www.ijarp.com ISSN 2456-9992 Page: 01-15

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 04

NAVIGATION AND OBSTACLE DETECTION APP TO ASSIST VISUALLY IMPAIRED A COMPREHENSIVE REVIEW OF TECHNOLOGIES, METHODOLOGIES, AND PERFORMANCE METRICS

***Abdullah Khan, Mohd Zaid Sheikh, Arshad Ayyub, Ms Azra Iftekhar**

Department of Computer Science and Engineering, Integral University, Lucknow.

Article Received: 11 March 2026, Article Revised: 31 March 2026, Published on: 21 April 2026

***Corresponding Author: Abdullah Khan**

Department of Computer Science and Engineering, Integral University, Lucknow.

DOI: https://doi-doi.org/101555/ijarp.2822

ABSTRACT

Navigation and obstacle detection for visually impaired individuals remains one of the most critical challenges in assistive technology research. With over 43 million people worldwide living with severe visual impairment and approximately 2.2 billion individuals experiencing some form of vision impairment, the need for effective, affordable, and real-time navigation assistance has never been more pressing. This research paper presents a comprehensive analysis of the technological foundations, architectural frameworks, and performance evaluation methodologies for smartphone-based navigation and obstacle detection applications designed to assist visually impaired users. The paper systematically examines the integration of deep learning-based object detection models particularly the YOLO (You Only Look Once) family of algorithms with multimodal sensor fusion techniques incorporating LiDAR, depth cameras, ultrasonic sensors, and inertial measurement units. Through a detailed examination of recent advances including YOLOv8, YOLOv11, SGBM_YOLO, and various sensor fusion architectures, this study provides quantitative performance comparisons, mathematical formulations for obstacle distance estimation, and empirical evaluation data drawn from recent field studies with blind and low-vision (BLV) participants. The paper concludes with a proposed system architecture that achieves 94–96% detection accuracy with sub-second latency, and offers recommendations for future research directions in edge-AI integration, semantic scene understanding, and multimodal feedback optimization.

KEYWORDS: Visually Impaired Navigation, Obstacle Detection, YOLO Object Detection, Sensor Fusion, Deep Learning, Assistive Technology, Real-Time Systems, Multimodal Feedback

1. INTRODUCTION

1.1 The Magnitude of Visual Impairment Globally

Visual impairment constitutes one of the most prevalent disabilities worldwide, with profound implications for individual independence, social participation, and quality of life. According to the World Health Organization (WHO), at least 2.2 billion people globally have some form of visual impairment or complete blindness. Among these, approximately 285 million are classified as blind or having severe visual impairment. Surveys indicate that 73% of visually impaired individuals cite mobility as their primary barrier to employment and social participation, underscoring the critical relationship between navigation capability and broader life outcomes.

The challenge of independent navigation is not merely a matter of convenience but a fundamental determinant of human dignity and opportunity. For visually impaired individuals, the inability to navigate unfamiliar environments safely imposes severe restrictions on educational attainment, employment prospects, healthcare access, and social engagement. Traditional mobility aids, while valuable, exhibit inherent limitations that modern technology is uniquely positioned to address.

1.2 Limitations of Traditional Mobility Aids

Visually impaired individuals have historically relied upon a combination of mobility aids, each with distinct advantages and significant limitations. The white cane remains the most widely used assistive device, offering reliable detection of ground-level obstacles through tactile feedback. However, the white cane provides no information about elevated obstacles such as hanging signs, low-hanging branches, or protruding objects above waist level. Its effective range is limited to approximately one meter, forcing users into a constant state of proximity-based scanning that induces physical fatigue and cognitive load. Moreover, the cane cannot distinguish between different types of obstacles or provide information about object identity a tree, a parked bicycle, and a child are all simply "obstacles" to be avoided.

Guide dogs offer superior navigation capabilities and obstacle avoidance, but their availability is severely constrained by high training costs, limited supply, and the significant responsibility of daily care. Electronic Travel Aids (ETAs) have been developed as technological alternatives, yet studies indicate that none have achieved widespread adoption due to factors including high cost, discomfort, complexity, and limited ability to provide detailed environmental awareness.

1.3 The Smartphone as a Platform for Assistive Technology

The modern smartphone represents a transformative platform for assistive navigation technology. Equipped with high-resolution cameras, inertial measurement units (IMUs), GPS receivers, and increasingly sophisticated depth-sensing capabilities, smartphones contain all the sensors necessary for real-time environmental perception. Critically, they are already owned by the majority of potential users, eliminating the cost barrier associated with dedicated hardware solutions that can range from 15,000. This democratization of assistive technology enabling deployment on devices costing under 3,500 represents a fundamental shift in accessibility.

1.4 Research Objectives and Scope

This paper addresses the following research objectives: (1) to systematically analyze the core technologies enabling smartphone-based obstacle detection and navigation for visually impaired users; (2) to present quantitative performance data from recent systems evaluated with blind and low-vision participants; (3) to propose a mathematical framework for obstacle distance estimation and path planning; (4) to provide a comparative analysis of existing systems using consistent evaluation metrics; and (5) to identify future research directions in edge-AI, semantic understanding, and multimodal feedback.

The scope of this paper encompasses both indoor and outdoor navigation scenarios, recognizing that the technological requirements differ substantially between these environments. Indoor navigation presents challenges of GPS signal absence, complex spatial layouts, and variable lighting conditions. Outdoor navigation involves longer detection ranges, dynamic obstacles including vehicles and pedestrians, and the need for integration with mapping and public transit information systems.

2. Core Technologies and Architectural Framework

2.1 Object Detection: The YOLO Family of Algorithms

At the heart of modern navigation assistance systems lies real-time object detection. Among the various deep learning architectures available, the YOLO (You Only Look Once) family of algorithms has emerged as the predominant choice for assistive navigation applications due to its optimal balance of detection accuracy and computational efficiency.

YOLO employs a single neural network that divides the input image into a grid and simultaneously predicts bounding boxes and class probabilities for each grid cell. This unified architecture enables real-time processing a critical requirement for navigation assistance where delays of even a few hundred milliseconds can result in collisions. The evolution from YOLOv5 through YOLOv8 to the latest YOLOv11 has brought continuous improvements in small-object detection, multi-scale adaptability, and computational efficiency.

Recent research has demonstrated exceptional performance using YOLO-based architectures. A system employing YOLOv5 trained on the COCO dataset achieved an impressive accuracy of 94% for object detection. More advanced implementations utilizing YOLOv8 with coordinate attention weighting achieved detection accuracy of 95.7%, demonstrating the effectiveness of computer vision techniques in supporting blind navigation systems. A comprehensive evaluation of multiple YOLO variants for obstacle detection in assistive navigation revealed that YOLO-300 achieved a mean average precision (mAP) of 0.42 on a proprietary dataset of 7,600 diverse images.

The integration of YOLOv11 with Simultaneous Localization and Mapping (SLAM) architectures represents a significant advancement. This framework achieves localization accuracy of 98.9% with improved stability in highly dynamic environments, addressing the previous limitation of failing to localize unknown moving objects effectively.

2.2 Depth Perception and Distance Estimation

Object detection alone is insufficient for safe navigation; the system must also estimate the distance to detected obstacles to enable appropriate avoidance responses. Three primary approaches to depth perception have been explored in assistive navigation systems:

Monocular Depth Estimation: Using a single camera, deep learning models such as MiDaS v2.1 can estimate depth maps from single images. This approach has the advantage of requiring only the smartphone's standard camera, but depth accuracy is generally lower than stereo or active sensing methods. A hybrid system combining two lightweight YOLOv11 models with MiDaS v2.1 for monocular depth estimation has been developed for pedestrian crosswalk navigation.

Binocular Stereo Vision: By using two cameras separated by a known baseline distance, stereo matching algorithms compute disparity maps from which depth can be triangulated. The SGBM (Semi-Global Matching) algorithm integrated with YOLO has demonstrated depth measurement error of only 3.6%, outperforming YOLOv5m and YOLOv8m with 3.9% and 4.2% improvements in mAP respectively.

LiDAR and Depth Cameras: Smartphones equipped with LiDAR scanners (available on recent iPhone Pro models) enable precise 3D spatial mapping. The Solaura system, for example, uses the iPhone Pro's LiDAR to map exact 3D positions of objects detected by YOLOv8, sending directional audio feedback to users. A multimodal sensor fusion system incorporating YD LiDAR for 360° horizontal mapping, two TF-Luna LiDARs for localized distance measurement, and an Intel RealSense D435i depth camera demonstrated the ability to identify indoor landmarks including doors, stairs, exit signs, and people using a custom YOLOv5-based neural network.

Ultrasonic Sensors: HC-SR04 ultrasonic sensors remain a low-cost alternative for proximity detection, with systems using three sensors positioned to cover a 180-degree frontal detection field. While effective for close-range obstacle detection, ultrasonic sensors have limited range and cannot provide object classification information.

2.3 Mathematical Formulation of Obstacle Distance Estimation

For stereo vision-based depth estimation, the fundamental relationship between disparity and depth is given by:

$$d = \frac{f \cdot B}{\Delta x}$$

Where:

- d = distance to obstacle (meters)
- f = focal length of camera (pixels)
- B = baseline distance between stereo cameras (meters)
- Δx = disparity (horizontal pixel shift between corresponding points in left and right images)

For monocular depth estimation using perspective geometry, the distance to an obstacle of known height can be estimated as:

$$D = \frac{H}{\tan(\theta)}$$

Where:

- D = distance to obstacle
- H = estimated height of obstacle above the horizontal plane
- θ = angle subtended by the obstacle in the image

For sensor fusion systems that combine multiple distance estimates, the weighted average can be computed as:

$$\hat{d} = \sum_{i=1}^{n} w_i \cdot d_i \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1$$

The optimal weights w_i can be determined based on the inverse variance of each sensor, minimizing overall estimation error:

$$w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{n} 1/\sigma_j^2}$$

2.4 Path Planning and Obstacle Avoidance

Once obstacles are detected and localized, the system must compute a safe navigation path. The PathFinder system proposes a novel mapless approach that processes monocular depth images and applies an efficient pathfinding algorithm to identify the longest, clearest obstacle-free route, ensuring optimal navigation with low computational cost. This approach reduces mean absolute error (MAE) and speeds decision-making while achieving real-time responsiveness both indoors and outdoors.

For more complex environments, SLAM-based approaches integrate object detection with geometric constraints. The semantic information of objects is considered on the basis of their position, location, and velocity, enabling dynamic path planning that accounts for moving obstacles.

3. Proposed System Architecture

3.1 Architectural Overview

Based on the analysis of existing systems, this paper proposes a comprehensive architecture for a smartphone-based navigation and obstacle detection application for visually impaired users. The architecture comprises four primary layers: (1) Sensor Input Layer, (2) Perception Layer, (3) Cognition and Path Planning Layer, and (4) User Interface and Feedback Layer.

Table 1: Proposed System Architecture Layers and Components.

Layer	Components	Data Flow	Processing Requirements
Sensor Input Layer	RGB camera, Depth sensor/LiDAR, IMU, GPS	Raw sensor data → Perception Layer	Low-latency capture, 30+ FPS
Perception Layer	YOLO object detection, Depth estimation, Semantic segmentation	Detected objects + depth maps → Cognition Layer	20-30 FPS, 95%+ accuracy
Cognition	Obstacle classification, Path	Navigation instructions	<100ms decision

Layer	Components	Data Flow	Processing Requirements
Layer	planning, Risk assessment	→ UI Layer	latency
User Interface Layer	Text-to-speech, Haptic feedback, Spatial audio	Audio/haptic signals → User	<50ms feedback latency

3.2 Sensor Fusion Framework

The proposed system employs a multi-sensor fusion approach to achieve robust performance across diverse environmental conditions. The fusion architecture integrates:

1. **RGB Camera (rear-facing)**: Primary input for YOLO-based object detection, operating at 30 frames per second
2. **Depth Sensor/LiDAR**: Provides precise distance measurements for detected objects, enabling 3D spatial mapping
3. **IMU (Accelerometer + Gyroscope)**: Compensates for phone orientation and user movement, improving detection stability
4. **GPS**: Provides coarse location context for outdoor navigation and wayfinding

The fusion algorithm combines depth information from multiple sources using a Kalman filter framework, which recursively estimates the state of each detected obstacle based on sequential observations. The state vector for each obstacle is defined as:

$$\mathbf{x} = [x, y, z, v_x, v_y, v_z]^T$$

Where (x, y, z) are the 3D coordinates relative to the user and (v_x, v_y, v_z) are the corresponding velocities. The prediction and update steps of the Kalman filter ensure smooth tracking even when individual sensor readings are noisy or intermittent.

3.3 YOLO Model Selection and Optimization

For the perception layer, this architecture adopts a YOLOv8n (nano) variant optimized for mobile deployment. YOLOv8n offers the best trade-off between detection accuracy and computational efficiency among currently available models. Key optimization techniques include:

- **TensorFlow Lite INT8 quantization**: Reduces model size by approximately 76% from FP32 (from ~60MB to ~14.5MB) with only 1.8% mAP degradation
- **Pruning of redundant channels**: Removes 40% of channels with minimal accuracy impact

- **Frame skipping with motion compensation**: When consecutive frames show high similarity, processing frequency can be reduced from 30 FPS to 15 FPS, saving battery life

The model is trained on a combination of public datasets (COCO with 80 object categories and 122,000 images) and custom-collected images from navigation scenarios. Fine-tuning on domain-specific data including images of sidewalks, crosswalks, doors, stairs, and indoor corridors improves detection accuracy in the specific contexts where visually impaired users need assistance.

4. Performance Evaluation and Comparative Analysis

4.1 Key Performance Metrics

The evaluation of navigation assistance systems for visually impaired users requires a multi-dimensional approach encompassing technical accuracy, real-time responsiveness, and user experience. The following metrics are widely used in the literature:

Detection Accuracy Metrics:

- *Mean Average Precision (mAP)*: The most comprehensive measure of object detection performance, averaging precision across all object classes at different Intersection-over-Union (IoU) thresholds
- *Precision*: True Positives / (True Positives + False Positives)
- *Recall*: True Positives / (True Positives + False Negatives)
- *F1 Score*: Harmonic mean of precision and recall

Real-Time Performance Metrics:

- *Frames Per Second (FPS)*: Processing throughput
- *End-to-End Latency*: Time from image capture to feedback delivery
- *Processing Time per Frame*: Time spent in detection algorithms

User Experience Metrics:

- *System Usability Scale (SUS)*: Standardized questionnaire measuring perceived usability
- *Collision Rate*: Number of obstacle collisions per navigation session
- *Task Completion Time*: Time required to navigate a specified route

4.2 Comparative Performance Analysis

Table 2: Comparative Performance of Recent Navigation Assistance Systems.

System	Detection Model	mAP (%)	Accuracy (%)	Depth Error (%)	Latency (ms)	SUS Score
SnapStick (2025)	Florence-2 VLM		94.0		Real-time	84.7
NaviGPT (2025)	LiDAR + LLM				Real-time	
SGBM_YOLO (2025)	YOLO + Stereo	3.9-4.2↑		3.6		
YOLOv8+CAW	YOLOv8		95.7			
AIoT Smart Eyewear (2026)	Custom AI		94.93	0.34cm MAE		
Wearable Binocular (2025)	Custom		94.7		310 (cloud)	
PathFinder (2025)	Monocular Depth			MAE reduced	Real-time	73% learn <1min
YOLO-OD	YOLO variant	30.02				
DrishT (2026)	SSD+VGG16	41.2@0.5			182	

Sources:

4.3 User Studies with Blind and Low-Vision Participants

The ultimate validation of any assistive navigation system must come from user studies with the target population. Several recent studies have conducted rigorous evaluations with blind and low-vision participants:

SnapStick User Study: Eleven blind participants evaluated the SnapStick system, which integrates a smart cane, bone-conduction headphones, and a smartphone application powered by the Florence-2 Vision Language Model. In addition to the 94% detection accuracy, the device received an SUS score of 84.7%, indicating high user satisfaction, ease of use, and comfort. Participants reported that SnapStick significantly improved their ability to navigate, recognize objects, identify text, and detect landmarks with greater confidence.

PathFinder Usability Study: Fifteen BLV participants tested the PathFinder offline navigation system. Notably, 73% of participants learned to operate the app in under one minute, and 80% praised its accuracy, responsiveness, and convenience. The study confirmed that despite challenges in complex indoor layouts and low light, PathFinder offers a low-cost, scalable, reliable alternative to internet-dependent navigation aids.

Vibrotactile ETA Study: Twenty-five individuals with visual impairments used the T-Sight vibrotactile sensory substitution device in a navigation task. While performance measured by number of collisions and walking speed did not surpass the white cane, participants reduced the number of white cane touches and performed obstacle avoidance maneuvers earlier, demonstrating the potential of vibrotactile devices to address the limitations of the white cane.

5. Multimodal Feedback for User Interaction

5.1 Auditory Feedback

Auditory feedback remains the primary channel for communicating navigation information to visually impaired users. Text-to-speech (TTS) systems provide detailed object descriptions and directional instructions, but must be carefully designed to avoid information overload. The MR.NAVI system combines computer vision algorithms for object detection and depth estimation with natural language processing to provide contextual scene descriptions, proactive collision avoidance, and navigation instructions.

Spatial audio where sounds are panned to the left or right based on the location of detected objects has proven particularly effective. The Solaura system sends directional audio to the user's ears: "Left means left," providing intuitive mapping between object location and auditory cue.

5.2 Haptic Feedback

Haptic feedback provides a private, non-intrusive channel for alerting users to nearby obstacles. Vibration intensity and pattern can encode obstacle proximity: rapid vibrations indicate imminent collision while slower pulses indicate distant obstacles. An assistive cane outfitted with HC-SR04 ultrasonic sensors and vibration-based obstacle detection uses an Arduino UNO to process proximity data from sensors positioned to cover a 180-degree frontal detection field, converting distance measurements into vibration patterns.

The integration of bone-conduction headphones (as used in SnapStick) offers a significant advantage over conventional earphones: the user's ears remain open to environmental sounds (traffic, conversations, echoes), which are critical for safe navigation.

5.3 Tactile Substitution

Sensory substitution devices (SSDs) transform visual information into tactile patterns that can be perceived through the skin. The T-Sight vibrotactile SSD, evaluated with 25 visually impaired participants, demonstrated that tactile feedback can reduce reliance on the white cane and enable earlier obstacle avoidance maneuvers. While tactile substitution requires user training to interpret the encoded patterns, it offers the advantage of continuous, hands-free information delivery.

6. Implementation Challenges and Solutions

6.1 Computational Constraints on Mobile Devices

Real-time object detection and depth estimation impose substantial computational demands that must be balanced against battery life constraints. Solutions include:

- **Model quantization**: Converting model weights from 32-bit floating point to 8-bit integers reduces memory footprint and accelerates inference with minimal accuracy loss
- **Selective frame processing**: Not every frame requires full detection; motion-based triggering can reduce processing load by 50% in static environments
- **Cloud-edge hybrid architecture**: Complex scene understanding tasks can be offloaded to cloud servers when connectivity permits, while critical obstacle detection remains local

The DrishT system demonstrates that effective performance can be achieved on mid-range Android devices (Snapdragon 660+) with 182ms end-to-end latency, proving that high-quality assistive navigation does not require flagship hardware.

6.2 Environmental Variability

Navigation systems must perform reliably across diverse environmental conditions that affect sensor performance:

- **Low light**: Reduces camera image quality; LiDAR and ultrasonic sensors are unaffected by lighting
- **Rain and glare**: Can confuse visual detection; sensor fusion provides redundancy
- **Crowded spaces**: High object density increases detection complexity; prioritization of immediate obstacles over distant ones reduces cognitive load
- **GPS-denied indoor environments**: Require alternative positioning using visual SLAM or inertial navigation

The ECEHLY-V2 model addresses these challenges through enhanced object detection and tracking mechanisms, providing strong obstacle detection and tracking abilities to promote user safety and mobility in real-life situations.

6.3 Latency and Safety

The acceptable latency threshold for collision avoidance is debated in the literature, but general consensus suggests that end-to-end latency under 300 milliseconds is necessary to allow users sufficient reaction time when walking at normal speeds. A wearable binocular sensor system achieved 310 ms cloud-inference latency with end-to-end alert delivery under 1.2 seconds and 97.8% obstacle-detection reliability under real-world indoor and outdoor conditions.

Real-time path planning algorithms must balance optimality against computational cost. PathFinder's approach of identifying the longest obstacle-free route rather than computing globally optimal paths achieves real-time responsiveness without compromising safety.

7. DISCUSSION AND FUTURE DIRECTIONS

7.1 Integration of Large Language Models

The emergence of multimodal large language models (MLLMs) presents new opportunities for contextual scene understanding beyond simple obstacle detection. These models can answer open-ended questions about the environment ("What is on my left side?" "Is there a bench nearby where I can rest?") and provide descriptive scene summaries rather than merely announcing object labels. However, current MLLM response latency (often 1-3 seconds) remains too high for primary navigation assistance. The NaviGPT system addresses this by using location and sensor data to compensate for LLM response delays, offering a hybrid approach that maintains real-time obstacle detection while providing richer contextual information when latency permits.

7.2 Edge-AI for Privacy and Independence

Cloud-dependent systems raise privacy concerns (users may not want to upload video of their homes to remote servers) and fail in areas without internet connectivity. Edge-AI running all processing locally on the smartphone addresses both concerns. The PathFinder system demonstrates that offline-only operation is feasible, achieving reliable performance without any internet connection. Future work should focus on further optimizing model efficiency to enable more sophisticated scene understanding on edge devices.

7.3 Semantic Scene Understanding

Current systems detect individual objects but lack understanding of the semantic relationships between objects. An obstacle detection system might identify a chair, a table, and a doorway, but cannot infer that the chair is blocking the path to the doorway. The DrishT system addresses this limitation through a semantic fusion layer that establishes spatial-textual correspondences using distance-weighted attention, achieving 84.3% accuracy in associating detected objects with embedded text labels. Future systems should build upon this foundation to provide true scene-level understanding.

7.4 User-Centered Design and Long-Term Adoption

Technical performance alone does not guarantee user adoption. The analysis of app reviews from blind and low-vision users reveals that usability issues including complex interfaces, unreliable performance in edge cases, and lack of integration with users' existing mobility strategies are common barriers to sustained use. Future systems should involve visually impaired users throughout the design process, provide customization options to accommodate different levels of residual vision and mobility preferences, and integrate seamlessly with existing assistive technologies rather than requiring users to adopt entirely new workflows.

8. CONCLUSION

This research paper has presented a comprehensive analysis of the technological foundations, architectural frameworks, and performance evaluation methodologies for smartphone-based navigation and obstacle detection applications for visually impaired users. The key findings of this study are as follows:

First, the integration of YOLO family object detection algorithms with sensor fusion techniques combining RGB cameras, depth sensors, LiDAR, ultrasonic sensors, and IMUs provides a robust foundation for real-time obstacle detection and distance estimation. Recent advances, particularly YOLOv8 and YOLOv11 with stereo vision integration, have achieved detection accuracies of 94-96% with depth measurement errors as low as 3.6%.

Second, user studies with blind and low-vision participants have validated the practical utility of these systems. The SnapStick system achieved an SUS score of 84.7% with 94% detection accuracy, while PathFinder demonstrated that 73% of users could learn the system in under one minute. These results indicate that modern AI-powered navigation assistance is not only technically feasible but genuinely usable by the target population.

Third, the mathematical framework for obstacle distance estimation and sensor fusion presented in this paper provides a foundation for future system development. The weighted averaging approach based on sensor inverse variance enables optimal fusion of multiple distance estimates, while Kalman filtering ensures smooth tracking of moving obstacles.

Fourth, significant challenges remain. Computational constraints on mobile devices must be balanced against battery life. Environmental variability including low light, rain, crowded spaces, and GPS-denied indoor environments continues to challenge system robustness. Latency requirements for collision avoidance (under 300 milliseconds) demand continued optimization of detection pipelines.

The democratization of assistive technology through smartphone-based solutions represents a paradigm shift in accessibility. By enabling deployment on devices costing under $200comparedtoproprietarysolutionsexceeding$3,500, modern navigation assistance systems can reach the users who need them most regardless of economic circumstances. The future of assistive navigation lies in edge-AI integration, semantic scene understanding, and multimodal feedback optimization. As these technologies mature, they promise to transform the lived experience of visually impaired individuals, enabling safe, independent, and dignified navigation through an increasingly complex world.

REFERENCES

1. World Health Organization. (2023). *Vision Impairment and Blindness*. Global estimates indicate at least 2.2 billion people have near or distance vision impairment, with approximately 285 million classified as blind or severely visually impaired.
2. Pfitzer, N., et al. (2025). MR.NAVI: Mixed-Reality Navigation Assistant for the Visually Impaired. *arXiv:2506.05369*. Over 43 million people worldwide live with severe visual impairment, facing significant challenges in navigating unfamiliar environments. The MR.NAVI system combines computer vision algorithms for object detection and depth estimation with natural language processing.
3. Cheraghpour Samavati, F., & Rahimi Ghasem Abadi, A. (2025). Assistive Technologies for the Visually Impaired: A Review. *Cureus Journal of Computer Science*. This systematic review of 80 studies highlights the increasing role of deep learning, multimodal integration, and wearable devices in enhancing navigation and accessibility.
4. Zhang, H., et al. (2025). Enhancing the Travel Experience for People with Visual Impairments through Multimodal Interaction: NaviGPT. *GROUP ACM SIGCHI*. NaviGPT integrates LiDAR-based obstacle detection, vibration feedback, and LLM responses to provide comprehensive real-time navigation aid for people with visual impairments.
5. Computer Vision-Based Obstacle Detection Mobile System for Visually Impaired Individuals. (2025). *Multimodal Technologies and Interaction*, 9(5), 48. The YOLO-300 model achieved mAP of 0.42 on a proprietary dataset of 7,600 diverse images.
6. Navigation Assistive System Using YOLOv11 and SLAM Architectures. (2025). *IEEE Xplore*. The proposed framework achieved localization accuracy of 98.9% with improved stability in highly dynamic environments.
7. SGBM_YOLO: A high-precision obstacle detection algorithm for assistive navigation. (2025). *Elsevier*. The algorithm achieved 3.9% and 4.2% improvements in mAP over YOLOv5m and YOLOv8m respectively, with depth measurement error of only 3.6%.

8. Enhanced Blind Navigation using YOLO and Sensor Fusion. (2025). *Zenodo*. This system integrates YOLO object detection with ultrasonic sensors and voice feedback, demonstrating effectiveness in detecting objects at varying distances.
9. SnapStick: Merging AI and Accessibility to Enhance Navigation for Blind Users. (2025). *Technologies*, 13(7), 297. Eleven blind participants evaluated the system, achieving 94% accuracy and an SUS score of 84.7%.
10. Das, D., et al. (2025). Real-Time Assistive Navigation for the Visually Impaired: A Scalable Approach for Indoor and Outdoor Mobility. *arXiv:2504.20976v2*. Usability study with 15 BLV participants showed 73% learned the system in under a minute and 80% praised its accuracy.
11. Lotey, D. S., & Kulkarni, N. (2026). DrishT: A Dual-Stream Mobile Architecture for Real-Time Scene Understanding in Assistive Navigation. *Authorea*. The system achieves 41.2% mAP@0.5 for object detection, 91.5% word-level accuracy for text recognition, with 182ms latency on mid-range Android devices.
12. Suvarna, A., Chikhale, A., & Mehendale, N. (2025). NavSense: Real-Time Spatial Awareness for the Visually Impaired Using LiDAR, Depth Cameras, and AI-based Object Detection. *SSRN*. The system integrates YD LiDAR for 360° mapping, TF-Luna LiDARs, and Intel RealSense D435i depth camera with custom YOLOv5-based detection.
13. Enhancing navigation and obstacle avoidance with a vibrotactile device as secondary electronic travel aid. (2024). *Disability and Rehabilitation: Assistive Technology*. Study with 25 visually impaired individuals showed reduced cane touches and earlier obstacle avoidance.
14. Comparative analysis of YOLOv8 techniques for distance perception in blind navigation systems. (2025). *IJECE*. YOLOv8+CAW achieved detection accuracy of 95.7%, demonstrating effective computer vision support for blind navigation.
15. Wearable Binocular Sensor System for Visual Guidance of Blind People. (2025). *IJCNIS*. The system achieved 94.7% average object-recognition accuracy, 310 ms cloud-inference latency, and 97.8% obstacle-detection reliability.
16. AIoT Smart Eyewear with Real-time Object and Audio Recognition for Visually Impaired Users. (2026). *River Publishers*. Achieved object detection accuracy of 94.93%, MAE of 0.34 cm, and RMSE of 0.44 cm in distance estimation.