Automation in Construction 172 (2025) 106053

Contents lists available at ScienceDirect

Automation in Construction

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

Comparing dynamic viewpoint control techniques for teleoperated robotic welding in construction

Sungboo Yoon [a], Moonseo Park [a], Changbum R. Ahn [b,*]

[a] Department of Architecture and Architectural Engineering, Seoul National University, Seoul 08826, Republic of Korea
[b] Department of Architecture and Architectural Engineering, Institute of Construction and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea

ARTICLE INFO

Keywords:
Teleoperation
Human-robot interface
Dynamic viewpoint control
Robotic welding
Virtual reality

ABSTRACT

Dynamic viewpoints offer effective visual feedback in teleoperation for construction, where tasks often require precise manipulation during frequent viewpoint adjustments. However, the comparative performance of various dynamic viewpoint control techniques remains unclear. This paper investigates the impact of dynamic viewpoint control techniques on task performance and user experience during teleoperation in construction. A user study was conducted in a remote welding-at-height scenario with 20 participants, including experienced welders and university students, to compare five techniques: (1) coupled vision-motion, (2) decoupled vision-motion with hand or head motion-based control, and (3) hybrid vision-motion with manual or automatic switching. Results showed that decoupled vision-motion with head motion-based control outperformed other techniques in task efficiency and user preference. Hybrid vision-motion with manual switching was more effective than decoupled vision-motion in contexts involving occlusions, reducing physical demand and enhancing welding quality. Based on these findings, guidelines are proposed for viewpoint control in teleoperated construction robots.

1. Introduction

Teleoperation has become integral to a wide range of construction applications [1–4], offering advantages such as adaptability to dynamic conditions by combining human and robotic capabilities [5] and relatively simple technological setup requirements [6]. Previous work has focused on enhancing teleoperation systems for construction robots by improving telerobotic control systems [7], developing methods for compensating for latency [8], and designing effective visual feedback interfaces [1,2]. In particular, researchers have identified visual feedback as a major factor affecting a user's situational awareness, which is critical for safety and productivity during teleoperation in construction [9–13].

Among visual feedback methods for supporting teleoperation, we focus on dynamic viewpoint control. This approach plays a crucial role in supporting teleoperation, as it determines how users view the remote workspace, where the viewpoint should be located, and how it should change during tasks. Coupled vision-motion links the user's viewpoint to the robot's movements, offering intuitive control and improved hand-eye coordination. However, it can lead to a limited field of view and increased user fatigue due to the “soda straw” effect, where the narrow viewing angle makes it challenging to understand the broader environment [14]. In contrast, decoupled vision-motion allows independent control of the camera and the robot, providing a more comprehensive view of the workspace and reducing cognitive load from constant adjustments. Yet, this approach can complicate control, as simultaneously managing both the camera and the manipulator can be demanding, especially for novice users. Hybrid vision-motion seeks to balance these trade-offs by allowing users to switch between coupled and decoupled control techniques. This flexibility provides opportunities in construction tasks that include various phases, potentially enabling users to leverage the advantages of both techniques depending on the task phase and specific needs.

Although the aforementioned viewpoint control techniques are deemed promising for teleoperation, selecting the appropriate viewpoint control technique for supporting construction tasks is not trivial. Previous work suggests both the strengths and weaknesses of each dynamic viewpoint control technique on telerobotic setups with short-duration and simple tasks, such as pick-and-place [15,16] or end-effector positioning [10], but there is still limited understanding of their comparative performance in construction settings. Construction tasks introduce unique challenges to viewpoint control due to their need

* Corresponding author.
E-mail addresses: yoonsb24@snu.ac.kr (S. Yoon), mspark@snu.ac.kr (M. Park), cbahn@snu.ac.kr (C.R. Ahn).

https://doi.org/10.1016/j.autcon.2025.106053
Received 27 August 2024; Received in revised form 31 January 2025; Accepted 7 February 2025
Available online 16 February 2025
0926-5805/

for high levels of hand-eye coordination and fine manipulation [17], issues that become more salient when applied to complex and skill-intensive tasks such as pipe-fitting [2], brick-laying [18], or welding [19]. Welding, in particular, involves frequent adjustments in viewpoint to maintain optimal visibility of the weld pool and surrounding areas including the wire tip [20–22]. Moreover, welding requires precise manipulation of the welding torch while adapting to changing viewpoints, which plays a crucial role in determining the quality of the weld [6]. Therefore, obtaining systematic knowledge of the comparative performance of dynamic viewpoint control techniques during telerobotic welding can not only help identify best practices for viewpoint control in remote welding, but also offer insights into addressing the broader challenges of teleoperation in construction, which require precise manipulation during frequent viewpoint adjustments.

In this work, we investigate the impact of dynamic viewpoint control techniques on task performance and user experience during teleoperation in construction. We designed an application scenario in a simulated environment using virtual reality (VR), which involves a robotic welding system at height on a construction site (Fig. 1A) and a worker equipped with a head-mounted display (HMD) operating the system from a remote location (Fig. 1B). The system includes a mobile lift and two robotic arms controlled by a single worker. The setup features a dynamic viewpoint with a second “camera-in-hand” robot arm (hereafter, camera arm) behind the “manipulation” robot arm (hereafter, manipulation arm), which is a similar physical setup to those used in previous studies on dynamic viewpoint using robotic arms [12,23]. The manipulation arm is equipped with a gas metal arc welding (GMAW) tool, commonly known as a metal inert gas (MIG) welding tool (Fig. 1C), and the camera arm is equipped with an RGB-D camera.

This study presents the results of our user study with 20 participants, including experienced welders ($N = 10$) and university students ($N = 10$), comparing five viewpoint control techniques: (1) coupled vision-motion, (2) decoupled vision-motion with hand motion-based viewpoint control, (3) decoupled vision-motion with head motion-based viewpoint control, (4) hybrid vision-motion with manual switching, and (5) hybrid vision-motion with automatic switching. These methods were compared based on task completion time, weld coverage ratio, path deviation, trajectory length, number of collisions, perceived cognitive load, perceived usability, and user preference.

There are four main contributions of this study. First, we provide an empirical evaluation to compare dynamic viewpoint control techniques specifically for teleoperation in construction, focusing on welding as an application scenario. Through both quantitative and qualitative analysis, we closely examine task performance and user experience during teleoperation with dynamic viewpoint control. Second, we present a technical demonstration of dynamic viewpoint control techniques for a telerobotic welding system, including the hybrid vision-motion. To our knowledge, this is among the first studies to explore hybrid vision-motion for viewpoint control in construction. Third, through a user study including both novice and expert users, we identify that differences in user preferences for viewpoint control techniques vary depending on the level of expertise in welding. Fourth, based on the results, we discuss practical implications for viewpoint control for teleoperated systems in construction.

2. Related works

Our work builds on previous work on teleoperation in construction, viewpoint control techniques for teleoperation, the opportunities and challenges of hybrid vision-motion, and telerobotic welding technologies in construction.

2.1. Teleoperation in construction

Teleoperation, the remote control of robots or machines by human operators [24], has been widely adopted in construction due to its ability to mitigate human risks in high-risk scenarios [24], such as demolishing structures [7] and working at heights [25]. By combining human adaptability with robotic precision, teleoperation is particularly advantageous in the dynamic and often unpredictable conditions of construction, where tasks frequently require on-the-fly adjustments [5,26]. Owing to these benefits, teleoperation has become a primary control technique in a wide range of construction applications, including excavators [1], cranes [4], and industrial manipulators or collaborative robots [2,3]. In this study, we focus on remote manipulation, also referred to as telemanipulation, tasks in construction.

However, implementing telemanipulation in construction presents several challenges. Challenges identified in previous work include time delays (latency) between control input and system output [27], variability in operator proficiency [9], and ensuring operator safety from potential collisions [28]. Moreover, due to the unstructured and dynamic nature of the work environment, providing environmental awareness without imposing cognitive load is particularly challenging for teleoperation systems in construction [9]. To address these challenges, previous work has presented various enhancements, such as improved control systems [7], latency compensation techniques [8], and more effective interface designs [2,3,29–31].

While these studies have contributed to the significant enhancement of the overall performance and user experience during telemanipulation for construction tasks, there has been relatively less emphasis on the user's dynamic viewing strategy in the construction domain. Although the user's perspective has been shown to significantly impact situational awareness and task performance [11], understanding of how users should view the remote workspace, where the viewpoint should be located, and to what extent the viewpoint should change during tasks has yet been limited to static viewing strategy, where the positional

Fig. 1. Application scenario in this study. (A) Robotic welding system with a mobile lift and two robotic arms (the camera arm and the manipulation arm), controlled by a single worker. (B) A welding worker with an HMD teleoperates the system from a remote location. (C) Operator's view through the HMD. The worker teleoperates the manipulation arm with an MIG-welding gun and is provided with a remote view from the camera arm with an RGB-D camera.

movement of the viewpoint remains relatively constrained [1,28,32]. One notable example is Kamezaki et al. [33], in which the authors present an autonomous multi-flying camera system (FCS) for supporting 4-degree-of-freedom (DoF) excavator during locomotion and manipulation tasks in a simulated disaster response site. The autonomous FCS ensures consistent image projection by avoiding frequent changes in capturing parameters and provides complementary images that reduces the need for mental rotation and translation. Our work builds upon their work by implementing higher DoF robots for both camera control and fine manipulation, utilizing a wider range of state-of-the-art dynamic viewpoint control techniques leveraging immersive technologies.

2.2. Viewpoint control techniques for teleoperation

Visual support is critical for the success of teleoperation, as it directly impacts situational awareness [13]. This is why the techniques to provide high-quality views of the remote workspace have long been a focus of study in the field of robotics, particularly in robot teleoperation. While multiple static cameras are widely applied for teleoperation of excavators [33] and cranes [4] in construction, this approach requires users to integrate information from various perspectives, which can increase cognitive load and task complexity [12]. Additionally, setting up multiple static cameras is often impractical due to variations in task and environments, and may not be feasible during robot deployment [13]. To address these challenges, previous studies have explored dynamic viewpoint techniques using external cameras on unmanned aerial vehicles (UAVs) [13,33] or robotic arms [12,15,16,23,34,35]. These techniques can be categorized into two main approaches: (1) decoupled vision-motion and (2) coupled vision-motion.

Decoupled vision-motion allows users to observe the robot's actions (motion) from a separate viewpoint (vision), independent of any tethering to the robot [11]. Manually controlling a decoupled camera has been widely used in scenarios such as the teleoperation of spacecraft docking [36], pick-and-place tasks [15,16], and UAV-UGV teams [37]. However, in many telerobotic applications, manually operating a dynamic camera increases operator workload and reduces teleoperation performance [13]. Moreover, managing both an external camera system and a manipulator can be confusing for novice users and may require substantial training for effective use. An alternative to manual control is to use head motion for dynamic viewpoint control, employing head tracking system based on vision sensors [38] or HMDs [39] to allow hands-free camera operation. However, head rotation relative to the display (i.e., view rotation) can increase cognitive load, especially when viewed on a screen [10,40].

In contrast, coupled vision-motion directly links the user's viewpoint (vision) to the robot's movement (motion), allowing users to view the scene through the "eyes" of the robot [11]. Viewpoint cameras can be attached and coupled to various parts of a robot, such as the elbow joint, wrist joint, and end-effector [41]. While this common coupled vision-motion approach simplifies coordination between viewpoint and control, the field of view is often restricted, which can lead to decreased task performance and user fatigue. This is widely known as the "soda straw" or "keyhole" effect, in which operators struggle to understand and comprehend the remote environment due to a limited viewing angle [14]. Moreover, controlling the foreign viewpoint of the robot, such as cameras attached to elbows or hands is counterintuitive for humans accustomed to the natural viewpoint control motions of human eyes, potentially causing frustration [41]. Orbital cameras are a common yet effective variant in the literature [10,42]. Kuitert et al. [10] proposed orbital head-mounted display that utilizes head rotation tracking in an HMD to command camera movement in azimuth and elevation directions around a fixation point located at a robot's end-effector, outperforming the standard HMD interface in perceived usefulness.

Recent research has introduced viewpoint adjustment methods that optimize viewpoints using a set of heuristics [12,13,15,23,33–35]. These methods have shown promise in providing effective viewpoints that avoid occlusions in cluttered remote workspaces and prevent potential collisions with the environment. However, in construction settings, the unstructured and fluctuating nature of tasks requires frequent workspace exploration, making these methods less effective when the camera viewpoint is constrained to the robot's end-effector throughout various task phases (e.g., mobile navigation, lift operation, and manipulation).

While previous work has extensively explored decoupled and coupled vision-motion techniques for effective dynamic viewpoint control, these studies have focused on short, simple tasks such as pick-and-place [15,16] or end-effector positioning [10], which do not require extensive manipulation skills. Moreover, previous studies rely on fixed-position setups, such as those in factory environments, where robotic systems do not need to navigate spatially dispersed target workspaces or workpieces [43]. We extend prior work by presenting the results of an experiment involving prolonged construction tasks with large, spatially dispersed workpieces, which not only require robot base mobility but also advanced skill, precision, and hand-eye coordination, particularly during frequent changes in viewpoint [17].

2.3. Opportunities and challenges of hybrid vision-motion in construction

Hybrid vision-motion aim to provide flexible viewpoint control by allowing transitions between the coupled and decoupled viewpoint control techniques. This hybrid approach differs from multi-user [1,44] or multi-view [45,46] teleoperation, which allows the user to switch between different views, in that it integrates both coupled and decoupled vision-motion within a single control system for a viewpoint camera. The individual benefits of coupled and decoupled vision-motion suggest that this hybrid approach can be effective for construction tasks involving multiple phases. For example, during navigation, a decoupled view that allows free exploration offers a comprehensive overview, aiding in obstacle avoidance and efficient path planning [47]. Fine manipulation tasks, such as welding, benefit from a coupled viewpoint, which allows simplified hand-eye coordination and provides a focused perspective directly on the task at hand [41]. However, despite these opportunities, the integration of coupled and decoupled vision-motion remains underexplored, and it remains relatively unclear whether a hybrid approach can leverage the advantages of both techniques to enhance teleoperation performance in complex, multi-phase construction tasks.

Moreover, hybrid vision-motion for construction teleoperation comes with a unique set of challenges. One primary challenge is determining where the viewpoint should be located when it is either coupled or decoupled. The optimal camera placement can vary significantly based on the context of the task [23] and human preferences [15]. Another challenge is determining the level of robot autonomy for viewpoint switching. In other words, it must be decided whether the transition between viewpoints should be initiated by a human or a robot. Previous studies [34,48] indicate that both manual and automated transitions have their own sets of advantages and challenges. Valiton et al. [34] observed that autonomous camera control significantly improved supervisory task performance and reduced workload but presented decreased operator awareness of task status and trust in the autonomy. However, this area remains an open question [49], suggesting further investigation to identify the appropriate autonomy for viewpoint transitions during construction tasks. Finally, if viewpoint transitions are to be handled by a robot, identifying the appropriate timing is critical. Transitions must occur at points that do not interfere with the operator's workflow and maintain situational awareness and precision. All of these factors impact the effectiveness of the hybrid vision-motion technique, which in turn affects task performance and user experience during teleoperation in construction.

2.4. Telerobotic welding in construction

Welding is commonly used in construction to fuse various metal components for applications such as on-site fabrication of metal frameworks [43] and the assembly of pre-fabricated components in off-site construction [19]. Traditionally, welding has relied on manual methods that require skilled labor in hazardous environments with exposure to high temperatures, fumes, and dust [50]. In comparison to manual welding, robotic welding offers more consistent quality [51,52], improves the efficiency of continuous and repetitive tasks [53], and reduces safety risks for workers [21,54].

The current control modes for robotic welding include teaching playback, offline programming, fully autonomous intelligent modes, and teleoperation [6,50]. Given the inherently complex, nonlinear, and unpredictable nature of welding, telerobotic welding systems offer distinct advantages, such as adaptability to complex environments and comparatively simple technological setup [6]. However, a key challenge in telerobotic welding in construction is that the operator must precisely control the welding torch maintaining a fixed angle and speed [20,21], continuously monitor the weld pool [24], and make real-time adjustments to robotic movements [6], while relying heavily on visual feedback from remote workspaces where visibility is often obstructed and spatially constrained.

To address these challenges, our work focuses on dynamic viewpoint system to support visual feedback for operators during telerobotic welding in construction. Previous studies employed decoupled viewpoints leveraging virtual replicas or reconstructed scenes for tasks such as gas tungsten arc welding (GTAW) on a V-grooved workpiece [55] and gas metal arc welding (GMAW) on a steel plate [54]. Some studies employed coupled viewpoints using an end-effector camera for laser welding on a steel plate [6] or an orbital camera for GTAW of a cooling pipe within the ITER nuclear fusion reactor [10]. While these studies have provided valuable advancements to diverse telerobotic welding systems, they mostly represent individual evaluation of applied viewpoint control techniques, with few studies directly comparing coupled and decoupled techniques for welding. Moreover, the current understanding of these techniques is still largely informed by the results of user studies involving only novice users. As expertise level has a significant impact on the performance of teleoperated tasks in various target domains [56–58], we argue that an evaluation should include task performance and user experience of experienced welders, which are potential end-users of telerobotic systems.

In summary, we contribute to prior work by presenting experimental results comparing state-of-the-art dynamic viewpoint control techniques in telerobotic welding for construction tasks, evaluating task performance, user experience, and expertise-dependent preferences, while introducing hybrid vision-motion and investigating appropriate autonomy for switching between coupled and decoupled techniques.

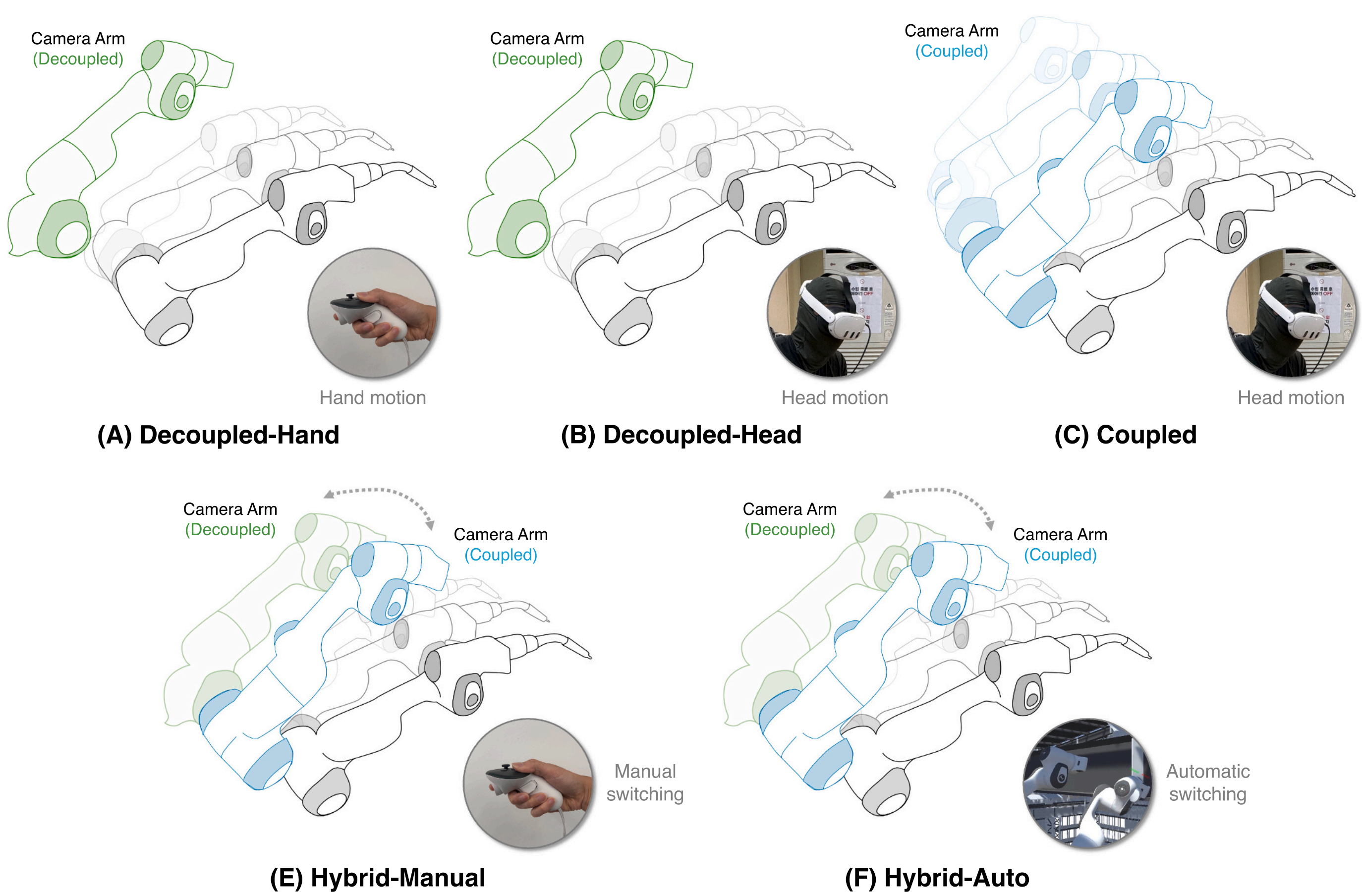

Fig. 2. Viewpoint control techniques compared in this study. The manipulation arm (grey) moves from left to right from the user's perspective. (A, B) Decoupled vision-motion: the camera arm (green) moves independently, controlled by hand (A) or head (B) motion. (C) Coupled vision-motion: the camera arm (blue) is linked to the manipulation arm, with the viewpoint directed by the user's head towards the welding gun tip. (E, F) Hybrid vision-motion: users can switch between decoupled (green) and coupled (blue) control techniques, manually (E) or automatically (F). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Viewpoint control techniques: Concepts and our implementations

In this work, we compare three viewpoint control approaches: (1) decoupled vision-motion (D), (2) coupled vision-motion (C), and (3) hybrid vision-motion (H). Additionally, we subdivide the decoupled vision-motion into two conditions based on the method of viewpoint control, namely hand motion-based (D-Hand) and head motion-based (D-Head). We also subdivide the hybrid vision-motion into two conditions based on the level of robot autonomy for switching, namely manual switching (H-Manual) and automatic switching (H-Auto). Video demonstrations of each viewpoint control technique are available at https://youtu.be/L5rUZPR6Q5s. This section describes our concepts and implementations of the resulting five viewpoint control techniques (Fig. 2). We also describe the implementation of teleoperation for the manipulation arm and mobile lift, both of which are applied across all conditions.

3.1. Decoupled vision-motion

Decoupled vision-motion makes the positional and orientational (6-DoF) movements of the camera arm's end-effector completely independent from that of the manipulation arm's end-effector. To control the camera arm's end-effector, which is a viewpoint camera shown to the user's HMD, the user can: (1) use hand motion to change the pose goal (Hand motion-based viewpoint control, Fig. 2A) or (2) use head motion to change the pose goal (Head motion-based viewpoint control, Fig. 2B). For hand motion-based control (D-Hand), we use a left-hand VR controller to track the 6-DoF movements of the hand in our implementation. Additionally, we implement the "clutching" technique proposed by Praveena et al. [59]. This technique allows the user to disengage control and move the controller independently of the robot (similar to lifting a finger and repositioning it on a trackpad). In other words, the relative 6-DoF movements of the controller are used rather than the absolute position. This approach ensures seamless viewpoint movement without abrupt positional changes. In our implementation, we use the grip button of the left-hand VR controller for clutching. For head motion-based control (D-Head), we use motion tracking sensors in the VR headset to track the 6-DoF movements of the user's head. The position and orientation of the user's head are used as the pose goal of the camera arm's end-effector.

3.2. Coupled vision-motion

Coupled vision-motion links the positional and orientational (6-DoF) movements of the camera arm's end-effector directly to those of the manipulation arm's end-effector (Fig. 2C). While the most common implementation of coupled vision-motion is the end-effector camera, it is widely recognized in previous studies [12,14] that the end-effector camera has limitations in complex teleoperation tasks and critical real-world scenarios. Thus, our implementation of coupled vision-motion follows the orbital head-mounted display proposed by Kuitert et al. [10], as described in Section 2.2. Using motion tracking sensors in a VR headset, the user controls the viewpoint, which is the pose goal of the camera arm's end-effector, with their head rotation (3-DoF). The viewpoint is always directed towards the fixation point, which in our implementation is set to the tip of the welding gun. Additionally, the user can move the viewpoint along the view direction using the thumbstick of the right-hand VR controller (up and down), creating the sensation that the camera zooms in and out. Based on our pilot testing with 10 participants [60], we adopted an initial viewpoint offset of 0.7 m from the fixation point.

3.3. Hybrid vision-motion

Hybrid vision-motion allows the user to switch between the decoupled and coupled control techniques. In our implementation, we integrated decoupled viewpoint with head motion-based control and coupled viewpoint with viewpoint adjustment using heuristic optimization. The choice of methods for each viewpoint control technique was motivated by prior work [13,34,39] and pilot testing with 10 participants [60]. Transitions between decoupled and coupled viewpoints are initiated either by the user (H-Manual, Fig. 2E) or by the robot (H-Auto, Fig. 2F). In H-Manual, the user initiates the transition by toggling the grab button on the left-hand VR controller. In H-Auto, we applied a velocity-based phase judgement method commonly used in prior work in telemanipulation [61], to determine the timing for the transition. Specifically, when the velocity of the manipulation arm end-effector remains low for a specific duration, the system assumes the user begins finer manipulation (e.g., main welding or precisely positioning an initial weld point) and switches to a coupled vision-motion. Conversely, if the velocity of the end-effector exceeds a predefined threshold, the system assumes the user is performing broader manipulation or exploration, and switches back to a decoupled vision-motion. Based on our pilot testing with two participants, we set the duration threshold to two seconds.

When transitioned to motion-coupled viewpoint, the goal pose of the camera arm is calculated using a weighted-sum nonlinear optimization, which is a common implementation in prior work [12,13,23,34].

$$\underset{\mathbf{g}}{\operatorname{argmin}} \sum_i w_i f_i(\mathbf{g}) \tag{1}$$

where $f_i(\mathbf{g})$ are weighted objective terms with corresponding weights w_i and $\mathbf{g}$ is the goal pose vector consisted of (x, y, z, θ), which are position and orientation variables. Let $\mathbf{c} = (x_c, y_c, z_c, \theta_c)$ denote the camera arm's current pose. The θ of the camera arm's goal pose $\mathbf{g}$ is always the same as the θ_c of the camera arm's current pose:

$$\theta = \theta_c \tag{2}$$

In our implementation, we adopted three of the collision, view, and motion-limiting objectives from Senft et al. [13]. First, the camera arm should maintain a distance from the manipulation arm end-effector that is greater than a predefined distance d_{min}.

$$f_1(\mathbf{g}) = g_\sigma^P\left(d_{g-ee}(\mathbf{g}) - d_{min}\right)_+ \tag{3}$$

where $d_{g-ee}(\mathbf{g})$ is the distance between the camera arm's goal position and manipulation arm end-effector and $g_\sigma^P(x)$ is the higher-order Gaussian with order p and a standard deviation of σ^2: $g_\sigma^P(x) = e^{-\left(x^2/2\sigma^2\right)^P}$, as per its definition by Senft et al. [13]. Next, the camera should be centered on the manipulation arm end-effector.

$$f_2(\mathbf{g}) = \left(sin\theta_{c-ee}(\mathbf{g})\right)^2 \tag{4}$$

where $\theta_{c-ee}(\mathbf{g})$ is the angle between the direction of camera arm's current pose and the vector from the camera to the manipulation arm end-effector. Lastly, the new goal position of the camera arm should be close to its current position.

$$f_3(\mathbf{g}) = \left(d_{c-g}(\mathbf{g})\right)^6 \tag{5}$$

where $d_{c-g}(\mathbf{g})$ is the distance between the new goal position of the

Table 1
Constant values used in optimization (adopted from Senft et al. [13]).

Variable		Value
Optimization weights	w_1	50
	w_2	10
	w_3	1000
Optimization constraints	p	3
	σ	0.1 m
	d_{min}	0.2 m

camera arm and its current position. Table 1 lists the constant values for weights and constraints used in the optimization, adopted from Senft et al. [13] and our pilot testing with two participants.

3.4. Teleoperation for manipulation arm

For teleoperation, a direct mapping technique was applied to align the VR controller's position and orientation with the manipulation arm's end effector configuration, providing direct control over the robot [62]. The orientation of the VR controller and its index trigger were aligned with those of a real MIG-welding gun, as shown in Fig. 3A. The desired joint positions were determined through a nonlinear inverse kinematics optimization to match the desired end-effector pose while limiting joint velocity and adhering to joint limits [13]. During the welding operation, participants fed the wire by pressing the right-hand VR controller's trigger, similar to pressing the trigger of a MIG-welding gun. Here, we assumed that the MIG-welding parameters, such as wire size, voltage, and wire feed speed, were pre-configured for optimal performance.

To assist users during telemanipulation, we implemented three assistive techniques: (1) constrained motion for welding, (2) haptic feedback for collision alarms and operational constraints, and (3) virtual visual cues for vision-motion coupling information. First, we applied constrained motion by imposing constraints on the angle of the welding gun to ensure welding quality. When the angle of the welding gun deviates beyond a certain threshold compared to the start of welding, the telemanipulation system adjusts the controller input to maintain the threshold angle. Based on prior work [20,63], we conservatively set the threshold at 10 degrees.

Furthermore, we incorporate haptic and virtual visual cues for assisting telemanipulation. Haptic feedback was provided through VR controllers, with vibrations alerting users to potential collisions, proximity to the workpiece, or commands exceeding each arm's reach. The intensity, interval, and timing of the vibrations were set based on pilot testing with two participants. Virtual visual cues provided real-time graphical feedback on welding status, including vision-motion mode (displayed as "Coupled" or "Decoupled") and tip-to-work distance, which is the distance between the tip of the welding gun and the workpiece [20], as shown in Fig. 3B. The tip-to-work distance indicator turned green when within a threshold, allowing wire feeding, and remained red otherwise. Based on prior work [20,63], we chose the threshold for tip-to-work distance at 20 mm.

3.5. Lift operation

As our implementation focuses on construction tasks at height, operating a lift or any kind of elevation equipment is necessary. We integrated the control system for lift operation into a single-user telemanipulation setup, using manual teleoperation with the non-dominant hand. In our implementation, the left-hand VR controller's thumbstick is used for moving forward/backward and left/right, and a trigger is employed to switch the thumbstick's function to moving up and down.

4. Evaluation

We conducted a user study to compare the performance and user experience of five viewpoint control techniques for teleoperation of welding robots in construction: decoupled (D-Hand and D-Head), coupled (C), and hybrid (H-Manual and H-Auto) vision motion. The user study utilized a within-subject design, meaning the participants tried all five conditions. The user study consisted of a series of paired welding tasks and post-task surveys, and semi-structured post-experiment interviews. This user study procedure was approved by the authors' Institutional Review Board (IRB 2306/002–002).

4.1. Setup and apparatus

The user study utilized a VR system including a Meta Quest 3 VR headset and Meta Quest Touch Plus controllers. The virtual robotic welding system integrated into the VR environment consists of a mobile lift and two robotics arms, both controlled by a single user. The mobile lift is a four-wheeled lift with a height range of 2.9 to 9.6 m. The two robotic arms are Franka Emika robots, controlled in joint position. The camera arm was equipped with a generic RGB-D camera, and the manipulation arm was equipped with a generic MIG-welding gun, modelled based on a design from the RoboDK library.

In this study, we focused on a robotic MIG-welding task involving the attachment of stiffeners to a standard wide-flange steel beam at 6.8 m in height within a virtual industrial hall (Fig. 4A). This simulated real-world construction setting assumes that the steel frame assembly has been completed, but the facilities and interior finishing remain unfinished. The virtual industrial hall used in this study is an open-source 3D model, which contains all parts of the steel supporting structure of the building, including steel columns, beams, roof trusses, and reinforced concrete spot-footings of the columns in real-scale [64]. The virtual building has dimensions of 62.9 m in length, 40.8 m in width, and 10.8 m in height. It features two naves (one main and one smaller adjacent) and two large entrance gates.

4.2. Participants

We recruited a total of 20 participants. Participants were divided into two groups (experienced welders and novice users) based on their welding experience. Our a priori power analysis indicated that the target sample size would be 9 participants per group (novice and expert) with a power of 0.8 and an alpha level of 0.05. We recruited 10 experienced welders (PE1 to PE10), with inclusion criteria being over 18 years old and having at least a month of welding experience. In addition, we recruited 10 novice users (PN1 to PN10) with no prior welding experience via the students' community of the authors' institution and online websites of the Department of Architecture and Architectural Engineering and the Department of Civil and Environmental Engineering. Table 2 lists the key characteristics of the participants based on their age, gender, experience levels of welding, robot operation, and VR games. All

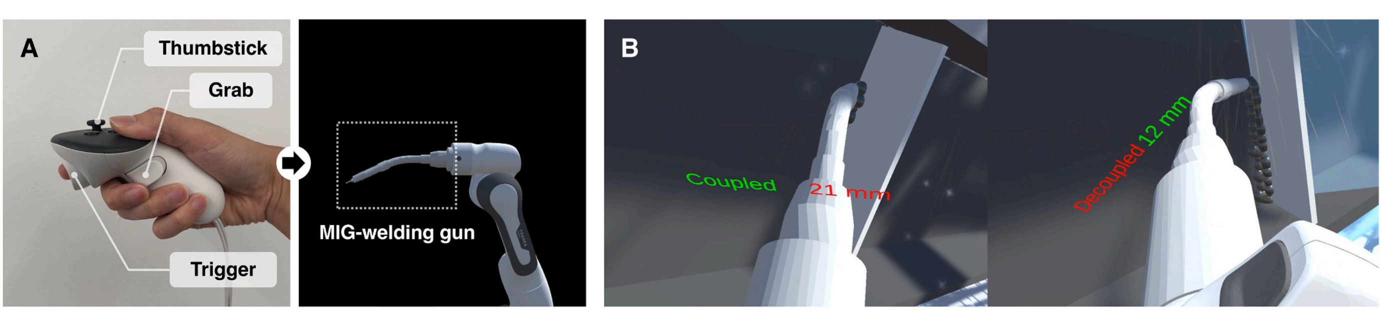

Fig. 3. (A) Direct mapping technique for teleoperation of the manipulation arm. (A–Left) Side view of the VR controller held in a user's hand. (A–Right) Side view of the robot's end effector in the simulated environment. (B) Virtual visual cues displaying the status of vision-motion and the tip-to-work distance.

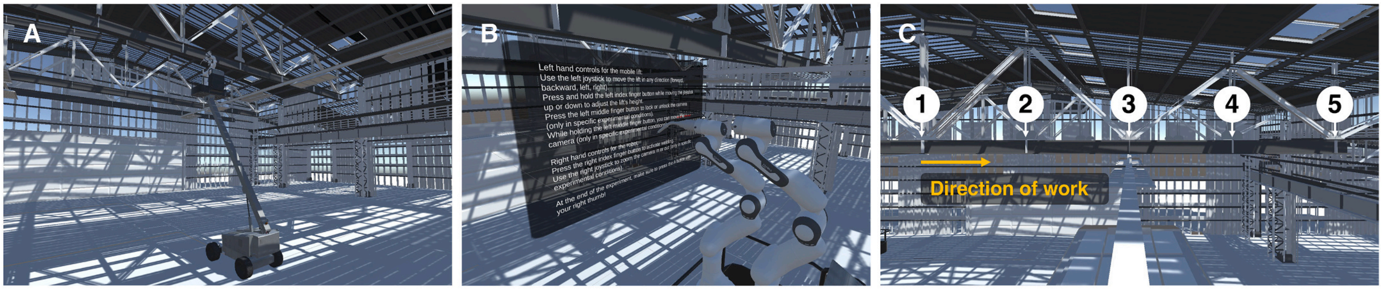

Fig. 4. Experimental setup. (A) Overview of the virtual industrial hall and the robotic welding system. (B) Tutorial text canvas shown to participants during the training phase. (C) Locations of the five weld beam stiffeners. Participants welded in sequence from 1 to 5.

Table 2
Demographic information and experience levels of the participants in the user study.

Variable	Novice users (*N* = 10)	Experienced welders (*N* = 10)
Age [21–43] (in years)	25.2 (*SD* = 4.3)	29.6 (*SD* = 6.4)
Gender		
Male	7	8
Female	3	2
Robot operation experience	–	4
VR games experience	9	3
Welding experience (in years)	–	2.1 (*SD* = 3.5)

20 participants reported they were right-handed.

4.3. Task and procedure

After obtaining consent from the participant, the experimenter introduced the purpose and scenario of the user study and tasks. The experimenter then conducted an initial survey to gather the participant's basic information, such as age, gender, and relevant experience in robot operation, VR games, and welding. Then, the participants with no prior welding experience watched a brief video tutorial featuring welding experts, demonstrating how to hold a welding gun with correct hand poses and how to weld a vertical joint with the vertical up welding technique. This step was skipped for experienced welders.

The experimenter then equipped participants with the VR headset and guided them through hands-on training on the robotic welding system within the VR environment, following on-screen instructions (Fig. 4B). The participants tried each of the five viewpoint control techniques once, in the order of D-Head, C, H-Manual, H-Auto, D-Hand. We chose this order to maximize training effectiveness and minimize dizziness during training, based on our pilot testing with two participants. All participants reported that they felt confident in using the VR system throughout the training process, which typically lasted 15 to 30 min. After completing the training, participants were asked to take a 3-min break without the VR headset. During the training and experiment, participants were asked to remain seated.

In the experiment, the order of conditions was randomized and fully counterbalanced according to the Balanced Latin Square method to control for possible within-subject effects. Participants were required to weld five weld beam stiffeners to a steel beam, resulting in 10 seams per trial (Fig. 4C). During the trial, the VR scene and the participant's HMD display were screen-recorded. Fig. 5 shows first- and third-person views of a sample trial of welding one seam in our experiment using the D-Head condition. Participants ended each trial by pressing the A button on the right-hand VR controller. After each trial, the experimenter administered a post-task survey consisting of two questionnaires (see Section 4.4). Participants were offered a 3-min break before proceeding to the next condition. After the final condition, participants were asked to rate their preferences for each condition and participated in a semi-structured post-experiment interview discussing their experiences and potential improvements to the teleoperation system. All interviews were audio-recorded, and all trial data, including user inputs, lift and end-effector locations, and robot joint positions, were recorded at 50 Hz. In total, the user study took approximately 60 to 90 min.

4.4. Measures

The measures for general task performance include task completion

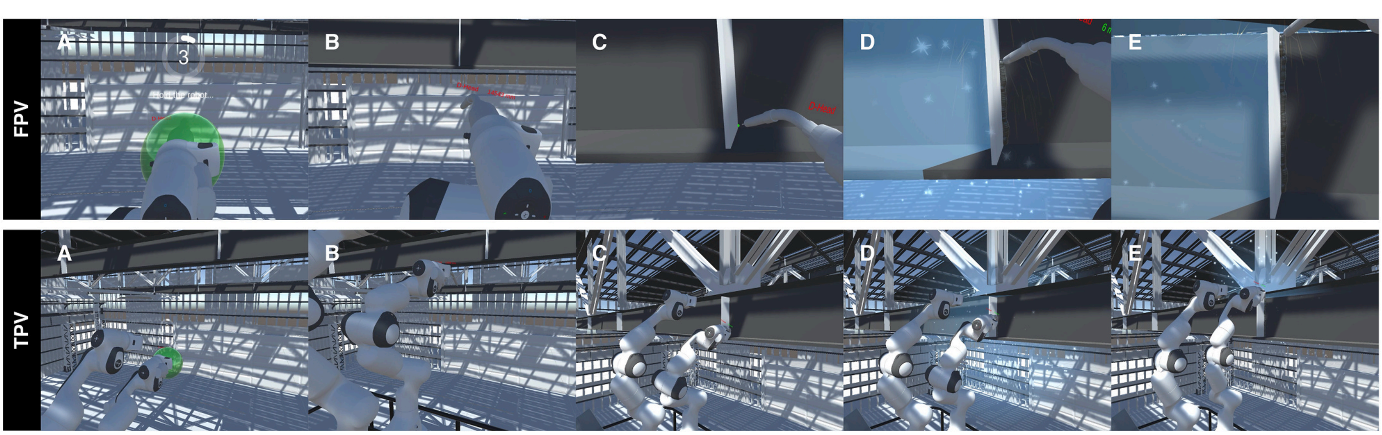

Fig. 5. (Top) First-person view (FPV) and (Bottom) Third-person view (TPV) of a sample trial of welding one seam using the D-Head condition. (A) The operator holds the manipulation arm with the right hand in the designated sphere for 3 s. (B) The operator navigates the mobile lift to approach the weld area and (C) position the tip of the welding gun at the initial point, indicated by the green reticle dot and tip-to-work distance text. (D, E) The operator performs welding using the vertical up technique.

time, weld coverage ratio, path deviation, number of collisions, and trajectory length. Task completion time is defined as the duration from the start of user interaction (after 3 s of holding the manipulation arm in the sphere) to the completion of all 10 welding seams (when the user presses the A button of the VR controller). Weld coverage ratio measures how successfully the welding seams were covered, as incomplete coverage or penetration can result in issues such as insufficient joint strength or potential weld defects [63]. This ratio is calculated by comparing the actual weld volume to the line segment of the target seam, in which a ratio of 1.0 indicates that the seam is fully covered. Path deviation, a typical measure of welding quality [6,22], is the standard deviation of distances between each point on the welding path and the nearest point on the ideal path, which is the actual seam location. For weld coverage and path deviation, to obtain a single data point for each condition, we averaged the results by seam and then calculated the average of the 10 data points. The number of collisions includes all collisions between the welding gun and any components of the workspace. Trajectory length is the total travel distance of the end-effector relative to the base position. We measured trajectory length for each robot arm.

We employed two post-task questionnaires to quantitatively understand participants' experiences with the five techniques: perceived workload was measured with the weighted NASA-TLX [65] and perceived usability was measured with the system usability scale (SUS) [66]. Participants were asked to respond on a 20-point scale for NASA-TLX and a 5-point Likert scale (from 1 = Strongly disagree to 5 = Strongly agree) for SUS. Additionally, we asked the participants to score their preference for each technique on a 5-point Likert scale (from 1 = Very unsatisfied to 5 = Very satisfied).

4.5. Analysis

To analyze the results of general task performance and subjective measures, we used a mixed analysis of variance (ANOVA) with expertise (Novice and Expert) as a between-subject factor and viewpoint control techniques (C, D-Hand, D-Head, H-Auto, and H-Manual) as a within-subject factor. In cases where Mauchly's test indicated a violation of sphericity, we used the Greenhouse-Geisser correction. *P*-values for post-hoc tests were corrected using Bonferroni correction to account for multiple comparisons. For trajectory length and time duration, trials were divided into three phases: lift operation, exploration, and manipulation phase. The phase was classified as lift operation, when there were any user inputs for lift operation using the left-hand VR controller. The phase was classified as manipulation when the user was pressing the trigger of the right-hand VR controller to feed the wire for welding. The phase was classified as exploration when the user was either exploring the workspace to get a better viewpoint or aligning the welding gun to the workpiece before welding, without any lift operation or manipulation.

5. Results

We present the results of our user study by first offering comparisons across techniques and expertise levels for general task performance and subjective measures. Following this, we compare task phases, focusing on measures of task efficiency such as task completion time and trajectory length. Finally, we examine hybrid vision-motion interactions, focusing on the selection rate of viewpoint control techniques and the duration of each technique.

5.1. General task performance

Fig. 6 shows the results of general task performance measures. Our mixed ANOVA revealed a significant main effect of viewpoint control technique on task completion time ($F(4,72) = 12.92$, $p < 0.0001$, partial $\eta^2 = 0.42$), path deviation ($F(4,72) = 5.63$, $p < 0.001$, partial $\eta^2 = 0.24$), and number of collisions ($F(4,72) = 6.76$, $p < 0.001$, partial $\eta^2 = 0.27$). Post-hoc analyses demonstrated that participants completed tasks significantly faster with the D-Head compared to the C ($p < 0.0001$) and D-Hand ($p < 0.001$). Moreover, H-Auto showed significantly faster task completion time compared to C ($p < 0.05$) and D-Hand ($p < 0.001$). H-Manual also showed significantly faster task completion time compared to D-Hand ($p < 0.001$). In terms of path deviation, H-Manual showed significantly lower values compared to C ($p < 0.01$). For the number of collisions, analyses revealed that the number of collisions was significantly fewer with the D-Head compared to the C ($p < 0.05$) and D-Hand ($p < 0.001$), and with the H-Manual compared to the C ($p < 0.05$) and D-Hand ($p < 0.05$).

Furthermore, expertise had a significant main effect on task completion time ($F(1,18) = 20.74$, $p < 0.001$, partial $\eta^2 = 0.54$), weld coverage ratio ($F(1,18) = 68.35$, $p < 0.0001$, partial $\eta^2 = 0.79$), and number of collisions ($F(1,18) = 6.12$, $p < 0.05$, partial $\eta^2 = 0.25$), and a marginal main effect on path deviation ($F(1,18) = 3.77$, $p = 0.07$, partial $\eta^2 = 0.17$). However, we found no significant interaction effect between viewpoint control technique and expertise on task completion time ($F(4,72) = 0.92$, $p = 0.46$, partial $\eta^2 = 0.05$), weld coverage ratio ($F(4,72) = 1.07$, $p = 0.38$, partial $\eta^2 = 0.06$), path deviation ($F(4,72) = 1.44$, $p = 0.23$, partial $\eta^2 = 0.07$), or number of collisions ($F(4,72) = 0.79$, $p = 0.53$, partial $\eta^2 = 0.04$). This suggests that the differences in general task

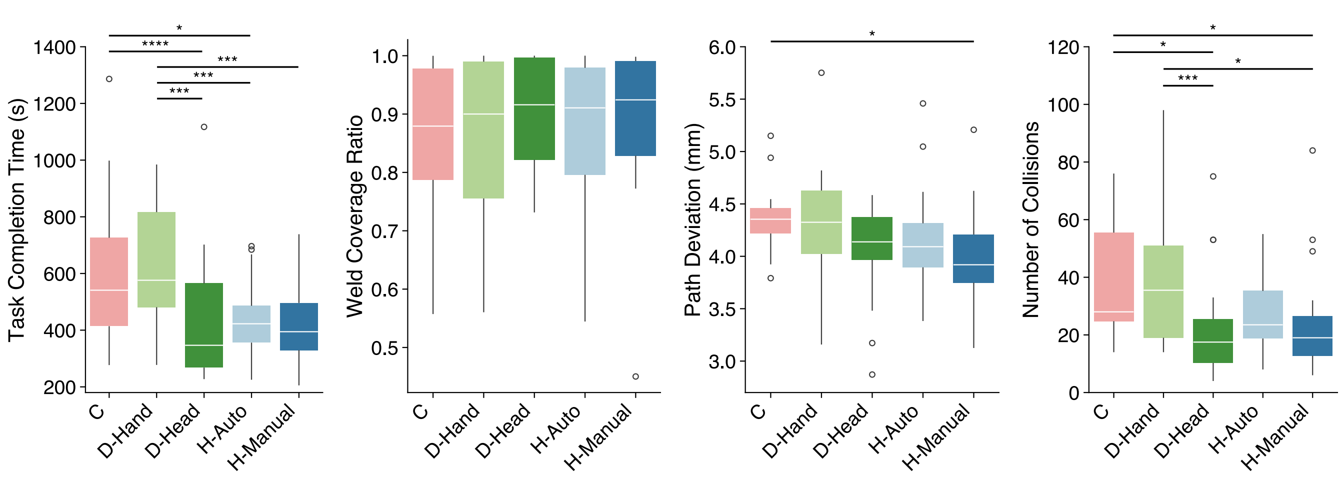

Fig. 6. General task performance by viewpoint control techniques. * $p < .05$, *** $p < .001$, **** $p < .0001$.

performance based on the viewpoint control technique are consistent for both novices and experts.

Fig. 7 shows the total trajectory length of the camera and manipulation arms. The mean values were 34.07 ($SD = 52.64$) and 52.32 ($SD = 35.30$) for the camera and manipulation arms, respectively. Our mixed ANOVA showed a significant main effect of viewpoint control technique on camera arm trajectory length ($F(4,72) = 20.67$, $p < 0.0001$, partial $\eta^2 = 0.54$) and manipulation arm trajectory length ($F(4,72) = 6.94$, $p < 0.01$, partial $\eta^2 = 0.28$). Post-hoc analyses demonstrated that the camera arm's movement was significantly lower in the D-Head compared to hybrid viewpoint control techniques: H-Auto ($p < 0.0001$) and H-Manual ($p < 0.05$). Hybrid viewpoint control techniques showed significantly lower trajectory length compared to D-Hand: H-Auto ($p < 0.001$) and H-Manual ($p < 0.05$). The C condition showed significantly longer trajectories of the camera arm compared to all other conditions: D-Hand ($p < 0.001$), D-Head ($p < 0.001$), H-Auto ($p < 0.01$), and H-Manual ($p < 0.01$). In terms of the manipulation arm, post-hoc analyses found that hybrid viewpoint control techniques significantly lowered the arm's movement compared to D-Hand: H-Auto ($p < 0.05$) and H-Manual ($p < 0.01$). Fig. 8 displays example trajectories of the two robot arms. These examples trajectories visually show significant differences in the trajectories of camera arm, displayed in yellow lines.

There was no significant main effect of expertise on both camera arm trajectory length ($F(1,18) = 2.11$, $p = 0.16$, partial $\eta^2 = 0.11$) or manipulation arm trajectory length ($F(1,18) = 1.51$, $p = 0.24$, partial $\eta^2 = 0.08$), and no significant interaction effect between viewpoint control technique and expertise on camera arm trajectory length ($F(4,72) = 0.42$, $p = 0.80$, partial $\eta^2 = 0.02$). We found a significant interaction effect on manipulation arm trajectory length ($F(4,72) = 3.91$, $p < 0.01$, partial $\eta^2 = 0.18$), but no significant pairwise differences were identified.

5.2. Subjective measures

Fig. 9 shows the results of subjective measures. Our mixed ANOVA revealed a significant main effect of viewpoint control technique on SUS scores ($F(4,72) = 14.47$, $p < 0.0001$, partial $\eta^2 = 0.45$), NASA-TLX scores ($F(4,72) = 11.62$, $p < 0.0001$, partial $\eta^2 = 0.39$), and preference scores ($F(4,72) = 14.39$, $p < 0.0001$, partial $\eta^2 = 0.44$). Post-hoc analyses demonstrated similar tendencies across all three subjective measures. Participants' average perceived usability (SUS scores) using the D-Head was significantly higher than when using C ($p < 0.01$), D-Hand ($p < 0.0001$), and H-Auto ($p < 0.001$). Moreover, the average SUS score of the H-Manual was significantly higher than the D-Hand ($p < 0.01$) and H-Auto ($p < 0.05$). Furthermore, post-hoc analyses revealed that the average scores of the NASA-TLX was significantly lower in the D-Head compared to the C ($p < 0.05$), D-Hand ($p < 0.01$), and H-Auto ($p < 0.05$). H-Manual was also showed lower scores compared to the D-Hand ($p < 0.05$) and H-Auto ($p < 0.05$). Lastly, post-hoc analyses demonstrated that the participants preferred the D-Head over C ($p < 0.001$), D-Hand ($p < 0.0001$), and H-Auto ($p < 0.001$). Furthermore, participnats significantly preferred H-Manual over D-Hand ($p < 0.05$) and H-Auto ($p < 0.001$).

There was no significant main effect of expertise on SUS ($F(1,18) = 0.92$, $p = 0.35$, partial $\eta^2 = 0.05$), NASA-TLX scores ($F(1,18) = 0.09$, $p = 0.77$, partial $\eta^2 = 0.01$), or preference scores ($F(1,18) = 1.43$, $p = 0.25$, partial $\eta^2 = 0.07$). We found no significant interaction effect between viewpoint control technique and expertise on SUS ($F(4,72) = 2.09$, $p = 0.09$, partial $\eta^2 = 0.10$), NASA-TLX scores ($F(4,72) = 1.72$, $p = 0.15$, partial $\eta^2 = 0.09$), and preference scores ($F(4,72) = 1.94$, $p = 0.11$, partial $\eta^2 = 0.10$). Although not statistically significant, we observed a trend in which experts ($M = 3.30$, $SD = 0.95$) showed a 37.5 % higher preference for the C condition compared to novices ($M = 2.40$, $SD = 1.26$), while all conditions except for C received similar or higher preference scores from novices.

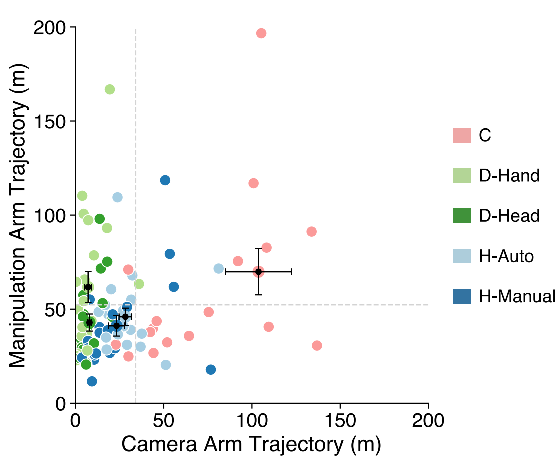

Fig. 7. Scatter plot of the trajectory length of the camera arm (x-axis) and manipulation arm (y-axis) by viewpoint control techniques. The error bars represent the standard error of the mean. The grey dashed lines indicate the overall mean trajectory length for each arm.

5.3. Task efficiency by phases

We further analyze the effects of viewpoint control techniques on the task efficiency for the different task phases defined in Section 4.5. We represent task efficiency using two measures: (1) time duration and (2) trajectory length of manipulation arm. Fig. 10 shows the task time duration by viewpoint control techniques. The average time for each phase was 65.03 s ($SD = 16.88$) for lift operation, 229.36 s ($SD = 141.82$) for exploration, and 211.14 s ($SD = 116.89$) for manipulation. Our one-way repeated measures analysis of variance (RM-ANOVA) revealed a significant difference among viewpoint control techniques in time duration for the phases of lift operation ($F(4,76) = 7.00$, $p < 0.0001$) and exploration ($F(4,76) = 21.27$, $p < 0.0001$). In the lift operation phase, post-hoc analyses found significant differences between C and D-Head ($p < 0.01$) and D-Hand and D-Head ($p < 0.05$). For the exploration phase, post-hoc analyses demonstrated that the participants took more time for exploration with C compared to D-Head ($p < 0.0001$), H-Auto ($p < 0.01$), and H-Manual ($p < 0.01$). Post-hoc analyses also showed longer exploration times for D-Hand compared to D-Head ($p < 0.0001$), H-Auto ($p < 0.001$), and H-Manual ($p < 0.001$). However, no significant differences were found in the time duration of the manipulation phases by viewpoint control techniques.

Fig. 11 shows the trajectory length of manipulation arm by viewpoint control techniques. The average trajectory length of each phase was 15.09 m ($SD = 10.68$) for lift operation, 43.35 m ($SD = 29.78$) for exploration, and 12.66 ($SD = 3.84$) for manipulation. In terms of trajectory length, our one-way RM-ANOVA revealed no significant difference among viewpoint control techniques across the three phases.

5.4. Selection of viewpoint control techniques in hybrid vision-motion

Finally, we investigate the hybrid vision-motion techniques, focusing on their effects on performance for the two different control techniques. In the exploration phase, the average rate of vision-motion coupling time relative the total time of each phase was 0.29 ($SD = 0.15$) and 0.25 ($SD = 0.20$) for H-Auto and H-Manual, respectively. In the manipulation phase, the average rate was and 0.92 ($SD = 0.15$) and 0.30 ($SD = 0.28$) for H-Auto and H-Manual, respectively (Fig. 12A). Our paired t-test found significant differences between H-Auto and H-Manual ($r(19) = 9.90$, $p < 0.0001$) in the manipulation phase but no significant differences between them in the exploration phase. We also observed that the use of vision-motion coupling for each phase was not consistent across participants, with higher coefficients of variation for both phases in H-Manual (exploration: 0.78, manipulation: 0.94) compared to H-Auto

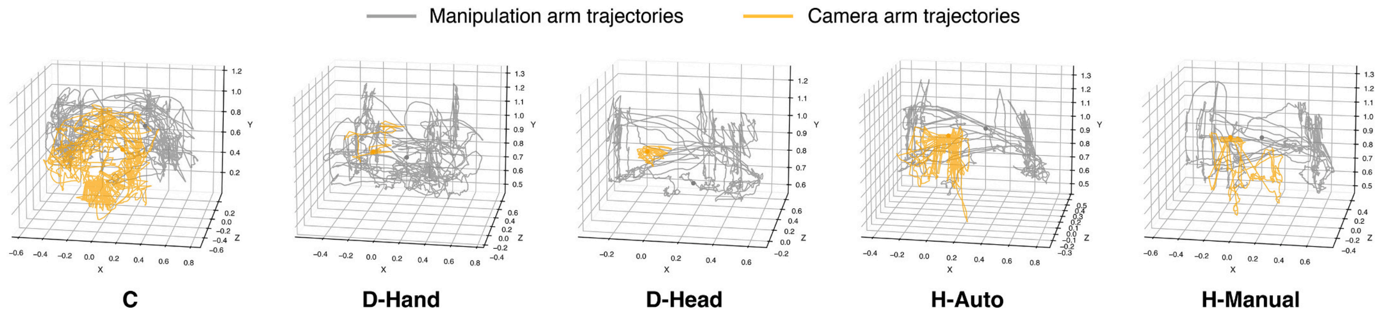

Fig. 8. Trajectories of the manipulation arm (grey line) and camera arm (yellow line) for the novice user PN4. Grey and yellow dots represent the initial positions of the manipulation arm and camera arm, respectively. The positions of the end-effectors are relative to the position of the mobile lift. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

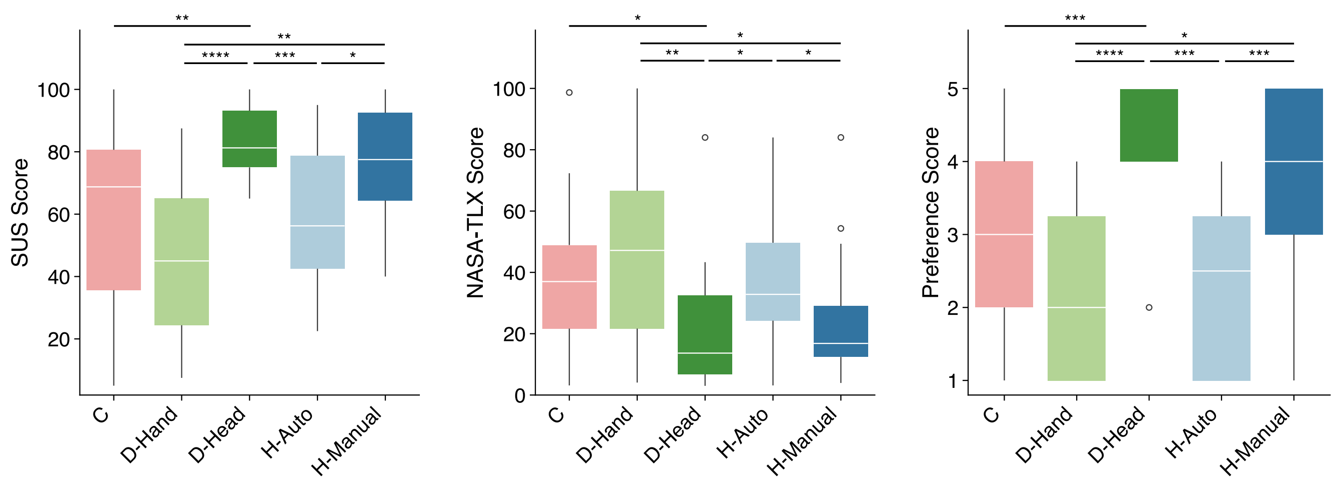

Fig. 9. Subjective measures by viewpoint control techniques. * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$.

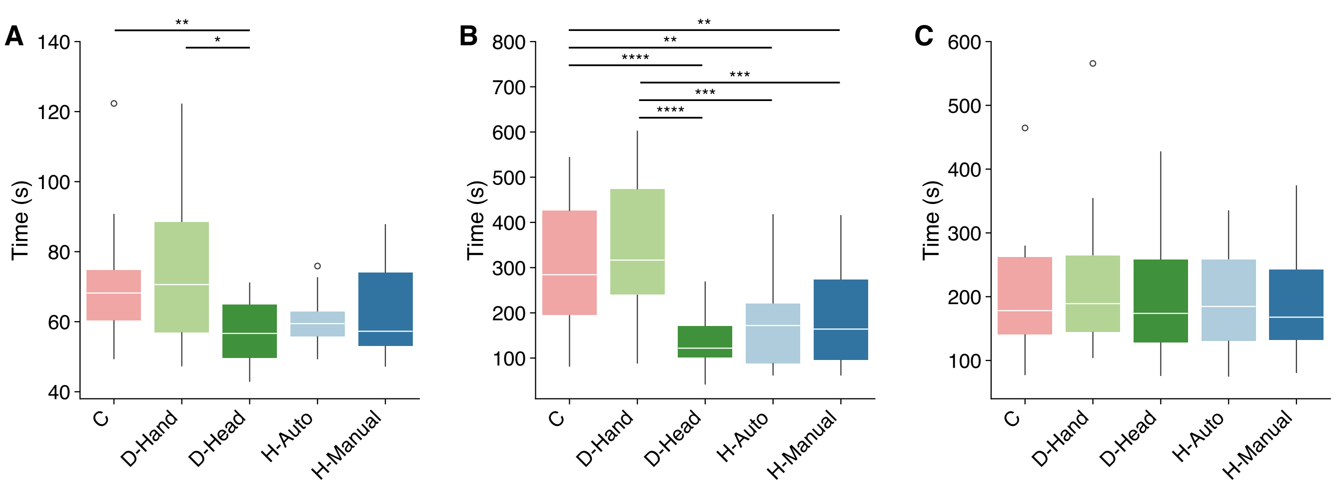

Fig. 10. Task time duration by viewpoint control techniques. (A) Lift operation phase, (B) exploration phase, and (C) manipulation phase. * $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$.

(exploration: 0.50, manipulation: 0.17). These results indicate that participants tended to use vision-motion coupling less for manipulation when they had the autonomy to switch the viewpoint control techniques. This preference for a decoupled viewpoint is reflected in Fig. 12B, which shows that when given the autonomy for vision-motion switching, selection rates varied among participants, but the decoupled viewpoint (highlighted in green in the bar plots) was generally preferred. The Pearson correlation between the rate of vision-motion coupling and total task completion time was not significant in the H-Manual but marginally significant in H-Auto ($r(18) = 0.44$, $p = 0.053$) (Fig. 12C). Specifically, we found a significant positive correlation between the rate of vision-motion coupling and the time for manipulation

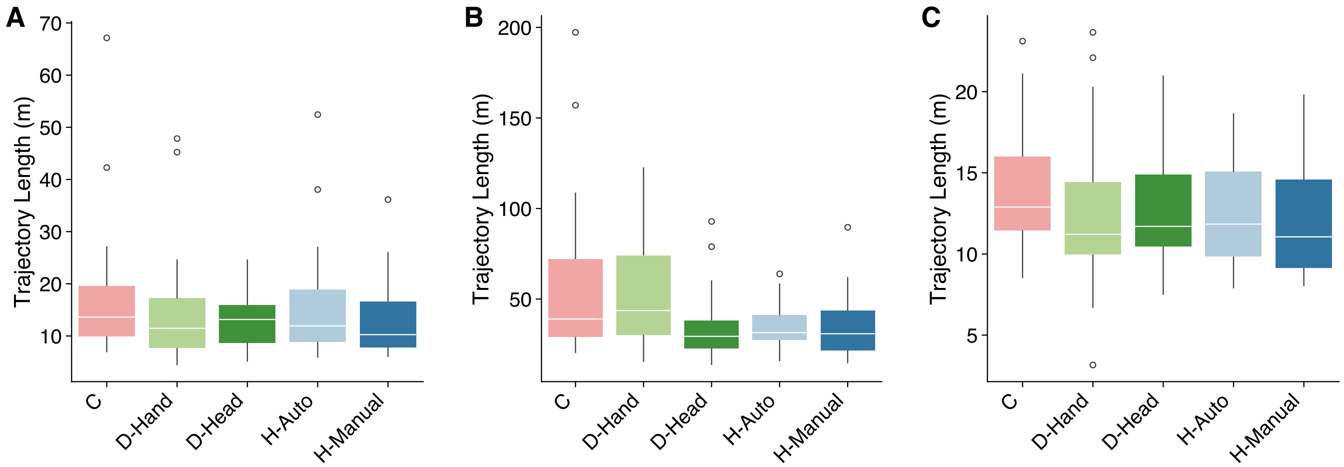

Fig. 11. Trajectory length of manipulation arm by viewpoint control techniques. (A) Lift operation phase, (B) exploration phase, and (C) manipulation phase.

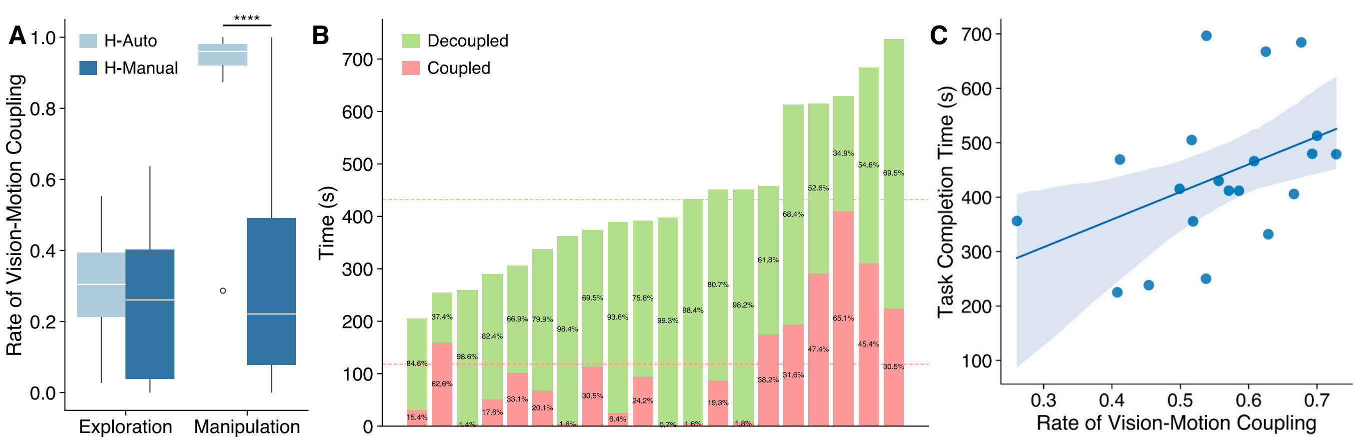

Fig. 12. (A) Rate of vision-motion coupling time relative to the total phase time for two hybrid techniques. **** $p < .0001$. (B) Accumulated time in coupled and decoupled states for H-Manual by participants. The horizontal lines indicate mean time of each state. The numbers on bars indicate the rate of each state. (C) Relationship between rate of vision-motion coupling and total task completion time in H-Auto. The shaded area represents 95 % confidence interval for the regression line.

in H-Auto ($r(18) = 0.85$, $p < 0.0001$).

6. Discussion

6.1. Main findings

In this study, we aim to investigate the effects of viewpoint control techniques on task performance and user experience during teleoperation of welding robots in construction. While previous work has focused less on user viewpoint compared to user interface designs in the construction domain, the findings of the current study demonstrate that the way the viewpoint is provided (viewpoint control technique) is indeed an important factor that significantly impacts teleoperation performance and user experience. In this section, we discuss our main findings with supporting qualitative findings and theoretical implications, as well as the challenges and opportunities for each viewpoint control technique.

6.1.1. Decoupled vision-motion

In general, decoupled vision-motion with head motion-based control (D-Head) outperformed coupled vision-motion for remote welding in construction. As demonstrated in the quantitative results, D-Head showed better performance compared to C in terms of task completion time, collision occurrences, camera arm movement, all subjective measures, and time for operating the lift and exploration. 14 participants (PN1, PN3–5, PN9, PN10, PE1, PE2, PE5–10) rated D-Head the highest scores in user preference (5 = Very satisfied).

Through our study, we empirically observed that welding in construction involves frequent changes in viewpoint across all task phases, particularly in environments where welding joints are scattered, necessitating constant base movement and exploration. Our phase-wise analysis supports this observation, revealing that the exploration phase accounts for an average of 45.4 % of the total task completion time. In this context, we found that participants instinctively tended to use head movements to control the viewpoint. This was reflected in the interviews, in which participants commented on the intuitiveness (PN1, PN6, PN10), comfort (PN2, PN7, PN8, PE9), and similarity to real-world conditions (PN3–5, PE1–8, PE10) of the head motion-based control. Participants with welding experience used head movements more actively for exploration, such as bending their backs or tilting their heads.

In particular, for the main welding phase, we conclude that the decoupled viewpoint was generally preferred by participants because it allowed the camera to move without significantly affecting other conditions, such as the location of the welding gun or the pose of the manipulation arm, and vice versa, which was not usually possible with the coupled viewpoint. This preference was also reflected in interviews, in which a few participants noted that they preferred to have their viewpoint "locked to the workspace" during the main welding phase (PN4, PN8, PE5, PE10).

Meanwhile, participants found controlling the camera with their left hand challenging, adding to the additional difficulties in control (PN6, PN7, PE5, PE9, PE10) and cognitive burden (PN1, PN10). In hand motion-based control, we observed that participants tried to use their heads to control the viewpoint even if they knew they could not do so. This led to increased perceived workload, such as frustration, as evidenced by the NASA-TLX results (Fig. 9–Middle). These findings align with the previous studies in which head movements were either suppressed [67] or did not affect the viewpoint [41]. Most participants also expressed dizziness while using the D-Hand condition, which they attributed to subtle changes in the orientation of the camera (PN6, PE1–5, PE7–10) and lift movement during the unnatural pose of the viewpoint (PN3, PN7). This implies that blurring or masking out the intermediate camera feeds during the transition of the camera by displaying a black screen [15] could be one possible solution for reducing dizziness and enhancing user experience for dynamic viewpoint control, especially when deployed on real-world construction sites.

6.1.2. Coupled vision-motion

Coupled vision-motion, as opposed to previous studies [10,42] that presented the effectiveness of such technique for teleoperation, was not preferred by participants, particularly those who were novices in welding. Our quantitative results indicate that coupled vision-motion underperformed compared to other techniques, in terms of task completion time, and collision occurrences. Moreover, the trajectory length of the camera arm was the longest among all viewpoint control techniques. We conclude that the primary cause of this is occlusions, both by the robots (i.e., the camera arm and the manipulation arm, see Fig. 13A) and the workpiece (i.e., the flange of the steel beam and the stiffeners, see Fig. 13B and C, respectively). The difficulties related to the misalignment between the control and view frames, requiring participants to perform mental transformations, was particularly salient in this condition (PN3, PN4, PN6, PN9). This was also reflected in the exploration time, in which coupled vision-motion showed a 116.7 % increase compared to the D-Head condition.

Moreover, coupled vision-motion demonstrated worse performance in subjective measures, including perceived workload, usability, and user preference. In particular, participants reported a 20.7 % and 27.1 % increase in mental and physical demand dimensions of the NASA-TLX, respectively, compared to their average scores. These results are not in line with previous studies [10,42] that suggested coupled vision-motion can achieve lower perceived workload and increased perceived usability and interface acceptance by providing robot-following viewpoints. Based on our findings, we conclude that while actively finding the proper viewpoint without occlusions can positively raise alertness during relatively simple tasks, such as aligning the end-effector to pipes [10] or pressing buttons [42], it can become a physical and cognitive burden during long-duration tasks in construction. Moreover, construction tasks often introduce various types of occlusions, making simple implementation of dynamic viewpoints that always follow the robot ineffective for avoiding occlusions.

Nevertheless, some participants expressed their preference to the coupled vision-motion. All participants who favored this technique commented on the expected effectiveness of this technique once they become accustomed to it (PN8–10, PE6). This preference was more common among experienced welders (see Section 5.2), who generally maintained a consistent welding posture and adjusted the lift to achieve their desired position (PE1–3, PE5, PE6, PE10), unlike novices who frequently changed postures and relied on moving the lift.

6.1.3. Hybrid vision-motion

Hybrid vision-motion was effective, but only when the location of the coupled viewpoint was optimized "appropriately". In general, participants expressed their favor of being able to leverage both views during welding, especially when the workspace is complex as in our settings. Six participants (PN1, PN6, PN7, PN8, PE3, PE4) gave the highest scores in user preference (5 = Very satisfied) for hybrid vision-motion, all of which were for H-Manual. PE3, who had more than 2 years of welding experience, linked his preference for hybrid vision-motion with manual switching to his welding posture:

> *"(I think the reason I prefer the H-Manual the most is) due to the posture when welding. When welding, I only focus on the tip of the wire, so I tend to bend forward to get a closer view and follow along with the flow of the weld pool. In such situations, it was convenient to change the viewpoint and couple with the welding gun using the (left-hand) button."*
>
> *(PE3)*

The appropriate viewpoint not only reduced the user's physical burden during welding but also offer them a clear, unobstructed view of both the target welding seam and the wire tip, resulting in improved welding quality. Manually switching to appropriate viewpoint with hybrid vision-motion showed a trend of reducing the physical demand dimension of the NASA-TLX by 46.3 % across all techniques and by 20 % compared to D-Head. We also observed a trend in which path deviation in H-Manual was 4.9 % lower compared to the average across all techniques and 2.7 % lower compared to D-Head.

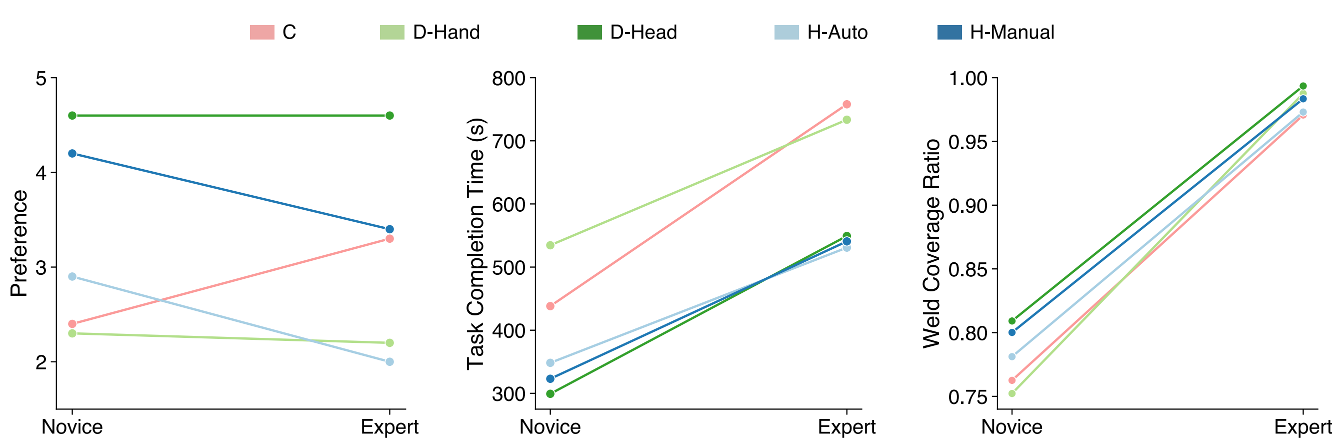

Fig. 13. Examples of occlusions occurred during the C condition (experienced welder PE5): (A) occlusion by the manipulation arm, (B) occlusion by the top flange of the steel beam, and (C) occlusion by the stiffeners.

Through our user study, we empirically observed that participants found the viewpoint "appropriate" when the optimized position closely matched their current view without occlusions. This typically occurred when the pre-coupling view was already aligned with the optimal viewpoint. However, challenges arose when the initial view did not align with the conditions for an appropriate view, leading to abrupt viewpoint adjustments despite being appropriate, as shown in Fig. 14. Participants who rated hybrid vision-motion lower (PN5, PE5, PE6, PE9, PE10) attributed their lower preference scores to the occasional changes to unnatural viewpoint locations. This issue arises from the second optimization function, which centers the camera on the manipulation arm end-effector, a method commonly used in autonomous viewpoint adjustment [13,34]. Despite this, some participants used this optimization as an opportunity to deal with occlusions. PN8, who primarily used the strategy of adjusting the height of the lift to avoid occlusions due to the top flange of the steel beam during welding, showed his preference for choosing alternate views with a single button instead of controlling the lift:

> *"Occlusions mainly occurred at the top part of the workpiece. To address this, in the D-Head, it was cumbersome as I had to move the lift additionally because I couldn't complete the welding in one go. However, in the H-Manual, it was more convenient as I could choose a different viewpoint instead of moving the lift."*
>
> *(PN8)*

Our implementation of velocity-based judgement for phase transitions in automatic switching led to controversy among participants about the timing of transitions. Participants who did not prefer the timing expressed frustration with maintaining the movements of their right hand (PN4, PN10) and wanted a certain time delay for switching after initial weld point positioning (PE6, PE8). In contrast, some participants stated that the timing was appropriate (PN1, PN9) and was within a predictable and controllable range (PN7).

Regardless of their thoughts on switching timing, participants generally preferred to have the autonomy for switching rather than letting robot have the autonomy (PN1, PN3–6, PN9, PN10, PE1–10). Our quantitative results indicate that the hybrid viewpoint with manual switching (H-Manual) outperformed automatic switching (H-Auto) in terms of all subjective measures. Moreover, H-Manual outperformed the coupled viewpoint (C) in collision occurrences, camera arm movement, and duration of the exploration phase, and outperformed the decoupled viewpoint (D-Hand) in task completion time, collision occurrences, exploration time, movements of both the camera and manipulation arm, and all subjective measures. These results suggest that the hybrid vision-motion not only combines both coupled and decoupled techniques, but also enhances viewpoint control by leveraging the benefits of both approaches.

6.2. Guidelines for viewpoint control in teleoperated construction robots

Based on our results, we uncover several guidelines for telerobotic systems working at height in construction, particularly with a dynamic viewpoint control technique to support complex, skill-intensive, and long-duration robotic tasks such as welding.

1. For general purposes and settings, decoupled vision-motion with head motion-based control should be prioritized.
2. If the task workspace has complex structures with frequent occlusions, hybrid vision-motion should be used. However, the transition of control techniques should be initiated by users, the coupled view should closely match their current viewpoint, and the system should enable users to decouple vision-motion as needed.
3. If leveraging the advantages of vision-motion coupling is more important than the efficiency of the task and robot arms, the coupled vision-motion technique can be considered, with adequate user training beforehand.

6.3. Limitations and future work

One limitation of our study is that, although we provided identical instructions on welding tasks and the teleoperation system to both experienced welders and novices, differences in their welding approaches, such as welding patterns and the use of dry runs, were observed. While we argue that these differences did not significantly

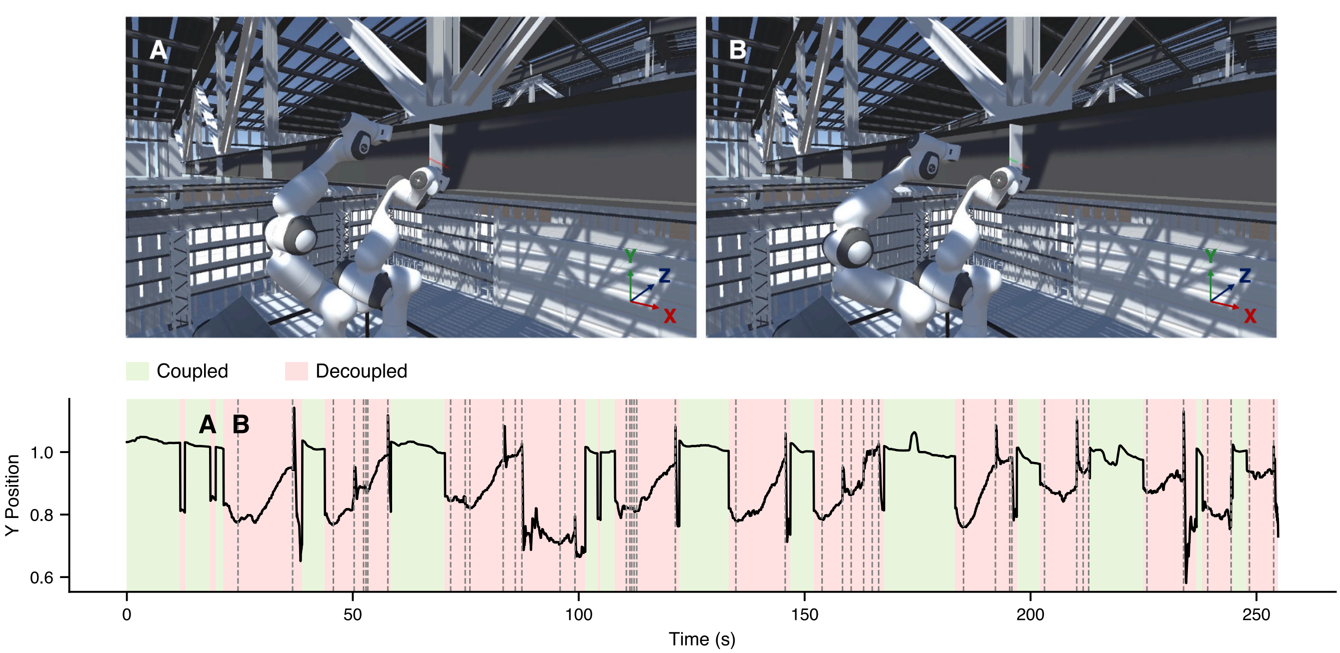

Fig. 14. The challenge of hybrid vision-motion: abrupt adjustment of the viewpoint. (Bottom) Camera arm's end-effector position in the Y-axis direction (vertical in world coordinates) from novice user PN5 in the H-Manual condition. The grey dashed lines indicate changes in the right-hand trigger state, which activates the welding gun. (Top–A) Third-person view before the user triggers vision-motion coupling. (Top–B) Third-person view after the user triggers vision-motion coupling. Note how the Y position of the camera arm changes.

affect our primary focus on analyzing the interaction effect between viewpoint control techniques and expertise, future work should explicitly control for these variations in welding skills between user groups for more robust and extensive evaluations. Next, as we focused on tool operations, bodily movements, and vision-motion coordination for welding in this study, other procedures of welding, such as adjusting voltage, changing wire in the wire-feeder, and changing shielding-gas cylinders, were simplified or omitted. Finally, our simulated environment cannot fully replicate the complexities and unpredictable nature of real-world construction sites. Challenges such as diverse types of collisions, dizziness, time delays between control input and system output [27], and visual or robot state errors [68] may become more significant in on-site implementations of our teleoperation system. Future work could benefit from addressing these sim-to-real gaps by validating findings through real-world deployments.

7. Conclusion

Effective viewpoint control is crucial for the success of teleoperation for construction tasks, yet it has been underemphasized in previous literature. In this paper, we investigated the impact of five different dynamic viewpoint control techniques on task performance and user experience during teleoperation of welding-at-height robots in construction through a user study with novice users and experienced welders. Our results show that viewpoint control technique significantly influences task performance and user experience, particularly with exploration time decreasing by over 53 % when using decoupled vision-motion with head motion-based control (D-Head) instead of coupled vision-motion. D-Head generally outperformed other techniques, providing better task efficiency and higher user preference. Hybrid vision-motion with manual switching (H-Manual) also showed promise, especially when users faced occlusions or needed to adjust their welding posture. Our work successfully contributes insight into developing viewpoint control systems for teleoperated robots in construction, particularly for skill-intensive tasks that require frequent viewpoint adjustments, such as welding, crane operation, and pipeline inspection.

Future research should expand our work by evaluating dynamic viewpoint control techniques in real-world construction settings. Incorporating more complex real-world scenarios will help bridge the gap between our simulated environment and actual construction sites, addressing factors such as repeatability, robotic platform stability, operator dizziness, and control-response delays. On-site validation will further enhance the applicability of viewpoint control techniques, including hybrid vision-motion, and ensuring their effectiveness for a wider range of teleoperated construction tasks.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Sungboo Yoon: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Moonseo Park:** Writing – review & editing, Supervision. **Changbum R. Ahn:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

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.

Acknowledgements

This research was supported by the New Faculty Startup Fund and the Institute of Engineering Research (IOER) at Seoul National University (SNU), the Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (RS-2023-00284237), and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2024-00415472).

Data availability

Data will be made available on request.

References

[1] D. Liu, J. Kim, Y. Ham, Multi-user immersive environment for excavator teleoperation in construction, Autom. Constr. 156 (2023) 105143, https://doi.org/10.1016/j.autcon.2023.105143.
[2] T. Zhou, P. Xia, Y. Ye, J. Du, Embodied robot teleoperation based on high-Fidelity visual-haptic simulator: pipe-fitting example, J. Constr. Eng. Manag. 149 (2023) 04023129, https://doi.org/10.1061/JCEMD4.COENG-13916.
[3] K. Duan, Z. Zou, Morphology agnostic gesture mapping for intuitive teleoperation of construction robots, Adv. Eng. Inform. 62 (2024) 102600, https://doi.org/10.1016/j.aei.2024.102600.
[4] Z. Zhang, M.O. Wong, W. Pan, Virtual reality enhanced multi-role collaboration in crane-lift training for modular construction, Autom. Constr. 150 (2023) 104848, https://doi.org/10.1016/j.autcon.2023.104848.
[5] S. Yoon, Y. Kim, M. Park, C.R. Ahn, Effects of spatial characteristics on the human–robot communication using deictic gesture in construction, J. Constr. Eng. Manag. 149 (2023) 04023049, https://doi.org/10.1061/JCEMD4.COENG-12997.
[6] H. Hu, A. Song, L. Wei, J. Mao, Development of a virtual reality-based teleoperated welding robot system for enhanced safety and efficiency, in: Proceedings of the 2024 4th International Conference on Robotics and Control Engineering, ACM, New York, NY, USA, 2024, pp. 77–82, https://doi.org/10.1145/3674746.3674758.
[7] H.J. Lee, S. Brell-Cokcan, Cartesian coordinate control for teleoperated construction machines, construction, Robotics 5 (2021) 1–11, https://doi.org/10.1007/s41693-021-00055-y.
[8] J. Du, W. Vann, T. Zhou, Y. Ye, Q. Zhu, Sensory manipulation as a countermeasure to robot teleoperation delays: system and evidence, Sci. Rep. 14 (2024) 4333, https://doi.org/10.1038/s41598-024-54734-1.
[9] 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.
[10] S. Kuitert, J. Hofland, C.J.M. Heemskerk, D.A. Abbink, L. Peternel, Orbital head-mounted display: a novel interface for viewpoint control during robot teleoperation in cluttered environments, in: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2023, pp. 1–7, https://doi.org/10.1109/IROS55552.2023.10341733.
[11] X. Wang, L. Shen, L.-H. Lee, A systematic review of XR-based remote human-robot interaction systems, ArXiv (2024) [Cs.HC], http://arxiv.org/abs/2403.11384.
[12] D. Rakita, B. Mutlu, M. Gleicher, An autonomous dynamic camera method for effective remote teleoperation, in: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Association for Computing Machinery, New York, NY, USA, 2018, pp. 325–333, https://doi.org/10.1145/3171221.3171279.
[13] E. Senft, M. Hagenow, P. Praveena, R. Radwin, M. Zinn, M. Gleicher, B. Mutlu, A method for automated drone viewpoints to support remote robot manipulation, in: 2022 IEEE/RSJ international conference on intelligent robots and systems (IROS), IEEE, 2022, pp. 7704–7711, https://doi.org/10.1109/IROS47612.2022.9982063.
[14] S.N. Young, R.J. Lanciloti, J.M. Peschel, The effects of Interface views on performing aerial telemanipulation tasks using small UAVs, Int. J. Soc. Robot. 14 (2022) 213–228, https://doi.org/10.1007/s12369-021-00783-9.
[15] R. Jia, L. Yang, Y. Cao, C.K. Or, W. Wang, J. Pan, Learning autonomous viewpoint adjustment from human demonstrations for telemanipulation, ACM Trans. Hum. Robot Interact. (2024), https://doi.org/10.1145/3660348.

[16] A. Naceri, D. Mazzanti, J. Bimbo, Y.T. Tefera, D. Prattichizzo, D.G. Caldwell, L. S. Mattos, N. Deshpande, The Vicarios virtual reality interface for remote robotic teleoperation, J. Intell. Robot. Syst. 101 (2021), https://doi.org/10.1007/s10846-021-01311-7.
[17] R.S. Johansson, G. Westling, A. Bäckström, J.R. Flanagan, Eye-hand coordination in object manipulation, J. Neurosci. 21 (2001) 6917–6932, https://doi.org/10.1523/jneurosci.21-17-06917.2001.
[18] Y. Liu, H. Jebelli, Human-robot co-adaptation in construction: bio-signal based control of bricklaying robots, in: Computing in Civil Engineering 2021, American Society of Civil Engineers, Reston, VA, 2022, pp. 304–312, https://doi.org/10.1061/9780784483893.038.
[19] P. Tavares, C.M. Costa, L. Rocha, P. Malaca, P. Costa, A.P. Moreira, A. Sousa, G. Veiga, Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality, Autom. Constr. 106 (2019) 102825, https://doi.org/10.1016/j.autcon.2019.04.020.
[20] A. Ipsita, L. Erickson, Y. Dong, J. Huang, A.K. Bushinski, S. Saradhi, A. M. Villanueva, K.A. Peppler, T.S. Redick, K. Ramani, Towards modeling of virtual reality welding simulators to promote accessible and scalable training, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2022, pp. 1–21, https://doi.org/10.1145/3491102.3517696.
[21] L. Wei, A. Song, H. Hu, J. Mao, Mixed reality-augmented remote welding system with virtual fixtures and autonomous agents, in: Proceedings of the 2024 4th International Conference on Robotics and Control Engineering, ACM, New York, NY, USA, 2024, pp. 131–137, https://doi.org/10.1145/3674746.3674767.
[22] Y.-P. Su, X.-Q. Chen, T. Zhou, C. Pretty, G. Chase, Mixed reality-enhanced intuitive teleoperation with hybrid virtual fixtures for intelligent robotic welding, NATO Adv. Sci. Inst. Ser. E Appl. Sci. 11 (2021) 11280, https://doi.org/10.3390/app112311280.
[23] D. Rakita, B. Mutlu, M. Gleicher, Remote telemanipulation with adapting viewpoints in visually complex environments, robotics, in: Science and Systems XV, 2019. https://par.nsf.gov/biblio/10104548.
[24] Y.-P. Su, X.-Q. Chen, C. Zhou, L.H. Pearson, C.G. Pretty, J.G. Chase, Integrating virtual, mixed, and augmented reality into remote robotic applications: a brief review of extended reality-enhanced robotic systems for intuitive telemanipulation and telemanufacturing tasks in hazardous conditions, NATO Adv. Sci. Inst. Ser. E Appl. Sci. 13 (2023) 12129, https://doi.org/10.3390/app132212129.
[25] K.H. Koh, M. Farhan, K.P.C. Yeung, D.C.H. Tang, M.P.Y. Lau, P.K. Cheung, K.W. C. Lai, Teleoperated service robotic system for on-site surface rust removal and protection of high-rise exterior gas pipes, Autom. Constr. 125 (2021) 103609, https://doi.org/10.1016/j.autcon.2021.103609.
[26] S. Yoon, Y. Kim, C.R. Ahn, M. Park, Challenges in deictic gesture-based spatial referencing for human-robot interaction in construction, in: Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC), International Association for Automation and Robotics in Construction (IAARC), Waterloo, Waterloo, Canada, 2021, pp. 491–497, https://doi.org/10.22260/isarc2021/0067.
[27] M. Seo, Y. Ham, The impact of time delay on construction teleoperation performance and human-robot interaction in space construction, J. Constr. Autom. Robot. 2 (2023) 9–14, https://doi.org/10.55785/jcar.2.3.9.
[28] T. Zhou, Q. Zhu, Y. Shi, J. Du, Construction robot teleoperation safeguard based on real-time human hand motion prediction, J. Constr. Eng. Manag. 148 (2022), https://doi.org/10.1061/(asce)co.1943-7862.0002289.
[29] S. Yoon, M. Park, C.R. Ahn, LaserDex: improving spatial tasks using deictic gestures and laser pointing for human–robot collaboration in construction, J. Comput. Civ. Eng. 38 (2024) 04024012, https://doi.org/10.1061/JCCEE5.CPENG-5715.
[30] S. Yoon, J. Park, M. Park, C.R. Ahn, A deictic gesture-based human-robot interface for in situ task specification in construction, in: Computing in Civil Engineering 2023, 2024, pp. 445–452, https://doi.org/10.1061/9780784485224.054.
[31] R. Ding, M. Cheng, Z. Han, F. Wang, B. Xu, Human-machine interface for a master-slave hydraulic manipulator with vision enhancement and auditory feedback, Autom. Constr. 136 (2022) 104145, https://doi.org/10.1016/j.autcon.2022.104145.
[32] P. Xia, H. You, J. Du, Visual-haptic feedback for ROV subsea navigation control, Autom. Constr. 154 (2023) 104987, https://doi.org/10.1016/j.autcon.2023.104987.
[33] M. Kamezaki, M. Miyata, S. Sugano, Video presentation based on multiple-flying camera to provide continuous and complementary images for teleoperation, Autom. Constr. 159 (2024) 105285, https://doi.org/10.1016/j.autcon.2024.105285.
[34] A. Valiton, H. Baez, N. Harrison, J. Roy, Z. Li, Active telepresence assistance for supervisory control: a user study with a multi-camera tele-nursing robot, in: 2021 IEEE international conference on robotics and automation (ICRA), IEEE, 2021, pp. 3722–3727, https://doi.org/10.1109/ICRA48506.2021.9561361.
[35] L. Chen, A. Naceri, A. Swikir, S. Hirche, S. Haddadin, et al., ArXiv (2024) [Cs.RO], http://arxiv.org/abs/2407.20156.
[36] M. Wilde, Z.K. Chua, A. Fleischner, Effects of multivantage point systems on the teleoperation of spacecraft docking, IEEE Trans. Hum. Mach. Syst. 44 (2014) 200–210, https://doi.org/10.1109/thms.2013.2295298.
[37] H. Stedman, B.B. Kocer, M. Kovac, V.M. Pawar, VRTAB-Map: a configurable immersive teleoperation framework with online 3d reconstruction, in: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), IEEE, 2022, https://doi.org/10.1109/ismar-adjunct57072.2022.00029.
[38] S. Rahnamaei, S. Sirouspour, Automatic viewpoint planning in teleoperation of a mobile robot, J. Intell. Robot. Syst. 76 (2014) 443–460, https://doi.org/10.1007/s10846-014-0028-7.
[39] Y. Chen, L. Sun, M. Benallegue, R. Cisneros, R.P. Singh, K. Kaneko, A. Tanguy, G. Caron, K. Suzuki, A. Kheddar, F. Kanehiro, Enhanced visual feedback with decoupled viewpoint control in immersive humanoid robot teleoperation using SLAM, ArXiv (2022) [Cs.RO], http://arxiv.org/abs/2211.01749.
[40] B.P. DeJong, J. Colgate, M. Peshkin, Mental Transformations in Human-Robot Interaction, 2011, pp. 35–51, https://doi.org/10.1007/978-94-007-0582-1_3.
[41] T.-C. Lin, A. Unni Krishnan, Z. Li, Perception-motion coupling in active telepresence: human behavior and teleoperation interface design, ACM Trans. Hum. Robot Interact. 12 (2023) 1–24, https://doi.org/10.1145/3571599.
[42] M. Talha, R. Stolkin, Preliminary evaluation of an orbital camera for teleoperation of remote manipulators, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 2754–2761, https://doi.org/10.1109/IROS40897.2019.8968218.
[43] D. Lee, G.-Y. Nie, K. Han, Automatic and real-time joint tracking and three-dimensional scanning for a construction welding robot, J. Constr. Eng. Manag. 150 (2024), https://doi.org/10.1061/jcemd4.coeng-14135.
[44] K.A. Szczurek, R.M. Prades, E. Matheson, J. Rodriguez-Nogueira, M. Di Castro, Multimodal multi-user mixed reality human–robot interface for remote operations in hazardous environments, IEEE Access 11 (2023) 17305–17333, https://doi.org/10.1109/ACCESS.2023.3245833.
[45] D. Wei, B. Huang, Q. Li, Multi-view merging for robot teleoperation with virtual reality, IEEE Robot. Autom. Lett. 6 (2021) 8537–8544, https://doi.org/10.1109/lra.2021.3109348.
[46] J. Betancourt, B. Wojtkowski, P. Castillo, I. Thouvenin, Exocentric control scheme for robot applications: an immersive virtual reality approach, IEEE Trans. Vis. Comput. Graph. 29 (2023) 3392–3404, https://doi.org/10.1109/TVCG.2022.3160389.
[47] A. Moniruzzaman, D. Rassau, S.M.S. Chai, Islam, teleoperation methods and enhancement techniques for mobile robots: a comprehensive survey, Robot. Auton. Syst. 150 (2022) 103973, https://doi.org/10.1016/j.robot.2021.103973.
[48] A. Valiton, Z. Li, Perception-action coupling in usage of telepresence cameras, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020, pp. 3846–3852, https://doi.org/10.1109/ICRA40945.2020.9197578.
[49] T.-C. Lin, A.U. Krishnan, Z. Li, Perception and action augmentation for teleoperation assistance in freeform telemanipulation, ACM Trans. Hum. Robot Interact. 13 (2024) 1–40, https://doi.org/10.1145/3643804.
[50] Q. Guo, Z. Yang, J. Xu, Y. Jiang, W. Wang, Z. Liu, W. Zhao, Y. Sun, Progress, challenges and trends on vision sensing technologies in automatic/intelligent robotic welding: state-of-the-art review, Robot. Comput. Integr. Manuf. 89 (2024) 102767, https://doi.org/10.1016/j.rcim.2024.102767.
[51] A. Rout, B.B.V.L. Deepak, B.B. Biswal, Advances in weld seam tracking techniques for robotic welding: a review, Robot. Comput. Integr. Manuf. 56 (2019) 12–37, https://doi.org/10.1016/j.rcim.2018.08.003.
[52] T. Lei, Y. Rong, H. Wang, Y. Huang, M. Li, A review of vision-aided robotic welding, Comput. Ind. 123 (2020), https://doi.org/10.1016/j.compind.2020.103326.
[53] X. Li, X. Li, M.O. Khyam, S.S. Ge, Robust welding seam tracking and recognition, IEEE Sensors J. 17 (2017) 5609–5617, https://doi.org/10.1109/jsen.2017.2730280.
[54] Y. Su, L. Lloyd, X. Chen, J.G. Chase, Latency mitigation using applied HMMs for mixed reality-enhanced intuitive teleoperation in intelligent robotic welding, Int. J. Adv. Manuf. Technol. 126 (2023) 2233–2248, https://doi.org/10.1007/s00170-023-11198-3.
[55] Q. Wang, Y. Cheng, W. Jiao, M.T. Johnson, Y. Zhang, Virtual reality human-robot collaborative welding: a case study of weaving gas tungsten arc welding, J. Manuf. Process. 48 (2019) 210–217, https://doi.org/10.1016/j.jmapro.2019.10.016.
[56] Y. Sharon, I. Nisky, Expertise, teleoperation, and task constraints affect the speed–curvature–torsion power law in RAMIS, J. Med. Robot. Res. 03 (2018) 1841008, https://doi.org/10.1142/s2424905x18410088.
[57] E.J. Lopez Pulgarin, O. Tokatli, G. Burroughes, G. Herrmann, Assessing tele-manipulation systems using task performance for glovebox operations, Front. Robot. AI 9 (2022) 932538, https://doi.org/10.3389/frobt.2022.932538.
[58] C. Pérez-D'Arpino, R.P. Khurshid, J.A. Shah, Experimental assessment of human–robot teaming for multi-step remote manipulation with expert operators, ACM Trans. Hum. Robot Interact. 13 (2024) 1–26, https://doi.org/10.1145/3618258.
[59] P. Praveena, L. Molina, Y. Wang, E. Senft, B. Mutlu, M. Gleicher, Understanding control frames in multi-camera robot telemanipulation, in: 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2022, pp. 432–440, https://doi.org/10.1109/hri53351.2022.9889543.
[60] S. Yoon, S. Shin, S. Lee, M. Park, C.R. Ahn, Evaluating viewpoint control techniques in virtual reality Interface for Teleoperating construction welding robots, in: B. Riveiro, P. Arias (Eds.), Proceedings of the 31st International Workshop on Intelligent Computing in Engineering, European Group for Intelligent Computing in Engineering (EG-ICE), Vigo, Spain, 2024, pp. 345–354, in: https://3dgeoinfoeg-ice.webs.uvigo.es/proceedings.
[61] K. Kawaharazuka, K. Okada, M. Inaba, Robotic constrained imitation learning for the peg transfer task in fundamentals of laparoscopic surgery, ArXiv (2024) [Cs.RO], http://arxiv.org/abs/2405.03440.
[62] D. Rakita, Methods for effective mimicry-based teleoperation of robot arms, in: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Association for Computing Machinery, New York, NY, USA, 2017, pp. 371–372, https://doi.org/10.1145/3029798.3034812.

[63] A. Althouse, Modern Welding, Goodheart-Wilcox Publisher, 2004.
[64] Industrial Hall Building Steel Structure 3d Model - TurboSquid 1591119, 2024. https://www.turbosquid.com/3d-models/industrial-hall-building-steel-structure-3d-model-1591119 (accessed June 21, 2024).
[65] S.G. Hart, L.E. Staveland, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research, in: P.A. Hancock, N. Meshkati (Eds.), Advances in Psychology, North-Holland, 1988, pp. 139–183, https://doi.org/10.1016/S0166-4115(08)62386-9.
[66] Brooke, SUS-A quick and dirty usability scale, in: Usability Evaluation in Industry, 1996, https://doi.org/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke.
[67] L. Ai, P. Kazanzides, E. Azimi, Mixed reality based teleoperation and visualization of surgical robotics, Healthc. Technol. Lett. 11 (2024) 179–188, https://doi.org/10.1049/htl2.12079.
[68] R. Xiao, C. Yang, Y. Jiang, H. Zhang, One-shot Sim-to-real transfer policy for robotic assembly via reinforcement learning with visual demonstration, Robotica 42 (2024) 1074–1093, https://doi.org/10.1017/s0263574724000092.