Navigating the Last Mile: Evaluating Head- and Cane-Mounted Cameras for Egocentric Spatial Awareness

Apurv Varshney*
University of California, Santa Barbara

Lucas Nadolskis†
University of California, Santa Barbara

Tobias Höllerer‡
University of California, Santa Barbara

Michael Beyeler§
University of California, Santa Barbara

2026 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3195-2565-9/26/$31.00 ©2026 IEEE | DOI: 10.1109/ISMAR70079.2026.00029

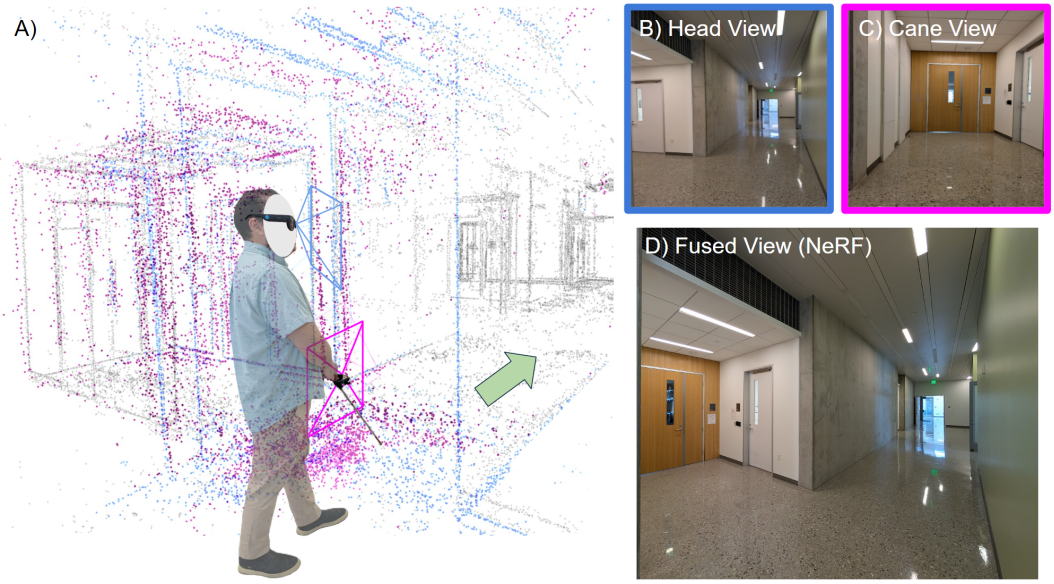


Figure 1: Evaluating head- vs. cane-mounted camera placements for assistive XR navigation in a complex "last-mile" indoor environment. A) Simultaneous Localization and Mapping (SLAM) point cloud reconstructions demonstrate how head-mounted (blue) and cane-mounted (pink) camera vantages capture complementary spatial geometry (gray points denote unmapped regions). B) Head-mounted view: forward-facing and gaze-aligned, providing the visual stability necessary for robust localization. C) Cane-mounted view: dynamic, ground-level perspective that captures critical low-lying geometry beyond the cane's tactile reach. D) Hybrid (head+cane) multi-view reconstruction via Neural Radiance Fields (NeRF): integrating both vantages synthesizes a high-fidelity 3D scene representation, combining tracking stability with dense geometric coverage to support downstream spatial reasoning.

Abstract

Robust navigational guidance is an important XR application for both sighted and non-sighted populations. In this paper, we mainly focus on blind pedestrians, who continue to face "last-mile" challenges such as locating entrances and navigating cluttered spaces. While smartglasses and wearables are maturing, a foundational design question remains underexplored: where on the body should cameras be placed to best support navigation? We present a mixed-methods investigation that focuses on the question of camera placement for generating spatial data supporting ego-centric navigation. A survey of 10 blind cane users surfaced practices for last-mile navigation and perceptions of body-mounted XR devices. A controlled case study with a blind co-author compared head- and cane-mounted cameras using synchronized Project Aria glasses while traversing five real-world environments. Using Simultaneous Localization and Mapping (SLAM) and Neural Radiance Fields (NeRFs) as diagnostic probes, we find the central trade-off appears to be not simply head versus cane but localization stability versus near-ground coverage, with the combined head+cane view providing complementary information for scene reconstruction. We translate these findings into architectural considerations for hybrid XR systems that extend the cane without interfering with tactile and auditory cues.

Index Terms: Accessibility, Assistive Technologies, Wearables, Orientation and Mobility

1 Introduction

Independent travel often breaks down for travelers at the "last mile": the final few meters of a journey, such as locating a doorway in a plaza or identifying the right bus at a stop (also known as the "last 10-meters/yards" problem [52, 15, 40]). These moments demand precise spatial awareness and robust environment tracking beyond the physical reach of the long cane or guide dog, remaining among the most challenging operational environments for assistive

*e-mail: apurv@ucsb.edu (Corresponding Author)
†e-mail: lgilnadolskis@ucsb.edu
‡e-mail: holl@cs.ucsb.edu
§e-mail: mbeyeler@ucsb.edu

2473-0726/26/$31.00 ©2026 IEEE
DOI 10.1109/ISMAR70079.2026.00029

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

Extended Reality (XR) systems [44, 21, 11, 19, 23].

Recent advances in wearable sensing have brought camera-equipped smartglasses and compact body-worn devices into everyday use [13, 12]. As this hardware matures, a foundational system design question remains underexplored: *where on the body should cameras be placed* to best support navigation for blind cane users? Prior work has explored head-, chest-, and cane-mounted configurations with distinct trade-offs: head-mounted cameras align with gaze and support landmarking [2, 34] but can miss low-lying hazards; chest or belt placements offer stability and discretion [42] yet often duplicate head-level coverage; and cane-mounted sensors couple to tactile exploration [46] but their sweeping motion can blur images and disrupt pose estimation [4]. Despite decades of assistive system proposals, camera placement itself has rarely been isolated as an independent design variable evaluated under controlled spatial computing conditions.

This paper treats camera placement as a first-class variable and examines its technical and experiential consequences for assistive XR navigation pipelines. We adopt a convergent mixed-methods design [10], collecting qualitative (formative survey) and quantitative (controlled vantage study) data in parallel and integrating results at the interpretation stage. We establish system requirements through an initial formative survey ($n = 10$) probing blind cane users' last-mile operational strategies, spatial information needs, and form-factor constraints for emerging wearables [50, 28, 14, 20]. We then conduct a controlled vantage study with an experienced blind co-author, directly comparing synchronized head- and cane-mounted Project Aria devices [12] across five complex indoor and outdoor locations.

Both studies were shaped through an iterative co-design process with the blind co-author [5, 41, 8, 35], an experienced independent cane user and long-time tester of assistive technologies. The co-design partner collaborated on survey question design, helped interpret and thematically code open-ended responses, and co-selected sites and landmarks for the vantage study to reflect authentic orientation and mobility (O&M) challenges. They were the sole participant in the data collection for the controlled Camera Vantage Study presented in Section 4, but had only limited involvement in the analysis of the collected data (helping with outcome-independent questions regarding the interpretation). This co-design approach ensured that our technical evaluations foregrounded authentic user priorities and practical form-factor limitations, rather than relying solely on isolated bench-testing, while also shielding against preferential bias.

We operationalize the effects of vantage using two core spatial computing pipelines widely used in XR navigation: Simultaneous Localization and Mapping (SLAM) as a proxy for maintaining a stable pose estimate and sparse geometric map, and Neural Radiance Fields (NeRF) as a proxy for generating high-fidelity, view-consistent 3D scene reconstructions suitable for downstream semantic reasoning. While coarse positioning (e.g., GPS) suffices for block-level guidance, last-mile decisions demand doorway-scale precision and robust indoor tracking [59, 23], requiring XR systems that can build and update persistent maps in situ.

Our analysis surfaces a core trade-off: head-mounted views provide the stability necessary for reliable localization, cane-mounted views capture complementary ground-level geometry useful for obstacle avoidance, and fusing both perspectives yields improved scene reconstruction quality. A stationary-cane control further suggests that motion, rather than the cane viewpoint itself, is the primary contributor to the reduced SLAM metrics. In short, our contributions are threefold:

- **Community-grounded system constraints:** a survey that surfaces operational barriers, spatial information needs, and form-factor requirements for assistive XR wearables.
- **Empirical characterization of vantage trade-offs:** a synchronized head-cane comparison across multiple environments, including a stationary-cane control study that isolates the impact of sweeping dynamics on computer vision pipelines.
- **Architectural considerations for hybrid or adaptive systems:** recommendations on multi-view integration and hybrid vantage strategies to balance localization stability with ground-level mapping for last-mile navigation.

2 BACKGROUND

We discuss related work on Navigation Challenges for Blind Travelers, Wearable XR Navigation Systems, and Emerging Spatial Pipelines before identifying Knowledge Gaps and reiterating our three main research questions.

2.1 Navigation Challenges for Blind Travelers

Navigating unfamiliar or complex environments remains persistently difficult for blind cane users, who rely primarily on tactile and auditory rather than visual strategies [50, 44, 62, 52, 22]. A long-running debate in assistive technology is whether systems should prioritize immediate *obstacle avoidance* or broader *spatial awareness and destination finding*. While early technical systems emphasized local obstacle detection [46, 38], Augmented Reality (AR) and XR are increasingly positioned as critical accessibility tools capable of delivering rich, global spatial context [9].

Environmental complexity poses significant challenges for these spatial computing pipelines. Indoors, wide-open layouts lack tactile anchors, prompting approaches that rely on pre-mapped AR environments[51] or crowdsourced visual route assistance [47]. Outdoors, traditional GPS lacks the doorway-scale precision required for the "last mile" [21]. To address this, modern XR interventions are exploring richer forms of support, including AR-driven tele-guidance [18], robotic companions [32, 24, 27], and conversational AI [25]. However, delivering this level of context-aware assistance requires highly robust computer vision localization, which fundamentally depends on where the tracking cameras are placed on the user's body. Our focus here is to *augment* proven tools (particularly the long cane) with computer vision capabilities rather than *replace* them.

2.2 Wearable XR Navigation Systems

A wide spectrum of wearable AR aids has been proposed, spanning head-, chest-, and cane-mounted placements. Each offers distinct camera vantages and affordances, yet few have been systematically compared regarding their impact on downstream computer vision tasks.

Head-mounted systems dominate both research and commercial accessibility applications, as evidenced by recent scoping reviews [37]. Devices like Envision Glasses [13, 45] and research platforms like Project Aria [12] utilize head-worn cameras for landmark detection, text recognition, and SLAM mapping. Head-mounted views inherently align with user orientation and gaze [2, 34], providing the stable viewpoints necessary for robust visual-inertial odometry. However, critics note their limited field of view for low-lying hazards and immediate ground geometry [38]. This trade-off between stable, gaze-aligned coverage and missed ground features is unresolved.

Chest- and belt-mounted systems offer alternatives, providing stable camera rigs without head-weight [56, 6]. While advocates highlight tracking stability and reduced conspicuousness [61], these placements often duplicate head-level coverage without significantly improving ground-plane reconstruction [31].

Cane-mounted systems embed vision or depth sensors directly into the mobility cane, promising a tight coupling with tactile exploration and unparalleled coverage of the ground plane [57, 16,

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

66, 1, 63]. However, this vantage point introduces severe technical challenges. The dynamic, sweeping motion of the cane causes motion blur and rapid viewpoint shifts that frequently break visual SLAM pose estimation [4, 39, 3]. While modern collaborative and visual-inertial SLAM architectures can handle complex, multi-agent tracking [54], the aggressive local motion of a cane often exceeds standard algorithmic tolerances, limiting its use as an independent tracking node.

Multimodal XR Feedback is the crucial counterpart to robust tracking. Once an XR system successfully maps an environment, it must convey spatial geometry without masking auditory cues. Recent work heavily explores multimodal and haptic AR interfaces to deliver this continuous spatial guidance, utilizing targeted audio [65, 55] alongside vibrotactile feedback [36] to communicate complex spatial information safely.

Taken together, the literature shows that head-mounted systems are stable but incomplete, chest-mounted ones offer redundancy with modest ergonomic gains, and cane-mounted ones provide unique but noisy perspectives. Crucially, prior work typically fixes a single head-, chest-, or cane-mounted device as one component of an end-to-end prototype and evaluates that system as a whole, so camera placement is confounded with the rest of the design — sensor model, algorithm, and interface all vary together. Few studies isolate placement itself under matched hardware and routes, leaving open how vantage alone shapes downstream computer vision tasks. We address exactly this gap.

2.3 Multi-View Sensing and Spatial Pipelines

Navigation performance is tightly coupled to the quality of sensor data and its integration. RGB-D, stereo, and Time of Flight (ToF) sensors provide crucial spatial depth cues[58, 26], while SLAM systems fuse vision and inertial data to track pose and build maps in real time[3]. Devices like Meta's Project Aria[12] offer tightly synchronized multi-sensor data streams optimized for such use cases.

Emerging view-synthesis methods, such as NeRF variants [17] promise to work on egocentric data and in dynamic environments, suggesting that unstable cane video could be reframed as a valuable, complementary signal for 3D reconstruction rather than simply defective input. Furthermore, co-design studies consistently show that blind travelers prefer technologies that extend and augment the long cane rather than replace it [14, 5, 48].

Our work bridges this gap between XR system design and user constraints. By evaluating SLAM and NeRF performance across synchronized head and cane vantages, we establish architectural guidelines for hybrid, multi-view camera placement. We aim to determine how to leverage the stability of head-mounted tracking alongside the rich geometric detail of cane-mounted cameras, ultimately supporting more resilient assistive XR pipelines.

2.4 Knowledge Gaps and Research Questions

Gap 1: Adoption and fit of emerging wearables. Prior surveys have examined general navigation needs of people who are blind [50, 28], but rarely focus on last-mile tasks or blind cane users' perceptions of emerging wearable placements. Key factors such as comfort, trust, and integration with established O&M techniques [14, 20] remain underexplored, yet strongly shape adoption and are essential for interpreting technical results.

Gap 2: Camera placement as a first-class design variable. Placement is usually treated as a given, yet its effect on the *type and quality of spatial information* remains untested. We isolate vantage (head vs. cane) under controlled conditions using two diagnostic probes: *(i) SLAM* for localization and sparse mapping, and *(ii) NeRF* for dense, view-consistent 3D reconstructions; applied here for the first time to cane-mounted video.

Guided by these gaps, we ask:

- **RQ1 (Survey on adoption/fit):** How do blind cane users approach last-mile navigation today, and what are their specific spatial information needs and form-factor constraints for integrating emerging wearables (head, cane, chest) with the traditional white cane?
- **RQ2 (Controlled study on vantage effects):** Holding hardware and environment constant, how does camera vantage (head vs. cane) affect localization stability and map quality (SLAM), and the fidelity of scene reconstructions (NeRF)? What is the specific role of cane-sweeping motion in any degradation?
- **RQ3 (Synthesis/design):** Given RQ1–RQ2, what design implications follow for assistive XR navigation; particularly for hybrid head+cane architectures that extend (rather than replace) the cane while minimizing cognitive load?

The survey (Section 3) addresses RQ1; the vantage study (Section 4) addresses RQ2 with a stationary-cane control that separates viewpoint from sweeping motion; the Discussion (Section 5) synthesizes RQ3 into actionable design guidance for hybrid assistive navigation.

3 FORMATIVE STUDY

To establish assistive XR requirements, we ran a formative survey of blind cane users' last-mile strategies and their perceptions of head-, chest-, and cane-mounted wearables.

3.1 Methods

Participants Ten blind adults (ages 18–65+) completed the survey (Table 1). Eligibility required age $\geq$18, self-reported total blindness, prior O&M training, and routine white cane use for independent travel. We recruited these experienced cane users via blindness community channels and word-of-mouth. All provided informed consent and received a $30 gift card.

Materials & Instrument The web-based survey comprised two parts. First, three co-designed scenario prompts (restaurant, crowded bus stop, shopping mall; full description in Appendix A) elicited open-ended responses regarding last-mile navigation strategies and pain points [10, 52, 44, 50, 21, 11]. These scenarios systematically varied in acoustics, spatial structure, and landmark availability. Second, targeted Likert items probed attitudes toward navigation priorities, environmental comfort, system trust, and device placement/form factor (Figure 2).

Analysis Open-ended responses underwent inductive thematic analysis, with two researchers iteratively developing a shared codebook until conceptual saturation. For 5-point Likert data, we report medians and 95% bias-corrected bootstrap confidence intervals (10,000 resamples). Given the sample size (n=10), inferential claims remain conservative; we avoid mean aggregation and instead treat individual items as reflective indicators for visualization (Figure 2) [7].

Ethics The survey protocol was reviewed by University of California, Santa Barbara Institutional Review Board (IRB) and determined to be exempt (Protocol No. 10-25-0112) under educational/benign behavioral research criteria. All procedures adhered to ethical standards for confidentiality and voluntary participation.

3.2 Cross-Cutting Navigation Challenges and Wearable Preferences

We evaluated participants' prior experience with assistive technologies alongside their camera placement preferences (Figure 2). Participants most frequently reported prior experience with smart glasses (6), followed by smart canes (4), chest-mounted devices (3),

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

ID	Age Range	Gender	Location	Years Since Blindness Onset	Years Using Cane	Travel Frequency	Cane Proficiency
P1	45-54	F	city	since birth	20+	weekly	moderate
P2	25-34	M	urban	since birth	11-20	weekly	moderate
P3	25-34	M	city	since birth	20+	daily	very proficient
P4	35-44	F	urban	since birth	20+	rarely	very proficient
P5	45-54	M	city	20+ years ago	20+	daily	very proficient
P6	65+	M	city	since birth	20+	monthly	very proficient
P7	18-24	M	urban	since birth	6-10	daily	very proficient
P8	25-34	M	urban	since birth	11-20	weekly	very proficient
P9	55-64	M	rural	20+ years ago	20+	rarely	extremely proficient
P10	35-44	M	city	20+ years ago	11-20	daily	extremely proficient

Table 1: Demographic and mobility characteristics of survey participants ($N = 10$). All participants self-identified as completely blind, reported receiving formal O&M training, and regularly use a white cane for independent navigation. All values are self-reported.

and wrist/hand-mounted solutions (2). Because participants emphasized destination-finding over mere obstacle avoidance, head-mounted placements received the highest comfort ratings. Conversely, direct cane integration was rated poorly due to added weight, sweep interference, and social conspicuousness, highlighting a strict requirement for lightweight, unobtrusive, hands-free designs.

Beyond wearable preferences, inductive coding of open-ended responses revealed three recurring navigation challenges across varying environments:

- **Sensory overload:** High ambient noise and reverberation (e.g., in malls or at bus stops) degraded essential auditory cues and increased cognitive load, leaving participants exhausted and disoriented (P1, P4).
- **Crowds and social dynamics:** Dense crowds necessitated constant path adjustments. Furthermore, crowded spaces introduced social vulnerabilities (e.g., harassment concerns, P4) and unsolicited bystander assistance that often disrupted orientation rather than helping.
- **Infrastructure inconsistency:** Homogeneous storefront designs (P7) and the absence of tactile landmarks, such as textural changes at bus stops (P4), hindered non-visual differentiation. This lack of standardization forces inefficient, trial-and-error navigation (P9) and prevents spatial memorization (P3).

3.3 Assistive XR Design Implications

By coding the open-ended scenario responses and integrating them with the Likert data, we distilled four critical design constraints for assistive XR systems:

- **C1: Unobstructed Audio (Mitigating Sensory Overload).** Participants described environments like malls and bus stops as acoustically chaotic. Because hearing is their primary tool beyond the cane's reach (used to track parallel traffic, echoes, and footfalls), XR audio interfaces must not occlude natural hearing. Spatial audio cues must be rendered sparsely and intelligently to prevent cognitive overload.
- **C2: Technology must complement rather than replace traditional aids.** The long cane is indispensable for near-field obstacle detection. Participants stressed that any cane-mounted sensors must not alter the sweeping rhythm or tactile feedback of the cane. Consequently, computer vision pipelines should not attempt to replace basic obstacle avoidance, but rather map the environment *beyond* the cane's physical reach [43, 30, 29, 57].
- **C3: Trust via Robust Localization.** Trust emerged as a non-negotiable barrier to adoption. In environments lacking stable tactile boundaries, participants rely on trial-and-error. If an XR system provides false spatial cues due to drifting pose estimation, it actively degrades safety. Therefore, underlying SLAM pipelines must maintain rigorous, high-confidence tracking before pushing spatial audio or haptic feedback.
- **C4: The Vantage Trade-off.** Participants highlighted a dichotomy: head-mounted views are comfortable and stable but miss low-lying hazards and textural boundaries; cane-mounted views capture the ground plane but introduce dynamic motion that users feared would degrade system performance.

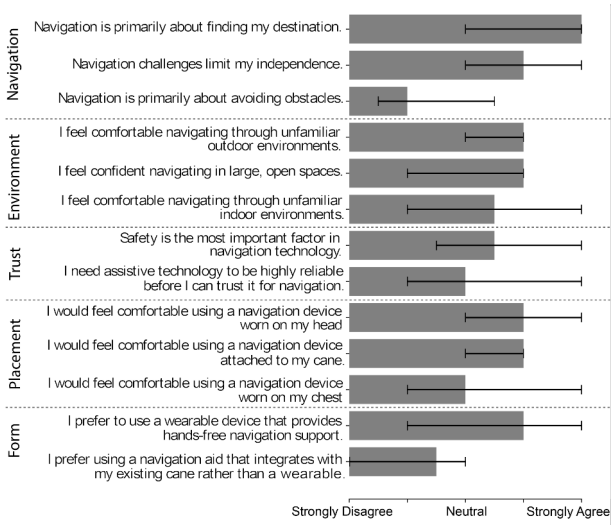


Figure 2: Participant responses to navigation and device preference questions grouped under five categories: Navigation, Environment, Trust, Placement, and Form, using a 5-point Likert scale (1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neither Agree Nor Disagree (Neutral), 4 = Somewhat Agree, 5 = Strongly Agree). Within each category, questions are presented in descending order based on their median score. These descriptive results suggest stronger comfort with head-mounted than cane-mounted devices, alongside a preference for hands-free aids that support destination finding.

These findings identify competing ergonomic and sensing requirements that motivate a controlled comparison of head- and cane-mounted recordings. However, the exact technical impact of placing tracking cameras on a sweeping cane which users identified as a major concern (C2 & C4) remains empirically under-explored. To move from user perception to systems evaluation, we next isolate vantage as an independent variable to test how standard XR tracking and mapping pipelines behave under these real-world constraints.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

4 CONTROLLED CAMERA VANTAGE STUDY

Building directly on the derived XR design constraints (specifically C3 and C4), we conducted a controlled study isolating camera vantage as an independent variable. To empirically evaluate how standard XR tracking and mapping pipelines perform under real-world O&M trajectories, we synchronized two Project Aria devices [12] - mounting one on the head and one on the sweeping cane of an experienced blind traveler. By holding the hardware and environment constant across five diverse indoor and outdoor last-mile scenarios, this quantitative evaluation [10, 60] moves beyond user perception to measure how head versus cane-mounted perspectives impact the foundational SLAM and spatial mapping capabilities required for reliable assistive XR. We frame it as a mechanistic case study: a single expert traveler, instrumented under tightly controlled conditions, isolates the effect of camera placement on the sensing pipeline rather than estimating how that effect varies across a population. This study characterizes the quality of spatial information prior to the design of any feedback layer; it does not benchmark deployed navigation systems, nor does it evaluate audio or haptic interfaces, which depend on the upstream spatial information we measure here.

4.1 Methods

4.1.1 Participant

The participant was a blind co-author who also served as a co-designer of the study, an experienced cane user with over a decade of independent travel experience. They have no light perception and travel independently on a daily basis in urban settings using a white cane, and were formally trained in O&M. Their deep familiarity with non-visual travel shaped the study to reflect authentic O&M practice. Their contributions included selecting study locations, identifying landmarks salient from a non-visual perspective, and collaborating on the interpretation of results, ensuring the evaluation remained grounded in lived experience.

4.1.2 Sensor Setup

We used two synchronized Project Aria devices [12] to record first-person video and inertial data (Figure 3A). We selected Project Aria because it provides synchronized egocentric RGB, wide-FOV monochrome cameras, and IMU streams, along with access to the Machine Perception Services (MPS) processing pipeline. This let us compare head- and cane-mounted placement while holding both the recording hardware and the processing tools constant. One device was mounted at the head using the standard glasses frame, while the other was attached to the cane shaft via a custom 3D-printed bracket (Figure 3B). The cane-mounted device was strapped to the bracket with zip ties, positioned so that with the cane held at its typical 30–45°operating angle the camera sat approximately 2 feet above the ground. The bracket was angled approximately 30°relative to the cane shaft so that, at this operating angle, the camera's field of view stayed directed forward along the direction of travel. Both devices recorded at 1408×1408 resolution and 20 frames per second, with internal clocks synchronized prior to each session. This setup allowed for tightly synchronized video and sensor streams across both vantage points while holding hardware constant.

The cane mount was designed in consultation with an O&M instructor to preserve natural sweep and minimize weight imbalance. The cane, originally 300 grams, now weighs 376 grams with the attached plastic filament mount and glasses. The blind co-author informally confirmed that the mount did not meaningfully disrupt the cane's balance or natural sweeping rhythm.

4.1.3 Environment Selection

Environments were chosen through a co-design process with our blind co-author, informed by the survey themes of entrance disambiguation, bus stop navigation, and large indoor orientation. We selected five sites to represent diverse acoustic and spatial conditions (indoor/outdoor, structured/open, static/dynamic), as described in Table 2: two outdoor environments and three indoor environments.

Routes ranged from 25–70 meters and were designed to incorporate both tactile and visual landmarks relevant to a blind traveler (Figure 4). Each environment presented realistic last-mile difficulties, including static obstacles (trash cans, furniture), dynamic elements (pedestrians), and variable lighting. This process ensured ecological validity while covering the range of last-mile challenges highlighted in Section 3.

4.1.4 Procedure

At each site, the participant completed one main trial, which was designed as a loop. Sessions followed a standardized routine at each site: (i) arrival at the location; (ii) a brief familiarization phase with an experimenter with the cane starting from a Start landmark till a Midpoint landmark and then back to the Start landmark. All the possible intermediate landmarks along this loop were shown to the participant; (iii) during the main trial, the participant retraced the same Start–Midpoint–Start loop independently, using standard O&M techniques and starting from the Start landmark. Whether to make use of the intermediate landmarks shown during familiarization was left to the participant's discretion. Both devices recorded simultaneously throughout the trial; (iv) recording was stopped once the participant completed the loop and came back to the Start landmark, marking the end of trial.

As an additional control, outside the main five-route comparison, we recorded a stationary-cane condition on one outdoor route (Outdoor2) and one indoor route (Indoor1). The participant walked the Start-to-Midpoint segment holding the cane forward and stationary rather than applying the sweeping motion. This condition shares the low cane-mounted viewpoint of normal cane use while removing the high-frequency motion sweeping introduces.

4.1.5 Data Collection and Analysis

Each session produced synchronized RGB video and IMU streams from both devices. Recordings were annotated with start and end markers, environment type, and vantage condition. Data were stored securely, preprocessed to ensure temporal alignment across sensors, and checked for integrity. Any corrupted or incomplete sessions were excluded prior to analysis.

To evaluate how vantage affects sensing performance, we analyzed the recordings using two complementary computer vision pipelines widely used in modern XR systems [4, 53, 33]:
SLAM (via Meta's proprietary MultiSLAM framework[1]) was used to estimate camera poses and recover sparse 3D map points, providing a measure of localization stability and geometric anchoring. Two metrics were derived from the system's outputs.

- **Filtered map points:** raw map points were filtered by nominal thresholds (inv_dist_std ≤ 0.005, dist_std ≤ 0.01) to retain only those the pipeline reports as having low positional uncertainty. We report both the count of points passing this filter and the *Map-point filter pass rate* (the proportion surviving); total raw counts additionally indicate potential coverage under more dynamic trajectories (e.g., cane sweeps).

- **Maximum-quality poses:** MPS assigns each camera pose a quality score in $[0, 1]$. We report the count of poses assigned the maximum score of 1.0 and the *Maximum-quality pose rate* (the proportion at that score), which we use as our primary indicator of localization stability.

[1] https://facebookresearch.github.io/projectaria_tools/docs/data_formats/mps/slam/mps_multi_slam

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

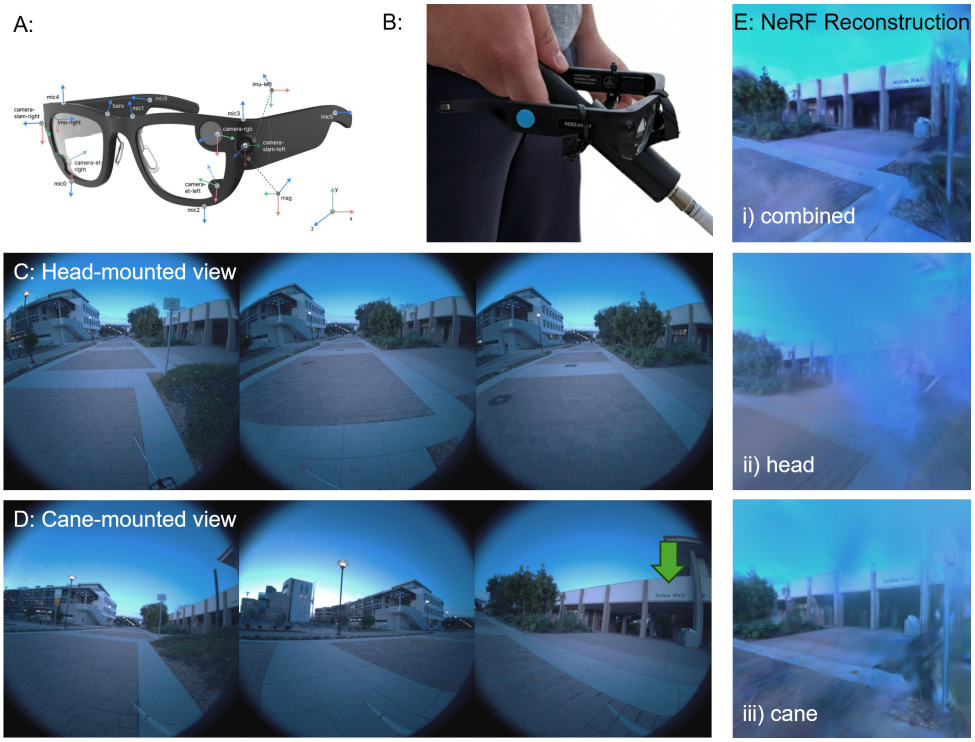


Figure 3: A) Meta's Project Aria smartglasses B) Cane Mount. The user navigates toward the entrance of Noble Hall (green arrow in D) in Outdoor2, situated next to an open plaza—a scenario where traditional canes offer limited spatial awareness. C) Head-mounted camera frames. D) Cane-mounted camera frames. E) NeRF-based 3D scene reconstructions compare head+cane (i), head-only (ii), and cane-only (iii) input streams.

Location	Distance	Description	Landmarks
Outdoor1	30m	Covered Walkway	Tactile warning pad, pillars, walls
Outdoor2	25m	Open Sidewalk	Sidewalk-to-plant area border, manhole cover
Indoor1	35m	Bright Modern Hallway	Floor mat at elevator, doors, railings, walls
Indoor2	70m	Simple layout Hallway	Garbage cans, walls, floor mats at exit door
Indoor3	45m	University Library	Garbage cans, hard-to-carpet floor transition, chairs, floor outlets, partitions

Table 2: Overview of navigation locations used in the study. Each corresponds to a specific physical site with a predefined walking path (distance indicates end-to-end route length) and notable environmental landmarks.

We treat these metrics as diagnostic outputs rather than independent ground-truth measures of mapping accuracy: both the pose quality score and the point-uncertainty estimates are computed internally by MPS using a proprietary calculation we cannot inspect. The filtering thresholds and quality criteria we apply follow MPS's general recommendations.

NeRFs were used to reconstruct view-consistent 3D scenes, offering a higher-fidelity substrate for semantic reasoning tasks such as entrance identification or free-space estimation. Specifically, we employed EgoLifter [17] as a representative modern egocentric reconstruction method and diagnostic probe — chosen because it targets egocentric video with complex motion and dynamic elements, not because it was designed for cane-mounted capture. Cane mounting is a demanding tool-mounted, body-adjacent case, with stronger motion and a lower viewpoint than the standard head-mounted egocentric video these methods assume. Reconstruction fidelity was quantified using peak signal-to-noise ratio (PSNR) and inspected qualitatively for coverage and coherence. For the combined Head+Cane condition, frames from both devices were supplied jointly as multi-view input to a single EgoLifter reconstruction; we did not perform any explicit runtime pose fusion between the two streams.

This dual analysis allowed us to examine how camera vantage affects two core spatial computing functions, localization and scene reconstruction, that downstream assistive XR systems may depend on.

4.1.6 Ethics

The vantage study involved data collection with a blind traveler in real environments and underwent full review and approval by University of California, Santa Barbara IRB (Protocol No. 9-24-0470).

The participant provided informed consent. Recordings were anonymized using EgoBlur [49] before analysis, and all procedures prioritized the participant's safety, comfort, and agency throughout the study.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.



Figure 4: Examples of key visual and tactile landmarks encountered during last-mile navigation tasks, including floor transitions, textured surfaces, doorways, tactile paving, and environmental boundaries.

4.2 Results

We evaluated vantage effects on SLAM performance and NeRF-based 3D scene reconstruction. Results are summarized in Table 3, Table 4, and Table 5, with qualitative examples shown in Figure 3 (see Appendix C for more). Below we highlight key findings.

4.2.1 SLAM Performance

As summarized in Table 3, the head-mounted camera consistently yielded higher values on both SLAM metrics. Its Maximum-quality pose rate exceeded 98% across all sites, while the cane-mounted rate dropped to 55–58% in outdoor settings. Although cane sweeps generated more raw map points overall, only 19–27% passed the uncertainty filter, compared to 27–39% for the head-mounted camera. These results suggest a consistent trade-off across the evaluated routes: head placement supports more stable localization, whereas cane placement captures broader coverage but with reduced reliability.

4.2.2 Stationary-Cane Control

To isolate the effect of motion, we compared three configurations: head-mounted, cane-mounted (dynamic sweep), and cane-mounted (stationary "no sweep"). Holding the cane steady improved the outdoor Maximum-quality pose rate from 39.5% to 83.2%, substantially narrowing the gap with the head-mounted rate (99.6%, Table 4). This suggests that the reduced pose quality was associated primarily with sweeping motion rather than the low viewpoint itself.

4.2.3 NeRF-Based 3D Scene Reconstruction

Reconstruction quality differed significantly by configuration (Table 5). A hierarchical mixed-effects model, with per-frame PSNR of held-out views as the unit of analysis and Scenario (indoor vs. outdoor) and Location (nested within Scenario) as random effects, showed a significant fixed effect of configuration on PSNR. [2] This structure captures variability at multiple levels of the environment hierarchy and isolates the effect of camera placement. Combined head+cane input achieved the highest PSNR, outperforming cane-only (Estimate = 1.204, SE = 0.193, $p < .001$) and head-only (Estimate = 1.811, SE = 0.194, $p < .001$). Cane-only also outperformed head-only (Estimate = 0.607, SE = 0.223, $p = .007$), suggesting its ground-level perspective adds complementary cues. Random effects indicated variability across Scenarios ($\sigma^2 = 7.346$) and Locations ($\sigma^2 = 5.906$), with PSNR generally lower outdoors, reflecting greater complexity and inconsistent lighting. Qualitative reconstructions (Figure 3 and Appendix C) likewise showed that hybrid input produced more coherent geometry, especially in cluttered indoor and sparse outdoor settings.

[2]Formula: `PSNR ~ Configuration + (1 | Scenario/Location)`

4.3 Summary of Findings

Together, these empirical results highlight a trade-off in assistive XR design: the tension between tracking stability and near-field spatial coverage. Head-mounted sensors yielded a higher Maximum-quality pose rate and a higher filter pass rate for map points, indicating the head as the more stable vantage in our evaluated pipelines for robust inside-out SLAM (addressing Constraint C3). Conversely, cane-mounted cameras suffered substantial tracking degradation, consistent with the erratic trajectories of the O&M sweep. A stationary-cane control indicated that this localization failure is associated primarily with sweeping motion rather than the low viewpoint itself (Constraint C2). However, despite poor tracking, the cane vantage successfully captured critical near-field, ground-level geometry largely missed by the head.

Crucially, the hybrid head-plus-cane configuration achieved the highest overall 3D reconstruction fidelity. This suggests that a hybrid architecture — anchoring SLAM pose estimation on the stable head while mapping the ground plane via the cane — is a promising direction for assistive spatial computing, though we evaluate it here only as a combined multi-view input to reconstruction rather than as a real-time fusion system. By grounding user perceptions (Constraint C4) in measured sensing pipeline performance, we show that head and cane vantages are empirically complementary in our setting rather than interchangeable.

5 GENERAL DISCUSSION

This work examines camera placement as a central design variable in wearable navigation aids for blind cane users. Integrating results across our two studies, we find that survey perceptions of cane instability were corroborated by the lower measured SLAM metrics, while survey skepticism about cane integration was challenged by the cane stream's complementary reconstruction benefits — a convergence that motivates the hybrid approaches discussed below.

5.1 Camera Placement Shapes Perception and Performance

Camera placement is not a one-size-fits-all decision but must be tailored to task and environment. Survey participants rated head-, chest-, and cane-mounted devices as somewhat acceptable (Figure 2), with preferences shaped by familiarity: head-mounted devices such as smartglasses were widely recognized, while skepticism toward cane-mounted systems reflected prior underperformance [46, 57, 28].

Quantitative evaluation underscored a trade-off between localization stability and spatial coverage. Head-mounted cameras consistently yielded higher values on SLAM metrics, aligning with prior work [39], while cane-mounted views captured unique low-level details [2, 38], evident in NeRF-based reconstructions (Figure 3). The stationary-cane control suggests that sweeping motion was a major contributor to reduced MPS pose quality on the two evaluated partial routes (Table 4). Rather than discarding cane-mounted vision, future systems should adapt to its dynamics with stabilization or motion-aware algorithms. Advancing tracking capabilities for this use case could benefit the broader XR community by making more robust inside-out tracking for headsets (e.g., VR sports games also have a lot of high frequency motion and need reliable tracking).

5.2 Implications for Assistive Technology Design

From a design perspective, our findings emphasize that navigation technologies must work in harmony with established practices. Participants emphasized the irreplaceable role of tactile cane feedback for near-field obstacle detection [64], noting that its utility cannot

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

Location	Filtered map points ↑		Maximum-quality poses ↑	
	Cane	Head	Cane	Head
Outdoor1	26.5% (361,373 / **1,364,893**)	**35.9%** (**390,267** / 1,087,275)	58.2% (69,125 / **118,771**)	**98.3%** (**111,789** / 113,676)
Outdoor2	27.3% (81,252 / **297,735**)	**34.5%** (**99,566** / 288,453)	55.1% (20,010 / 36,300)	**99.9%** (**42,759** / **42,810**)
Indoor1	27.3% (**449,647** / **1,645,114**)	**39.7%** (442,985 / 1,116,486)	84.7% (166,447 / 196,584)	**99.9%** (**196,912** / **197,014**)
Indoor2	19.6% (**297,709** / **1,520,318**)	**27.4%** (281,707 / 1,029,381)	73.2% (150,316 / **205,414**)	**99.4%** (**200,814** / 201,988)
Indoor3	22.4% (**498,057** / **2,223,797**)	**34.1%** (471,954 / 1,382,517)	94.6% (201,190 / **212,647**)	**99.9%** (**207,038** / 207,089)

Table 3: Comparison of SLAM performance across five navigation environments, showing the percentage (pass rate) and raw counts of Filtered map points and Maximum-quality poses (↑ higher is better), reported as a fraction of total map points or pose estimates. Filtered map points indicate potential coverage, and the Maximum-quality pose rate indicates localization stability. Bolded values indicate the best-performing configuration within each environment.

Location	Performance Metric	Head-Mounted Camera	Cane-Mounted Camera	Cane-Mounted (No Sweep)
Outdoor	Filtered map points ↑	**37.6%** (**50,531** / 135,285)	26.6% (40,846 / **153,566**)	30.1% (44,744 / 148,832)
	Maximum-quality poses ↑	**99.6%** (**13,315** / 13,366)	39.5% (6,868 / **17,408**)	83.2% (11,095 / 13,338)
Indoor	Filtered map points ↑	38.0% (82,448 / 217,227)	34.2% (101,556 / **296,521**)	**39.2%** (**115,779** / 295,671)
	Maximum-quality poses ↑	**99.8%** (25,561 / 25,613)	92.6% (**27,437** / **29,628**)	98.5% (26,167 / 26,575)

Table 4: Comparison of SLAM performance across indoor and outdoor environments, evaluating standard head- and cane-mounted configurations against a cane-mounted condition with sweeping motion disabled ("No Sweep"). Metrics include the pass rate (percentage) and raw counts of Filtered map points and Maximum-quality poses (↑ higher is better), reported as a fraction of total points or pose estimates. Filtered map points are those passing the MPS uncertainty thresholds; Maximum-quality poses are those assigned the maximum MPS quality score (1.0). Bolded percentages indicate the best-performing configuration for each metric within each environment.

Location	Cane	Head	Head+Cane
Outdoor1	12.95 ± 3.34	13.32 ± 3.85	$\mathbf{13.89 \pm 3.91}$
Outdoor2	12.66 ± 1.75	16.23 ± 3.35	$\mathbf{17.87 \pm 2.82}$
Indoor1	19.56 ± 2.41	19.73 ± 2.00	$\mathbf{22.21 \pm 2.74}$
Indoor2	21.25 ± 2.18	18.44 ± 2.13	$\mathbf{21.50 \pm 2.46}$
Indoor3	16.22 ± 2.33	15.76 ± 2.55	$\mathbf{16.46 \pm 2.19}$
Outdoor	12.89 ± 3.04	14.12 ± 3.92	$\mathbf{14.90 \pm 4.05}$
Indoor	18.96 ± 3.13	17.94 ± 2.78	$\mathbf{19.99 \pm 3.57}$

Table 5: The table reports PSNR (↑ higher is better) for each location under three input configurations: Cane-mounted, Head-mounted, and Head+Cane. Bold values indicate the highest mean PSNR in each row.

be replicated by other modalities. At the same time, tasks such as bus identification or navigating large open indoor spaces extend beyond the cane's reach. Participants expressed a preference for hands-free solutions that complement, not override, existing cane use. Head-mounted sensing offers stability and ease of adoption, while cane-mounted sensing contributes complementary ground-level information that improves scene reconstruction when supplied jointly with head-level input. This openness to cane-mounted sensing, despite mixed prior experiences, underscores a crucial design principle: technologies that align with established O&M practice are more likely to be trusted and adopted, while tools that interfere with sweep or attempt to replace tactile feedback risk rejection and cognitive overload. Future assistive systems should therefore adopt adaptive or hybrid vantage strategies that explicitly account for cane feedback, balancing localization stability with broader environmental coverage. More broadly, the need to combine stable and dynamically moving viewpoints is relevant to XR systems that rely on distributed body-worn sensors.

5.3 Toward Adaptive, Multisensor Navigation Systems

Because we measure the sensing layer rather than an interface, our clearest implications are constraints on what a feedback layer can draw upon, supporting tasks such as entrance identification, free-space estimation, and path planning. The head-mounted stream's stable localization is the more trustworthy basis for *global* cues such as heading, orientation, and route-level guidance, where drift would directly mislead the traveler (Constraint C3), while cane-derived geometry, though poor for localization, uniquely captures the near-ground structure the head misses, making it the better source for *local*, selective alerts surfaced only when relevant. This mirrors strategies cane users already employ, drawing simultaneously on tactile, auditory, and spatial-memory cues.

This division also governs how information reaches the traveler without overloading it: because the physical cane already conveys ground textures and curbs by touch, the XR system need not re-render these via audio, and can reserve spatial audio for objects beyond the cane's reach, such as overhanging branches, approaching vehicles, or specific storefronts. This addresses the fragility of auditory reliance our survey identified (Constraint C1), and the same principles of modality offloading and selective rendering apply to AR for sighted users, whose audio or haptic cues must not compete with visual attention on the physical world.

Realizing this will require advances in motion compensation, lightweight multi-view integration, and context-aware feedback, together with co-design to ensure such systems reduce cognitive burden and integrate with established O&M technique.

5.4 Limitations and Future Work

Despite these contributions, several limitations point to opportunities for future work.

The controlled vantage study involved a single blind participant and should be read as a mechanistic case study of camera placement under authentic cane use, not as evidence for population-level differences across blind travelers. This design let us hold participant behavior, hardware, routes, and environments constant while

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

comparing synchronized head- and cane-mounted streams across five real-world locations, isolating the sensing question in a way a multi-user study cannot at this level of control. It is therefore appropriate for characterizing how vantage shapes the signal available to SLAM and NeRF pipelines, but it does not establish how those effects vary across travelers, cane techniques, or gaits, and does not replace a broader user study.

Second, our study environments were intentionally bounded. The survey used three analytic scenarios (restaurant, bus stop, mall) and the vantage study examined five real-world routes, chosen for clarity and comparability. While they captured variation in spatial and sensory demands, they do not reflect the full diversity of blind travel; future work should extend evaluation to dynamic, crowded settings that surface the broader social and sensory complexities raised in the survey [7].

Third, we evaluated vantage using SLAM and NeRF as diagnostic probes rather than integrated real-time systems. This allowed us to isolate placement effects, but stops short of demonstrating end-to-end deployment. Follow-up work should validate these trends in real-time assistive pipelines, advancing lightweight SLAM, multi-view reconstruction, and camera stabilization.

Finally, differences in scale between the survey (multi-user) and vantage study (single-user) limit direct comparability. While this constrains integration validity, triangulating across methods strengthens confidence in the core finding: head- and cane-mounted perspectives are complementary.

6 CONCLUSION

This work shows that camera placement is as critical to assistive navigation design as the data processing techniques and algorithms. By combining community-grounded perspectives with a controlled evaluation of vantage, our results indicate that head-mounted cameras provide stable localization while cane-mounted cameras contribute complementary ground-level detail; together offering richer, more reliable scene understanding than either alone.

These findings underscore that future assistive XR systems should augment, not replace, the tactile foundation of the cane, integrating head- and cane-mounted perspectives to balance stability with coverage. Hybrid approaches that respect established O&M practices hold promise for enabling confident, independent last-mile travel.

Looking ahead, large-scale, multi-user datasets collected in diverse real-world conditions will be essential to refine these hybrid designs, enabling spatial AI systems that fuse multiple viewpoints, adapt to the dynamics of cane use, and earn the trust of blind travelers. By designing for the demanding constraints of blind pedestrians, this work points towards assistive XR systems that better balance localization stability with near-ground scene coverage, while also suggesting broader lessons for robust embodied sensing in XR.

ACKNOWLEDGMENTS

We thank Brianna Pettit for giving insights into O&M instruction and our undergrad Research Assistants for helping fabricate cane mount for the glasses. Supported by the National Library of Medicine of the National Institutes of Health (NIH) under Award Number DP2-LM014268. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. In accordance with IEEE guidelines, generative AI tools were used in limited ways during manuscript preparation. Gemini 3 Pro (Google) was used for light editing and language refinement of portions of the manuscript. All AI-generated outputs were reviewed and verified by the authors, who take full responsibility for the content of the article.

SUPPLEMENTAL MATERIALS

Supplemental materials include the full survey instrument, additional qualitative results, SLAM point clouds and NeRF reconstructions for all five locations, and documentation of the co-design process. A video shows synchronized head- and cane-mounted recordings from one representative route. We also release the full dataset and derived analysis outputs under a CC BY 4.0 license at https://huggingface.co/datasets/apurv4166/LastMileEvaluation.

REFERENCES

[1] S. Agrawal, M. E. West, and B. Hayes. A novel perceptive robotic cane with haptic navigation for enabling vision-independent participation in the social dynamics of seat choice. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9156–9163. IEEE, 2022. 2

[2] K. W. Arthur. *Effects of field of view on performance with head-mounted displays*. phd, The University of North Carolina at Chapel Hill, 2000. AAI9968542 ISBN-10: 0599733721. 2, 7

[3] M. Bamdad, H.-P. Hutter, and A. Darvishy. Deep learning-powered visual slam aimed at assisting visually impaired navigation. *arXiv preprint arXiv:2510.20549*, 2025. 3

[4] M. Bamdad, D. Scaramuzza, and A. Darvishy. Slam for visually impaired people: A survey. *IEEE Access*, 12:130165–130211, 2024. doi: 10.1109/access.2024.3454571 2, 3, 5

[5] C. L. Bennett and D. K. Rosner. The promise of empathy: Design, disability, and knowing the "other". In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300528 2, 3

[6] Biped.ai. Noa. https://biped.ai/. 2

[7] A. Blandford, D. Furniss, and S. Makri. *Qualitative Hci Research: Going Behind the Scenes*. Morgan & Claypool Publishers, Mar. 2016. 3, 9

[8] E. Brulé, B. J. Tomlinson, O. Metatla, C. Jouffrais, and M. Serrano. Review of quantitative empirical evaluations of technology for people with visual impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020. 2

[9] J. M. Coughlan and J. Miele. Ar4vi: Ar as an accessibility tool for people with visual impairments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 288–292, 2017. doi: 10.1109/ISMAR-Adjunct.2017.89 2

[10] J. W. Creswell and V. L. P. Clark. *Designing and Conducting Mixed Methods Research*. SAGE Publications, Inc, Los Angeles, 2011. 2, 3, 5

[11] F. E.-Z. El-Taher, L. Miralles-Pechuán, J. Courtney, K. Millar, C. Smith, and S. Mckeever. A survey on outdoor navigation applications for people with visual impairments. *IEEE Access*, 11:14647–14666, 2023. doi: 10.1109/ACCESS.2023.3244073 2, 3

[12] J. Engel, K. Somasundaram, M. Goesele, A. Sun, A. Gamino, A. Turner, A. Talattof, A. Yuan, B. Souti, B. Meredith, C. Peng, C. Sweeney, C. Wilson, D. Barnes, D. DeTone, D. Caruso, D. Valleroy, D. Ginjupalli, D. Frost, E. Miller, E. Mueggler, E. Oleinik, F. Zhang, G. Somasundaram, G. Solaira, H. Lanaras, H. Howard-Jenkins, H. Tang, H. J. Kim, J. Rivera, J. Luo, J. Dong, J. Straub, K. Bailey, K. Eckenhoff, L. Ma, L. Pesqueira, M. Schwesinger, M. Monge, N. Yang, N. Charron, N. Raina, O. Parkhi, P. Borschowa, P. Moulon, P. Gupta, R. Mur-Artal, R. Pennington, S. Kulkarni, S. Miglani, S. Gondi, S. Solanki, S. Diener, S. Cheng, S. Green, S. Saarinen, S. Patra, T. Mourikis, T. Whelan, T. Singh, V. Balntas, V. Baiyya, W. Dreewes, X. Pan, Y. Lou, Y. Zhao, Y. Mansour, Y. Zou, Z. Lv, Z. Wang, M. Yan, C. Ren, R. D. Nardi, and R. Newcombe. Project aria: A new tool for egocentric multi-modal ai research, 2023. 2, 3, 5

[13] Envision. Envision smart glasses. https://www.letsenvision.com/, 2023. 2

[14] 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. In *Proceedings of the 25th International ACM SIGACCESS Conference on*

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

Computers and Accessibility, ASSETS '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3597638.3608955 2, 3

[15] C. Gleason, A. J. Fiannaca, M. Kneisel, E. Cutrell, and M. R. Morris. FootNotes: Geo-referenced Audio Annotations for Nonvisual Exploration. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3):109:1–109:24, Sept. 2018. doi: 10.1145/3264919 1

[16] I. Gonsher, A. Salazar, S. Mehta, S. Shulman, N. Gaitanis, A. Khosla, D. D. Tamesis, and J. Sun. The Smart Cane Project: Integrating Screen Interfaces and Physiological Sensors into Mobility Devices. In *Emerging Technologies in Healthcare and Medicine*, vol. 116. AHFE Open Acces, 2023. ISSN: 27710718 Issue: 116. doi: 10.54941/ahfe1004383 2

[17] Q. Gu, Z. Lv, D. Frost, S. Green, J. Straub, and C. Sweeney. Egolifter: Open-world 3d segmentation for egocentric perception. *arXiv preprint arXiv:2403.18118*, 2024. 3, 6

[18] R. Guarese, E. Pretty, and F. Zambetta. Xr towards tele-guidance: mixing realities in assistive technologies for blind and visually impaired people. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 324–329, 2023. doi: 10.1109/VRW58643.2023.00074 2

[19] M. Gupta, A. Abdolrahmani, E. Edwards, M. Cortez, A. Tumang, Y. Majali, M. Lazaga, S. Tarra, P. Patil, R. Kuber, and S. M. Branham. Towards More Universal Wayfinding Technologies: Navigation Preferences Across Disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Apr. 2020. doi: 10.1145/3313831.3376581 2

[20] 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*, CHI '23, pp. 1–17. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581249 2, 3

[21] G. Jain, B. Hindi, Z. Zhang, K. Srinivasula, M. Xie, M. Ghasemi, D. Weiner, S. A. Paris, X. Y. T. Xu, M. Malcolm, M. K. Turkcan, J. Ghaderi, Z. Kostic, G. Zussman, and B. A. Smith. Streetnav: Leveraging street cameras to support precise outdoor navigation for blind pedestrians. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3654777.3676333 2, 3

[22] G. Jain, Y. Teng, D. H. Cho, Y. Xing, M. Aziz, and B. A. Smith. "i want to figure things out": Supporting exploration in navigation for people with visual impairments. *Proc. ACM Hum.-Comput. Interact.*, 7(CSCW1), Apr. 2023. doi: 10.1145/3579496 2

[23] H. Kacorri, E. Ohn-Bar, K. M. Kitani, and C. Asakawa. Environmental Factors in Indoor Navigation Based on Real-World Trajectories of Blind Users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–12. Association for Computing Machinery, New York, NY, USA, Apr. 2018. doi: 10.1145/3173574.3173630 2

[24] R. Kamikubo, S. Kayukawa, Y. Kaniwa, A. Wang, H. Kacorri, H. Takagi, and C. Asakawa. Beyond Omakase: Designing Shared Control for Navigation Robots with Blind People, Mar. 2025. arXiv:2503.21997 [cs]. doi: 10.1145/3706598.3714112 2

[25] KaniwaYuka, KuribayashiMasaki, KayukawaSeita, SatoDaisuke, TakagiHironobu, AsakawaChieko, and MorishimaShigeo. ChitChatGuide: Conversational Interaction Using Large Language Models for Assisting People with Visual Impairments to Explore a Shopping Mall. *Proceedings of the ACM on Human-Computer Interaction*, Sept. 2024. Publisher: ACMPUB27New York, NY, USA. doi: 10.1145/3676492 2

[26] R. K. Katzschmann, B. Araki, and D. Rus. Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3):583–593, 2018. doi: 10.1109/TNSRE.2018.2800665 3

[27] S. Kayukawa, D. Sato, M. Murata, T. Ishihara, H. Takagi, S. Morishima, and C. Asakawa. Enhancing blind visitor's autonomy in a science museum using an autonomous navigation robot. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581220 2

[28] I. Khan, S. Khusro, and I. Ullah. Technology-assisted white cane: evaluation and future directions. *PeerJ*, 6:e6058, Dec. 2018. doi: 10.7717/peerj.6058 2, 3, 7

[29] S. Y. Kim and K. Cho. Usability and design guidelines of smart canes for users with visual impairments. *international Journal of Design*, 7(1), 2013. 4

[30] Á. Kristjánsson, A. Moldoveanu, Ó. I. Jóhannesson, O. Balan, S. Spagnol, V. V. Valgeirsdóttir, and R. Unnthorsson. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative neurology and neuroscience*, 34(5):769–787, 2016. 4

[31] M. Kubota, M. Kuribayashi, S. Kayukawa, H. Takagi, C. Asakawa, and S. Morishima. Snap&Nav: Smartphone-based Indoor Navigation System For Blind People via Floor Map Analysis and Intersection Detection. *Proc. ACM Hum.-Comput. Interact.*, 8(MHCI):275:1–275:22, Sept. 2024. doi: 10.1145/3676522 2

[32] M. Kuribayashi, K. Uehara, A. Wang, S. Morishima, and C. Asakawa. WanderGuide: Indoor Map-less Robotic Guide for Exploration by Blind People, Feb. 2025. arXiv:2502.08906 [cs]. doi: 10.1145/3706598.3713788 2

[33] O. Kwon, J. Park, and S. Oh. Renderable neural radiance map for visual navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9099–9108, 2023. 5

[34] K. Lee, D. Sato, S. Asakawa, C. Asakawa, and H. Kacorri. Accessing Passersby Proxemic Signals through a Head-Worn Camera: Opportunities and Limitations for the Blind. *ASSETS. Annual ACM Conference on Assistive Technologies*, 21:10.1145/3441852.3471232, 2021. doi: 10.1145/3441852.3471232 2

[35] J. Li, S. Kim, J. A. Miele, M. Agrawala, and S. Follmer. Editing spatial layouts through tactile templates for people with visual impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–11. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300436 2

[36] J. Li, Z. Zhang, Z. Yan, Y. Zhao, and H. Peng. Comparing vibrotactile and skin-stretch haptic feedback for conveying spatial information of virtual objects to blind vr users. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 686–696, 2025. doi: 10.1109/VR59515.2025.00091 3

[37] Y. Li, K. Kim, A. Erickson, N. Norouzi, J. Jules, G. Bruder, and G. F. Welch. A scoping review of assistance and therapy with head-mounted displays for people who are visually impaired. *ACM Trans. Access. Comput.*, 15(3), Aug. 2022. doi: 10.1145/3522693 2

[38] Y. Lin, K. Wang, W. Yi, and S. Lian. Deep learning based wearable assistive system for visually impaired people. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 2549–2557, 2019. doi: 10.1109/ICCVW.2019.00312 2, 7

[39] P. Liu, X. Zuo, V. Larsson, and M. Pollefeys. MBA-VO: Motion Blur Aware Visual Odometry. pp. 5530–5539. IEEE Computer Society, Oct. 2021. doi: 10.1109/ICCV48922.2021.00550 3, 7

[40] J. C. Lock, G. Cielniak, and N. Bellotto. A Portable Navigation System with an Adaptive Multimodal Interface for the Blind. In *AAAI Spring Symposium Series*, 2017. 1

[41] K. Mack, E. McDonnell, D. Jain, L. Lu Wang, J. E. Froehlich, and L. Findlater. What do we mean by "accessibility research"? a literature survey of accessibility papers in chi and assets from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2021. 2

[42] F. Mathis and J. Schöning. LifeInsight: Design and Evaluation of an AI-Powered Assistive Wearable for Blind and Low Vision People Across Multiple Everyday Life Scenarios. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, pp. 1–25. Association for Computing Machinery, New York, NY, USA, Apr. 2025. doi: 10.1145/3706598.3713486 2

[43] R. Milallos, V. Tibdewal, Y. Wang, A. Ogueh Udegbe, and T. Oh. "would the smart cane benefit me?": Perceptions of the visually impaired towards smart canes. In *Proceedings of the 23rd International*

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.

ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3441852.3476524 4

[44] K. Müller, C. Engel, C. Loitsch, R. Stiefelhagen, and G. Weber. Traveling more independently: A study on the diverse needs and challenges of people with visual or mobility impairments in unfamiliar indoor environments. *ACM Trans. Access. Comput.*, 15(2), May 2022. doi: 10.1145/3514255 2, 3

[45] OrCam. Orcam myeye 3 pro. https://www.orcam.com/en-us/orcam-myeye-3-pro. 2

[46] J. Pariti, V. Tibdewal, and T. Oh. Intelligent mobility cane - lessons learned from evaluation of obstacle notification system using a haptic approach. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–8. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3375217 2, 7

[47] D. Plikynas, A. Indriulionis, A. Laukaitis, and L. Sakalauskas. Indoor-guided navigation for people who are blind: Crowdsourcing for route mapping and assistance. *Applied Sciences*, 12(1), 2022. doi: 10.3390/app12010523 2

[48] X. Qi and J. Yu. Participatory design in human-computer interaction: Cases, characteristics, and lessons. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25. Association for Computing Machinery, New York, NY, USA, 2025. doi: 10.1145/3706598.3713436 3

[49] N. Raina, G. Somasundaram, K. Zheng, S. Miglani, S. Saarinen, J. Meissner, M. Schwesinger, L. Pesqueira, I. Prasad, E. Miller, et al. Egoblur: Responsible innovation in aria. *arXiv preprint arXiv:2308.13093*, 2023. 6

[50] S. Real and A. Araujo. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors*, 19(15), 2019. doi: 10.3390/s19153404 2, 3

[51] P. Ruvolo. Considering spatial cognition of blind travelers in utilizing augmented reality for navigation. In *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pp. 99–104, 2021. doi: 10.1109/PerComWorkshops51409.2021.9430997 2

[52] M. Saha, A. J. Fiannaca, M. Kneisel, E. Cutrell, and M. R. Morris. Closing the gap: Designing for the last-few-meters wayfinding problem for people with visual impairments. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, p. 222–235. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3308561.3353776 1, 2, 3

[53] M. Sakashita, B. Thoravi Kumaravel, N. Marquardt, and A. D. Wilson. Sharednerf: Leveraging photorealistic and view-dependent rendering for real-time and remote collaboration. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2024. 5

[54] P. Schmuck, T. Ziegler, M. Karrer, J. Perraudin, and M. Chli. COVINS: Visual-Inertial SLAM for Centralized Collaboration . In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 171–176. IEEE Computer Society, Los Alamitos, CA, USA, Oct. 2021. doi: 10.1109/ISMAR-Adjunct54149.2021.00043 3

[55] J. Sessner, F. Dellert, and J. Franke. Multimodal feedback to support the navigation of visually impaired people. In *2022 IEEE/SICE International Symposium on System Integration (SII)*, pp. 196–201, 2022. doi: 10.1109/SII52469.2022.9708751 3

[56] S. Shoval, I. Ulrich, and J. Borenstein. Navbelt and the guidecane [obstacle-avoidance systems for the blind and visually impaired]. *IEEE Robotics & Automation Magazine*, 10(1):9–20, 2003. doi: 10.1109/MRA.2003.1191706 2

[57] P. Slade, A. Tambe, and M. J. Kochenderfer. Multimodal sensing and intuitive steering assistance improve navigation and mobility for people with impaired vision. *Science Robotics*, 6(59):eabg6594, 2021. doi: 10.1126/scirobotics.abg6594 2, 4, 7

[58] P. Theodorou, K. Tsiligkos, and A. Meliones. Multi-sensor data fusion solutions for blind and visually impaired: Research and commercial navigation applications for indoor and outdoor spaces. *Sensors*, 23(12), 2023. doi: 10.3390/s23125411 3

[59] L. M. Turkstra, T. Bhatia, A. Van Os, and M. Beyeler. Assistive technology use in domestic activities by people who are blind. *Scientific Reports*, 15(1):7486, Mar. 2025. Publisher: Nature Publishing Group. doi: 10.1038/s41598-025-91755-w 2

[60] K. Van Turnhout, A. Bennis, S. Craenmehr, R. Holwerda, M. Jacobs, R. Niels, L. Zaad, S. Hoppenbrouwers, D. Lenior, and R. Bakker. Design patterns for mixed-method research in hci. In *Proceedings of the 8th Nordic conference on human-computer interaction: Fun, fast, foundational*, pp. 361–370, 2014. 5

[61] H.-C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré, and D. Rus. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6533–6540, 2017. doi: 10.1109/ICRA.2017.7989772 2

[62] M. Wang, N. Zhu, V. Renaudin, A. Dommes, and M. Servières. Understanding and using spatial landmarks of visually impaired people for navigation applications. In *2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, Sep. 2023. doi: 10.1109/IPIN57070.2023.10332513 2

[63] WeWALK. Wewalk smart cane. https://wewalk.io/en/, 2020. 2

[64] M. A. Williams, C. Galbraith, S. K. Kane, and A. Hurst. "just let the cane hit it": how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, ASSETS '14, pp. 217–224. Association for Computing Machinery, New York, NY, USA, Oct. 2014. doi: 10.1145/2661334.2661380 7

[65] X. Zhang, X. Yao, Y. Zhu, and F. Hu. An arcore based user centric assistive navigation system for visually impaired people. *Applied Sciences*, 9(5), 2019. doi: 10.3390/app9050989 3

[66] Y. Zhao, C. L. Bennett, H. Benko, E. Cutrell, C. Holz, M. R. Morris, and M. Sinclair. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–14, 2018. 2

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on November 14,2026 at 18:11:55 UTC from IEEE Xplore. Restrictions apply.