Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Phantom Hand: Research on Dual-Arm Robotic Fine Teleoperation Based on VR Immersive Interaction and Adaptive Shared Control

**Mocun Yang**

jef_mocun_yang@163.com

The Second High School Affiliated to Capital Normal University

**Research Article**

**Keywords:** dual-arm teleoperation; virtual reality; wearable IMU; motion reconstruction; quaternion; shared control

**Posted Date:** November 9th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11611893/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:**
Consent for publication Written informed consent was obtained from the author's legal guardian for the author's participation in this study and for the publication of this manuscript, including any images and test data in which the author appears as the test participant. The author also provided assent to participate and to the publication of these materials.

参赛学生姓名：杨默存

中学：北京师范大学第二附属中学

省份：北京

国家/地区：中国

指导老师姓名：张瑜

指导老师单位：北京师范大学第二附属中学

论文题目：Phantom Hand: Research on Dual-Arm Robotic Fine Teleoperation Based on VR Immersive Interaction and Adaptive Shared Controlal

# Phantom Hand: Research on Dual-Arm Robotic Fine Teleoperation Based on VR Immersive Interaction and Adaptive Shared Control

Mocun Yang

**Abstract:** Operations in chemical accident sites, nuclear facilities, and other hazardous environments may expose personnel to toxic substances, radiation, high temperatures, or unstable structures. Teleoperation offers a way to keep the operator at a safe distance, but controlling two robotic arms naturally remains difficult because multiple joints, grippers, and viewing directions must be coordinated at the same time. This study develops Phantom Hand, a dual-arm robotic teleoperation system that combines wearable inertial sensing with virtual reality interaction. Five IMUs worn on the chest, upper arms, and forearms capture the operator's upper-body motion. The reconstructed arm movements are mapped to the two robotic arms, while a VR headset controls the remote camera and two handheld controllers operate the grippers.

Several practical problems arise when human motion is transferred directly to the robot. The five IMUs share one RS-485 bus and are read sequentially, so their measurements are aligned using timestamps and short-term quaternion extrapolation. Natural-pose calibration and a chest-centered reference frame are used to reduce mounting errors and the influence of torso rotation. Swing–twist decomposition is then applied to separate coupled shoulder and forearm rotations. The control system also checks data validity, joint limits, command velocity, and communication status. Valid commands follow the operator's motion, whereas unsafe or incomplete commands are constrained or held.

Experiments showed that the average sampling-time span across the five IMUs was 112.43 ms. After calibration, spatial compensation, and temporal alignment were applied, the mean motion crosstalk decreased from 18.7% to 5.4%, while static jitter decreased from 0.684° to 0.241°. The robot also completed a grasping-and-transfer sequence under combined arm, gripper, and visual control. These results show that the proposed system can support natural and fine-grained dual-arm teleoperation using relatively low-cost sensing and interaction hardware.

**Keywords:** dual-arm teleoperation; virtual reality; wearable IMU; motion reconstruction; quaternion; shared control

# Table of Contents

1 Introduction ..... 5

1.1 Research Background and Significance ..... 5

1.2 Research Status ..... 6

1.3 Existing Methods and Challenges Faced by the System ..... 7

2 System Design ..... 9

2.1 Overall System Architecture ..... 9

2.2 Mechanical Structure and Actuator Configuration ..... 10

2.3 Multi-IMU Orientation Sensing and Wearable Configuration ..... 13

2.4 VR Interaction and Human-to-Robot Functional Mapping ..... 14

2.5 Control Computer and Hardware Connections ..... 15

2.6 Software Workflow and Adaptive Shared-Control Architecture ..... 16

3 Five-IMU Pose Reconstruction and Robotic-Arm Mapping Method ..... 17

3.1 Multi-IMU Acquisition and Temporal Alignment ..... 17

3.2 Natural-Pose Calibration and Chest-Reference Transformation ..... 19

3.3 Bilateral Joint-Angle Extraction ..... 20

3.3.1 Shoulder Flexion and Abduction ..... 21

3.3.2 Upper-Arm Axial Rotation ..... 21

3.3.3 Elbow Flexion and Forearm Rotation ..... 22

3.4 Human-to-Robot Mapping and Synchronized Control ..... 23

3.4.1 Joint Calibration and Mapping ..... 23

3.4.2 Filtering and Command Constraints ..... 23

3.4.3 Synchronized Dual-Arm Transmission ..... 24

4 VR Interaction and System Implementation ..... 25

4.1 Headset-Controlled Camera Following ..... 25

4.2 Controller-Based Dual-Gripper Control ..... 27

4.3 State-Adaptive Shared Control and Safety Mechanisms ..... 29
5 Experiments and Results Analysis ..... 30
5.1 Experimental Setup and Protocol ..... 31
5.1.1 Experimental Platform ..... 31
5.1.2 Experimental Procedure ..... 32
5.1.3 Evaluation Metrics ..... 32
5.2 Pose Reconstruction Results ..... 32
5.2.1 Sampling Timing and Static Stability ..... 32
5.2.2 Single-Joint Response and Motion Crosstalk ..... 33
5.3 Ablation Results ..... 34
5.4 Grasping Experiment ..... 35
6 Conclusions and Future Work ..... 36
6.1 Summary ..... 36
6.2 Future Work ..... 36
References ..... 37
Acknowledgments ..... 39
Appendix A: Bill of Materials ..... 41

# 1 Introduction

## 1.1 Research Background and Significance

Hazardous chemical operations, disaster response, and nuclear facility maintenance often expose personnel to toxic gases, high temperatures, radiation, oxygen deficiency, and unstable environments. Studies have shown that accident severity is influenced by facility layout, risk isolation, emergency-response capability, time pressure, and operational coordination [1–3]. Therefore, teleoperated robots that perform observation, grasping, and object-handling tasks can reduce direct human exposure while preserving human judgment and operational flexibility.

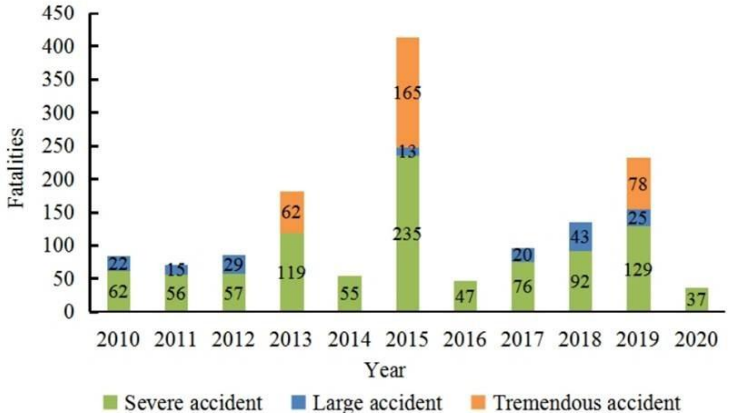
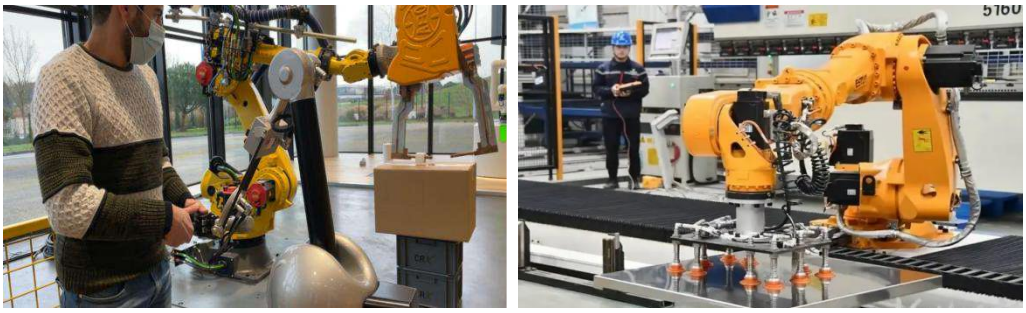



Figure 1-1 Fatalities Caused by Different Levels of Hazardous Chemical Accidents in China from 2010 to 2020

In high-risk or inaccessible environments, robots can undertake physical contact and hazardous exposure while human perception, judgment, and decision-making remain in the remote-control loop. This division of roles provides the practical motivation for teleoperation.



(a) Dual-Arm Robotic Manipulation (b) Robot-Assisted Surgery



(c) Remote Industrial Robot Operation (d) High-Risk Equipment Maintenance

Figure 1-2 Typical Application Scenarios of Robot Teleoperation in Hazardous Environments

Teleoperation transmits human control commands to a remote robot while returning robot and environmental information to the operator. Communication delay can disrupt the synchronization of commands and feedback, increase operator workload, and reduce control performance [4,5]. These challenges are more significant in dual-arm systems, where operators must coordinate multiple joints, two end effectors, and the remote viewpoint. Previous research has therefore explored force feedback, integrated perception and control, ergonomic interfaces, and delay-compensation methods to improve teleoperation performance [6–9].

Against this background, this study develops a dual-arm teleoperation system integrating five wearable IMUs with VR interaction. The IMUs reconstruct bilateral upper-limb motion, the VR headset controls the remote camera, and two controllers independently operate the grippers. The system focuses on reducing sequential-sampling misalignment, mounting-orientation errors, torso-motion interference, shoulder-motion coupling, and human–robot structural differences, thereby establishing a complete control chain from natural human motion to dual-arm robot execution.

## 1.2 Research Status

Early teleoperation systems mainly relied on joysticks, teach pendants, keyboards, and master manipulators. Although these interfaces provide clear control boundaries, their operation becomes less intuitive as robot degrees of freedom increase. Recent research has therefore introduced VR, motion capture, gesture recognition, and wearable inertial sensors to improve interaction naturalness.

**(1) VR and immersive teleoperation.**

VR integrates head-pose tracking, controller input, and immersive visual feedback. Previous studies have demonstrated VR-based navigation and robotic-arm control [10],

while augmented-reality research has shown that virtual–physical registration and information overlays can improve operators' understanding of robot states and environments [11]. Functional separation of motion and auxiliary control can further reduce operational burden [12]. Controller-based interaction, gesture recognition, and hybrid motion mapping have also been investigated [13–16]. However, controllers or gestures alone cannot continuously capture complete torso, upper-arm, and forearm orientations.

**(2) Cooperative dual-arm teleoperation.**

Dual-arm robots can stabilize, manipulate, or transport objects cooperatively. Existing studies have investigated learning-based dual-arm grasping and visually guided manipulation [17,18]. However, bilateral operation increases control dimensionality and synchronization requirements. Human-motion-driven systems must therefore capture both arms accurately while addressing mirrored joint configurations, command synchronization, and mechanical joint limits.

**(3) IMU-based upper-limb pose reconstruction.**

Wearable IMUs are compact, inexpensive, resistant to visual occlusion, and suitable for continuous motion acquisition. Previous studies have used inertial sensors to estimate shoulder and elbow angles [19], reconstruct seven-degree-of-freedom upper-limb motion [20], and estimate upper-body pose through multi-sensor fusion [21]. These studies confirm the feasibility of IMU-based motion reconstruction but also identify sensor mounting errors, relative-segment calibration, magnetic interference, and drift as important sources of error.

**(4) Quaternion-based orientation processing.**

Quaternions represent three-dimensional rotations continuously and avoid the gimbal-lock problem associated with Euler angles. Quaternion averaging supports multi-frame pose calibration [22], while spherical interpolation provides a mathematical basis for smooth orientation interpolation and short-term extrapolation [23]. Methods such as VQF further demonstrate how bias estimation and magnetic-disturbance rejection can improve IMU orientation robustness [24]. These studies provide the theoretical basis for the quaternion calibration and temporal-alignment methods used in this study.

Shared control combines human intent with machine supervision [13]. Here, the operator commands the task, motion, and grasp, while a rule-based controller limits or holds unsafe or invalid inputs.

## 1.3 Existing Methods and Challenges Faced by the System

The system addresses four main challenges.

(1) Multi-IMU temporal misalignment. Sequential RS-485 polling gives the five IMUs different sampling times, so response timestamps and quaternion extrapolation are used for alignment [24].

(2) Mounting deviations and torso interference. Natural-pose calibration and a chest-centered reference compensate for fixed sensor offsets and whole-torso rotation [20–22].

(3) Shoulder-motion coupling. Quaternion processing and swing–twist decomposition separate limb direction from axial rotation and reduce Euler-angle crosstalk [23,24].

(4) Human–robot structural differences. Independent zero positions, direction coefficients, mapping ratios, and limits are required for each joint and mirrored arm.

# 2 System Design

## 2.1 Overall System Architecture

An immersive interface with functionally separated control channels can reduce mode-switching demands and improve operators' understanding of the remote workspace [10–12]. Accordingly, five IMUs provide fine-grained bilateral arm commands, two VR controllers operate the grippers, and the VR headset controls the camera viewpoint. These human inputs form the intent layer of the adaptive shared-control architecture.

The system consists of operator, control, and robot sides. The operator side supplies continuous arm pose, grasping intent, and viewing direction. The Raspberry Pi acts as the supervisory layer by reconstructing pose, checking data validity, applying joint and velocity constraints, selecting tracking or holding behavior, and scheduling communication. The robot side executes the resulting safe targets, while the camera stream closes the visual feedback loop.

**human input → pose computation → robot execution → visual feedback**

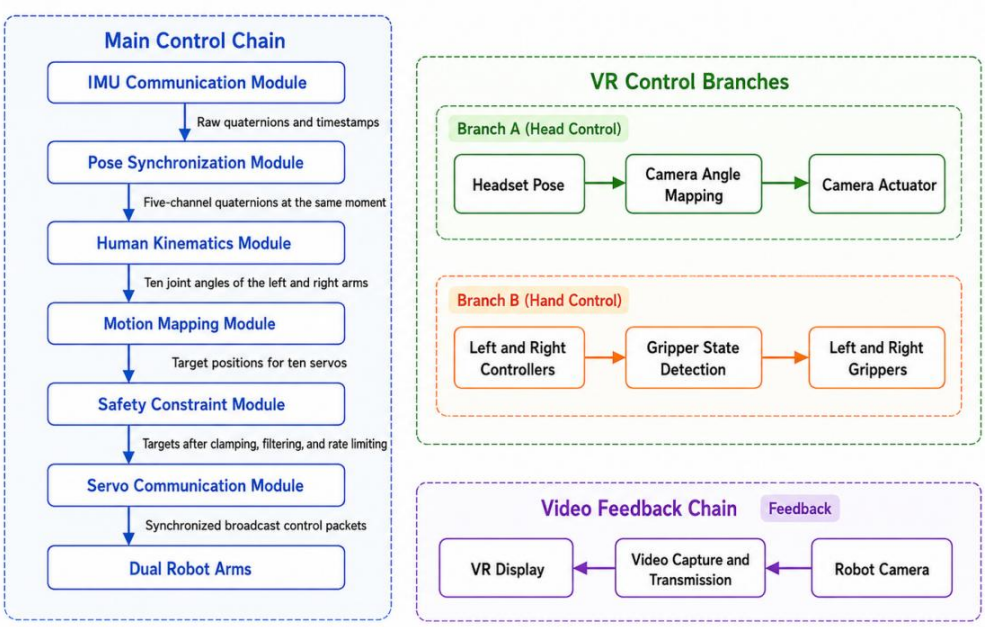



Figure 2-1 Overall System Architecture and Data Flow

At the functional level, the main control chain converts five quaternion streams into target positions for the two robotic arms. Separate VR branches control the head viewpoint

and the left and right grippers, while the video link returns images from the robot to the operator.

Processing continuous body motion separately from discrete grasping actions helps reduce unintended interactions between control channels and follows the principle of assigning different input devices according to task function [12].

## 2.2 Mechanical Structure and Actuator Configuration

The robot adopts a left-right symmetric dual-arm structure.

Each arm contains six actuated joints. One HLS3950 servo drives shoulder flexion/extension, corresponding to the main forward/backward motion of the arm, while the remaining five joints are driven by STS3215-C018 servos, including shoulder abduction/adduction, upper-arm internal/external rotation, elbow flexion/extension, forearm rotation, and wrist-related motion. Five joints of each arm are controlled by reconstructed human motion; the remaining wrist-related joint is maintained at a predefined position during the present experiments.

At the distal end of each arm, one additional servo controls gripper opening and closing, and another controls lateral swinging of the hand. The robot head uses two servos to provide yaw and pitch motion, retaining only two rotational degrees of freedom.

The complete system therefore uses 18 servos, including 2 HLS3950 servos and 16 STS3215-C018 servos.

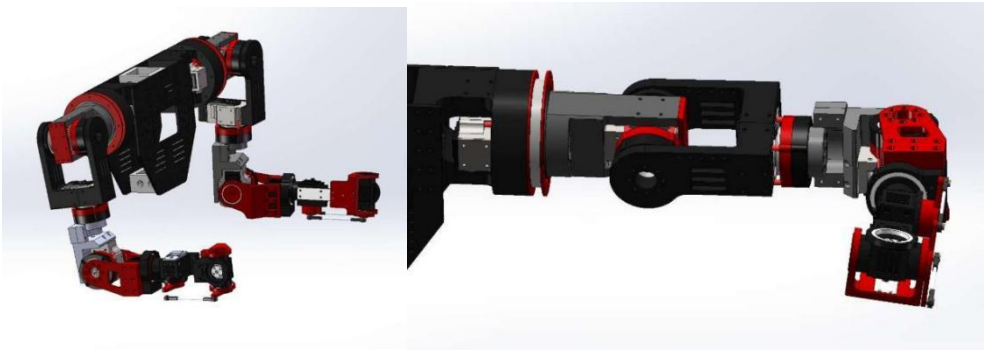


Figure 2-2 Dual-Arm 3D Model Motion Demonstration

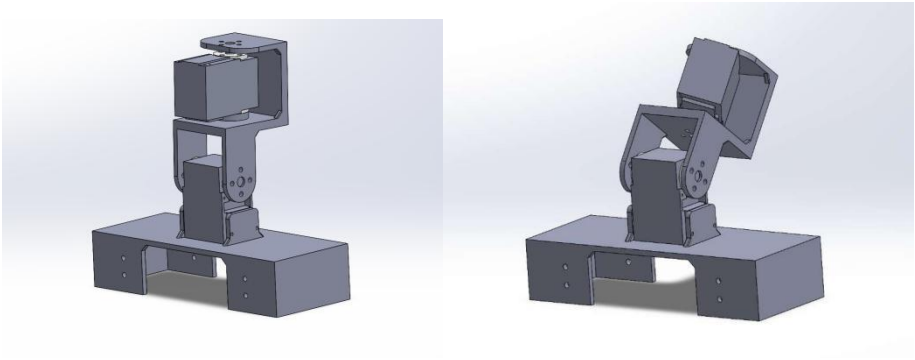


Figure 2-3 Two-DOF Head 3D Model Motion Demonstration



Figure 2-4 Structural Components and Actuator Modules

The structural components are fabricated using a combination of three-dimensional modeling and 3D printing. Modular brackets, connecting rings, and servo mounts can be combined according to the joint axes, allowing adjustment of servo orientation, installation spacing, and mechanical zero positions.

The mirrored structure of the two arms provides a basis for bimanual coordination. However, because the installation directions of corresponding servos on the left and right sides may be opposite, the control program cannot simply apply identical encoder commands to both sides. Instead, the direction coefficient, zero position, and motion limits must be defined separately for each joint.

The HLS3950 provides greater stall torque and is therefore suitable for the relatively highly loaded forward/backward arm joints located close to the robot torso.

The STS3215-C018 is smaller and lighter, making it suitable for the remaining distributed actuator positions, including the other arm joints, grippers, wrists, and head.

During cooperative dual-arm transportation, the two end effectors must maintain their relative positions and orientations. Therefore, the symmetry of the mechanical structure, synchronized servo updates, and joint constraints jointly influence coordination

performance [17][18].




Figure 2-5 Dual-Arm Mechanical Structure and Servo Allocation

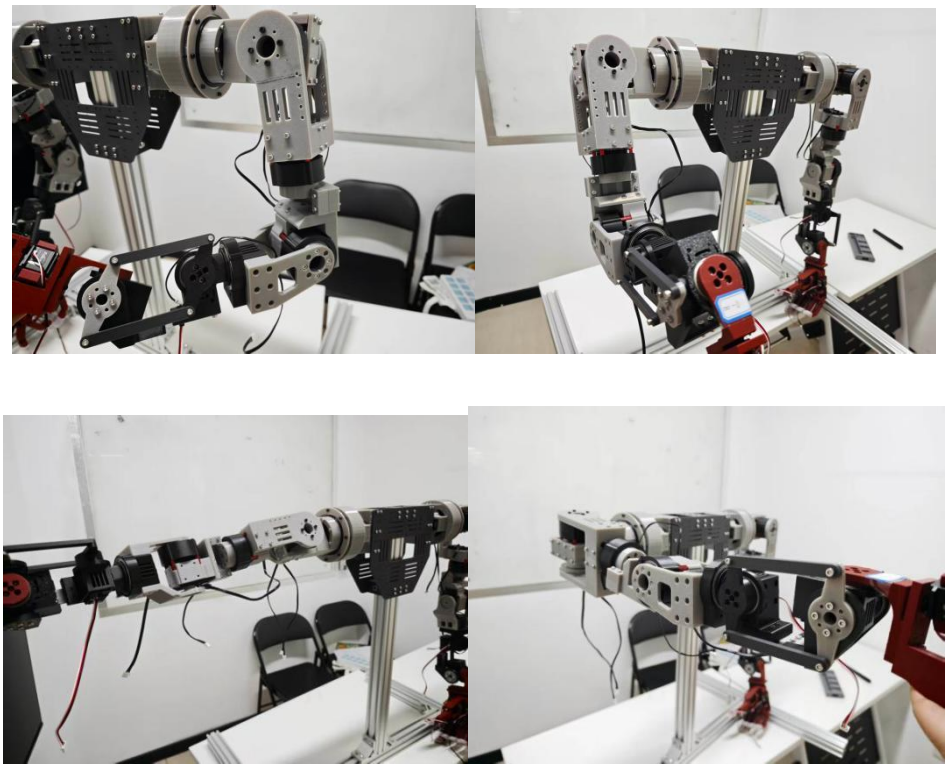


Figure 2-6 Physical Dual-Arm Motion Demonstration

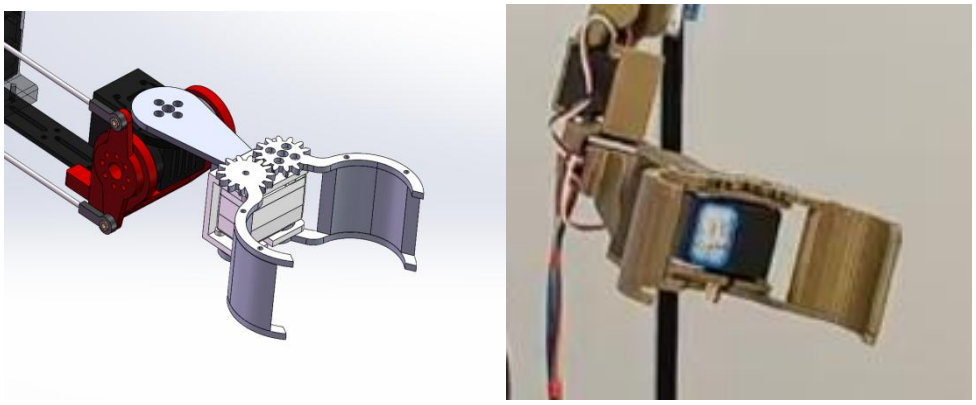


Figure 2-7 Structural Design of the Hand

## 2.3 Multi-IMU Orientation Sensing and Wearable Configuration

The human-motion sensing subsystem uses five WT901C-485 sensors mounted on the chest, left upper arm, left forearm, right upper arm, and right forearm.

The upper-arm and forearm sensors are used to reconstruct shoulder motion, elbow motion, and forearm rotation, while the chest sensor establishes a local body reference frame to compensate for the influence of whole-body turning on arm-joint calculations.

Reconstructing upper-limb joint motion using the relative orientations of adjacent body segments is a common approach in wearable inertial motion capture [19]–[21].

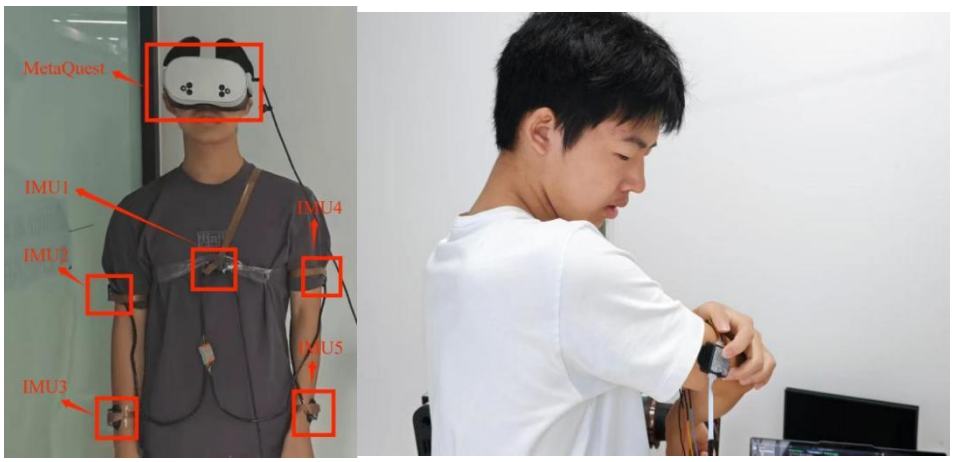



Figure 2-8 Five-IMU Wearing Positions and Sensor Attachment

Table 2-1 Installation Positions and Functions of the Five IMUs

| IMU ID | Installation Position | Function |
|---|---|---|

| | | |
|---|---|---|
| IMU 1 | Torso/Chest | Establishes the local body reference frame and compensates for overall torso rotation. |
| IMU 2 | Left Upper Arm | Measures left shoulder movement and upper-arm axial rotation. |
| IMU 3 | Left Forearm | Measures left elbow flexion/extension and forearm pronation/supination. |
| IMU 4 | Right Upper Arm | Measures right shoulder movement and upper-arm axial rotation. |
| IMU 5 | Right Forearm | Measures right elbow flexion/extension and forearm pronation/supination. |

The five IMUs have independent addresses and share a single RS-485 bus. Sharing one bus reduces the number of communication interfaces and the amount of wearable-side wiring. However, the Raspberry Pi must query the sensors sequentially by address, meaning that the five sensor measurements obtained within one acquisition cycle correspond to different sampling times.

The system therefore records a timestamp for each response and subsequently performs temporal alignment through quaternion-increment extrapolation.

During natural-standing calibration, multiple quaternion frames are averaged to obtain a stable initial orientation. This approach avoids the periodic-boundary problems that can arise from directly averaging Euler angles [22].

## 2.4 VR Interaction and Human-to-Robot Functional Mapping

The VR interaction subsystem uses a Meta Quest headset together with left and right controllers.

The headset provides yaw and pitch changes relative to its initial orientation, which are used to control the two servos in the robot head. The left and right controllers independently generate opening and closing commands for the left and right grippers.

Compared with assigning all degrees of freedom to controller joysticks, separating the functions of the body-mounted IMUs, headset, and hand controllers gives clear physical meanings to continuous body motion, viewing direction, and grasping actions.

Related studies indicate that immersive displays, natural-motion input, and intuitive master-side interface design can improve spatial understanding and task interaction in teleoperation [10]–[12][13].

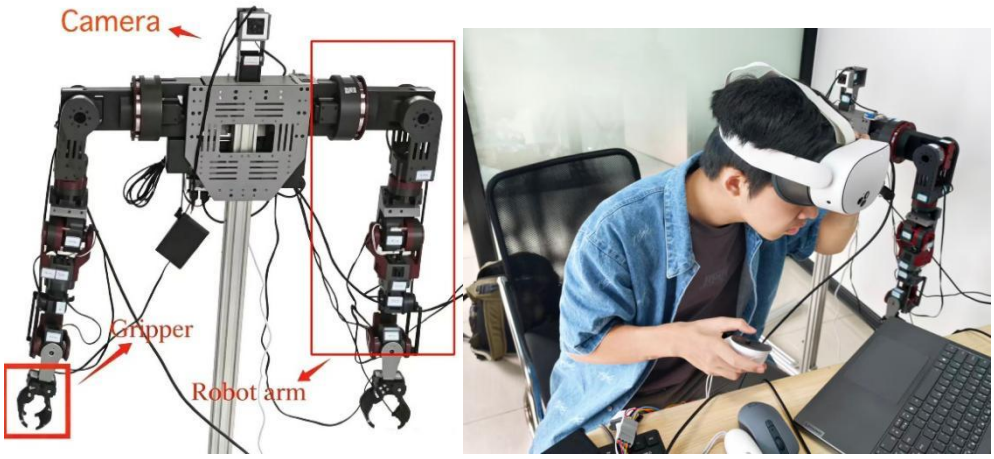



Figure 2-9 Robot-Side Hardware and VR Teleoperation Setup

Table 2-2 Human Motion to Robot Joint Mapping

| Human Motion | Left Robot Arm | Right Robot Arm |
|---|---|---|
| Shoulder Flexion / Extension | ID1 | ID21 |
| Shoulder Abduction / Adduction | ID2 | ID22 |
| Upper-Arm Internal / External Rotation | ID3 | ID23 |
| Elbow Flexion / Extension | ID4 | ID24 |
| Forearm Pronation / Supination | ID5 | ID25 |

The human body and robotic arms are not identical in terms of joint axes, link lengths, or ranges of motion. The system therefore adopts joint-space mapping and separately stores the servo ID, mechanical zero position, rotation direction, mapping scale, and allowable range for the left and right sides.

Wearable motion mapping can improve the naturalness of control, but it must still account for the physical constraints of the robot so that a pose achievable by the human body does not generate a target configuration that the robot cannot execute [14]–[16].

## 2.5 Control Computer and Hardware Connections

The system uses a Raspberry Pi as its control computer.

The Raspberry Pi connects to the five WT901C-485 sensors through a USB-to-RS-485 interface, sends target positions to the 18 servos through a separate serial-servo bus adapter, and receives data from the VR headset and controllers over a wireless network.

The video stream from the robot-mounted camera is transmitted over the network to the VR display. At the same time, headset pose data are returned to the Raspberry Pi to control the two-degree-of-freedom head.

The power system supplies the actuators separately from the computing platform. The servos and IMUs are connected to a 12 V DC power bus, while the Raspberry Pi and low-voltage peripherals are powered from a regulated 5 V supply. All communication modules share a common reference ground.

Because the peak current drawn by the servos is significantly higher than that of the computer and sensors, the actuator power lines and logic communication lines must be appropriately arranged, with sufficient current margin reserved.

## 2.6 Software Workflow and Adaptive Shared-Control Architecture




Figure 2-10 System Software and Control Flow

System operation consists of device checking and homing, natural-pose calibration, and real-time teleoperation. At startup, the Raspberry Pi verifies the IMUs, servo bus, and VR devices before gradually moving the servos to their predefined zero positions. The operator then maintains a natural posture while multiple IMU samples are averaged in quaternion space to establish the initial sensor orientations [23].

During real-time operation, the five IMUs are aligned to a common timestamp using quaternion-based short-term extrapolation [24]. Human commands are passed to the robot when the input state is valid; otherwise, the supervisory layer applies range limiting, velocity

limiting, channel holding, or abnormal-state holding. This state-dependent intervention constitutes the adaptive shared-control mechanism used in the present prototype. Related orientation-fusion methods may further reduce bias and magnetic interference [25].

# 3 Five-IMU Pose Reconstruction and Robotic-Arm Mapping Method

This chapter describes the computational chain that converts five sequentially acquired IMU orientations into synchronized commands for the dual robotic arms. The method contains four stages: multi-IMU temporal alignment, natural-pose calibration and chest-reference transformation, bilateral joint-angle extraction, and human-to-robot mapping with command constraints. All orientations are processed as unit quaternions to avoid the discontinuities and singularities associated with direct Euler-angle mapping [19–23].

Let a unit quaternion be

$$q=(w, v)=(w, x, y, z), \| q \| =1, \tag{3-1}$$

where $w$ is the scalar component and $v$ is the vector component. Its inverse is $q^{-1}=(w, -x, -y, -z)$. For a three-dimensional vector $r$ represented as the pure quaternion $(0, r)$, rotation by $q$ is written as

$$Rot(q, r)=\overrightarrow{[q \otimes (0, r) \otimes q^{-1}]}, \tag{3-2}$$

where $\otimes$ denotes quaternion multiplication and $\mathrm{vec}(\cdot)$ extracts the vector component.

## 3.1 Multi-IMU Acquisition and Temporal Alignment

The five WT901C-485 sensors share one RS-485 bus and are queried sequentially in the order of chest, left upper arm, left forearm, right upper arm, and right forearm. Consequently, measurements collected in one control cycle correspond to different instants. This difference is small during slow motion but can cause visible phase errors during rapid, bilateral, or composite movements. The system therefore estimates the acquisition time of each response and extrapolates the earlier orientations to the latest time in the cycle.

Because the internal sampling instant of each sensor is unavailable, the midpoint of the request-response transaction is used as its estimated timestamp:

$$t_{i,k}=\frac{t_{i,k}^{start}+t_{i,k}^{end}}{2}, \tag{3-3}$$

where $i \in \{1, \ldots, 5\}$ denotes the sensor and $k$ denotes the acquisition cycle. This timestamp does not reproduce the sensor's exact internal sampling time, but it provides a consistent estimate of the relative acquisition order on the shared bus.

For two consecutive valid orientations of sensor $i$, the quaternion increment is

$$\Delta q_{i,k}=q_i(t_{i,k}) \otimes q_i^{-1}(t_{i,k-1}). \tag{3-4}$$

Writing $\Delta q_{i,k}$ in axis-angle form as

$$\Delta q_{i,k}=\left(cos\frac{\phi_i}{2}, u_i sin\frac{\phi_i}{2}\right), \tag{3-5}$$

its real-valued power is

$$\Delta q_{i,k}^{\lambda_i}=\left(cos\frac{\lambda_i\phi_i}{2}, u_i sin\frac{\lambda_i\phi_i}{2}\right). \tag{3-6}$$

The latest timestamp in the current cycle is selected as the common target time, $t^*=max_i t_{i,k}$. Under the short-term constant-angular-velocity assumption, the extrapolation ratio and aligned orientation are

$$\lambda_i=\frac{t^*-t_{i,k}}{t_{i,k}-t_{i,k-1}}, \tag{3-7}$$

$$\hat{q}_i(t^*)=\Delta q_{i,k}^{\lambda_i} \otimes q_i(t_{i,k}). \tag{3-8}$$

The most recently acquired sensor has $\lambda_i=0$ and therefore requires no extrapolation. The operation is applied only when two consecutive valid samples are available and $0 \le t^* - t_{i,k} \le 0.15$ s. A control frame is accepted only when all five sensors respond and the raw temporal span

$$T_{span}=\max_i t_{i,k} - \min_i t_{i,k} \tag{3-9}$$

does not exceed 0.16 s. Otherwise, the pose update is rejected and the previous valid robot target is retained. This prevents incomplete or outdated segment orientations from being combined in the same reconstruction cycle.

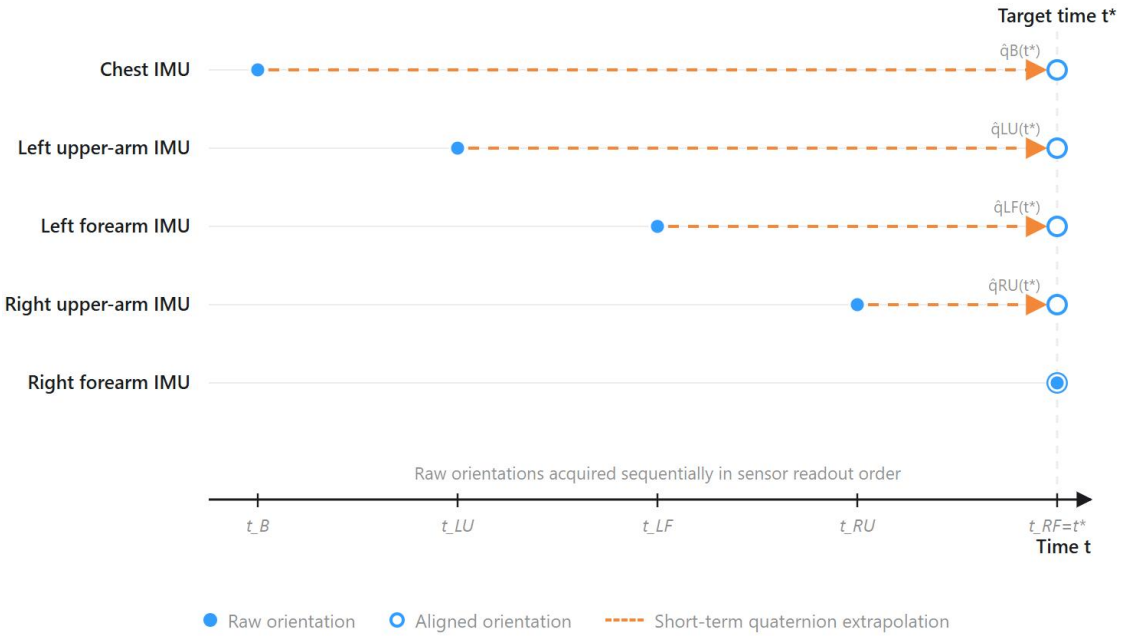



Figure 3-1 Sequential Sampling and Temporal Alignment of the Five IMUs

## 3.2 Natural-Pose Calibration and Chest-Reference Transformation

After temporal alignment, the sensor orientations correspond to a common time but remain expressed in the world frame. Two additional errors must be removed before joint reconstruction. First, strap position and installation angle cause a fixed offset between each sensor frame and its body segment. Second, torso rotation changes the world-frame orientations of all arm-mounted sensors even when the arms do not move relative to the body.

At startup, the operator stands naturally with both arms relaxed. The system collects 20 complete five-IMU frames. Because $q$ and $-q$ represent the same rotation, each calibration quaternion is first sign-aligned with the first sample:

$$q'_{i,n}=\begin{cases} q_{i,n}, & q_{i,1}^{T}q_{i,n} \geq 0, \\ -q_{i,n}, & q_{i,1}^{T}q_{i,n}<0. \end{cases} \tag{3-10}$$

The natural orientation of sensor $i$ is then estimated by the normalized mean

$$q_{i0}=normalize\left(\sum_{n=1}^{20} q'_{i,n}\right). \tag{3-11}$$

This multi-frame estimate reduces the influence of sensor noise and small involuntary movements during calibration. Quaternion-averaging methods such as the Markley method provide the mathematical basis for this operation [22].

The change of each sensor relative to its natural pose is

$$q_i^{\Delta}(t)=\hat{q}_i(t) \otimes q_{i0}^{-1}. \tag{3-12}$$

Let $q_b^{\Delta}$ denote the chest rotation and $q_{b0}$ the natural chest orientation. The orientation of limb segment $i$ after removing common torso motion and expressing the remaining rotation in the natural chest coordinate system is

$$q_i^c=q_{b0}^{-1} \otimes \left(q_b^{\Delta}\right)^{-1} \otimes q_i^{\Delta} \otimes q_{b0}, \tag{3-13}$$

where $i \in \{lu, lf, ru, rf\}$ denotes the left upper arm, left forearm, right upper arm, and right forearm. If the torso and arms rotate together as a rigid body, $q_i^{\Delta} \approx q_b^{\Delta}$ and the common rotation is largely cancelled. Thus, subsequent shoulder and elbow parameters describe limb motion relative to the operator rather than changes in global facing direction.

The chest coordinate system uses the $X$-axis vertically upward, the $Y$-axis toward the operator's right, and the $Z$-axis forward. In the natural hanging posture, the longitudinal reference direction of an upper arm or forearm is $d_0=(-1,0,0)^T$.

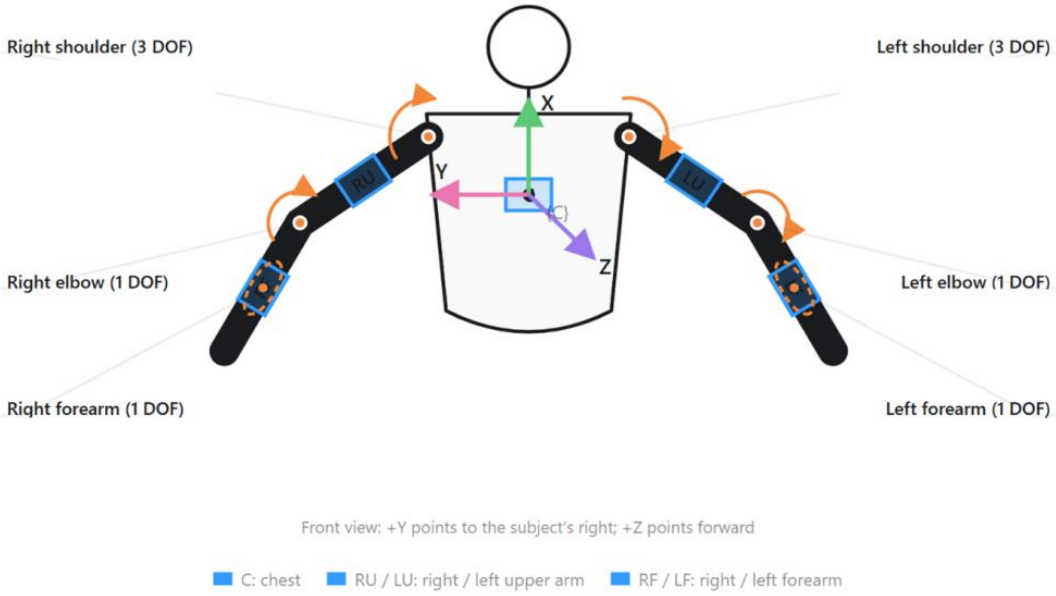



Figure 3-2 Chest-Centered Coordinate System and Five-IMU Upper-Limb Model

## 3.3 Bilateral Joint-Angle Extraction

Five motion parameters are reconstructed for each arm: shoulder flexion/extension, shoulder abduction/adduction, upper-arm internal/external rotation, elbow flexion/extension, and forearm pronation/supination. The two arms therefore provide ten control parameters. Directional motion is obtained from rotated limb vectors, while rotation about a limb's longitudinal axis is isolated through swing-twist decomposition.

### 3.3.1 Shoulder Flexion and Abduction

For an upper-arm orientation $q_u^c$, the current limb direction is

$$d_u=Rot(q_u^c, d_0)=\left(d_x, d_y, d_z\right)^T. \tag{3-14}$$

Shoulder flexion is calculated from the projection of $d_u$ onto the sagittal plane. Shoulder abduction uses a side-dependent sign so that outward movement is positive on both sides:

$$\theta_{flex}=atan2(d_z, -d_x), \tag{3-15}$$

$$\theta_{abd}=atan2\left(c_{side}, \sqrt{d_x^2+d_z^2}\right), c_{side}=\begin{cases} -d_y, & \text{left arm,} \\ d_y, & \text{right arm.} \end{cases} \tag{3-16}$$

This formulation relates the two shoulder angles to the direction of the upper arm rather than to a fixed Euler-angle sequence.

### 3.3.2 Upper-Arm Axial Rotation

The longitudinal direction alone cannot determine internal or external rotation because the upper arm can twist without changing its primary axis. The method therefore separates the change in limb direction (swing) from the residual rotation about the current limb axis (twist).

Let the natural longitudinal axis and its reference direction be $a_0=(1,0,0)^T$ and $r_0=(0,1,0)^T$. For the current upper-arm orientation,

$$a=Rot(q_w^c, a_0), r=Rot(q_w^c, r_0). \tag{3-17}$$

The swing quaternion and corresponding zero-twist reference direction are

$$q_{swing}=FromTwoVectors(a_0, a), r_{swing}=Rot\left(q_{swing}, r_0\right). \tag{3-18}$$

The upper-arm axial rotation is the signed angle from $r_{swing}$ to $r$ about the current longitudinal axis:

$$\theta_{upper}=atan2\left[a \cdot \left(r_{swing} \times r\right), r_{swing} \cdot r\right]. \tag{3-19}$$

This decomposition reduces the tendency of shoulder flexion or abduction to be interpreted as upper-arm axial rotation.

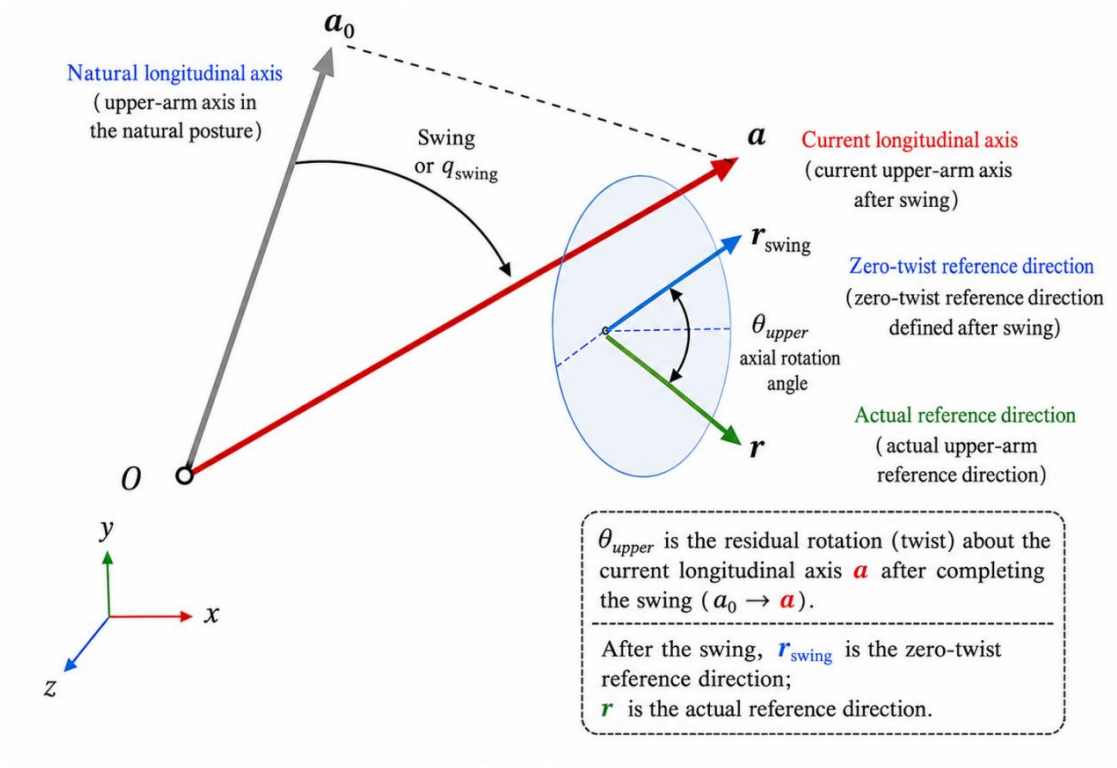



Figure 3-3 Geometric Decomposition of Upper-Arm Swing and Twist

### 3.3.3 Elbow Flexion and Forearm Rotation

Elbow and forearm motion are calculated from the forearm orientation relative to the upper arm:

$$q_e=(q_u^c)^{-1} \otimes q_f^c. \tag{3-20}$$

Rotating the natural limb direction by $q_e$ gives $d_f=(f_x, f_y, f_z)^T$, from which elbow flexion is obtained as

$$\theta_{elbow}=atan2(f_z, -f_x). \tag{3-21}$$

Because the upper-arm orientation has already been removed, shoulder movement alone does not directly generate an elbow command. Forearm pronation/supination is obtained by applying the same swing-twist operation to $q_e$:

$$\theta_{forearm}=Twist(q_e, a_0, r_0). \tag{3-22}$$

Finally, side- and joint-specific sign coefficients standardize the anatomical meaning of the reconstructed angles:

$$\theta_{a,j}^h=\sigma_{a,j}\theta_{a,j}, \sigma_{a,j} \in \{-1,1\}, a \in \{L, R\}. \tag{3-23}$$

For example, outward rotation is defined as positive for both arms even though the physical

sensor and servo directions are mirrored.

## 3.4 Human-to-Robot Mapping and Synchronized Control

The human and robotic arms differ in zero position, joint direction, transmission ratio, and allowable range. Each reconstructed human angle is therefore mapped using joint-specific parameters rather than being sent directly as a servo command.

**Table 3-1 Mapping Between Reconstructed Human Motions and Robot Servos**

| Human Joint Parameters | Left Robot Arm Servo | Right Robot Arm Servo |
|---|---|---|
| Shoulder Flexion / Extension | ID1 | ID21 |
| Shoulder Abduction / Adduction | ID2 | ID22 |
| Upper-Arm Internal / External Rotation | ID3 | ID23 |
| Elbow Flexion / Extension | ID4 | ID24 |
| Forearm Pronation / Supination | ID5 | ID25 |

### 3.4.1 Joint Calibration and Mapping

Let $p_{j,0}$ be the servo zero position, $s_j \in \{-1,1\}$ the installation-direction coefficient, and $k_j$ the mapping scale in encoder counts per degree. The raw target for joint $j$ is

$$p_j^{raw}=p_{j,0}+s_j k_j \theta_j. \tag{3-24}$$

For a measured calibration movement from $p_{j,0}$ to $p_j^{cal}$ corresponding to a reference angle $\theta_j^{cal}$, the mapping scale is

$$k_j=\frac{|p_j^{cal}-p_{j,0}|}{|\theta_j^{cal}|}. \tag{3-25}$$

The direction coefficient $s_j$ is determined separately for each side to account for the mirrored mechanical installation. Similar joint-specific mapping strategies have been used to improve the correspondence between human motion and structurally different robots [16].

### 3.4.2 Filtering and Command Constraints

Before mapping, each reconstructed human angle is restricted to the verified control range:

$$\theta_j^c=clip\left(\theta_j, \theta_{j,min}, \theta_{j,max}\right). \tag{3-26}$$

**Table 3-2 Verified Human-Side Control Ranges**

| Joint Motion | Minimum Angle | Maximum Angle |
|---|---|---|

| | | |
|---|---|---|
| Shoulder Flexion / Extension | −30° | 90° |
| Shoulder Abduction / Adduction | −10° | 80° |
| Upper-Arm Internal / External Rotation | −60° | 60° |
| Elbow Flexion / Extension | 0° | 90° |
| Forearm Pronation / Supination | −70° | 70° |

These are conservative ranges verified for the prototype rather than physiological limits of human motion. Exponential smoothing suppresses small fluctuations:

$$\theta_{j,k}^{f}=(1-\alpha)\theta_{j,k-1}^{f}+\alpha\theta_{j,k}^{c},\ \alpha=0.35. \tag{3-27}$$

The change in filtered human angle is limited to $\omega_{max}=120^{\circ}$/s:

$$\theta_{j,k}^{o}=\theta_{j,k-1}^{o}+clip\left(\theta_{j,k}^{f}-\theta_{j,k-1}^{o},-\omega_{max}\Delta t,\omega_{max}\Delta t\right). \tag{3-28}$$

The mapped servo target is then restricted to the joint's calibrated mechanical interval:

$$p_{j,k}^{target}=clip\left(p_{j,0}+s_j k_j \theta_{j,k}^{o},p_{j,min},p_{j,max}\right). \tag{3-29}$$

A second rate limit, $v_{max}=500$ encoder counts/s, is applied directly in servo space:

$$p_{j,k}^{cmd}=p_{j,k-1}^{cmd}+clip\left(p_{j,k}^{target}-p_{j,k-1}^{cmd},-v_{max}\Delta t,v_{max}\Delta t\right). \tag{3-30}$$

The first limit constrains abrupt changes in the reconstructed human motion, while the second constrains the final actuator command. If a complete and timely five-IMU frame is unavailable, no new target is calculated and the previous valid command is retained.

### 3.4.3 Synchronized Dual-Arm Transmission

The ten IMU-controlled joint targets and the positions of any joints that must remain fixed are assembled into one synchronized broadcast packet. Each servo reads the command associated with its own ID, allowing the left and right arms to receive targets calculated in the same control cycle with minimal additional bus delay. This is particularly important during cooperative transport, where the two end effectors must maintain compatible motion [17,18].

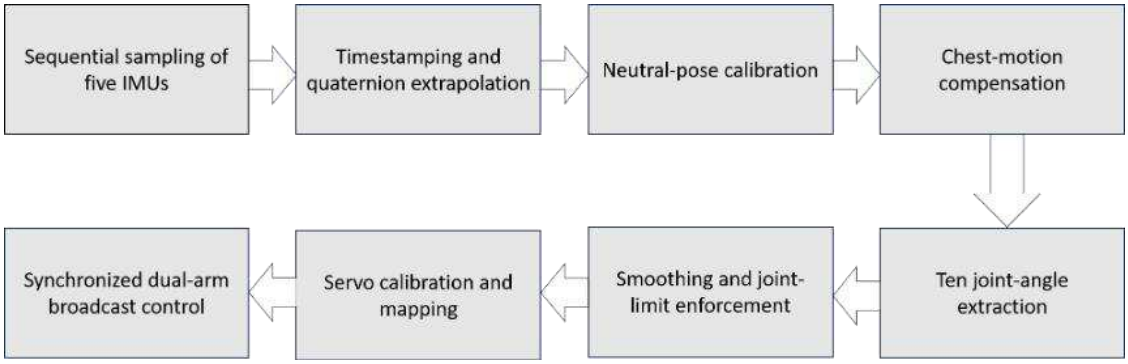



Figure 3-3 Complete Five-IMU Reconstruction and Dual-Arm Mapping Process

In summary, the proposed method aligns the sequential IMU measurements in time, removes fixed mounting offsets and common torso rotation, decomposes bilateral upper-limb motion into ten joint parameters, and converts these parameters into constrained, synchronized robot commands.

# 4 VR Interaction and System Implementation

Following the five-IMU pose reconstruction and human-to-robot motion mapping described in Chapter 3, the system further integrates a Meta Quest headset and two controllers to provide immersive teleoperation with visual feedback, dual-arm control, and grasping capability. VR integrates remote visual feedback and natural interaction within a unified interface [10,11]. Human upper-limb motion controls the continuous motion of the robotic arms, the headset controls the viewing direction of the remote camera, and the left and right controllers independently operate the corresponding grippers. By assigning pose control, viewpoint control, and grasping to separate input channels, the system avoids overloading a single device and improves the naturalness and controllability of teleoperation.

## 4.1 Headset-Controlled Camera Following

Traditional robot teleoperation commonly relies on either fixed cameras or keyboard-based camera control. A fixed viewpoint cannot cover the complete workspace around the robotic arms, while keyboard operation can divert the operator's attention from manipulation. Therefore, the Meta Quest headset is used to acquire the operator's head pose in real time and map head rotation to the yaw and pitch of the remote camera. This head-pose-to-viewpoint mapping allows the operator to change the viewing direction naturally, reducing additional button operations and improving spatial perception during teleoperation [10,11].

At system startup, the initial headset orientation quaternion $q_{h,0}$is recorded, while the current head pose $q_h(t)$ is continuously acquired during operation. The head orientation

relative to the initial direction is

$$q_h^{\Delta}(t)=q_h(t)q_{h,0}^{-1}. \tag{4-1}$$

Because the camera mechanism has only yaw and pitch degrees of freedom, headset roll is not transmitted to the camera. Let the head yaw and pitch extracted from $q_h^{\Delta}$ be $\psi_h$and $\phi_h$, respectively. The camera target angles are

$$\begin{aligned}\theta_{pan}^{*}&=clip\left(\theta_{pan,0},+,s_{pan}k_{pan}\psi_h\theta_{pan}^{min}\theta_{pan}^{max}\right),\\ \theta_{tilt}^{*}&=clip\left(\theta_{tilt,0},+,s_{tilt}k_{tilt}\phi_h\theta_{tilt}^{min}\theta_{tilt}^{max}\right).\end{aligned} \tag{4-2}$$

Here, $\theta_{pan,0}$and $\theta_{tilt,0}$are the initial camera positions; $k_{pan}$and $k_{tilt}$are the mapping ratios from head rotation to camera rotation; $s_{pan}, s_{tilt} \in \{-, 1+1\}$ compensate for differences between the headset coordinate system and the mechanical rotation directions of the camera; and $clip(\cdot)$denotes angular limiting.

The mapping ratio does not need to remain fixed at 1:1. It can be reduced when the mechanical range of the camera is smaller than the operator's comfortable head-rotation range, or increased when a larger viewing range is required from a smaller head movement. Camera yaw and pitch are also restricted by the mechanical limits of the structure to prevent collision with the limiting mechanism.

Images captured by the camera are transmitted in real time and displayed within the Meta Quest virtual field of view. As the operator turns the head, the headset drives the remote camera, which changes the returned image and forms the closed loop

Head Rotation → Camera Rotation → Remote Image Change → VR Visual Feedback.

This provides an observation experience similar to viewing the workspace from the robot's position while eliminating the need for separate viewpoint-control buttons.

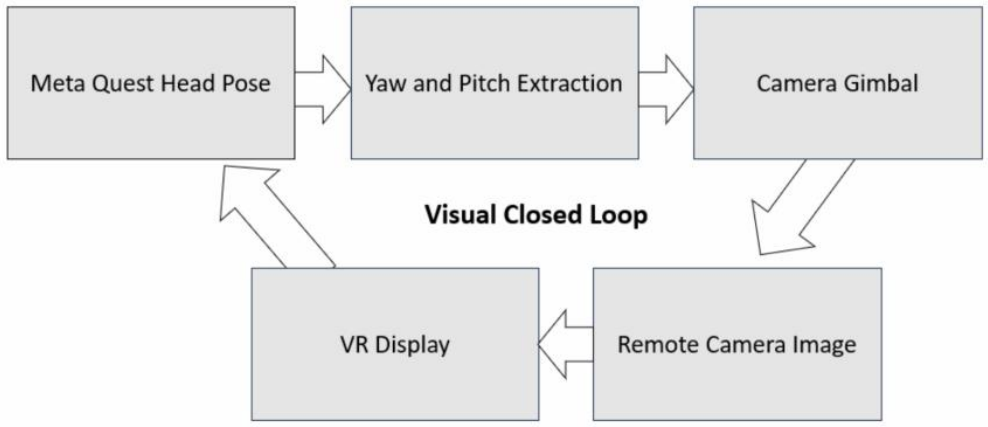



Figure 4-1 Closed-Loop Headset-to-Camera View Control

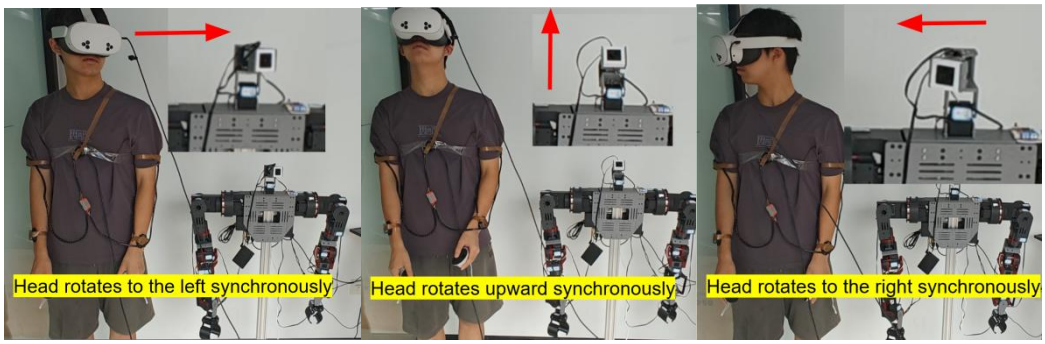



Figure 4-2 Camera Following Test

## 4.2 Controller-Based Dual-Gripper Control

The five IMUs continuously capture the poses of the chest, upper arms, and forearms, but they cannot directly and reliably represent finger-level grasping intent. Adding additional hand-mounted IMUs would increase both system complexity and the burden of wearing and calibration. The left and right Meta Quest controllers are therefore used to control the corresponding grippers independently, separating continuous dual-arm pose control from discrete grasping commands. Such functional separation reduces the operational burden that would otherwise result from assigning excessive degrees of freedom to a single input channel [12,13].

Let the input states of the left and right controllers be $u_L$ and $u_R$. For button-based control, $u_s$ is represented as a binary variable:

$$u_s = \begin{cases} 0, & \text{gripper open,} \\ 1, & \text{gripper closed,} \end{cases} \quad s \in \{L, R\}. \tag{4-3}$$

The target gripper position is mapped between the predefined open and closed positions:

$$p_{g,s}^{*} = p_{g,s}^{open} + u_s\left(p_{g,s}^{close} - p_{g,s}^{open}\right). \tag{4-4}$$

Here, $p_{g,s}^{open}$ and $p_{g,s}^{close}$ are the servo positions corresponding to the fully open and safely closed states of gripper $s$, respectively. Both positions are determined through mechanical calibration and remain within the actual travel range of the gripper, preventing sustained closing commands from causing servo stall or structural damage.

When the controller trigger value is used as input, small fluctuations may occur near the switching threshold. To prevent repeated transitions between open and closed states, a hysteresis interval is introduced:

$$u_s(k)=\begin{cases} 1, & z_s(k) \geq \tau_{close}, \\ 0, & z_s(k) \leq \tau_{open}, \\ u_s(k-1), & \tau_{open}<z_s(k)<\tau_{close}, \end{cases} \tag{4-5}$$

where $z_s(k)$is the controller input value at frame $k$, with

$$\tau_{open}<\tau_{close}.$$

The hysteresis interval suppresses false switching caused by fluctuations around a single threshold. State-change detection is also used so that a new gripper command is generated only when the gripper state changes:

$$Send_s(k)=\begin{cases} 1, & u_s(k) \neq u_s(k-1) \text{ and } t_k - t_s^{last} \geq T_d, \\ 0, & \text{otherwise.} \end{cases} \tag{4-6}$$

Here, $T_d$ is the minimum command interval. This prevents the same command from being repeatedly transmitted across consecutive rendering frames, thereby reducing communication load and mechanical vibration caused by unnecessarily frequent servo responses.

The left and right controllers maintain independent state variables and target positions, allowing either independent or simultaneous bimanual grasping. The division of functions is therefore explicit: the five IMUs control continuous dual-arm pose, the controllers provide discrete grasping commands, and the headset controls the remote viewpoint. Although the three inputs are processed independently, they act together within the same teleoperation cycle.

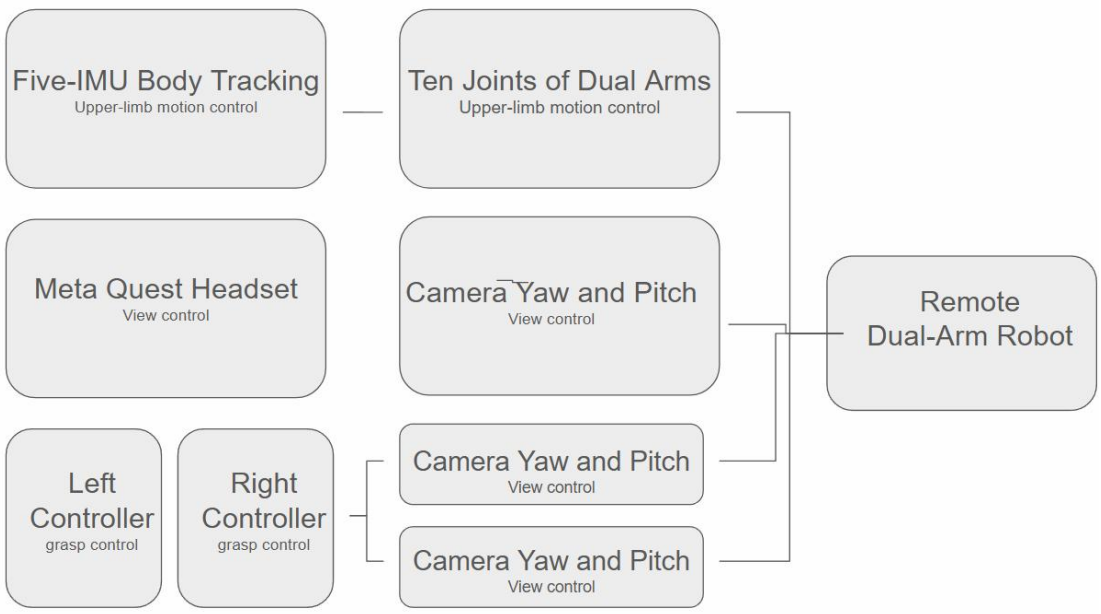



Figure 4-3 Functional Allocation of IMUs, Headset, and Controllers

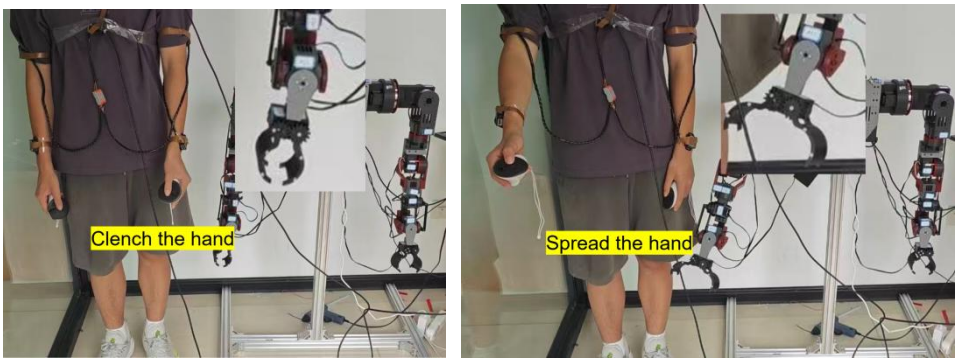



Figure 4-4 Independent Gripper Control Using the Left and Right Controllers

## 4.3 State-Adaptive Shared Control and Safety Mechanisms

The shared-control mechanism contains two roles. The human operator selects the task, arm motion, gripper state, and viewing direction; the controller acts as a rule-based supervisor. At startup, the supervisor verifies communication, reads the current servo positions, and performs slow homing before accepting real-time commands. No autonomous target selection, trajectory planning, or grasp synthesis is used.

During valid operation, control authority remains with the operator and the system tracks the reconstructed targets after filtering and synchronized transmission. When a target approaches a verified joint range or changes too rapidly, the supervisor modifies it through joint and velocity limits. When an IMU frame is incomplete, the sampling-time span exceeds the threshold, or a VR channel times out, the affected command is rejected and its previous valid target is held. The resulting transitions among normal tracking, constrained tracking, and position holding make the shared-control behavior adaptive to system state.

Persistent abnormal frames keep the controller in position holding and report the affected channel. Tracking resumes only after valid data return and the safety conditions are satisfied. On exit, the last safe target is retained unless the robot has been mechanically supported and torque release is deliberately requested. Figure 4-5 summarizes these state transitions.

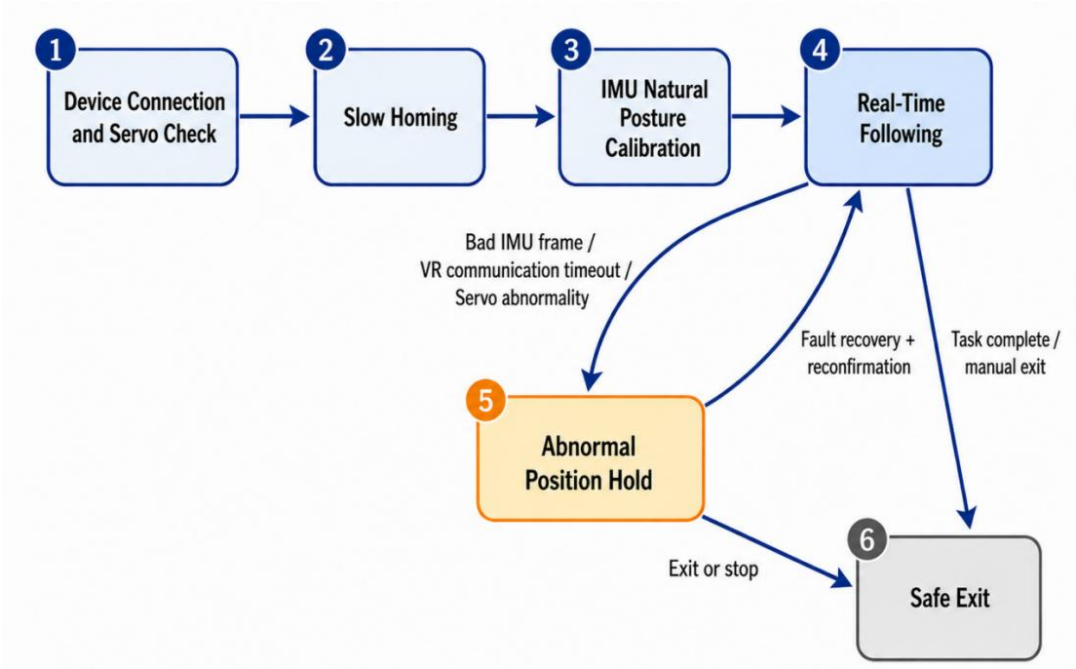



Figure 4-5 Safety State Transitions of the Real-Time Teleoperation System

The VR and IMU channels provide human intent, while the supervisory layer validates and constrains the resulting targets before execution. Figure 4-6 confirms that both arm target sets can be issued within the same synchronized update cycle. This architecture supports fine-grained multimodal operation without claiming autonomous manipulation.




Figure 4-6 Dual-Arm Synchronization Test

# 5 Experiments and Results Analysis

This chapter evaluates multi-IMU acquisition timing, static output stability, joint-motion decoupling, the contributions of the reconstruction modules, and the functional feasibility of VR-assisted fine teleoperation with supervisory shared control. The ablation study evaluates the sensing and reconstruction portion of the control chain; the grasping experiment evaluates integrated operation qualitatively.

## 5.1 Experimental Setup and Protocol

### 5.1.1 Experimental Platform

The platform comprised five WT901C-485 IMUs, a Raspberry Pi control computer, a dual-arm robot, and a Meta Quest 3S headset with two controllers. The IMUs were mounted on the chest, upper arms, and forearms and were queried sequentially over one RS-485 bus. The headset controlled the remote camera, while the two controllers operated the grippers. Table 5-1 lists the components relevant to the experiments; the detailed mechanical and electrical architecture is given in Chapter 2.

**Table 5-1 Main components of the experimental platform**

| Module | Specification |
|---|---|
| Attitude sensing | Five WT901C-485 IMUs |
| IMU communication | RS-485, 9600 bit/s |
| Arm actuation | HLS3950 and STS3215 servos |
| Servo communication | 1 Mbit/s bus with synchronized broadcast commands |
| VR interaction | Meta Quest 3S headset and two controllers |
| Control software | Python 3.13 and Unity 6.3 LTS |

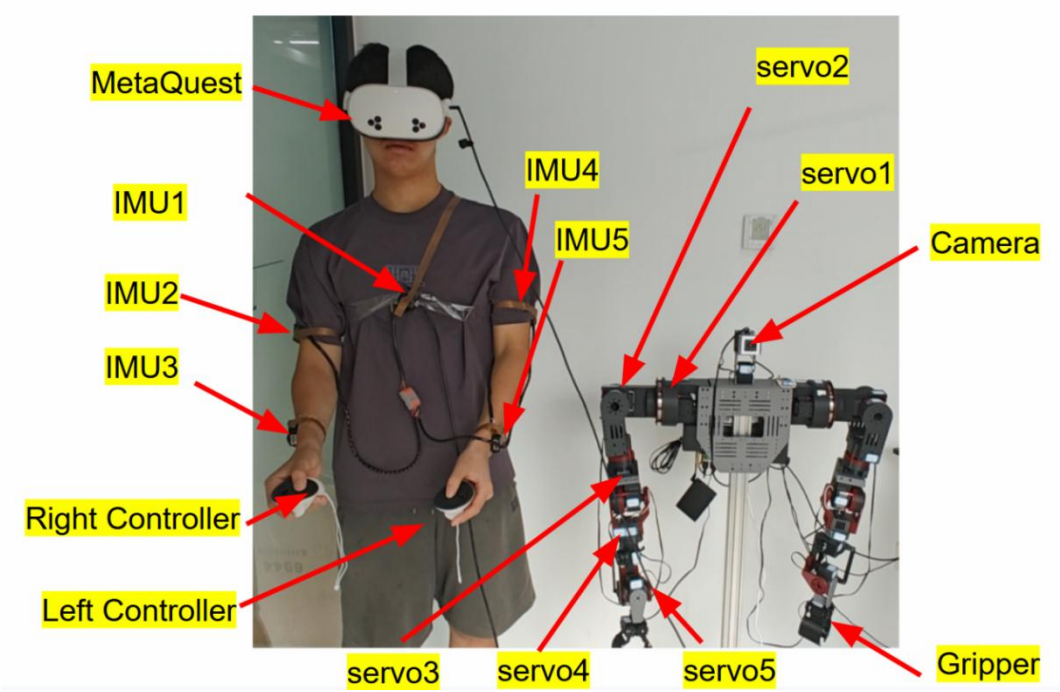



Figure 5-1 Experimental setup

### 5.1.2 Experimental Procedure

A single 17-year-old operator completed all measurements. Sensor placement and strap tightness were kept unchanged within each session. Before testing, the operator maintained the natural standing posture described in Chapter 3 while the system collected 20 complete IMU frames for quaternion averaging and mounting-offset calibration [19–22].

The operator then performed shoulder flexion, shoulder abduction, upper-arm external rotation, elbow flexion, and forearm supination with each arm, producing ten single-joint conditions. Each condition lasted 12 s and contained three slow repetitions. The raw quaternions, response timestamps, aligned orientations, and reconstructed joint angles were stored for analysis.

Functional teleoperation was examined through a grasping experiment in which the robotic arm approached a bottle, closed the gripper, lifted the object, and transferred it to a new position. The experiment was used to verify the integrated control chain rather than to estimate a population success rate.

### 5.1.3 Evaluation Metrics

During prescribed motion $i$, the robust variation of output joint $j$ was calculated from the 5th and 95th percentiles. The mean normalized off-diagonal response defines the crosstalk rate:

$$A_{ij}=P_{95}(\theta_j)-P_5(\theta_j),\ C_i=\frac{1}{9}\sum_{j\neq i}\frac{A_{ij}}{A_{ii}}\times 100\%. \tag{5-1}$$

The temporal span within one five-sensor acquisition cycle was

$$\Delta T_{raw}(k)=\max_i t_i(k)-\min_i t_i(k). \tag{5-2}$$

Static stability was represented by the standard deviation of each reconstructed joint angle in the natural resting posture:

$$\sigma_j=\sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left[\theta_j(k)-\overline{\theta}_j\right]^2}. \tag{5-3}$$

These metrics were selected because they can be calculated from the recorded IMU and reconstructed joint-angle streams without an external motion-capture reference.

## 5.2 Pose Reconstruction Results

### 5.2.1 Sampling Timing and Static Stability

The raw sampling span across the five sequentially queried IMUs was 112.433 ms on average, with a

standard deviation of 0.320 ms, a 95th percentile of 112.641 ms, and a maximum of 113.552 ms. The small dispersion shows that the bus cycle was stable, but the approximately 112 ms difference between the earliest and latest measurements was large enough to produce segment misalignment during dynamic motion. The temporal-alignment step therefore extrapolated all five quaternion streams to the latest response time using short rotational increments [23].

With the complete reconstruction method, the mean static standard deviation across the ten joint outputs was 0.241°. The largest joint standard deviation was 0.665° for right shoulder abduction. These values indicate stable output in the natural posture, while the comparatively larger shoulder variation suggests that multi-axis shoulder reconstruction remains more sensitive than elbow reconstruction.

### 5.2.2 Single-Joint Response and Motion Crosstalk

For each of the ten prescribed movements, the robust variation of all ten reconstructed joint outputs was normalized by the target-joint variation to form the response matrix in Figure 5-2. Diagonal cells represent the intended response, whereas off-diagonal cells show undesired coupling.

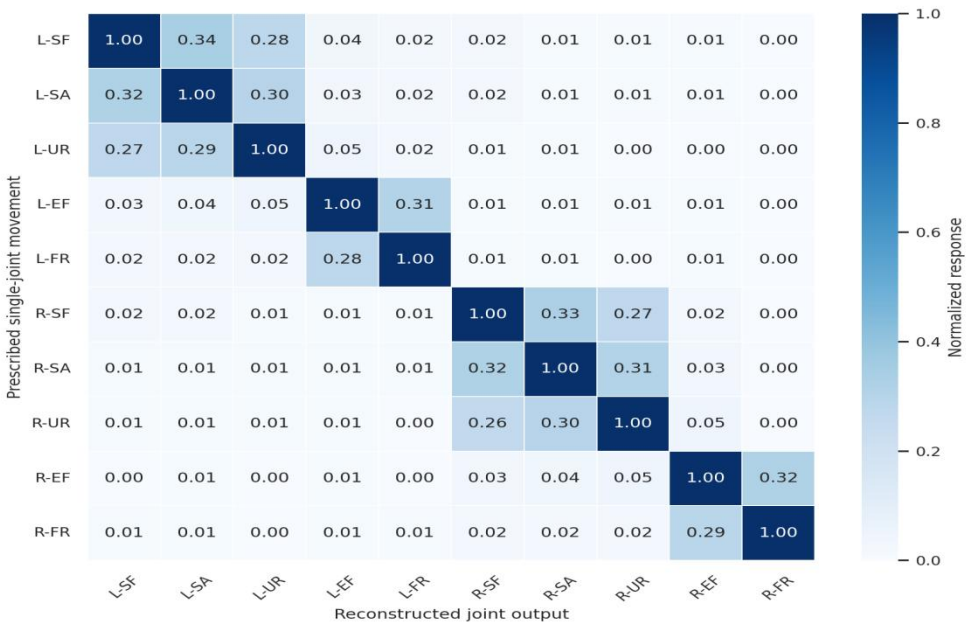



Figure 5-2 Normalized response heatmap for ten single-joint movements

The concentration of the response along the diagonal shows that the five-IMU method distinguished the main bilateral joint motions. Residual off-diagonal responses were most evident around the multi-degree-of-freedom shoulder. Chest-reference compensation reduced motion caused by torso rotation, while swing–twist decomposition separated changes in limb direction from axial rotation and thereby reduced shoulder crosstalk.

## 5.3 Ablation Results

Four configurations were compared to separate the contributions of relative-pose calibration, chest-reference and geometric compensation, and temporal alignment: direct Euler mapping (A), relative-pose mapping (B), spatial compensation (C), and the complete method (D). Their module composition is summarized in Table 5-2. All four configurations were evaluated using the same raw IMU recordings, valid time intervals, and single-joint motion segments. These configurations do not compare autonomous policies; they isolate the pose-processing modules that supply reliable human commands to the supervisory shared-control layer.

**Table 5-2 Ablation configurations**

| Method | Relative pose | Chest reference | Swing–twist decomposition | Time alignment |
|---|---|---|---|---|
| A | — | — | — | — |
| B | ✓ | — | — | — |
| C | ✓ | ✓ | ✓ | — |
| D | ✓ | ✓ | ✓ | ✓ |

The comparison in Table 5-3 reports the two metrics that can be calculated directly from the recorded joint outputs.

**Table 5-3 Comparison of the four reconstruction methods**

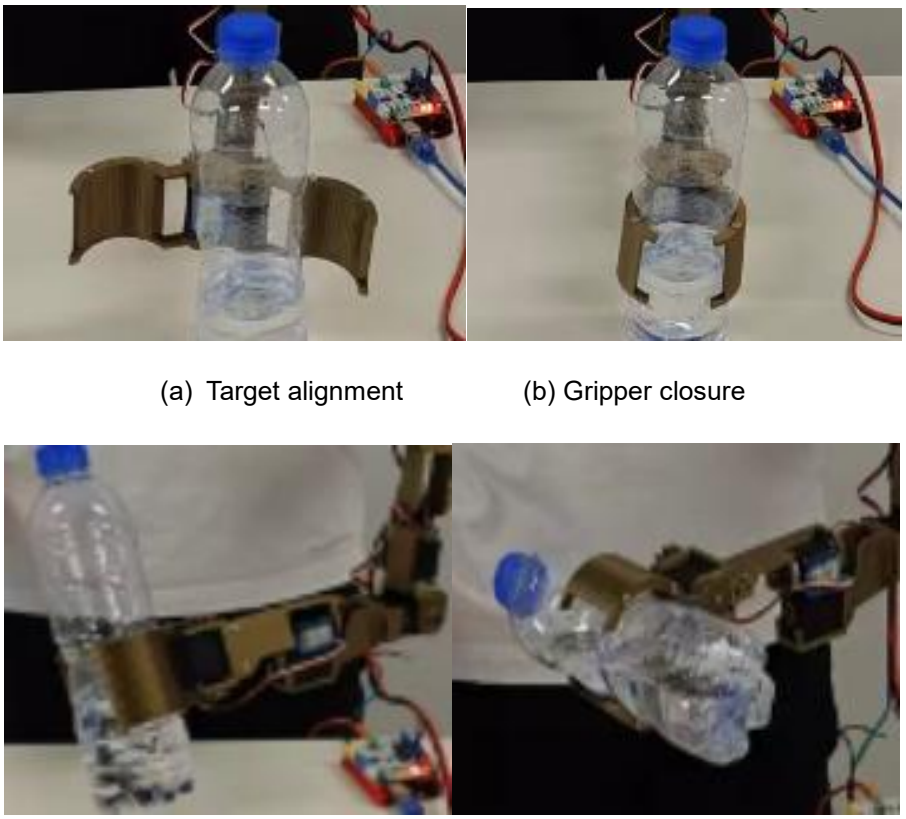
| Method | Mean crosstalk rate (%) | Mean static jitter (°) |
|---|---|---|
| A | 18.7 | 0.684 |
| B | 14.1 | 0.512 |
| C | 7.3 | 0.296 |
| D | 5.4 | 0.241 |

Relative-pose calibration reduced crosstalk from 18.7% to 14.1% and static jitter from 0.684° to 0.512°, showing the benefit of removing fixed mounting and initial-pose offsets. Adding the chest reference and swing–twist decomposition produced the largest improvement: Method C reduced crosstalk to 7.3% and static jitter to 0.296°. This result is consistent with the intended role of spatial compensation in rejecting torso rotation and separating coupled shoulder components.

Temporal alignment further reduced crosstalk from 7.3% to 5.4% and static jitter from 0.296° to 0.241°. The improvement is smaller than that produced by spatial compensation in these slow single-joint tests, but it supports retaining temporal alignment for bilateral and composite movements, where the 112 ms raw sampling span is more consequential. Overall, each stage improved the recorded output, and the complete method achieved the lowest crosstalk and static variation.

## 5.4 Grasping Experiment

The grasping experiment evaluated whether fine-grained arm motion, controller-based grasping intent, and supervisory command processing could operate in one continuous sequence. A plastic bottle was placed within the reachable workspace. The operator guided the arm toward the bottle using upper-limb motion and visual feedback, while the controller validated and constrained the commands before transmission. The corresponding VR controller then commanded the gripper to close around the object.



(a) Target alignment (b) Gripper closure

(c) Object lifting (d) Object transfer

Figure 5-3 Grasping experiment sequence: (a) target alignment, (b) gripper closure, (c) object lifting, and (d) object transfer

As shown in Figure 5-3, the arm aligned the open gripper with the bottle, closed the gripper, lifted the object, and transferred it while maintaining the grasp. The sequence confirms functional cooperation among five-IMU pose reconstruction, human-to-robot mapping, supervisory command constraints, synchronized servo control, gripper commands, and visual feedback. All task-level decisions remained human-generated.

In summary, the experiments show stable sequential acquisition, reduced crosstalk and static variation after reconstruction and compensation, and completion of a grasping-and-transfer sequence under the supervisory control chain.

# 6 Conclusions and Future Work

## 6.1 Summary

This study developed Phantom Hand, a dual-arm robotic fine-teleoperation prototype integrating five wearable IMUs, immersive VR interaction, and state-adaptive supervisory shared control. Fine teleoperation is realized through ten reconstructed upper-limb joint parameters, independent gripper commands, and headset-controlled visual orientation. Human intent remains the primary command source, while the supervisory layer validates, limits, holds, or rejects commands according to sensing, communication, and mechanical conditions.

Prototype measurements showed a mean raw five-sensor sampling span of 112.433 ms. Relative-pose calibration, spatial compensation, and temporal alignment progressively reduced mean motion crosstalk from 18.7% to 5.4% and mean static jitter from 0.684° to 0.241°. The grasping-and-transfer experiment further demonstrated continuous cooperation among human motion input, adaptive supervisory processing, controller-based gripper commands, synchronized execution, and visual feedback.

## 6.2 Future Work

The current system still has several limitations.

First, the five IMUs are sequentially acquired over a shared RS-485 bus, resulting in an effective update rate of approximately 7.10 Hz. Future work could improve update frequency through higher baud rates, optimized query protocols, parallel acquisition, or hardware synchronization, and could incorporate angular velocity and acceleration for more accurate short-term prediction.

Second, long-term IMU performance may be affected by gyroscope bias, magnetic interference, and strap slippage. Magnetic-anomaly detection, online bias estimation, automatic recalibration, sensor-loosening detection, and kinematic constraints could improve long-term stability.

Third, the current human-to-robot mapping is mainly linear and does not fully capture nonlinearities in robot linkages, transmissions, and servo output. Piecewise-linear or nonlinear calibration, together with forward and inverse kinematics, could improve mapping accuracy. For dual-arm tasks, constraints on the relative position, orientation, and velocity of the two end effectors could help reduce object tilting, slipping, and dropping.

Finally, the current grippers provide no direct force feedback. Future versions could integrate pressure, force/torque, or tactile sensing and return grasp information through controller vibration or visual cues. Video transmission, camera-motion prediction, and viewpoint stabilization could also be improved to reduce visual latency and enhance operation in complex environments.

# References


[1] Li, S., Zhao, C., & Zhang, L. (2025). Several Issues and Suggestions Regarding the Future Intrinsically Safe Chemical Industrial Park: Insights from Severe Hazardous Chemical Accident Analysis and Industrial Park Field Investigations. Processes, 13(11), 3630. https://doi.org/10.3390/pr13113630

[2] Yang, D., Zheng, Y., Peng, K., Pan, L., Zheng, J., Xie, B., & Wang, B. (2022). Characteristics and Statistical Analysis of Large and above Hazardous Chemical Accidents in China from 2000 to 2020. International Journal of Environmental Research and Public Health, 19(23), 15603. https://doi.org/10.3390/ijerph192315603

[3] Jiang, W., Huang, Z., Wu, Z., Su, H., & Zhou, X. (2022). Quantitative Study on Human Error in Emergency Activities of Road Transportation Leakage Accidents of Hazardous Chemicals. International Journal of Environmental Research and Public Health, 19(22), 14662. https://doi.org/10.3390/ijerph192214662

[4] Patiño, K., Slawiñski, E., Moran-Armenta, M., Mut, V., Rossomando, F. G., & Moreno-Valenzuela, J. (2025). Neural Networks in the Delayed Teleoperation of a Skid-Steering Robot. Mathematics, 13(13), 2071. https://doi.org/10.3390/math13132071

[5] Eishi Kim, Vsevolod Peysakhovich, and Raphaëlle N. Roy. 2021. Impact of communication delay and temporal sensitivity on perceived workload and teleoperation performance. In ACM Symposium on Applied Perception 2021 (SAP '21). Association for Computing Machinery, New York, NY, USA, Article 4, 1–8. https://doi.org/10.1145/3474451.3476233

[6] Okamura A. M. (2009). Haptic feedback in robot-assisted minimally invasive surgery. Current Opinion in Urology, 19(1), 102–107. https://doi.org/10.1097/MOU.0b013e32831a478c

[7] Luo, J., He, W., & Yang, C. (2020). Combined perception, control, and learning for teleoperation: key technologies, applications, and challenges. Cognitive Computation and Systems, 2(2), 33-43.

[8] Wong, S. W., & Crowe, P. (2024). Cognitive ergonomics and robotic surgery. Journal of Robotic Surgery, 18(1), 110. https://doi.org/10.1007/s11701-024-01852-7.

[9] Yang, B., Liu, C., Zhang, L., Teng, L., Tian, J., Xu, S., & Zheng, W. (2025). Novel Design of Three-Channel Bilateral Teleoperation with Communication Delay Using Wave Variable Compensators. Electronics, 14(13), 2595. https://doi.org/10.3390/electronics14132595

[10] Galarza, B. R., Ayala, P., Manzano, S., & Garcia, M. V. (2023). Virtual Reality Teleoperation System for Mobile Robot Manipulation. Robotics, 12(6), 163. https://doi.org/10.3390/robotics12060163

[11] Makhataeva, Z., & Varol, H. A. (2020). Augmented Reality for Robotics: A Review. Robotics, 9(2), 21. https://doi.org/10.3390/robotics9020021

[12] Chon, S. U., Seo, J., Kim, J., Han, S., Park, S., Kim, J. T., Kim, J., & Cho, J. (2022). Design of an Intuitive Master for Improving Teleoperation Task Performance Using the Functional Separation of Actuators: Movement and Gravity Compensation. Actuators, 11(7), 204. https://doi.org/10.3390/act11070204

[13] Zhu, Y., Jiang, B., Chen, Q., Aoyama, T., & Hasegawa, Y. (2023). A shared control framework for enhanced grasping performance in teleoperation. IEEE Access, 11, 69204-69215.

[14] Saran M. (2025). Comparing hand-based and controller-based interactions in virtual reality learning: effects on presence and interaction performance. PeerJ Computer Science, 11, e3168. https://doi.org/10.7717/peerj-cs.3168

[15] Xie, J., Xu, Z., Zeng, J., Gao, Y., & Hashimoto, K. (2025). Human–Robot Interaction Using Dynamic Hand Gesture for Teleoperation of Quadruped Robots with a Robotic Arm. Electronics, 14(5), 860. https://doi.org/10.3390/electronics14050860

[16] Chen, L., Li, C., Fahmy, A., & Sienz, J. (2024). GestureMoRo: an algorithm for autonomous mobile robot teleoperation based on gesture recognition. Scientific Reports, 14(1), 6199.

[17] Lv, H., Kong, D., Pang, G., Wang, B., Yu, Z., Pang, Z., & Yang, G. (2022). GuLiM: A Hybrid Motion Mapping Technique for Teleoperation of Medical Assistive Robot in Combating the COVID-19 Pandemic. IEEE Transactions on Medical Robotics and Bionics, 4(1), 106–117. https://doi.org/10.1109/TMRB.2022.3146621

[18] Wang, Y., & Kasaei, H. (2025, May). Learning Dual-Arm Coordination for Grasping Large Flat Objects. In 2025 IEEE International Conference on Robotics and Automation (ICRA) (pp. 7997-8003). IEEE.

[19] Ramon-Canyameres, P., Palomo-Avellaneda, L., Zaplana, I., & Rosell, J. (2024, September). Dual-arm robotic manipulation using visual guidance. In 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 1-4). IEEE.

[20] M. El-Gohary and J. McNames, "Shoulder and Elbow Joint Angle Tracking With Inertial Sensors," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 9, pp. 2635-2641, Sept. 2012, doi: 10.1109/TBME.2012.2208750.

[21] L. Peppoloni, A. Filippeschi, E. Ruffaldi and C. A. Avizzano, "A novel 7 degrees of freedom model for upper limb kinematic reconstruction based on wearable sensors," 2013 IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 2013, pp. 105-110, doi: 10.1109/SISY.2013.6662551.

[22] T. Lisini Baldi, F. Farina, A. Garulli, A. Giannitrapani and D. Prattichizzo, "Upper Body Pose Estimation Using Wearable Inertial Sensors and Multiplicative Kalman Filter," in IEEE Sensors Journal, vol. 20, no. 1, pp. 492-500, 1 Jan.1, 2020, doi: 10.1109/JSEN.2019.2940612.

[23] Markley, F. L., Cheng, Y., Crassidis, J. L., & Oshman, Y. (2007). Averaging quaternions. Journal of Guidance, Control, and Dynamics, 30(4), 1193-1197. https://doi.org/10.2514/1.28949.

[24] Ken Shoemake. 1985. Animating rotation with quaternion curves. SIGGRAPH Comput. Graph. 19, 3 (Jul. 1985), 245–254. https://doi.org/10.1145/325165.325242

[25] Laidig, D., & Seel, T. (2023). VQF: Highly accurate IMU orientation estimation with bias estimation and magnetic disturbance rejection. Information Fusion, 91, 187-204. https://doi.org/10.48550/arXiv.2203.17024, https://doi.org/10.1016/j.inffus.2022.10.014.

# Acknowledgments

My interest in engineering and robotics began with building and experimenting with machines from an early age. Over time, I came to see engineering not simply as a way to make machines perform more complex tasks, but as a means of helping people and extending human capabilities. Experiences in my family and in social-practice projects made me increasingly aware of the difficulties people face when their physical abilities or access to an environment are limited.

This project grew from that motivation. In hazardous environments such as chemical accidents, fire-rescue sites, or other areas that are difficult or unsafe for direct human access, robots can undertake physical operations while teleoperation preserves human judgment and decision-making. Through this research, I explored virtual reality, motion capture, robotic motion control, and dual-arm coordination, and began to think more deeply about how human motion and intention can be transferred to machines in a natural way.

I am especially grateful to my teacher, Zhang Yu of The Second High School Attached to Beijing Normal University, for guidance throughout the project. From research planning and technical implementation to paper preparation, the advice I received helped me gradually develop the work from an initial idea into a functioning system. I also thank the teachers who assisted with experimental filming and technical support, as well as my parents for providing the time and space needed for sustained learning and experimentation. Open learning resources on Bilibili, YouTube, and GitHub also supported my independent study of robotics, programming, and related technologies.

This research also taught me that a practical teleoperation system is far more than a direct transfer of human motion to a robot. Motion capture, pose computation, communication, visual feedback, and robot execution must work together reliably. In addressing problems involving coordinate transformation, motion mapping, communication, and synchronization, I became more aware of the distance between a functioning prototype and a reliable engineering system. The current system still has limitations in dual-arm coordination, latency, and operational accuracy, leaving substantial room for further improvement.

For me, the significance of this project lies not only in building a robotic teleoperation system, but also in clarifying the direction I hope to continue exploring: using mechanical engineering, robotics, and human–robot interaction to design technologies that extend human capabilities and help solve practical problems.

Generative artificial intelligence tools were used only as auxiliary tools during this

research and manuscript preparation, including language editing and English translation, code-debugging assistance, and draft generation for some figures and charts. AI was not used to generate or fabricate experimental data, invent references, or replace my work in research design, experimental implementation, data processing, or results analysis. All AI-assisted content was manually checked against the original materials, experimental records, and code.

# Appendix A: Bill of Materials

| No. | Component | Model or Specification | Quantity | Function |
|---|---|---|---|---|
| 1 | Inertial Measurement Unit | WT901C-485 | 5 | Measures the orientations of the torso, left/right upper arms, and left/right forearms. |
| 2 | USB-to-RS-485 Converter | Isolated RS-485 converter | 1 | Connects the five addressed IMUs to the Raspberry Pi for sequential quaternion acquisition. |
| 3 | Control Computer | Raspberry Pi 5, 4 GB | 1 | Performs IMU acquisition, time alignment, joint-angle calculation, motion mapping, communication, and servo control. |
| 4 | VR Headset | Meta Quest 3s | 1 | Measures the operator's head orientation and displays the first-person camera view. |
| 5 | VR Controllers | Meta Quest left and right controllers | 1 pair | Sends independent opening and closing commands to the left and right grippers. |
| 6 | Serial Servo Controller | Feetech-compatible bus controller | 1 | Sends synchronized position and speed commands to the 18 servos. |
| 7 | Shoulder Forward/Backward Servos | HLS3950, HL-3950-C001 | 2 | Drives the forward and backward motion of the left and right arms under relatively high loads. |
| 8 | Other Left-Arm Servos | STS3215-C018 | 5 | Drives the remaining five joints of the left arm. |
| 9 | Other Right-Arm Servos | STS3215-C018 | 5 | Drives the remaining five joints of the right arm. |
| 10 | Gripper Servos | STS3215-C018 | 2 | Opens and closes the left and right grippers. |
| 11 | Hand Lateral-Swing Servos | STS3215-C018 | 2 | Controls the left-right swinging motion of the two hand and gripper assemblies. |

| | | | | |
|---|---|---|---|---|
| 12 | Head Yaw Servo | STS3215-C018 | 1 | Rotates the robot head and camera to the left and right. |
| 13 | Head Pitch Servo | STS3215-C018 | 1 | Rotates the robot head and camera upward and downward. |
| 14 | Robot-Mounted Camera | USB or network camera | 1 | Captures the first-person video transmitted to the VR headset. |
| 15 | Servo Power Supply | 12 V high-current DC power supply | 1 | Supplies power to the HLS3950 and STS3215-C018 servos. |
| 16 | Power Distribution Module | High-current distribution board | 1 | Distributes power to the arms, hands, and two-axis head. |
| 17 | RS-485 Communication Cable | Shielded twisted-pair cable | Several | Connects the five WT901C-485 sensors. |