Real-Time Object Detection and Augmented Reality to Support Low-Vision Navigation and Object Localization: A Demonstration

Yong-Joon Thoo
Human-IST Institute, Department of Computer Science
University of Fribourg
Fribourg, Switzerland
yongjoon.thoo@unifr.ch

Karim Aebischer
Human-IST Institute, Department of Computer Science
University of Fribourg
Fribourg, Switzerland
karim.aebischer@unifr.ch

Nicolas Ruffieux
Department of Special Education
University of Fribourg
Fribourg, Switzerland
nicolas.ruffieux@unifr.ch

Denis Lalanne
Human-IST Institute, Department of Computer Science
University of Fribourg
Fribourg, Switzerland
denis.lalanne@unifr.ch

Figure 1: First-person PoV footage taken from the HoloLens 2 headset with visual icons anchored to the detected objects in the environment.

Abstract

Low-vision (LV) people often face difficulties in navigating complex environments and recognizing objects due to reduced visual acuity, contrast sensitivity, or field of view. In this paper, we present a prototype that combines real-time object detection with augmented reality (AR) visualization to enhance spatial awareness for LV users, supporting safe navigation and object localization in indoor spaces. The system integrates an RGB-D camera and a Microsoft HoloLens 2 headset via ROS and Unity, using a YOLOv11-based perception pipeline to detect and localize static and dynamic objects in 3D and display high-contrast AR icons above the detected objects within the user's field of view, conveying both their position and identity.

CCS Concepts

• **Human-centered computing** → *Accessibility systems and tools*; *Accessibility technologies*; • **Social and professional topics** → **People with disabilities**.

AlpCHI '26, Ascona, Ticino, Switzerland

Keywords

low-vision, augmented reality, accessibility, navigation assistance, object detection

ACM Reference Format:
Yong-Joon Thoo, Karim Aebischer, Nicolas Ruffieux, and Denis Lalanne. 2026. Real-Time Object Detection and Augmented Reality to Support Low-Vision Navigation and Object Localization: A Demonstration. In *Proceedings of Human-Computer Interaction Conference of the Alpine Region (AlpCHI '26).* ACM, New York, NY, USA, 4 pages.

1 Introduction

Low vision (LV) can result from mild to severe eye conditions and can affect various aspects of visual perception (*e.g.,* visual acuity, contrast sensitivity, field of view) making it difficult to perceive or identify objects and/or navigate in complex environments [8, 10]. However, most LV individuals may prefer to rely on their residual vision [8, 20, 24, 25]. In line with this, augmented reality (AR) has emerged as a promising approach to visually enhance environmental perception by overlaying digital cues directly within the user's field of view [4].

For instance, AR has been explored as a navigation aid to enhance spatial awareness for LV users by visually augmenting their surroundings. One line of work has focused on improving the visibility of environmental objects and obstacles. Htike *et al.* [8] evaluated eight visual enhancement techniques—such as color or edge overlays and bounding boxes—on a HoloLens v1, emphasizing the need for customizable visual augmentations. Similarly, Zhang *et al.* [23] proposed *Trans4Trans*, a transformer-based segmentation model integrated into a wearable system to detect transparent obstacles (*e.g.*, glass doors, windows). Other studies have investigated AR-based navigation and spatial guidance, such as Zhao *et al.* [24], who explored projection- and smartglasses-based visualizations for stair navigation, and Zhao *et al.* [25], who compared audio and visual feedback for wayfinding tasks. More recently, Chen *et al.* [1] introduced *VisiMark*, an AR interface that augments predefined navigation landmarks (*e.g.*, stairs, elevators) with world-anchored visual cues and "signboards" at hallway intersections to depict spatial layout and upcoming landmarks.

Beyond navigation, AR and computer vision have also been used to support object recognition and interaction. For instance, Zhao *et al.* [26] presented *CueSee*, an AR application that recognizes products and employs directional visual cues to guide users toward target items during visual search tasks. Lee *et al.* [2] developed *ARTennis*, a wearable AR prototype enhancing the visual saliency of tennis balls, and later introduced *CookAR* [3], which highlights the affordances of kitchen tools (*e.g.*, knives, pans) using real-time segmentation to support safe and efficient interactions.

Building on prior work—and insights from our user understanding studies involving LV individuals [18, 19]—we present a prototype system (cf. Figure 1) that advances AR-based assistance for LV users through real-time detection and visualization of static and dynamic objects in indoor environments (*e.g.,* rooms, open spaces). The system uses an RGB-D camera and real-time object recognition to estimate each object's distance, position, and label, which are transmitted to a HoloLens 2 headset for in-situ AR visualization. Unlike approaches that rely on predefined landmarks or pre-mapped spaces, our system autonomously identifies environmental elements and displays them as visual icons that convey both their location and nature. This design aims to support spatial awareness during navigation while enabling users to localize and recognize objects that may afford interaction (*e.g.,* a chair, sofa, person, bag).

Although the prototype targets assistive indoor navigation and object localization, the underlying real-time detection and AR visualization techniques may also complement emerging AR-based training systems for low-vision training, including therapist-guided approaches explored in our prior work [18, 19].

A short video demonstration illustrating the system and example use cases is available at: https://visar-alpchi26.human-ist.ch/.

2 System Implementation

The prototype integrates real-time 3D perception with AR visualization to support spatial awareness for LV users. It combines an Intel RealSense D455 RGB-D camera [12] with a custom ROS [11, 14]–based perception pipeline using the Ultralytics YOLOv11 model [21], and a Unity [22] application running on a Microsoft HoloLens 2 [5] headset with the MRTK3 toolkit [6]. The system was deployed on a Lenovo Legion T5 desktop (AMD Ryzen 7 7700 CPU, 32 GB RAM, NVIDIA GeForce RTX 4070 Super, 1 TB SSD) running Ubuntu 20.04 with ROS Noetic [15]. Figure 2 illustrates the overall flow from sensing and detection to AR visualization.

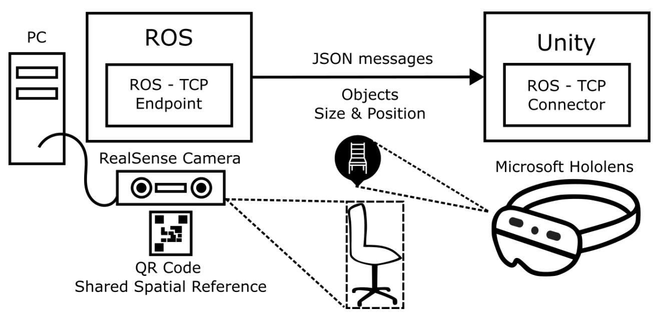
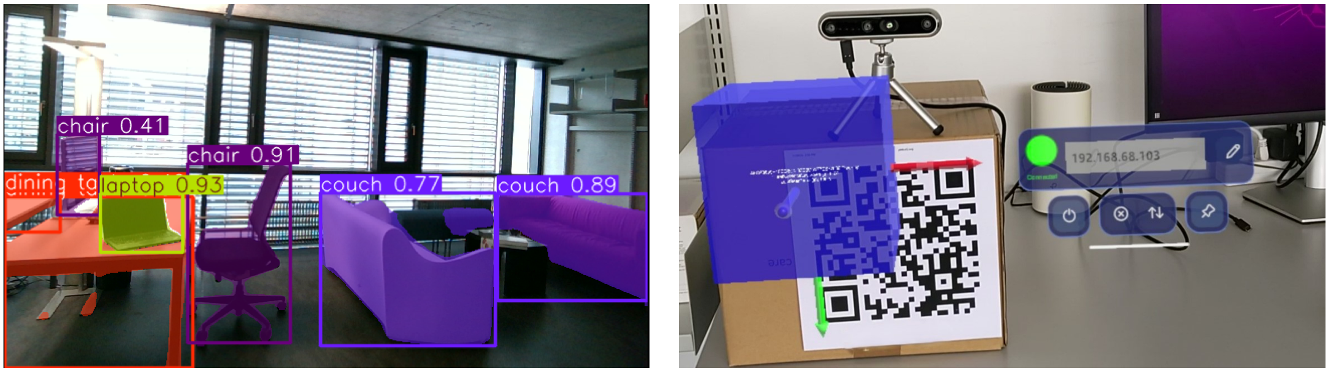


Figure 2: Overview of the sensing, detection, and AR visualization pipeline using ROS, a RealSense camera, and the HoloLens 2 with QR-based spatial alignment.

2.1 Perception and Depth Estimation

We implemented a ROS node that fuses RGB and depth data from the *realsense-ros* wrapper [13] to detect, segment, and localize objects in 3D. The node dynamically retrieves the camera intrinsics (f_x, f_y, c_x, c_y) and processes synchronized RGB–D frames through the YOLOv11 segmentation model (cf. Figure 3a). For each detection above a given confidence threshold, the node extracts depth values from the segmented region, computes the average object distance Z, and estimates its position relative to the camera using the pinhole model:

$$(X, Y, Z) = (\frac{(u - c_x)Z}{f_x}, \frac{(v - c_y)Z}{f_y}, Z) \qquad (1)$$



(a) (b)

Figure 3: Illustration of: (a) objects detected and segmented by the YOLO-v11 model; and (b) the AR interface for registration and TCP communication between the headset and the ROS node.

where (u, v) denotes the centroid of the segmentation mask. Estimated object positions, sizes, and confidence scores are formatted as JSON messages and published to the corresponding topic at 2 Hz. Annotated segmentation images are concurrently streamed for debugging or visualization in RViz.

2.2 Communication, Registration, and AR Visualization

To enable real-time communication between ROS and Unity, we used the *ROS–TCP Endpoint* [17] (ROS side) and *ROS–TCP Connector* [16] (Unity side) packages, where the Endpoint creates a server for Unity to connect to, and the Connector manages the serialization and exchange of ROS messages directly within the Unity scene. Through this link, the Unity application running on the HoloLens 2 subscribes to the detection topic and receives JSON messages describing the detected objects and their 3D positions.

To enable registration between the RealSense camera and the headset—and thereby ensure the accurate positioning of virtual objects from the headset's perspective—we used a printed QR-code marker placed at a fixed position next to the RealSense camera with a predefined spatial offset (cf. Figure 3b). This marker is detected by the HoloLens 2 using the *Microsoft.MixedReality.QR* SDK [7, 9], which estimates the headset's pose relative to the QR code after establishing the connection to the ROS endpoint, thereby creating a shared spatial reference. This alignment ensures consistent registration between the camera and headset referentials, allowing virtual content to be accurately anchored within the physical environment. The Unity application then projects high-contrast virtual icons above the detected objects (*e.g.*, chairs, tables, or bags), using the information contained within the JSON message, assisting users in perceiving their nature and location within the real environment through the AR headset (cf. Figure 1).

3 Future Work and Conclusion

This paper presented a prototype system that integrates real-time object detection with AR visualization to support spatial awareness for LV users by conveying both the position and identity of nearby objects and/or obstacles. While the current setup uses a single RGB-D camera in a room-scale environment, in future work we aim to extend the architecture to multi-camera configurations to increase spatial coverage, improve robustness against occlusions, and support multi-room deployments. Future iterations will also explore task-oriented interactions—such as querying specific object types through speech or other modalities—and adaptive visualization techniques to prevent visual overload. For instance, as highly cluttered environments may increase cognitive load due to excessive visual cues, we plan to explore selectively displaying icons or objects within a defined distance or relevance threshold from the user. Future work will also include user evaluations in real-world scenarios to assess usability, perceived usefulness, and integration into indoor environments.

References

[1] Ruijia Chen, Junru Jiang, Pragati Maheshwary, Brianna R Cochran, and Yuhang Zhao. 2025. VisiMark: Characterizing and Augmenting Landmarks for People with Low Vision in Augmented Reality to Support Indoor Navigation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 298, 20 pages. doi:10.1145/3706598.3713847

[2] Jaewook Lee, Devesh P. Sarda, Eujean Lee, Amy Lee, Jun Wang, Adrian Rodriguez, and Jon E. Froehlich. 2023. Towards Real-time Computer Vision and Augmented Reality to Support Low Vision Sports: A Demonstration of ARTennis. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 81, 3 pages. doi:10.1145/3586182.3615815

[3] Jaewook Lee, Andrew D. Tjahjadi, Jiho Kim, Junpu Yu, Minji Park, Jiawen Zhang, Jon E. Froehlich, Yapeng Tian, and Yuhang Zhao. 2024. CookAR: Affordance Augmentations in Wearable AR to Support Kitchen Tool Interactions for People with Low Vision. *UIST 2024 - Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (10 2024). doi:10.1145/3654777.3676449

[4] Yifan Li, Kangsoo Kim, Austin Erickson, Nahal Norouzi, Jonathan Jules, Gerd Bruder, and Gregory F. Welch. 2022. A Scoping Review of Assistance and Therapy with Head-Mounted Displays for People Who Are Visually Impaired. *ACM Trans. Access. Comput.* 15, 3, Article 25 (aug 2022), 28 pages. doi:10.1145/3522693

[5] Microsoft. 2025. Microsoft HoloLens. https://learn.microsoft.com/en-us/hololens/?WT.mc_id=mixedreality_product. Accessed: 20 October 2025.

[6] Microsoft. 2025. Mixed Reality Toolkit 3. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/. Accessed: 25 September 2025.

[7] Microsoft. 2025. QR code tracking overview. https://learn.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/qr-code-tracking-overview. Accessed: 30 September 2025.

[8] Hein Min Htike, Tom H. Margrain, Yu-Kun Lai, and Parisa Eslambolchilar. 2021. Augmented Reality Glasses as an Orientation and Mobility Aid for People with

Low Vision: a Feasibility Study of Experiences and Requirements. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 729, 15 pages. doi:10.1145/3411764.3445327

[9] Nuget. 2025. Microsoft.MixedReality.QR NuGet Package. https://www.nuget.org/packages/Microsoft.MixedReality.QR. Accessed: 30 September 2025.

[10] World Health Organization. 2019. World report on vision. https://www.who.int/publications/i/item/world-report-on-vision.

[11] Morgan Quigley, Josh Faust, Tully Foote, Jeremy Leibs, et al. [n. d.]. ROS: an open-source Robot Operating System.

[12] RealSense. 2025. RealSense Depth Camera D455. https://realsenseai.com/stereo-depth-cameras/real-sense-depth-camera-d455/. Accessed: 20 October 2025.

[13] Intel RealSense. 2025. ROS Wrapper for Intel RealSense Devices. https://github.com/IntelRealSense/realsense-ros/tree/ros1-legacy. Accessed: 20 October 2025.

[14] Open Robotics. 2025. ROS - Robot Operating System. https://www.ros.org/. Accessed: 20 October 2025.

[15] Open Robotics. 2025. ROS Noetic Ninjemys. https://wiki.ros.org/noetic. Accessed: 20 October 2025.

[16] Unity Technologies. 2025. ROS TCP Connector. https://github.com/Unity-Technologies/ROS-TCP-Connector. Accessed: 20 October 2025.

[17] Unity Technologies. 2025. ROS TCP Endpoint. https://github.com/Unity-Technologies/ROS-TCP-Endpoint. Accessed: 20 October 2025.

[18] Yong-Joon Thoo, Karim Aebischer, Nicolas Ruffieux, and Denis Lalanne. 2025. Enhancing Therapist-Guided Low-Vision Training with Projected Gaze Behaviors in Co-Located Shared AR. In *Proceedings of the 2025 ACM Symposium on Spatial User Interaction (SUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 15, 15 pages. doi:10.1145/3694907.3765924

[19] Yong-Joon Thoo, Karim Aebischer, Nicolas Ruffieux, and Denis Lalanne. 2025. Exploring Shared Augmented Reality for Low-Vision Training of Activities of Daily Living. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*. Association for Computing Machinery, New York, NY, USA, Article 31, 19 pages. doi:10.1145/3663547.3746371

[20] Yong-Joon Thoo, Maximiliano Jeanneret Medina, Jon E. Froehlich, Nicolas Ruffieux, and Denis Lalanne. 2023. A Large-Scale Mixed-Methods Analysis of Blind and Low-vision Research in ACM and IEEE. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 20, 20 pages. doi:10.1145/3597638.3608412

[21] Ultralytics. 2025. Ultralytics YOLO11. https://docs.ultralytics.com/models/yolo11/. Accessed: 20 October 2025.

[22] Unity. 2025. Unity Game Engine. https://unity.com/. Accessed: 20 October 2025.

[23] Jiaming Zhang, Kailun Yang, Angela Constantinescu, Kunyu Peng, Karin Muller, and Rainer Stiefelhagen. 2022. Trans4Trans: Efficient Transformer for Transparent Object and Semantic Scene Segmentation in Real-World Navigation Assistance. *IEEE Transactions on Intelligent Transportation Systems* 23, 10 (2022), 19173 – 19186. doi:10.1109/TITS.2022.3161141 46.24.01; LK 01.

[24] Yuhang Zhao, Elizabeth Kupferstein, Brenda Veronica Castro, Steven Feiner, and Shiri Azenkot. 2019. Designing AR Visualizations to Facilitate Stair Navigation for People with Low Vision. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 387–402. doi:10.1145/3332165.3347906

[25] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. 2020. The Effectiveness of Visual and Audio Wayfinding Guidance on Smartglasses for People with Low Vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376516

[26] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. 2016. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 73–84. doi:10.1145/2971648.2971730