Human-Robot Collaborative Manipulation via Switchable Teleoperation and Autonomous Grasping

Jia Guo[1], Jiacheng Li[2], Kenta Urano[3], Takuro Yonezawa[3], , Nobuo Kawaguchi[1]
[1]Inst. of Innov. for Future Society, Nagoya University, Nagoya, Japan
[2]Department of Applied Mathematics and Systems, Faculty of Informatics, Kanagawa University, Yokohama, Japan
[3]Grad. Sch. of Eng., Nagoya University, Nagoya, Japan

Abstract: Teleportation of robotic arms over long distances places a significant cognitive and physical burden on human operators, particularly during the precision-demanding final phase of grasping. This work proposes a human-robot collaborative control framework that integrates large-range teleportation with proximity-triggered autonomous grasping. In the proposed system, the operator maintains full control during coarse positioning, while an automatic mode transition is activated when the end-effector reaches a predefined distance threshold from the target object. This work employs a distance-aware switching mechanism to seamlessly hand over control from the human operator to an autonomous grasping module, enabling efficient and reliable object acquisition without requiring continuous high-precision input from the operator. Experimental results demonstrate that the proposed approach significantly reduces operator workload while improving grasping success rate and overall task efficiency. This work provides a practical and scalable solution for human-robot collaboration in remote manipulation scenarios.

Introduction

Remote teleoperation of robotic manipulators has emerged as a critical technology for executing tasks in hazardous, inaccessible, or distant environments, including nuclear decommissioning, deep-sea exploration, surgical intervention, and space operations. However, pure teleoperation imposes significant cognitive and physical burdens on human operators, particularly during precision manipulation tasks that demand sustained attention and fine motor control over extended periods. Shared autonomy, which strategically combines human intelligence with robot automation, offers a promising paradigm to address these limitations. In this framework, the human operator provides high-level guidance and spatial reasoning—capabilities at which humans excel—while the autonomous system handles low-level, repetitive, and precision-demanding subtasks such as final approach and grasping. This division of labor leverages the complementary strengths of humans and machines: operators can focus on global navigation, obstacle avoidance, and task-level decision-making without being burdened by the fine-grained control required for stable object acquisition. Recent studies have demonstrated that such shared control architectures can significantly reduce task completion time, decrease operator workload, and improve overall manipulation success rates compared to both pure teleoperation and full autonomy approaches.

Despite its potential, implementing effective shared autonomy for remote manipulation presents several fundamental challenges. First, the design of an appropriate control handover mechanism remains an open problem: determining when, where, and how to transition control authority from the human to the autonomous system requires careful consideration of task context, environmental uncertainty, and operator intent. A premature handover may result in autonomous grasping attempts when the end-effector is poorly positioned, while delayed transitions negate the benefits of automation. Second, robust autonomous grasping in unstructured environments continues to pose significant difficulties, as the system must reliably estimate object pose, generate collision-free grasp configurations, and adapt to variations in object geometry, surface properties, and scene clutter—all in real-time and often with limited or noisy sensory information. Third, communication latency inherent in remote operation scenarios introduces additional complexity, as time delays can destabilize the control loop and create mismatches between operator expectations and robot behavior during the transition phase.

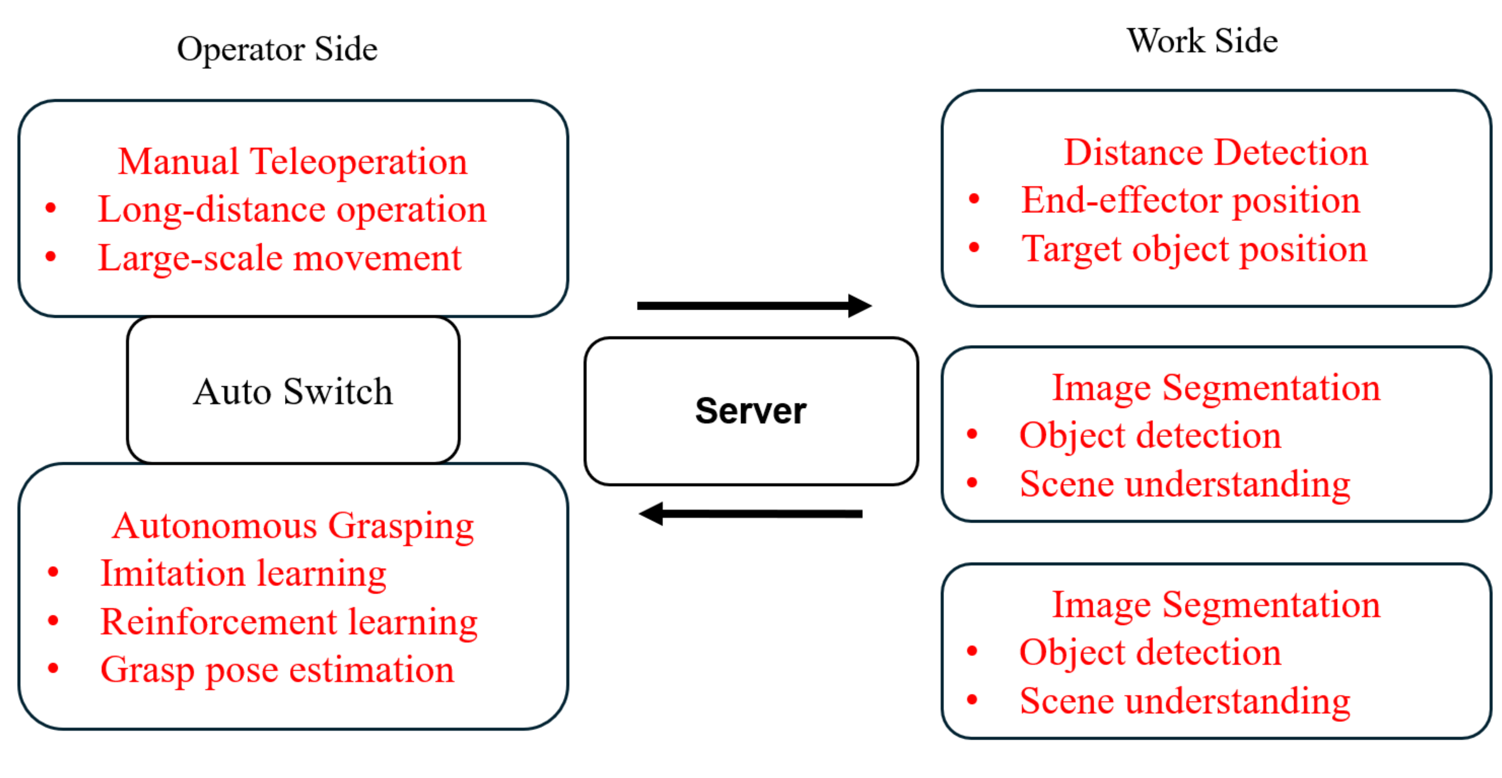

Fig. 1 Overview of the proposed shared autonomy framework

Methods:

System Overview:

The proposed system implements a shared autonomy framework for remote robotic manipulation, combining human teleoperation with autonomous grasping capabilities. As illustrated in Fig. 1, the system architecture consists of three main components: the operator side, the server, and the work side. The operator side handles manual teleoperation for large-scale movements and executes autonomous grasping after mode switching. The work side performs real-time perception tasks including distance detection and image segmentation, while providing visual feedback to the operator. The server acts as a communication bridge, relaying control commands and video streams between the two sides. Details are shown in Fig. 1.

Manual Teleoperation Mode:

In manual teleoperation mode, the operator controls the robot arm remotely using a VR controller. This mode is designed for long-distance operations and large-scale movements, where human spatial reasoning and decision-making are essential. The operator observes the remote workspace through real-time video streaming and directly commands the end-effector position and orientation. Control commands are transmitted through the server via low-latency communication protocols.

Proximity-based Mode Switching:

The work side continuously monitors the distance between the end-effector and the target object. When this distance falls below a predefined threshold, the system automatically switches from manual teleoperation to autonomous grasping. This proximity-based switching mechanism ensures that the operator handles coarse positioning, while the autonomous system takes over for precision grasping where consistent execution is required.

Perception Module:

The work side employs deep learning-based image segmentation to identify and localize target objects. An RGB-D camera mounted on the robot provides color and depth information for 3D position estimation. The segmentation module generates object masks to compute target positions, which serve both distance measurement for mode switching and target pose generation for autonomous grasping.

Autonomous Grasping:

Once switched to autonomous mode, the grasping module controls the robot arm through learning-based approaches. The grasping policy is trained via imitation learning from human demonstrations collected during teleoperation. Given the segmented point cloud, the system estimates optimal grasp poses using neural network-based detection methods, then executes collision-free trajectories with appropriate velocity and force control.

Conclusion: This paper presented a shared autonomy framework for remote robotic manipulation that combines manual teleoperation with proximity-triggered autonomous grasping. The proposed system allows operators to perform large-scale movements through direct control, while automatically switching to autonomous mode when the end-effector approaches the target object. This division of labor leverages human capabilities for global navigation and decision-making, while utilizing machine precision for the grasping task, thereby reducing operator workload and improving manipulation reliability. Future work will focus on optimizing the mode switching threshold through adaptive learning methods, improving grasp success rates under varying object geometries and environmental conditions, and conducting comprehensive user studies to evaluate system per-

Acknowledgements
This work is partially supported by JSPS KAKENHI (JP22K18422), NEDO (JPNP23003), NICT (222C01, 22609) and CSTI SIP3 (JPJ012495).
The research was conducted under the auspices of the Hosei International Fund (HIF) Foreign Scholars Fellowship.

References
[1] Shu, Y., Xiong, C., & Chen, C. (2024). A Study on Robotic Arm Target Recognition and Grasping Method Based on Deep Learning. International Journal of Pattern Recognition and Artificial Intelligence, 38(05).
[2] Xu, X., Li, Y., & Lei, G. (2025). Robotic Arm Grasping of Invisible Objects in Cluttered Environments. Transactions of FAMENA, 49(3), 73–94.