POLITECNICO DI TORINO
Repository ISTITUZIONALE

A Comparison Between Two Assistive Indoor Navigation Applications

Original
A Comparison Between Two Assistive Indoor Navigation Applications / Manuri, F., Perseo, M., Giugliano, S., Sanna, A.. - (2026), pp. 7-12. (IEEE Smart World Congress 2026 Rende (ITA) 7-11 Settembre 2026) [10.1109/SWC69865.2026.11763769].

Availability:
This version is available at: 11583/3017712 since: 2026-11-09T16:36:22Z

Publisher:
IEEE

Published
DOI:10.1109/SWC69865.2026.11763769

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

21 November 2026

A Comparison Between Two Assistive Indoor Navigation Applications

979-8-3315-5021-9/26/$31.00 ©2026 IEEE

1st Federico Manuri
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
federico.manuri@polito.it

2nd Michele Perseo
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
michele.perseo@studenti.polito.it

3rd Salvatore Giugliano
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
salvatore.giugliano@studenti.polito.it

4th Andrea Sanna
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
andrea.sanna@polito.it

***Abstract*—Over 2 billion people live with vision impairment, and existing mobility aids, such as white canes, guide dogs, and GPS-based apps, all fail indoors or present severe limitations, whether due to limited environmental feedback, cost, or satellite signal attenuation. These constraints motivated a paradigm shift toward vision-based wearable systems, positioning egocentric vision and smart glasses as a promising technological direction. This research explores the capabilities of the Meta Project Aria Gen 1 platform for developing egocentric-vision-based assistive indoor navigation applications. Two prototypes were developed to evaluate two indoor navigation paradigms: the first relies on pre-mapped spatial data of the indoor environment to be navigated; the second interprets environmental cues and translates them into actionable navigation instructions, thereby reducing the dependency on prior knowledge of the environment. Preliminary results with two groups of ten users show that both systems enabled independent route completion, with comparable System Usability Scale scores (83 and 88.5).**

***Index Terms*—Indoor Navigation, Egocentric Vision, Assistive Technology, Human-Machine Interaction, Project Aria**

I. Introduction

Navigation within unfamiliar environments was identified by the World Health Organization as a substantial challenge for more than 2 billion individuals experiencing near or distance vision impairment [1]. Among this population, an estimated 43 million individuals were classified as completely blind and 300 million suffered from moderate or severe vision impairment in 2020, conditions that were found to significantly disrupt daily routines [2]. These rates are expected to increase, with up to 700 million individuals to experience blindness or moderate-to-severe vision impairment by 2050 [3].

Traditional mobility aids include white canes and guide dogs; however, guide dogs require considerable training and financial costs, while white canes offer limited environmental interpretation. In the early 2000s, the proliferation of GPS infrastructure and smartphones enabled the development of novel digital navigation aids. Applications such as BlindSquare and Microsoft's Seeing AI were introduced, providing GPS-based audio guidance and AI-driven environmental narration. Nevertheless, the accuracy of GPS-dependent systems was significantly compromised in indoor environments due to satellite signal attenuation. To mitigate this limitation, infrastructure-dependent alternatives, including Bluetooth Low Energy (BLE) beacons, RFID tags, and Wi-Fi fingerprinting, were investigated. Although operational effectiveness was demonstrated in controlled deployments, extensive installation requirements and maintenance overheads were identified, rendering large-scale implementation impractical. These constraints motivated a paradigm shift toward vision-based wearable systems, positioning egocentric vision and smart glasses as a promising technological direction.

The integration of artificial intelligence with wearable hardware has been leveraged to advance egocentric vision frameworks that use onboard cameras and sensors to reconstruct the user's perspective [4]. Early implementations, such as Google Glass, established foundational hardware paradigms, despite limited commercial adoption due to privacy and usability concerns [5]. Recent consumer-oriented devices, such as Ray-Ban Meta glasses, achieved broader market penetration through conventional aesthetic designs [6]. However, these systems were not originally engineered for assistive purposes; accessibility features, such as the "Be My Eyes" integration [1], relied on remote human assistance rather than autonomous environmental interpretation. Dedicated assistive smart glasses, including Envision Glasses [7] and OrCam MyEye [8], were developed to provide object detection, text reading, and facial recognition. While valuable insights regarding camera integration and vocal interfaces were derived from these systems, significant barriers were identified. High financial costs, subscription-based licensing models, and narrow task specialization were frequently observed, limiting widespread accessibility.

Overall, existing assistive smart glasses focus on object

979-8-3315-5021-9/26/$31.00 ©2026 IEEE

[1] https://www.bemyeyes.com/

and text recognition in infrastructure-free settings, but none provide users with actionable, real-time route guidance to a specific destination within an indoor space. This represents a specific and, to the best of the authors' knowledge, unaddressed gap at the intersection of egocentric vision, wearable AI, and indoor wayfinding for BLV users.

To address this gap, this work makes two contributions. First, it presents an evaluation of the Meta Project Aria [9] Gen 1 platform for the development of egocentric-vision-based assistive indoor navigation applications, to our knowledge, the first such investigation of this platform for this purpose. Second, two prototypal navigation applications were developed to directly compare two architecturally distinct indoor navigation paradigms: the first relies on pre-mapped spatial data of the indoor environment to be navigated (a spatial localization approach), while the second interprets environmental cues in real time and translates them into actionable navigation instructions, thereby reducing dependency on prior environmental knowledge (an environment understanding approach). This comparison provides preliminary evidence on the relative usability and practical trade-offs of localization-based versus perception-based navigation strategies, a comparison not previously reported for assistive indoor navigation.

The proposed work addresses the following research questions:

- **RQ1**: Do the two proposed paradigms allow users to effectively navigate indoor environments?
- **RQ2**: Do the users prefer one of the two navigation paradigms?

This work contributes to accessibility research by evaluating a novel AI-driven wearable platform for indoor navigation and by providing a preliminary comparative analysis of two distinct interaction paradigms, along with an examination of translational challenges associated with real-world deployment.

II. State of the Art

Outdoor navigation systems are usually constrained indoors by the attenuation of Global Navigation Satellite System (GNSS) signals within built environments, rendering traditional positioning methods ineffective, whereas infrastructure-dependent localization frameworks [10] present severe limitations such as continuous hardware maintenance and signal degradation [11], [12].

In response to these constraints, infrastructure-independent localization strategies were developed. Techniques such as Pedestrian Dead Reckoning (PDR), visual Simultaneous Localization and Mapping (SLAM), and sensor fusion were implemented to facilitate navigation within uninstrumented spaces [11], [12]. While deployment costs were eliminated and environmental adaptability was enhanced, distinct technical challenges were introduced. Cumulative drift errors in inertial-based methods were documented, progressively degrading positional accuracy over extended periods [12]. Additionally, vision-based SLAM frameworks were found to be highly sensitive to illumination variations, dynamic obstacles, and texture-poor surfaces, while imposing substantial computational overhead [13]. For assistive navigation applications, where continuous reliability was prioritized, temporary localization failures were considered significant usability risks [14]. Hybrid architectures combining opportunistic environmental cues with onboard sensing were subsequently explored to mitigate these drift-related limitations [12], [14]. Onboard sensors integrated into personal devices, including inertial measurement units (IMUs), cameras, and magnetometers, were leveraged to enable positioning without external hardware [14].

To address the trade-offs between infrastructure-dependent localization and pure sensor-based localization, egocentric vision frameworks have been increasingly investigated. Mapless, reactive navigation systems were initially developed to prioritize immediate safety over global mapping. RGB-D sensing was employed for real-time obstacle detection and depth expansion [15], while monocular depth processing was implemented on consumer smartphones to generate obstacle-free trajectories without prior environmental modeling [16]. Dedicated hardware combinations, including RGB-D tablets coupled with haptic feedback canes, were utilized to align visual odometry with pre-extracted architectural maps, enabling dynamic waypoint generation [17]. To bridge the gap between reactive and mapped approaches, wearable devices integrating real-time Semantic Visual SLAM were engineered. Environmental features were actively segmented and interpreted at an object level during navigation, eliminating reliance on static cartography [18]. More recently, a paradigm shift toward user-driven customization was observed. Vision and Large Language Models were integrated into do-it-yourself programming interfaces, empowering end-users to configure personalized semantic filters and interactive scene queries [19], [20]. Real-time object detection and scene understanding were demonstrated to provide richer navigational feedback than geometric mapping alone [17], [18]. Augmented reality interventions were also evaluated to enhance landmark perceptibility, thereby improving orientation and decision-making processes within unfamiliar indoor settings [18], [21], [22]. These vision-based approaches were recognized as a viable compromise, balancing scalability with the delivery of human-interpretable environmental cues essential for visually impaired navigation.

Concurrently, a critical misalignment between academic research priorities and the expressed requirements of blind and low-vision (BLV) populations was documented. An analysis of the existing literature revealed that BLV individuals were included in fewer than 20% of the investigated studies, and a weak correlation was observed between researcher-developed functionalities and user-perceived utility. Through structured interviews and systematic literature reviews, specific assistive requirements were cataloged [23]. Text recognition, scene querying, obstacle avoidance, and spatial navigation were consistently ranked as primary necessities, whereas previously emphasized research directions were frequently deemed secondary by end-users. Furthermore, pathology-specific variations in functional requirements were identified across distinct clinical groups, indicating that a generalized assistive frame-

work was insufficient [24]. Facilitation of social interaction, particularly facial recognition and conversational awareness, was highlighted as a significantly underserved domain. The necessity of adopting user-centered design methodologies was thereby established, wherein device development was strictly guided by empirically validated BLV priorities rather than technologically driven assumptions.

III. System Design and Implementation

The experimental applications were designed and developed to provide real-time, hands-free navigation assistance for visually impaired individuals within indoor environments. The design was constrained by hardware compatibility limitations, network bandwidth requirements, and user-centric accessibility standards. Consequently, a distributed architecture was implemented, decoupling data acquisition from computational processing.

A. Sensor Platform and Hardware Configuration

Data acquisition was performed using Meta Project Aria Gen 1 smart glasses, which integrate a synchronized multi-sensor array. The hardware configuration included a forward-facing RGB camera, two fisheye SLAM cameras, event-based cameras (ET), dual inertial measurement units (IMU), a magnetometer, a barometer, and an eight-microphone MEMS array. Although Meta's Machine Perception Services (MPS) were evaluated for SLAM and trajectory estimation, they were excluded from the pipeline due to their reliance on cloud-based post-processing and delayed inference windows. Consequently, all localization and perception workloads were relegated to a localized, on-premise computational framework.

B. Distributed Architecture and Network Topology

Direct deployment of the Project Aria SDK on ARM-based mobile or edge platforms was restricted by official platform support. To circumvent this constraint while preserving user portability, a three-tier distributed architecture was established. The Aria glasses served as the data-acquisition unit; a dedicated x86 server hosted the SDK and managed stream routing; and an Android mobile device served as the user endpoint. Communication between the server and the mobile client was established via a persistent WebSocket connection operating over a local Wi-Fi 6 network.

C. Environment Understanding Application

1) Pipeline: The core computational workload was offloaded to an Android application developed in Kotlin. Object detection was performed using a custom-trained YOLOv11n model, converted to TensorFlow Lite for hardware-accelerated inference via the NNAPI and GPU delegates. A custom dataset of indoor signage was collected at the Politecnico di Torino campus, capturing multiple video sequences from different angles and under varying lighting conditions using Project Aria glasses. After a thorough review of the navigation-relevant visual elements present throughout the campus, the following object classes were identified: room sign, room directions, exit sign, stair sign, and staircase. Ultralytics' recommendations were adopted for the dataset size, the split among training/validation/test, and the training parameters. However, after preliminary evaluation, an additional 500 frames per class were collected and annotated to improve robustness. Directional distinctions (left/right) were encoded directly into the classification space to reduce dependency on subsequent text recognition for navigational routing.

Detected bounding boxes exceeding a confidence threshold of 0.25 were subjected to non-maximum suppression and spatial rescaling. Regions of interest corresponding to textual signs were extracted, boundary-clamped, and upscaled by a factor of $5\times$ if dimensions fell below 100 pixels. Binarization was applied to enhance contrast under variable illumination. Text recognition was executed using Google ML Kit's on-device OCR engine. To mitigate character-level misreads in short alphanumeric identifiers, a fuzzy matching module was implemented. Tokenized OCR outputs were compared against the target destination using a weighted Levenshtein distance. Reduced substitution costs (0.2) were assigned to typographically similar character pairs (e.g., $0/o$, $1/l/i$, $5/s$). Contextual exclusion filters were applied to prevent false positives from adjacent signage containing overlapping room identifiers. An example of sign processing and recognition is shown in Fig. 1.

2) Interaction: A fully hands-free vocal interface was integrated to accommodate user accessibility requirements. Speech-to-text recognition was handled by the offline Vosk engine, utilizing a closed-grammar configuration to restrict the vocabulary to predefined destinations and control commands. A wake-word detection routine operated continuously in the background, triggering active listening upon recognition of initiation tokens. Destination input was facilitated through phonetically unambiguous spoken forms, which were subsequently mapped to compact system identifiers. Text-to-speech feedback was generated using the Google TTS API, with strict temporal gating applied to prevent instruction flooding. A cooldown mechanism ($5s$ default) suppressed redundant outputs, while exception handling allowed immediate auditory feedback upon successful OCR resolution following failed attempts. In the absence of detectable signage, a timeout guidance protocol was activated, escalating verbal prompts based on the duration of the detection gap. Navigation was terminated automatically upon destination confirmation or via manual voice command.

D. Spatial Localization Application

1) Pipeline: Visual localization and environmental mapping were achieved through ORB-SLAM3 (see Fig. 2), an open-source framework selected for its robust multi-map atlas management and support for visual-inertial sensor configurations. The system was configured to process rectified grayscale images alongside high-frequency IMU measurements, enabling real-time six-degree-of-freedom pose estimation and continuous trajectory optimization. Despite known limitations in low-texture environments and during slow-motion initialization, the framework's loop-closure and map-



Fig. 1. An example of sign processing with the Environment Understanding Application.

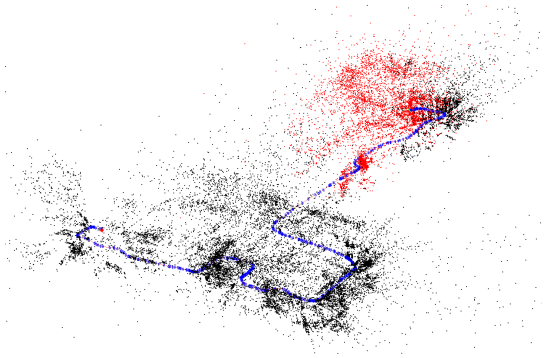


Fig. 2. An example of environmental mapping achieved through ORB-SLAM3.

merging capabilities were deemed sufficient for structured indoor navigation. For dynamic object detection, a YOLOv9 tracking architecture was deployed for its optimized gradient information propagation and efficient parameter utilization, thereby facilitating real-time object classification and bounding box regression. Concurrently, the ZoeDepth monocular depth estimation framework was utilized to generate metric depth maps from monocular frames, effectively bridging relative scale estimation with physical distance measurements. Detected objects falling outside a depth range of 0.8, m to 5.0, m were systematically filtered to mitigate computational overhead and suppress false positives in cluttered environments.

The server-side application was developed in Python and orchestrated four concurrent threads: an input manager, a SLAM request handler, an AI inference engine, and a synchronization module. To minimize inter-process communication latency, a C++ process executing ORB-SLAM3 interfaced with the Python environment via POSIX shared memory, utilizing a 20-slot ring buffer and semaphore-based synchronization to prevent data corruption and blocking. Prior to transmission, incoming frames were rectified to compensate for fisheye lens distortion, downscaled to 380×380 pixels, and converted to grayscale for SLAM processing, while a parallel RGB stream was routed to the AI inference thread. The synchronization module aggregated pose estimates and valid object detections, asynchronously matching the most recent SLAM output with corresponding neural network predictions. To maintain real-time performance under variable network conditions, outdated frames were systematically discarded, and a congestion-control mechanism was enforced to prevent buffer overflow.

2) Interaction: The client-side application was developed in Unity and deployed to an Android tablet to serve as the user interface and navigation terminal. The software architecture consists of two primary operational phases: an initialization sequence and a continuous navigation loop. During initialization, a persistent TCP connection to the server was established, native speech-to-text and text-to-speech engines were configured, and vocal destination input was awaited.

Upon successful localization, the user's position was continuously updated within a precomputed *NavMesh* representation of the indoor environment, and an optimal path was dynamically recalculated using an A*-based pathfinding algorithm. Directional instructions were synthesized and delivered via the text-to-speech module according to a strict filtering protocol: routine guidance was suppressed at 10-second intervals to reduce auditory fatigue, while immediate feedback was triggered only upon directional changes, proximity to critical waypoints, or state transitions.

Dynamic obstacles were instantiated as *NavMesh* obstacles using persistent tracking identifiers to enable real-time path avoidance without inducing excessive cognitive load. Manual camera manipulation was supported through multi-touch gestures, which automatically reset to a top-down orthographic view after a period of inactivity. To prevent input conflicts between gestural controls and voice recognition, speech activation was gated by a prolonged touch gesture, a constraint explicitly communicated to the user at session initiation.

IV. Experimental Evaluation and Results

A. Experimental Setup and Methodology

The experimental evaluation was conducted to assess both the technical performance and user-centric utility of the proposed navigation systems. A two-phase protocol was adopted: an initial familiarization sequence was performed under unobstructed vision conditions to allow users to acclimate to the auditory feedback mechanism, followed by a blindfolded navigation phase to simulate real-world assistive use cases. Quantitative metrics, including spatial accuracy, processing latency, frame throughput, object detection precision, and optical character recognition match rates, were systematically logged. Qualitative usability was assessed via the System Usability Scale questionnaires administered post-session.

B. Protocol

Trials were conducted at the Politecnico di Torino. All users were volunteers, and they did not receive any reward for participating in the tests. Users have been divided into two homogeneous groups of 10 people, balanced by gender, familiarity with wearable devices, and familiarity with assistive technology. The average age was, respectively, 25.2 and

25.3. Before the tests, participants provided written informed consent to participate and could discontinue participation at any time during the experiments. The experimental evaluation protocols followed the declaration of Helsinki [2]. Data were recorded in accordance with the European General Data Protection Regulation 2016/679, stored in local repositories with anonymized identities, and used only for the post-processing aggregate evaluation procedure. No sensitive data was recorded. Due to the limited number of testers, the study is conform with the Politecnico di Torino's Institutional Research Ethics Committee guidelines for pilot studies.

C. System Latency and Processing Throughput

Real-time processing capabilities were evaluated through comprehensive latency and throughput analyses for both applications. Computational benchmarks indicated that local neural network inference consistently remained below 150 ms, with optical character recognition averaging approximately 30 ms. Consequently, the mean end-to-end processing latency was approximately 260 ms, thereby satisfying the predefined temporal constraints. Frame drop rates remained minimal across all sessions, averaging approximately 3.3%, indicating that pipeline saturation was effectively avoided. The asynchronous architecture enabled continuous dissemination of spatial tracking data even when visual inference was temporarily delayed, with outdated detection frames accounting for less than 12% of the total stream. This design choice was determined to preserve navigational fluidity without compromising user safety.

D. Object Detection and Signage Recognition Performance

The accuracy of the object detection pipeline was quantified using standard computer vision metrics, including precision, recall, F1-score, and mean average precision at varying Intersection over Union (IoU) thresholds. Overall detection performance was deemed robust, with a mean precision exceeding 0.9 and an mAP@0.5 score of 0.939. Directional signage classes were identified with exceptional reliability, achieving recall values of 1.00 and mAP@0.5 scores over 0.97. Signage recognition was further evaluated through an optical character recognition pipeline, which incorporated sequential rejection gates based on bounding-box area, distortion metrics, and confidence thresholds. A functional match rate of approximately 0.62 was observed, while character-level accuracy reached 0.88 when restricted to destination-relevant text. The primary bottleneck was identified as the *TOO_SMALL* rejection category, which was attributed to the mounting heights and distances of indoor signage relative to the camera field of view. Reducing the area threshold was deliberately avoided to prevent the admission of low-resolution crops that would degrade optical character recognition fidelity.

E. Spatial Localization Accuracy and System Robustness

Spatial localization performance was assessed through measurements of accuracy and precision across multiple navigation trials. Precision, defined as the repeatability of the estimated position relative to ground truth, was consistently maintained at approximately 0.2 meters, indicating highly stable tracking behavior. Accuracy, representing the final distance to the designated target, showed greater variability, ranging from 0.3 to 3 meters, with an average value of about 2 meters. System robustness was further evaluated by monitoring tracking loss events and relocalization dynamics. Positioning failures can occur, typically triggered by abrupt head movements, transient network latency, or the traversal of visually homogeneous environments. Initial map initialization required between 150 and 350 frames under standard conditions, translating to an approximate setup duration of 15 to 30 seconds. Relocalization exhibited high variance depending on environmental context; rapid recovery (under two frames) was achieved when the field of view intersected previously mapped, high-feature regions, whereas prolonged reacquisition periods (exceeding 400 frames) were recorded in geometrically ambiguous or unexplored zones. These characteristics were deemed acceptable for semi-autonomous indoor navigation.

F. Usability Assessment

The two systems performed similarly in execution time, and users always successfully completed the proposed navigation tasks. Moreover, the blindfolded navigation phase confirmed that users could complete routing tasks independently, demonstrating that the system effectively bridged the information gap typically encountered in complex indoor environments. In terms of usability, both systems obtained SUS scores greater than 80, corresponding to an excellent rating. The *spatial localization* paradigm obtained an average SUS score of 83 ($\sigma = 8.12$), whereas the *environment understanding* paradigm obtained an average SUS score of 88.5 ($\sigma = 7.7$).

Based upon these preliminary results, it is possible to affirm that the two proposed paradigms allow users to effectively navigate indoor environments (**RQ1**). Despite the higher score achieved by the *spatial localization* paradigm, further tests with a larger user base would be necessary to statistically prove these results (**RQ2**).

G. Limitations

The Project Aria 1 device was primarily optimized for offline data acquisition rather than low-latency real-time inference. Thermal throttling was observed during extended high-fidelity streaming, and network bandwidth constraints occasionally introduced transmission jitter. Furthermore, the system was intentionally designed as a contextual sensory augmentation tool rather than a precise metric localization engine. Absolute coordinate mapping and exact distance computation were omitted to prioritize responsiveness and computational efficiency. Consequently, navigational success was contingent upon the user's cognitive mapping abilities, with the system functioning as a dynamic safety net that provided critical directional cues upon detecting relevant environmental landmarks.

The absence of BLV participants and the limited depth of paradigm comparison are both consequences of this study's

[2]https://www.wma.net/policies-post/wma-declaration-of-helsinki/

early-stage, exploratory scope. In line with staged HCI evaluation practices for assistive technologies, healthy users were involved in simulated blindfolded navigation to identify and resolve major usability issues before progressing to BLV-user studies. A more rigorous comparative evaluation, incorporating metrics such as localization accuracy under occlusion, processing latency across hardware configurations, and scalability to larger environments, is planned as future work.

V. Conclusions and Future Directions

The proposed work aimed to evaluate the capabilities of Meta Project Aria smart glasses for assistive technology applications tailored to individuals with blindness or low vision (BLV). Two primary navigational paradigms were evaluated: spatial localization and contextual environmental awareness through real-time object detection. Overall, both paradigms enabled independent completion of navigation tasks, confirming their viability for indoor wayfinding (**RQ1**). Users showed a slight preference for the *environment understanding* paradigm (SUS score of 88.5). (**RQ2**). However, since these results are based on preliminary evaluations with 20 participants at a single university campus, further validation should be conducted before generalizing to real-world BLV navigation needs.

Moreover, some technical enhancements were identified for subsequent development. First of all, combining the two approaches may improve system usability: a persistent global mapping layer will enable proactive guidance during periods of visual occlusion or when signage is absent. Furthermore, dynamic obstacle avoidance was identified as a critical area for expansion, wherein depth-sensing data or secondary detection models could be incorporated to alert users to transient hazards. Beyond the immediate scope of indoor navigation, the architectural framework established in this research was designed to serve as a modular foundation for comprehensive assistive platforms, contributing to a broader transition toward holistic assistive technologies.

Acknowledgment

The authors thank the Meta Project Aria staff for providing the devices used in this research project.

References

[1] Vision impairment and blindness. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] P. Ackland, S. Resnikoff, and R. Bourne, "World blindness and visual impairment: Despite many successes, the problem is growing," vol. 30, no. 100, pp. 71–73. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5820628/

[3] Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the Global Burden of Disease Study — Institute for Health Metrics and Evaluation. [Online]. Available: https://www.healthdata.org/research-analysis/library/trends-prevalence-blindness-and-distance-and-near-vision-impairment-over

[4] X. Li, H. Qiu, L. Wang, H. Zhang, C. Qi, L. Han, H. Xiong, and H. Li, "Challenges and Trends in Egocentric Vision: A Survey," vol. 23, no. 1, pp. 1–33. [Online]. Available: http://arxiv.org/abs/2503.15275

[5] Smriti. Google Glass Failure: Why did it fail and why was it discontinued? [Online]. Available: https://inspireip.com/google-glass-failure-why-was-it-discontinued/

[6] Meta Ray-Ban Glasses vs. Traditional Smart Glasses: What's the Differe. [Online]. Available: https://www.daposieyewear.com/blog/meta-ray-ban-glasses-vs-traditional-smart-glasses-whats-difference

[7] Envision Glasses - Smart Glasses for People who are Blind or Low Vision. [Online]. Available: https://www.letsenvision.com/glasses/home

[8] R. Cohen. Experience the Power of Assistive Technology with OrCam's AI Devices. OrCam Techologies. [Online]. Available: https://www.orcam.com/en-us/home

[9] Meta. Meta Project Aria. https://www.projectaria.com/. Accessed: 28/05/2026.

[10] W. C. S. S. Simões, G. S. Machado, A. M. A. Sales, p. u. family=Lucena, given=Mateus M., N. Jazdi, and J. Vicente F. de Lucena, "A Review of Technologies and Techniques for Indoor Navigation Systems for the Visually Impaired," vol. 20, no. 14. [Online]. Available: https://www.mdpi.com/1424-8220/20/14/3935

[11] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," vol. 37, no. 6, pp. 1067–1080. [Online]. Available: https://ieeexplore.ieee.org/document/4343996

[12] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," vol. 21, no. 3, pp. 2568–2599. [Online]. Available: https://ieeexplore.ieee.org/document/8692423

[13] M. Bamdad, D. Scaramuzza, and A. Darvishy. SLAM for Visually Impaired People: A Survey. [Online]. Available: http://arxiv.org/abs/2212.04745

[14] T. G. Hailu, X. Guo, and H. Si, "Indoor Positioning Systems as Critical Infrastructure: An Assessment for Enhanced Location-Based Services," vol. 25, no. 16. [Online]. Available: https://www.mdpi.com/1424-8220/25/16/4914

[15] A. Aladrén, G. López-Nicolás, L. Puig, and J. J. Guerrero, "Navigation Assistance for the Visually Impaired Using RGB-D Sensor With Range Expansion," vol. 10, no. 3, pp. 922–932. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6819807

[16] D. Das, A. Das, and F. Sadaf, Real-Time Wayfinding Assistant for Blind and Low-Vision Users.

[17] B. Li, J. P. Muñoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Arditi, and M. Yousuf, "Vision-Based Mobile Indoor Assistive Navigation Aid for Blind People," vol. 18, no. 3, pp. 702–714. [Online]. Available: https://ieeexplore.ieee.org/document/8370712

[18] Z. Chen, X. Liu, M. Kojima, Q. Huang, and T. Arai, "A Wearable Navigation Device for Visually Impaired People Based on the Real-Time Semantic Visual SLAM System," vol. 21, no. 4. [Online]. Available: https://www.mdpi.com/1424-8220/21/4/1536

[19] J. Herskovitz, A. Xu, R. Alharbi, and A. Guo, "Hacking, Switching, Combining: Understanding and Supporting DIY Assistive Technology Design by Blind People," in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, ser. CHI '23. Association for Computing Machinery, pp. 1–17. [Online]. Available: https://doi.org/10.1145/3544548.3581249

[20] C. Morrison, M. Grayson, R. F. Marques, D. Massiceti, C. Longden, L. Wen, and E. Cutrell, "Understanding Personalized Accessibility through Teachable AI: Designing and Evaluating Find My Things for People who are Blind or Low Vision," in Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility, ser. ASSETS '23. Association for Computing Machinery, pp. 1–12. [Online]. Available: https://dl.acm.org/doi/10.1145/3597638.3608395

[21] M. Bamdad, H.-P. Hutter, and A. Darvishy, "InCrowd-VI: A Realistic Visual–Inertial Dataset for Evaluating Simultaneous Localization and Mapping in Indoor Pedestrian-Rich Spaces for Human Navigation," vol. 24, no. 24. [Online]. Available: https://www.mdpi.com/1424-8220/24/24/8164

[22] J. Qiu, F. P.-W. Lo, X. Gu, Y. Sun, S. Jiang, and B. Lo. Indoor Future Person Localization from an Egocentric Wearable Camera. [Online]. Available: http://arxiv.org/abs/2103.04019

[23] B. Gamage, T.-T. Do, N. S. C. Price, A. Lowery, and K. Marriott. What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. [Online]. Available: http://arxiv.org/abs/2505.19325

[24] S. Ruffieux, C. Hwang, V. Junod, R. Caldara, D. Lalanne, and N. Ruffieux, "Tailoring assistive smart glasses according to pathologies of visually impaired individuals: An exploratory investigation on social needs and difficulties experienced by visually impaired individuals," vol. 22, no. 2, pp. 463–475. [Online]. Available: https://doi.org/10.1007/s10209-021-00857-5