

Article

Mixed Reality-Assisted Multimodal Teleoperation System for Mobile Robotic Arms

Chen Zhang ¹, Chengjian He ¹, Bo Wang ², Le Zheng ¹, Jin Zhao ¹ and Guangwei Wang ^{1,*}

¹ School of Mechanical Engineering, Guizhou University, Guiyang 550025, China

² Department of Mechanical Engineering, The City College of New York, New York, NY 10031, USA

* Correspondence: gwwang@gzu.edu.cn

How To Cite: Zhang, C.; He, C.; Wang, B.; et al. Mixed Reality-Assisted Multimodal Teleoperation System for Mobile Robotic Arms. *Adv. Mechatron.* **2026**, *1*(1), 6.

Publication History

Received: 25 June 2026

Revised: 20 August 2026

Accepted: 22 September 2026

Published: 29 September 2026

Abstract: Teleoperation is essential for manipulation in hazardous or inaccessible environments, but is often limited by reduced perception, limited control flexibility, and low operator immersion. To overcome these issues, we present a mixed-reality (MR)-assisted teleoperation system for mobile robotic arms. The system integrates stereoscopic vision via a head-mounted display and an augmented-reality (AR) interface providing real-time feedback and safety alerts. A multimodal control strategy further supports complex maneuvers. In the experiments, participants achieved a 92.42% success rate in object retrieval and placement, and success rates in narrow corridor tasks were 84.08% with conventional displays and 89.41% under MR-based collaborative control. These findings suggest potential usability benefits for the remote operation of mobile robotic arms.

Keywords: teleoperation; mobile robotic arm; multimodal control

1. Introduction

The growing demand for tasks in hazardous environments such as firefighting, disaster management, and inspection of power infrastructure has heightened the need to protect human operators from direct exposure [1]. Mobile robots equipped with robotic arms have been widely adopted due to their maneuverability and mobility [2–4]. Complex tasks may still require human intervention [5], despite advances in fault-tolerant and safety-critical control of mobile robots [6,7]. Human–robot collaboration combines robotic capabilities with human cognitive abilities, providing a promising approach to complex tasks [8]. In this paradigm, operators provide real-time guidance and monitoring, allowing robots to dynamically adapt to various requirements, thereby enhancing operational efficiency and reducing error rates.

To tackle the challenges of increasingly complex mission scenarios, various control frameworks have been proposed. A task priority-based control framework facilitating collaborative movement between a mobile base and robotic arm in teleoperation is introduced in [9]. A hybrid motion telecontrol method for specific collaborative mobile robots using weighted Cartesian impedance control is developed in [10]. In [11], a finite state machine approach is used to create a modular ROS-based system, enabling autonomous state transitions based on

predefined conditions to reduce operator workload. However, this method lacks feedback mechanisms and does not ensure operational safety. Moreover, these approaches do not consider operators' limited access to the robot's environment in remote settings, rely on simple decision-making processes, and lack rapid response to complex events. Hence, establishing a highly immersive and multimodal remote operation mechanism that prioritizes safety and enhances the efficiency of human-robot collaboration (HRC) is essential [12].

Visual feedback is also crucial in teleoperation, providing operators with direct operational information and enhancing sensory experience. Many studies focus on improving visual feedback in teleoperation tasks [13,14]. Traditional teleoperation methods rely on 2D interfaces like liquid crystal displays (LCDs), offering a third-person perspective of the environment. However, these methods lack immersion and may cause operator discomfort [15]. To overcome these limitations, there is a growing interest in simulating natural visual systems that enable depth perception. Virtual reality (VR) environments have been developed to study human–robot interaction in assistive robotics [16]. Whitney et al. [17] integrated a VR system with ROS, allowing operators to remotely control robotic arms for gripping tasks using VR devices. Stotko et al. [18] developed an immersive operating system for inaccessible contaminated areas, capturing and reconstructing environmental information



in 3D using a depth camera, facilitating immersive remote operations. While VR enhances immersion in human-robot interaction, these methods primarily focus on virtual representations of remote environments. In contrast, capturing natural objects through stereo vision via video streaming offers greater convenience in estimating distances and object sizes, effectively meeting task accuracy requirements.

Additionally, the complexity of the work environment significantly affects safe teleoperation. Challenges may arise due to the presence of shelves, stacks, narrow aisles, and corridors. When operators use head-mounted displays (HMDs), the user interface is linked to the camera's field of view, and a lack of a comprehensive environmental overview can hinder effective control of the robot's posture. AR can enhance the experience with stereoscopic devices [19]. For instance, Vogel et al. [20] proposed a projection and camera approach projecting a rectangular safety space on the ground as visual guidance. Hietanen et al. [21] designed a shared space model for human-robot cooperation, presenting the robot workspace in real time via 2D projection and 3D HMD AR glasses to enhance safety. Papanastasiou et al. [22] introduced an AR-based system for collaborative assembly, visually alerting the operator when the robot operates in hazardous areas. MR-based interfaces have been investigated for remote control of mobile manipulators [23], while AR-based eye-gaze interfaces have enabled hands-free mobile robot operation [24].

This study focuses on integrating stereoscopic visual feedback with arm-extension warnings and switchable collaborative and supervisory control for mobile manipulation in confined spaces. The main contributions of this paper are:

- (i) Providing highly immersive stereoscopic images to remote users using VR technology, and integrating AR to deliver early warning information about the robot arm's posture, enhancing operator situational awareness.
- (ii) Implementing a human-robot collaborative control strategy that leverages the operator's cognitive advantages to monitor and manage robot movements, processes, and components, thereby supporting task execution and improving the operational flexibility and usability of remote manipulation.
- (iii) Demonstrating the practicality of the developed system through verification experiments involving robotic arm teleoperation tasks.

The remainder of the paper is organized as follows: Section 2 introduces the framework of the proposed teleoperation system; Section 3 details the control strategy and human-robot interaction interface; Section 4 presents feasibility and performance results; and Section 5 concludes the paper.

2. Framework of Multimodal Teleoperation System

In teleoperation, operators often face limited situational awareness, hindering effective manipulation and interaction with the robot and its environment. Our system aims to establish a highly immersive multimodal teleoperation mecha-

nism that enhances the operator's perceptual range and adaptability to complex environments. By integrating multidimensional visual information—including monocular and stereo cameras—we expand the operator's perceptual range. A natural hand-guided motion control system enables intuitive control of the robotic arm, enhancing operational adaptability. The use of stereoscopic display improves object detection and grasping, increasing operator immersion. Additionally, AR technology provides real-time warning information about the robotic arm, empowering operators to effectively understand and respond to the environment, thus improving overall performance and decision-making capabilities.

The system architecture is depicted in Figure 1, illustrating the collaborative relationship between the human operator and the mobile robotic arm. The system consists of two primary components: the master side and the slave side. The master side is based on the HTC VIVE PRO2 virtual reality device, including an HMD, controllers, and Base Stations. The slave side is developed around the SMART mobile robotic arm, which includes an EAI differential drive chassis and a UFACTORY xArm 6-axis robotic arm. The robotic arm is equipped with an electric gripper and integrates an Intel RealSense D435 camera and a monocular camera for visual information acquisition, as shown in Figure 2. The mobile chassis is equipped with an industrial computer, LiDAR, IMU, and ultrasonic sensors for environmental perception, path planning, motion control, and data transmission. The master and slave sides communicate via a wireless local-area network (WLAN), enabling remote control of the robotic arm and environmental feedback.

Master Side: It provides an immersive operating environment for the operator and transmits control commands to the slave side. The operator can monitor the environment and interact with the robotic arm using intuitive hand movements through the HTC VIVE PRO2. To enhance immersion and manipulation precision, the MR visualization interface is designed using the Vuforia engine in Unity 3D. This interface presents stereoscopic video from the robot-mounted camera through the HMD, allowing for a more realistic experience. Additionally, the status of the robotic arm is displayed as AR warning information, enabling the operator to quickly identify potential issues.

Slave Side: It executes remote operation commands and provides feedback to the master side. To enhance perception of the surrounding environment, both binocular and monocular cameras are integrated to capture visual information. The mobile platform has basic autonomous capabilities based on ROS, receiving pose and action commands from the controller and transforming them into end-effector poses and actions in Cartesian coordinates. The gripping action of the electric gripper is controlled by receiving trigger commands from the VIVE controller. Visual data is read using *mjpg-streamer* and converted into video streams, which are streamed via ROS topics through WLAN to the master side GUI for display.

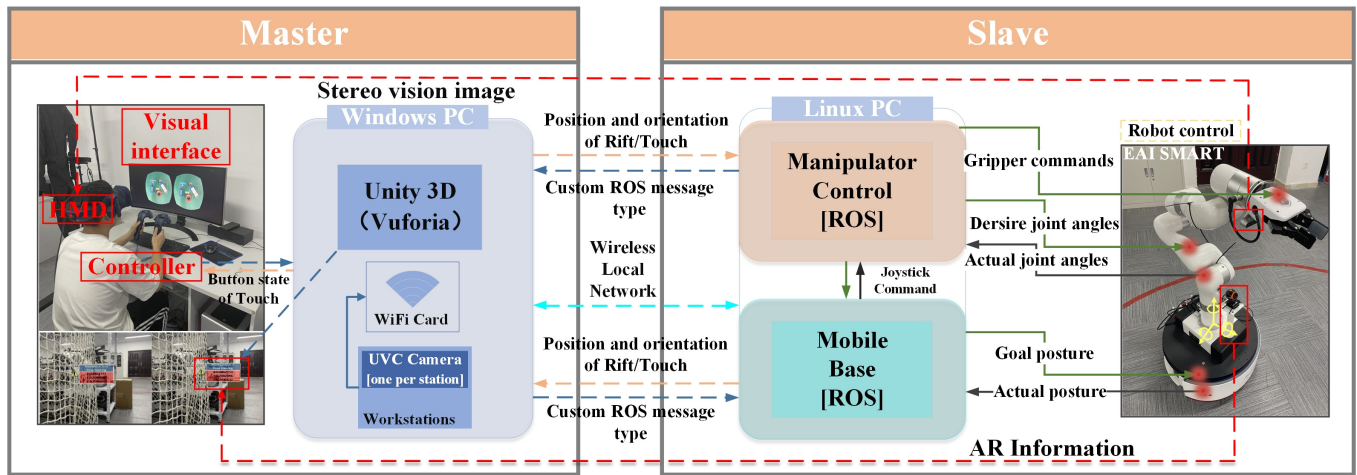


Figure 1. Architecture of the proposed MR-based multimodal assistive teleoperation system. Dashed and solid lines indicate wireless and wired connections, respectively, while arrows indicate the transmission of visual feedback, robot-state information, and control commands.

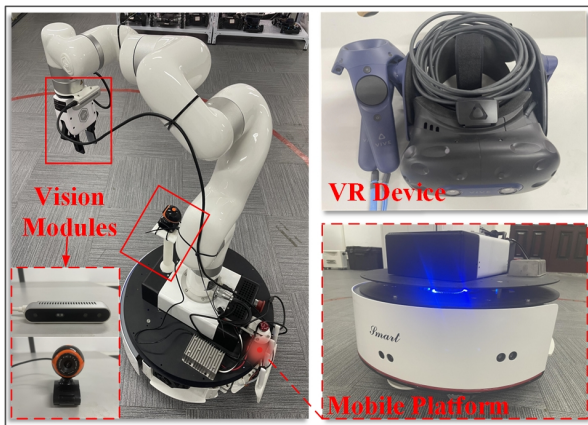


Figure 2. Hardware components of the system: Mobile manipulator platform (left), monocular camera and RealSense D435 camera (bottom-left), HTC VIVE HMD and VR controllers (top-right), EAI differential drive chassis (bottom-right).

3. Design of Multimodal Teleoperation System

The multimodal interaction implemented in the current system mainly consists of stereoscopic visual feedback through the head-mounted display, AR-based textual and color-coded warning information, and robot motion and task commands provided through the VR controller. These visual feedback and control modalities are further integrated with collaborative and supervisory control modes to support remote operation under different task conditions. This section presents the implemented multimodal teleoperation functions and human-robot interaction interface, with the objective of improving environmental perception and operational adaptability for remote operators. Auditory and force feedback are not included in the current prototype and will be considered as extensions in future work.

3.1. Multimodal Teleoperation

To facilitate operator supervision and adapt to complex task scenarios, the system is designed with two control modes: collaborative control mode and supervisory control mode. The

collaborative control mode focuses on complex and unstructured tasks in unknown environments, requiring operator input at the action level to ensure flexibility and safety. The supervisory control mode is suitable for structured environments where full operator involvement is unnecessary. In this mode, the operator leverages cognitive advantages to monitor and manage robotic movements and processes, supporting teleoperation performance and flexible operator supervision.

Figure 3 illustrates the implemented human-robot collaboration paradigms that assist operators in effectively performing manipulation tasks. In collaborative control mode, the robot autonomously handles coarse operations while the operator is responsible for overall control and error correction in finer movements. The VR controller enables precise control of both the mobile base and the robotic arm. In supervisory control mode, a state machine control allows the robot to autonomously govern navigation and manipulation while the operator provides macro-level supervision. For example, during a task, the mobile robotic arm initially operates autonomously, moving towards the target position and gripping objects in a specific sequence. If issues arise, the operator can swiftly switch to collaborative control mode to address complex problems. This switch facilitates operator supervision and enables flexible responses to intricate scenarios.

3.1.1. Collaborative Control Mode

The remote manipulation module enhances ease of use, as demonstrated in Figure 4. The system maps the VR controller's buttons to control both the mobile chassis and the robotic arm, providing comprehensive control. By pulling the trigger on the controller, the operator can quickly switch between collaborative and supervisory control modes.

This operation mode divides control into robot movement control and robot arm control, using the same control button to switch between them. In robot movement control, directional buttons are used to move the chassis, and the menu button adjusts the robot's speed for precise positioning. In robot arm control mode, the controller manipulates each joint individu-

ally. The control panel and system menu buttons control the six joints of the robotic arm, and the side grip button controls the gripper. The system menu button also executes the robotic arm motion planner. To facilitate complex manipulation, the operator can adjust the value and duration of the controller's pressing force in real time with visual feedback.

3.1.2. Supervisory Control Mode

The SMACH (State Machine) enables autonomous control of the mobile robotic arm by managing the states of manipulation, navigation, and other common tasks. Figure 5 illustrates this supervisory control mode. When a task is initiated, the robot automatically carries out navigation and manipulation in the order specified by the predefined state machine, alternating between manipulation and navigation modes.

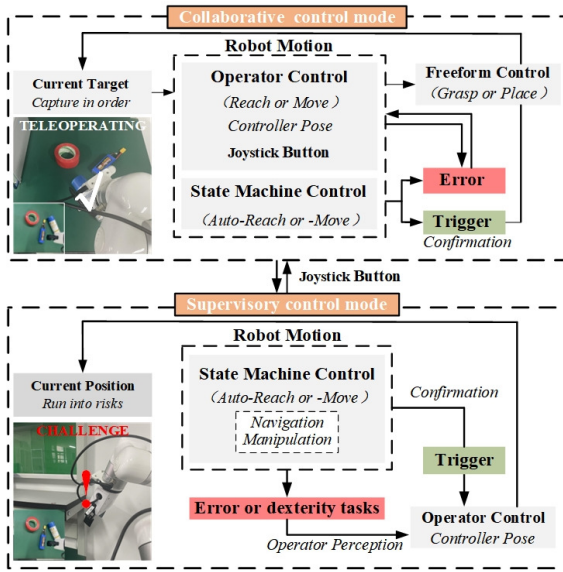


Figure 3. Framework of multimodal teleoperation system for assisted remote robot manipulation.

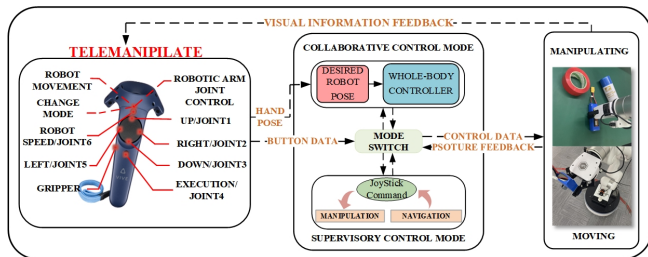


Figure 4. Details of the remote manipulation module.

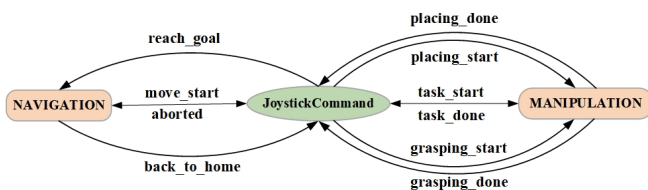


Figure 5. State-machine task architecture under the supervisory control mode.

During state transitions, the system autonomously addresses potential issues such as inaccurate position estimation or object detection failures. Once errors are resolved, the system transitions to a safe state to continue task execution. This handling increases robot fault tolerance while reducing the need for continuous manual intervention during task execution. The operator can monitor movements using visual feedback, maximizing cognitive advantages and ensuring effectiveness while maintaining safety and flexibility.

Navigation and Manipulation: The navigation module uses the following state and observation models:

$$x_k = f(x_{k-1}, u_k, w_k), \quad \forall k \in \mathbb{Z}^+ \quad (1)$$

$$z_k = h(y_k, x_k, v_k), \quad \forall k \in \mathbb{Z}^+ \quad (2)$$

where x_k denotes the robot state at time step k , y_k represents the environment map information, u_k is derived from IMU and odometry data, and z_k is obtained from LiDAR measurements. The terms w_k and v_k represent process and measurement noise, respectively. The SLAM problem is then formulated as $P(x, y | z, u)$, where x and y denote the robot states and environment map, respectively.

The manipulation task focuses on the precise grasping and placing of objects. In the camera orientation state, the robotic arm moves the D435 camera to capture images and point clouds. The YOLOv5 algorithm aids in target detection. The estimated grasping posture is calculated by processing the point cloud of the target object, using the RANSAC algorithm to filter noise, and applying the method proposed in [25].

Operational Task Flow: Upon initiating a task, the operator assesses the task type. If it involves repetitive tasks in structured environments, the system defaults to supervisory control mode. For intricate operations, the operator uses collaborative control mode. The adaptive operation task under state machine control consists of six sub-states, as shown in Figure 6. The robot sets initial parameters in the preprocessing state, including camera posture and target numbers. It then performs camera positioning, object detection, and grasp posture estimation. If the grasp posture is correct, the state machine transitions to the object grasp and place state. Each task includes grasping and placing postures. The system checks the task queue upon completion; if not empty, a new task is created. In case of errors or the need for dexterity, the operator can enter real-time commands via the VR controller to cancel tasks.

3.2. Multimodal Human-Robot Interactive Interface

3.2.1. Stereoscopic Vision

To improve operator situational awareness and control accuracy, stereoscopic video technology processes the environment in 3D and presents it through an HMD. A depth camera captures left and right video streams, wirelessly transmitted to the host computer. To reduce bandwidth requirements during real-time transmission, video streams are compressed using H.264 encoding before RTP transmission. The host computer decodes the video using the FFmpeg library and renders it into

a format displayable by the HMD via the graphics card. The video transmission framework is illustrated in Figure 7. The

real-time stereoscopic display during target grasping is shown in Figure 8.

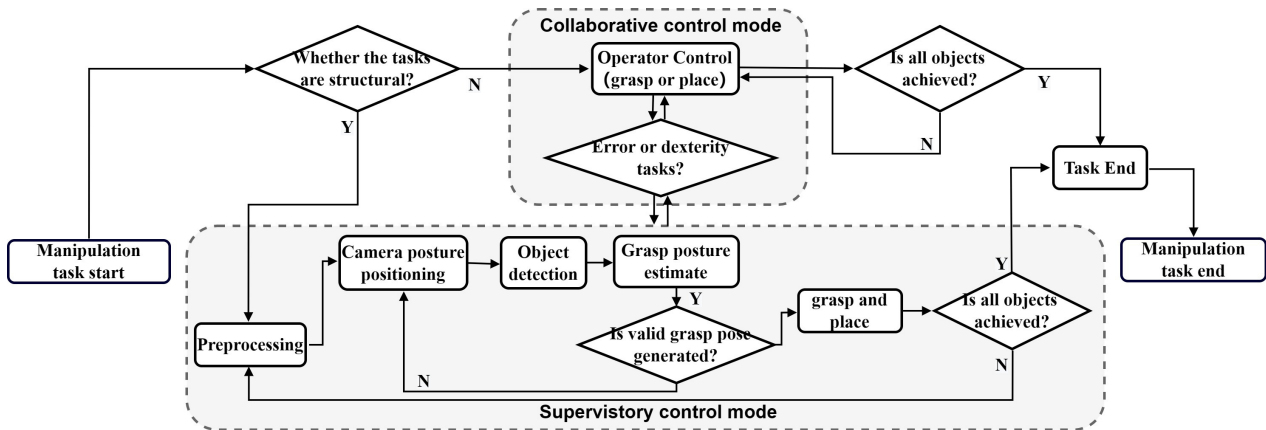


Figure 6. State-machine adaptive operation flowchart. Arrows indicate task-state transitions and switching between supervisory and collaborative control.

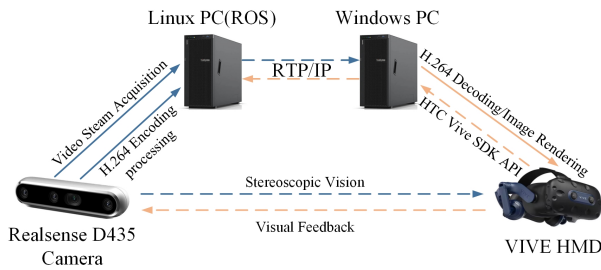


Figure 7. Video transmission framework.

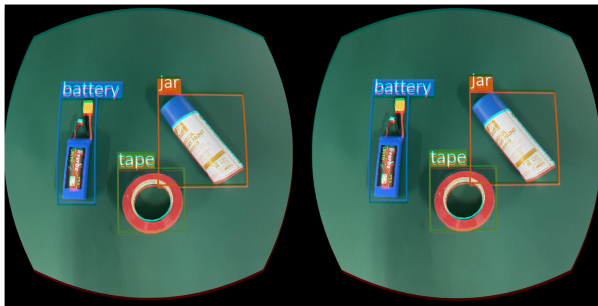


Figure 8. Stereoscopic display during target grasping.

3.2.2. Interface Design for AR Warning Messages

Augmented reality enhances the user experience by mapping a virtual information box to the stereoscopic display interface, using the area target method for registration. Target areas for warnings focus on narrow aisle areas with complex environments. These targets are simulated using the Vuforia Creator in Unity 3D, creating a laboratory environment with five target areas, as illustrated in Figure 9.

A custom external camera driver captures images and matches features to the real-world viewpoint of the robotic arm. AR info boxes are designed and placed according to the virtual coordinate system. A real-time connection between the ROS system and Unity 3D maps the transformation frames of the robotic arm's end-effector to the AR objects, updating the

3D coordinates in real time, as shown in Figure 10.

To improve operators' perception of the robotic arm's posture in visually restricted environments, we designed a graded hazard warning system around the mobile robotic arm. The operational space of the arm's end effector is divided into safety zones with different colors based on distance. A custom program on the ROS platform dynamically modifies the AR info box color. Table 1 shows the relationship between the distance of the end effector and the AR warning colors. Here, α denotes the extension distance of the robotic-arm end effector relative to the reference point of the robotic-arm base, which is used to characterize the extension of the manipulator during operation. For the experimental platform used in this study, the monitored range of 0–0.762 m was divided into three levels at intervals of 0.254 m, corresponding to the green, yellow, and red warning regions, respectively, to provide graded posture warnings to the operator.

Operators wearing the HMD can view a Mechanical Arm Pose Warning displayed on the MR interface, presented in different colors as the robot moves into predefined hazardous areas, as shown in Figure 11. This alert helps operators efficiently adjust the arm pose as necessary, establishing a bidirectional communication system between the physical robot and virtual data.



Figure 9. Simulated laboratory environment with target areas.

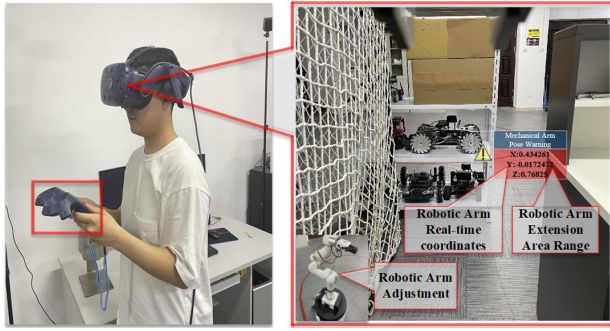


Figure 10. Real-time mapping of robotic arm frames to AR objects.

Table 1. Zones of robotic arm end-effector locations.

Region	Gap (m)	Color
Area 1	$0 \leq \alpha < 0.254$	Green
Area 2	$0.254 \leq \alpha < 0.508$	Yellow
Area 3	$0.508 \leq \alpha \leq 0.762$	Red

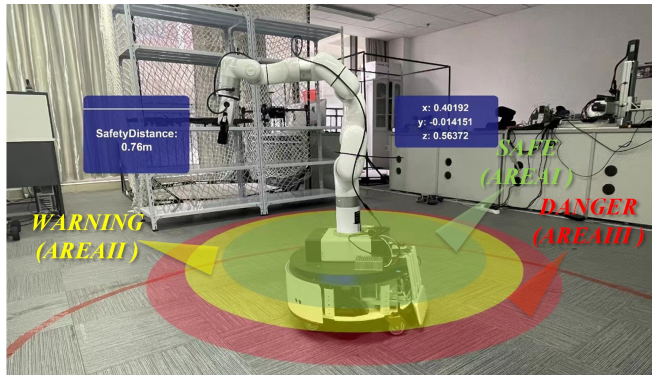


Figure 11. Diagram of the robotic arm's extension areas.

4. Experimental Validation and Performance Evaluation

We evaluated the performance of the supervisory control mode through autonomous navigation, target collection, and delivery tasks. We also evaluated task performance and perceived situational awareness in the collaborative control mode. Throughout the mission, the operator utilized the developed teleoperation system and interacted with the visual feedback interface using VIVE controllers and visual feedback from multiple cameras.

4.1. Experimental Setup

Participants: A total of 15 volunteers participated in the study, including 10 men and 5 women aged 22–30 years. None of the participants had previous practical experience with the teleoperation system. All 15 participants took part in the autonomous navigation test, the narrow-corridor passing test, and the user acceptance survey, while a subset of 10 participants participated in the target collection and delivery test. All participants had normal or corrected-to-normal vision and were in good health. Before the formal experiments, the experimental procedure was thoroughly explained to the participants, followed by standardized training on control commands and safety procedures.

Implementation: The teleoperation system includes a computer running the reconstruction client and the server component, and the SMART mobile robotic arm acting as the teleoperation client. The robotic arm is equipped with an NVIDIA Jetson AGX Xavier; the desktop computer has an Intel Core i7-9700K CPU, 32GB RAM, and an NVIDIA GeForce RTX 2070 GPU. Data visualization is performed using the HTC Vive HMD with a native resolution of 2448×2448 pixels per eye. Both computers were connected via WiFi during all experiments.

4.2. Navigation Test under Supervisory Control Mode

Under the supervisory control mode, the participants initiated the navigation task and monitored the robot through the visual feedback interface, while the robot autonomously navigated from point A to point B according to the predefined goal position. The participants did not continuously provide directional or path commands during autonomous navigation. Obstacles were arranged along the navigation route, as illustrated in Figure 12. Four obstacle-layout conditions were tested (Figure 13), and each participant attempted the task once for each condition. Success was defined as the robot avoiding all obstacles.

All participants successfully managed the robot's autonomous navigation without collisions, achieving a 100% success rate. Table 2 presents the average performance under the different obstacle conditions, further demonstrating the task performance of the autonomous navigation module.

4.3. Target Collection and Delivery Test under the Supervisory Control Mode

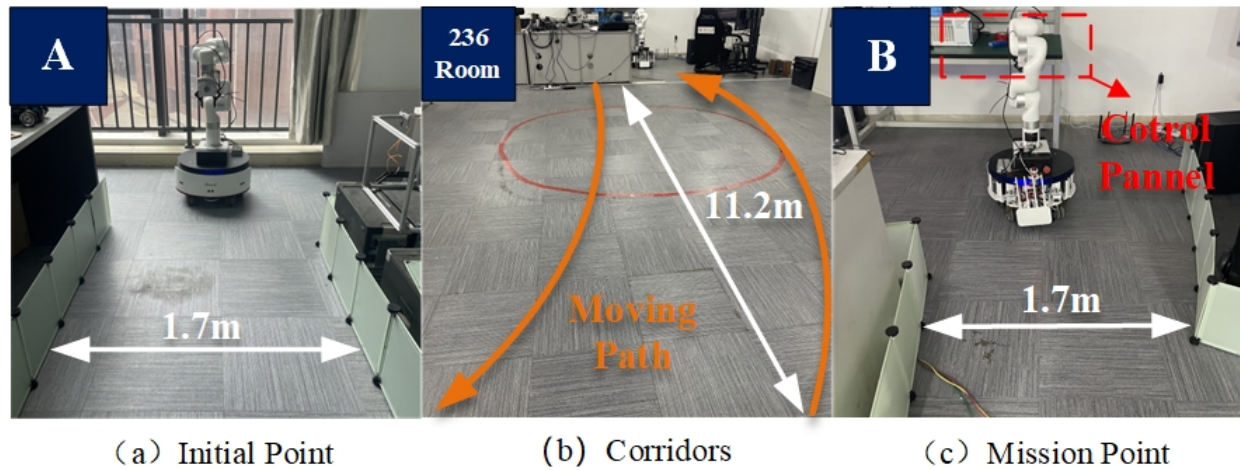
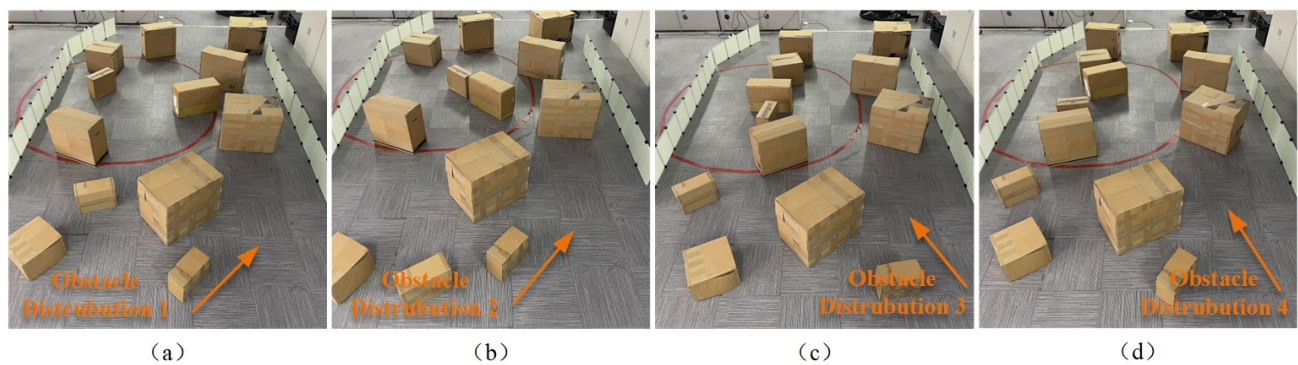
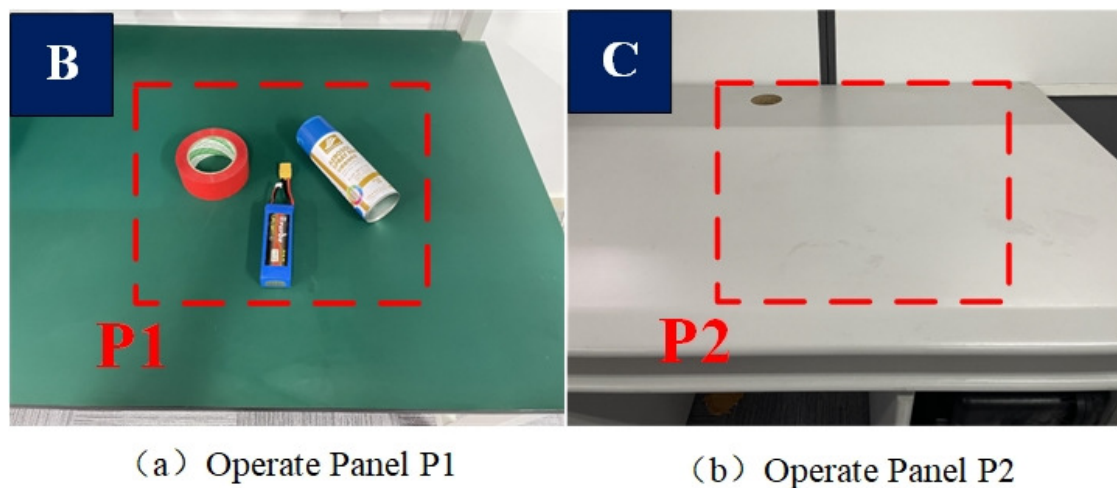
This experiment was conducted to evaluate the task performance of the robot in collecting and delivering targets under the supervisory control mode. The target objects used in the experiment included batteries, spray cans, and adhesive tapes. Three specific tasks that involved grasping, delivering, and placing these objects were tested. The experimental procedure was as follows: The target collection and delivery experiment was conducted with a subset of 10 participants selected from the 15 volunteers described in Section 4.1. During the experiment, the mobile manipulator autonomously collected and delivered three types of target objects—batteries, spray cans, and adhesive tapes—from control panel P_1 at location B to control panel P_2 at location C according to the commands issued by the participants. The operating panels are depicted in Figure 14.

The results were divided into two cases: success and failure. Success was defined as the successful grasping of target objects and their delivery to the control panel P_2 . Failure was determined if the target objects were not recognized, if they were not grasped successfully, or if they were dropped during the delivery process. The task execution process is illustrated in Figure 15.

The results of the experiment are displayed in Table 3. The system was able to collect and deliver the target objects with a task success rate of 92.42% under supervisory control. Moreover, when it was in autonomous mode, the system achieved a success rate of over 90%, which is consistent with the expectations of the state machine control strategy.

Table 2. Average performance of autonomous navigation.

Route	Success Rate (%)	Avg. Distance (m)	Avg. Time (s)	Avg. Speed (m/s)
Case 1	100	16.32	48.37	0.3374
Case 2	100	16.43	49.58	0.3314
Case 3	100	16.54	49.67	0.3330
Case 4	100	16.76	49.75	0.3369

**Figure 12.** Predetermined navigation route with obstacles.**Figure 13.** Four obstacle layouts used in the autonomous navigation experiments: (a) Case 1; (b) Case 2; (c) Case 3; and (d) Case 4.**Figure 14.** Schematic diagram of the operating panel.

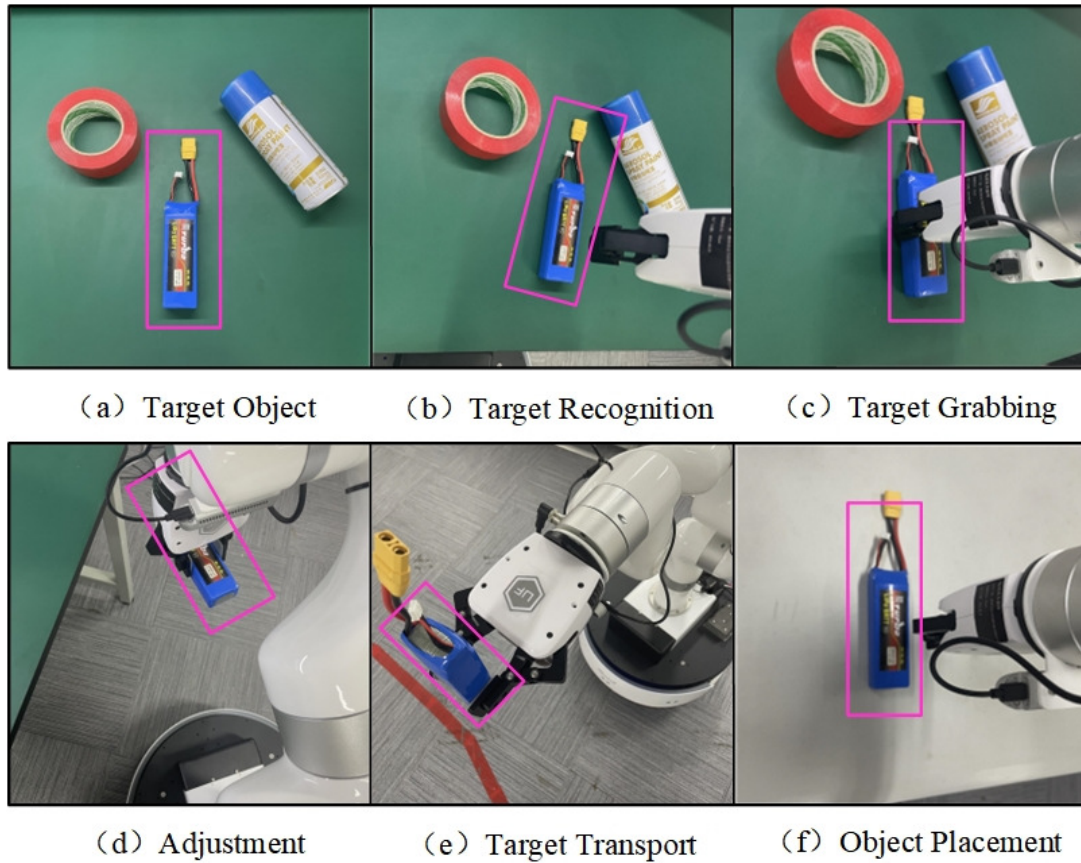


Figure 15. The process of task execution.

Table 3. Average performance of target collection and delivery.

Target Objects	Success Rate (%)	Failure Rate (%)
battery	94.18%	5.82%
can	92.21%	7.79%
tape	90.87%	9.13%
overall	92.42%	7.58%

4.4. Narrow Corridor Passing Test under the Collaborative Control Mode

To evaluate the effectiveness of the proposed MR-based operator interface in improving situational awareness and facilitating complex maneuvering in collaborative control mode, an experiment was conducted in a controlled laboratory environment. The experiment involved 15 participants, and different conditions were tested in the narrow corridor setting. The objective of the experiment was for the participants to operate the robot from position P_1 through the narrow DE corridor to position P_2 for target collection. The widths of the narrow corridors were set incrementally from 0.6 to 1.2 m, as illustrated in Figure 16.

Three manipulation modes were used: autonomous operation, 2D video, and MR-based operator interface. After reaching position P_2 and collecting the target, participants were required to return to position P_1 and place the target in a similar manner. The experiment results were classified as either success or failure. The task was considered a fail-

ure if one of the following conditions occurred: (a) failure to place the target correctly or (b) contact with an obstacle during movement. In contrast, the task was considered successful if none of these conditions occurred. The initial position of the robotic arm was consistently set at the same location before each experiment. It is important to note that for safety reasons, if there was a risk of the arm colliding with an obstacle, it was quickly stopped by pressing the emergency button. Figure 17 illustrates the operating screen of the eighth randomly selected participant while navigating to the target position D using the MR-based operation interface.

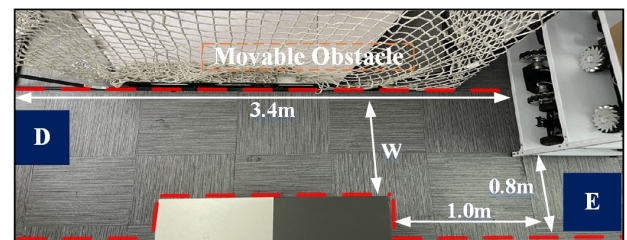
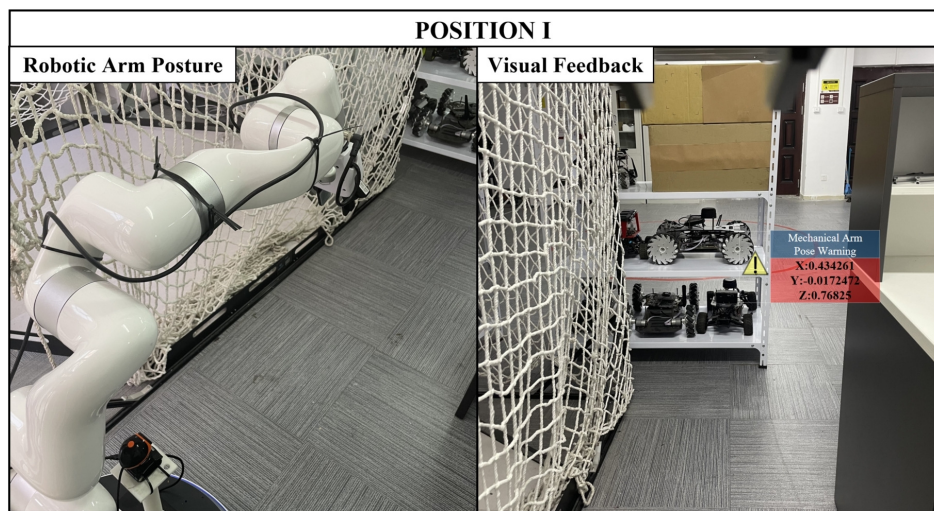
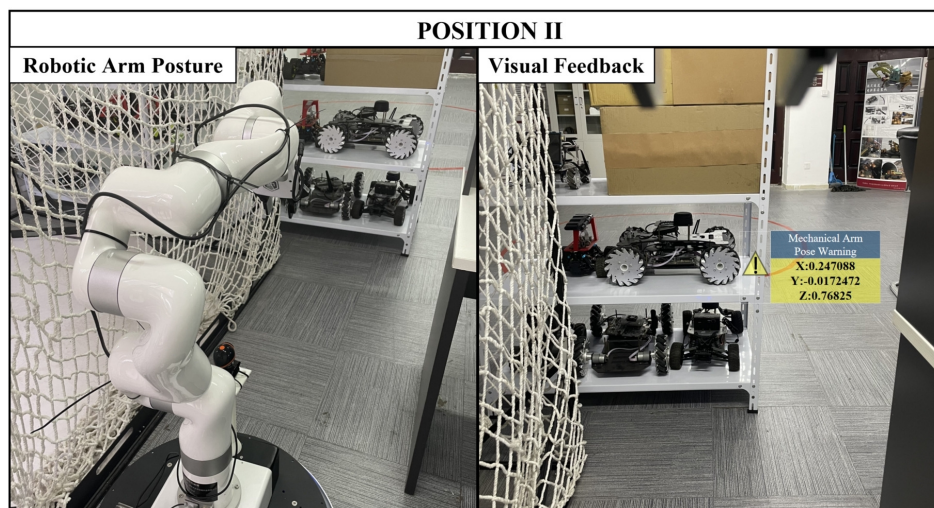


Figure 16. The constructed narrow corridor map environment.

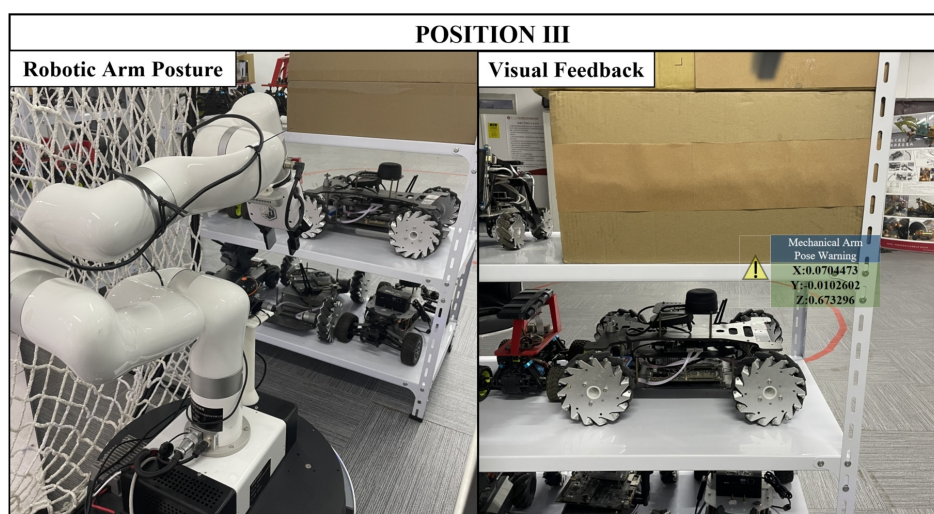
The right side of the screen presents AR warning information on the MR interface, including the current Cartesian coordinates of the end-effector and color-coded warnings related to the extension range. On the left side of the screen, the recording shows the participant's effective adjustment of the robotic arm's posture throughout the process.



(a) An obstacle is in front of the robotic arm, and the arm is extended (the AR information box shows red); the arm's posture should be adjusted promptly.



(b) The robotic arm is closer to an obstacle; the operator continues to adjust the robotic arm's attitude until the AR information box turns yellow.



(c) The robotic arm approaches the obstacle and adjusts its attitude again until the AR information box turns green and then turns in place, successfully passing through the narrow corridor.

Figure 17. Actual operation of the passing test in the narrow corridor.

In this experiment, each participant was asked to perform multiple maneuvers over a period of two days, involving different maneuver patterns and narrow lane widths. The results of these tasks are presented in Figure 18. For the 0.6-m narrow corridor experiment, the task success rates for the three methods were as follows: 82.67% for the MR-based operator interface, 73.83% for the 2D video mode, and 67% for the autonomous mode. One possible explanation for the observed differences in success rates among the three modes is that when the mobile robot enters a narrow lane in autonomous mode, if the extension range of the robotic arm is too extensive, self-adjustment is not timely, resulting in a higher likelihood of collision with obstacles during autonomous steering, leading to task failure. In 2D video mode, although manual real-time adjustments can be made, visual feedback in this mode lacks depth perception, and the grasped target object can visually obscure the view of the actuator at the end of the robotic arm, making it difficult to effectively adjust the movements of the robotic arm. This can result in the target falling or colliding with obstacles.

As the width of the narrow path increases, the success rates of the three modes exhibit an upward trend, and the difference between them gradually decreases, with mean success rates exceeding 88% at a width of 1.2 m.

It is important to note that when using the MR-based operator interface, participants were able to effectively adjust the posture of the robotic arm based on their situational awareness of the operating environment. They consistently maintained a high success rate throughout the experiment, with an average success rate of 89.41%, which is higher than that of the 2D video mode, which achieved an average success rate of 84.08%. These results suggest that the MR interface may help operators maintain situational awareness.

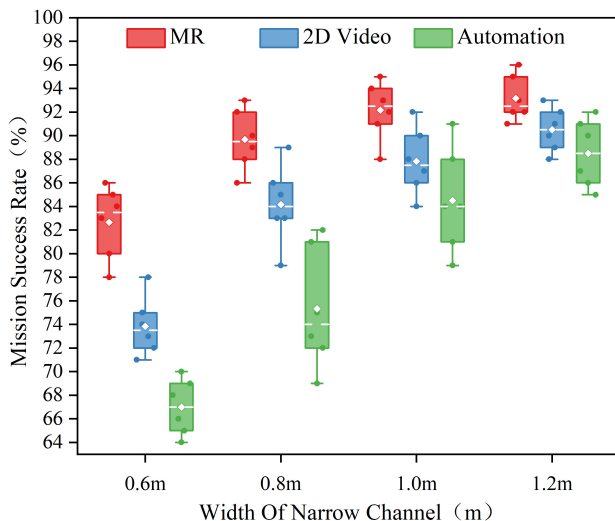


Figure 18. Mission test results. Numerical values of task success in the narrow corridor environment using different modes of operation for various participants are recorded in the results (marked with ●), upper and lower limits, upper and lower quartiles, and means (marked with ◇), and median line (marked with —).

4.5. User Acceptance Survey

At the end of the experiment, all participants completed a 7-point Likert-scale questionnaire to subjectively evaluate their experience with the system. The questionnaire consisted of six items rated from -3 to $+3$, where -3 represented “strongly disagree”, 0 represented “neutral”, and $+3$ represented “strongly agree”. The survey is presented in Table 4. After completing the tasks under the 2D video and MR-based modes, the participants rated the two operation modes separately using the same six questionnaire items. The testing order of the two operation modes was counterbalanced across participants to reduce potential order effects.

As shown in Figure 19, participants generally agreed that the MR-based operator interface proposed in this paper is more intuitive and immersive than traditional 2D video. Participants reported benefits in several aspects of teleoperation, although the 2D video mode received higher ratings for ease of view control. These subjective ratings suggest potential benefits of the MR-based interface for perceived situational awareness and complex teleoperation tasks. It should be noted that all experiments were conducted in a controlled laboratory environment using a wireless local-area network; therefore, the current results mainly reflect the system performance under short-range and relatively stable communication conditions. Complex outdoor environments, long-distance communication, and network latency or fluctuations may affect the real-time visual feedback and control responsiveness, which will be further evaluated in future work.

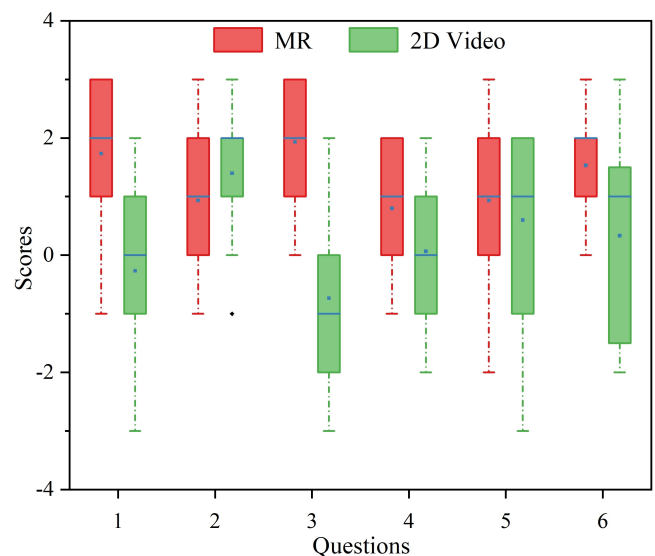


Figure 19. User survey study. For each question in the questionnaire, mean scores, upper and lower limits, upper and lower quartiles (including interquartile range), outliers (marked with ·), and medians (marked with —) were recorded. In most respects, the MR-based operation interface presented in this paper scored higher on the 7-point Likert scale than the 2D video-based method.

Table 4. User acceptance survey.

NO.	Survey Items	Scores
1	Easy to stay situationally aware?	−3 ~ +3
2	Easy to control the views?	−3 ~ +3
3	Ease of access to information?	−3 ~ +3
4	Easy to perform complex teleoperation tasks?	−3 ~ +3
5	More direct coordination of robot movement and manipulation?	−3 ~ +3
6	Integrated user immersion experience?	−3 ~ +3

5. Conclusions

This paper has presented a novel multimodal assistive teleoperation system for mobile robotic arms, integrating mixed reality (MR) technologies with collaborative and supervisory control modes. A stereoscopic virtual reality (VR)-based interface enables intuitive pose control, while augmented reality (AR)-based feedback provides real-time posture warnings. Experiments with novice operators indicated favorable task performance and perceived usability of the proposed system, with higher questionnaire ratings for perceived situational awareness and immersion during complex teleoperation tasks. These findings highlight the system's potential to extend the accessibility and safety of teleoperation in hazardous or remote scenarios. Future work will focus on mitigating communication delays and incorporating additional sensory modalities, such as force and audio feedback, to further broaden the applicability of multimodal teleoperation.

Author Contributions

C.Z.: conceptualization, methodology, software, investigation, data curation, writing—original draft preparation; C.H.: methodology, validation, investigation; B.W.: software, validation; L.Z.: investigation, data curation; J.Z.: validation, writing—reviewing and editing; G.W.: conceptualization, supervision, project administration, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China, grant number 52265070, the Guizhou Provincial Science and Technology Projects, grant number 2024100 MS(2026)083, and the Zhuhai Industry-Academia Cooperation R&D Project, grant number 23 200 0400 2732.

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Human Medical Experimental Ethics Committee of Guizhou University (Approval No. HMEEC-GZU-2026-7011; 3 March 2026).

Informed Consent Statement

All participants were informed of the experimental procedures and relevant safety requirements and voluntarily provided informed consent before participating in the study.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used ChatGPT to assist with language polishing and improving the clarity of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Purushottam, A.; Xu, C.; Jung, Y.; et al. Dynamic Mobile Manipulation via Whole-Body Bilateral Teleoperation of a Wheeled Humanoid. *IEEE Robot. Autom. Lett.* **2024**, *9*, 1214–1221. <https://doi.org/10.1109/lra.2023.3334677>.
2. Li, X.; Yue, H.; Yang, D.; et al. A Large-Scale Inflatable Robotic Arm toward Inspecting Sensitive Environments: Design and Performance Evaluation. *IEEE Trans. Ind. Electron.* **2023**, *70*, 12486–12499.
3. Wang, B.; Nersesov, S.G.; Ashrafiuon, H. Formation Regulation and Tracking Control for Nonholonomic Mobile Robot Networks Using Polar Coordinates. *IEEE Control Syst. Lett.* **2022**, *6*, 1909–1914. <https://doi.org/10.1109/lcsys.2021.3135753>.
4. Wang, B.; Ashrafiuon, H.; Nersesov, S.G. Leader–Follower Formation Stabilization and Tracking Control for Heterogeneous Planar Underactuated Vehicle Networks. *Syst. Control Lett.* **2021**, *156*, 105008. <https://doi.org/10.1016/j.sysconle.2021.105008>.
5. Moniruzzaman, M.; Rassau, A.; Chai, D.; et al. Teleoperation Methods and Enhancement Techniques for Mobile Robots: A Comprehensive Survey. *Robot. Auton. Syst.* **2022**, *150*, 103973.
6. Guo, B.; Dian, S.; Zhao, T.; et al. Dynamic Event-Driven Neural Network-Based Adaptive Fault-Attack-Tolerant Control for Wheeled Mobile Robot System. *ISA Trans.* **2023**, *140*, 71–83. <https://doi.org/10.1016/j.isatra.2023.06.010>.
7. Han, T.; Wang, B. Safety-Critical Stabilization of Force-Controlled Nonholonomic Mobile Robots. *IEEE Control Syst. Lett.* **2024**, *8*, 2469–2474. <https://doi.org/10.1109/lcsys.2024.3492999>.
8. Ajoudani, A.; Zanchettin, A.M.; Ivaldi, S.; et al. Progress and Prospects of the Human-Robot Collaboration. *Auton. Robots* **2018**, *42*, 957–975. <https://doi.org/10.1007/s10514-017-9677-2>.
9. Teng, T.; Fernandes, M.; Gatti, M.; et al. Whole-Body Control on Non-holonomic Mobile Manipulation for Grapevine Winter Pruning Automation. In Proceedings of the 2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM), Chongqing, China, 3–5 July 2021; pp. 37–42. <https://doi.org/10.1109/icarm52023.2021.9536083>.
10. Wu, Y.; Lamon, E.; Zhao, F.; et al. Unified Approach for Hybrid Motion Control of MOCA Based on Weighted Whole-Body

- Cartesian Impedance Formulation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3505–3512. <https://doi.org/10.1109/lra.2021.3062316>.
11. Luo, R.C.; Lee, S.L.; Wen, Y.C.; et al. Modular ROS Based Autonomous Mobile Industrial Robot System for Automated Intelligent Manufacturing Applications. In Proceedings of the 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Boston, MA, USA, 6–9 July 2020; pp. 1673–1678. <https://doi.org/10.1109/aim43001.2020.9158800>.
 12. Aoki, J.; Sasaki, F.; Yamashina, R.; et al. Teleoperation by Seamless Transitions in Real and Virtual World Environments. *Robot. Auton. Syst.* **2023**, *164*, 104405. <https://doi.org/10.1016/j.robot.2023.104405>.
 13. Szczurek, K.A.; Prades, R.M.; Matheson, E.; et al. Multimodal Multi-User Mixed Reality Human-Robot Interface for Remote Operations in Hazardous Environments. *IEEE Access* **2023**, *11*, 17305–17333. <https://doi.org/10.1109/access.2023.3245833>.
 14. Aguirre, O.A.; Ñacato, J.C.; Andaluz, V.H. Virtual Simulator for Collaborative Tasks of Aerial Manipulator Robots. In Proceedings of the 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), Sevilla, Spain, 24–27 June 2020; pp. 1–6. <https://doi.org/10.23919/cisti49556.2020.9141092>.
 15. Naceri, A.; Mazzanti, D.; Bimbo, J.; et al. The Vicarios Virtual Reality Interface for Remote Robotic Teleoperation. *J. Intell. Robot. Syst.* **2021**, *101*, 80. <https://doi.org/10.1007/s10846-021-01311-7>.
 16. Bustamante, S.; Peters, J.; Schölkopf, B.; et al. ArmSym: A Virtual Human-Robot Interaction Laboratory for Assistive Robotics. *IEEE Trans. Hum.-Mach. Syst.* **2021**, *51*, 568–577. <https://doi.org/10.1109/thms.2021.3106865>.
 17. Whitney, D.; Rosen, E.; Ullman, D.; et al. ROS Reality: A Virtual Reality Framework Using Consumer-Grade Hardware for ROS-Enabled Robots. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–9. <https://doi.org/10.1109/iros.2018.8593513>.
 18. Stotko, P.; Krumpfen, S.; Schwarz, M.; et al. A VR System for Immersive Teleoperation and Live Exploration with a Mobile Robot. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 3630–3637. <https://doi.org/10.1109/IROS40897.2019.8968598>.
 19. Fidalgo, C.G.; Yan, Y.; Cho, H.; et al. A Survey on Remote Assistance and Training in Mixed Reality Environments. *IEEE Trans. Vis. Comput. Graph.* **2023**, *29*, 2291–2303. <https://doi.org/10.1109/tvcg.2023.3247081>.
 20. Vogel, C.; Walter, C.; Elkmann, N. Safeguarding and Supporting Future Human-robot Cooperative Manufacturing Processes by a Projection- and Camera-based Technology. *Procedia Manuf.* **2017**, *11*, 39–46. <https://doi.org/10.1016/j.promfg.2017.07.127>.
 21. Hietanen, A.; Pieters, R.; Lanz, M.; et al. AR-Based Interaction for Human-Robot Collaborative Manufacturing. *Robot. Comput.-Integr. Manuf.* **2020**, *63*, 101891. <https://doi.org/10.1016/j.rcim.2019.101891>.
 22. Papanastasiou, S.; Kousi, N.; Karagiannis, P.; et al. Towards Seamless Human Robot Collaboration: Integrating Multimodal Interaction. *Int. J. Adv. Manuf. Technol.* **2019**, *105*, 3881–3897. <https://doi.org/10.1007/s00170-019-03790-3>.
 23. Li, Q.; Sun, M.; Song, Y.; et al. Mixed Reality-Based Brain Computer Interface System Using an Adaptive Band-pass Filter: Application to Remote Control of Mobile Manipulator. *Biomed. Signal Process. Control* **2023**, *83*, 104646. <https://doi.org/10.1016/j.bspc.2023.104646>.
 24. Lee, J.; Lim, T.; Kim, W. Investigating the Usability of Collaborative Robot Control Through Hands-Free Operation Using Eye Gaze and Augmented Reality. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Detroit, MI, USA, 1–5 October 2023; pp. 4101–4106. <https://doi.org/10.1109/IROS55552.2023.10342045>.
 25. Sekkat, H.; Tigani, S.; Saadane, R.; et al. Vision-Based Robotic Arm Control Algorithm Using Deep Reinforcement Learning for Autonomous Objects Grasping. *Appl. Sci.* **2021**, *11*, 7917. <https://doi.org/10.3390/app11177917>.