3. The USB camera captures the pictures of the surrounding environment of the UGV from top to bottom and transmits them to the PC through wired transmission, providing the teleoperator with 2D real-time images in the VR world. The VICON system obtains the position and attitude information of the UGV in real-time, synchronizes it to the surrogate in the miniature map in the virtual world, and provides the teleoperator with a real-time 3D virtual WiM through the Unity3D environment. The teleoperator can grab the virtual surrogate through the VR handle and manipulate it. The actions are synchronized to the movement of the actual UGV. The teleoperator can also directly control the UGV with the joysticks through the Xbox controller.

To simulate the natural environment as a virtual one, there was a need to capture the spatial features of the former and the movement of the UGV in real-time. For example, the UGV could scan its surrounding 3D environment using its onboard cameras or sensors and send the data to the remote site. In our design, we decided to utilize the VICON system and Unity3D to simulate the natural scanned environment. Our simulation was able to reconstruct the environment with high precision to focus on the research on the control performance rather than the acquisition of 3D environment information. The algorithm is designed to support them to share the same bodily awareness, embodied action, and social activity. The interactive interface is the setup for mediating the interaction between teleoperators and the physical UGV. That means the virtual surrogates work in real-time to cope with the teleoperators' actions. The planning algorithm running on the physical UGV can constantly "chase" the surrogate, configuring positions in both the virtual and physical worlds. In order to accurately map the teleoperator's WiM Control to the UGV in the physical world, we also use the Proportional Integral Derivative (PID) control method to realize the synchronization of the user's hand movements to those of the robot. In our case, we can reduce the body movement error to less than 0.5cm and the rotation angle error to less than 0.5 degrees when controlling the UGV. Thus, we proposed the factor of Control Methods (Joystick control vs. WiM control) (see details in subsection 4.2).

The premise of our system design is to provide a miniature map from real-time environmental scans on which to interact with the miniature UGV on the map. However, we considered that in the absence of a map, for example, caused by significant errors or failures in the scanning system, the miniature UGV can still be interacted with and controlled remotely. We, therefore, proposed the factor of Map Visibility (Visible vs. Invisible) (see details in subsection 4.2). This activity enables our WiM to relate to the

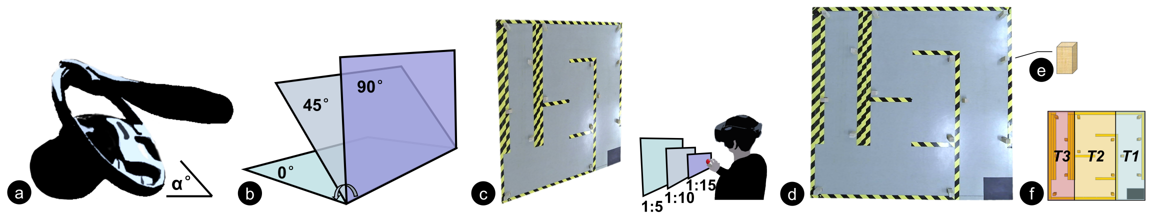

Figure 2: (a) An example of a steering wheel of the car placed at an angle; (b) the interface placed at different angles (0, 45, and 90 degrees); (c) the interface of different sizes due to different scaling ratio (1:5, 1:10, and 1:15); (d) a picture of the bird's view of our experimental site; (e) the target is a 5cm×5cm×10cm wood block; (f) three local tasks (*T1 - T3*).

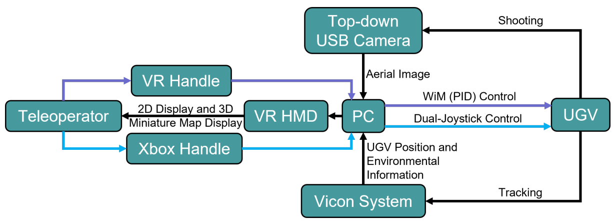

Figure 3: Flow Chart of Our System.

teleoperators' presence and experiences towards the worlds we perceive together through the immediate environment (physical world) and the world around us (the virtual world, including a third-person bird's eye view).

4 USER STUDY

The user study aimed to explore how control methods (joystick vs. gestures) with different WiM map visibility (visible vs. hidden) would affect the teleoperation of UGV in a bird's-eye view. The interface would display the real-time images to the user on a 2D screen in VR. The user would use this screen to monitor and operate the UGV in the four conditions.

4.1 Tasks

A maze was built for the UGV to complete a comprehensive task where the teleoperator must hit 14 wooden target blocks as quickly as possible, but without hitting or going over the warning lines made up of yellow and black tapes. It consisted of 3 (*T1-T3*) different local tasks (see Figure 2f): (1) *T1*. Four targets with 32cm minimum path width; (2) *T2*. Five targets with 28cm minimum path width; (3) *T3*. Five targets with a 24cm minimum path width. From *T1* to *T3*, the width of the passage was gradually reduced, and more movement (translation and rotation movement) was required.

4.2 Conditions

In our pilot study, we found that the placement angle of the 2D screen and the miniature map would impact the teleoperator's posture and comfort level, which has determined the most appropriate parameters (e.g., interface placed at an angle of 45 degrees, the interface size of 1:10 and the actual map size ratio; for more details see subsubsection 3.1.2). After employing the above parameters, we have the following four conditions (see Figure 4) derived from two independent variables (Control Methods: *Joystick Control* vs. *WiM Control* and Map Visibility: *Invisible Map* vs. *Visible Map*):

- *Joystick Control* with *Invisible Map*. Users monitored the 2D screen and used an Xbox Controller to control the body movement and rotation of the UGV;
- *Joystick Control* with *Visible Map*. Users monitored the 2D screen and checked the miniature map at the same time. They controlled the UGV using an Xbox controller;
- *WiM Control* with *Visible Map*. Users monitored the 2D screen and checked the miniature map at the same time. They used a Vive controller to interact with the miniature UGV in a miniature map to control the real UGV;
- *WiM Control* with *Invisible Map*. Users used a Vive controller to interact with the miniature UGV but the map was invisible. The approach to interacting with UGV was the same.

4.3 Evaluation Metrics

4.3.1 Performance Measures. To evaluate the performance and usability of the WiM technique, we measured the time of collisions on the black and yellow warning tapes and completion time to finish each of the three tasks (*T1 - T3*) from the data capture via the VICON system and within Unity3D, the platform used to develop the testing application and run it. (1) *Collisions Time*. The Unity3D program would automatically record the time when the UGV hit the black and yellow warning tapes for each trial in each task. (2) *Completion Time*. We measured the completion time for each trial in each task.

4.3.2 Subjective Measures. (1) *NASA-TLX Workload Questionnaire* [Hart 2006]. The NASA-TLX was used to the measure workload demands of each task. This questionnaire contained questions with 11-point scales (from 0 to 10), which assessed six elements of users' workload (Mental, Physical, Temporal, Performance, Effort, and Frustration). (2) *User Experience Questionnaire (UEQ)* [Laugwitz et al. 2008]. UEQ was used to measure the preference level for each condition. This questionnaire contained questions with 7-point scales (from -3 to 3), which assessed eight elements of users' experience.

4.4 Procedure

The participants were required to drive two rounds for each condition in our within-subjects study. The order of conditions is counterbalanced using a Latin Square design to mitigate carry-over effects. Before starting the actual trials, there were training sessions for

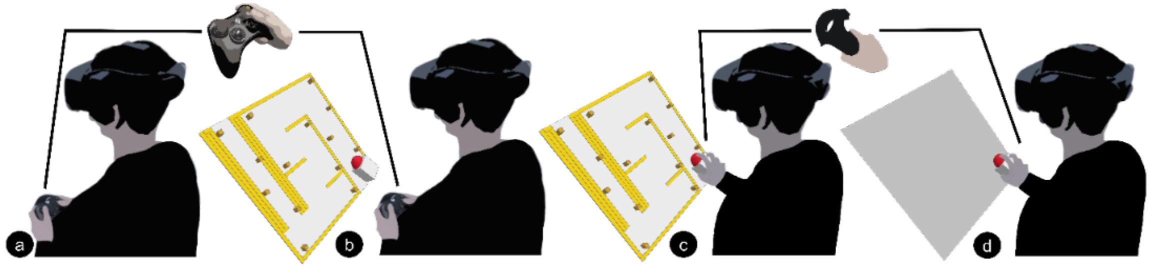

Figure 4: (a) *Joystick Control with Invisible Map*; (b) *Joystick Control with Visible Map*; (c) *WiM Control with Visible Map* and (d) *WiM Control with Invisible Map*. *The invisible map is highlighted for the readability of readers.

participants to let them become familiar with the VR device, UGV, and controls. Before starting, they needed to fill in a questionnaire to collect demographic data and past VR and UGV teleoperation experience. Participants were required to fill a NASA-TLX workload and UEQ after each condition.

4.5 Participants

Sixteen participants (8 males and 8 females, aged between 19-30, mean = 23) from a university campus were recruited for this experiment. They all declared to be healthy and had no health issues, physical and otherwise. They had normal or corrected-to-normal vision and did not suffer from any known motion sickness issues in their normal daily activities. None of them had any experience driving a UGV using an HMD in TPV (short for third-person view).

4.6 Hypotheses

Based on our review of the literature and experiment design, we formulated the following four hypotheses:

- $H_{1.1}$: WiM Control would lead to a better overall performance than Joystick Control;
- $H_{1.2}$: WiM Control would lead to better local tasks performance in *T1 - T3* than Joystick Control;
- $H_{2.1}$: Visible Map would lead to better overall performance than Invisible Map;
- $H_{2.2}$: Visible Map would lead to better local tasks performance in *T1 - T3* than Invisible Map;
- $H_{3.1}$: There would be interaction effects showing that the combination of WiM Control and Visible Map would lead to better overall performance;
- $H_{3.2}$: There would be interaction effects showing that the combination of WiM Control and Visible Map would lead to better performance in local tasks;
- $H_{4.1}$: The combination of WiM Control and Visible Map would lead to lower workload demands;
- $H_{4.2}$: The combination of WiM Control and Visible Map would lead to higher user preferences.

5 RESULTS

All participants understood the nature of the tasks, and all recorded data were valid. A Shapiro-Wilk test for normality was performed on each measure separately for each condition and showed that they followed a normal distribution. To examine interaction effects for non-parametric data, we applied Aligned Rank Transform [Elkin et al. 2021] on NASA-TLX and UEQ data before performing repeated measures ANOVAs (RM-ANOVAs) with them.

5.1 Objective Results

5.1.1 Task Performance. A two-way RM-ANOVA showed two main effects on the time of collisions for Control Methods ($F_{1,15}$ = 26.879, p <.0001) and Map Visibility ($F_{1,15}$ = 27.685, p <.0001) respectively (see also Figure 5a). Another RM-ANOVA found two main effects on completion time for Control Methods ($F_{1,15}$ = 26.811, p <.0001) and Map Visibility ($F_{1,15}$ = 22.337, p <.0001) respectively. However, there was no interaction effect between Control Methods × Map Visibility.

Table 1: All Simple effects for workload data.

	Control Methods		Map Visibility	
Demands	Invisible Map	Visible Map	Joystick	WiM
Mental	p >0.5	Joystick >WiM F = 35,952, p <0.001	p >0.5	p >0.5
Physical	p >0.5	Joystick >WiM F = 69.222, p <0.001	p >0.5	p >0.5
Temporal	p >0.5	Joystick >WiM F = 31.095, p <0.001	Without Map >Visible Map F = 20.077 p <0.001	p >0.5
Performance	p >0.5	Joystick >WiM F = 72.142, p <0.001	Without Map >Visible Map F = 41.404, p <0.001	p >0.5
Effort	p >0.5	Joystick >WiM F = 26.825, p <0.001	Without Map >Visible Map F = 20.622, p <0.001	Without Map >Visible Map F = 6.377, p <0.05
Frustration	p >0.5	Joystick >WiM F = 72.617, p <0.001	Without Map <Visible Map F = 33.644, p <0.001	p >0.5

Table 2: All Simple effects for teleoperator preference data.

	Control Methods		Map Visibility	
Preferences	Invisible Map	Visible Map	Joystick	WiM
Attractiveness	p >0.5	Joystick <WiM F = 72.617, p <0.001	Without Map >Visible Map F = 125.088, p <0.001	Without Map >Visible Map F = 5.55, p <0.001
Perspicuity	p >0.5	Joystick >WiM F = 48.267, p <0.001	p >0.5	p >0.5
Efficiency	p >0.5	Joystick <WiM F = 43.334, p <0.001	Without Map >Visible Map F = 17.983 p <0.001	p >0.5
Dependability	p >0.5	Joystick <WiM F = 73.308, p <0.001	Without Map >Visible Map F = 16.054, p <0.001	Without Map <Visible Map F = 16.397, p <0.001
Stimulation	Joystick >WiM F = 5.866, p <0.05	Joystick <WiM F = 21.908, p <0.001	Without Map >Visible Map F = 121.884, p <0.001	p >0.5
Novelty	p >0.5	Joystick <WiM F = 54.126, p <0.001	p >0.5	p >0.5

Two Bonferroni post-hoc tests revealed that the time of collisions and completion time were significantly lower for WiM Control compared to Joystick Control (Control Methods, p <.0001). The collision time of the Joystick Control group was 8.473s higher than that of the WiM Control group (95% confidence interval: 4.898 - 11.956s). The collision time of the Invisible Map group was 11.216s higher

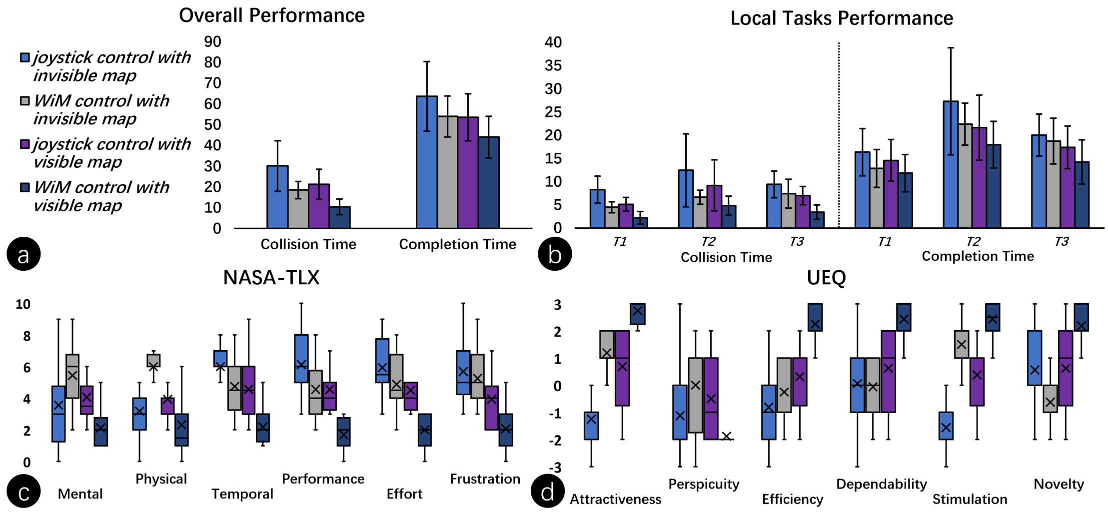

Figure 5: Mean collision time, and mean completion times of (a) overall tasks and (b) each local task; (c) box Plots of workload demands, and of (d) user preferences. The error bars represent 95% confidence intervals. '×' in box plots represents the mean value.

than that of the Visible Map group (95% confidence interval: 6.672 - 15.759s). They were also significantly lower for Visible Map compared to Invisible Map (Map Visibility, p <.0001). The completion time of the Joystick Control group was 9.998s higher than that of the WiM Control group (95% confidence interval: 5.882 - 14.113s). The completion time of the Invisible Map group was 9.658s higher than that of the Visible Map group (95% confidence interval: 5.302 - 14.013s).

5.1.2 Local Task Performance. We found similar effects in the local tasks. RM-ANOVAs showed the main effects on the time of collisions and completion time for both Control Methods and Map Visibility in all local tasks. There were no interaction effects between Control Methods × Map Visibility in all local tasks (*T1-T3*) (see also Figure 5b). In *T1*, the RM-ANOVA showed the main effects on the time of collision (Control Methods ($F_{1,15}$ = 19.651, p <.0001) and Map Visibility ($F_{1,15}$ = 19.736, p <.0001)). In *T2*, the RM-ANOVA showed the main effects on the time of collision (Control Methods ($F_{1,15}$ = 26.879, p <.0001) and Map Visibility ($F_{1,15}$ = 27.685, p <.0001)). In *T3*, the main effects showed Control Methods ($F_{1,15}$ = 26.879, p <.0001) and Map Visibility ($F_{1,15}$ = 27.685, p <.0001).

For WiM Control compared to Joystick Control (Control Methods), a Bonferroni post-hoc test revealed that the time of collisions was significantly lower in *T1* (p <.0001), *T2* (p <.05) and *T3* (p <.01); and the completion time was also significantly lower for Visible Map compared to Invisible Map in *T1* (p <.05), *T2* (p <.01) and *T3* (p <.0001).

For Visible Map compared to Invisible Map (Map Visibility), a Bonferroni post-hoc test revealed that the time of collisions was significantly lower in *T1* (p <.0001), *T2* (p <.01) and *T3* (p <.01); and the completion time was only significantly lower for Visible Map compared to Invisible Map in *T1* (p <.01) and *T2* (p <.05).

5.2 Subjective Results

Figure 5c and d showed the box plots of all NASA-TLX workload data and all UEQ data, respectively. No outliers were found by studentizing whether the residuals exceeded ±3. After applying Aligned Rank Transform to the subjective data, RM-ANOVAs showed interaction effects for all elements of the NASA-TLX workload and UEQ data. Then, we analyzed the data again to find simple main effects for each variable.

5.2.1 Workload Demands. Table 1 summarizes the results of the workload data. There was no significant difference (p >0.5) between the two control methods (Joystick vs. WiM) when the miniature map was invisible for all demand categories. However, WiM Control led to a significantly lower workload (p <0.001) than Joystick Control when the miniature map was visible for all demand categories.

There was no significant difference (p >0.5) between two map visibility (Invisible Map vs. Visible Map) whether it is Joystick Control or WiM Control for Mental and Physical demands. For Temporal demands and Performance demands, Invisible Map led to a significantly higher workload (p <0.001) than Invisible Map when using Joystick Control, but no difference was found when using WiM Control (p >0.5). For Effort demands, Invisible Map led to a significantly higher workload than Invisible Map when using Joystick Control (p <0.001) or WiM Control (p <0.05). For Frustration demands, we found that Invisible Map led to a lower workload than Visible Map (p <0.001) when using Joystick Control but no difference was found when using WiM Control.

5.2.2 User Preferences. Summary results are shown in Table 2. There was no significant difference (p >0.5) between the two control methods (Joystick vs. WiM) when the miniature map was invisible for all preferences except Stimulation (Joystick Control >WiM Control, p <0.05). There was a significant difference between the two control methods (Joystick Control vs. WiM Control, p <0.001) when the miniature map was visible. All the elements of UEQ showed that Joystick Control was preferred by participants except Perspicuity.

When using Joystick Control, there were significantly higher preferences for Invisible Map in Attractiveness, Efficiency, Dependability, and Stimulation (p <0.001) but no difference was found in Perspicuity. When using WiM Control, results showed higher preferences for Invisible Map in Attractiveness (p <0.001); but lower preferences for Visible Map in Dependability (p <0.001). There were no significant differences in the other four elements of UEQ (p >0.05).

6 DISCUSSION

In terms of overall task performance, whether using WiM Control or providing a Visible Map can reduce the collision times and completion time during the teleoperation of the UGV, which supports $H_{1.1}$ and $H_{2.1}$. However, we did not find any interaction effect and, as such, we cannot confirm whether the combination of two variables can significantly improve user performance—that is, part of $H_{3.1}$ is not supported. We found the same main effects but no interaction effect in each local task, which confirmed $H_{1.2}$ and $H_{2.2}$, but the entire $H_{3.2}$ could not be confirmed.

The results show that the two factors (Control Methods and Map Visibility) are independent of each other affecting the teleoperators' performance. From the collision performance point of view, the reduction of the collision time of WiM Control relative to Joystick Control (8.473s) is lower than that of Visible Map relative to Invisible Map (11.216s). In terms of efficiency, there is little difference in the reduction of completion time (9.998s vs. 9.658s) between them. Therefore, these results indicate that using WiM Control or providing a visible miniature map can improve the accuracy and efficiency of users' teleoperation of UGV. Providing visible maps can significantly enhance the accuracy (i.e., reduction of errors and improvement in collision performance). The objective results in the local tasks showed similar effects as the overall task. Our results point to the observations of the overall task, where the task difficulty would gradually increase with decreased width of the pathway, which would then require performing more movements of the UGV.

Regarding workload demands, we found that using WiM Control rather than Joystick Control could significantly reduce demands on teleoperators in all aspects of workload when the map was visible. On the other hand, when using Joystick Control, Visible Map led to increased teleoperators' work demands, which was reflected in making participants more sensitive to how much time they used, increasing their amount of effort, and reducing their confidence in their overall performance. Also, providing a miniature map without interacting with it reduced frustration. Because the miniature map would provide teleoperators with additional spatial information, which helped reduce frustration and increase confidence. When using WiM Control, Visible Map significantly reduced the effort level of the teleoperator. These observations give strong support to the $H_{4.1}$ related to workload.

In terms of user preferences, participants thought it was more exciting and motivating to use Joystick Control when the map was invisible. In contrast, when the map was visible, teleoperators had a better overall impression of WiM Control; they thought WiM Control was more efficient, easier to use, more exciting, and novel. However, due to its novelty, WiM Control also required some initial learning, especially for those with limited experience with VR. When using Joystick Control, teleoperators preferred Invisible Map rather than having the map visible in overall impression, efficiency, sense of control, and degree of excitement. Their preference was understandable as the map could represent a distracting factor. This observation confirms two non-significant results (Mastery and Novelty). When using WiM Control, teleoperators found it easier to control when they could see the miniature map but they suggested that they could also do well without the map visible. While the operation was relatively more complex (than Joystick Control), it was considered more natural and closer to how they would instruct the UGV's movements in real life. They further thought it was more attractive when there was no miniature map but had access to only the miniature UGV for control, which gave them a better overall impression of WiM control with an invisible map. These conclusions also support $H_{4.2}$ related to user preferences.

Regarding images, our method fruitfully supports the variable of map size. When the task requires fast, flexible, and collision-free access to a designated location, a teleoperator only needs to plan the route of the surrogate in mind and let the UGV track the following movements of the hand in real-time. However, suppose a teleoperator wants to perform slow, precise action at the designed location. In that case, our miniature map allows the teleoperator to zoom in and out of the immersive environment. In this manner, our method provides the teleoperator with a flexible and efficient way to control the UGV compared to the traditional dual-joystick control method. Our method also offers the teleoperator rotatable and orthogonal views of the miniature map. For instance, the teleoperators can hold the UGV by turning a miniature map if the UGV's position exceeds the rotatable range of their wrists. Moreover, the converted perspective view from the aerial view of the drone can precisely support the teleoperators in determining the distance in the cyber-physical world.

In terms of UGV control, the UGV in the present work is designed to keep up with the movement speed of the hand as fast as possible based on its performance. However, the UGV chasing the hand may occur with fast hand movements that exceed a specific speed even when our site setting is based on regular level ground (e.g., inside a building). If it is in the mud after rain or on a rough mountain road with a slope, the UGV might experience a nonlinear movement speed caused by slippage or insufficient power, and so the UGV may also chase the hand. Chasing behavior can be thought of as short-distance, straight-line waypoint movement manipulation, which means that the UGV would reach the target location at a straight-line distance with full power. However, if the path of the teleoperator's hand is not in a straight line, chasing behavior may cause the UGV to move in the wrong direction. Therefore, our method requires the UGV to be flexible and performant; that is, it has the ability to adapt to different environments.

7 CONCLUSION

In this paper, we proposed customized WiM interfaces in VR to enable the teleoperation of UGVs. We started with a flat 2D miniature map and UGV to enable the remote operation of the UGV from a third-person view (i.e., a bird's-eye view). Our results from an experiment involving precise remote control of the UGV showed that the miniature maps and UGV represent a promising framework for VR interfaces. Their use in the VR interface led to more efficient and accurate teleoperation performance, lower workload demands, and higher user preference. Since our work was successful, we have opened the door for other researchers to deal with various unexplored areas, such as converting a 2D plane into a 3D space for operating UGVs or UAVs, using simple hand operations for deploying multiple UGVs remotely, and the use miniature map(s) and UGVs by different users.

ACKNOWLEDGMENTS

We thank the participants for their time and the reviewers for their thoughtful feedback. This work was supported in part by XJTLU Key Special Fund (#KSF-A-03) and XJTLU Research Development Fund (#RDF-21-20-008).

REFERENCES

Andrew Bluff and Andrew Johnston. 2019. Don't Panic: Recursive Interactions in a Miniature Metaworld. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. 1–9.

Pierre Bourdieu. 2020. Outline of a Theory of Practice. In *The new social theory reader*. Routledge, 80–86.

Junshen Chen, Marc Glover, Chenguang Yang, Chunxu Li, Zhijun Li, and Angelo Cangelosi. 2017. Development of an Immersive Interface for Robot Teleoperation. In *Towards Autonomous Robotic Systems*, Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lekakou (Eds.). Springer International Publishing, Cham, 1–15.

Kurtis Danyluk, Barrett Ens, Bernhard Jenny, and Wesley Willett. 2021a. A Design Space Exploration of Worlds in Miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 122, 15 pages.

Kurtis Danyluk, Barrett Ens, Bernhard Jenny, and Wesley Willett. 2021b. A design space exploration of worlds in miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

Olaf Diegel, Aparna Badve, Glen Bright, Johan Potgieter, and Sylvester Tlale. 2002. Improved mecanum wheel design for omni-directional robots. In *Proc. 2002 Australasian conference on robotics and automation, Auckland*. 117–121.

Christophe Domingues, Mouna Essabbah, Nader Cheaib, Samir Otmane, and Alain Dinis. 2012. Human-Robot-Interfaces based on Mixed Reality for Underwater Robot Teleoperation. *IFAC Proceedings Volumes* 45, 27 (2012), 212–215. 9th IFAC Conference on Manoeuvring and Control of Marine Craft.

John V. Draper, David B. Kaber, and John M. Usher. 1998. Telepresence. *Human Factors* 40, 3 (1998), 354–375. PMID: 9849099.

Adam Drogemuller, Andrew Cunningham, James Walsh, Bruce H Thomas, Maxime Cordeil, and William Ross. 2020. Examining virtual reality navigation techniques for 3D network visualisations. *Journal of Computer Languages* 56 (2020), 100937.

Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 754–768.

Mohamed Elobaid, Yue Hu, Giulio Romualdi, Stefano Dafarra, Jan Babic, and Daniele Pucci. 2019. Telexistence and teleoperation for walking humanoid robots. In *Proceedings of SAI Intelligent Systems Conference*. Springer, 1106–1121.

Sara Falcone, Gwenn Englebienne, Jan Van Erp, and Dirk Heylen. 2022. Toward Standard Guidelines to Design the Sense of Embodiment in Teleoperation Applications: A Review and Toolbox. *Human–Computer Interaction* 0, 0 (2022), 1–30.

Gal Gorjup, Anany Dwivedi, Nathan Elangovan, and Minas Liarokapis. 2019. An intuitive, affordances oriented telemanipulation framework for a dual robot arm hand system: On the execution of bimanual tasks. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3611–3616.

Simone Grassini, Karin Laumann, and Martin Rasmussen Skogstad. 2020. The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does). *Frontiers in Psychology* 11 (2020).

Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

Rebecca Hetrick, Nicholas Amerson, Boyoung Kim, Eric Rosen, Ewart J. de Visser, and Elizabeth Phillips. 2020. Comparing Virtual Reality Interfaces for the Teleoperation of Robots. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*. 1–7.

Matthias Hirschmanner, Christiana Tsiourti, Timothy Patten, and Markus Vincze. 2019. Virtual reality teleoperation of a humanoid robot using markerless human upper body pose imitation. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, 259–265.

Muthukkumar S. Kadavasal and James H. Oliver. 2009. Virtual Reality Interface Design for Multi-Modal Teleoperation *(World Conference on Innovative Virtual Reality, Vol. ASME-AFM 2009 World Conference on Innovative Virtual Reality)*. 169–174.

Peter Kazanzides, Balazs P Vagvolgyi, Will Pryor, Anton Deguet, Simon Leonard, and Louis L Whitcomb. 2021. Teleoperation and Visualization Interfaces for Remote Intervention in Space. *Frontiers in Robotics and AI* 8 (2021).

Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76.

Ziming Li, Yiming Luo, Jialin Wang, Yushan Pan, Lingyun Yu, and Hai-Ning Liang. 2022. Collaborative Remote Control of Unmanned Ground Vehicles in Virtual Reality. In *2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET)*. 1–8. https://doi.org/10.1109/IMET54801.2022.9929783

Yiming Luo, Jialin Wang, Hai-Ning Liang, Shan Luo, and Eng Gee Lim. 2021. Monoscopic vs. Stereoscopic Views and Display Types in the Teleoperation of Unmanned Ground Vehicles for Object Avoidance. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 418–425.

Yiming Luo, Jialin Wang, Rongkai Shi, Hai-Ning Liang, and Shan Luo. 2022. In-Device Feedback in Immersive Head-Mounted Displays for Distance Perception During Teleoperation of Unmanned Ground Vehicles. *IEEE Transactions on Haptics* 15, 1 (2022), 79–84. https://doi.org/10.1109/TOH.2021.3138590

Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems* 77, 12 (1994), 1321–1329.

Abdeldjallil Naceri, Dario Mazzanti, Joao Bimbo, Yonas T Tefera, Domenico Prattichizzo, Darwin G Caldwell, Leonardo S Mattos, and Nikhil Deshpande. 2021. The Vicarios Virtual Reality Interface for Remote Robotic Teleoperation. *Journal of Intelligent & Robotic Systems* 101, 4 (2021), 1–16.

Nicolas Nostadt, David A. Abbink, Oliver Christ, and Philipp Beckerle. 2020. Embodiment, Presence, and Their Intersections: Teleoperation and Beyond. *J. Hum.-Robot Interact.* 9, 4, Article 28 (may 2020), 19 pages.

R. Ott, M. Gutierrez, D. Thalmann, and F. Vexo. 2005. VR haptic interfaces for teleoperation: an evaluation study. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.* 789–794.

Yushan Pan. 2021. Reflexivity of Account, Professional Vision, and Computer-Supported Cooperative Work: Working in the Maritime Domain. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 370 (oct 2021), 32 pages. https://doi.org/10.1145/3479514

Yushan Pan, Arnfinn Oksavik, and Hans Petter Hildre. 2021. Making Sense of Maritime Simulators Use: A Multiple Case Study in Norway. *Technology, Knowledge and Learning* 26, 3 (2021), 661–686.

Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 265–272.

Michail Theofanidis, Saif Iftekar Sayed, Alexandros Lioulemes, and Fillia Makedon. 2017. Varm: Using virtual reality to program robotic manipulators. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*. 215–221.

Nhan Tran, Josh Rands, and Tom Williams. 2018. A hands-free virtual-reality teleoperation interface for wizard-of-oz control. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*.

Ramón Trueba, Carlos Andujar, and Ferran Argelaguet. 2009. Complexity and occlusion management for the world-in-miniature metaphor. In *International Symposium on Smart Graphics*. Springer, 155–166.

Michael E. Walker, Hooman Hedayati, and Daniel Szafir. 2019. Robot Teleoperation with Augmented Reality Virtual Surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 202–210.

Xian Wang, Diego Monteiro, Lik-Hang Lee, Pan Hui, and Hai-Ning Liang. 2022. VibroWeight: Simulating Weight and Center of Gravity Changes of Objects in Virtual Reality for Enhanced Realism. In *2022 IEEE Haptics Symposium (HAPTICS)*. 1–7. https://doi.org/10.1109/HAPTICS52432.2022.9765609

Chadwick A Wingrave, Yonca Haciahmetoglu, and Doug A Bowman. 2006. Overcoming world in miniature limitations by a scaled and scrolling WIM. In *3D User Interfaces (3DUI'06)*. IEEE, 11–16.

Murphy Wonsick and Taşkın Padır. 2021. Human-humanoid robot interaction through virtual reality interfaces. In *2021 IEEE Aerospace Conference (50100)*. IEEE, 1–7.