Beyond Physical Reach: Comparing Head- and Cane-Mounted Cameras for Last-Mile Navigation by Blind Users

Apurv Varshney
apurv@ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

Lucas Nadolskis
lgilnadolskis@ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

Tobias Höllerer
holl@cs.ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

Michael Beyeler
mbeyeler@ucsb.edu
University of California, Santa Barbara
Santa Barbara, CA, USA

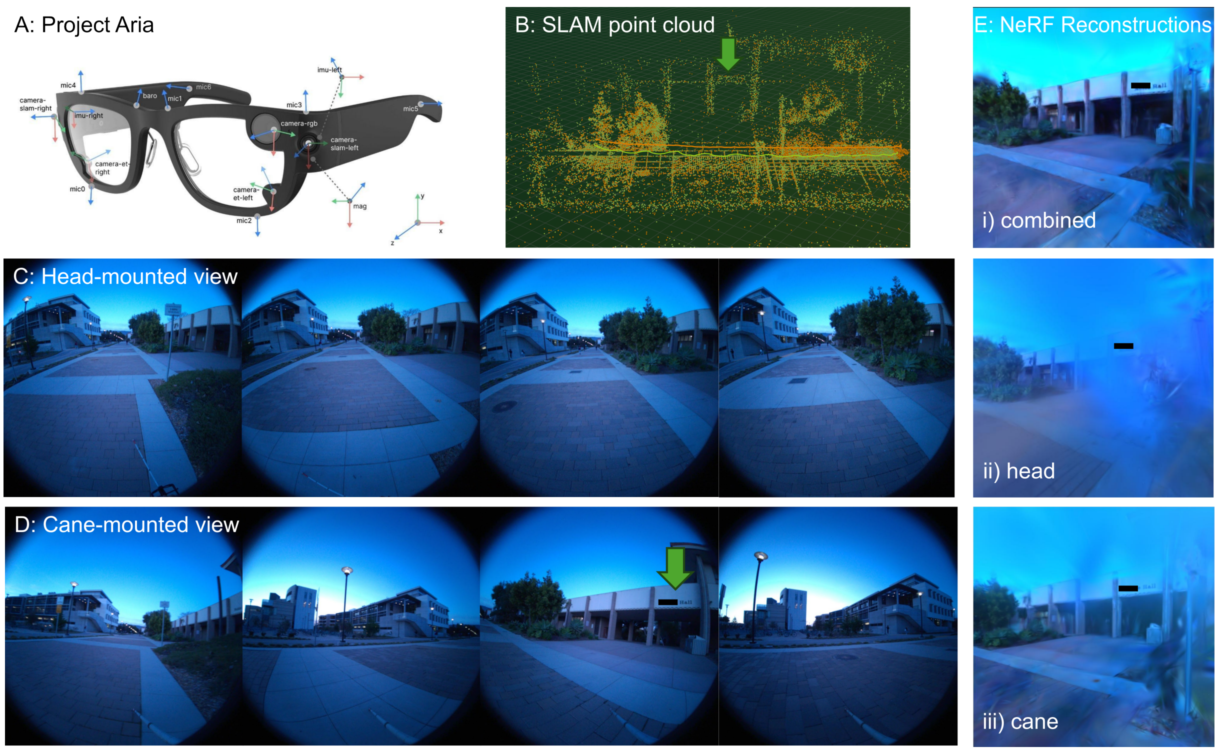

Figure 1: Head- vs cane-mounted cameras for last-mile navigation. The user navigates toward the entrance of Anonymous Hall (green arrow in D), situated next to an open plaza—a scenario where traditional canes offer limited spatial awareness. A) Meta's Project Aria smartglasses include five cameras (two $150 \times 120^{\circ}$ mono scene, one $110 \times 110^{\circ}$ RGB, and two eye-tracking), plus nonvisual sensors (IMUs, magnetometer, barometer, GPS, Wi-Fi/Bluetooth beacons, microphones). B) SLAM-generated point cloud from head-mounted (orange) and cane-mounted (green) cameras. C) Head-mounted camera frames show forward-facing views ideal for landmark recognition. D) Cane-mounted camera frames capture a broader, ground-level FOV, detecting obstacles beyond both the cane's tip and the head-mounted camera's reach. E) NeRF-based 3D scene reconstructions compare head+cane (i), head-only (ii), and cane-only (iii) input streams for path planning.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ABSTRACT

Blind individuals face persistent challenges in last-mile navigation, including locating entrances, identifying obstacles, and navigating complex or cluttered spaces. Although wearable cameras are increasingly used in assistive systems, there has been no systematic, vantage-focused comparison to guide their design. This paper addresses that gap through a two-part investigation. First, we surveyed ten experienced blind cane users, uncovering navigation strategies, pain points, and technology preferences. Participants stressed the importance of multi-sensory integration, destination-focused travel, and assistive tools that complement (rather than replace) the cane's tactile utility. Second, we conducted controlled data collection with a blind participant navigating five real-world environments using synchronized head- and cane-mounted cameras, isolating vantage placement as the primary variable. To assess how each vantage supports spatial perception, we evaluated SLAM performance (for localization and mapping) and NeRF-based 3D reconstruction (for downstream scene understanding). Head-mounted sensors delivered superior localization accuracy, while cane-mounted views offered broader ground-level coverage and richer environmental reconstructions. A combined (head+cane) configuration consistently outperformed both. These results highlight the complementary strengths of different sensor placements and offer actionable guidance for developing hybrid navigation aids that are perceptive, robust, and user-aligned.

CCS CONCEPTS

• **Human-centered computing → Accessibility technologies**; • **Computing methodologies → 3D imaging**; **Tracking**.

KEYWORDS

visual impairment, orientation and mobility, assistive technology, navigation, computer vision

ACM Reference Format:

Apurv Varshney, Lucas Nadolskis, Tobias Höllerer, and Michael Beyeler. 2025. Beyond Physical Reach: Comparing Head- and Cane-Mounted Cameras for Last-Mile Navigation by Blind Users. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Navigating unfamiliar environments independently remains a significant challenge for blind and low vision (BLV) individuals. While mobility aids such as canes and guide dogs provide essential support, they often fall short in enabling successful "last-mile" navigation: locating a specific entrance, finding an empty seat in a crowded room, or identifying small obstacles in cluttered environments [37, 39]. These tasks demand spatial awareness that extends beyond the physical reach of traditional tools.

Recent advances in wearable computer vision and inertial sensing have opened new possibilities for enhancing navigation [18, 21, 32, 38]. Wearable systems can now interpret surroundings in real time and deliver context-aware feedback. However, a fundamental question remains unresolved: *where* on the body should sensors be placed to best support real-world navigation?

Head-mounted systems, such as smartglasses, align with the user's gaze and are effective for detecting distant or eye-level landmarks [1]. One promising example is Meta's Project Aria, which integrates multiple cameras, IMUs, and cloud-based services for advanced real-time environment mapping (Fig. 1A). Chest- and Waist-mounted systems like the NavBelt [40], Tactile Wayfinder [19], and devices explored by Katzschmann et al. [26] and Wang et al. [44], Commercial examples like Biped [5], offer slightly broader and more stable perspectives, but cover largely overlapping regions. In contrast, cane-mounted configurations provide a radically different view: lower to the ground, naturally swept across space, and tightly coupled with tactile exploration. This vantage is well-suited for detecting curbs, low-lying obstacles, and terrain transitions, but is also susceptible to motion blur and potential SLAM instability.

All three configurations are being actively explored in assistive navigation. Commercial systems like Envision Glasses [12] and OrCam MyEye [35] rely on head-mounted cameras for object recognition. Recent efforts have revived the idea of augmenting canes with cameras and depth sensors [3, 15, 41, 42], though issues with motion artifacts, real-time compute demands, and localization accuracy have limited adoption.

Despite these parallel development tracks, there has been no principled, vantage-focused comparison of sensor placements under consistent conditions. To address this gap, we conducted a two-part investigation combining qualitative and quantitative methods:

- **Survey of experienced BLV cane users.** We collected real-world navigation strategies, pain points, and device preferences. Participants detailed how they tackle complex settings like busy sidewalks, bus stops, and shopping malls. The goal of the survey was to surface unmet needs, overlooked cues, and strong opinions on what makes a navigation aid trustworthy and usable.
- **Controlled data collection in diverse, real-world environments.** Using synchronized Project Aria glasses mounted on the head and cane, we compared sensor vantage with a blind participant across five indoor and outdoor locations. As illustrated in Fig. 1, a head-mounted camera (Fig. 1C) may miss critical wayfinding cues (like a doorframe) captured only in the cane-mounted stream (Fig. 1D). This setup isolates the effect of sensor placement from user variation, enabling a direct comparison of each configuration's impact on Simultaneous Localization and Mapping (SLAM) quality for real-time localization and sparse mapping (Fig. 1B), and dense 3D scene reconstruction using Neural Radiance Fields (NeRF) useful for fine-grained path planning (Fig. 1E).

This paper thus provides the first comprehensive vantage-based comparison in this domain. Our results reveal that cane motion, rather than the low vantage itself, is the primary source of SLAM degradation, yet that vantage captures spatial cues the head-mounted view lacks, thereby leading to richer 3D reconstructions. In sum, our contributions are:

- **Qualitative insights** from expert cane users, identifying design requirements, challenges, and strong preferences that highlight how new systems should integrate with (rather than replace) existing tools and strategies.
- **Quantitative evaluation** of head- and cane-mounted sensors under identical conditions, showing how vantage affects SLAM-based localization and downstream tasks like NeRF-based 3D scene reconstruction, and demonstrating a strong benefit for combined perspectives.
- **Design implications** for wearable navigation aids that extend the cane's spatial reach, grounded in both technical performance and user-driven needs.

By bridging user insights with consistent, vantage-focused performance data, this study reframes sensor placement as both a technical constraint and an opportunity. Our findings point toward hybrid solutions that harness the stability of head-mounted views and the spatial richness of cane-mounted input, moving closer to assistive technologies that are both robust and deeply aligned with the practices of blind travelers.

2 RELATED WORK

This work builds on a rich history of research in assistive navigation for BLV individuals, spanning tactile tools, wearable systems, and multimodal sensor platforms. We situate our study within five key areas: navigation challenges, wearable systems, cane-mounted sensors, sensor processing, and user-centered design.

2.1 Navigation Challenges for BLV Individuals

Navigating environments, particularly unfamiliar ones, remains a major barrier to independence for BLV individuals [10, 34, 38]. Traditional aids like white canes and guide dogs provide essential proximal feedback but often fail to convey broader environmental context, such as layout, landmark location, or dynamic obstacles [7, 27]. These challenges are especially acute during "last-meter" tasks like identifying a building entrance, which fall beyond the spatial range of existing tools and often occur in noisy, cluttered spaces [22, 39].

Unfamiliar indoor settings pose additional hurdles due to inconsistent cues and the scarcity of accessible maps [24, 34]. Outdoor navigation, while aided by GPS, suffers in urban environments and lacks the precision needed for goal-directed mobility [2, 10, 22]. These deficits create unmet needs not just for obstacle avoidance but for goal-oriented navigation, spatial understanding, and exploration support [7, 8, 23, 30, 32].

Effective systems must balance rich feedback with low cognitive load [7, 36], and align with user priorities, which may not match technological focus areas [13]. This underscores the need for tools that augment existing strategies without disrupting user autonomy.

2.2 Wearable Navigation Systems

Wearable systems integrate visual and depth sensors into head-, chest-, or belt-mounted configurations to enhance environmental awareness [9, 25].

2.2.1 Body-Mounted Configurations. Head-mounted devices, such as Envision Glasses and OrCam MyEye [12, 35], are effective for tasks like text recognition or detecting distant landmarks [8, 25], and often align with user preferences [13]. However, their limited field-of-view and gaze-level perspective may miss ground-level hazards crucial for safe ambulation [1, 29].

Chest-mounted [5, 44] or belt-mounted systems [19, 26, 40] offer broader, more stable views or deliver directional cues via haptic feedback. Yet, these setups can suffer from occlusion, feedback overload, or poor detection of obstacles at non-standard heights [26, 36]. Across all configurations, achieving situational awareness and real-time responsiveness remains a challenge.

2.2.2 Cane-Mounted Configurations. The white cane remains the gold standard for mobility, prompting efforts to augment it with sensors for richer feedback [27, 41, 43]. Smart canes have incorporated ultrasonic sensors, cameras, LiDAR, and even physiological monitors [15, 45], often paired with haptic or auditory alerts [20, 36]. The cane's proximity to the ground makes it uniquely suited to detecting surface changes and low obstacles.

However, technical hurdles have hindered adoption, including power constraints, form factor limitations, and unstable camera motion due to cane sweeping [3, 15, 31]. Real-time vision-based processing from the cane tip remains particularly difficult due to motion blur and inconsistent inertial measurement unit (IMU) readings. Prior work has highlighted the cane's potential, but few studies have rigorously compared cane-mounted cameras to head-mounted alternatives for tasks like localization and scene reconstruction. Our work addresses this gap through controlled, head-to-head evaluation.

2.3 Sensor Processing and Data Quality

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 [26, 42], while SLAM systems fuse vision and inertial data to track pose and build maps in real time [3]. Devices like Meta's Project Aria offer tightly synchronized multi-sensor data streams optimized for such use cases [11] (Fig. 1A).

Sensor placement strongly influences data quality. Head-mounted systems, while stable, may miss ground-level cues [1, 29]. Cane-mounted systems provide a complementary viewpoint but face SLAM degradation due to dynamic motion [3, 31]. New NeRF-based scene reconstruction like EgoLifter [16] promise to work on egocentric data and in dynamic environments. But they add further constraints: accurate input poses and viewpoint diversity are key to dense, photorealistic models. This makes evaluating how placement affects downstream reconstruction tasks essential for robust system design.

2.4 User-Centered Design in Navigation Tools

Ultimately, successful adoption depends on usability and integration with existing mobility strategies. Prior work underscores how new technologies can bridge critical gaps in blind navigation, yet also face significant obstacles to real-world uptake [6, 14, 17, 25, 33]. For instance, overly complex or intrusive solutions may fail to align with established cane skills, creating more friction than benefit. Co-design approaches (i.e., engaging BLV individuals early and often) help identify these real-world constraints, reduce cognitive burden, and tailor systems to users' lived experience [4, 13, 28]. Studies further emphasize the need for concise feedback, seamless integration with the cane, and support for diverse needs like exploration, not just obstacle avoidance [18, 23, 30, 34].

While surveys show a preference for wearable systems that are hands-free and minimally intrusive [13], few existing technologies fully meet these criteria. This gap between technical innovation and practical usability underscores the importance of grounding system development in genuine user input, ensuring that the resulting tools add value without undermining the autonomy or expertise of blind travelers.

3 FORMATIVE STUDY: USER NEEDS & NAVIGATIONAL STRATEGIES

Designing effective navigation aids requires grounding development in the lived experiences of end-users and domain experts. To this end, we conducted a formative study that combined focus groups and surveys with blind individuals and a certified orientation and mobility (O&M) instructor. This dual-pronged approach allowed us to explore everyday navigation strategies, surface pain points in last-mile mobility, and establish guiding principles for user-centered design of wearable navigation technologies.

3.1 Focus Group Insights

We began with small-scale focus groups involving a blind participant, an O&M instructor, and HCI researchers. These discussions emphasized co-design and ethics, encouraging a free exchange of perspectives between stakeholders. The blind participant described real-world challenges, such as frustration with smartcanes that failed to outperform a standard white cane. The O&M instructor contributed insights into common strategies like shorelining and tactile cue detection, while researchers helped bridge these experiences with technical considerations.

Themes from these sessions highlighted the irreplaceability of the cane's tactile feedback and the need for technologies that *augment*, rather than replace, existing tools. Participants emphasized trust, real-time environmental awareness, and the importance of cognitive load in technology adoption. These insights directly informed the design of the subsequent survey.

3.2 Survey Methods

Building on the focus group findings, we developed a scenario-based survey to explore how blind cane users approach complex navigation challenges.

The survey featured three real-world scenarios:

(1) navigating to a restaurant after being dropped off,
(2) navigating a crowded bus stop, and
(3) exploring a shopping mall.

Each scenario included five targeted questions probing orientation, obstacle avoidance, environmental cues, and challenges.

Participants were recruited via community networks and screened for eligibility: all self-identified as completely blind, were 18 years or older, had received O&M training, and were proficient in cane use. Ten individuals completed the survey (demographics in Table 1). Responses included Likert-scale items and open-ended text fields. All values were self-reported. This study was deemed exempt by our Institutional Review Board.

3.3 Scenario-Based Results

3.3.1 Scenario 1: Navigating to a Restaurant. Participants consistently reported relying on auditory cues (e.g., traffic flow, vehicle sounds) and tactile input (e.g., cane detecting curbs or pavement changes) to establish orientation. Sound reflections and environmental structure played key roles in determining sidewalk versus street boundaries.

For obstacle avoidance, participants emphasized consistent cane sweeps, spatial memory, and use of auditory cues. Crossing decisions were typically made using parallel traffic sounds or tactile paving, with soundless intersections described as particularly challenging. To locate the restaurant, participants scanned for open doors, increased foot traffic, or ambient sounds from inside. Frustrations included poorly marked entrances, unexpected street furniture, and signalless crossings. As P10 noted, "There are so many things to trip on before I even get to the door."

3.3.2 Scenario 2: Navigating a Crowded Bus Stop. Identifying the bus stop often involved listening for bus engines or pedestrian clusters. Participants described using shoreline techniques and tactile cues (e.g., textured sidewalk changes) to approach and stay oriented. Positioning themselves safely while waiting varied: some used poles or benches as anchors, others stood close to the curb for easier bus access. Recognizing the correct bus was a common concern, with participants relying on announcements, route calls, or driver confirmation. Major challenges included crowd noise, inconsistent announcements, and obstacles like temporary signage or planters. As P4 remarked, "I've been left behind because I didn't realize the bus had pulled up behind another."

3.3.3 Scenario 3: Navigating a Shopping Mall. To get oriented in a mall, participants used sound reflections, airflow near

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 Orientation and Mobility (O&M) training, and regularly use a white cane for independent navigation. All values are self-reported.

entrances, and cane-detected flooring changes. Maintaining direction involved tracking unique sounds (e.g., music, water features), tactile landmarks (e.g., carpet transitions), and scanning for entrances. Distinguishing store entries from open spaces relied on changes in soundscape or structural cues like automatic doors. Elevators/escalators were found by following foot traffic or auditory beacons. Escalator boarding was described as a high-stress task. Participants identified disorienting layouts, ambient noise, and a lack of consistent cues as key challenges. As P7 shared, "It's like a maze with no signs I can read."

3.4 Cross-Cutting Navigation Challenges

In addition to scenario-specific strategies, participants identified recurring challenges that cut across environments and situations. These themes reflect both physical and social barriers to navigation, as well as personal adaptations developed through experience.

3.4.1 Device Attitudes and Preferences. Participants also rated their agreement with a series of statements about assistive technology preferences (Table 2). Responses suggest a high level of comfort with both head- and cane-mounted cameras. However, participants expressed a clear preference for devices that integrate with, rather than replace, the white cane. Wearability and hands-free operation were highly valued, as was the ability to maintain control using familiar O&M techniques.

3.4.2 Environmental and Structural Barriers. Participants frequently reported difficulties in visually uniform or unfamiliar environments with limited tactile or auditory landmarks. Malls and parking lots were repeatedly cited as disorienting due to wide-open layouts and inconsistent cues. Identifying specific entrances or amenities was a persistent challenge.

3.4.3 Sensory Overload and Interference. Crowded or noisy environments posed major difficulties. Competing sounds interfered with participants' ability to detect helpful cues or maintain spatial awareness. Bus stops, malls, and city streets were all flagged as problematic in this regard.

3.4.4 Social and Safety Concerns. Several participants described negative interactions with sighted individuals offering unsolicited or confusing help. These interactions disrupted orientation and led to a lack of confidence in unfamiliar settings. Feelings of vulnerability in crowded public spaces were also common.

3.4.5 Personalization and Cognitive Strategies. Many participants reported adapting their navigation techniques based on environment familiarity, including switching cane types, using apps like Soundscape, and employing time-based estimations. Strategies like pre-planning, repetition, and active spatial mapping were central to their independence.

3.5 Key Findings and Design Implications

Our survey revealed critical insights about last-mile navigation strategies, preferences, and persistent challenges for blind users:

3.5.1 Navigation strategies depend heavily on multi-sensory integration. Participants consistently used a sophisticated blend of auditory, tactile, and occasionally olfactory cues. Auditory information (e.g., traffic sounds, bus announcements, store music) frequently guided orientation and destination confirmation. Tactile feedback, particularly through cane exploration (e.g., identifying curbs, ground textures, walls), was crucial in maintaining direction, detecting obstacles, and confirming precise locations such as bus stops or store entrances.

3.5.2 Obstacle avoidance is essential but secondary to destination-focused navigation. Users prioritized confidently finding their destination rather than merely avoiding hazards. Techniques like shorelining were often strategically employed to support broader spatial orientation rather than exclusively for obstacle detection.

3.5.3 Each scenario posed unique but overlapping challenges. Common difficulties included unclear or non-distinctive landmarks, unpredictable environmental layouts, dynamic obstacles, and sensory overload in crowded or noisy settings. Participants also frequently highlighted specific frustrations,

Table 2: Participant responses to ten navigation and device preference questions grouped under four categories: Navigation, 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). While formal statistical analysis was not conducted, responses suggest general comfort with head- and cane-mounted devices, and a preference for navigation aids that integrate with the existing cane over additional wearable devices.

Category	Question	Mean	Std. Error
Navigation	Navigation is primarily about finding my destination.	4.125	1.246
	Navigation challenges limit my independence.	3.875	1.126
	Navigation is primarily about avoiding obstacles.	2.375	1.187
Trust	Safety is the most important factor in navigation.	3.500	1.195
	I need assistive technology to be highly reliable before I can trust it.	3.375	1.506
Placement	I would feel comfortable using a navigation device worn on my chest.	4.000	1.069
	I would feel comfortable using a navigation device worn on my head.	3.375	1.061
	I would feel comfortable using a navigation device worn on my cane.	3.125	1.553
Form	I prefer to use a wearable device that provides hands-free navigation support.	3.750	1.389
	I prefer using a navigation aid that integrates with my existing cane rather than a wearable.	2.250	0.886

such as ambiguous storefront entrances, confusing intersections, difficulty hearing critical auditory cues, and unexpected obstacles in busy, dynamic environments.

3.5.4 Technology must complement rather than replace traditional aids. Participants were generally open to new technologies, but emphasized the cane's irreplaceable tactile feedback and reliability. They expressed a strong preference for wearable, hands-free solutions, with head-mounted devices considered most comfortable. However, they also showed skepticism about devices requiring additional wearable components, stressing the importance of seamless integration with established navigation techniques.

4 SYSTEM DESIGN AND EVALUATION

The formative study revealed key insights about the navigation needs and technology preferences of blind cane users. Participants emphasized the irreplaceable role of the cane, the importance of multi-sensory strategies, and the desire for wearable aids that extend (rather than disrupt) their established O&M techniques. These findings motivated a deeper investigation into the technical trade-offs of sensor placement for camera-based navigation systems.

In this section, we present a structured, quantitative evaluation of two promising configurations: head-mounted and cane-mounted cameras. While head-mounted cameras benefit from stable positioning and alignment with user gaze, cane-mounted configurations offer a unique, ground-level viewpoint with broader spatial coverage due to natural sweeping motion. However, prior work and participant feedback raised concerns about motion-induced artifacts and instability, particularly in real-time systems. To empirically assess these trade-offs, we collected synchronized sensor data from both configurations as a blind co-author navigated five real-world environments using standard O&M techniques.

4.1 Methods

4.1.1 Co-Design Process. Grounded in user-centered design principles [4, 28], our data collection methodology was shaped through a co-design process involving a co-author (18+) who is blind, proficient in independent cane navigation, and has extensive experience with various navigation aids. Through iterative planning and feedback sessions, we collaboratively selected study locations to ensure both environmental diversity and real-world relevance, identified key landmarks aligned with common O&M strategies, and co-defined the loop-based navigation task. This close collaboration was essential for ensuring ecological validity, grounding the experimental design in the lived experiences of proficient cane users, and confirming the practical value and appropriateness of the selected tasks and locations.

4.1.2 Locations. The study was conducted across five environments (two outdoor and three indoor) chosen to reflect a diverse range of last-mile navigation challenges (Table 3). Locations included both structured settings (e.g., hallways, doorways) and unstructured or cluttered spaces (e.g., open plazas, furniture clusters). Each route was populated with common wayfinding elements like curbs, planters, floor transitions, and ambient pedestrian traffic. This diversity enabled us to evaluate performance across a range of spatial layouts, obstacle types, and sensory conditions, including variable lighting and acoustic environments.

4.1.3 Data Collection Procedure. At each location, the blind co-author first completed a brief familiarization phase with an experimenter. Together, they walked the predefined looped route (beginning and ending at the same location) while the sighted experimenter described and pointed out key intermediate landmarks (e.g., textured flooring, architectural features, obstacles) that had been selected during the co-design process. Example landmarks are shown in Fig. 2. This process ensured the blind co-author had a consistent frame of

Scenario	Distance	Landmarks
Outdoor1	30m	Tactile warning pad, pillars, walls
Outdoor2	25m	Sidewalk-to-plant area border, manhole cover
Indoor1	35m	Floor mat at elevator, doors, railings, walls
Indoor2	70m	Garbage cans, walls, floor mats at exit door
Indoor3	45m	Garbage cans, hard-to-carpet floor transition, chairs, floor outlets, partitions

Table 3: Overview of navigation scenarios used in the study. Each scenario corresponds to a specific physical location with a predefined walking path (distance indicates end-to-end route length) and notable environmental landmarks.

Figure 2: Examples of key visual and tactile landmarks encountered during last-mile navigation tasks, including floor transitions, textured surfaces, doorways, tactile paving, and environmental boundaries. While this figure emphasizes features detectable via vision or touch, participants also described using auditory and olfactory cues.

reference for each environment before independent navigation began.

During the main trial, the co-author navigated the route independently using standard O&M techniques. Two identical Project Aria smartglasses devices were used: one worn on the head and one mounted securely to the co-author's mobility cane. Each device featured an RGB camera, two monochrome scene cameras, dual IMUs, magnetometer, barometer, GPS, and audio sensors. This setup ensured synchronized and comparable sensor streams from both perspectives. Navigation was entirely self-directed; the co-author determined pace, route-finding strategies, and landmark identification using auditory, tactile, and proprioceptive cues. Minimal verbal assistance was provided, and only when necessary for safety.

Each environment presented realistic last-mile challenges, including static obstacles (e.g., trash cans, furniture), dynamic elements (e.g., pedestrians), and variable lighting. This protocol enabled direct comparison of head- and cane-mounted perspectives under consistent, real-world navigation conditions.

4.2 Results

4.2.1 SLAM Performance. Effective path planning requires reliable spatial awareness, which in assistive systems is typically provided by SLAM. SLAM enables a system to generate a map of its surroundings (via 3D point estimation) while simultaneously tracking its own position and orientation (pose estimation) within that map. This functionality is especially critical in unfamiliar environments where no prior map is available.

To evaluate how sensor placement impacts SLAM quality, we compared outputs from Meta's MultiSLAM framework using their Machine Perception Services (MPS) using synchronized data from head-mounted and cane-mounted Project Aria devices.

We evaluated SLAM performance across the five real-world environments (examples in Fig. 3) using two key metrics derived from the system's outputs: the proportion of "accurate" 3D map points and the proportion of "accurate" camera pose estimations, as defined below. First, raw 3D point cloud data generated by SLAM often includes points with high positional uncertainty or geometric inaccuracies. To assess map quality, we filtered these raw points based on nominal quality thresholds (i.e., a maximum inv_dist_std of 0.005 and a maximum dist_std of 0.01) provided by the SLAM system. The proportion of points remaining after filtering constitutes our first metric, termed **high-quality 3D points**. Second, the SLAM system assigns a quality score (ranging from 0 to 1) to each estimated camera pose, indicating confidence in the localization result. To evaluate the accuracy and stability of the localization, we applied a strict criterion, considering only poses assigned the maximum possible quality score. Therefore, we filtered out any pose estimate with a quality score below 1.0. The proportion of poses meeting this high-confidence threshold constitutes our second metric, referred to as **accurate pose estimations**. When interpreting the SLAM results, it is important to differentiate between raw counts and performance ratios. For pose estimations, a higher total count, particularly observed for the cane configuration, likely reflects the more dynamic trajectory sampled rather than indicating superior SLAM performance itself; therefore, the ratio of accurate poses serves as the primary indicator of localization stability in our analysis. In contrast, for the 3D map points, while the ratio of high-quality points indicates the reliability of the generated map features, the total number of points generated (often higher for the cane) may offer insights into the extent of environmental coverage achieved due to the sensor's motion. Consequently, both the total point count (suggesting potential coverage) and the accuracy ratio (indicating reliability) can be valuable considerations when evaluating the overall mapping capabilities of each configuration for different downstream tasks.

Location	High-quality 3D points ↑		Accurate Pose Estimations ↑	
	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 4: Comparison of SLAM performance across five navigation environments, showing the percentage and raw counts of high-quality 3D points and accurate pose estimates (↑ higher is better), reported as a fraction of total 3D points or pose estimates. High-quality points are those with low positional uncertainty, and accurate poses are those assigned the maximum confidence score (1.0) by the SLAM system. Bolded values indicate the best-performing configuration within each environment.

Location	Performance Metric	Head-Mounted Camera	Cane-Mounted Camera	Cane-Mounted (No Sweep)
Outdoor	High-quality 3D points ↑	**37.6%** (**50,531** / 135,285)	26.6% (40,846 / **153,566**)	30.1% (44,744 / 148,832)
	Accurate pose estimations ↑	**99.6%** (**13,315** / 13,366)	39.5% (6,868 / **17,408**)	83.2% (11,095 / 13,338)
Indoor	High-quality 3D points ↑	38.0% (82,448 / 217,227)	34.2% (101,556 / **296,521**)	**39.2%** (**115,779** / 295,671)
	Accurate pose estimations ↑	**99.8%** (25,561 / 25,613)	92.6% (**27,437** / **29,628**)	98.5% (26,167 / 26,575)

Table 5: 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 percentage and raw counts of high-quality 3D points and accurate pose estimates (↑ higher is better), reported as a fraction of total points or estimates. High-quality points are those with low positional uncertainty; accurate poses are those assigned the maximum confidence score (1.0) by the SLAM system. Bolded percentages indicate the best-performing configuration for each metric within each environment.

As summarized in Table 4, the head-mounted configuration consistently outperformed the cane-mounted one in both metrics. Pose estimation from the head-mounted camera was highly stable across all locations, with accuracy exceeding 98%. In contrast, pose accuracy for the cane-mounted camera was notably lower and more variable, dropping to 55–58% in outdoor settings. These deficits likely stem from the dynamic, sweeping motion of the cane, which introduces erratic trajectories and motion blur—conditions that challenge SLAM algorithms.

While the cane-mounted configuration often generated more raw 3D points—owing to its broader and more dynamic field of view—these points were less likely to be accurate. The head-mounted camera produced a higher proportion of high-quality 3D points, reflecting more consistent visual features and motion patterns better suited for SLAM.

In short, although the cane-mounted configuration offers greater environmental coverage, the head-mounted configuration provides more reliable localization and mapping—suggesting a fundamental trade-off between spatial breadth and signal stability.

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

Table 6: 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 significantly highest PSNR for each row, highlighting the improved performance of Head+Cane configuration in all scenarios.

4.2.2 Analyzing Motion-Induced Degradation. To isolate the factors contributing to degraded SLAM performance in the cane-mounted condition, we conducted an ablation study using three sensor configurations: (1) head-mounted, (2) cane-mounted (typical dynamic sweep), and (3) cane-mounted (stationary, "No Sweep"). The stationary cane-mounted setup involved holding the cane still in a natural position, eliminating motion-induced blur while preserving the altered point of view. Results showed a clear improvement: pose estimation accuracy rose from 39.5% to 83.2% in outdoor

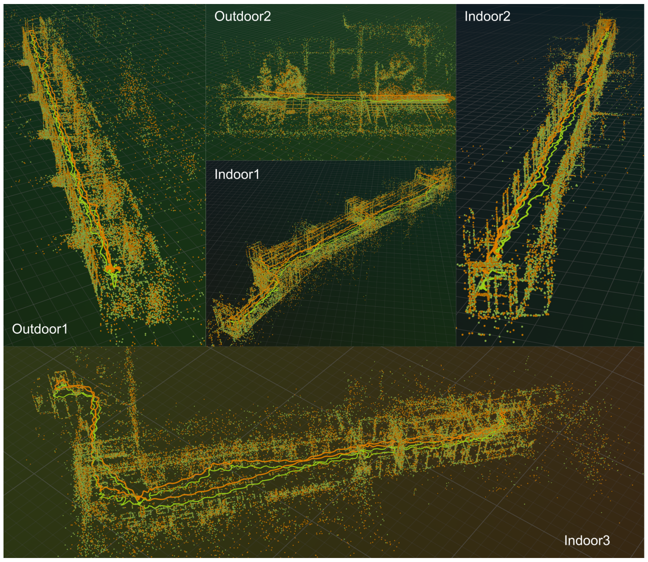

Figure 3: Semi-dense 3D point clouds and global closed-loop trajectories captured across five locations (Outdoor1–2, Indoor1–3). Point clouds were generated using Meta's MultiSLAM. Trajectories are overlaid in orange (head-mounted camera) and green (cane-mounted camera), illustrating differences in environmental coverage and motion patterns between sensor placements.

settings—substantially narrowing the gap with the head-mounted configuration (99.6%, Table 5).

These findings suggest that the primary cause of SLAM degradation was not the viewpoint of the cane *per se*, but the erratic motion patterns introduced by sweeping. Rapid, irregular motion likely overwhelmed the SLAM algorithm's ability to track consistent visual features and contributed to poor pose convergence, particularly in environments with inconsistent lighting or sparse structure.

This reinforces the idea that cane-mounted cameras offer valuable spatial information but require stabilization or filtering techniques to become viable for real-time mapping.

4.2.3 NeRF-Based 3D Scene Reconstruction. While SLAM provides sparse maps and pose estimates for basic navigation, richer 3D representations can better support downstream tasks like spatial reasoning and fine-grained path planning. To explore this potential, we used Neural Radiance Fields (NeRF) to generate dense, photorealistic 3D reconstructions from each sensor configuration. Specifically, we employed the EgoLifter [16] model, which is optimized for egocentric video with complex motion and dynamic elements typical of real-world navigation.

Reconstructions were generated using three configurations: `head`-mounted, `cane`-mounted, and `head+cane` (synchronized data from both). Reconstruction quality was quantified using Peak Signal-to-Noise Ratio (PSNR), a standard metric for evaluating visual fidelity in novel view synthesis.

To statistically evaluate the effects of configuration and environment, we fit a hierarchical mixed-effects model with `configuration` as a fixed effect and `scenario` (indoor vs. outdoor) and `location` (nested within scenario) as random effects. This model structure captures variability at multiple levels of the environment hierarchy and isolates the effect of sensor placement.

The results are shown in Table 6. The analysis revealed a significant fixed effect of configuration on PSNR. The `head+cane` configuration yielded the highest PSNR overall, significantly outperforming both `cane` (Estimate = 1.204, SE = 0.193, $p < .001$) and `head` (Estimate = 1.811, SE = 0.194, $p < .001$). Interestingly, `cane` outperformed `head` (Estimate = 0.607, SE = 0.223, $p = .007$), suggesting that while less stable, the cane-mounted viewpoint may capture more spatially informative content. Random effects confirmed substantial variability across scenarios ($\sigma^2 = 7.346$) and individual locations within scenarios ($\sigma^2 = 5.906$).

PSNR was consistently lower in outdoor environments, reflecting greater environmental complexity, inconsistent lighting, and visual sparsity. Model fit improved over simpler baselines (log-likelihood = -2934.81 vs. -3186.39), and residual variance was substantially reduced ($\sigma^2 = 7.51$ vs. 11.57), indicating that the hierarchical structure captured meaningful variability.

Qualitative results in Figure 4 further illustrate these findings: reconstructions from the head+cane configuration produced more coherent geometry and visual detail, especially in cluttered indoor scenes and visually sparse outdoor spaces.

4.3 Summary of Quantitative Findings

Our results revealed a core trade-off between sensor stability and spatial coverage. Head-mounted cameras consistently produced accurate pose estimates and high-quality 3D point maps across all environments, confirming their strength for SLAM-based localization. In contrast, cane-mounted cameras showed lower SLAM performance due to motion artifacts, but captured complementary spatial information from a ground-level perspective.

This advantage became more pronounced in downstream NeRF-based scene reconstructions. Despite its SLAM limitations, the cane-mounted configuration outperformed head-mounted input in reconstruction quality. The combined head+cane configuration yielded the highest PSNR overall. A hierarchical mixed-effects model confirmed that sensor placement had a significant effect on reconstruction quality, with additional variability explained by scenario (indoor vs. outdoor) and specific location.

Together, these findings suggested that while head-mounted sensors are better suited for stable localization, cane-mounted sensors contribute valuable spatial cues. Hybrid systems that integrate both perspectives may offer the best of both by balancing localization accuracy with broader environmental understanding for assistive navigation tasks.

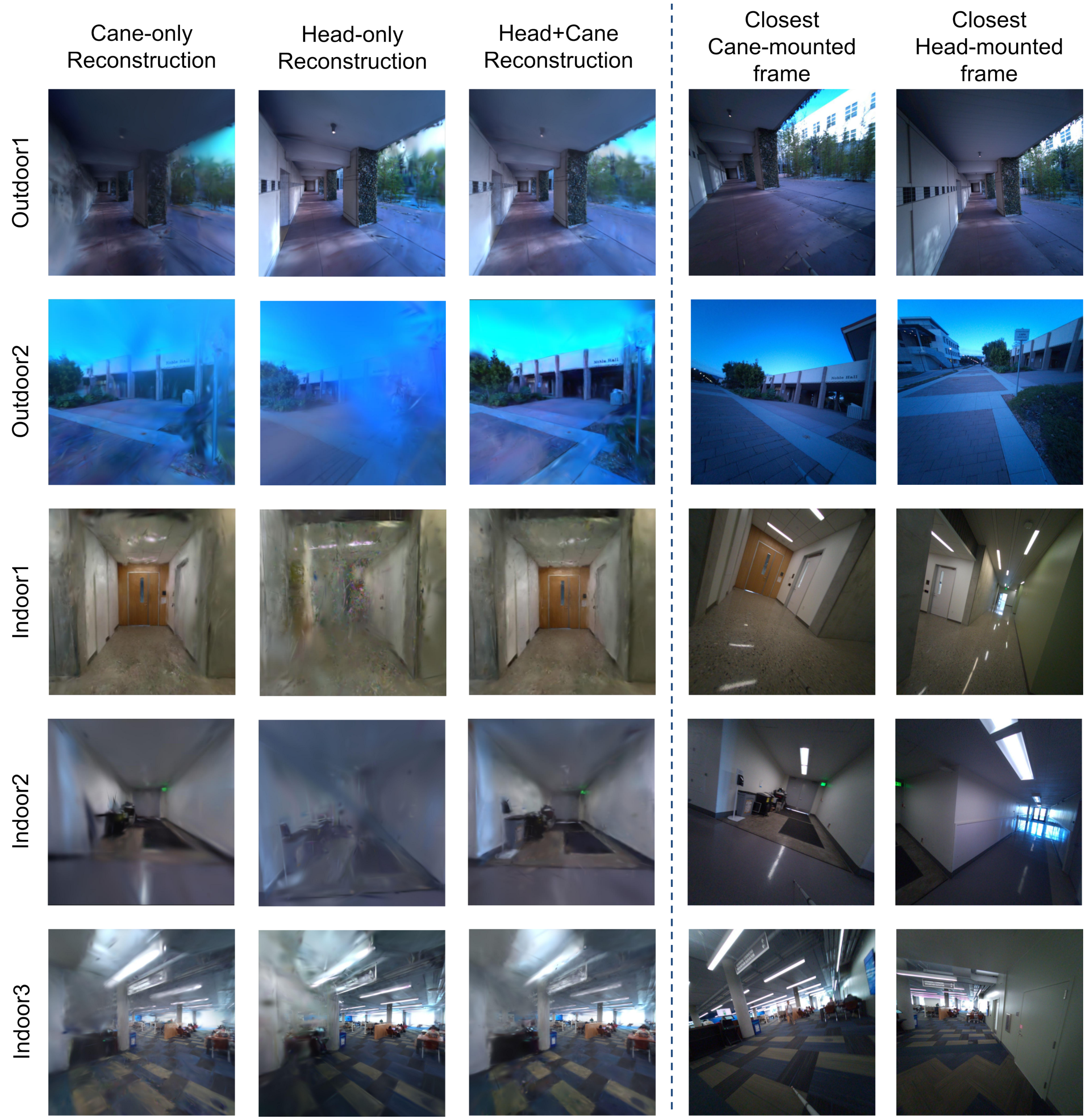

Figure 4: NeRF-based reconstructions using three input configurations—`head-only`, `cane-only`, and `head+cane`—across multiple locations. Each row shows one scene: the first three columns depict synthesized views generated from each configuration. The final two columns show the closest ground-truth frames captured by the head-mounted and cane-mounted cameras, respectively, for visual reference. The combined configuration consistently produces sharper, more complete reconstructions, capturing both ground-level geometry and high-frequency environmental details.

5 DISCUSSION

This work offers a fresh look at sensor placement in wearable navigation aids for blind cane users. Through a qualitative study grounded in the lived experiences of cane users and O&M instructors, we identified key design values: augmenting (not replacing) existing strategies, integrating seamlessly with mobility tools, and minimizing cognitive load. We then conducted a first-of-its-kind, head-to-head comparison of head- and cane-mounted cameras under consistent experimental conditions, isolating sensor vantage from user variation.

Our findings reveal that vantage alone is not inherently detrimental to SLAM. Rather, the dynamic motion inherent in cane sweeping drives SLAM degradation. Yet, that same vantage yields more spatially informative views for 3D reconstruction and complements the cane's tactile function. Below, we discuss how these results broaden the design space and highlight potential pathways for future systems.

5.1 Sensor Placement Shapes Perception and Performance

Our results confirm that sensor placement is not a one-size-fits-all decision, but that it must be tailored to the intended task and environment. All three sensor placements (head, chest, cane) were rated as somewhat acceptable by our survey participants (Table 2), but preferences appeared shaped by prior exposure to commercial technologies. Head-mounted devices (e.g., smartglasses) were familiar and widely discussed, likely contributing to their higher comfort ratings. In contrast, cane-mounted systems have historically underperformed [27, 36, 41], and that skepticism was reflected in the survey responses.

A key takeaway from our quantitative evaluation is the trade-off between localization stability and richer spatial coverage. Head-mounted cameras consistently delivered superior SLAM estimates, especially in structured indoor environments (Table 4), aligning with prior observations that head-level perspectives are optimal for stable feature tracking [31]. However, cane-mounted views revealed valuable low-level scene details often missed by head-mounted sensors [1, 29], which is particularly evident in our NeRF-based reconstructions (Fig. 4).

This vantage-based benefit may have been obscured in prior studies by motion artifacts from cane sweeping. Our ablation showed that motion, not viewpoint, is the main culprit behind degraded SLAM performance (Table 5). Holding the cane still improved pose estimation substantially. Thus, while dynamic cane use injects noise, it also yields unique angles on the environment. By explicitly comparing head and cane vantage in the same loops, we demonstrate that cane-based data can be highly advantageous when processed effectively.

However, technical challenges remain. The instability introduced by natural cane sweeping led to degraded SLAM performance, particularly in outdoor environments with fewer visual anchors (Table 4). Addressing these issues will require robust motion compensation techniques and algorithms that can parse useful signal from inherently noisy input. Rather than discarding cane-mounted vision, future systems should invest in better understanding and adapting to its unique dynamics.

5.2 Technology Should Extend or Complement the Cane

Our qualitative study and survey data confirm prior work [37–39] that obstacle avoidance alone is insufficient (Table 2); participants emphasized goal-focused navigation and the irreplacable role of cane feedback for near-field obstacle detection and tactile exploration. Obstacle detection is already handled effectively by the cane [46], whose tactile feedback remains essential and difficult to replicate through other means. While historically underperforming cane-based sensors may have tempered expectations, participants nonetheless welcomed solutions that enhance their existing techniques.

Survey participants expressed a clear preference for hands-free systems that do not supersede the cane's function, but rather extend it. This openness to cane-mounted sensing (despite past disappointments) underscores a crucial design principle: technologies that align with the cane's established utility and user autonomy are more likely to be trusted and adopted. Conversely, tools that attempt to replicate or override the cane's tactile feedback risk user skepticism and high cognitive load.

5.3 Toward Adaptive, Multisensor Navigation Systems

Our NeRF analysis showed that combining head- and cane-mounted inputs produced the highest-fidelity reconstructions (Table 6), reinforcing the notion that each vantage offers distinct benefits. Future systems could capitalize on this complementary coverage through adaptive multisensor architectures. Rather than choosing between head or cane, advanced solutions might selectively fuse data, thereby enabling stable localization from the head and rich scene detail from the cane.

These findings echo broader user strategies observed in our qualitative study, where blind cane users draw on multiple sensory channels (auditory, tactile, and spatial memory) to form holistic situational awareness. Similarly, sensor fusion could balance the strengths of each vantage in real time, or switch dynamically based on environmental cues (e.g., prioritizing head-mounted input for straightforward corridor navigation, then augmenting with cane-mounted data when scanning for near-ground features).

Achieving this vision demands algorithmic innovations in motion compensation, real-time sensor fusion, and context-aware feedback. Equally important is co-design with the blind community to ensure that these systems reduce cognitive burden and align with established O&M practices, rather than imposing new workflows.

5.4 Limitations and Future Work

This study has several limitations that must be acknowledged.

First, our evaluation involved a single blind co-author who participated in the design, data collection, and interpretation process. This collaborative approach reflects established co-design methodologies in HCI [4, 28] and was chosen to ensure ecological validity and alignment with real-world navigation strategies. Working closely with an experienced cane user allowed us to explore complex scenarios, iterate on our protocol, and focus our evaluation on isolating the effects of sensor placement—something difficult to achieve with a larger, short-term participant pool. However, this approach necessarily limits the generalizability of our findings. Future studies should expand to include a broader range of blind and low vision users to assess usability, comfort, and performance variability at scale.

Second, our study focused on short, relatively simple routes in predefined indoor and outdoor environments. While these scenarios reflect common last-mile challenges, they do not fully capture the complexity of open-ended urban navigation or long-distance wayfinding. Longer-term, in-the-wild deployments will be critical for understanding how wearable systems perform in messy, real-world conditions.

Third, while our analysis focused on vision-based SLAM and 3D reconstruction, we did not evaluate real-time system feedback, user experience, or long-term adoption. And although we identified promising trends (particularly the benefits of combining perspectives) our work stops short of training or deploying adaptive multisensor systems.

Looking forward, a large-scale, multi-user dataset collected across real-world conditions could transform the field. Such data would enable the training of more robust perception algorithms tailored to blind users' movement patterns and sensor perspectives, improving SLAM, spatial awareness, and ultimately trust in AI-driven navigation systems.

6 CONCLUSION

Our work reshapes the conversation around sensor placement in wearable navigation aids by demonstrating that vantage point is as pivotal as algorithmic design. Our head-to-head comparison of head- and cane-mounted cameras (validated under consistent, real-world conditions) revealed that cane-mounted views, frequently discounted due to motion instability, can capture indispensable ground-level details for last-mile navigation. In fact, merging head and cane perspectives consistently produced the most robust reconstructions, underscoring the potential of hybrid sensor architectures.

Beyond establishing a technical benchmark, our study integrates user-centered insights to highlight a deeper takeaway: practical solutions must respect the cane's established role while extending its reach. Co-design with blind travelers, rather than imposing top-down assumptions, enables technologies that are not merely functional, but truly empowering. By balancing sensor stability with rich environmental coverage, and by aligning with existing O&M techniques, future navigation systems can deliver both robust localization and meaningful real-world adoption.

ACKNOWLEDGMENTS

Brianna Pettit for giving insights into O&M instruction and our undergrad Research Assistants for helping fabricate cane mount for the glasses.

REFERENCES

[1] Kevin Wayne Arthur. 2000. *Effects of field of view on performance with head-mounted displays.* phd. The University of North Carolina at Chapel Hill. AAI9968542 ISBN-10: 0599733721.

[2] Mauro Avila and Limin Zeng. 2017. A Survey of Outdoor Travel for Visually Impaired People Who Live in Latin-American Region. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments* (Island of Rhodes, Greece) *(PETRA '17)*. Association for Computing Machinery, New York, NY, USA, 9–12. doi:10.1145/3056540.3064953

[3] Marziyeh Bamdad, Davide Scaramuzza, and Alireza Darvishy. 2024. SLAM for Visually Impaired People: A Survey. *IEEE Access* 12 (2024), 130165–130211. doi:10.1109/access.2024.3454571

[4] Cynthia L. Bennett and Daniela K. Rosner. 2019. The Promise of Empathy: Design, Disability, and Knowing the "Other". In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300528

[5] Biped.ai. [n. d.]. NOA. https://biped.ai/.

[6] Lucie Brunet, Françoise Darses, and Malika Auvray. 2018. Strategies and needs of blind pedestrians during urban navigation. *Le travail humain* 81 (June 2018), 141. doi:10.3917/th.812.0141

[7] Shaojun Cai, Ashwin Ram, Zhengtai Gou, Mohd Alqama Wasim Shaikh, Yu-An Chen, Yingjia Wan, Kotaro Hara, Shengdong Zhao, and David Hsu. 2024. Navigating Real-World Challenges: A Quadruped Robot Guiding System for Visually Impaired People in Diverse Environments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 44, 18 pages. doi:10.1145/3613904.3642227

[8] Ruijia Chen, Junru Jiang, Pragati Maheshwary, Brianna R. Cochran, and Yuhang Zhao. 2025. VisiMark: Characterizing and Augmenting Landmarks for People with Low Vision in Augmented Reality to Support Indoor Navigation. doi:10.1145/3706598.3713847 arXiv:2502.10561 [cs].

[9] Dimitrios Dakopoulos and Nikolaos G. Bourbakis. 2010. Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1 (2010), 25–35. doi:10.1109/TSMCC.2009.2021255

[10] Fatma El-Zahraa El-Taher, Luis Miralles-Pechuán, Jane Courtney, Kristina Millar, Chantelle Smith, and Susan Mckeever. 2023. A Survey on Outdoor Navigation Applications for People With Visual Impairments. *IEEE Access* 11 (2023), 14647–14666. doi:10.1109/ACCESS.2023.3244073

[11] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brighid Meredith, Cheng Peng, Chris Sweeney, Cole Wilson, Dan Barnes, Daniel DeTone, David Caruso, Derek Valleroy, Dinesh Ginjupalli, Duncan Frost, Edward Miller, Elias Mueggler, Evgeniy Oleinik, Fan Zhang, Guruprasad Somasundaram, Gustavo Solaira, Harry Lanaras, Henry Howard-Jenkins, Huixuan Tang, Hyo Jin Kim, Jaime Rivera, Ji Luo, Jing Dong, Julian Straub, Kevin Bailey, Kevin Eckenhoff, Lingni Ma, Luis Pesqueira, Mark Schwesinger, Maurizio Monge, Nan Yang, Nick Charron, Nikhil Raina, Omkar Parkhi, Peter Borschowa, Pierre Moulon, Prince Gupta, Raul Mur-Artal, Robbie Pennington, Sachin Kulkarni, Sagar Miglani, Santosh Gondi, Saransh Solanki, Sean Diener, Shangyi Cheng, Simon Green, Steve Saarinen, Suvam Patra, Tassos Mourikis, Thomas Whelan, Tripti Singh, Vasileios Balntas, Vijay Baiyya, Wilson Dreewes, Xiaqing Pan, Yang Lou, Yipu Zhao, Yusuf Mansour, Yuyang Zou, Zhaoyang Lv, Zijian Wang, Mingfei Yan, Carl Ren, Renzo De Nardi, and Richard Newcombe. 2023. Project Aria: A New Tool for Egocentric Multi-Modal AI Research. arXiv:2308.13561 [cs.HC]

[12] Envision. 2023. Envision Smart Glasses. https://www.letsenvision.com/.

[13] Bhanuka Gamage, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur Lowery, and Kim Marriott. 2023. 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 Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 30, 21 pages. doi:10.1145/3597638.3608955

[14] Nicholas A. Giudice and Gordon E. Legge. 2008. Blind Navigation and the Role of Technology. In *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*. John Wiley & Sons, Ltd, 479–500. doi:10.1002/9780470379424.ch25 Section: 25 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470379424.ch25.

[15] Ian Gonsher, Adriana Salazar, Shrey Mehta, Samantha Shulman, Nicholas Gaitanis, Arshiya Khosla, Denise Danielle Tamesis, and Jillian Sun. 2023. 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. doi:10.54941/ahfe1004383 ISSN: 27710718 Issue: 116.

[16] Qiao Gu, Zhaoyang Lv, Duncan Frost, Simon Green, Julian Straub, and Chris Sweeney. 2024. EgoLifter: Open-world 3D Segmentation for Egocentric Perception. *arXiv preprint arXiv:2403.18118* (2024).

[17] Maya Gupta, Ali Abdolrahmani, Emory Edwards, Mayra Cortez, Andrew Tumang, Yasmin Majali, Marc Lazaga, Samhitha Tarra, Prasad Patil, Ravi Kuber, and Stacy M. Branham. 2020. Towards More Universal Wayfinding Technologies: Navigation Preferences Across Disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376581

[18] Hwei-Shin Harriman, Dragan Ahmetovic, Sergio Mascetti, Darren Moyle, Michael Evans, and Paul Ruvolo. 2021. Clew3D: Automated Generation of O&M Instructions Using LIDAR-Equipped Smartphones. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, USA) *(ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 54, 3 pages. doi:10.1145/3441852.3476564

[19] Wilko Heuten, Niels Henze, Susanne Boll, and Martin Pielot. 2008. Tactile wayfinder: a non-visual support system for wayfinding. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges* (Lund, Sweden) *(NordiCHI '08)*. Association for Computing Machinery, New York, NY, USA, 172–181. doi:10.1145/1463160.1463179

[20] Mohd Heikal Husin and Yang Kwang Lim and. 2020. InWalker: smart white cane for the blind. *Disability and Rehabilitation: Assistive Technology* 15, 6 (2020), 701–707. doi:10.1080/17483107.2019.1615999 arXiv:https://doi.org/10.1080/17483107.2019.1615999 PMID: 31729282.

[21] Glidance Inc. [n. d.]. Glide: AI Mobility and Navigation Aid for Blind & Low Vision. https://glidance.io/.

[22] Gaurav Jain, Basel Hindi, Zihao Zhang, Koushik Srinivasula, Mingyu Xie, Mahshid Ghasemi, Daniel Weiner, Sophie Ana Paris, Xin Yi Therese Xu, Michael Malcolm, Mehmet Kerem Turkcan, Javad Ghaderi, Zoran Kostic, Gil Zussman, and Brian A. Smith. 2024. 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* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 139, 21 pages. doi:10.1145/3654777.3676333

[23] Gaurav Jain, Yuanyang Teng, Dong Heon Cho, Yunhao Xing, Maryam Aziz, and Brian A. Smith. 2023. "I Want to Figure Things Out": Supporting Exploration in Navigation for People with Visual Impairments. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 63 (April 2023), 28 pages. doi:10.1145/3579496

[24] Watthanasak Jeamwatthanachai, Mike Wald, and Gary Wills. 2019. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment* 37, 2 (2019), 140–153.

[25] Justin Kasowski, Byron A. Johnson, Ryan Neydavood, Anvitha Akkaraju, and Michael Beyeler. 2023. A systematic review of extended reality (XR) for understanding and augmenting vision loss. *Journal of Vision* 23, 5 (05 2023), 5–5. doi:10.1167/jov.23.5.5

[26] Robert K. Katzschmann, Brandon Araki, and Daniela Rus. 2018. 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 (2018), 583–593. doi:10.1109/TNSRE.2018.2800665

[27] Izaz Khan, Shah Khusro, and Irfan Ullah. 2018. Technology-assisted white cane: evaluation and future directions. *PeerJ* 6 (Dec. 2018), e6058. doi:10.7717/peerj.6058

[28] Jingyi Li, Son Kim, Joshua A. Miele, Maneesh Agrawala, and Sean Follmer. 2019. Editing Spatial Layouts through Tactile Templates for People with Visual Impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3290605.3300436

[29] Yimin Lin, Kai Wang, Wanxin Yi, and Shiguo Lian. 2019. Deep Learning Based Wearable Assistive System for Visually Impaired People. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. 2549–2557. doi:10.1109/ICCVW.2019.00312

[30] Guanhong Liu, Tianyu Yu, Chun Yu, Haiqing Xu, Shuchang Xu, Ciyuan Yang, Feng Wang, Haipeng Mi, and Yuanchun Shi. 2021. Tactile Compass: Enabling Visually Impaired People to Follow a Path with Continuous Directional Feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 28, 13 pages. doi:10.1145/3411764.3445644

[31] Peidong Liu, Xingxing Zuo, Viktor Larsson, and Marc Pollefeys. 2021. MBA-VO: Motion Blur Aware Visual Odometry. IEEE Computer Society, 5530–5539. doi:10.1109/ICCV48922.2021.00550

[32] Fernando Merchan, Martin Poveda, Danilo Cáceres-Hernández, and Javier Sanchez Galan F. 2021. Indoor Navigation Aid Systems for the Blind and Visually Impaired Based on Depth Sensors. 187–223. doi:10.4018/978-1-7998-6522-3.ch007

[33] Hein Min Htike, Tom H. Margrain, Yu-Kun Lai, and Parisa Eslambolchilar. 2021. Augmented Reality Glasses as an Orientation and Mobility Aid for People with Low Vision: a Feasibility Study of Experiences and Requirements. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 729. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445327

[34] Karin Müller, Christin Engel, Claudia Loitsch, Rainer Stiefelhagen, and Gerhard Weber. 2022. 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, Article 13 (May 2022), 44 pages. doi:10.1145/3514255

[35] OrCam. [n. d.]. OrCam MyEye 3 Pro. https://www.orcam.com/en-us/orcam-myeye-3-pro.

[36] Jagannadh Pariti, Vinita Tibdewal, and Tae Oh. 2020. 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* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3334480.3375217

[37] Jun Park and Subeh Chowdhury. 2018. Investigating the barriers in a typical journey by public transport users with disabilities. *Journal of Transport & Health* 10 (2018), 361–368. doi:10.1016/j.jth.2018.05.008

[38] Santiago Real and Alvaro Araujo. 2019. Navigation Systems for the Blind and Visually Impaired: Past Work, Challenges, and Open Problems. *Sensors* 19, 15 (2019). doi:10.3390/s19153404

[39] Manaswi Saha, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2019. 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* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 222–235. doi:10.1145/3308561.3353776

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

[41] Patrick Slade, Arjun Tambe, and Mykel J. Kochenderfer. 2021. Multimodal sensing and intuitive steering assistance improve

navigation and mobility for people with impaired vision. *Science Robotics* 6, 59 (2021), eabg6594. doi:10.1126/scirobotics.abg6594 arXiv:https://www.science.org/doi/pdf/10.1126/scirobotics.abg6594

[42] Paraskevi Theodorou, Kleomenis Tsiligkos, and Apostolos Meliones. 2023. 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

[43] I. Ulrich and J. Borenstein. 2001. The GuideCane-applying mobile robot technologies to assist the visually impaired. *Trans. Sys. Man Cyber. Part A* 31, 2 (March 2001), 131–136. doi:10.1109/3468.911370

[44] Hsueh-Cheng Wang, Robert K. Katzschmann, Santani Teng, Brandon Araki, Laura Giarré, and Daniela Rus. 2017. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 6533–6540. doi:10.1109/ICRA.2017.7989772

[45] WeWALK. 2020. WeWALK Smart Cane. https://wewalk.io/en/.

[46] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. 2014. "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)*. Association for Computing Machinery, New York, NY, USA, 217–224. doi:10.1145/2661334.2661380