REMOTE 6D OBJECT POSE ESTIMATION FOR AGILE GRASPING: LEVERAGING CLOUD COMPUTING THROUGH WIRELESS COMMUNICATION

Oleksii Zamozhskyi[1], Alexander Werner[1], Marie Charbonneau[1,2], and William Melek[1]
[1] Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Canada.
[2] Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, Canada.
Email: ozamozhs@uwaterloo.ca

INTRODUCTION

The transition toward Industry 4.0 is driving demand for agile robotic systems capable of vision-guided manipulation [1]. Advanced perception methods, particularly those based on deep learning, enable 6D pose estimation but remain computationally demanding for onboard hardware [2]. Offloading perception to remote servers offers a promising alternative [3], yet introduces latency and reliability challenges [4] that threaten the strict timing requirements of closed-loop control.

This work investigates the feasibility of performing closed-loop robotic grasping reliably when 6D object pose estimation is offloaded over a wireless network.

.MATERIALS AND METHODS

A robotic grasping system was designed as shown in Fig 1. Images were streamed to a remote server with an NVIDIA RTX 4090 GPU running the FoundationPose model [5] for 6D object pose estimation. Estimated poses were returned to the robot to guide conveyor-based pick-and-place grasps. Grasp success rate, latency, and reliability were evaluated on four wireless links (Wi-Fi 60 GHz, Wi-Fi 5 GHz, 5G NSA 24 GHz, and 5G NSA 3.5 GHz), with 50 trials conducted for each network.

RESULTS AND DISCUSSION

The results show that a visual servoing loop can be closed over a wireless network using remote perception, provided that latency and variability remain within acceptable bounds (Table 1). System performance was strongly tied to the wireless link: higher latency reduced the maximum feasible conveyor speed, while greater stability consistently improved control reliability.

The experiments also offer practical insights into the wireless technologies themselves (Table 1). Wi-Fi 60 GHz delivered the best balance of low latency and stable throughput. Wi-Fi 5 GHz achieved low latency but suffered from high variability, limiting consistency. 5G NSA 24 GHz offered moderate latency with excellent stability, indicating strong potential for deployment. 5G NSA 3.5 GHz, though stable, exhibited high latency, making it less suitable for time-critical tasks.

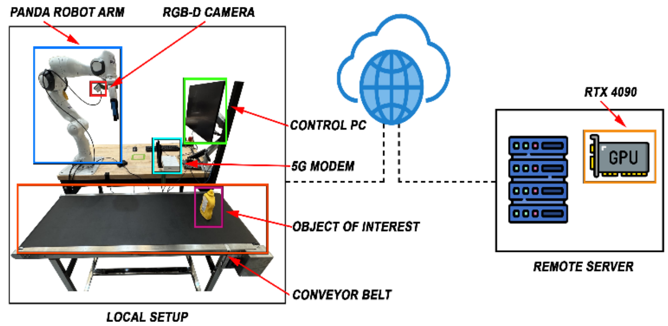

Fig 1 System components: robot arm, RGBD camera, control PC, 5G modem, conveyor, target object, and remote inference server.

CONCLUSIONS

This study shows that closed-loop robotic grasping is feasible when perception is offloaded to a remote server over a wireless network. Performance was shaped by communication latency and stability, with lower and steadier values enabling more reliable operation.

These results highlight the need for further work on improving wireless communication, strengthening local control, and advancing faster, more adaptable perception models. By addressing these challenges, cloud-augmented robotic systems can move from feasibility studies toward practical deployment, where they are expected to play a critical role in future industrial automation.

REFERENCES

[1] Fourmy M et al. arXiv:2311.05344, 2023.
[2] Huang Y et al. arXiv:2505.20814, 2025.
[3] Zahidi U et al. J Field Robot 41(8): 2771–89, 2024.
[4] Lee K et al. Reliable high-performance wireless systems for factory automation, 2020.
[5] Wen B et al. arXiv:2312.08344, 2023.

Table 1: Grasp success and failure counts, with average and standard deviation of RTT and server processing time for all four networks.

	Success	Failures	Mean Round-Trip Time [ms]	Mean Server Processing Time [ms]
Wi-Fi at 60 GHz	47	3	41.16 ± 3.46	21.45 ± 1.65
Wi-Fi 5 at 5 GHz	44	6	44.05 ± 32.96	20.04 ± 1.70
5G NSA at 24 GHz	43	7	58.26 ± 6.95	20.23 ± 1.77
5G NSA at 3.5 GHz	23	27	88.51 ± 14.74	20.36 ± 1.80