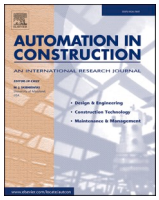
Automation in Construction 156 (2023) 105143

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon



Multi-user immersive environment for excavator teleoperation in construction

Di Liu [a], Jeonghee Kim [b,c], Youngjib Ham [a,*]

[a] Department of Construction Science, Texas A&M University, 3137 TAMU, College Station, TX 77843, USA
[b] Department of Electronic Engineering, the Department of Biomedical Engineering, and Department of Artificial Intelligence, Hanyang University, 04763, South Korea
[c] Department of Engineering Technology & Industrial Distribution, Texas A&M University, College Station, TX 77840, USA

ARTICLE INFO

Keywords:
Human-robot interaction
Teleoperation
Excavator
Virtual reality

ABSTRACT

Excavation on jobsites is a collaborative effort, with a spotter serving as an extra set of eyes for the operator to ensure safety. However, current training methods using simulators are limited in immersion and primarily designed for individual operators to practice basic skills. This paper presents a multi-user teleoperation system featured with hybrid-immersive interface and between-user communication. The effectiveness of the system is evaluated through human subject experiments. The study's findings indicate that operators experienced greater immersion, a stronger sense of presence, and improved user interaction when using an immersive interface. This paper has important implications for the future of excavation training, with potential benefits including reduced utility strike damage and the opportunity to investigate human-robot collaboration in jobsites with multiple workers performing various roles within a highly immersive operational environment.

1. Introduction

Excavators, being among the most commonly used construction machines, represent a significant contributor to construction accidents. One of the most severe accidents in excavation is the collision between the excavator bucket and underground utility lines. The human factor plays an important role in such accident occurrences. In CGA white paper [1], 52% of damages have been reported due to root cause "Excavation Practices Not Sufficient". Operating excavators under challenging work environments and unskilled workforce worsen this situation. The increasing labor shortages have become obstacles to overcoming these challenges. In recent decades, there have been efforts to leverage automation in construction, particularly on deployment of construction robots in jobsites. Unmanned robotics could be implemented in dangerous conditions. Despite the advantages, raising problems cannot be overlooked. First, it is well known that the construction workplace, especially urban jobsites, is particularly dynamic and unstructured, which exposes tremendous challenges in relying on the automation [2,3]. Second, compared to industrial robotic systems typically separated from human workforce, the construction robot usually shared the workplace with humans. It is common that certain tasks in a jobsite (e.g., excavation) involve multiple workers and robotic entities. The interaction between multiple workers and construction robots brings more challenges, affecting the task performance, safety, and human operator's cognitive activities in a more complicated way. Lastly, it is noticeable that robots cannot simply fulfill the roles of the human workforce. Instead of taking over jobs from human counterparts, the goal of a healthy human-robot partnership should augment the potential of human workers, and free workers up to higher-level activities. With this regard, it is necessary to investigate a worker-centered practice that sheds light on resolving problems during operating the robotic excavator.

To improve a human-robot partnership, it is necessary to understand that real-life excavation is often a worker-centered practice performed by a team including multiple workers rather than a single operator. More than one worker is engaged in the excavation since the jobsite requires non-operator personnel (e.g., spotter) to oversee challenges from the workspace and surrounding environment as well as to communicate with the operator simultaneously. Various unpredictable factors may contribute to a challenging environment, including task related factors (e.g., inaccurate locations & depth of buried pipeline, limited workspace, task difficulty levels) and distractors from jobsites (e.g., surrounding work activities, other workers, noise) and surrounding environment (e.g., visual distractor such as traffic and pedestrians,

* Corresponding author.
E-mail addresses: catsquito@tamu.edu (D. Liu), jkim448@hanyang.ac.kr (J. Kim), yham@tamu.edu (Y. Ham).

https://doi.org/10.1016/j.autcon.2023.105143
Received 19 May 2023; Received in revised form 16 October 2023; Accepted 17 October 2023
Available online 21 October 2023

auditory distractors such as noise). These factors compromise the task performance and safety. Moreover, previous studies showed that the operation systems including interfaces and control method often yield cross effect on the operator cognitive workload and further compromise safety [4,5]. Unlike a single worker being responsible for both task execution and safety monitoring, involving multiple workers not only assist on tasks, but respond to jobsite challenges better and reduces operator's workloads.

Among all the non-operator personnel in excavation, a spotter is a human role who can provide real-time signals to an operator while an operator identifies and executes these signals by performing excavator control accordingly. A spotter serves as an extra set of eyes and ears for the operator and plays an important role on jobsite safety [6]. A spotter can be considered as a real-human interface that delivers task execution or safety information to an operator via speech or hand gestures. The successful collaboration between multiple human counterparts and robot(s) with a comprehensive spatial awareness is significant to avoid work accidents, such as underground utility strike during excavation. Thus, there is a necessity to explore between-worker interaction and shared spatial awareness, such as between an operator and a spotter, towards the development of a more efficient and worker-centered human-robot partnership.

This paper focuses on (1) developing a multi-user system, featured with hybrid-immersive and intuitive interface for communication; and (2) identifying challenging factors in an urban excavation site including buried utility lines and testing the effectiveness of the system and analyzing how individual factors affect work performance. The effectiveness of the proposed system is assessed by a user experiment from two aspects, (1) the user evaluation of immersion, sense of presence, and multi-user interaction, (2) the performance assessment of accuracy, efficiency, and safety in terms of underground utility strikes. The proposed hybrid-immersive VR system is assessed and compared with a non-immersive monitor-based system. This study has the potential to make contributions to the body of knowledge in the following aspects: (1) assess the communication between users with different roles in the context of human-robot interaction in jobsites, (2) bridge research gaps through leveraging immersion and intuitiveness of teleoperation interface as well as enhancing the environmental reality of the simulated scenario, and (3) lay out the foundation work for the further investigation on human factors in worker-centered multi-user human-robot teaming contexts. In addition, this study could contribute to the practice building on the developed simulator for multi-user team-based excavation training.

2. Research background

2.1. Immersive multi-user human-robot collaboration system in construction

According to McKinsey 2022 Report - Technology Trends Outlook that lists 14 technology trends affecting the world in the next two decades, immersive-reality technology (IRT), along with advanced connectivity, applied AI, sustainable energy and consumption, shows particular high relevance to the construction industry among all 20 industry sectors [7]. Immersive-reality technology (IRT) will greatly affect 265 million deskless workers in the global construction workforce by shifting the new wave of remote work, scalability of training, saving cost, and testing simulations more efficiently. According to the taxonomy of virtuality-reality, IRT includes AR, VR, MR, XR. There have been research efforts regarding multi-user implementation in various sub-areas. First, construction safety and training are the major sub-area of IRT implementation [8]. Second, IRT implementation with multi-users involved in building management allows to simulate indoor building environment and to study the related occupant behaviors such as emergency evacuation [9]. Third, IRT allows multiple users from different locations to co-work remotely in the same virtual environment for the design and education purposes [10]. Multiple users using the same VR model in multiple viewports between site works can reduce the time to identify the anomalies and take effective actions [11]. Lastly, it is well-accepted that IRT enables greater leeway in remote control such as teleoperation and accelerates the automation progress as well as the in-depth studies of human-robot collaboration, as a close-to-real work environment and seamless interface design are essential for human-in-the-loop machine control process [12,13]. With the implementation of simulating the virtual entities, virtual workflows, and virtual environments, [14] developed a real-time immersive user interface to allow users to perform crane operation and avoid blind spots. Research on VR-based multi-user construction robot operation, often using non-immersive construction simulator based on monitors or mobile devices [14–18], has brought an acceptable yet inadequate experience of immersion. The inadequate level of immersion may cause less accurate task performance. Implementing robust IRT into excavator remote control has the potential to involve multiple users, and further augment the workforce with the inclusion of divergent groups such as women, elderly, and people with disabilities to be part of the excavation. Some research efforts have been done on implementing IRT in multi-user teleoperation. A previous study proposed a construction equipment training framework which allows trainees to use immersive headsets (HMD) and joysticks to interact with other users or interactable virtual objects such as construction vehicles [19,20]. Nevertheless, these user studies were conducted with limited immersive experience. The studies were often simplified to simulate real work scenarios and rarely designed the workflow for multi users with multiple roles [21]. There is a need for facilitating different users with a proper level of immersion or assessing human factors in a team-based context in robot operation. This paper aims to provide a close-to-real excavator teleoperation experience for multiple users with different roles.

2.2. Multi-user teleoperation system requirements and related immersive technologies

According to the level of robot autonomy for human-robot collaboration [2], human efforts in teleoperation dominate high level cognitive activities such as sensing and planning, and in this regard, robotic excavator and human operator co-act on each motion. This process requires carefully assigning tasks to different users and robots to level up efficiency and safety. As a continual evolving loop, cognitive activities in a multi-user teleoperation workflow, include three aspects, (1) sensing: sense the overall situational context as the work proceeds, such as selectively perceiving task-related signals and distractable stimuli, (2) acting: control a robotic excavator or communicate with other entities; and (3) planning: make decisions for the next step work, which involves judgement and memory. Therefore, in an immersive reality system suitable for multi-user teleoperation, the integration of hardware and software should (1) fulfill different user's roles related to sensing and acting, and (2) respond to the team collaboration such as between-user communication and shared situational awareness to facilitate high-level cognitive operations in the next step work.

IRTs have the potential to satisfy these requirements: hybrid-immersive interface, virtual humans, and intuitive between-user communication. Studies in different fields show that visual sense has the primacy above all other senses in the human information processing [22–24]. One reason is that retinas of human eye hold 70% of the body entire set of sensory receptors, thus when multiple sources of sensual input rival, visual information often receives priority [25]. In a construction jobsite, human workers proceed most information through visual perceptions and operate the construction robot through visual-motor integration. Visual inputs contribute the most to cognitive overload or cognitive tunneling [26], and further lead to poor performance and safety issues. Visual inputs delivered by different types of interfaces greatly affect human operators' visual perception. Studies showed that enhanced visual perception can be achieved through immersive visual

interface, which commonly uses wearable Head-Mounted Display (HMD) or CAVE-like interface to provide a first-person view (FPV), as it produces a high level of immersion, intuitiveness, and realistic experience compared to non-immersive interface [4,27], despite drawbacks of motion sickness and physical discomfort [28]. In fact, prior studies evaluated the usability of different visual interface, e.g., immersive visual interface vs. non-immersive visual interface [29,30], often coupled with an secondary assessment of haptic methods. Kotek et al. evaluated single user's operation error rate on performing a control-button task on a virtual panel with two spatial arrangements (vertical arrangement, horizontal arrangement) and two sensory configurations (visual stimulus, auditory stimulus), by comparing multiple types of operation environments, VR CAVE-like environment with a flystick, VR headset with a glove, an office PC, touchscreen tablet, and a standard control panel. Findings showed that the operation error with control buttons were higher when using VR interface. As this finding did not investigate the interplay effect of different visual interfaces with haptic interfaces, it is hard to simply conclude that the immersive visual interface leads to the variation of operation errors. Morosi et al. investigated several performance metrics and mental workload of operating a virtual excavator with a customized haptic control method in a comparison of two types of visual interfaces (immersive VR headset and flat monitor). And this study provided in-depth discussions that the stereoscopic vision and auditory stimuli enhanced the depth perception of the task environment so that caused improvements regarding making control errors and damages, and further confirmed that excavator simulator design can be benefited from utilizing IRT. In addition, prior studies proposed visual alerts superimposing on a FPV interface, for the purpose of collision avoidance and improving task performance while mitigating the influence of visual data overloaded [31,32]. Despite these efforts of assessing immersive VR interfaces, the following still needs to be further investigated. First, the interplay between multiple sensory configurations, for instance, visual interface and control interface, is rarely investigated, which leads to an insufficient understanding how multi-sensory input methods affect performance outcomes altogether. Second, although conventional non-immersive displays, such as desktop monitors, has been discussed in terms of the limited level of immersion [32,33], it is notable that non-immersive interface has the advantage of allowing users to access a broader range of visual information from multiple viewports simultaneously. For instance, multiple displays represent the entire machine operation process and environment from different view aspects [34,35]. Third, in a multi-user team-based collaborative operation, interface design should be operation-dependent, in other words, interfaces with different levels of immersion should be designed to fulfill different user demands based on specific operations, and there is no one-fits-all solution.

Intuitive between-user communication can be achieved through virtual humans (avatar) and natural communication mode such as speech and gesture. Virtual human (avatar) can be AI-powered or directly controlled by a user to perform various gestures and actions that facilitate communication with other virtual entities or objects [36,37]. The input system that enables the avatar in the virtual environment includes conventional input hardware such as keyboard and mouse, gamepad, touch screen, and motion tracking devices. The avatar has a significant impact on user behaviors in a virtual environment, which has been used for the hazard recognition, safety training, and collaborative education [38]. Meanwhile, human communication in the real world is naturally delivered across channels via speech, body gestures, or facial expression. Similarly, as the virtual replica, intuitive communication can be achieved through natural interfaces such as speech and motion recognition between different entities. Speech was studied for designing intuitive modeling interface and human-building interaction [39,40], and is often considered to be suitable for descriptive tasks. On the other hand, human body movements in the real world can be detected and decoded by the recognition system or accurately remapped onto a 3D virtual model in real time. Human motions can be registered via visual and non-visual methods [41] [42,43], which have been implemented on the ergonomic studies of construction workers' physical activities and risk behaviors or training machines to understand human motions [44–48]. For example, [49] developed a vision-based framework to recognize hand gestures of workers in a jobsite and tested it for the communication between a human worker and a dump truck.

To summarize, on the purpose of investigating the real-life team-based excavation practice in the context of a worker-center human robot teaming, knowledge gaps were found in the excavation simulator design, specifically in the aspects of immersion, intuitiveness, and work contexts. First, as for the level of immersion, existing excavation simulators provide acceptable yet less-immersive operation environment, which is available for a single-user only. Second, despite the availability of various immersive reality technologies, the prior studies primarily focused on validating the advantage of immersive display (e.g., VR headset) and they lack efforts dedicatedly designed to fulfill user's needs. Effective interface design could be benefited from combinations of utilizing both immersive and non-immersive interfaces, and there is no one-fits-all solution. Third, as for simulating the multi-user collaborative operation, it is necessary to enhance the intuitiveness of between-user communication. Fourth, the simulation work scenario of prior works is often inadequate on providing close-to-real excavation job sites which often include various challenging environmental factors, such as buried utility lines. In addition, they lack the engagement and feedbacks from non-operator personnel (e.g., spotter) who actively participates in operating a construction machine and this could lead to the insufficient team-wise understanding about the performance of the operator and the system.

3. Hybrid-immersive interface for excavation simulation

3.1. Excavation simulation platform

The primary goal of the hardware design is to serve as the physical excavator simulator. As the main part of the proposed platform, the VR headset (HTC Vive Pro) was functional as the user display (for operator) with a resolution of 1440 by 1600 per eye. The excavator joysticks that have a USB connection that could be plugged into the PC were selected for the operation. The joystick movements were emulated as keystrokes to tie the joystick movements to key presses and set these key press inputs as the inputs in the virtual model. One of the standard control patterns, ISO control pattern [50], is utilized in the proposed system. The two pedals of the excavator have been chosen to replicate the control in the simulator to best imitate a real excavation experience. We also mimic the actual pedal control of an excavator, i.e., the pedal threads move forward or backward as the user presses each pedal forward or backward, respectively. Unlike the joysticks that have a built-in signal conversion to the PC, we developed a specialized hardware interface to convey the information of the pedals as the inputs of the virtual model. The hardware interface was designed on a custom-designed printed circuit board (PCB) that consists of a voltage regulator, a microcontroller module, and connections for pedals and PC. The voltage regulator steps down the 12 V coming from the wall transformer to the 5 V required input to power each of the pedals. The microcontroller module is also mounted on the PCB and being powered through the USB connection, and also communicated through a serial communication protocol. The analog output signals that represent the position of the pedals are sent to two analog-to-digital pins of the microcontroller. The digitized signals of the pedals are delivered to the PC from the microcontroller module. The signals from the pedals and joysticks could be delivered to the virtual model through Uduino, which helps simplify the communication between the Arduino UNO and the Unity Game Engine. By doing so, the excavator simulator that is designed in the Unity can read the interactive information from joysticks and pedals of the excavator. The overall block diagram of the hardware interface is shown in Fig. 1.

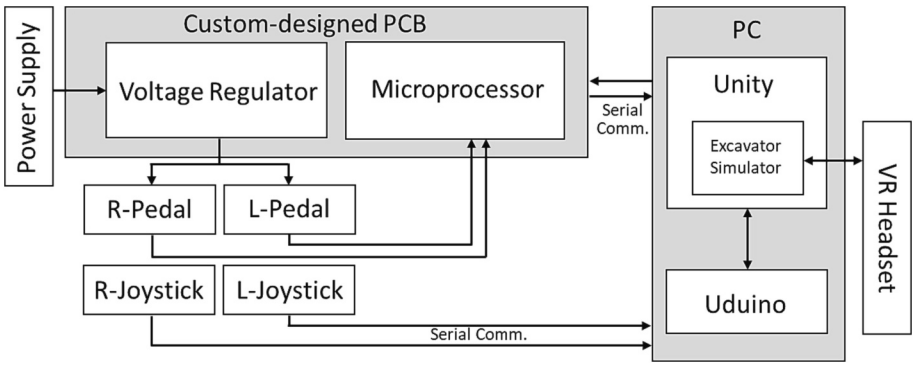


Fig. 1. Block diagram of the custom-designed hardware interface in this study.

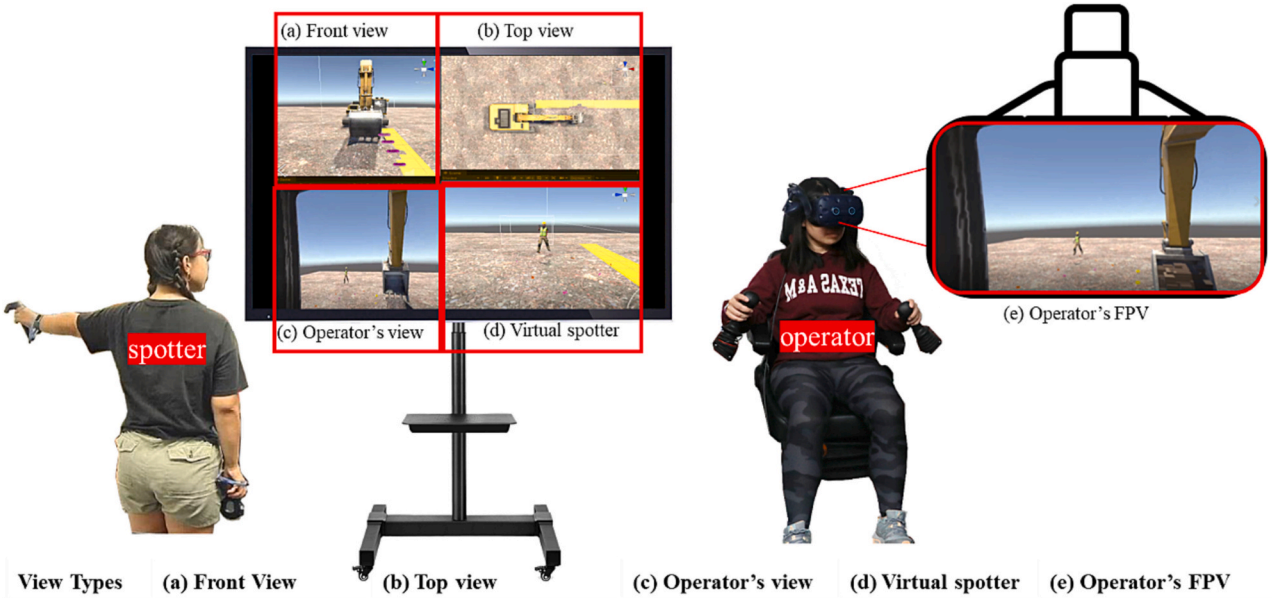
3.2. Hybrid-immersive visual interface and communication system

3.2.1. Visual interface

To properly allocate visual inputs, two attributes of the visual interface, namely, level of immersion (high, low) and types of visual awareness (comprehensive, concentrated), are adopted in the interface design to fulfill specific demands of different users depending on their roles.

The role of the operator is to follow spotter's signals as well as to control the excavator for completion of the given task while avoiding underground utility strikes. In this sense, the operator is primarily expected to have a high level of concentration on excavator bucket and spotter's signals along with a necessary amount of awareness on essential safety cues, such as flags that mark the approximate horizontal location of buried pipelines. To achieve adequate visual awareness which allows the operator to concentrate on a safe excavation and to avoid distraction from unnecessary visual information, a high level of immersion and concentration are critical. Thus, a full-immersive visual display with a first-person view was provided for the operator [Fig. 2].

The role of a spotter, on the other hand, who sends signals to the operator, monitors the excavation, detects potential utility strike risks and other environmental distractors, demanding a holistic spatial visibility. Unlike the operator, the spotter doesn't directly interact with the excavator or directly concentrate on excavation task, and it is suitable to facilitate the spotter with a visual interface that allows to access multiple view perspectives so that the spotter can achieve comprehensive visual awareness. To this end, a set of four viewports, composed of a top view of the entire task space, a front view, a first-person view allowing the spotter to see what the operator was seeing simultaneously, was displayed in a monitor [Fig. 2]. Multiple viewports allow the spotter to monitor the excavator and task executions. In addition, from the front view, top view and virtual spotter view, the spotter monitored the accurate position of buried utility lines marked by the visually enhanced cue which set invisible to the operator. The spotter can also check the avatar's motion signals from the virtual spotter view and operator's visual awareness from the operator's view. Overall, the spotter is



View Types	(a) Front View	(b) Top view	(c) Operator's view	(d) Virtual spotter	(e) Operator's FPV
User	Spotter	Spotter	Spotter	Spotter	Operator
Purpose	Oversee task and buried utility line	Oversee task, environment, and buried utility line	Oversee operator's visual awareness	Oversee hand signals	Perform excavation, check spotter's signals
Interface Types	Non-immersive display	Non-immersive display	Non-immersive display	Non-immersive display	Immersive VR headset

Fig. 2. Hybrid-Immersive visual interface.

facilitated with visual inputs of the excavator, task zone, and the location of buried utility lines. The visual interface with multiple viewports by a large display allows the spotter to enhance spatial visibility to supervise the operator on each step.

To summarize the interface design, the FPV interface designed for the operator provides a high level of immersion and a concentrated awareness that ensure the direct operation with the excavator while trying to minimize distractions from unnecessary inputs; the multi-viewport interface designed for the spotter provides a comprehensive awareness with a low level of immersion that ensure a holistic spatial visibility to monitor the operator, excavator, task space, and surrounding environment altogether [Fig. 3]. To be noted, such interface design ensures different types of awareness (e.g., concentrated, comprehensive) tailored to meet the needs of the different roles of users.

3.2.2. Between-user communication interface

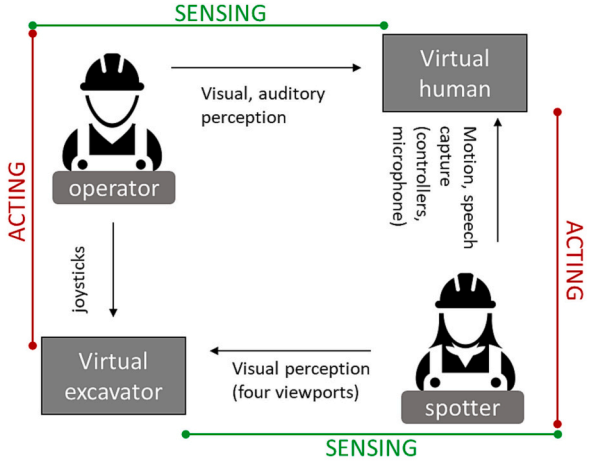
The between-user communication happens between the operator and the spotter in real time. Due to their different tasks and roles, they sense different aspects of the environment and act accordingly. For the operator, visual sensing occurred through one FPV delivered by an immersive display, which allowed a high level of immersion and ensured the operator to concentrate on operating the excavator [Fig. 3]. For the spotter, visual sensing occurred through four different views delivered by a monitor, which allowed low immersion yet high comprehensiveness, and this ensured the spotter to monitor the excavation, workspace, surrounding environment, and potential risks simultaneously, as well as to send hand/verbal signals responsively [Fig. 3]. Spotter's motion signals were captured by motion controllers in real time, and the coordinates of body joints were mapped onto a virtual avatar at each frame, which was modeled with human appearance [Fig. 4]. The avatar is the representative of the spotter, and the operator communicated with the avatar, i.e., receiving signals [Fig. 4]. Avatar (spotter's virtual replica) was placed right in front of the digging workspace, which ensured that it was perceived in the operator's field of view when the operator performs given tasks. To better mimic the restriction of limited workspace in an urban excavation scenario, the avatar was not able to walk around and stand at the same position to deliver different signals to the operator during the entire experiment. Moreover, the spotter's speech signals were captured by a microphone and transferred to the operator. Overall, the communication interface between the operator and spotter takes account of natural human communication (hand and speech signals) to ensure robust communication when both sensing and acting occur during work [Fig. 3].

4. Human subject experiments

4.1. Experimental setup

Table 1 presents the information of a total of 57 participants in the experiments approved by the University Institutional Review Boards (IRB). All participants were recruited via the bulk mail system (for those who are in construction-related majors) and completed an online screen process to ensure the eligibility of operating two joysticks and no vision or hearing impairment. Prior to the experiment, a pilot study has been conducted to test the protopype functionality with employing a small group of users ($N < 10$) [51].

A within-group experiment was conducted in a single day. The entire experiment for each participant lasted for 45–70 min approximately and included the following steps. After arrived for the experiment, participants reviewed and signed the consent form first. A background questionnaire was completed including the demographic information, construction work experience, VR/video game experience. Then, an instruction session was provided by explaining a standard excavator structure, the ISO control method, and showing a tutorial video of excavator control. After that, participants received the first training session by operating an excavator with two joysticks following the ISO control. Participants took as long as they needed to practice until they were confident to operate the excavator. The next two sessions were two trials where the participants performed an excavation task solely without working with a spotter. In each trial, the participants accessed the virtual environment using either the headset or the monitor respectively [Fig. 5]. The participants completed the questionnaire once they finished the trial. Then, the participants were provided with the second training session by following the spotter's hand signals to operate the excavator. The spotter introduced each signal to the participant and answered any related questions. The participant practiced ten hand



	Level of Immersion	Awareness/View	Communication
Sensing	[Operator] Immersive display, [Spotter] Non-immersive display	[Operator] Concentrated, first-person-view [Spotter] Comprehensive, multiple viewports	[Operator] visual, auditory perception [Spotter] visual perception
Acting	[Operator] Interaction with virtual world: remote control of an excavator	--	[Operator] control interface (e.g., joystick)
	[Spotter] Interaction with virtual world: providing signals through an avatar	--	[Spotter] motion/speech capture

Fig. 3. Hybrid-Immersive communication interface.

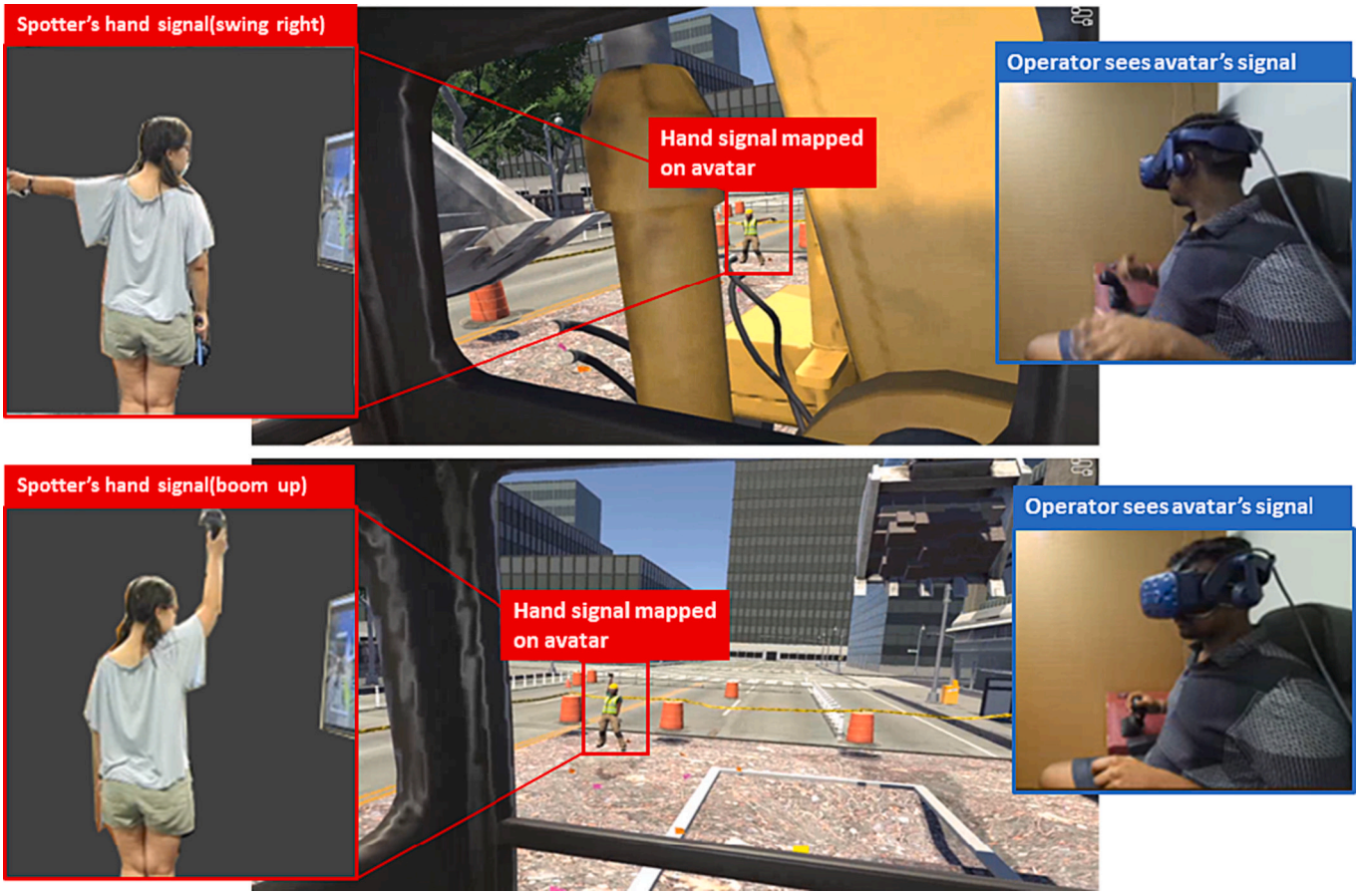


Fig. 4. Motion capture of hand signals for spotter-operator communication.

signals guided by the spotter's avatar in the virtual environment. The training session was conducted using an excavation scenario, and the spotter practiced three times with each participant. After training with the spotter, two experiment trials were conducted. In each trial, participants followed the hand signals from the spotter, and performed the excavation task using either the headset or the monitor respectively [Fig. 5]. Once the second set of two trials were completed, the participants provided responses to the questionnaire. In addition, upon arrival at the experiment location, the participants were required to review graphic instruction for the preparation purposes. The instruction illustrated the standard excavator components, the ISO control pattern of an excavator, a tutorial video about operating a standard excavator in a real jobsite, as well as different signals. This session helped the participants to familiarize and memorize the excavator control and signals.

There was a total of four experiment trials for each participant [Fig. 5]. In the first set of two independent trials, the participant was asked to perform an excavation task three times in an urban construction jobsite independently. The participants were required to excavate in the task zone located in front of the excavator without hitting buried utility lines. Yellow flags marked the horizontal coordinates of the buried utility lines. Upon completion of excavating one load of soil, the participant was required to dump it into a marked dumping zone located on the left side of the excavator then returned to the beginning position for excavation. The participant operated two joysticks to control the bucket, stick, and boom to complete digging action, then to swing the excavator cabinet 90 degree approximately to the left side to dump the soil. This task was repeated three times in trials #1 and #2 respectively. In the second set of two collaborative trials, the participants followed hand signals from the spotter's avatar to perform a similar excavation task. The spotter avatar was set in a standing pose in front of the task zone within the participant's direct field-of-view (FOV) during work. To check signals from the spotter avatar and workspaces in a fully immersive FPV, the participants as the operator were allowed to rotate their heads during the trials. Further, task-related challenging factors described in the introduction section were designed in the urban jobsites to represent the real-world settings. The first challenging factor is underground utility lines within the digging zone, and the operator needs to avoid collisions with the buried utility line and identify approximate positions by checking ground flags. The second factor is the narrow dumping space, and the operator needs to carefully estimate the spatial distance when dumping soils to ensure the task accuracy.

4.2. *Performance measurements*

The user evaluation of this study is to evaluate the effectiveness of hybrid-immersive interface and communication system in aspects of immersion, presence, and between-user interaction, compared to the monitor-based simulator. This evaluation is composed of two parts. Part I is to evaluate the quality of immersion and sense of presence of the entire task scenario based on the first set of two independent trials. As Table 2 shows, a total of nine items of Part I are based on two evaluations to measure the immersion and sense of presence of the virtual reality developed by Witmer et al. and Schwind et al. [52,53]. The immersion was evaluated by naturalness, compelling, visual involvement, visual flexibility, and display quality that affect required task performance while the sense of presence was evaluated by the user experience in a virtual space compared to looking at an image. Participants evaluated these aspects between immersive display and a monitor by using a 7-point Likert scale after they finished two independent excavation trials. As Table 3 shows, Part II is to evaluate the between-user collaboration. A total of three items are included, namely, the duration of visual contact with the spotter avatar, the visual easiness of checking hand

Table 1
Participant Background Information ($n = 57$).

Categories		Response Ranges	Percentage
Gender		Female, Male, Other	F - 37.1%, M - 61.3% Other - 1.6%
Age		18–55 years old	18–29: 80.7% 30–39: 12.9% 40–55: 6.7%
Race		White, African American, Native American, Asian, Pacific islander, multi-races, other	Asian: 35.5% White: 58.1% Multi-races: 4.8% Other: 1.6%
Work experience	Work in a construction job site	0–10 years	No experience - 59.7% Below 5 years - 36.5%
	Operating a construction machine/vehicle	0–10 years, or above 10 years	No experience: 72.6% 0–10 years: 25.8% 10 years or above: 1.6%
VR/Video game experience	Frequency to use VR application	• Never • Monthly or less often • Weekly or a few times a week • Daily	Never = 40.3% Monthly or less often = 40.3% Weekly/a few times a week = 9.7%
	Frequency to play video game	• Never • Monthly or less often • Weekly or a few times a week • Daily	Never = 12.9% Monthly or less often = 54.8% Weekly/a few times a week = 19.4% Daily = 12.9%
	Previous Motion sickness	Yes/No	No = 82.3%, Yes = 17.7%

signals from the avatar, the visual flexibility from multiple view directions. Similarly, the experience of using VR display and monitor was compared and evaluated by a 7-point Likert scale. Part II evaluation was performed after the participants completed two collaborative excavation trials with a spotter.

The task performance was measured based on the task accuracy and the completion time of each trial. The task accuracy is evaluated by performance errors of digging and dumping respectively. The error of digging was defined by the amount of soil loads dug out of the excavation task zone. A higher error contributes to the lower accuracy of digging. The error of dumping was defined by the amount of soil in the bucket dumped out of the dumping zone. A higher dumping error contributes to lower accuracy of dumping. Both digging and dumping errors were calculated based on soil-bucket mesh collision. As the primary

Table 2
VR Effectiveness Survey – Part I.

Immersion Evaluation
1. [Naturalness] How natural did your interactions with the environment seem?
2. [Visual Involvement] How much did the visual aspects of the environment involve you?
3. [Compelling] How compelling was your sense of moving around inside the virtual environment?
4. [Closeness] How closely were you able to examine objects?
5. [Display Quality Interfere] How much did the visual display quality interfere with or distract you from performing assigned tasks or required activities?
6. [Concentration] How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
7. [Flexibility on View Directions] How flexible was your sense of checking objects from different view directions?
Sense of Presence Evaluation
1. [Images vs. Somewhere] When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?
2. [Within vs. Out of the Virtual Environment] During the time of your experience, did you often think to yourself that you were actually in the virtual environment?

Table 3
VR Effectiveness Survey – Part II.

Collaboration Evaluation
1. [Duration of Visual Contact] How long could you maintain visual contact with the spotter's avatar? Could you see the spotter's avatar in a limited amount of time or most of the time during the experiment?
2. [Visual Easiness] How easy was your sense of checking the behaviors from the spotter's avatar? Was it easy for you to detect the avatar's behaviors or was it difficult for you to detect the avatar's behaviors?
3. [Visual Flexibility] How flexible was your sense of checking the signals from the spotter's avatar from multiple view directions?

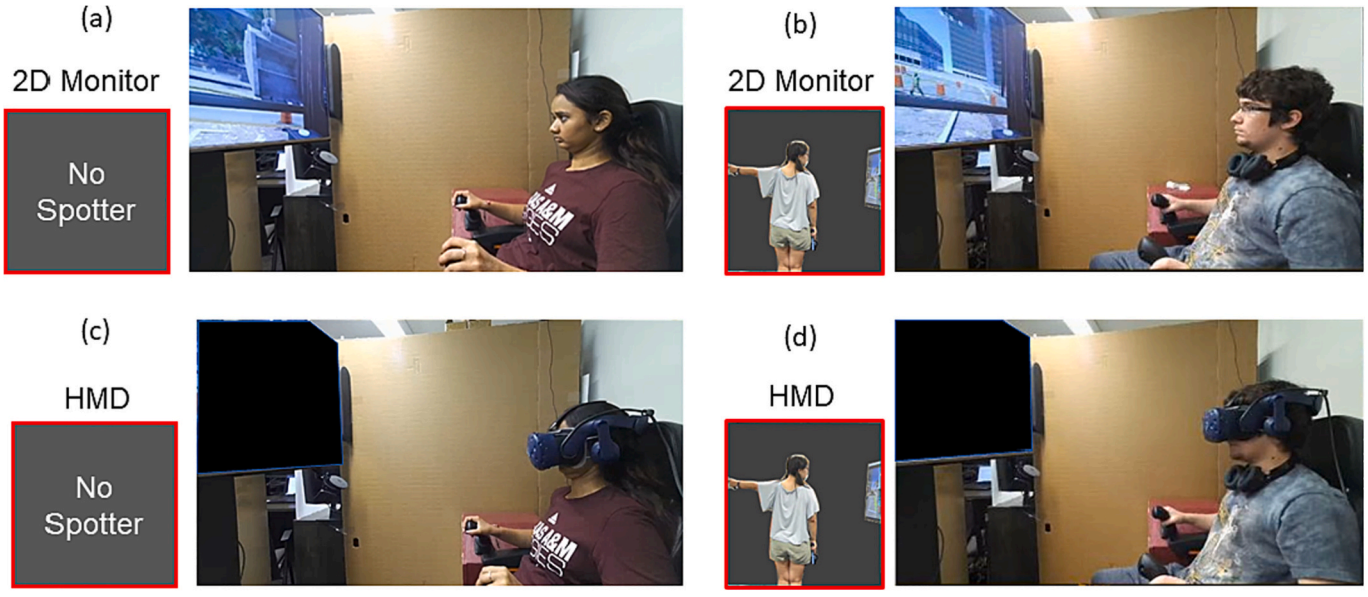


Fig. 5. Experimental trials: (a) independent trial with a monitor, (b) collaborative trial with a monitor, (c) independent trial with HMD, and (d) collaborative trial with HMD.

accident during excavation is the collision between the bucket and underground utility lines buried within the excavation zone, the total number of bucket-utility line collisions for each trial was considered as the metric of the unsafe behavior. In each trial, once such collisions occurred, the collision number (COL#) was automatically counted.

Previous studies attempted to quantify the operator's overall excavation performance by calculating soil excavation productivity. For example, [54] defined the task productivity by soil volume divided by execution time as an indicator of participant's digging skills. Despite the benefits, the factor of avoiding accident occurrence was not integrated. In this study, the overall excavation performance is analyzed by taking account of accident-avoidance building upon the multi-attribute utility theory [55], which allows to weigh multiple attributes (e.g., accuracy, time, and safety) based on their utility (importance). The first step is to normalize each factor and assign weights based on importance. Here, the normalized accuracy, time, and collision rate could be defined as (200 – individual's error) / 200, (individual's task completion time – Min task completion time) / (Max task completion time – Min task completion time), and (individual's collision number – Min collision number) / (Max collision number – Min collision number), respectively. Where 200 is a constant number of the volume of soil particles per load defined in a mesh deformation in this experiment, and (200 – individual's error) is the completed soil volume upon finishing excavation one load of soil per individual. To be noted, in the experiment, the constant number (200) is subject to virtual modeling resolution, such as refined soil particles. A Min-Max scaling method is built upon for normalization to control the sensitivity to outliers in the dataset. The normalized factors could be combined with weights as follows.

$$Performance = (a^{*}\ normalized\ accuracy)/(b^{*}normalized\ time)- (c^{*}\ normalized\ collision\ rate) \quad (1)$$

In Eq. (1), the weights (a, b, c) for each factor (accuracy, time, safety) are adjustable to control the importance of each factor. In this study, we consider accuracy, completion time, and collision-avoidance are equally important to the performance, and thus weights are equal to 1. So, the overall performance can be summarized as.

$$P = \overline{A}/\overline{T} - \overline{COL} \quad (2)$$

As for Eq. (2), a higher value of P indicates better task performance, defined by a higher excavation productivity and a lower collision-occurrence.

Lastly, to further evaluate the effectiveness of immersive visual interface and between-user communication, we investigated if two independent variables, display types and operation types, cause any mixed effect, and if they are equally attributable to the performance outcomes (i.e., completion time, errors, collisions). The mixed effect of two independent variables, namely, display types, operation types, was investigated building on a generalized linear mixed effect model (GLMM) [56] which is defined as below.

$$T_{ij} = \beta_0 + \beta_1{}^{*} DISPLAY_i + \beta_2{}^{*} OPERATION_i + \beta_{ij} + e_{ij} \quad (3)$$

$$E_{ij} = \gamma_0 + \gamma_1{}^{*} DISPLAY_i + \gamma_2{}^{*} OPERATION_i + \gamma_{ij} + \varepsilon_{ij} \quad (4)$$

$$C_{ij} = \delta_0 + \delta_1{}^{*} DISPLAY_i + \delta_2{}^{*} OPERATION_i + \delta_{ij} + \zeta_{ij} \quad (5)$$

As for the Eq. (3), T_{ij} is the completion time for the participant j (j = 1, 2, …57) in each condition (i = 1,2,3,4). β_0 is the intercept, representing the average completion time when both display type and operation type are at their reference levels. β_1 is the coefficient for display type, representing the effect of switching from one display type to the other on completion time (display types: HMD—H, Monitor-M). β_2 is the coefficient for operation type, representing the effect of switching from one operation type to the other on completion time (operation types: operator-only vs. spotter-direct). β_{ij} is the coefficient for the participant j in condition i, representing the variation of individual random effect. e_{ij} is the residual error for the participant j in condition i, representing the variation in time after accounting for the fixed effects (e.g., display types, operation types). As for the Eq. (4), E_{ij} is the error for the participant j (j = 1, 2, …57) in each condition (i = 1,2,3,4). γ_0 is the intercept, representing the average errors when both display type and operation type are at their reference levels. γ_1 is the coefficient for display type, representing the effect of switching from one display type to the other on errors. γ_2 is the coefficient for operation type, representing the effect of switching from one operation to the other on errors. ε_{ij} is the residual error for the participant j in condition i, representing the variation in error after accounting for the fixed effects (e.g., display types, operation types). As for the Eq. (5), C_{ij} is the number of collisions for the participant j (j = 1, 2, …57) in each condition (i = 1,2,3,4). δ_0 is the intercept, representing the average collisions when both display type and operation type are at their reference levels. δ_1 is the coefficient for display type, representing the effect of switching from one display type to the other on collisions. δ_2 is the coefficient for operation type, representing the effect of switching from one operation to the other on collisions. ζ_{ij} is the residual error for the participant j in condition i, representing the variation in in the number of collisions after accounting for the fixed effects (e.g., display types, operation types).

5. Experimental results analysis and interpretation

5.1. VR effectiveness

The Shapiro-Wilk test was first performed to test the normality for the subjective evaluation and concluded that data are not normally distributed. Thus, the nonparametric Wilcoxon Signed Rank Test was selected for the analysis. Significant differences ($p < 0.001$) between the monitor and HMD are found in all twelve items regarding immersion, sense of presence, and between-user interactions [Table 4]. As for Immersion, Fig. 6. shows that immersive display (HMD) was evaluated with higher scores in naturalness, closeness, compelling, visual involvement, visual flexibility, concentration on tasks and activities compared to the monitor. The quality of the immersive display also showed less interference that affects required task performance than the monitor. The immersive display achieved a higher mean score (45.12) than the monitor (24.70) regarding the immersion and sense of presence [Fig. 6]. In addition, results of using 2D-Monitor display show higher variability than the results of wearing HMD on all nine dimensions of evaluating Immersion and sense of presence [Fig. 6] as well as on all three dimensions of evaluating collaboration [Fig. 7]. The difference of variability of using two display types indicates that using a 2D-monitor

Table 4
VR Effectiveness – Immersion, Sense of Presence, Between-user Interaction.

	Naturalness	Visual Involvement	Compelling	Closeness
Monitor vs. HMD	Statistic = 1.0 p = 1.93e-10***	Statistic = 7.5 p = 1.21e-10***	Statistic = 16.0 p = 4.93e-10***	Statistic = 7.5 p = 3.12e-10***
	Display Quality Interfere	**Concentration**	**Flexibility on view directions**	**Images vs. Somewhere**
Monitor vs. HMD	Statistic = 257.5 p = 0.001***	Statistic = 43.5 p = 3.58e-09***	Statistic = 16.0 p = 2.29e-10***	Statistic = 43.5 p = 3.58e-09***
	Within vs. Out of the virtual environment	**Duration of visual contact**	**Interactive Visual Easiness**	**Interactive Visual Flexibility**
Monitor vs. HMD	Statistic = 76.0 p = 2.02e-08***	Statistic = 4.0 p = 3.53e-11***	Statistic = 23.5 p = 5.91e-09***	Statistic = 15.0 p = 1.43e-10***

Significant codes: < 0.001 '***' < 0.01'**' < 0.05 '*' < 0.1 '.'.

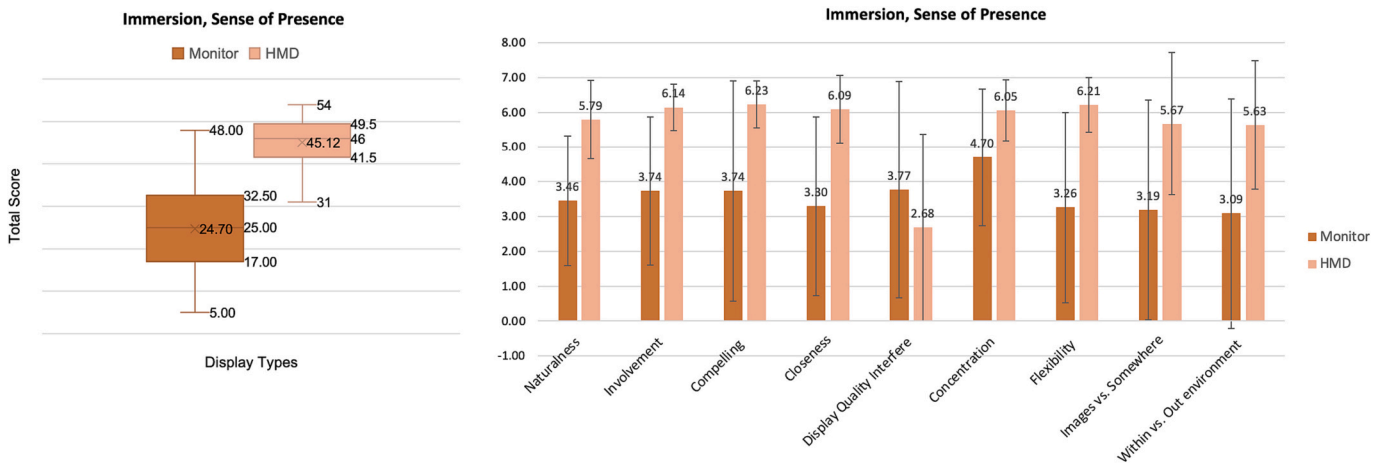


Fig. 6. Results of VR survey – Part I: Immersion, Sense of Presence.

may result a wide range of experiences in terms of immersion, sense of presence, and collaboration. Some participants may feel familiar or comfort of using 2D monitor, while others may struggle to feel presence or easy on collaboration. Results of collaboration evaluation show that the immersive display achieved higher scores in terms of the duration of visual contact with the spotter avatar, the visual easiness of checking hand signals from the avatar, the visual flexibility from multiple view directions [Fig. 7]. Overall, the effectiveness of the multi-user hybrid-immersive system is demonstrated in the aspects of immersion, sense of presence, and collaboration, by comparing it to the monitor-based system, which was validated by the VR effectiveness surveys.

5.2. Task performance

The performance of excavation was measured by errors at work and completion time. As the sample size ($n = 57$) is relatively small but still larger than 50, the Kolmogorov-Smirnov test and the Shapiro-Wilk test were performed to test the data normality in a numerical way. To reduce the oversensitivity caused by the numerical methods, graphical methods including the histogram and Q-Q plot were used to facilitate the normality test. As the Shapiro-Wilk test and the Kolmogorov-Smirnov test showed that data regarding the task completion time, digging errors, and dumping errors are not normally distributed ($p < 0.05$), which was also confirmed by Q-Q plot and histogram. Thus, the nonparametric Wilcoxon Signed Rank Test was selected to determine whether there were significant differences in task performance between the different operation and display types.

5.2.1. Task accuracy

Despite that Table 6 and Fig. 8 show that independent tasks using a monitor yields a slightly higher mean digging error and collaborative tasks using a monitor yields a slightly lower mean digging error, as Table 5 shows, there was no significant difference between the results of digging errors between different types of displays ($p > 0.05$). Meanwhile, significant difference was found in mean dumping error when considering the display types ($p < 0.05$), regardless of independent or collaborative tasks. Results of dumping error demonstrated that when wearing an HMD, the participants tend to perform more accurately than using a monitor. It indicates that increasing the level of display immersion would enhance the user's adaptability to the work environment especially with challenging factors such as a narrow workspace. This can be explained that, during the dumping task, due to the narrow workspace commonly found in such a crowded urban jobsite, the dumping zone was partially occluded in a monitor which provided a fixed viewport not allowing the participant to have a broader vision by rotating the head. While using HMD, the participants could easily adapt to the challenging environment by rotating their heads freely, and finally achieve a broader view of dumping zone, which is not possible when they were using a monitor with a fixed viewport. Consequently, for the simulated scenario in a busy and crowded area with challenging task

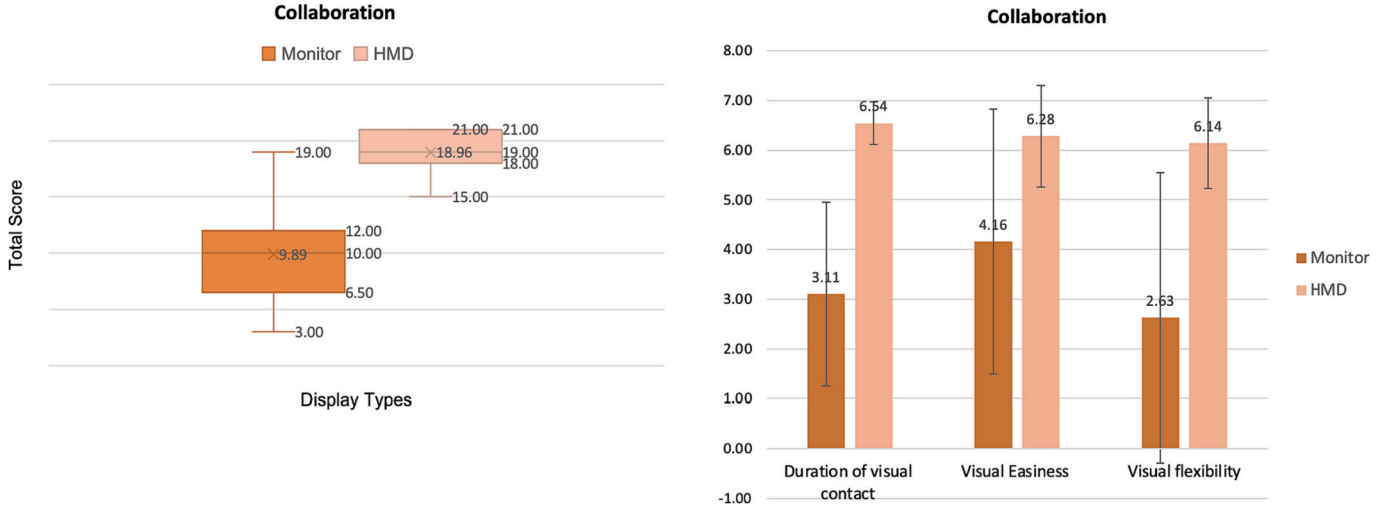


Fig. 7. Results of VR survey – Part 2: Collaboration.

Fig. 8. FPV of the operator during the dumping operation: (a) Independent, 2D monitor trial (spotter stand in front of digging zone without sending any signals) (b) Independent, HMD trial (spotter stand in front of digging zone without sending any signals) (c) Collaborative, HMD trial (spotter stand in front of digging zone and was showing one hand signal) (d) Collaborative, 2D monitor trial (spotter stand in front of digging zone and was showing one hand signal).

Table 5
Results with four different conditions: NS_M (no spotter, monitor), NS_H (no spotter, HMD), S_M (spotter, monitor), S_H (spotter, HMD).

Accuracy – Digging Error			Accuracy – Dumping Error		
	NS_M	S_H		NS_M	S_H
NS_H	stat = 471.0 p = 0.44	stat = 425.5 p = 0.14	NS_H	stat = 340.0 p = 0.0003***	stat = 128.5 p = 0.0000067***
S_M	stat = 487.0 p = 0.21	stat = 471.0 p = 0.78	S_M	stat = 258.0 p = 0.000018***	stat = 181.0 p = 0.000085***
Efficiency - Completion Time(s)			Safety – Number of Collisions		
	NS_M	S_H		NS_M	S_H
NS_H	stat = 548.5 p = 1.31e-06***	stat = 279.0 p = 2.04e-10***	NS_H	stat = 291.0 p = 0.0063**	stat = 9.0 p = 9.47e-06***
S_M	stat = 464.5 p = 4.99e-08***	stat = 60.5 p = 1.12e-13***	S_M	stat = 0 p = 1.04e-08***	stat = 9.0 p = 0.72
Performance score (P)					
	NS_M	S_H			
NS_H	stat = 382 p = 0.0004***	stat = 724 p = 0.4177			
S_M	stat = 143 p = 5.75e-08***	stat = 107 p = 1.11e-08***			

Significant codes: < 0.001 '***' < 0.01'**' < 0.05 '*' < 0.1 '.'.

factor such as a narrow workplace, facilitating with a higher level of immersion provided by HMD allows the participants to have a better visual sense, and eventually were able to reduce task errors and achieve a better performance accuracy than using a monitor. Thus, it is conclusive that with a task related factor (i.e., narrow task space), utilizing an immersive display would be more beneficial in terms of improving the performance accuracy than a monitor because higher level of visual immersion allows the operator to sense the environment better and further adapt to the challenging environment easier. During two trials of wearing HMD, as for the different operation types, result of independent trial shows more dumping errors than collaborative trial. This can be explained that in the collaborative trial multiple viewports allow the spotter to have a comprehensive visual understanding of the task environment and provide the operator with more accurate spatial information through signal communication. While working independently, the lower level of visual comprehensiveness of FPV and the narrow task area were challenging for the operator to self-estimate the spatial distance to accurately dump soils, even though the operator could perceive the dumping scene fully with free head rotations. Overall, the increased level of immersion as well as the between-user collaboration can enhance the user adaptability to the challenging work environment, reduce the obstacle of spatial distance estimation, thus improve performance accuracy in a jobsite with challenging factors. When using 2D-monitor, however, the operator was not able to see the spotter during the dumping operation, yet the dumping accuracy in collaborative trials was significantly improved. One explanation is that since the operator experienced HMD trials between 2D-monitor trials, using HMD was likely to improve the spatial perception and leads to a better dumping accuracy in the second 2D-monitor trial although the operator could not see the spotter during the dumping operation in both 2D-monitor trials [Fig. 8]. Fig. 8 shows the FPV of the operator in four trials. To be noted, in 2D-Monitor trials [Fig. 8a, Fig. 8d], the operator could not see the spotter through FPV during dumping, so the separated views outlined in red and blue at the bottom left corners are added only to clarify the status of spotter.

5.2.2. Task efficiency

Results of task completion time show that significant difference (p < 0.05) is found between different operation types, as well as different display types [Table 5]. Additionally, when using a monitor, independent trials (190 s) took longer time than collaborative trials (168.11 s) [Table 6] [Fig. 9]. In independent trials wearing an HMD took less time (176.88 s) to complete than using a monitor (190 s). Interestingly, collaborative trials wearing an HMD took much longer time (222.56 s) to complete than that using a monitor (168.11 s). This can be explained that dumping soil requires the operator to rotate 90 degrees to the left

Table 6
Results of Efficiency, Accuracy, Accidents, and overall performance (NS_M (no spotter, monitor), NS_H (no spotter, HMD), S_M (spotter, monitor), S_H (spotter, HMD)).

	NS_M	NS_H	S_M	S_H
Efficiency (T) Task Completion Time(s)	Mean = **190.00** (50.9, 6.74) Median = 177.0	Mean = **176.88** (53.12, 7.04) Median = 164.0	**Mean = 168.11 (34.21,4.53)** Median = 209.0	Mean = 222.56 (54.25, 7.19) Median = 162.0
Accuracy (e) Dumping errors	Mean = 25.40 (21.99, 2.91) Median = 18.0	Mean = 13.95 (16.39, 2.17) Median = 10.0	Mean = 10.42 (10.53, 1.40) Median = 0.0	**Mean = 3.79 (8.76, 1.16)** Median = 9.0
Digging errors	Mean = 2.91(7.06, 0.94) Median = 1.0	Mean = 2.51(2.89, 0.38) Median = 2.0	Mean = 1.63 (1.80, 0.24) Median = 1.0	Mean = 1.70(1.95, 0.26) Median = 1.0
Accidents (COL) Number of Collisions	Mean = 4.74(4.79, 0.63) Median = 4.0	Mean = 2.84 (5.34, 0.71) Median = 0.0	Mean = 0.12 (0.43, 0.06) Median = 0.0	**Mean = 0.09(0.39, 0.05)** Median = 0.0
Performance Score (P) (200-e)/T – COL/16	Mean = 0.70 (1.94, −0.43) Median = 0.72	Mean = 0.96 (2.63, 0.59) Median = 0.93	**Mean = 1.16** (1.83, 0.59) Median = 1.15	Mean = 0.93 (1.54, 0.31) Median = 0.96

and cause the spotter was out of operator's front view in a monitor. In this case, in HMD trials, the operator could shift back to the front view to the spotter by rotating the head to the right, and check the spotter's hand signals; however, in monitor trials, the operator could not freely change the direction of front view and check hand signals thus rarely spent time on signal checking during dumping. The fixed direction of field-of-view in monitor trials disables the operator to freely switch front view to a perpendicular direction to check signals, so less time spent on signal checking. Further, results of the mixed effect demonstrate that the display types and operation types are not equally attributable to the task completion time, which will be discussed in 5.2.4.

5.2.3. Collisions

As Table 5 shows, there is a significant difference in the number of collisions made between independent trials and collaborative trials ($p < 0.05$). The operators made more collisions when they worked independently. For independent trials, significant difference ($p < 0.05$) was found on the collision numbers between HMD trials (2.84) and monitor trials (4.74) [Table 6] [Fig. 9]. Results indicate that collaborative trials could lead to safer practice with less collisions made. This can be explained by that multiple viewports and visual cue allow the spotter to see the accurate location of buried utility line, while the operator saw the utility line only by checking the ground flags which were easily removed during the excavation. Results also indicate that with a higher level of immersion, the user senses the environment better and acts safer. Nevertheless, although the level of immersion affects the numbers of collisions, the communication with the spotter played a major role in collision avoidance as the location information of underground utility lines could be delivered via the between-user communication.

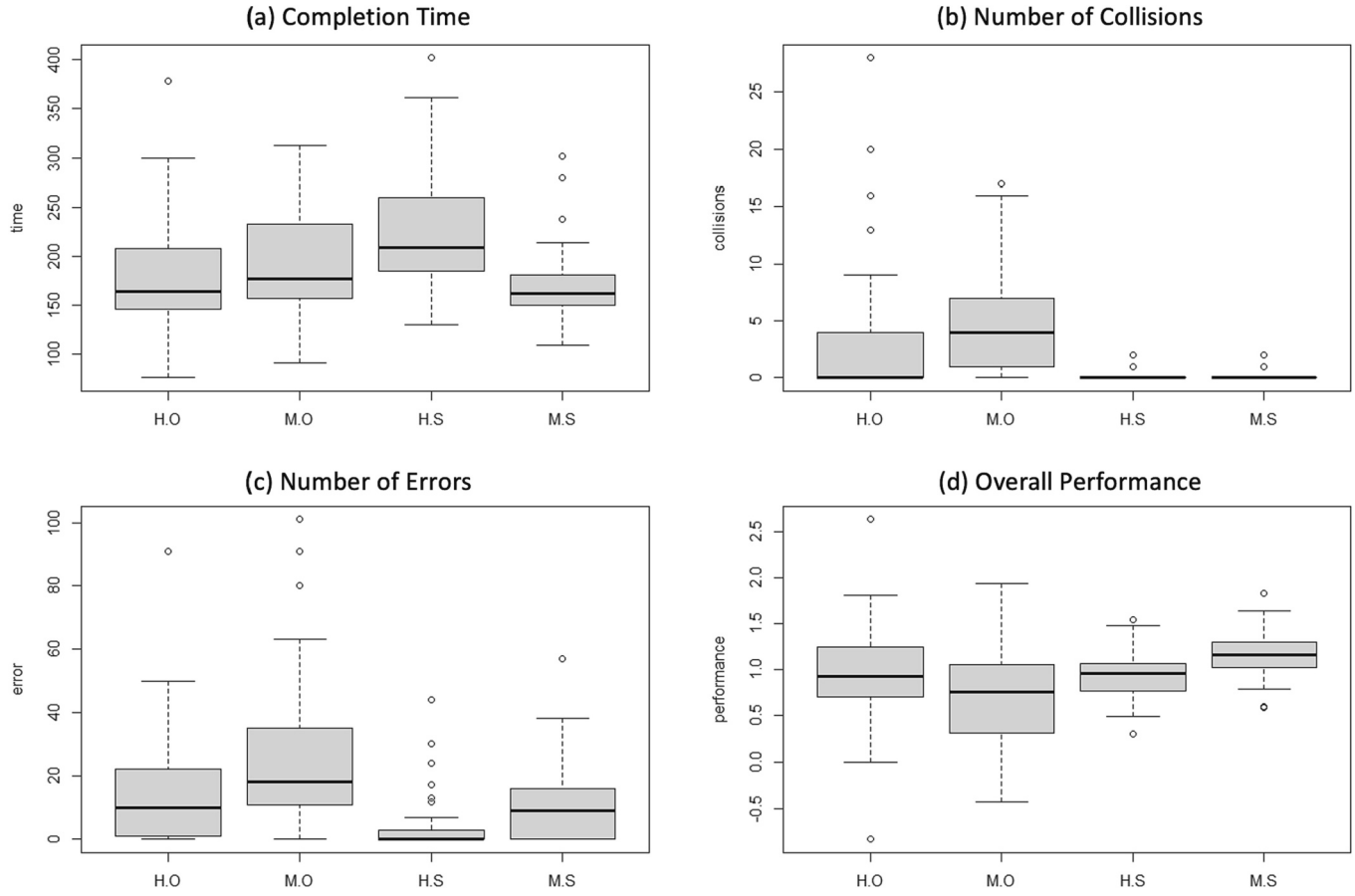


Fig. 9. Results under four different conditions - Independent trial with HMD [H·O], Independent trial with 2D-Monitor [M.O], Collaborative trial with HMD [H·S], Collaborative trial with 2D-Monitor [M.S]: (a) Completion Time (b) Number of Collisions (c) Number of Errors (d) Overall Performance.

5.2.4. Mixed effect

Three performance metrics (i.e., completion time, error, collision) as dependent variables were regressed on two independent variables, as well as their interaction, in a generalized linear mixed-effects model. The reference level in this mode is defined by the outcome of the trial in which the operator was using HMD display to perform the task without a spotter. The significances between the outcome of the reference level and outcomes from other three trials were analyzed [Table 7]. This model included random intercepts for participants, as well as random slopes for display_types and operation_types at the participant level. As Table 7 shows, as for the mixed effect on time, a significant interaction ($p < 0.001$) was found between the display types and operation types, with the estimated effect of using display_type_Monitor with operation_type_SpotterDirect being significantly different from the combined effects of using display_type-Monitor with operation_type_OperatorOnly and using display_type_HMD with operation_type_SpotterDirect. The correlation between the fixed effects was low to moderate, with a correlation of −0.527 between intercept and display_types_Monitor, and 0.334 between display types and operation types. As for the mixed effect on performance errors, a significant interaction ($p < 0.01$) was also found between the display types and operation types, with the estimated effect of using display_type_Monitor with operation_type_SpotterDirect being significantly different from the combined effects of using display_type_Monitor with operation_type_OperatorOnly and using display_type_HMD with operation_type_SpotterDirect. Residuals represent the variation in four dependent variables (time, error, collision, performance) that is not accounted for by the fixed effects (display_type, operation_type, and their interactions) or the random effects (variation across subjects). The fitted model of residuals shows the random scattering around the horizontal line at zero [Fig. 10], indicating that the model is properly accounting for the variation in the data.

To further investigate which independent variable has a greater effect on the performance metrics, average marginal effects were evaluated. As Table 8 shows, first of all, the time for task completion decreases by 20.6667 s when the display type is changed from HMD (reference level) to a monitor, which is a significant effect ($p < 0.001$); and the time for task completion increased by 11.8947 s when the operation type is changed from operator_only (reference level) to spotter_direct, which is a significant effect ($p < 0.1$). Changing display types has a greater average marginal effect on task completion time than changing operation types [Fig. 11a]. Second, the dumping error increased by 0.9649 when the display types are changed from HMD (reference level) to a monitor, which is a significant effect ($p < 0.05$); and errors decreased 3.6842 when the operation type is changed from operator_only (reference level) to spotter_direct, which is a significant effect ($p < 0.001$). Lastly, the number of collisions increased by 9.0439 when the display type is changed from HMD (reference level) to a monitor, which is a significant effect ($p < 0.001$); and the number of collisions decreased by 12.5702 when the operation type is changed from operator_only (reference level) to spotter_direct, which is a significant effect ($p < 0.001$). It was observed that changing operation types has a greater average marginal effect on task errors, collisions, and overall performance than changing display types [Fig. 11b, c, d].

The results of a GLMM supported the mixed effect of four combinations of display types and operation types on completion time, errors, collisions. First, results of average marginal effects demonstrated that task efficiency (i.e., time) is affected more by display types than task-related independent variables such as operation types. This finding is particularly important as it indicates that to assess the performance of a virtual excavation, when evaluating the completion time, display type should be carefully specified as it may cause higher average marginal effect on performance than other task related variables. To be noted, it was observed that time as a major performance metric is more sensitive to the display types than other performance metrics in this experiment. This finding also indicates that some attributes of visual interface (e.g., display types) may play as a major confounding factor when we simulate a close-to-real virtual excavation process. Second, results indicate that the task accuracy and safety in a virtual excavation are mainly affected by the task-related variables such as the operation types (e.g., independent vs. collaborative), and they are less sensitive to the changes of display types.

5.2.5. Overall performance taking account of accident-avoidance

As shown in Table 5, for the overall performance (P), when using a monitor, there is a significant difference ($p < 0.05$) between independent and collaborative trials as collaborative trials have higher mean performance score (1.15) [Table 6], which indicates that the performance was improved by collaborating with a spotter. The operation types also show a greater average marginal effect on the overall performance than the display types [Table 8] [Fig. 11d]. The performance metric integrating task completion time, accuracy and safety generally reflects the outcome and has the potential to measure the operator/trainee's performance under different task-related variables such as operation types. Furthermore, to adjust the weights of each performance attribute, this metric can be implemented under different work scenarios to emphasize different operation aspects, such as safety, productivity, or task skills. Assigning appropriate weights to different attributes also allows to benchmark similar studies and provide insights about the best practice.

5.3. User feedbacks and limitations

User feedbacks regarding the devices and overall experience were collected in the post experiment debrief. As for the operator, most participants reported a better sense of presence when using VR headset than using 2D monitor which limited the visual flexibility as needed, which is consistent with the experiment outcome. Nevertheless, some participants, especially female users, reported that they felt more comfortable of using 2D monitor due to a higher familiarity than using a wearable VR headset. These feedbacks are valuable indicators that to optimize the simulation system design there might be a necessity to conduct assessments on individual difference such as gender and work experience, which could be further investigated in future works. Another consideration is that when wearing VR headset for the first time in practice session, most participants tended to spend some time to visually explore the virtual environment instead of directly focusing on performing the task. This observation indicates that it is of importance to arrange practice session prior to the formal trials and allow participants to gain the familiarity with VR devices and reduce the distraction due to the excitement.

Table 7
Fixed effects on performance.

Fixed effects	time		error		collision	
	t value	Pr (> | t |)	t value	Pr (> | t |)	t value	Pr (> | t |)
(Intercept)	25.142	<2e-16 ***	4.679	1.57e-05 ***	7.40	3.44e-10 ***
display_types - M	1.914	0.0581	3.684	0.000355 ***	4.332	2.97e-05 ***
operation_types - S	5.644	1.26e-07 ***	−4.308	4.27e-05 ***	−3.736	0.000289 ***
display_types_M: operation_types_S	−7.188	7.82e-11 ***	−3.138	0.002173 **	−1.444	0.151618

Significant codes: < 0.001 '***' < 0.01'**' < 0.05 '*' < 0.1 '.'

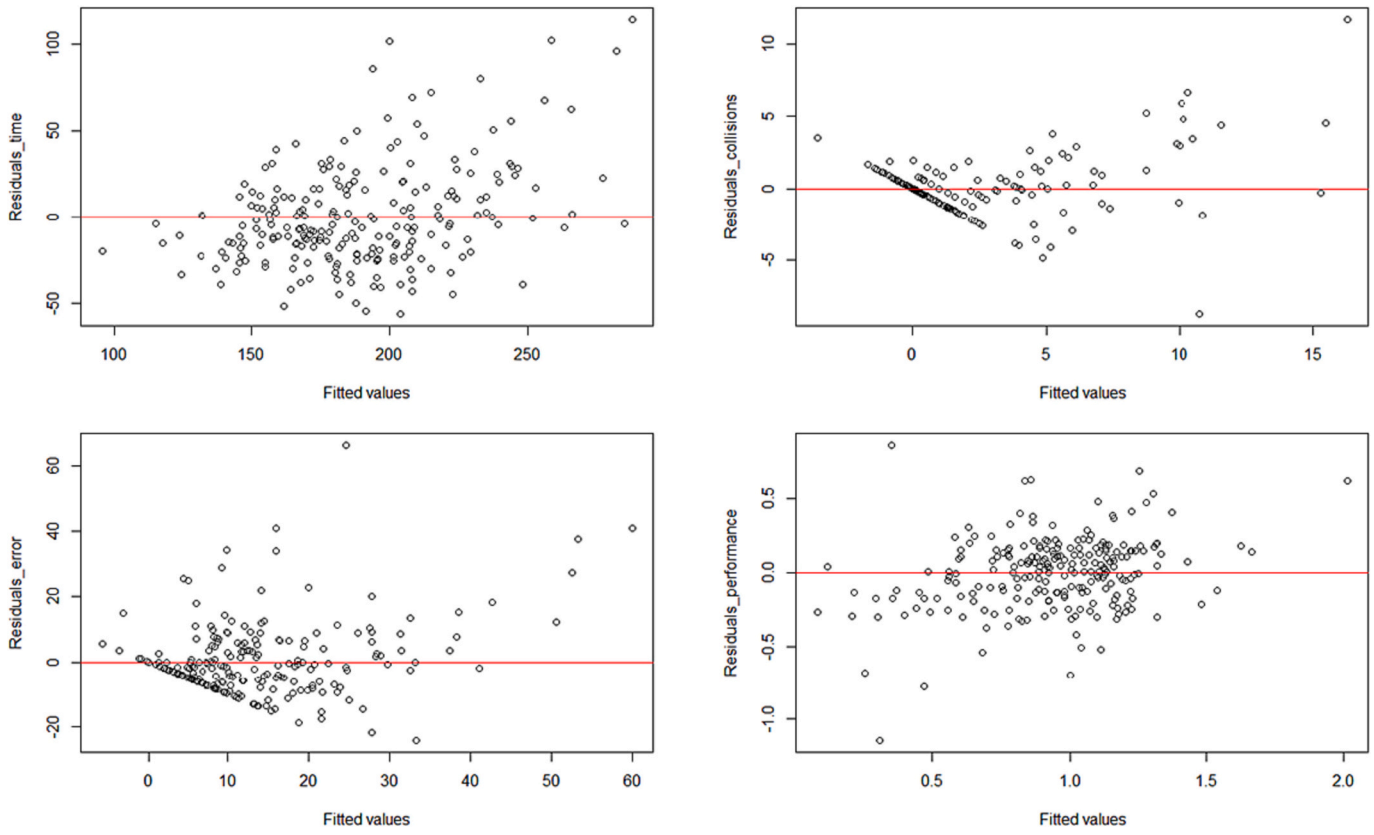


Fig. 10. Fitting residuals of the mixed effect model.

Table 8
Average Marginal effects of display types and operation types on performance.

	time		error		collision		performance	
Factor	AME	p	AME	p	AME	p	AME	p
display_types - M	−20.6667	0.0000 ***	0.9649	0.0217 *	9.0439	0.0000 ***	−0.0153	0.7218
operation_types - S	11.8947	0.0711 .	−3.6842	0.0000 ***	−12.5702	0.0000 ***	0.2105	0.0000 ***

Significant codes: < 0.001 '***' < 0.01'**' < 0.05 '*' < 0.1 '.'

Based on the proposed system, it could be possible for the spotter to examine the way to effectively communicate with the operator by testing different communication channels. The spotter provided the following feedback regarding the overall performance of the collaborative excavation. First, the non-immersive virtual environment with multiple viewports allowed a comprehensive understanding on the excavator, task space, and surrounding environment. Among four viewports, a duplicated view of the operator's FPV was especially helpful to understand excavation process in real-time and deliver signals accordingly. On the other hand, it is admitted that there is a likelihood that the non-immersive interface may lead to less realism for the spotter, and the visual interface design traded off the realism for the situational comprehensiveness. Second, although the motion capture technique was able to simulate real-time communication signals accurately, calibration every time before a new task trial is cumbersome. Third, the spotter participated in the work with a fixed standing position in an urban jobsite with limited workspace. Very often in a crowded urban jobsite, a spotter may constantly change the physical position to support the task. Hence it is admitted that fixed position decreased the spotter's mobility and flexibility of monitoring the excavation in a dynamic environment. In this experiment, however, maintaining a fixed position allowed to minimize the confounding effect caused by the spotter's mobility on the performance outcome.

Based on the post experiment comments, it would be helpful to evaluate the performance of spotter in future works by including other situational variables, such as changing spotter's physical mobility, in light of enhancing the overall realistic level. Also, there is a necessity to conduct an in-depth investigation on the trade-off effect between the level of immersion and comprehensiveness when selecting the interface for the spotter and how these attributes would affect the spotter's communication performance and the overall team performance. These findings are particularly valuable to guide future experiment design and improve the simulation system when assessing the performance of a non-operator personal (e.g., a spotter) in such collaborative operation.

6. Conclusions and future work

This study proposed a multi-user excavator simulation system composed of robotic control, motion capture, and hybrid-immersive interfaces. The within-group experiments were conducted to evaluate the system effectiveness by comparing it with the conventional single-user monitor-based simulator. We found that HMD-based simulator created a more close-to-real excavation experiment than monitor-based simulator, measured by the higher degree of immersion, sense of presence, and user interaction. Furthermore, the results of task performance including unsafe behaviors were analyzed, and a performance metric

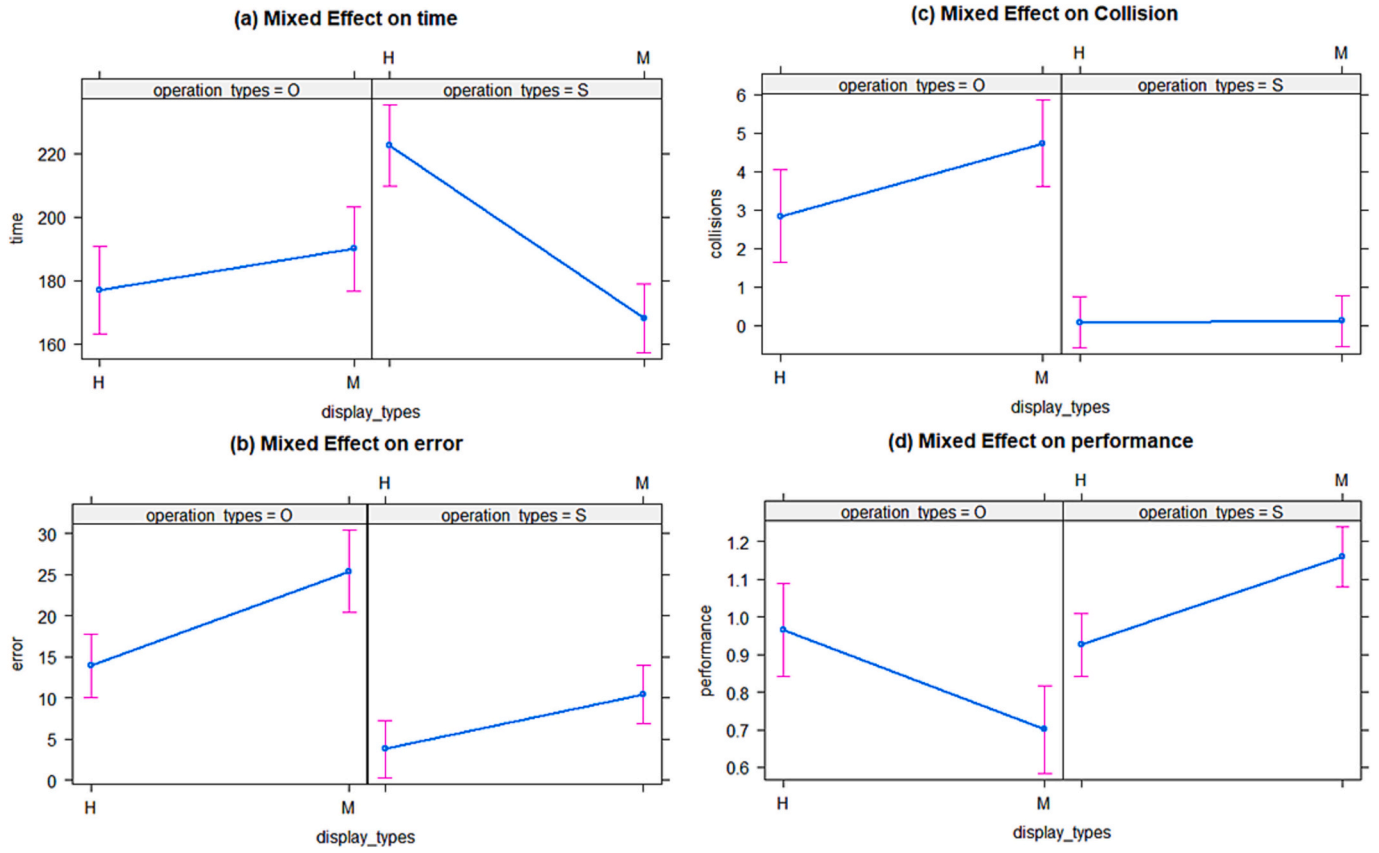


Fig. 11. Estimated changes of mean values of time, collision, error, and performance.

integrated with excavation productivity and accident-avoidance rate was used to measure trainee's skills taking account of their safety performance. Results from a GLMM demonstrate that the task accuracy and unsafe behaviors were affected by a mixed effect of display types and spotter's guidance, and primarily by the interaction with spotter. By collaborating with a spotter, we could observe improvements on reducing task errors and unsafe behaviors. On the other hand, the completion time is affected by a mixed effect of display types and interaction with the spotter, and display types cause a greater average marginal effect. Additionally, the overall performance integrated by task completion time, accuracy, and unsafe behaviors is analyzed to assess the operation outcome. The outcome can be a step forward in studying the human factors of the multi-user interaction with multi-roles in the human-machine teaming in jobsites. Moreover, the research outcome can inform the human teammate's capacity in terms of information acquisition and decision selection in a teleoperated teamwork process, which is of importance of reducing Out-Of-The-Loop (OOTL) performance error and augmenting worker-centered practices in human-machine teaming. Applications foreseen are not only in designing the highly immersive simulator for team-based training, but also in advancing human-centered study in the human-machine teaming in jobsites. The next-step work will be related to the improvement of investigating the human factors between the operator and the spotter with increasing complexity of the scene.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under Grant No.2026574. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] CGA White Paper, Common Ground Alliance. https://commongroundalliance.com/Portals/0/CGA%20White%20Paper%202019%20-%20FINAL.pdf?ver=2020-11-10-201648-153, 2019.
[2] C.-J. Liang, X. Wang, V.R. Kamat, C.C. Menassa, Human–robot collaboration in construction: classification and research trends, J. Constr. Eng. Manag. 147 (10) (2021), https://doi.org/10.1061/%28ASCE%29CO.1943-7862.0002154 (accessed August 11, 2021).
[3] X. Wang, C.-J. Liang, C.C. Menassa, V.R. Kamat, Interactive and immersive process-level digital twin for collaborative human–robot construction work, J. Comput. Civ. Eng. 35 (2021) 04021023, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000988.
[4] Y. Su, X. Chen, T. Zhou, C. Pretty, G. Chase, Mixed reality-integrated 3D/2D vision mapping for intuitive teleoperation of mobile manipulator, Robot. Comput. Integr. Manuf. 77 (2022), 102332, https://doi.org/10.1016/j.rcim.2022.102332.
[5] X. Tang, D. Zhao, H. Yamada, T. Ni, Haptic interaction in tele-operation control system of construction robot based on virtual reality, in: 2009 Int. Conf. Mechatron. Autom, 2009, pp. 78–83, https://doi.org/10.1109/ICMA.2009.5246470.
[6] J. Hollingsworth, Spotters: A Critical Element of Site Safety. https://safetymanagementgroup.com/spotters-a-critical-element-of-site-safety/, 2015 (accessed March 15, 2023).
[7] McKinsey Technology Trends Outlook 2022, McKinsey&Company, 2022. https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%

20digital/our%20insights/the%20top%20trends%20in%20tech%202022/mckinsey-tech-trends-outlook-2022-full-report.pdf.
[8] V. Getuli, P. Capone, A. Bruttini, Planning, management and administration of HS contents with BIM and VR in construction: an implementation protocol, Eng. Constr. Archit. Manag. 28 (2020) 603–623, https://doi.org/10.1108/ECAM-11-2019-0647.
[9] Y. Shi, J. Du, E. Ragan, K. Choi, S. Ma, Social influence on construction safety behaviors: A multi-user virtual reality experiment, in: Constr. Res. Congr. 2018, American Society of Civil Engineers, New Orleans, Louisiana, 2018, pp. 174–183, https://doi.org/10.1061/9780784481288.018.
[10] J. Du, Z. Zou, Y. Shi, D. Zhao, Zero latency: real-time synchronization of BIM data in virtual reality for collaborative decision-making, Autom. Constr. 85 (2018) 51–64, https://doi.org/10.1016/j.autcon.2017.10.009.
[11] J.M. Davila Delgado, L. Oyedele, P. Demian, T. Beach, A research agenda for augmented and virtual reality in architecture, engineering and construction, Adv. Eng. Inform. 45 (2020), 101122, https://doi.org/10.1016/j.aei.2020.101122.
[12] J.S. Lee, Y. Ham, H. Park, J. Kim, Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation, Autom. Constr. 135 (2022), 104119, https://doi.org/10.1016/j.autcon.2021.104119.
[13] J.S. Lee, Y. Ham, Exploring Human-Machine Interfaces for Teleoperation of Excavator, 2022, pp. 757–765, https://doi.org/10.1061/9780784483961.079.
[14] Y. Fang, J. Teizer, A Multi-User Virtual 3D Training Environment to Advance Collaboration among Crane Operator and Ground Personnel in Blind Lifts, 2014, pp. 2071–2078, https://doi.org/10.1061/9780784413616.257.
[15] H. Guo, H. Li, G. Chan, M. Skitmore, Using game technologies to improve the safety of construction plant operations, Accid. Anal. Prev. 48 (2012) 204–213, https://doi.org/10.1016/j.aap.2011.06.002.
[16] H. Voordijk, F. Vahdatikhaki, Virtual reality learning environments and technological mediation in construction practice, Eur. J. Eng. Educ. 47 (2022) 259–273, https://doi.org/10.1080/03043797.2020.1795085.
[17] H. Li, G. Chan, M. Skitmore, Multiuser virtual safety training system for tower crane dismantlement, J. Comput. Civ. Eng. 26 (2012) 638–647, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000170.
[18] B. Kim, C. Kim, H. Kim, Interactive modeler for construction equipment operation using augmented reality, J. Comput. Civ. Eng. 26 (2012) 331–341, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000137.
[19] F. Vahdatikhaki, A.K. Langroodi, L. Olde Scholtenhuis, A. Dorée, Feedback support system for training of excavator operators, Autom. Constr. 136 (2022), 104188, https://doi.org/10.1016/j.autcon.2022.104188.
[20] F. Vahdatikhaki, K. El Ammari, A.K. Langroodi, S. Miller, A. Hammad, A. Doree, Beyond data visualization: a context-realistic construction equipment training simulators, Autom. Constr. 106 (2019), 102853, https://doi.org/10.1016/j.autcon.2019.102853.
[21] X. Li, W. Yi, H.-L. Chi, X. Wang, A.P.C. Chan, A critical review of virtual and augmented reality (VR/AR) applications in construction safety, Autom. Constr. 86 (2018) 150–162, https://doi.org/10.1016/j.autcon.2017.11.003.
[22] T. Kassuba, C. Klinge, C. Hölig, B. Röder, H.R. Siebner, Vision holds a greater share in visuo-haptic object recognition than touch, NeuroImage. 65 (2013) 59–68, https://doi.org/10.1016/j.neuroimage.2012.09.054.
[23] C. Spence, C. Parise, Y.-C. Chen, The Colavita Visual Dominance Effect, CRC Press/Taylor & Francis, 2012. https://www.ncbi.nlm.nih.gov/books/NBK92851/ (accessed November 1, 2022).
[24] H. Mcgurk, J. Macdonald, Hearing lips and seeing voices, Nature. 264 (1976) 746–748, https://doi.org/10.1038/264746a0.
[25] S. Cole, E. Balcetis, Chapter Three - Motivated perception for self-regulation: How visual experience serves and is served by goals, in: B. Gawronski (Ed.), Adv. Exp. Soc. Psychol, Academic Press, 2021, pp. 129–186, https://doi.org/10.1016/bs.aesp.2021.04.003.
[26] G.R. Dirkin, Cognitive tunneling: use of visual information under stress, Percept. Mot. Skills 56 (1983) 191–198, https://doi.org/10.2466/pms.1983.56.1.191.
[27] H.A. Al-Jundi, E.Y. Tanbour, A framework for fidelity evaluation of immersive virtual reality systems, Virtual Reality (2022), https://doi.org/10.1007/s10055-021-00618-y.
[28] S. Martirosov, M. Bureš, T. Zítka, Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality, Virtual Reality 26 (2022) 15–32, https://doi.org/10.1007/s10055-021-00507-4.
[29] L. Kotek, Z. Tuma, K. Subrt, J. Kroupa, P. Blecha, J. Rozehnalova, R. Blecha, P. Heinrich, Testing human errors in virtual reality training, MM Sci. J. (2022) 6263–6268, https://doi.org/10.17973/MMSJ.2022_12_2022128.
[30] F. Morosi, G. Caruso, Configuring a VR simulator for the evaluation of advanced human–machine interfaces for hydraulic excavators, Virtual Reality 26 (2022) 801–816, https://doi.org/10.1007/s10055-021-00598-z.
[31] Z. Hong, Q. Zhang, X. Su, H. Zhang, Effect of virtual annotation on performance of construction equipment teleoperation under adverse visual conditions, Autom. Constr. 118 (2020), 103296, https://doi.org/10.1016/j.autcon.2020.103296.
[32] M. Wallmyr, T.A. Sitompul, T. Holstein, R. Lindell, Evaluating mixed reality notifications to support excavator operator awareness, in: D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, P. Zaphiris (Eds.), Hum.-Comput. Interact. – INTERACT 2019, Springer International Publishing, Cham, 2019, pp. 743–762, https://doi.org/10.1007/978-3-030-29381-9_44.
[33] Y. Shi, J. Du, E. Ragan, Review visual attention and spatial memory in building inspection: toward a cognition-driven information system, Adv. Eng. Inform. 44 (2020), 101061, https://doi.org/10.1016/j.aei.2020.101061.
[34] M. Kamezaki, J. Yang, R. Sato, H. Iwata, S. Sugano, A situational understanding enhancer based on augmented visual prompts for teleoperation using a multi-monitor system, Autom. Constr. 131 (2021), 103893, https://doi.org/10.1016/j.autcon.2021.103893.
[35] M. Kamezaki, J. Yang, H. Iwata, S. Sugano, Visibility enhancement using autonomous multicamera controls with situational role assignment for teleoperated work machines, J. Field Robot. 33 (2016) 802–824, https://doi.org/10.1002/rob.21580.
[36] M. Peterson, Learning interaction in an avatar-based virtual environment: a preliminary study, PacCALL J. 1 (2005) 29–40.
[37] J. Wen, M. Gheisari, Using virtual reality to facilitate communication in the AEC domain: a systematic review, Constr. Innov. 20 (2020) 509–542, https://doi.org/10.1108/CI-11-2019-0122.
[38] R. Eiris, M. Gheisari, Research trends of virtual human applications in architecture, engineering and construction, J. Inf. Technol. Constr. ITcon. 22 (2017) 168–184. https://www.itcon.org/2017/9.
[39] S. Khan, B. Tunçer, Gesture and speech elicitation for 3D CAD modeling in conceptual design, Autom. Constr. 106 (2019), 102847, https://doi.org/10.1016/j.autcon.2019.102847.
[40] A.M. Malkawi, R.S. Srinivasan, A new paradigm for human-building interaction: the use of CFD and augmented reality, Autom. Constr. 14 (2005) 71–84, https://doi.org/10.1016/j.autcon.2004.08.001.
[41] X. Wang, Z. Zhu, Vision-based hand signal recognition in construction: a feasibility study, Autom. Constr. 125 (2021), 103625, https://doi.org/10.1016/j.autcon.2021.103625.
[42] M. Yahya, J.A. Shah, K.A. Kadir, Z.M. Yusof, S. Khan, A. Warsi, Motion capture sensing techniques used in human upper limb motion: a review, Sens. Rev. 39 (2019) 504–511, https://doi.org/10.1108/SR-10-2018-0270.
[43] W. Chang, L. Dai, S. Sheng, J. Too Chuan Tan, C. Zhu, F. Duan, A hierarchical hand motions recognition method based on IMU and sEMG sensors, in: 2015 IEEE Int. Conf. Robot. Biomim. ROBIO, 2015, pp. 1024–1029, https://doi.org/10.1109/ROBIO.2015.7418906.
[44] Y. Ye, Y. Shi, Y. Lee, G. Burks, D. Srinivasan, J. Du, Exoskeleton Training through Haptic Sensation Transfer in Immersive Virtual Environment, 2022, pp. 560–569, https://doi.org/10.1061/9780784483961.059.
[45] T. Stranick, C. Lopez, Adaptive virtual reality exergame: promoting physical activity among workers, J. Comput. Inf. Sci. Eng. 22 (2021), https://doi.org/10.1115/1.4053002.
[46] Y. Shi, J. Du, C.R. Ahn, E. Ragan, Impact assessment of reinforced learning methods on construction workers' fall risk behavior using virtual reality, Autom. Constr. 104 (2019) 197–214, https://doi.org/10.1016/j.autcon.2019.04.015.
[47] V.M. Manghisi, A.E. Uva, M. Fiorentino, M. Gattullo, A. Boccaccio, A. Evangelista, Automatic ergonomic postural risk monitoring on the factory shopfloor –the ergosentinel tool, Procedia Manuf. 42 (2020) 97–103, https://doi.org/10.1016/j.promfg.2020.02.091.
[48] M. Kurien, M.-K. Kim, M. Kopsida, I. Brilakis, Real-time simulation of construction workers using combined human body and hand tracking for robotic construction worker system, Autom. Constr. 86 (2018) 125–137, https://doi.org/10.1016/j.autcon.2017.11.005.
[49] X. Wang, Z. Zhu, Vision–based framework for automatic interpretation of construction workers' hand gestures, Autom. Constr. 130 (2021), 103872, https://doi.org/10.1016/j.autcon.2021.103872.
[50] 14:00-17:00, ISO 10968, ISO (n.d.), https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/11/31188.html, 2004 (accessed June 8, 2022).
[51] D. Liu, Y. Ham, J. Kim, H. Park, Towards a collaborative future in construction robotics: A human-centered study in a multi-user immersive operation and communication system for excavation, in: Proc. 1st Future Constr. Workshop Int. Conf. Robot. Autom. ICRA 2022, International Association for Automation and Robotics in Construction (IAARC), 2022, https://doi.org/10.22260/ICRA2022/0016.
[52] B.G. Witmer, M.J. Singer, Measuring presence in virtual environments: a presence questionnaire, Presence Teleoperators Virtual Environ. 7 (1998) 225–240, https://doi.org/10.1162/105474698565686.
[53] V. Schwind, P. Knierim, N. Haas, N. Henze, Using Presence Questionnaires in Virtual Reality, CHI, 2019, https://doi.org/10.1145/3290605.3300590.
[54] X. Su, P.S. Dunston, R.W. Proctor, X. Wang, Influence of training schedule on development of perceptual–motor control skills for construction equipment operators in a virtual training system, Autom. Constr. 35 (2013) 439–447, https://doi.org/10.1016/j.autcon.2013.05.029.
[55] R.L. Keeney, H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Trade-Offs, Cambridge University Press, Cambridge, 1993, https://doi.org/10.1017/CBO9781139174084.
[56] A.F. Zuur, E.N. Ieno, N. Walker, A.A. Saveliev, G.M. Smith, Mixed effects models and extensions in ecology with R, Springer, New York, NY, 2009, https://doi.org/10.1007/978-0-387-87458-6.