IntuitCap: A 60-DOF Upper-body Motion Capture System for Dexterous Robot Manipulation

Jing Dai[1], Jianbo Yuan[1,2], Yiwen Lu[2], Haohua Zhu[2], Sheng Yi[2], Weixin Yan[1,*]

***Abstract*—Addressing the challenge of limited flexibility and precision in complex task execution by current robotic systems, this paper presents an innovative human-machine interaction solution. We develop an upper-limb 60-DOF motion capture system based on an exoskeleton and smart gloves with integrated tactile feedback, achieving an optimized balance between precision, cost, and ergonomics in structural design. The core innovation lies in our dual-mode kinematic mapping framework, supporting both telexistence in homomorphic virtual environments and teleoperation of heterogeneous mechanical systems. Through systematic experiments, we quantitatively validate that the system achieves millimeter-level end-effector positioning for both arms and fingers. Tests on an experimental platform with dual-arm robots and multi-fingered dexterous hands demonstrate that the system can achieve a high success rate across diverse task scenarios, providing a viable pathway for endowing robots with human-level operational capabilities.**

***Index Terms*—motion capture, teleoperation, telexistence, exoskeleton, dexterous manipulation**

I. Introduction

Despite significant advances in robotics, achieving precise decision-making and motion control in highly redundant, multi-modal, nonlinear robotic systems remains challenging. Full autonomy in complex manipulation tasks often falls short of satisfactory solutions. Teleoperation has emerged as an effective approach to address these limitations, leveraging human intuition and experience to complement robotic capabilities in complex scenarios. This approach supports various applications including semi-autonomous operation [1], imitation learning [2], [3], and direct control [4]. Concurrently, the concept of telexistence [5], which enables human operators to project their presence through robotic avatars, has become increasingly relevant. A fundamental technology underpinning these approaches is the efficient and accurate acquisition of human motion data, which remains a significant technical challenge.

Recent global initiatives, including the DARPA Robotics Challenge and ANA Avatar XPRIZE competition, have underscored these challenges. These competitions demonstrate the critical importance of human-robot collaboration in information-constrained environments, aiming to provide operators with immersive remote experiences while systematically capturing and analyzing human strategies for complex task execution.

[1]School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China

[2]Zhejiang Dexterous Intelligent Tech. Co., Ltd., Zhejiang, China

[*]Corresponding author: xiaogu4524@sjtu.edu.cn

Building on these insights, researchers have developed various human data acquisition systems for teleoperation. Current sensor-based approaches rely on precise, stable multi-modal sensors for bilateral control and environmental adaptation [6]–[8]. While these works demonstrate the potential of motion capture approaches, they often require expensive equipment, complex calibration procedures, or are limited to specific robot morphologies. They also typically introduce latency in remote transmission [9]. In response to these limitations, alternative methodologies have emerged. More economical and lightweight data acquisition systems are being developed to implement teleoperation [10].

This paper presents a low-cost, high-precision 60-DOF upper-body motion capture system combining an ergonomic exoskeleton with smart gloves for real-time teleoperation. Our key contributions include: (1) a comprehensive motion capture system balancing accuracy, cost, and ergonomics; (2) a novel kinematic framework enabling cross-embodiment mapping between heterogeneous robotic systems; (3) real-time performance with minimal latency; and (4) systematic validation demonstrating quantitative improvements across diverse manipulation scenarios.

II. Related Work

Motion capture for teleoperation includes vision-based methods [11], [12] that suffer from occlusion and latency issues [13], and structure-based systems offering higher accuracy. Recent works include GELLO [14] (limited to arm-only), AirExo [15] (homomorphic design), ACE [16] (end-effector only), and DexCap [17] (hybrid approach with occlusion issues). Our system addresses these limitations through a comprehensive wearable design with passive joints and anatomically-adapted gloves, enabling robust upper-limb and finger motion capture with minimal calibration.

III. System Description

Our system (Fig. 1) integrates a 3D-printed exoskeleton with smart gloves featuring gallium-indium liquid metal sensors. The gloves capture 20-DOF finger data at 120Hz with 7.5ms latency and provide fingertip vibrotactile feedback. The exoskeleton uses magnetic encoders and weighs 3.75kg at $<$\$5,000 cost, which is lighter than most existing systems [18].

A. Multi-DOF Motion Capture System

At the core of our motion capture architecture lies the exoskeleton design, which fundamentally determines the system's capabilities for capturing comprehensive upper-body

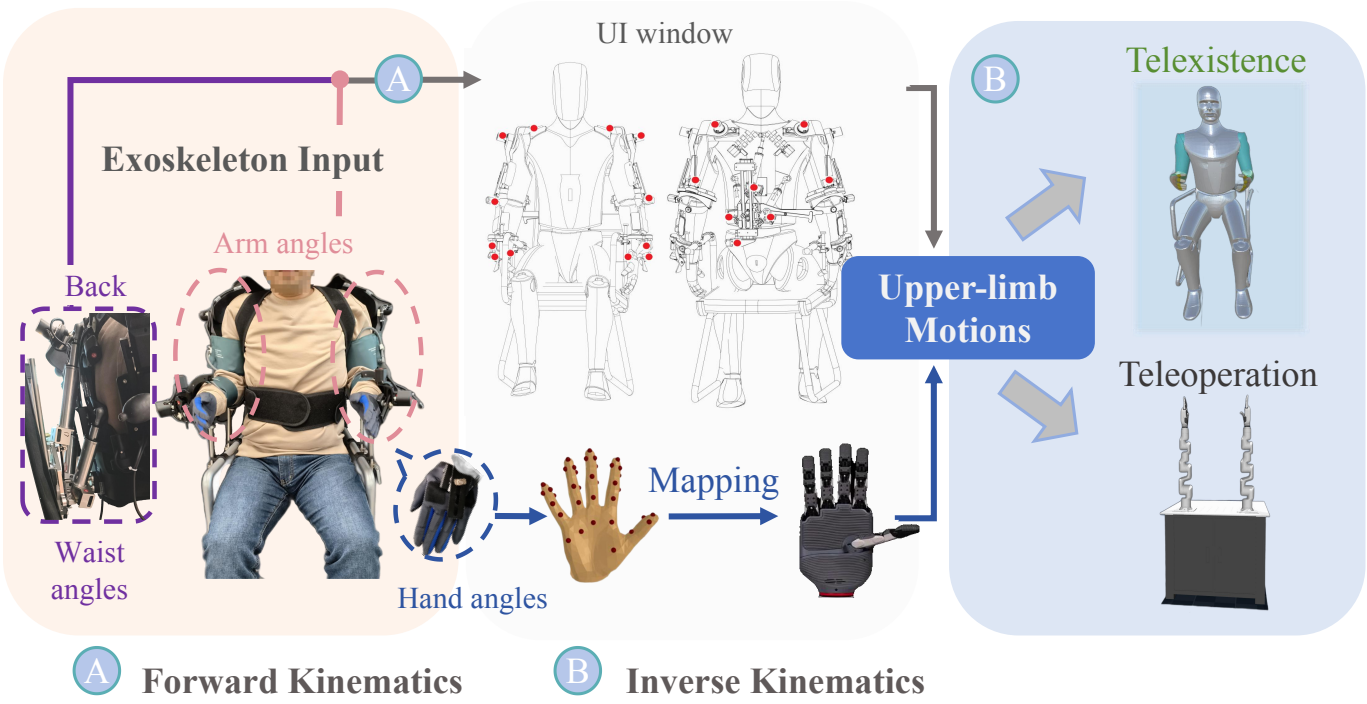

Fig. 1. Overview of the proposed teleoperation framework: motion capture, kinematic mapping, and robot control through digital twin interface.

kinematics. Two mainstream approaches exist in exoskeleton design: designing homomorphic mechanical arms that match target robots, offering better mapping but limited to one-to-one applications [6], [14], [15] and using simple linkages to capture operator's end-effector pose, which may introduce workspace inconsistency and singularity issues [16].

Joint Configuration Design: Our design prioritizes human upper-limb biomechanics through a physical Human-Robot Interaction (pHRI) approach [19]. The mechanical design incorporates several key innovations in degrees-of-freedom (DOF) configuration and structural adaptation, including scapula rotation (q_1), 3-DOF shoulder (q_2-q_4), elbow (q_5), and a 3-DOF wrist mechanism (q_6-q_8). To mitigate joint interference between q_4 and q_6, we implement a novel gear transmission system (Fig. 2).

For finger data acquisition and feedback, we employ a high-precision smart glove system providing 20-DOF finger data (4 DOF per finger) at 120Hz. Kalman filtering reduces noise while preserving motion dynamics before mapping to the robot's kinematic model. The acquired data is constrained by biomechanical limits: MCP joint flexion/extension ($-35°$ to $10°$), MCP abduction/adduction ($0°$ to $60°$), and PIP/DIP joint limits ($0°$ to $110°$). Tactile feedback from remote teleoperation activates linear actuators in the glove's fingertips; once stable contact is established at the controlled end-effector, the corresponding fingertips vibrate to provide feedback, ensuring fine manipulation and stable object grasping.

Dual-Base Configuration System: The system implements a dual-base architecture to address the trade-off between precision and mobility. This design, enabled by magnetic coupling mechanisms shown in Fig. 3(a), allows operators to select between a desktop-oriented setup and a mobile configuration based on task requirements. The desktop configuration incorporates a customized chair enabling full upper-body tracking with 60 DOF monitoring (4-DOF lumbar joint, 8×2-DOF arm joints, and 20×2-DOF hand joints) for tasks demanding high positional stability. The mobile configuration (Fig. 3(b)) prioritizes arm and hand movements while eliminating torso constraints, extending the operational workspace for space-constrained environments where complete range of motion is prioritized over absolute precision.

Real-time Data Acquisition: The system utilizes CAN communication protocol for real-time control and data collection, maintaining a consistent 100Hz sampling frequency for comprehensive 20-DOF upper-limb tracking. Unity3D digital twin validation demonstrates real-time motion mapping for bimanual manipulation.

B. Motion Mapping Framework

Converting raw kinematic data into functional control signals for diverse target platforms requires a sophisticated software architecture bridging human physiological movements with robotic actuation. As illustrated in Fig. 4, our system pipeline connects physical and virtual domains through a layered processing framework. The master controller acquires hand positions via forward kinematics from exoskeleton data, while instrumented gloves provide comprehensive finger gesture information.

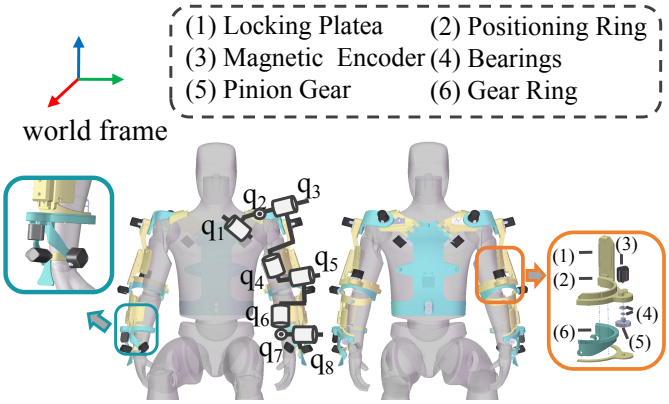

Fig. 2. Exploded structural diagram of the exoskeleton's 8-DOF passive joint configuration with orthogonal joint pairs and gear transmission mechanism between q4 and q6.

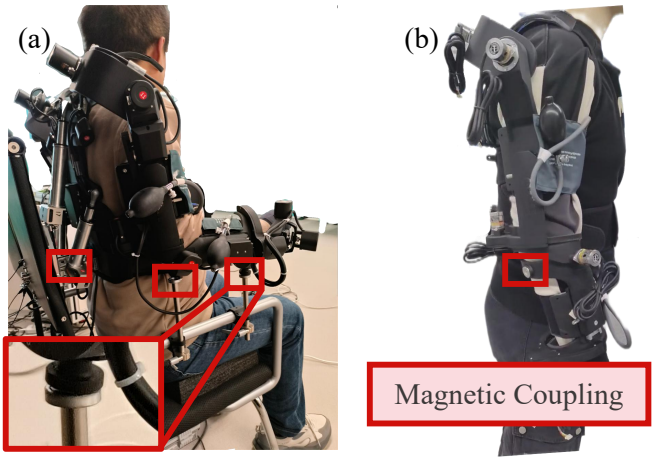

Fig. 3. Dual operational configurations of the proposed exoskeleton system: a) stationary setup with integrated chair for full-body motion capture, featuring lumbar joint tracking capabilities optimized for high-precision manipulation tasks; b) portable implementation prioritizing upper-limb tracking, providing users with enhanced mobility and wider field of view during operation.

Our system employs a dual-pathway mapping strategy to address kinematic correspondence challenges between heterogeneous structures. For telexistence applications involving analogous morphology, we utilize an absolute alignment methodology with calibrated finger joint alignment that preserves spatial relationships between operator and virtual embodiment, enabling direct mapping between corresponding joint spaces. When controlling non-anthropomorphic robotic systems, a relative mapping approach establishes end-effector correspondence with optimized finger mapping while accommodating workspace disparities between human and robot domains.

This bifurcated architecture ensures seamless adaptation across diverse operational scenarios while maintaining intuitive control relationships. Bidirectional information flow is achieved through contact-aware feedback via five fingertip-mounted vibration motors and I2C multiplexers, creating a closed-loop control system capable of transferring precise manipulation motions across virtual and physical environments.

1) End-effector Pose Mapping: Our inverse kinematics framework employs optimization-based approaches [20] using the damped least squares method with Denavit-Hartenberg (DH) parameters. The end-effector transformation is:

$$T_c = T_1(q_1) \cdot T_2(q_2) \cdot \ldots \cdot T_8(q_8) \tag{1}$$

where $T_c \in SE(3)$ is the end-effector pose, q_i the ith joint angle, and T_i the transformation matrix.

The inverse kinematics is solved iteratively using:

$$q_{k+1} = q_k + \alpha(J^T J + \lambda I)^{-1} J^T \Delta x \tag{2}$$

subject to $q_{\min} \leq q \leq q_{\max}$, where λ, α are damping and step parameters, Δx the pose error, and J the Jacobian.

For anthropomorphic hand mapping, we employ:

$$q_h^{(c)} = k_j \cdot \min(\max(q_f^{(c)}, q_f^{\min}), q_f^{\max}) \tag{3}$$

where q_h are simulated hand angles (20 DOF), $q_f^{(c)}$ the calibrated human hand angles, k_j the mapping direction (± 1), and $q_f^{\min}, q_f^{\max}$ the joint limits.

For 12-DOF robotic hands, we employ weighted optimization with joint reduction [21]:

$$\theta^* = \arg\min f(\theta), \quad \text{s.t. } c_i(\theta) \leq 0, c_e(\theta) = 0, l_i \leq \theta_i \leq u_i \tag{4}$$

where θ are the 12-DOF joint angles, $f(\theta)$ the objective function, $c_i(\theta), c_e(\theta)$ the constraints, and l_i, u_i the joint limits. The objective combines:

$$f(\theta) = w_p f_p(\theta) + w_o f_o(\theta) + w_v f_v(\theta) \tag{5}$$
$$f_p(\theta) = \|p_g - \text{FK}(\theta)\|_2^2 \tag{6}$$
$$f_o(\theta) = \log(1 + \|q_g - \text{FK}(\theta)\|_2^2) \tag{7}$$
$$f_v(\theta) = \|\dot{\theta}\|_2^2 \tag{8}$$

where $f_p(\theta)$, $f_o(\theta)$, and $f_v(\theta)$ are position error, orientation error, and velocity minimization terms.

The system uses magnetic calibration with seated upright posture as zero position.

2) Workspace Alignment: In teleoperation, workspace mismatch requires precise calibration. F. Zacharis [22] analyzed spatial matching for human-worn exoskeletons but did not address motion mapping to heterogeneous controlled robotic arms. We employ absolute correspondence for similar workspaces and relative mapping for heterogeneous systems to maintain spatial consistency.

Similar workspace scenarios: We developed a digital twin system in Unity3D to validate the 60-DOF motion data quality, which supports high-quality URDF model rendering, enabling real-time comparison between actual and simulated

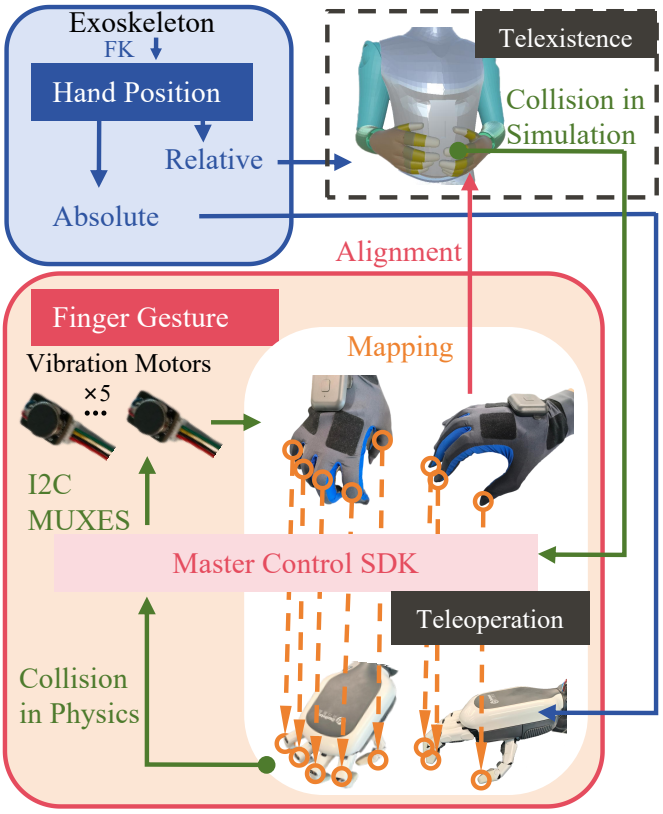

Fig. 4. System architecture with dual-pathway mapping: direct spatial correspondence for virtual environments and adaptive kinematic translation for robotic platforms. Vibrotactile feedback completes the control loop.

movements. During initialization, the exoskeleton's initial state is aligned with the virtual human model to ensure consistent relative poses. For pose mapping in the similar workspace scenarios, we follow:

$$T_h^{\text{aim}} = T_{hb}^c \cdot T_c \cdot T_{hc}^e \tag{9}$$

where T_{hb}^c represents the fixed pose transformation from the base virtual humanoid frame to base exoskeleton frame, T_{hc}^e denotes the fixed end-point transformation from end-effector frame of exoskeleton end-effector frame to wirtual humanoid frame, and T_h^{aim} represents the pose transformation matrix for humanoid telexistence.

Dissimilar workspace scenarios: In teleoperation scenarios with heterogeneous robotic arms, we need to align the initial postures of the exoskeleton system and the robotic arm. This alignment allows for absolute angle control between the two systems. Due to the inconsistency between the operator's workspace and the robot arm's end-effector space, and the inability to obtain relative positions between base coordinate frames, we implement relative movement control when manipulating the robotic arm. This approach maintains positional consistency between the control and slave endpoints relative to their initial states, rather than attempting absolute spatial correspondence. Specifically, we follow:

$$R_m^{\text{aim}} = R_{hb}^c \cdot R_{hc}^e \tag{10}$$

$$P_m^{\text{aim}} = P_m^o + R_{hb}^c (P_{hc}^e - P_{hc}^{e0}) \tag{11}$$

where R and P denote rotation and position respectively, and superscript 0 represents the exoskeleton and robot arm's initial end-effector pose. This mapping approach demonstrates robust performance across different robot models and workspace sizes, offering improved practicality compared to direct pose matching.

IV. Experimental Evaluation

We conducted comprehensive experiments focusing on precision and portability.

A. Precision Validation

We employed the Chingmu MC1300 high-precision motion capture system (accuracy: ±0.02mm) as ground truth to validate our system's accuracy. Data was collected over 10 minutes encompassing diverse upper-limb poses and movements.

For the exoskeleton component, we tracked bilateral hand markers and compared end-effector positions derived through our system's forward kinematics against ground truth measurements. TABLE I shows evaluation metrics including Median Absolute Error (MAE), 95th percentile absolute error, and Truncated Root Mean Square Error (T-RMSE).

The results demonstrate consistent motion capture performance at centimeter-level precision across all spatial dimensions. The left arm exhibited slightly better overall accuracy (9.19mm MAE) compared to the right arm (10.59mm MAE), with individual coordinate errors below 6.52mm MAE.

Fig. 5 provides a temporal comparison between our exoskeleton system (blue) and motion capture ground truth (red) across three spatial dimensions. While some oscillations occur during large ranges of motion, the movement tendency consistently aligns with ground truth measurements, confirming the system's suitability for teleoperation applications.

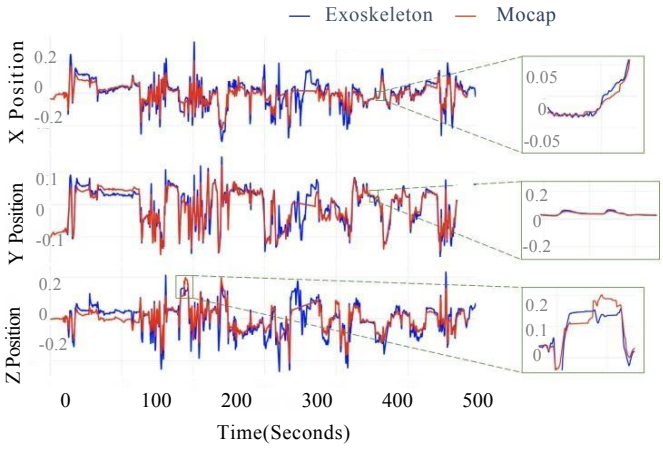

Fig. 5. Temporal comparison of exoskeleton system (blue) vs. optical motion capture ground truth (red) showing X, Y, Z position trajectories over 10-minute trial.

TABLE I
END-EFFECTOR POSITIONING ACCURACY VS. OPTICAL MOTION CAPTURE GROUND TRUTH.

Coordinate	MAE (mm)	95% (mm)	T-RMSE (mm)
Right x	6.31	19.36	11.24
Right y	5.46	13.86	8.59
Right z	6.52	25.07	13.69
Right overall	10.59	34.57	19.69
Left x	5.85	17.27	10.08
Left y	4.11	11.07	6.49
Left z	5.77	18.37	10.65
Left overall	9.19	27.54	16.04

For the hand tracking component, we evaluated positioning error of each finger joint (TABLE II). The data shows consistent millimeter-level precision across all fingers and joint types. MCP joints demonstrated highest accuracy (0.35mm-2.57mm), with errors gradually increasing toward distal joints (TIP joints: 5.87mm-6.45mm).

TABLE II
SMART GLOVE JOINT-LEVEL ACCURACY MEASUREMENTS ACROSS ALL FINGERS AND JOINT TYPES.

	MCP	PIP	DIP	TIP
Thumb	2.57	3.32	3.88	6.22
Index	1.16	3.07	3.21	5.87
Middle	0.35	3.12	3.15	5.96
Ring	1.24	2.95	3.17	6.02
Pinky	1.77	3.71	3.96	6.45

This high-fidelity positioning for both arms and fingers enables our platform to support precise manipulation tasks requiring fine control. These validation results confirm that our system maintains average errors within acceptable ranges for practical teleoperation applications, successfully balancing lightweight, ergonomic design and precision motion capture capabilities.

B. Cross-platform Adaptability Tests

We evaluated the system's portability through integration with dual JAKA mini dual-arms equipped with DexHand 021 dexterous hands (12 active DOF), shown in Fig. 6. A single trained operator with 2 hours of prior system training performed all experiments. **Experimental Setup:** The experimental protocol comprised pick-and-place operations using: (1) tennis ball (67mm diameter), (2) doll (200×150×80mm), (3) towel (300×200mm), and (4) water bottle (500ml capacity). Additional tasks included: (5) liquid pouring (200ml water), and (6) finger trajectory tracking along a measuring tape (20cm-50cm). Each task was repeated 20 times.

Task success criteria: For tasks 1-3, success required lifting objects 0.15m while maintaining stable grasp. For task 4 (Bottle Grasping), success was defined as achieving stable grasp within a 30cm × 40cm area without toppling. Task 5 (Bimanual Liquid Pouring) required complete liquid transfer without spillage [23]. Task 6 (Measuring Tape Tracking) evaluated precision by measuring positional deviation at 50cm position. Success rate (SR) was defined as: SR = 1 - (e/3), where e represents measured positional deviation in centimeters and 3cm constitutes maximum tolerable error.

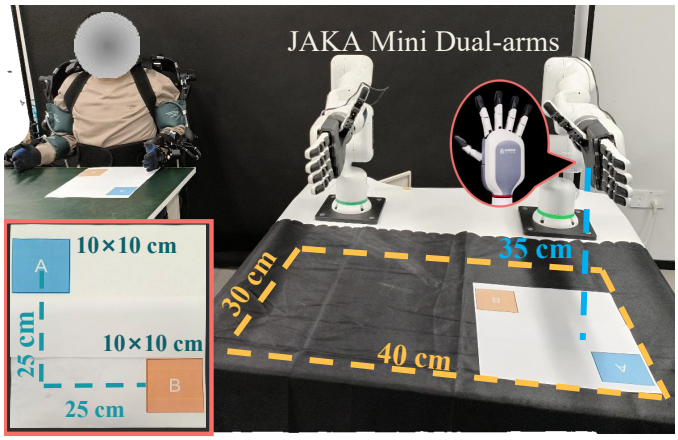

Fig. 6. Experimental teleoperation platform with JAKA Mini dual-arm system and five-fingered dexterous hands. Workspaces are spatially mapped between operator and robot domains.

Results Summary: Selected experimental scenarios are shown in Fig. 7. Analysis reveals consistently high success rates across all scenarios (80-95%), with task completion times scaling proportionally with manipulation complexity. TABLE III summarizes the results.

TABLE III
TASK PERFORMANCE EVALUATION ACROSS DIFFERENT MANIPULATION SCENARIOS.

Task	Success Rate (%)	Avg. Time (s)	Trials
Tennis Ball	90	17.5	20
Doll	95	15.8	20
Towel	85	26.2	20
Bottle Grasp	85	28.4	20
Liquid Pouring	80	74.5	20
Tape Tracking	92	12.3	20

The system demonstrates efficient performance in structured object manipulation tasks, achieving 90% and 95% success rates for tennis ball and doll pick-and-place operations with completion times of 17.5s and 15.8s respectively. This high performance extends to deformable object manipulation (towel: 85% success rate, 26.2s) and precision grasping tasks (bottle: 85% success rate, 28.4s), indicating robust cross-domain applicability.

The bimanual liquid pouring task yielded an 80% success rate with longer execution time (74.5s). This task presents unique challenges in coordinating precise relative positioning between dual robotic arms while maintaining stable grasp forces. Despite these challenges, the consistently high success rate across all manipulation scenarios demonstrates the practical utility and operational robustness of our teleoperation framework.

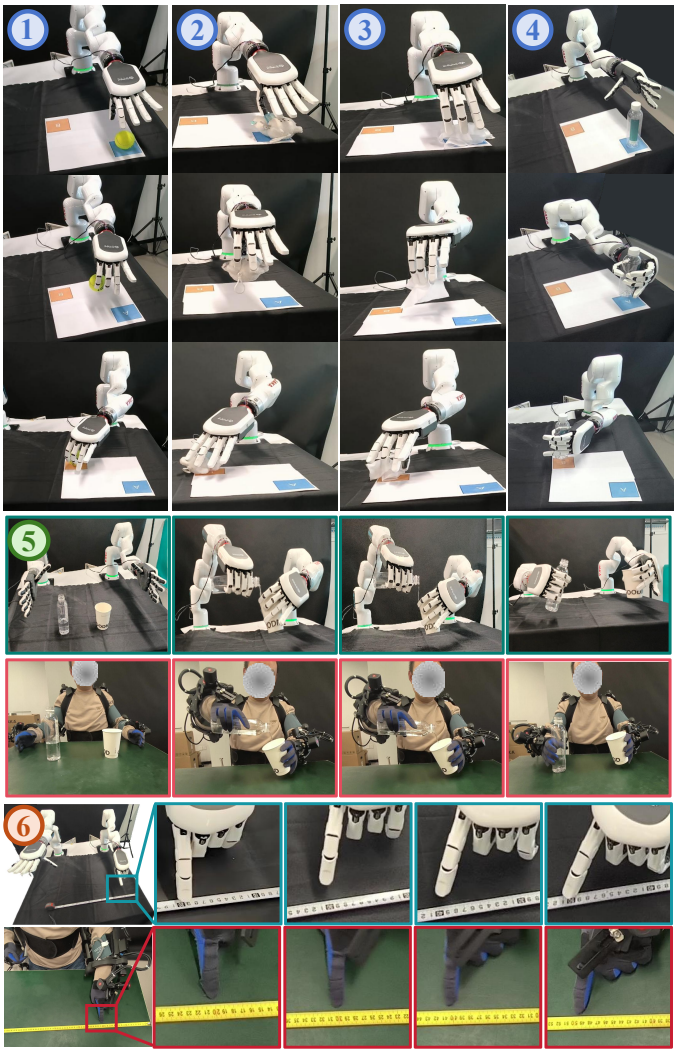

Fig. 7. Experimental validation across multiple manipulation tasks: (1-4) pick-and-place operations, (5) bimanual liquid transfer, and (6) precision tracking.

V. Conclusion

We have developed an experimental platform that effectively bridges human anthropomorphic movement patterns with robotic kinematic structures, enabling intuitive manipulation of objects with diverse physical properties. Our contributions include: (1) a lightweight 60-DOF upper-body motion capture system balancing accuracy, cost, and ergonomics; (2) a kinematic framework addressing cross-embodiment mapping between non-homomorphic configurations; (3) an integrated architecture delivering real-time performance with minimal latency; and (4) systematic validation showing quantitative accomplishments across diverse manipulation scenarios. Arm and finger joint tracking with quantitatively validated precision, coupled with high success rates observed in teleoperation tasks, confirms the system's ability to capture subtle hand movements while maintaining mobility and workspace flexibility essential for practical applications.

References

[1] S. Chen et al., "ASHA: Assistive Teleoperation via Human-in-the-Loop Reinforcement Learning," in 2022 International Conference on Robotics and Automation (ICRA), 2022, pp. 7505–7512.

[2] T. Z. Zhao et al., "Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware," arXiv preprint arXiv:2304.13705, 2023.

[3] M. Seo et al., "Deep Imitation Learning for Humanoid Loco-manipulation Through Human Teleoperation," in 2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids), 2023, pp. 1–8.

[4] Y. Ishiguro et al., "Bilateral Humanoid Teleoperation System Using Whole-Body Exoskeleton Cockpit TABLIS," IEEE Robot. Autom. Lett., vol. 5, no. 4, pp. 6419–6426, 2020.

[5] K. Darvish et al., "Teleoperation of Humanoid Robots: A Survey," IEEE Trans. Robot., vol. 39, no. 3, pp. 1706–1727, 2023.

[6] M. Schwarz et al., "NimbRo Avatar: Interactive Immersive Telepresence with Force-Feedback Telemanipulation," in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 5312–5319.

[7] H. Liu et al., "High-Fidelity Grasping in Virtual Reality using a Glove-based System," in 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 5180–5186.

[8] S. Park et al., "A Whole-Body Integrated AVATAR System: Implementation of Telepresence With Intuitive Control and Immersive Feedback," IEEE Robot. Autom. Mag., pp. 2–10, 2023.

[9] P. F. Hokayem and M. W. Spong, "Bilateral teleoperation: An historical survey," Automatica, vol. 42, no. 12, pp. 2035–2057, 2006.

[10] M. Forouhar et al., "A Tactile Lightweight Exoskeleton for Teleoperation: Design and Control Performance," in 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024, pp. 178–183.

[11] S. Li et al., "A Mobile Robot Hand-Arm Teleoperation System by Vision and IMU," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020, pp. 10900–10906.

[12] S. Han et al., "Online optical marker-based hand tracking with deep labels," ACM Trans. Graph., vol. 37, no. 4, pp. 1–10, 2018.

[13] A. Handa et al., "DexPilot: Vision-Based Teleoperation of Dexterous Robotic Hand-Arm System," in 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 9164–9170.

[14] P. Wu et al., "GELLO: A General, Low-Cost, and Intuitive Teleoperation Framework for Robot Manipulators," arXiv preprint arXiv:2309.13037, 2024.

[15] H. Fang et al., "Airexo: Low-cost exoskeletons for learning whole-arm manipulation in the wild," in 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024, pp. 15031–15038.

[16] S. Yang et al., "ACE: A Cross-Platform Visual-Exoskeletons System for Low-Cost Dexterous Teleoperation," arXiv preprint arXiv:2408.11805, 2024.

[17] C. Wang et al., "DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation," arXiv preprint arXiv:2403.07788, 2024.

[18] L. Zhao, T. Yang, Y. Yang, and P. Yu, "A Wearable Upper Limb Exoskeleton for Intuitive Teleoperation of Anthropomorphic Manipulators," Machines, vol. 11, no. 4, Art. no. 4, Apr. 2023, doi: 10.3390/machines11040441.

[19] M. A. Gull et al., "A Review on Design of Upper Limb Exoskeletons," Robotics, vol. 9, no. 1, 2020.

[20] S. Chiaverini, "Singularity-robust task-priority redundancy resolution for real-time kinematic control of robot manipulators," IEEE Trans. Robot. Autom., vol. 13, no. 3, pp. 398–410, 1997.

[21] D. Rakita et al., "RelaxedIK: Real-time Synthesis of Accurate and Feasible Robot Arm Motion," in Robotics: Science and Systems XIV, 2018.

[22] F. Zacharias et al., "Workspace comparisons of setup configurations for human-robot interaction," in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010, pp. 3117–3122.

[23] B. Calli et al., "Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set," IEEE Robot. Autom. Mag., vol. 22, no. 3, pp. 36–52, 2015.