MATER: Mutually Aware Framework for Teleoperated-robot with Extended Reality

Ziliang Zhang
University of California, Riverside
Riverside, CA, USA
zzhan357@ucr.edu

Cong Liu
University of California, Riverside
Riverside, CA, USA
congl@ucr.edu

Hyoseung Kim
University of California, Riverside
Riverside, CA, USA
hyoseung@ucr.edu

Abstract

Robot teleoperation with extended reality (XR teleoperation) enables intuitive interaction by mapping user motions to remote robots with real-time 3D feedback. However, existing systems suffer from large completion delays and trajectory deviations under prolonged network latency, rooted in their exclusive reliance on network communication and strict synchronous execution architecture. Moreover, network fluctuations destabilize teleoperation accuracy, while dynamic user motions amplify teleoperation errors.

We present MATER, an end-to-end XR teleoperation framework that introduces a mutually-aware architecture in which each side reconstructs its counterpart's delayed or missing state to decouple the execution from network dependency. MATER includes latency-adaptive input window and user motion gap interpolation techniques to handle unstable network communication. It also proposes motion-driven robot state rollback and robot trajectory coordination to handle complex motions. The key idea behind these techniques is to adapt on the fly by reshaping reconstruction and filling gaps as network conditions fluctuate, and by realigning states when motions become fast or complex. Together with lightweight local synchronization and bandwidth optimizations, these system-level advances make MATER resilient to both network and motion dynamics.

We implement MATER across three hardware settings, including simulated and physical robots, and evaluate it on 9,500 real-world teleoperation trials from the RoboSet dataset [1], covering single- and multi-step missions. Compared to state-of-the-art XR teleoperation frameworks, MATER reduces teleoperation error by up to 69.8% on WLAN and 73.1% on cellular networks with only 6.7% maximum runtime overhead. It also shortens mission completion time by up to 47.7%, enabling smoother teleoperation. A real-world case study on ten stationary and mobile missions further shows MATER achieves up to 37.7% faster completion while lowering average teleoperation error by up to 57.2%. MATER code is available at: https://github.com/rtenlab/mater

CCS Concepts

• **Computer systems organization** → *Robotics*; • **Human-centered computing** → *Virtual reality*; *Mixed / augmented reality*; • **Networks** → *Network performance evaluation.*

Keywords

Software Architecture, Extended Reality, Robotics, Teleoperation.

ACM Reference Format:
Ziliang Zhang, Cong Liu, and Hyoseung Kim. 2026. MATER: Mutually Aware Framework for Teleoperated-robot with Extended Reality. In *Abstracts of the 2026 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS Abstracts '26), June 08–12, 2026, Ann Arbor, MI, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3801489.3806876

1 Introduction

Robot teleoperation via Extended Reality (XR teleoperation) emerges as a transformative technology in industrial automation, medical robotics, autonomous robot training, and humanoid robotics [2–11]. In XR teleoperation, an XR device uses onboard sensors to capture user motions and generate user poses, which are transmitted to a remote robot. The robot imitates these poses and sends back its own state and the surrounding environment. The XR device then fuses the latest user pose with the returned robot state and environment to render a 3D frame on a head-mounted display (HMD). By referencing both user and robot poses in this frame, the user adjusts next motion in real time, repeating the cycle until the mission is completed.

Today's XR teleoperation systems remain fragile, suffering from significant delays between user motion and robot feedback, primarily due to the network latency and the long control-feedback pipeline [2, 6, 7, 10, 12–16]. Unlike traditional teleoperation, immersive 3D rendering in XR teleoperation makes network delays highly perceptible as persistent discrepancies between the user pose and the robot pose in the displayed frame, known as teleoperation error. Continuing motion despite this error makes the robot drift from the user's intent, while pausing to wait slows mission completion. The root cause is *the synchronous execution architecture*: current systems require both sides to wait for network updates to maintain state consistency. This creates tight coupling, where the robot needs XR-to-robot messages to follow user poses, while the XR device requires robot-to-XR updates to render the robot state and environment. This dual reliance leaves the control-feedback loop highly vulnerable to even the modest network delay or fluctuation (**C1**), and highly sensitive to even the slightest increase in user motion speed or curvature (**C2**).

Contributions. We present **MATER**: **M**utually **A**ware **T**eleoperated-**E**xtended **R**eality, an end-to-end XR teleoperation framework that breaks the exclusive network reliance through bidirectional state reconstruction. MATER allows each side to locally reconstruct the other's delayed or missing information using local sensing and prior received information, decoupling robot control

---

SIGMETRICS Abstracts '26, Ann Arbor, MI, USA

ACM ISBN 979-8-4007-2570-8/2026/06
https://doi.org/10.1145/3801489.3806876

and XR visualization from round-trip communication. Therefore, each side maintains its own local state, enabling globally asynchronous execution. This architectural shift from the conventional, strict network-dependent XR teleoperation makes MATER naturally resilient to prolonged network delay.

Rather than relying on any single component, MATER's contribution lies in a coordinated system-level design that jointly addresses both challenges under network-induced latency, bandwidth variations, and motion-induced dynamics. Specifically, to address network fluctuation in (C1), MATER introduces a latency-adaptive input window technique on both sides that dynamically adjusts the input sizes based on observed latency, stabilizing reconstruction accuracy under unstable networks. On the robot side, a user motion gap interpolation component further mitigates faulty planner outputs when pose data is dropped or arrives out-of-order. To handle motion dynamics in (C2), a motion-driven robot state rollback reverts the reconstructed robot state according to the current user motion, while a robot trajectory coordination mechanism re-adjusts the robot's actual movement command based on reconstructed robot states and robot physical constraints, keeping robot trajectories aligned with XR displayed frames.

In addition, while MATER's mutually-aware paradigm naturally leads to an asynchronous execution model between XR and robot globally, MATER employs a synchronous execution model for local components on each device to eliminate instability from resource contention, where competing GPU kernels and CPU threads cause jitters, skipped executions, and stale data propagation. Finally, to address bandwidth variation that risks further pose drops, MATER implements bandwidth-adaptive data flow scaling, which prioritizes pose transmission by reducing the data that has little impact on the current cycle of teleoperation. The key idea is to adapt on the fly: MATER reshapes the system parameters to reconstruct motions when networks fluctuate, and realigns states and trajectories when user motions become faster or more complex, providing resilience with minimal runtime overhead.

We evaluate the MATER framework on three hardware settings over five motion patterns derived from 9,500 real robot teleoperation trials in the RoboSet dataset [1], a widely adopted dataset for autonomous robot training that covers 54 single- and multi-step missions. We experiment with four different network environments covering ubiquitous WLAN and cellular networks. Compared to the state-of-the-art XR teleoperation framework [13], MATER reduces the teleoperation error by up to 69.8% in WLAN and 73.1% in cellular networks, while shortening the mission completion time by up to 47.7% on a real robot manipulator [17]. We further conduct a real-world case study with 10 missions, including both stationary and mobile scenarios. In these trials, MATER achieves up to 37.7% faster mission completion time while maintaining on average 25.4% lower teleoperation error in stationary missions and 57.2% lower error in mobile missions. This abstract summarizes our full paper [18].

2 XR Teleoperation Background

Fig. 1 illustrates that a user performs a hand sweeping mission using the conventional software architecture that are widely adopted in state-of-the-art XR teleoperation [2, 7, 10, 13, 19, 20]. Initially, the Head CAM (camera) and Inertial Measurement Unit (IMU) capture the surrounding environment and the user's head inertial data, and the Hand CAM captures the user's hand. These sensor data are fed into a **control pipeline** (blue lines in the figure), which comprises *tracking* and *control* modules on the XR and robot, respectively. The tracking module constructs the *user pose* information by calculating the user's head movement from Head CAM and IMU data and the user's hand movement from Hand CAM data. Then, the control module on the robot receives the user pose from the XR and generates control commands to move the robot's end-effector accordingly.

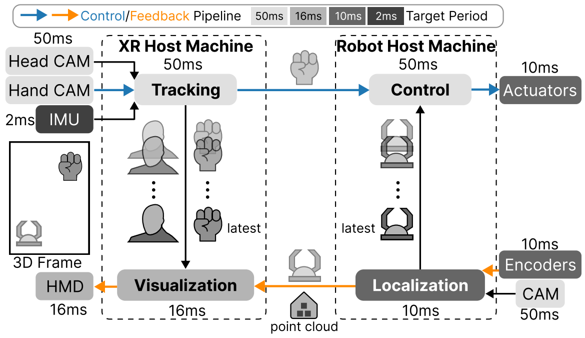

Figure 1: Existing XR teleoperation framework.

Since the user relies on 3D frames to observe the remote robot's pose and perform subsequent motions, the **feedback pipeline** (orange lines) comprises *localization* and *visualization* modules on the robot and XR, respectively. The localization module produces robot pose for the end-effector, and point cloud data for surrounding objects in the environment [2, 13]. Then, the visualization module on the XR side combines the latest user pose (captured and processed in the latest sampling period from the tracking module) with the received robot pose and point cloud to render 3D frames (typically at 60 FPS).

3 MATER Design Overview

We present **MATER**, which takes a mutually-aware paradigm to decouple the execution pipelines from the network dependency. Today's XR and robot systems already have sufficient sensing and computation resources to locally reconstruct the state of their counterpart, allowing each side to continue execution without relying solely on network communication. Building on this insight, MATER decouples both control and feedback pipelines from network dependency through the following major ideas: (i) Bidirectional reconstruction of time-shifted states, (ii) asynchronous global execution, and (iii) control/data plane separation.

Bidirectional Reconstruction of Time-Shifted States. Illustrated in Fig. 2 blue and orange arrows, MATER lets each side maintain a locally reconstructed, time-shifted state of its counterpart by running local reconstruction pipelines. This differs from existing prediction-based techniques that only project a time-forwarded version of a single side's state [21–24], as MATER allows both sides to reconstruct the missing information. The XR reconstructs the remote robot state based on the latest local user pose and a series of previously received robot poses, while the robot reconstructs

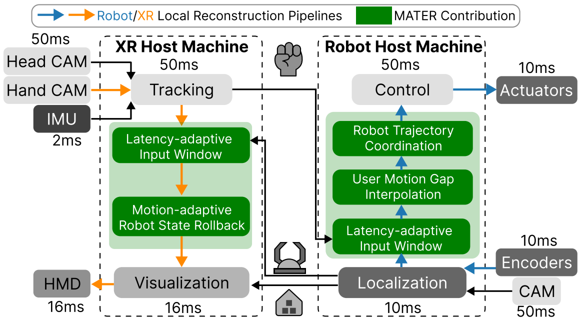

Figure 2: MATER Software Architecture

user motion with the latest local robot pose and a series of received user poses. MATER's architecture does not restrict the choice of reconstruction algorithm; Time-forwarding [22], Extended Kalman Filter (EKF) [25], and Multilayer Perceptron [26] can all serve as the reconstruction model. Our implementation uses EKF by default due to its less intensive computation and low runtime overhead. However, since reconstruction relies on poses transmitted over the network, its quality degrades as latency increases (C1). To maintain reconstruction quality under changing network conditions, we introduce a latency-adaptive input window technique that adjusts input sizes based on the runtime network latency.

Asynchronous Global and Synchronized Local Execution. The bidirectional reconstruction approach makes it possible to switch to asynchronous execution between the XR and robot. Unlike prior frameworks tied to a single, tightly synchronized state, MATER enables *asynchronous global states*, where XR visualization and robot control can progress independently and synchronize only when network messages arrive. With this asynchronous global execution, the challenges C1 (network-dependent performance) and C2 (motion-driven accuracy) can be addressed by optimizing the local reconstruction methods separately without causing interference to the global execution. On the XR side, a motion-adaptive robot state rollback selects the reconstructed robot state based on current motion to address C2. On the robot side, MATER implements a user motion gap interpolation technique that generates missing or delayed user poses during network communication to address C1. It then proposes a robot trajectory coordination technique to adjust the control command based on the poses displayed to the user and the physical constraints of the robot to address C2. While MATER enables asynchronous execution between XR and robot, it enforces the synchronous execution of local components on each side to avoid contention on local computation resources that causes runtime delay.

Acknowledgment

This work was supported by the National Science Foundation (NSF) grants 1943265, 2312395, 2300525, 2312397, and 2343653.

References

[1] Vikash Kumar, Rutav Shah, Gaoyue Zhou, Vincent Moens, Vittorio Caggiano, Abhishek Gupta, and Aravind Rajeswaran. Robohive: A unified framework for robot learning. *Advances in Neural Information Processing Systems*, 36, 2024.

[2] Abdeldjallil Naceri, Dario Mazzanti, Joao Bimbo, Domenico Prattichizzo, Darwin G Caldwell, Leonardo S Mattos, and Nikhil Deshpande. Towards a virtual reality interface for remote robotic teleoperation. In *2019 19th International Conference on Advanced Robotics (ICAR)*, pages 284–289. IEEE, 2019.

[3] Ruishuo Xu, Weijun Wang, Wei Feng, Zhaokun Zhou, Boyoung An, Ruizhen Gao, and Kaichen Zhou. Design of a human-robot interaction system for robot teleoperation based on digital twinning. In *2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*, pages 720–726. IEEE, 2022.

[4] Intuitive. Da Vinci 5, 2025. https://www.intuitive.com/en-us/products-and-services/da-vinci/5/.

[5] Kateryna Zinchenko and Kai-Tai Song. Autonomous endoscope robot positioning using instrument segmentation with virtual reality visualization. *IEEE Access*, 9:72614–72623, 2021.

[6] David Black, Yas Oloumi Yazdi, Amir Hossein Hadi Hosseinabadi, and Septimiu Salcudean. Human teleoperation-a haptically enabled mixed reality system for teleultrasound. *Human–Computer Interaction*, 39(5-6):529–552, 2024.

[7] Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. Open-television: Teleoperation with immersive active visual feedback. *arXiv preprint arXiv:2407.01512*, 2024.

[8] Aadhithya Iyer, Zhuoran Peng, Yinlong Dai, Irmak Guzey, Siddhant Haldar, Soumith Chintala, and Lerrel Pinto. Open teach: A versatile teleoperation system for robotic manipulation. *arXiv preprint arXiv:2403.07870*, 2024.

[9] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. *arXiv preprint arXiv:2406.08858*, 2024.

[10] Matthias Hirschmanner, Christiana Tsiourti, Timothy Patten, and Markus Vincze. Virtual reality teleoperation of a humanoid robot using markerless human upper body pose imitation. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pages 259–265. IEEE, 2019.

[11] Mohammad Bakhshalipour and Phillip B Gibbons. Agents of autonomy: A systematic study of robotics on modern hardware. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 7(3):1–31, 2023.

[12] Yunpeng Su, Leo Lloyd, Xiaoqi Chen, and J Geoffrey Chase. Latency mitigation using applied hmms for mixed reality-enhanced intuitive teleoperation in intelligent robotic welding. *The International Journal of Advanced Manufacturing Technology*, 126(5):2233–2248, 2023.

[13] David Whitney, Eric Rosen, Daniel Ullman, Elizabeth Phillips, and Stefanie Tellex. Ros reality: A virtual reality framework using consumer-grade hardware for ros-enabled robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. IEEE, 2018.

[14] Minzhao Lyu, Rahul Dev Tripathi, and Vijay Sivaraman. Metavradar: Measuring metaverse virtual reality network activity. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 7(3):1–29, 2023.

[15] Shihan Lin, Yi Zhou, Xiao Zhang, Todd Arnold, Ramesh Govindan, and Xiaowei Yang. Tiered cloud routing: Methodology, latency, and improvement. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 9(1):1–41, 2025.

[16] Sandeepa Bhuyan, Shulin Zhao, Ziyu Ying, Mahmut T Kandemir, and Chita R Das. End-to-end characterization of game streaming applications on mobile platforms. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(1):1–25, 2022.

[17] Kinova Robotics. Kinova ros. https://github.com/Kinovarobotics/kinova-ros.

[18] Ziliang Zhang, Cong Liu, and Hyoseung Kim. MATER: Mutually aware framework for teleoperated-robot with extended reality. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 10(2), June 2026.

[19] Marco Gallipoli, Sara Buonocore, Mario Selvaggio, Giuseppe Andrea Fontanelli, Stanislao Grazioso, and Giuseppe Di Gironimo. A virtual reality-based dual-mode robot teleoperation architecture. *Robotica*, pages 1–24, 2024.

[20] Dinh Tung Le, Sheila Sutjipto, Yujun Lai, and Gavin Paul. Intuitive virtual reality based control of a real-world mobile manipulator. In *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 767–772. IEEE, 2020.

[21] Henrikke Dybvik, Martin Løland, Achim Gerstenberg, Kristoffer Bjørnerud Slåttsveen, and Martin Steinert. A low-cost predictive display for teleoperation: Investigating effects on human performance and workload. *International Journal of Human-Computer Studies*, 145:102536, 2021.

[22] Thomas B Sheridan. Space teleoperation through time delay: Review and prognosis. *IEEE Transactions on robotics and Automation*, 9(5):592–606, 1993.

[23] Neil McHenry, Jason Spencer, Patrick Zhong, Jeremy Cox, Michael Amiscaray, KC Wong, and Gregory Chamitoff. Predictive xr telepresence for robotic operations in space. In *2021 IEEE Aerospace Conference (50100)*, pages 1–10. IEEE, 2021.

[24] Hem Regmi and Sanjib Sur. Argus: Predictable millimeter-wave picocells with vision and learning augmentation. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(1):1–26, 2022.

[25] Jay H Lee and N Lawrence Ricker. Extended kalman filter based nonlinear model predictive control. *Industrial & Engineering Chemistry Research*, 33(6):1530–1541, 1994.

[26] Hind Taud and Jean-Franccois Mas. Multilayer perceptron (mlp). In *Geomatic approaches for modeling land change scenarios*, pages 451–455. Springer, 2017.