

Executive Summary

- This study evaluated how monoscopic monitor, immersive stereoscopic VR, and co-located walk-with-robot affect operator performance, workload, and situation awareness during mobile robot teleoperation and LiDAR mapping.
- Participants supervised an OMRON LD-90 robot through an indoor corridor, identifying visual targets while the robot performed LiDAR-based mapping.
- Stereoscopic and co-located viewing led to faster completion compared to monoscopic viewing, with more continuous movement and less need for camera adjustments.
- Workload (NASA-TLX) was comparable across modes but most consistent in the co-located condition, while stereoscopic viewing showed the highest engagement and active scanning behavior.
- Findings suggest that interface design for teleoperation should balance perceptual richness with cognitive demand, rather than assuming more immersion always improves performance.

Background

- Traditional teleoperations with 2D displays has been the standard for remote operations, but it often fails to convey spatial depth accurately. [1][2]
- This limitation increases cognitive load and slows spatial judgements, particularly in tasks requiring obstacle avoidance and target identification.
- Stereoscopic displays and VR headsets have been shown to improve navigation accuracy and reduce task time by restoring depth cues. [6][8]
- A limited field of view further burdens operators in standard setups. 360° camera systems combined with VR headsets have shown improvements in collision-related performance. [7]
- However, Head-Mounted Displays (HMDs) can increase discomfort and alter scanning strategies. [4][10]

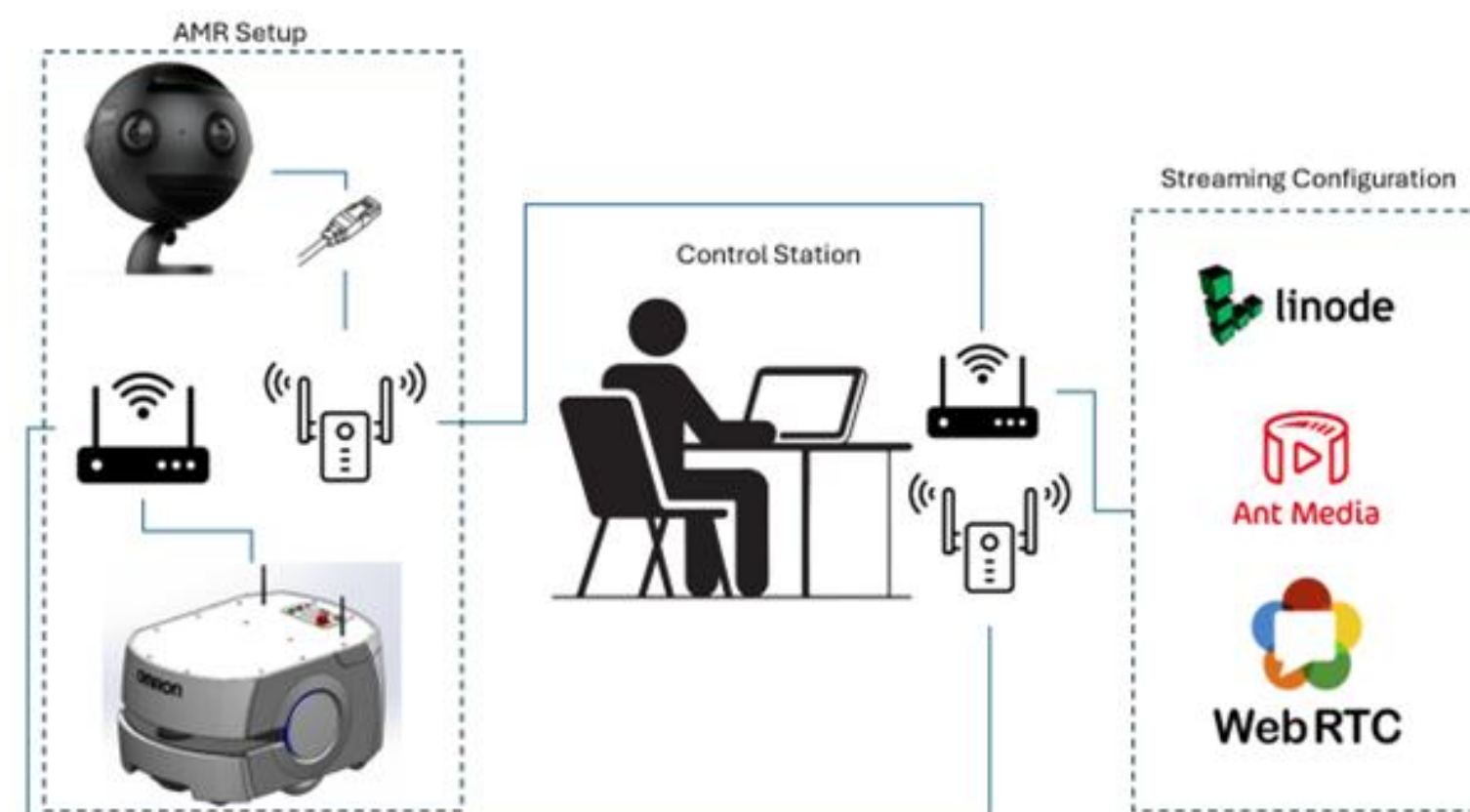


Figure 1: System architecture for supervised mapping and teleoperation. The OMRON LD-90 autonomous mobile robot is equipped with a 360-degree camera and connects to two separate Wi-Fi networks.

Experimental Setup & Test Plan

- Robot Platform:** OMRON LD-90 autonomous mobile rover with onboard LiDAR; mapping performed using the vendor software stack without modification.
- Camera system:** Insta360 Pro 2 mounted on a custom aluminum frame at ~1.2 m (seated eye height), providing a stitched equirectangular 360° view.
- Streaming pipeline:** Low-latency RTMP-to-WebRTC architecture via a cloud-hosted Ant Media Server; end-to-end video delay ~0.2 s under typical settings.
- Dual-network design:** Video streaming and robot control traffic separated across two Wi-Fi networks to prevent bandwidth interference.
- Three operator viewpoints tested:**
 - Mode A:** 2D monitor – monoscopic feed, keyboard control
 - Mode B:** VR headset (Meta Quest 3) – immersive equirectangular 360°, keyboard control.
 - Mode C:** Co-located – natural vision, pendant joystick controller (baseline)
- Robot speed capped at 1 m/s; a supervisor was present during all sessions.

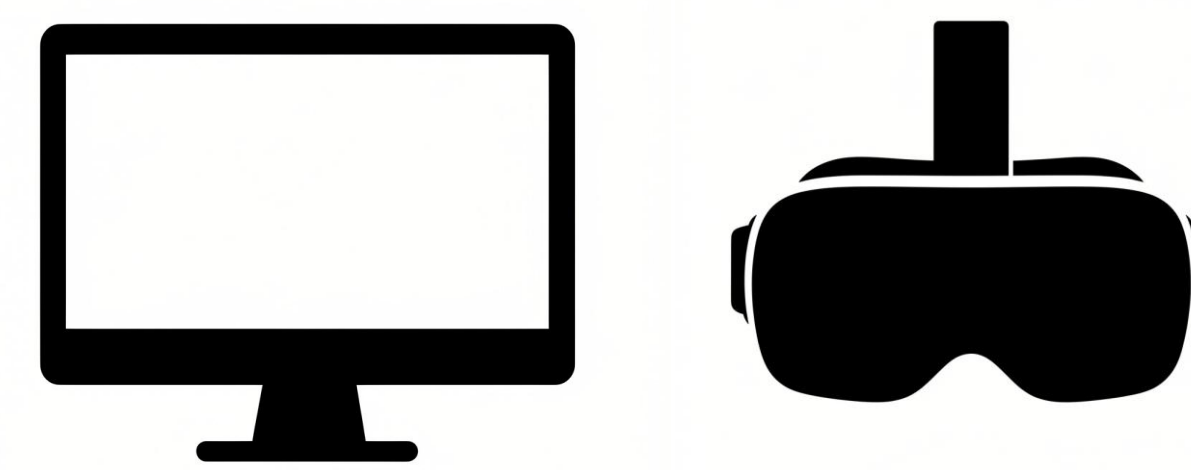


Figure 2: Three modalities used for mapping: co-located walk-with-robot, monoscopic 2D monitor, and stereoscopic VR headset. Each participant was tasked to identify color-and-shape visual targets along the way, and return to the starting position, while the robot simultaneously built a LiDAR map of the environment.

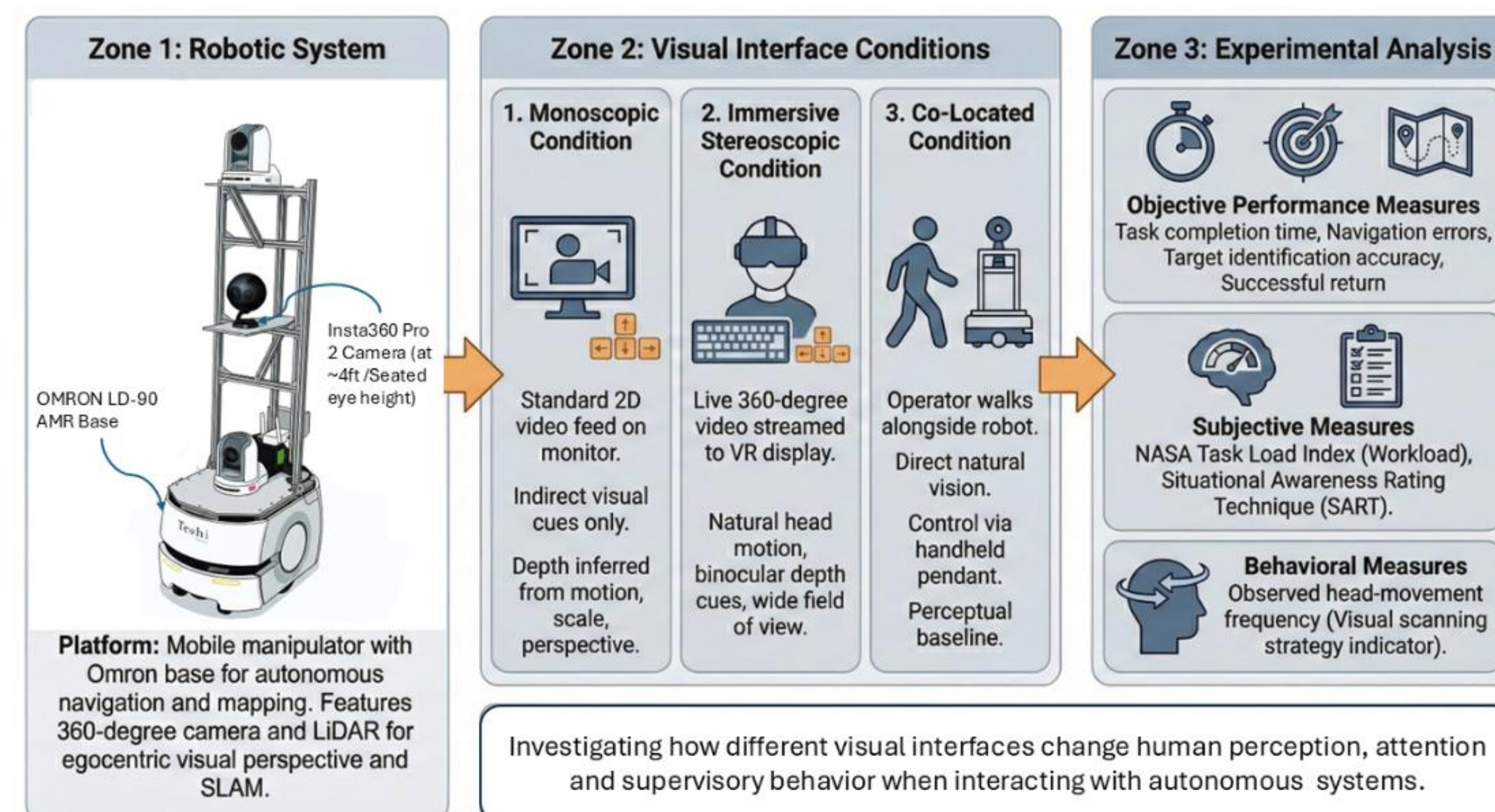


Figure 3: Experimental overview showing the robotic platform, three viewing conditions (monoscopic, immersive stereoscopic, and co-located), and the performance, workload, and behavioral measures used to evaluate how visual interface modality affects human supervision during mobile robot operation.

Experimental Results

- Task Completion Time:** Monoscopic viewing had the longest median task time (~300 s), while stereoscopic (~195 s) and walk-with-robot (~175 s) were significantly faster, nearly a 40% reduction, with co-located having the smallest spread.
- Workload (NASA-TLX):** Perceived workload was moderate across all three modes (medians: Monoscopic ~29, Stereoscopic ~22, walk-with-robot ~20), with stereoscopic and co-located conditions trending lower.
- Situation Awareness (SART):** Total SART scores were comparable across all display modes; monoscopic viewing showed slightly higher attentional demand than others.

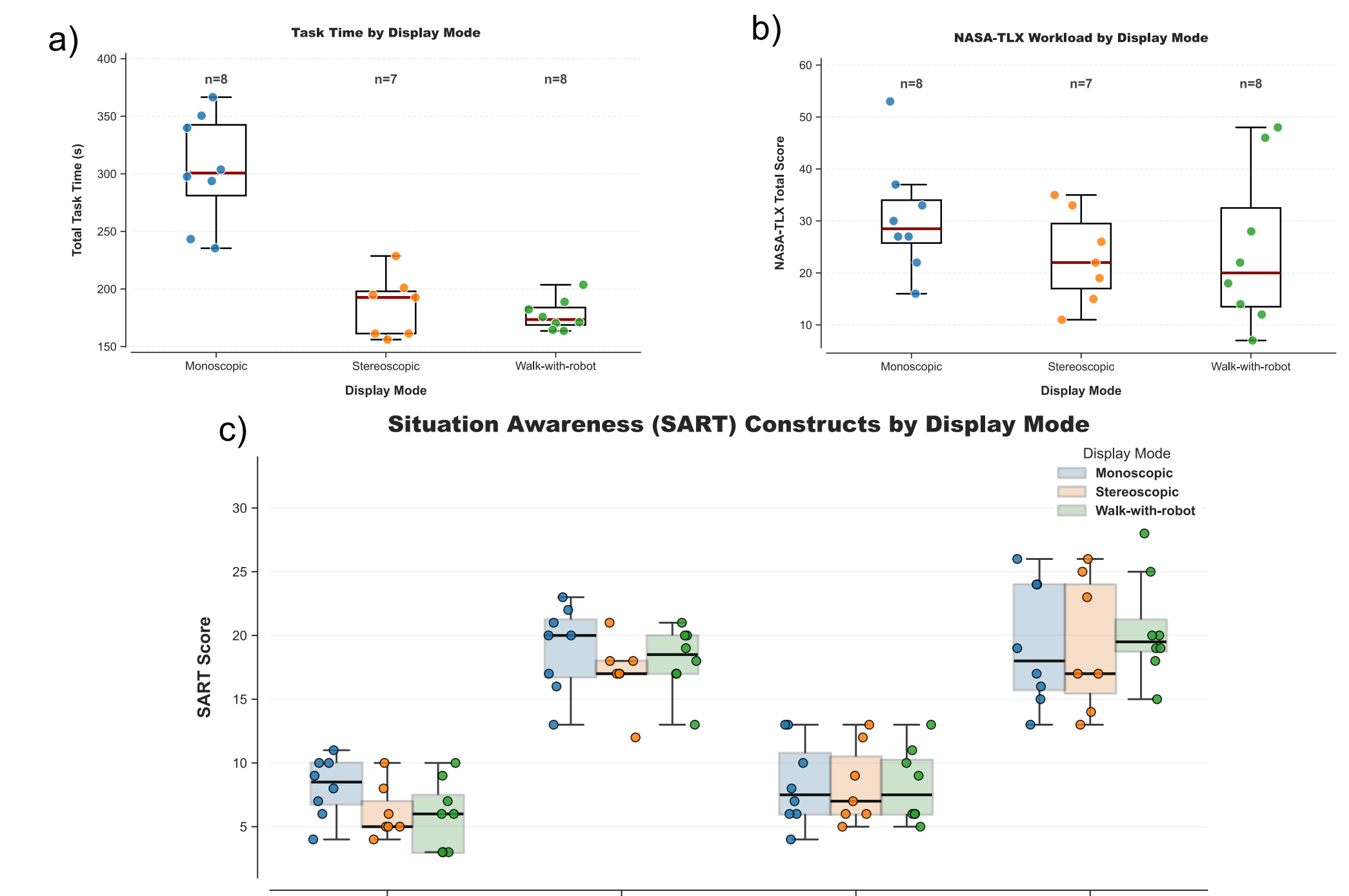


Figure 4: a) Task Completion time by display mode ($n = 8, 7, 8$). b) NASA-TLX workload scores by display mode. c) Situation Awareness (SART) constructs by display mode, showing Demand (D), Supply (S), Understanding (U), and Total SART $[U - (D - S)]$.

Conclusions

- Changing visual feedback modality alters how operators supervise mobile robots, even when the robot, task, and environment remain constant.
- Monoscopic displays were associated with slower performance and more compensatory behavior, suggesting greater effort to reconstruct spatial relationships from limited depth cues.
- Immersive HMDs enabled more natural, continuous exploration but increased mental effort and potential discomfort for some users.
- Co-located viewing provided the most consistent workload and serves as a valuable perceptual baseline for comparing mediated interfaces.
- Interface design for teleoperation should balance perceptual richness with cognitive demand. More immersion does not uniformly improve performance.

References

- D. Drascic, "Skill Acquisition and Task Performance in Teleoperation Using Monoscopic and Stereoscopic Video Remote Viewing," *Proceedings of the Human Factors Society Annual Meeting*, vol. 35, no. 19, pp. 1367–1371, Sep. 1991, doi: 10.1177/154193129103501906.
- I. Iehisa, M. Ayaki, K. Tsubota, and K. Negishi, "Factors affecting depth perception and comparison of depth perception measured by the three-rod test in monocular and binocular vision," *Heliyon*, vol. 6, no. 9, Sep. 2020, doi: 10.1016/j.heliyon.2020.e04904.
- M. Martinez Escobar, B. Junke, J. Holub, K. Hisey, D. Elliot, and E. Winer, "Evaluation of monoscopic and stereoscopic displays for visual-spatial tasks in medical contexts," *Computers in Biology and Medicine*, vol. 61, pp. 138–143, Jun. 2015, doi: 10.1016/j.combiomed.2015.03.026.
- S. Livatino and F. Privitera, "3D visualization technologies for teleoperated robots," in *Proceedings of the ACM symposium on Virtual reality software and technology*, in VRST '06, New York, NY, USA: Association for Computing Machinery, Nov. 2006, pp. 240–243, doi: 10.1145/1180495.1180544.
- V. Nirmal, O. Ngwu, E. P. Sridhar, M. D. A. Habib, and M. Rahman, "Depth Perception and Task Performance in Robotic Teleoperation," *IJSE Annual Conference. Proceedings*, pp. 1–6, 2024, doi: 10.21872/2024IJSE.6954.
- J. Y. Chen, R. V. Oden, C. Kenny, and J. O. Merritt, "Effectiveness of Stereoscopic Displays for Indirect-Vision Driving and Robot Teleoperation," Art. no. ARLTR5259, Accessed: Jan. 24, 2026. [Online]. Available: <https://apps.dtic.mil/sti/html/ADA526325/>
- S. Wibowo, I. Siradjuddin, F. Ronilaya, and M. N. Hidayat, "Improving teleoperation robots performance by eliminating view limit using 360 camera and enhancing the immersive experience utilizing VR headset," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1073, no. 1, p. 012037, Feb. 2021, doi: 10.1088/1757-899X/1073/1/012037.
- Y. Luo, J. Wang, H.-N. Liang, S. Luo, and E. G. Lim, "Monoscopic vs. Stereoscopic Views and Display Types in the Teleoperation of Unmanned Ground Vehicles for Object Avoidance," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Aug. 2021, pp. 418–425, doi: 10.1109/RO-MAN50785.2021.9515455.
- W. Fung, W. Lo, Y. Liu, and N. Xi, "A case study of 3D stereoscopic vs. 2D monoscopic tele-reality in real-time dexterous teleoperation," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug. 2005, pp. 181–186, doi: 10.1109/IROS.2005.1545299.
- L. Meng, J. Liu, W. Chai, J. Wang, and M. Q.-H. Meng, "Virtual Reality Based Robot Teleoperation via Human-Scene Interaction," Aug. 02, 2023, *arXiv: arXiv:2308.01164*, doi: 10.48550/arXiv.2308.01164.