AI-Enhanced Navigational Glasses for Blind

Mrs. S. A. Bagal[1], Shardul Mhetre[2], Pranav Naik[3], Soham Walimbe[4]

[1]Guide, Department of Artificial Intelligence and Data Science,AISSMS Institute of Information Technology

[2,3,4]Department of Artificial Intelligence and Data Science,AISSMS Institute of Information Technology

Abstract

Visually impaired individuals encounter significant challenges in achieving safe and independent navigation in dynamic environments. Over the past decade, advancements in Artificial Intelligence (AI), computer vision, and sensor technologies have accelerated the development of intelligent assistive systems designed to enhance spatial awareness and mobility. This survey provides an in-depth look at modern navigational assistance solutions, encompassing hardware-based and AI-driven approaches. It explores the integration of embedded platforms, such as Raspberry Pi and Arduino, with camera modules, and LiDAR for real-time obstacle detection and path guidance. In addition, deep learning techniques, including convolutional neural networks (CNNs), transformer-based vision models, and simultaneous localization and mapping (SLAM), have been examined for their roles in perception, scene understanding, and decision-making. This study further reviews multimodal data fusion methods, user interaction mechanisms such as audio and haptic feedback, and wearable implementations aimed at intuitive guidance. This study identifies key challenges related to latency, energy efficiency, cost, and user adaptability and outlines emerging research directions in edge AI and context-aware systems to enable robust, affordable, and intelligent navigation assistance for the visually impaired community.

Keywords: Artificial Intelligence, Assistive Technology, Navigational Assistance, Visually Impaired, Computer Vision, Obstacle Detection, Deep Learning, Sensor Fusion, Edge Computing, Wearable Devices, YOLO, Large Language Models (LLM)

1. INTRODUCTION

Vision is a primary sense that facilitates effortless interaction with and navigation through our complex environments. For the over 43 million people worldwide who are blind or visually impaired (according to the World Health Organization), the simple act of moving through an unfamiliar or even a familiar space can present a significant challenge [1]. This challenge extends beyond mere mobility, often impacting independence, social integration, and overall quality of life. For centuries, the white cane and guide dogs have served as the primary tools for aiding navigation, offering tactile and guided support. While invaluable, these tools have inherent limitations; they are primarily effective for detecting immediate, ground-level obstacles and require extensive training to use effectively. As urban environments become increasingly complex, the limitations of conventional aids have driven researchers and engineers to explore advanced solutions that combine Artificial Intelligence (AI), computer vision, and embedded systems to enhance environmental understanding and autonomous navigation.

Over the past decade, significant progress has been made in developing electronic travel aids (ETAs) that integrate hardware and AI-based techniques to provide real-time feedback and guidance [2]. These systems typically employ sensors such as ultrasonic modules, LiDAR, infrared detectors, and RGB or depth cameras to detect and classify obstacles in the user's surroundings. Embedded computing platforms like Raspberry Pi, Arduino, and NVIDIA Jetson have enabled on-device data processing and decision-making, reducing dependency on external computation and allowing for portable, low-power implementations. AI and deep learning have revolutionized the capabilities of these systems by enabling robust perception and scene understanding. Convolutional Neural Networks (CNNs) and transformer-based models are increasingly used for tasks such as object detection, semantic segmentation, and depth estimation, allowing systems to identify pedestrians, vehicles, or other obstacles with high accuracy [10] [transformer]. In parallel, Simultaneous Localization and Mapping (SLAM) techniques and path planning algorithms have contributed to spatial mapping and autonomous navigation, enabling users to receive context-aware instructions in real time [3]. Moreover, advances in multimodal data fusion combining data from multiple sensors have enhanced reliability in complex or cluttered environments [21] [24].

Despite these advancements, several challenges persist in achieving widespread adoption of AI-based navigational aids. Key issues include computational latency in real-time processing, power consumption on embedded devices, environmental adaptability under varying lighting or weather conditions, and ensuring user privacy and data security [4]. Furthermore, high production costs and the need for user-specific customization often limit the affordability and accessibility of such solutions.

2. Literature Review

Research on navigational assistance for visually impaired users spans several decades and a wide spectrum of approaches, from early electronic travel aids (ETAs) to recent AI-driven, multimodal systems.

Sensor-Based Navigational Systems

The foundational work in Electronic Travel Aids (ETAs) focused on electronically extending the perceptual range of the traditional white cane. These early systems primarily leveraged non-visual sensors like ultrasonic, infrared (IR), and LiDAR to detect obstacles beyond the cane's physical reach, adapting obstacle-avoidance strategies from mobile robotics for human wearables.

Ultrasonic sensors were a cornerstone of this era due to their low cost, simplicity, and reliable short-range detection. They operate by emitting high-frequency sound waves and calculating distance based on the time-of-flight of the reflected echo. Pioneering devices like the Sonic Pathfinder used ultrasonic transducers mounted on a headband to provide auditory feedback on obstacle proximity [5]. A significant evolution of this concept was the GuideCane, which integrated an array of ultrasonic sensors onto a wheeled cane-like device [5]. Upon detecting an obstacle, it would physically steer the user around it, providing directional guidance through tactile force. Similarly, the NavBelt utilized sensors, often including more advanced LiDAR for higher spatial resolution, to construct a simple environmental map and provide auditory directional cues, effectively creating a wearable obstacle-avoidance system [5] [16]. Subsequent generations of these devices, often categorized as "smart canes," refined this approach by integrating ultrasonic sensors with simple microcontrollers and vibration motors to provide graded, haptic proximity feedback [7]. Commercial and research prototypes like the SmartCane demonstrated the practical viability of enhancing the standard white cane with electronics to detect knee-to-head-level obstacles [7]. Several works integrate haptics with vision and range sensors to convey prioritized, context-aware alerts [8] [9]. Infrared (IR) sensors were also explored in some low-cost implementations for their short-range precision, though their susceptibility to ambient light and reflective surfaces limited their reliability [15].

Despite their contribution in proving the feasibility of active obstacle avoidance, these classical ETAs and smart canes faced significant limitations. They often struggled with ergonomics, false positives in cluttered environments, and power/weight constraints.

Vision- and Learning-Based Perception

Vision-based systems represent a significant evolution in Electronic Travel Aids (ETAs), leveraging camera modules—including RGB, stereo, and depth sensors—to capture the user's surroundings. Unlike classical ETAs that primarily measure proximity, these systems aim to extract semantically rich information about the environment, enabling object recognition, path identification, and a higher degree of contextual awareness.

The progression of these systems mirrors advances in the field of computer vision. Early prototypes relied on traditional techniques such as edge detection, optical flow, and feature extraction for basic obstacle identification. A paradigm shift occurred with the rise of deep learning, particularly Convolutional Neural Networks (CNNs) [10]. The adoption of pretrained models like YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN dramatically improved the accuracy and robustness of real-time object detection and scene understanding [17] [18] [19]. For instance, a system proposed by Kuriakose et al. (2019) utilized a CNN model on embedded hardware for pedestrian and obstacle detection, providing users with voice-based feedback. Further advancing this concept, Tian et al. (2020) developed an indoor navigation system that employed depth cameras and semantic segmentation to generate spatial layouts for obstacle-free path planning.

To move beyond 2D recognition and achieve true 3D spatial understanding, researchers frequently integrate depth-sensing cameras, such as the Intel RealSense and Microsoft Kinect. These sensors are crucial for estimating precise distances to objects. This capability is often combined with Simultaneous Localization and Mapping (SLAM) techniques—including monocular SLAM—to construct dynamic environmental maps, thereby significantly improving localization and path prediction accuracy [3] [13]. Recent research continues to push the boundaries of perception, exploring transformer-based vision models for enhanced performance in tasks like depth estimation and semantic segmentation [11] [12]. A consistent finding in the literature is that multi-modal sensing improves system robustness. Combining visual data with complementary inputs from depth sensors (stereo, structured light) or LiDAR mitigates challenges in complex lighting and texture-poor environments where cameras alone may fail [14]. However, a critical challenge for deployment is the high computational demand of these sophisticated models, which leads to latency, thermal issues, and energy consumption on portable hardware. Consequently, a major research thrust focuses on optimizing these systems for real-time performance through model compression techniques (quantization, pruning), efficient neural architectures (e.g., MobileNet, Tiny-YOLO variants), and the use of edge AI accelerators to strike a practical balance between analytical power and operational efficiency [15] [19].

Hybrid and AI-Driven Approaches

Recent research trends emphasize hybrid systems that combine multiple sensing modalities with AI algorithms to overcome individual limitations of single-sensor designs [20]. By fusing data from ultrasonic, LiDAR, and camera sensors, these systems achieve more reliable obstacle detection and contextual awareness.

Khan et al. (2021) developed a multimodal navigation aid integrating stereo vision and ultrasonic sensors with a CNN-based object recognition module [21]. The system leveraged sensor fusion to handle diverse environments, using deep learning for classification and ultrasonic data for precise distance estimation.

Similarly, Zhang et al. (2022) introduced a wearable navigation framework using LiDAR, GPS, and IMU data combined with a transformer-based model for spatial scene interpretation [22]. This integration allowed real-time localization and obstacle avoidance in outdoor scenarios.
Advancements in edge AI have also enabled real-time inference on low-power embedded platforms such as NVIDIA Jetson Nano, Google Coral TPU, and Raspberry Pi 4. These platforms can process neural network models locally, reducing latency and ensuring privacy by avoiding cloud-based computation. Moreover, lightweight deep learning architectures like MobileNet, EfficientNet, and Tiny-YOLO have been adopted to optimize performance without sacrificing accuracy [23].

3. Limitations and Challenges

Despite the significant progress in assistive navigation systems for visually impaired users, there remain several persistent limitations that constrain real-world adoption and performance. These can be grouped under hardware constraints, algorithmic and perception challenges, user interaction and usability issues, and socio-economic/accessibility barriers.

Hardware and Embedded System Constraints

Embedded assistive prototypes frequently struggle with weight, power consumption, heat dissipation, and portability. Wearable devices that combine cameras, LiDAR or depth sensors, and compute modules often exceed acceptable ergonomic limits for everyday use. For instance, reviews highlight that devices still impose significant bulk or frequent charging needs, which discourage continuous use [24]. Moreover, on-device computing of deep learning models for real-time guidance frequently incurs latency and thermal limitations when deployed on low-power platforms [25]. Perception, Scene Understanding and Environment Adaptation

Many systems perform well in controlled indoor settings but fail to generalize to dynamic, unstructured outdoor environments. Challenges such as varying lighting, shadows, reflective surfaces, rain, texture-poor regions and overhead obstacles degrade sensor and vision system performance. For example, SLAM-based assistive navigation reviews point out limitations in dynamic scene changes and feature-poor environments [25]. Furthermore, semantic understanding—such as recognizing crosswalks, curbs or moving hazards—is still fragile in real time, especially on resource-constrained embedded systems [24].

Algorithmic Efficiency and Real-Time Responsiveness

The trade-off between accuracy of perception algorithms and computational/energy cost remains a major bottleneck. Deep learning models for object detection, semantic segmentation or depth estimation are often too heavy for wearables without significant optimizations. Studies recommend model compression, pruning, and lightweight architectures, but real-world prototypes often still incur high latency or fail to meet the strict timing needed for safe navigation [26].

Usability, Feedback Modalities and User Adaptation

While many systems offer audio or haptic feedback, delivering information without overwhelming or confusing the user remains a delicate design challenge. Multimodal fusion systems (audio + haptic + tactile) are promising, but balancing feedback intensity, modality and relevance to the user's context is under-explored. For example, users may ignore audio cues in noisy outdoor environments or get fatigued with continuous haptic feedback [24]. Additional issues include user training time, learning curves, and device acceptance. A recent study reported that visually impaired users cite comfort, intuitiveness and adaptability as key barriers to adoption.

Cost, Accessibility and Real-World Deployment

Commercialization and widespread adoption of assistive navigation technologies are still limited by cost, manufacturing scalability, infrastructure dependencies (e.g., Internet connectivity, GPS accuracy) and maintenance. Many research prototypes remain laboratory demonstrations rather than field-tested deployments. A systematic review emphasises that fewer than half of surveyed systems report real-world user trials, and many assume ideal conditions [27]. Moreover, the gap between research and user need remains: devices may not align with the daily mobility patterns, preferences or constraints of visually impaired populations [24].

Privacy, Data Security and Ethical Considerations

Assistive navigation systems often rely on cameras, microphones or other sensors to monitor the environment, which raises significant privacy and ethical concerns. Users' environments may be recorded or streamed, raising questions about data consent, continuous surveillance and personal safety. Though less frequently addressed in technical literature, these are essential impediments for trust and acceptance.

Interoperability and Standardisation

A further challenge lies in the lack of interoperability standards for assistive navigation systems. Many research systems are bespoke, proprietary, or require custom hardware, making integration into existing mobility aids or user ecosystems difficult. The fragmentation of platforms, sensor types, feedback modes and interfaces limits scalability and user choice.

4. Comparative Study

System	Hardware	Core Technique	Feedback	Strengths	Limitations
GuideCane	Ultrasonic Sensors	Obstacle avoidance algorithm	Audio/ Tactile	Early proof of concept for autonomous navigation	Bulky, limited accuracy in cluttered areas
Smart Cane	Ultrasonic and IR Sensors	Microcontroller-based processing	Vibration	Low-cost, simple short-range detection	No scene understanding or localization
Vision-Based Systems	Camera + IMU	CNN-based object detection	Audio	Object recognition and semantic awareness	High computational cost, latency issues
Multimodal fusion	Camera + Ultrasonic/LiDAR	Sensor fusion	Audio + Haptic	Reliable in diverse environments	Complex synchronization, higher power needs
Edge AI Wearables	Raspberry Pi + Depth Camera	Compressed DNN (MobileNet, Tiny-YOLO)	Audio	Real-time, portable	Accuracy trade-offs due to compression

Table 1: Comparative Study of AI-Based Navigational Systems for the Visually Impaired

5. Proposed Solution

To address the limitations of existing navigational aids, this work proposes a smart wearable goggle-based assistive system designed to provide real-time obstacle detection and navigational guidance for visually impaired individuals. The architecture integrates computer vision, embedded computing, and language models to achieve intelligent environmental awareness and human-friendly feedback.

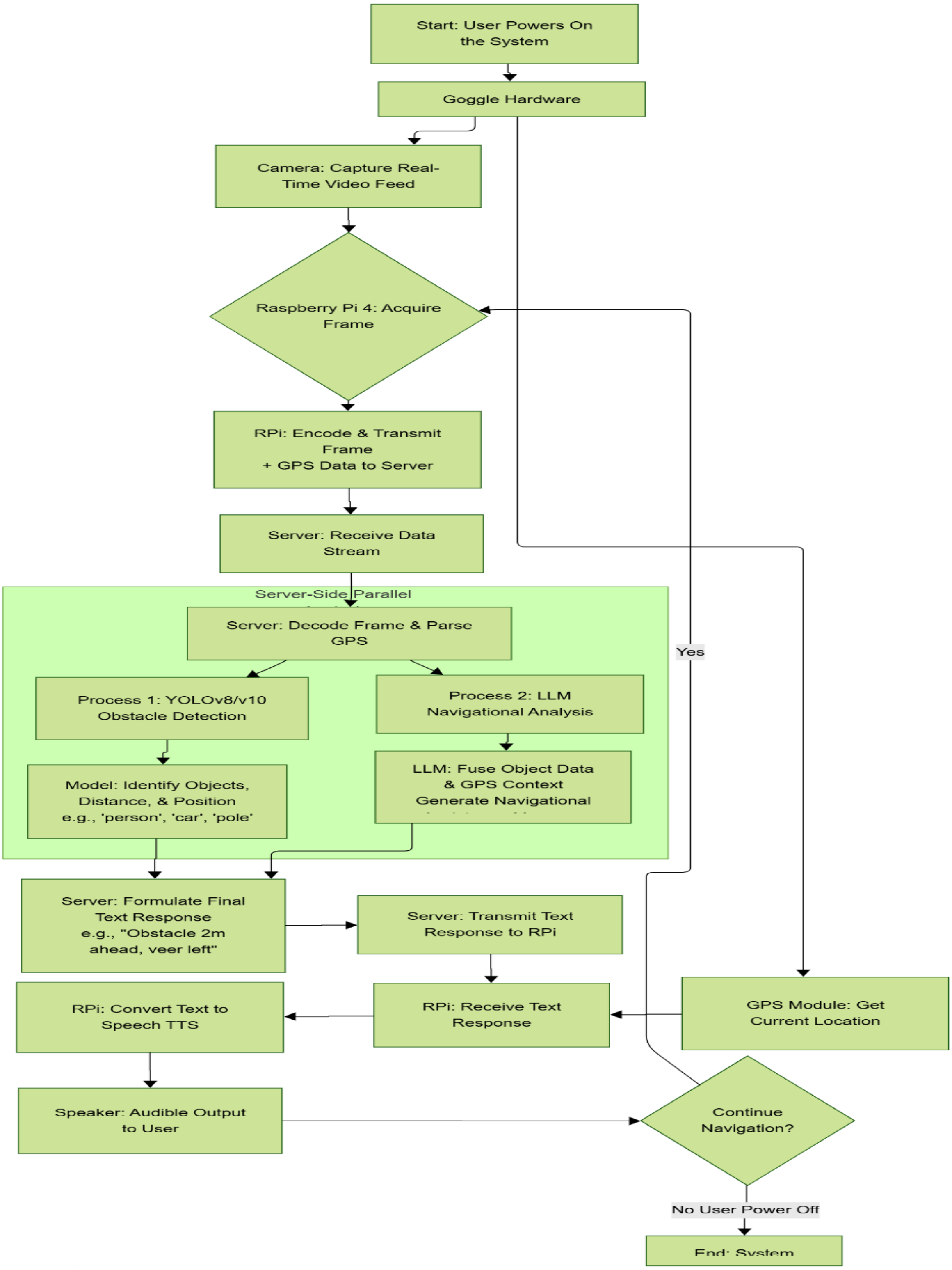

Image 1: System Architecture & Workflow

System Overview

The proposed system consists of a camera-mounted goggle, a Raspberry Pi 4 as the local processing and communication unit, and a cloud or edge server responsible for advanced AI inference. The camera continuously captures the user's forward-facing scene, transmitting the live video feed to the server via a wireless network.

Perception and Object Detection

At the server end, the video stream is processed using a YOLOv8/YOLOv10-based deep learning model trained for object and obstacle detection. These models are selected for their superior accuracy-speed trade-off, enabling effective identification of static and dynamic obstacles such as pedestrians, vehicles, poles, and uneven terrain. The detected objects are further analyzed to estimate distance and relative motion to the user.

Contextual Reasoning and Navigational Assistance

To provide high-level situational understanding, the obstacle data and scene context are passed to a Large Language Model (LLM) integrated into the server pipeline. The LLM interprets the detected scene, infers spatial relationships, and generates natural-language navigational instructions (e.g., "Obstacle ahead, move slightly left"). This enables adaptive guidance rather than rigid rule-based feedback.

Real-Time Feedback Delivery

The generated navigational message is transmitted back to the Raspberry Pi in real time. The Pi converts the text output into speech feedback using a Text-to-Speech (TTS) engine. A miniature speaker embedded in the goggle delivers clear and immediate audio cues to the user, ensuring seamless situational awareness without distracting sensory overload.

Localization and Mobility Support

A GPS module is integrated into the system to enable continuous location tracking and potential route planning in future iterations. The GPS data can also assist in generating context-aware alerts, such as warnings near road intersections or construction zones.

System Advantages

The proposed architecture combines real-time vision processing, semantic reasoning via LLM, and speech-based feedback, creating a robust assistive ecosystem. By offloading computationally intensive tasks to the server while retaining minimal edge operations on the Raspberry Pi, the design maintains low latency, energy efficiency, and affordability—key criteria for real-world usability.

6. Future Research Directions

The proposed AI-driven navigational assistant offers a solid foundation for safe and intelligent mobility support for visually impaired users. However, several enhancements can be explored to improve adaptability, scalability, and user experience in future research and development

Integration of Edge and Cloud Intelligence

Future iterations can adopt a hybrid edge–cloud framework to balance computation between the Raspberry Pi and remote servers. Incorporating edge accelerators such as the Google Coral TPU or NVIDIA Jetson Nano enables real-time inference using lightweight models like quantized YOLO or MobileViT. Meanwhile, complex reasoning tasks can be offloaded to the cloud. This integration reduces latency, enhances responsiveness, and ensures reliable operation even under limited network connectivity.

Enhanced Scene Understanding with Multimodal Learning

The next generation of systems can incorporate multimodal fusion, combining visual, auditory, and spatial data for deeper environmental comprehension. Integrating depth sensors, IMUs, and LiDAR can improve obstacle detection accuracy in complex or low-light environments. Future models can also employ Vision-Language Models (VLMs) to provide richer semantic interpretations of surroundings—such as describing crosswalks, signboards, or moving hazards in natural language.

Integration with Navigation and IoT Ecosystems

The inclusion of GPS-based route planning can be expanded into full-fledged geo-aware navigation, integrating with Google Maps APIs or OpenStreetMap datasets. Coupling this system with smart city infrastructure (IoT beacons, pedestrian signals, and connected transport systems) could provide context-aware alerts such as traffic light status or public transport proximity.

Personalized and Adaptive Feedback Mechanisms

Further advancements can introduce adaptive feedback strategies that tailor guidance based on individual user behavior, walking speed, and environment type. Machine learning-based personalization can help reduce cognitive load and deliver feedback in a preferred modality—audio, vibration, or even bone-conduction output—to enhance comfort and usability in public spaces.

7. Conclusion

This paper presented a comprehensive survey of AI-based navigational assistance systems developed to enhance the mobility and independence of visually impaired individuals. The review covered classical electronic travel aids, smart canes, wearable haptic interfaces, and modern AI-driven systems employing deep learning and multimodal fusion. While significant progress has been made through the integration of vision-based perception and sensor technologies, challenges such as computational efficiency, latency, affordability, and user adaptability persist.

To address these limitations, a goggle-based assistive system was proposed, combining camera vision, YOLOv8/YOLOv10-based obstacle detection, and LLM-driven navigational reasoning, supported by a Raspberry Pi for edge processing and text-to-speech feedback. The system further incorporates GPS-based localization to improve situational awareness. This architecture aims to deliver real-time, context-aware, and user-friendly guidance in both indoor and outdoor environments.

8. Acknowledgement

The authors want to thank Mrs. Bagal for her useful remarks and suggestions, as well as thank everyone who helped plan this job.

References

1. Vision Loss Expert Group of the Global Burden of Disease Study. “Global estimates on the number of people blind or visually impaired by cataract: a meta-analysis from 2000 to 2020.” Eye 38.11 (2024): 2156.
2. Kim, In-Ju. “Recent advancements in indoor electronic travel aids for the blind or visually impaired: a comprehensive review of technologies and implementations.” Universal Access in the Information Society 24.1 (2025): 173-193.
3. Bamdad, Marziyeh, Davide Scaramuzza, and Alireza Darvishy. “Slam for visually impaired people: A survey.” IEEE Access (2024).
4. Real, Santiago, and Alvaro Araujo. “Navigation systems for the blind and visually impaired: Past work, challenges, and open problems.” Sensors 19.15 (2019): 3404.

5. Shoval, Shraga, Iwan Ulrich, and Johann Borenstein. “NavBelt and the Guide-Cane [obstacle-avoidance systems for the blind and visually impaired].” IEEE robotics & automation magazine 10.1 (2003): 9-20.
6. Pérez, Andrés, et al. “Intelligent Jacket for Monitoring Mobility of People with Reduced Disabilities.” International Conference on Information Technology & Systems. Cham: Springer International Publishing, 2021.
7. Ramisetti, Chandu, et al. “An Ultrasonic Sensor-based blind stick analysis with instant accident alert for Blind People.” 2022 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2022.
8. Sorgini, Francesca, et al. “Haptic-assistive technologies for audition and vision sensory disabilities."”Disability and Rehabilitation: Assistive Technology 13.4 (2018): 394-421.
9. Kleinberg, Dvir, et al. “A haptic feedback system for spatial orientation in the visually impaired: A comprehensive approach.” IEEE Sensors Letters 7.9 (2023): 1-4.
10. Ashiq, Fahad, et al. “CNN-based object recognition and tracking system to assist visually impaired people.” IEEE access 10 (2022): 14819-14834.
11. Zhang, Jiaming, et al. “Trans4Trans: Efficient transformer for transparent object segmentation to help visually impaired people navigate in the real world.” Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
12. Yogaraj, A., et al. “Vision Transformer Based Vision Enhancement for Visually Impaired Individuals.” 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT). Vol. 1. IEEE, 2024.
13. Bamdad, Marziyeh, Davide Scaramuzza, and Alireza Darvishy. “Slam for visually impaired people: A survey.” IEEE Access (2024).
14. Mahendran, Jagadish K., et al. “Computer vision-based assistance system for the visually impaired using mobile edge artificial intelligence.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
15. Nada, Ayat A., Mahmoud A. Fakhr, and Ahmed F. Seddik. “Assistive infrared sensor based smart stick for blind people.” 2015 Science and Information Conference (SAI). IEEE, 2015.
16. Jain, Mayuri, and Warish Patel. “Review on lidar-based navigation systems for the visually impaired.” SN Computer Science 4.4 (2023): 323.
17. Joshi, Rakesh Chandra, Saumya Yadav, and Malay Kishore Dutta. “YOLO-v3 based currency detection and recognition system for visually impaired persons.” 2020 International Conference on Contemporary Computing and Applications (IC3A). IEEE, 2020.
18. Arora, Adwitiya, et al. “Real time multi object detection for blind using single shot multibox detector.” Wireless Personal Communications 107.1 (2019): 651-661.
19. Kumar, Nitin, and Anuj Jain. “A Deep Learning Based Model to Assist Blind People in Their Navigation.” J. Inf. Technol. Educ. Innov. Pract. 21 (2022): 95-114.
20. Cho, Jun-Dong. “Multi-sensory interaction for blind and visually impaired people.” Electronics 10.24 (2021): 3170.
21. Khan, Sulaiman, Shah Nazir, and Habib Ullah Khan. “Analysis of navigation assistants for blind and visually impaired people: A systematic review.” IEEE access 9 (2021): 26712-26734.
22. Zhang, Jiachen, et al. “GNSS-RTK adaptively integrated with LiDAR/IMU odometry for continuously global positioning in urban canyons.” Applied Sciences 12.10 (2022): 5193.

23. Maya-Martínez, Sergio-Uriel, et al. "Pedestrian detection model based on Tiny-Yolov3 architecture for wearable devices to visually impaired assistance." Frontiers in robotics and AI 10 (2023): 1052509.1
24. Okolo, G.I.; Althobaiti, T.; Ramzan, N. "Assistive Systems for Visually Impaired Persons: Challenges and Opportunities for Navigation Assistance". Sensors 2024, 24, 3572.
25. M. Bamdad, D. Scaramuzza and A. Darvishy, "SLAM for Visually Impaired People: A Survey," in IEEE Access, vol. 12, pp. 130165-130211, 2024.
26. Messaoudi MD, Menelas BJ, Mcheick H. Review of Navigation Assistive Tools and Technologies for the Visually Impaired. Sensors (Basel). 2022 Oct 17;22(20):7888.
27. Casanova E, Guffanti D, Hidalgo L. Technological Advancements in Human Navigation for the Visually Impaired: A Systematic Review. Sensors (Basel). 2025 Apr 1;25(7):2213.