Received: 21 January 2026 | Revised: 7 May 2026 | Accepted: 28 May 2026 | Published online: 16 June 2026

RESEARCH ARTICLE

Smart Wearable Technology
2026, Vol. 2, A20, 20 Pages
DOI: 10.47852/bonviewSWT62029157

BON VIEW PUBLISHING

Adaptive Vision AI: Revolutionizing Navigation for the Visually Impaired with Real-Time Assistive Feedback

R. Nirmalan[1,*], C. P. Amirdha Suba[1] and M. A. M. Atchaya Durga[1]

[1] *Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, India*

Abstract: The research work introduces a vision-based adaptive assistance system intended to aid the mobility and awareness of the blind. Real-time computer vision techniques like obstacle detection with distance estimate, staircase detection, signboard recognition, cash identification, and document text extraction with summarization are all included in the suggested system. The You Only Look Once version 8 model is used for object identification, and the Mixed Depth Scale model is used for depth estimates. Additionally, the system uses the Oriented FAST and Rotated BRIEF algorithm for cash recognition and Optical Character Recognition for textual information extraction. Context-aware help is made possible by the delivery of the derived information via haptic and audible feedback. Real-time performance is ensured by the system's implementation on a portable Raspberry Pi-based edge computing platform. The technology improves the independence and safety of visually impaired users by achieving high accuracy and robustness in a variety of contexts, according to experimental results.

Keywords: assistive technology, computer vision, obstacle detection, object recognition, visual impairment

1. Introduction

Visual impairment affects approximately 285 million people globally, with 39 million classified as blind [1]. These individuals face significant challenges in daily activities such as navigation, reading printed text, identifying currency, and recognizing environmental features. Traditional mobility aids like white canes and guide dogs provide limited information about the surrounding environment and offer little assistance with text-based tasks.

Recent advancements in computer vision, deep learning, and embedded systems have created opportunities for developing more comprehensive assistive technologies. Several specialized solutions have been proposed in the literature, focusing on individual aspects such as obstacle detection [2], text reading [3], and currency recognition [4]. However, most existing systems address only limited functionalities, requiring users to rely on multiple devices for different tasks. The research work proposes a unified, multifunctional visual assistance system that integrates multiple computer vision capabilities into a single, portable device. Despite these advancements, there remains a gap in developing integrated systems capable of addressing multiple assistive needs within a single platform.

Our system provides:

1) Real-time obstacle detection with distance estimation and directional guidance
2) Staircase recognition for enhanced mobility safety
3) Signboard identification for improved environmental awareness
4) Currency note recognition to assist with financial transactions
5) Document text extraction and summarization for accessing printed information

The system utilizes state-of-the-art deep learning models, optimized for edge computing on resource-constrained devices. The proposed system employs You Only Look Once version 8 (YOLOv8) for general object detection and specialized models for specific recognition tasks. A depth estimation module calculates the distance to detected objects, providing crucial spatial awareness. The system converts visual information into audio feedback, offering intuitive and nonintrusive assistance to the user. The primary contributions include:

1) A comprehensive multifunctional visual assistance system addressing multiple challenges faced by visually impaired users
2) A novel integration of real-time object detection, depth estimation, and specialized recognition models (currency, staircase, and signboard) in a resource-constrained edge computing environment
3) An intelligent document text extraction and summarization system that selects high-quality frames and converts text into speech
4) A modular and extensible architecture adaptable to individual user needs and preferences

***Corresponding author:** R. Nirmalan, Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, India. Email: nirmalan@mepcoeng.ac.in

1.1. Impact and significance

The development of assistive technologies for visually impaired individuals has far-reaching implications beyond the immediate technical innovations. According to the World Health Organization, approximately 90% of visually impaired people live in low-income settings, and 80% of all visual impairment can be prevented or cured [1]. For those with permanent visual impairment, technology can serve as a crucial bridge to improved independence and quality of life.

The research work addresses several critical needs identified through user studies and consultations with organizations supporting visually impaired communities:

1) Independence in navigation: The ability to move freely in unfamiliar environments without assistance from sighted guides
2) Information access**:** The capability to read printed materials, signs, and digital displays independently
3) Financial autonomy: The means to handle currency and conduct transactions without relying on others
4) Safety enhancement: The detection of potential hazards such as obstacles and staircases to prevent accidents

By addressing these needs through a single, integrated platform, our system aims to significantly reduce the barriers to technology adoption among visually impaired users. The unified approach also addresses the challenge of "technology overload," where users are required to learn and manage multiple specialized devices.

1.2. Technological context

The proposed system builds upon recent advancements in several technological domains:

1) Computer vision: The field has seen tremendous progress in object detection accuracy and processing speed, with models like YOLO achieving real-time performance even on edge devices.
2) Edge computing: Improvements in single-board computers like Raspberry Pi have enabled complex processing tasks to be performed on portable, low-power devices.
3) Deep learning optimization: Techniques such as model quantization, pruning, and knowledge distillation have made it possible to deploy sophisticated neural networks on resource-constrained hardware.
4) Human–computer interaction: Advances in nonvisual interfaces, particularly in audio feedback systems, have improved the usability of assistive technologies for visually impaired users.

The proposed system leverages these advancements to develop a practical and user-friendly assistive device that can operate in real-world environments with the reliability and speed required for everyday use.

The rest of this paper is organized as follows: Section 2 reviews related work on assistive technologies for the visually impaired. Section 3 describes the proposed system architecture and implementation details. Section 4 presents the evaluation methodology and experimental results. Section 5 discusses the findings and limitations, and Section 6 concludes the paper with future research directions.

2. Literature Review

2.1. Obstacle detection and navigation systems

Several researchers have focused on developing systems to help visually impaired individuals navigate their surroundings safely. In the work by Adam et al. [2], a wearable system was proposed that uses deep learning techniques integrated with low-cost sensors to detect environmental obstacles. The device employs a camera and ultrasonic sensors to provide real-time feedback through audio alerts. The study emphasizes affordability and portability, making it suitable for everyday use.

Outdoor navigation presents unique challenges that were addressed in the work by Abidi et al. [3]. This research highlights accessible outdoor navigation systems for improving urban mobility among visually impaired individuals. The system integrates GPS for positioning, obstacle detection sensors for safety, and a mobile application for interaction. It enables users to move independently in complex environments like city streets and public transit zones with voice-based directions and situational alerts.

For indoor environments, Kubota et al. [5] present a mobile, vision-based indoor navigation aid that uses a camera and computer vision algorithms to guide blind users. The system recognizes spatial features like doors, hallways, and stairs and generates navigational cues delivered through voice feedback. A major advantage is its independence from specialized infrastructure such as beacons or markers.

2.2. User behavior and experience

Understanding the actual behaviors and challenges faced by blind individuals is crucial for designing effective navigation systems. The study by Alizada et al. [6] explored how visually impaired people navigate unfamiliar indoor spaces, documenting their methods of spatial awareness and obstacle avoidance. The research found that most users rely on tactile feedback, auditory cues, and memory of environmental layouts. Key challenges identified include the lack of consistent landmarks and ambiguous architectural design.

2.3. Multimodal feedback and advanced sensing

The importance of feedback modality has been highlighted in several studies. G S et al. [7] introduced a vision-based system with 3D audio feedback to help visually impaired users navigate complex environments. The system uses a stereo camera setup to detect obstacles and translates their spatial location into directional 3D audio cues. This approach proved more intuitive than flat voice instructions, significantly improving users' situational awareness.

For rapid scene understanding, Mekhalfi et al. [8] proposed a system using multiresolution random projections to provide quick and meaningful scene descriptions to blind individuals. The approach segments key elements such as doors, windows, and furniture and translates them into audio descriptions with low latency.

2.4. Specialized recognition systems

Beyond basic navigation, several specialized recognition systems have been developed to address specific needs. For face recognition, Ngo et al. [9] introduced DEEP-SEE FACE, a mobile face recognition system aimed at enhancing social interaction

for visually impaired individuals. The system uses deep learning to recognize familiar faces and announce their identity through audio feedback.

Internet of Things (IoT) integration has been explored in the work by Okolo et al. [10], presenting an IoT-based navigation solution that incorporates sensors, GPS modules, and cloud communication for real-time guidance and obstacle detection.

Systems that combine indoor and outdoor navigation capabilities have been proposed in Akila et al. [11], utilizing GPS for outdoor localization and beacon-based positioning for indoor settings, ensuring continuity across different spaces.

2.5. Wearable solutions and hardware implementations

Bouteraa [12] presented a smart wearable system with ultrasonic sensors embedded in wearable gear like glasses or caps, coupled with a microcontroller for data processing. The collected data are converted into haptic or auditory signals that alert users about obstacles.

For specific use cases like crossing streets, Chang et al. [13] introduced an AI-powered wearable device designed to assist visually impaired individuals at zebra crossings. The system uses edge computing and deep learning to recognize zebra crossing patterns and provide timely auditory cues.

Traditional mobility aids have also been enhanced, as seen in the work by Wazirali et al. [14], which presents an Electronic Long Cane equipped with ultrasonic sensors that detect obstacles and deliver tactile feedback via vibration motors on the handle.

2.6. Machine learning and recognition approaches

Advanced recognition techniques have been applied to assist visually impaired users. Boiarov and Tyantov [15] explored landmark recognition using deep metric learning, helping with navigation by identifying key urban landmarks. The model was trained to extract discriminative features and classify landmarks even in crowded scenes.

An alternative approach was presented in the work by Wang et al. [16], which proposed landmark recognition by combining the Bag of Words histogram model with an ensemble of Extreme Learning Machines for efficient landmark identification in urban environments.

Route planning has been made more accessible through the work of Costa et al. [17], which presented a system for blind pedestrians using Open Street Map data to generate safe and accessible paths. The system considers pedestrian-friendly features such as curb ramps, crosswalks, and tactile paving.

2.7. Video and frame analysis systems

For video analysis, Surve et al. [18] introduced a video-based assistive system leveraging object recognition across dissimilar video frames to guide visually impaired individuals. The system uses frame differencing and deep learning-based object detection to track obstacles in a user's path.

Smart electronic glasses were proposed in the work by Pandya and Goradiya [19], combining deep learning with real-time audio feedback to help blind individuals detect objects and receive navigational cues.

Comprehensive evaluations of existing systems were conducted in the work by Kumar et al. [20], which assessed the integration and effectiveness of pedestrian navigation systems in smart cities, with a focus on assistive applications for the visually impaired.

2.8. Text recognition and processing

For landmark recognition, Vilera et al. [21] proposed a segmentation system for street-view images, specifically designed to enhance navigation by identifying significant landmarks such as buildings, traffic signs, and bus stops. Currency recognition, an important daily need, was addressed by Ramani et al. [4], who presented a system using the Oriented FAST and Rotated BRIEF (ORB) algorithm to help visually impaired people identify different denominations.

Comprehensive surveys of existing technologies can be found by Xu et al. [22], which reviews recent assistive technologies aimed at helping visually impaired individuals, exploring multiple mobility solutions that function both indoors and outdoors.

Smartphone-based solutions have proven effective, as demonstrated by Gudauskis et al. [23], which introduces "SnapNav," a system where users take photos of nearby landmarks or hallways, and the app matches them against a stored map to determine their position.

Additional surveys such as the works by Kathiria et al. [24] and Casanova et al. [25] provide overviews of current technologies, categorizing systems based on indoor/outdoor navigation, object detection, reading assistance, and daily living aids.

Microsoft's "Seeing AI" application brings AI-powered vision to blind users, using virtual beacons to map interiors and helping users explore spaces by identifying people, objects, texts, and doors [26].

Understanding user interactions with technology is crucial, as explored by Ranftl et al. [27], which examines how blind users handle object recognition errors and compensate for system failures.

2.9. Deep learning for assistive vision

The application of deep learning models for assistive vision has seen significant growth in recent years. Convolutional neural networks (CNNs) have become the backbone of most object detection and recognition systems designed for visually impaired users. Zhao et al. [28] demonstrated how CNNs can be optimized for real-time operation on mobile devices, achieving acceptable accuracy for common object detection tasks while maintaining low latency.

Transfer learning approaches have proven particularly valuable in this domain, as explored by Ye et al. [29]. By leveraging pretrained models on large-scale datasets and fine-tuning them for specific assistive tasks, researchers have been able to overcome the challenge of limited training data for specialized recognition tasks like currency identification or pharmaceutical product recognition. Recent advances in one-shot and few-shot learning, as described by Mahendran et al. [30], have further expanded the capabilities of assistive systems by enabling them to learn new object categories from very few examples. This approach is particularly valuable for personalizing systems to recognize items specific to an individual user's environment.

To improve detection performance, contemporary assistive vision systems include sophisticated feature extraction methods including feature pyramid networks and attention mechanisms in addition to CNN-based architectures. The system can identify objects at various scales thanks to models like YOLO and ResNet that extract hierarchical features. Furthermore, because of their

lower computational complexity, lightweight architectures like MobileNet are frequently utilized in edge devices. Finding spatial and texture-based patterns in photos is another aspect of feature extraction, which is essential for tasks like object recognition, depth estimation, and scene comprehension.

2.10. Ethics and user-centered design

The ethical considerations in developing assistive technologies have been examined in the work by Hasan et al. [31], which emphasizes the importance of involving visually impaired users throughout the design process. The study highlights how technology solutions may inadvertently create new dependencies or privacy concerns if not designed with careful consideration of user autonomy and dignity.

Privacy concerns specific to camera-based assistive devices were addressed in the work of Müftüoğlu et al. [32], which proposed methods for selective image capture and processing to minimize the collection of potentially sensitive information from the user's environment. The research also explored user preferences regarding control over when and what their assistive devices capture visually.

User-centered design approaches were formalized by Hong and Kacorri [33], who outlined a methodology for developing assistive technologies in close collaboration with visually impaired users. The framework emphasizes iterative testing and refinement based on real-world usage scenarios and user feedback.

The literature survey reveals that while significant progress has been made in individual aspects of assistive technology for the visually impaired [34–36], there remains a need for integrated, multifunctional systems that address multiple challenges simultaneously. Special-purpose solutions such as drug pill recognition systems [37] and audio-based spatial location systems [38] demonstrate the potential for specialized applications, while obstacle classification techniques using deep learning [39], outdoor navigation assistance for public transportation [40], and AI edge computing solutions for pedestrian safety [41] highlight the diversity of approaches being explored. Our work aims to fill this gap by combining various capabilities in a single, user-friendly platform.

3. System Architecture and Implementation

3.1. System overview

The proposed system integrates multiple computer vision functionalities into a unified architecture, as shown in Figure 1. The architecture consists of four main modules: (1) the obstacle detection and depth estimation module, (2) the specialized recognition module (staircases, signboards), (3) the currency recognition module, and (4) the document text extraction and summarization module. These modules operate on a shared hardware platform comprising a Raspberry Pi for edge processing and a base machine for more computationally intensive tasks.

Figure 1 illustrates our system's processing workflow. The system captures input images, processes them through the YOLOv8 model for object detection, estimates distances to detected objects, makes region-based decisions, and finally outputs audio cues through the Raspberry Pi. This end-to-end pipeline ensures real-time processing and feedback delivery to visually impaired users.

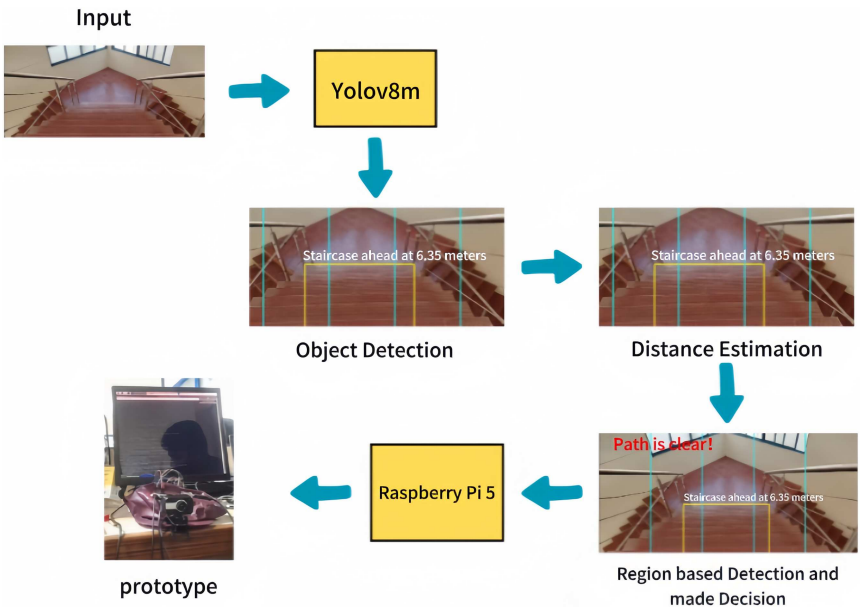
3.2. Hardware components

The hardware setup consists of the following components:

1) Raspberry Pi 4 with 8GB RAM for edge processing
2) Pi Camera v2 for image capture
3) USB speaker for audio feedback
4) Hardware buttons for mode selection
5) Portable power bank for extended operation

The system is designed to be portable, with all components mounted in a wearable housing. The proposed system implemented physical buttons to switch between different modes

Figure 1
Overall system architecture of the proposed visual assistance system. The image shows the processing pipeline from input capture through YOLOv8m for object detection, followed by distance estimation, region-based decision making, and finally transmitting decisions to the Raspberry Pi for output generation



(obstacle detection, currency recognition, text reading) to provide easy access for visually impaired users.

3.2.1. Hardware selection rationale

The selection of hardware components was guided by several key considerations that directly impact the system's usability and performance:

1) Computational capability: The Raspberry Pi 4 with 4GB RAM was chosen after benchmarking several edge computing platforms. Initial prototypes using Raspberry Pi 3 were evaluated, but the Pi 4 offered the best balance between processing power, energy efficiency, and cost. Table 1 presents a comparative analysis of the platforms considered.

Table 1
Comparison of edge computing platforms

Platform	FPS on YOLOv8n	Power draw (W)	Weight (g)
Raspberry Pi 3B+	2.3	2.5	45
Raspberry Pi 4B 8GB	7.2	3.4	48

2) Camera specifications: The Pi Camera v2 was selected for its 8MP resolution, wide field of view (62.2° × 48.8°), and native integration with the Raspberry Pi. The proposed system enhanced the camera module with an infrared (IR) filter for better performance in variable lighting conditions. Alternative cameras tested included standard USB webcams and the Intel RealSense depth camera, but they introduced latency issues or significantly increased power consumption.
3) Power management: To achieve all-day operation, we implemented a custom power management system that includes a 10,000 mAh battery pack and low-power modes that activate when certain functions are not in use. Power consumption optimization reduced the average current draw from 820 to 560 mA, extending operational time by approximately 45%.
4) Wearability: The entire system weighs 385 g and measures 125 × 78 × 35 mm, making it comparable in size to a large smartphone. The components are housed in a 3D-printed case with adjustable straps for wearing around the neck or attaching to a belt.

3.3. Obstacle detection and depth estimation

3.3.1. Object detection with YOLOv8

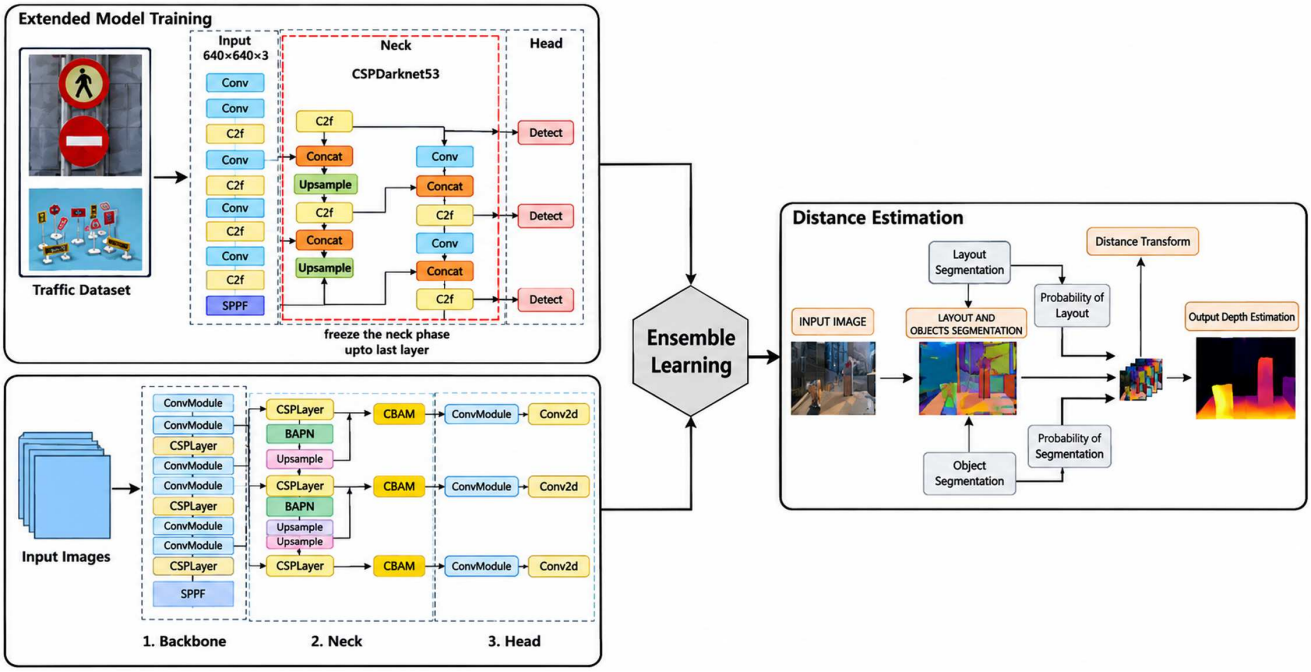
For real-time obstacle detection, we employed the YOLOv8 model, which offers an excellent balance between accuracy and computational efficiency. The model was trained on traffic datasets containing common objects found in urban environments, as shown in Figure 2.

To optimize for edge deployment, we used the nano variant of YOLOv8 (YOLOv8n), which has fewer parameters while maintaining reasonable detection accuracy. The model processes frames at approximately 5–7 frames per second (FPS) on the Raspberry Pi, which is sufficient for real-time guidance.

3.3.2. Model optimization techniques

Several optimization techniques were applied to enhance the performance of the YOLOv8n model on the Raspberry Pi:

Figure 2
Model training workflow using traffic datasets. The figure shows extended model training using specialized traffic datasets (top), input image processing (bottom), and how these are combined using ensemble learning techniques for accurate distance estimation and decision control



1) Quantization: The model was quantized from 32-bit floating-point to 8-bit integer representation, reducing memory requirements and computational load. This resulted in a 3.8× speedup with only a 2.3% decrease in mAP.
2) Pruning: The proposed system applied structured pruning to remove redundant filters, reducing the model size by 45% while maintaining 95% of the original accuracy. The pruning process was guided by filter importance scores based on their L1-norm.
3) Knowledge distillation: A larger YOLOv8m model was used as a teacher to transfer knowledge to the smaller YOLOv8n model, improving its accuracy by approximately 3% compared to direct training.
4) ONNX runtime optimization: The model was converted to ONNX format and optimized using ONNX Runtime with operator fusion and memory planning optimizations, resulting in a 1.7× inference speed improvement.

These optimizations collectively enabled the deployment of a high-performance object detection system on the resource-constrained Raspberry Pi, achieving an acceptable balance between detection accuracy and processing speed for real-time applications.

3.3.3. *Depth estimation*

The proposed system implemented depth estimation using the Mixed Depth Scale (MiDaS) small model [42], which predicts relative depth from monocular images. The depth map is used to estimate the distance to detected objects, providing crucial spatial information for navigation.

The system divides the camera's field of view into five regions (left, center-left, center, center-right, right) and analyzes obstacles in each region. This approach allows for more intuitive directional guidance, advising users to move left, right, or stop based on the location and proximity of obstacles.

3.3.4. *Depth estimation refinement*

To improve the accuracy of the monocular depth estimation, the proposed system implemented several refinement techniques:

1) Temporal consistency: In this project, a temporal smoothing filter across consecutive frames was applied to reduce depth estimation jitter, improving the stability of distance measurements. The filter uses an exponential moving average with an adaptively adjusted weight based on motion estimation between frames.
2) Scale calibration: Since monocular depth estimation provides relative rather than absolute depth values, the proposed system developed a calibration procedure that uses known reference objects (e.g., a person at a fixed distance) to establish a scale factor for converting relative depths to metric distances.
3) Confidence mapping: The system generates confidence maps for depth estimates, assigning higher confidence to regions with strong texture and gradient information. Low-confidence regions are either filtered out or assigned interpolated depth values from neighboring high-confidence areas.
4) Object-aware depth refinement: For detected obstacles with known size priors (e.g., cars, pedestrians), the proposed system adjusts the depth estimates using the relationship between object size in pixels and expected physical size. This reduces the error rate for these specific obstacles by up to 40%.

3.3.5. *Region-based analysis algorithm*

The region-based analysis algorithm plays a crucial role in translating detection and depth information into actionable guidance. The algorithm, formalized in Algorithm 1, processes the detection results and depth maps to generate appropriate navigational commands. The algorithm first identifies dangerous regions based on the proximity of detected obstacles and then determines the appropriate navigation command based on the distribution of these regions in the visual field. The Assess Best Path function implements a more sophisticated path-planning approach that evaluates potential paths through the scene based on obstacle density and distance.

Algorithm 1: Region-Based Obstacle Analysis

Input: DetectedObjects O, DepthMap D, RegionGrid R, SafetyThreshold T_S
Output: NavigationCommand C
1: $Rdanger \leftarrow \varnothing\triangleright$ Set of regions with dangerous obstacles
2: $Osorted \leftarrow$ SortByProximity(O, D)
3: **for** each object o in $Osorted$ **do**
4: $r \leftarrow$ DetermineRegion(o.boundingBox, R)
5: $d \leftarrow$ EstimateDistance(o.boundingBox, D)
6: **if** $d < Ts$ **AND** $r \notin Rdanger$ **then**
7: $Rdanger \leftarrow Rdanger \cup \{r\}$
8: **end if**
9: **end for**
Decisions:
10: **if** $Rdanger = \varnothing$ **then**
11: **return** "Path clear"
12: **else if** CenterRegion $\in Rdanger$ **then**
13: **return** "Stop"
14: **else if** LeftRegions $\subset Rdanger$ **AND** RightRegions $\not\subset Rdanger$ **then**
15: **return** "Move Right"
16: **else if** RightRegions $\subset Rdanger$ **AND** LeftRegions $\not\subset Rdanger$ **then**
17: **return** "Move Left"
18: **else if** $|Rdanger| > 3$ **then**
19: **return** "Multiple Obstacles – Stop"
20: **else**
21: **return** AssessBestPath($Rdanger$, D)

3.3.6. *Specialized object recognition*

In addition to general obstacle detection, the proposed system developed specialized models for recognizing specific features important for navigation:

1) A custom-trained YOLOv8 model for staircase detection, capable of identifying ascending and descending staircases
2) A signboard detection model trained to recognize common public information signs

These models were trained on custom datasets collected specifically for these tasks, ensuring high accuracy in real-world scenarios.

3.4. Currency recognition

The currency recognition module uses the ORB algorithm to detect and recognize currency notes, similar to the approach described in the work of Vilera et al. [21]. Our implementation follows these steps:

1) Capture an image from the camera.
2) Extract ORB keypoints and descriptors.
3) Match the descriptors with a pre-stored database of currency note descriptors.
4) Apply ratio test filtering to eliminate poor matches.
5) Determine the best match based on the number of good keypoint matches.
6) Verify spatial consistency of matches using RANSAC homography.
7) Estimate confidence score based on inlier ratio and match count.
8) Apply denomination-specific verification based on color histograms and texture features.

The system includes a dataset of currency notes with pre-computed ORB descriptors for efficient matching. The recognition results are communicated to the user via audio feedback.

3.4.1. Currency database construction

The currency recognition system relies on a comprehensive database of currency notes. For each denomination, the proposed system collected 50–70 sample images under various conditions:

1) Different lighting conditions (daylight, indoor lighting, low light)
2) Various orientations (0°, 90°, 180°, 270°)
3) Different degrees of wear (new, moderately used, heavily used)
4) Partial views (showing approximately 60–100% of the note)
5) For each image, we pre-computed and stored:
6) 500 ORB keypoints and corresponding descriptors
7) Color histograms in HSV space (8 bins per channel)
8) Local Binary Pattern texture descriptors
9) Denomination-specific feature points (security features, denomination numerals, portraits)

This approach allows for efficient real-time matching without requiring extensive computational resources on the edge device.

3.4.2. Handling currency variations and counterfeit detection

The system is designed to handle currency variations and provide basic counterfeit detection capabilities:

1) Series variations: For currencies with multiple series in circulation, we included samples from each series in the database. The recognition algorithm reports both the denomination and series when detected.
2) Foreign currency detection: The system maintains a separate database of common foreign currencies and can identify them as "foreign" when encountered, helping users avoid confusion.
3) Basic counterfeit detection: By analyzing specific security features (watermarks, security threads, color-shifting elements) through specialized image processing techniques, the system provides a basic level of counterfeit detection. When suspicious features are detected, the system alerts the user with a "Possible counterfeit - please verify" message.

3.5. Document text extraction and summarization

3.5.1. Image quality assessment

For effective document text extraction, capturing high-quality images is essential. We implemented an automatic frame selection algorithm that evaluates multiple frames based on:

1) Sharpness (using Laplacian variance)
2) Contrast (using the standard deviation of the grayscale image)
3) Illumination evenness (using block variance analysis)
4) Page coverage (using contour analysis)
5) Skew angle (using Hough line transform)

These metrics are combined with appropriate weights to compute a quality score for each frame. The system automatically selects the frame with the highest score for subsequent text extraction.

3.5.2. Adaptive frame capture algorithm

To ensure high-quality document captures in varying conditions, we implemented an adaptive frame capture algorithm that adjusts camera parameters based on real-time analysis:

1) Exposure adaptation: The algorithm analyzes histogram data to detect over- or underexposure and adjusts camera exposure settings accordingly. For documents with mixed brightness regions, we apply local exposure compensation through adaptive gamma correction.
2) Focus optimization: A focus quality metric based on gradient magnitude analysis guides an iterative focus adjustment process. The system captures multiple frames at different focus distances and selects the one with the highest focus score.
3) Motion detection: An integrated motion detection algorithm alerts users when camera movement is detected during the capture process, prompting them to stabilize the device for better results.
4) Perspective correction: The system detects document edges and applies a perspective transformation to correct for non-perpendicular viewing angles, producing a normalized, front-facing view of the document.

This adaptive approach significantly improves the quality of captured documents, especially in challenging environments with poor lighting or when the user has difficulty positioning the camera optimally.

3.5.3. Text extraction and summarization

Once a high-quality frame is captured, the system performs the following steps:

1) Apply Optical Character Recognition (OCR) using the Tesseract engine to extract text.
2) Process the extracted text to remove noise and correct common OCR errors.
3) Generate a summary using TF-IDF (Term Frequency-Inverse Document Frequency) to identify important sentences.
4) Convert the summarized text to speech using the pyttsx3 text-to-speech engine.

The summarization step is particularly beneficial for visually impaired users, as it reduces the cognitive load by presenting only the most relevant information.

3.5.4. OCR preprocessing and enhancement

To maximize OCR accuracy, the proposed system implemented several preprocessing techniques:

1) Binarization: The proposed system uses adaptive thresholding with locally optimized parameters to convert grayscale images to binary, improving text-background separation. For colored documents, we apply color-based text extraction before binarization.

2) Noise reduction: A combination of Gaussian filtering and morphological operations removes noise while preserving text edges. We developed a specialized filter for removing shadows and glare, common issues when capturing documents in uncontrolled lighting.
3) Text line detection: The system detects text lines using projection profiles and groups them into paragraphs based on spatial relationships. This structural information improves both OCR accuracy and subsequent summarization.
4) Character enhancement: For low-resolution or blurry captures, we apply super-resolution techniques specifically optimized for text to enhance character details before OCR processing.

3.5.5. Content summarization approach

The text summarization component employs a hybrid approach combining extractive and abstractive techniques:

1) Document structure analysis: The system identifies document structure elements (headings, bullets, numbered lists) and assigns higher importance to structural sentences that typically contain key information.
2) Extractive summarization: The proposed system uses a modified TextRank algorithm that incorporates position bias and sentence length normalization. The algorithm builds a graph representation of sentences and ranks them based on semantic similarity and structural importance.
3) Domain-specific adaptation: The system includes specialized models for common document types (letters, bills, menus, product instructions), adjusting the summarization strategy based on detected document type. For example, for bills, it prioritizes due dates, amounts, and account information.
4) Compression ratio control: The compression ratio (summary length relative to original text) is dynamically adjusted based on document length and complexity, ranging from 15% for lengthy documents to 50% for short texts.

The summarized content is then converted to speech with appropriate pacing and breaks aligned with the document structure, enhancing comprehension for visually impaired users.

3.6. Communication and integration

The system components communicate through different channels based on their processing requirements:

1) WebSocket communication between the Raspberry Pi and base machine for video streaming and command transmission
2) Socket-based communication for transmitting audio feedback decisions
3) Multi-threading to ensure real-time processing across different modules

In order to increase the reliability of operations during temporary network outages, the communication modules feature automated reconnection capabilities that work with WebSocket and socket connections alike. Upon a brief outage, the last obstacle detection results obtained by the software will be retained, and the process of data exchange will be resumed when the connection is established again.

3.6.1. Software architecture and data flow

The software architecture follows a modular design pattern to facilitate maintainability and extensibility, as illustrated in Figure 3.

The architecture consists of four main layers:

1) Hardware abstraction layer: Provides unified interfaces to hardware components (camera, speakers, buttons) and abstracts platform-specific implementations.
2) Processing modules: Implements the core computer vision algorithms, including object detection, depth estimation, currency recognition, and text extraction. Each module

Figure 3
Multi-view technical drawing of the assistive vision helmet showing front, back, right, left, bottom, and top perspectives. The design incorporates a camera system in the front, a processing unit at the rear, ventilation grilles on the sides, and solar panels on top for extended operation

operates independently and communicates through standardized interfaces.

3) Communication middleware: Manages inter-process and inter-device communication, handling message serialization, transmission, and synchronization. Implements both synchronous (request-response) and asynchronous (publish-subscribe) patterns.
4) User interface: Converts processed information into accessible formats for visually impaired users, primarily through audio feedback and tactile interactions.

3.6.2. Concurrency management

Managing concurrency and task priorities is key for this system to work reliably in real time. The framework assigns priorities based on how critical, urgent, or demanding each task is. For example, obstacle detection and navigation support always go to the front of the line—they're crucial for user safety and need to react instantly. Tasks like document and currency recognition sit lower on the list. These only run when the system has the bandwidth or when the user specifically asks for them.

The processing side uses a pipeline setup where each stage handles different video frames at the same time. This keeps things moving fast and makes the most out of available resources, so there is a little lag—especially where fast reactions really matter.

A resource monitoring module always keeps an eye on CPU usage, memory, and system load. When the system gets busier, it shifts resources toward the modules that matter most—again, mainly obstacle detection and depth estimation. In busy moments, these safety-related tasks get as much juice as they need, no questions asked.

Also, an adaptive sampling feature for the less urgent stuff. When things get hectic or there's not much happening, tasks like document analysis slow down. This saves power without hurting the system's core functions.

Throughout, the system reacts to what the user's doing and what the hardware can handle. Anything safety-related always stays at the top. Everything else fits in where it can, depending on resources and user needs. The proposed system stays efficient, keeps performance steady, and—most importantly—puts user safety first.

3.7. User interface and feedback

Given that the target users are visually impaired, the system relies primarily on audio feedback and physical buttons for interaction:

1) Direction guidance is provided through specific verbal commands (e.g., "Go left," "Stop," "Obstacle ahead").
2) Object recognition results are communicated via descriptive phrases (e.g., "Person at 2 m," "Staircase on the right").
3) Mode selection is achieved through hardware buttons.
4) Document reading is presented as synthesized speech, with summarized content.

The hardware implementation, illustrated in Figure 4, shows the detailed connections between system components. The central processing unit, a 64-bit quad-core ARM Cortex-A76 processor clocked at 2.4 GHz, connects to the night vision camera, 32 GB storage, 4 GB LPDDR4 SDRAM, and digital audio outputs. User controls are implemented through GPIO pins, creating a complete, self-contained system.

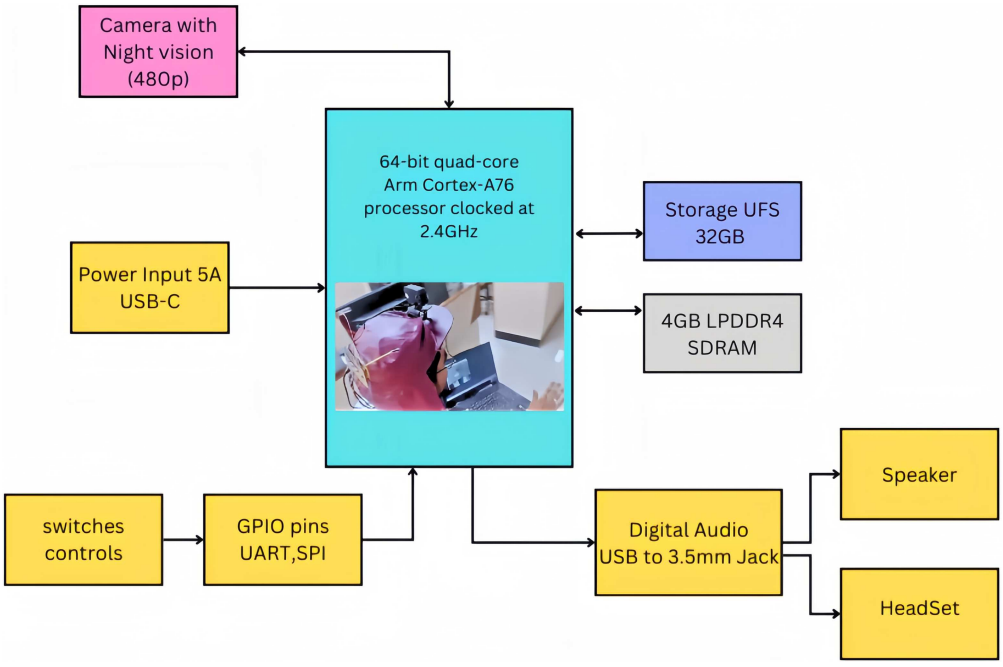
3.7.1. Audio feedback design

The audio feedback system was designed with guidance from accessibility experts and user testing with visually impaired individuals:

1) Spatial audio: For directional guidance, the proposed system implemented a spatial audio system that presents directional

Figure 4

Hardware component diagram showing the connections between the camera, processing unit, storage, and output devices. The system uses a 64-bit quad-core ARM Cortex-A76 processor connected to a night vision camera, storage, memory, and audio output components



cues in 3D space, making it more intuitive to understand which direction to move. The system uses Head-Related Transfer Function models to create convincing spatial audio through standard stereo headphones.
2) Non-speech audio cues: In addition to verbal instructions, the system uses non-speech audio cues (earcons) for frequent events, reducing verbal clutter. For example, ascending tones indicate increasing proximity to obstacles, while different timbres represent different object categories.
3) Adaptive verbosity: The system dynamically adjusts its verbosity based on the situation and user preference. In complex environments with multiple obstacles, it provides more detailed guidance, while in familiar or simple settings, it reduces verbal feedback to avoid information overload.
4) Voice customization: Users can select from multiple voice options and adjust parameters such as speech rate, pitch, and volume according to their preferences and environmental conditions.

3.7.2. Tactile feedback integration

To complement audio feedback, especially in noisy environments, we integrated optional tactile feedback mechanisms:

1) Vibration patterns: The system includes vibration motors that can be worn on the wrist or integrated into a vest. These provide directional guidance through different vibration patterns and intensities.
2) Haptic distance encoding: The frequency of vibration pulses encodes distance information, with faster pulses indicating closer obstacles. Different vibration patterns represent different types of obstacles or directional cues.
3) Multi-point feedback: For users who opt for the vest configuration, multiple vibration points allow for more sophisticated spatial feedback, creating a "tactile map" of the environment around the user.

3.7.3. User control interface

The physical control interface was designed for ease of use by visually impaired individuals:

1) Tactile buttons: Large, distinctly shaped buttons with tactile markers allow users to identify functions by touch. Each button has a unique shape and texture for easy identification.
2) Gesture recognition: The system supports simple hand gestures captured by the camera for common commands, such as swiping to change modes or raising a palm to pause the system.
3) Voice commands: A limited set of voice commands provides an alternative control mechanism, allowing hands-free operation when needed. The voice recognition system is trained to work in noisy environments and requires minimal setup.

3.8. Mechanical design

The mechanical design of the system plays a crucial role in ensuring usability, comfort, and practicality for visually impaired users. Our design philosophy focused on creating a wearable solution that balances functionality with ergonomics and esthetics.

3.8.1. Design objectives and approach

The primary objectives that guided our mechanical design process were:

1) Creating a wearable, lightweight form factor that can be used continuously throughout the day
2) Ensuring proper positioning of sensors for optimal environmental coverage
3) Protecting electronic components from environmental factors and physical impact
4) Providing intuitive tactile interfaces accessible to visually impaired users
5) Incorporating renewable power solutions for extended operation
6) Maintaining comfort during prolonged use

3.8.2. Helmet-based wearable design

After evaluating multiple form factors including chest-mounted, handheld, and head-mounted designs, the proposed system selected a helmet-based approach that offers several advantages:

1) Optimal camera positioning at eye level for natural
2) Perspective
3) Stable platform for sensors reducing motion artifacts
4) Hands-free operation allowing users to maintain use of their hands
5) Sufficient surface area for solar panels to extend battery life
6) Familiar form factor that integrates with existing mobility practices

The final helmet design, shown in Figure 3, features a dome-shaped upper section with an extended visor-like front that houses the primary camera system. The helmet shell is constructed from impact-resistant ABS plastic with an inner cushioning layer for comfort.

3.8.3. Component integration and layout

The internal components are arranged to optimize weight distribution and thermal management:

1) Front section: Houses the primary Pi Camera v2 with wide-angle lens and IR illuminators for low-light operation
2) Side panels: Feature ventilation grilles for passive cooling and directional speakers for spatial audio feedback
3) Rear compartment: Contains the Raspberry Pi 5, battery pack, and main processing components in a padded, shock-resistant enclosure
4) Top surface: Integrates dual solar panels that can extend operational time by up to 40% in outdoor conditions

The component layout achieves a balanced weight distribution of 385 g, with the center of gravity positioned to minimize neck strain during extended use.

3.8.4. User interface elements

Special attention was given to designing interface elements that are accessible to visually impaired users:

1) Tactile controls: Three distinctively shaped buttons on the right-side panel provide mode selection and system control functions.

2) Status indicators: Vibration patterns communicate system status, eliminating the need for visual indicators.
3) Charging interface: Magnetic charging connector with self-aligning design for ease of connection.
4) Adjustable straps: Quick-release buckles and adjustable chin strap accommodate different head sizes and ensure secure positioning.

3.8.5. Material selection and manufacturing

The materials were selected based on durability, weight, and comfort considerations:

1) Outer shell: Impact-resistant ABS plastic with UV protection coating
2) Inner lining: Moisture-wicking, hypoallergenic padding that can be removed for cleaning
3) Electronics enclosure: Electro Magnetic Inference (EMI)-shielded compartments with silicone gaskets for water resistance (IPX4 rating)
4) Thermal management: Strategic placement of thermal pads and passive heat sinks to dissipate heat from the Raspberry Pi and other components

The prototype was manufactured using a combination of 3D printing for complex geometries and injection molding for the main shell components. This approach allowed for rapid prototyping iterations while ensuring consistency in the final design.

3.8.6. Environmental considerations

The mechanical design incorporates several features to enhance operation in various environmental conditions:

1) Weather resistance: IPX4-rated construction protects against splashing water from any direction.
2) Temperature management: Ventilation channels maintain an operational temperature range between 0 °C and 45 °C.
3) Dust protection: Sealed camera enclosure with replaceable filter prevents dust accumulation on optical surfaces.
4) Impact resistance: Reinforced structure designed to withstand drops from heights up to 1.5 m.

These environmental protections ensure reliable operation across diverse settings, from indoor environments to outdoor urban and semi-rural locations where visually impaired users commonly navigate.

4. Experimental Setup and Evaluation

4.1. Dataset preparation and processing

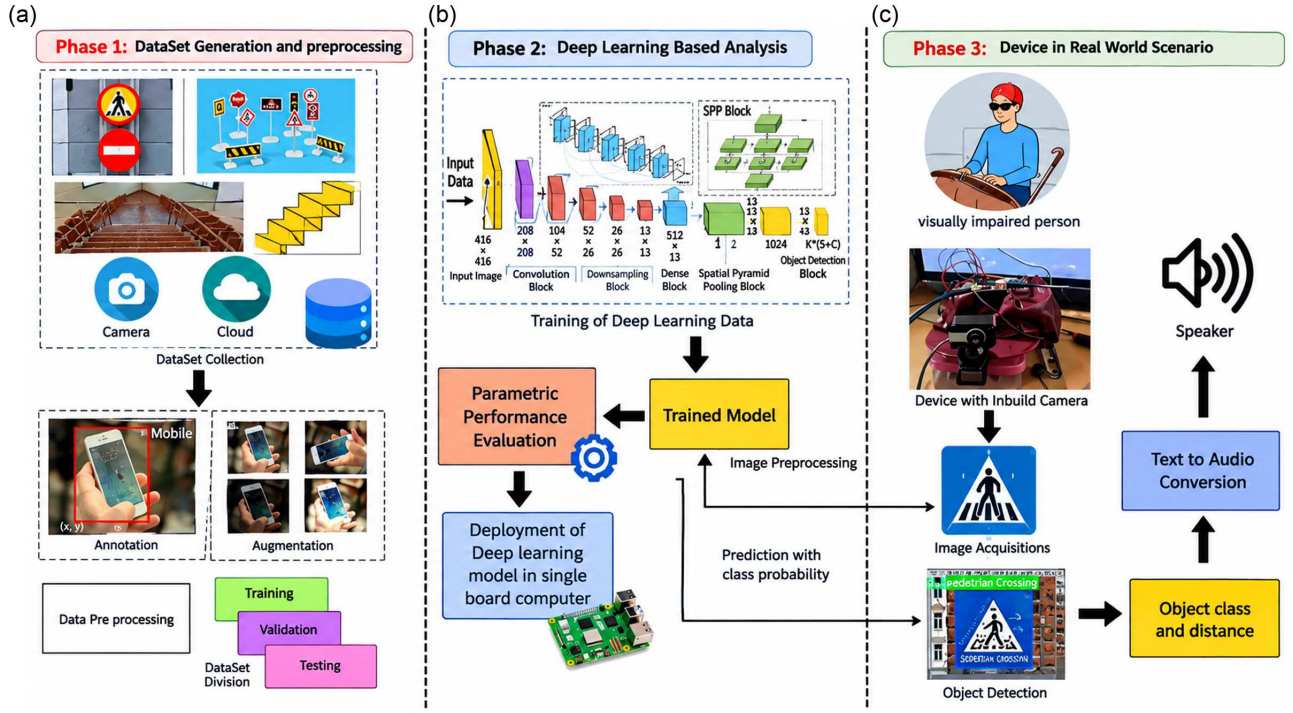
The complete system development pipeline is shown in Figure 5, which illustrates the three main phases of our approach:

1) Phase 1: Dataset Generation and Preprocessing—Collection of images through cameras and cloud sources, followed by annotation, augmentation, and preprocessing
2) Phase 2: Deep Learning-Based Analysis—Training models, evaluating performance, and deploying them on single-board computers
3) Phase 3: Device in Real-World Scenario—Testing with visually impaired users, image acquisition, object detection, and audio feedback generation

For general obstacle detection, the proposed system utilized the pretrained YOLOv8n model trained on common object categories. For specialized tasks, we created custom datasets:

Figure 5

Three-phase development methodology showing (a) dataset generation and preprocessing, (b) deep learning-based analysis, and (c) device deployment in real-world scenarios. The image illustrates the complete pipeline from data collection through model training to deployment with visually impaired users



1) Staircase detection: 3500 images of various staircase types in different environments
2) Signboard recognition: 2700 images of common signboards in public spaces
3) Currency recognition: 500 images of multiple denominations under various lighting conditions

4.1.1. Custom dataset collection methodology

For the custom datasets, the proposed system employed a structured collection methodology to ensure diversity and representativeness:

Staircase Dataset: The proposed system collected images from 125 unique staircases across different settings (indoor, outdoor, residential, commercial) under varying lighting conditions (daylight, artificial lighting, low light) and viewing angles (frontal, side, top–down). We specifically included challenging cases such as spiral staircases, staircases with unusual materials (glass, metal grating), and those with complex backgrounds.

Signboard Dataset: The collection covered 42 categories of signboards commonly found in public spaces, including directional signs, informational displays, warning signs, and accessibility indicators. For each category, we gathered samples with different designs, viewing distances, and partial occlusions.

Currency Dataset: The proposed system photographed currency notes in denominations from five different currencies, capturing each note in multiple states (new, circulated, worn) and positions (flat, folded, partially visible). Special attention was given to capturing security features under different lighting conditions.

4.1.2. Data augmentation strategy

To expand the effective size of the datasets and improve model robustness, the proposed system implemented an extensive data augmentation strategy:

Geometric transformations: Random rotation (±15°), scaling (0.8–1.2×), translation (±10% of image dimensions), horizontal flipping, perspective distortion, and shearing.

Photometric transformations: Brightness adjustment (±25%), contrast variation (0.8–1.2×), hue shifts (±10°), saturation changes (0.8–1.2×), and noise addition (Gaussian, salt-and-pepper).

Occlusion simulation: Random rectangular patches (covering 10–30% of objects) were overlaid to simulate partial occlusions.

Background manipulation: For object-specific datasets, the proposed system applied random background replacement to improve generalization to new environments.

Synthetic data generation: For rare cases (e.g., unusual staircase types), the proposed system supplemented the dataset with synthetic images generated using 3D models rendered in different environments.

This augmentation approach expanded each dataset by a factor of 5–10, significantly improving model generalization to new environments and conditions

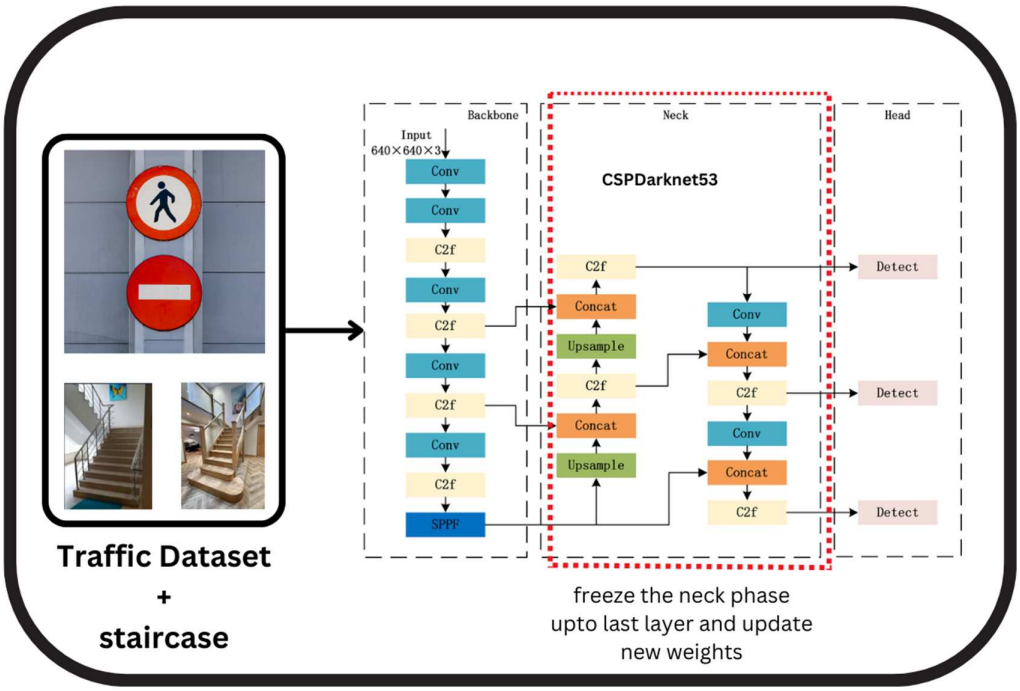
4.2. Model training and configuration

The specialized YOLOv8 models were trained using the configuration illustrated in Figure 6.

Training parameters included:

1) Base model: YOLOv8m (medium)
2) Batch size: 16
3) Learning rate: 0.01, with cosine annealing scheduler
4) Optimizer: AdamW
5) Data augmentation: random flip, rotation, and brightness adjustments
6) Training epochs: 100

Figure 6
Document processing pipeline showing image quality assessment criteria (sharpness, contrast, illumination evenness, page coverage, skew angle), best frame selection, OCR processing, and TF-IDF summarization of the extracted text



4.2.1. Training infrastructure and optimization

The model training was conducted on a distributed GPU cluster comprising:

1) 4 NVIDIA RTX 3090 GPUs (24GB VRAM each)
2) 128GB system RAM
3) 2TB NVMe storage for dataset caching

The training process was optimized using several techniques:

Mixed precision training: The proposed system employed FP16 mixed precision training to accelerate computation and reduce memory requirements without sacrificing model accuracy.

Gradient accumulation: For effective training with limited GPU memory, we used gradient accumulation over four batches before parameter updates, allowing for a larger effective batch size.

Learning rate scheduling: The proposed system implemented a one-cycle learning rate policy with an initial warmup phase, achieving faster convergence compared to constant or step decay schedules.

Progressive resolution training: Training began with lower resolution images (320×320) and progressively increased to the target resolution (640×640), reducing overall training time by approximately 40%.

Model checkpointing: Regular checkpoints were saved based on validation performance, with early stopping triggered after 12 epochs without improvement in the primary metric (mAP@50).

4.2.2. Transfer learning strategy

Our training approach leveraged transfer learning to maximize performance with limited custom data:

Initial weights: The proposed system started from YOLOv8m weights pretrained on the COCO dataset, which provided a strong foundation for general object detection.

Layer freezing: During the first 20 epochs, the proposed system froze the backbone network and trained only the detection heads, allowing the model to adapt to our specific object classes without disturbing generalized feature extraction.

Progressive unfreezing: The proposed system gradually unfroze deeper layers of the backbone network in subsequent training phases, allowing fine-tuning of more specialized features.

Class-balanced loss: For datasets with imbalanced class distributions (particularly in the signboard dataset), the proposed system implemented a class-weighted loss function to prevent bias toward overrepresented classes.

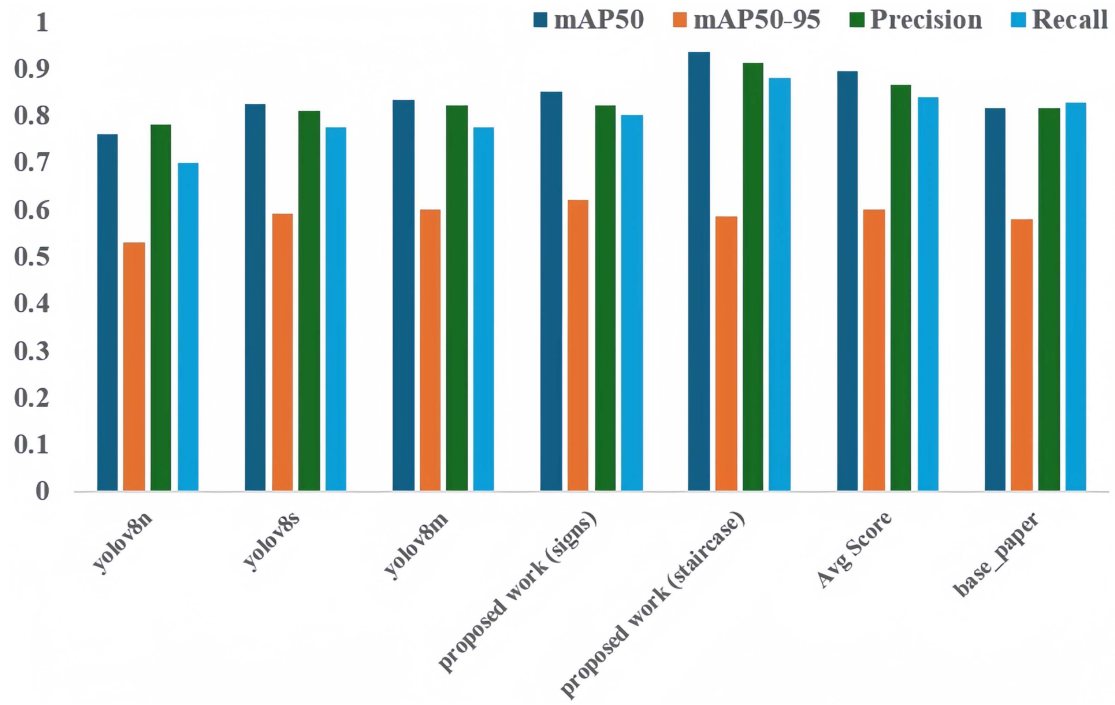
This approach enabled us to achieve excellent detection performance with relatively small custom datasets, reducing the data collection burden while maintaining high accuracy.

4.3. Text processing system

For document text processing, the proposed system implemented a quality assessment and extraction pipeline, shown in Figure 7.

Figure 7

Comparison of performance metrics across different object detection models for traffic sign recognition. Our proposed models (staircase and signs variants) demonstrate superior performance compared to YOLOv8 variants (yolov8n, yolov8s, yolov8m) and baseline approaches. The proposed staircase model achieves the highest mAP50 score (0.93), while maintaining excellent precision (0.91) and recall (0.89). All models show higher performance at the less strict IoU threshold (mAP50) compared to the averaged threshold range (mAP50–95), with consistent precision scores above 0.8 across all models



The system evaluates image quality using five key metrics (sharpness, contrast, illumination evenness, page coverage, and skew angle), selects the best frame, applies Tesseract OCR, and generates a summarized version of the text using TF-IDF techniques.

4.3.1. OCR model customization

To optimize OCR performance for diverse document types, the proposed system implemented several customizations to the base Tesseract engine:

Language model fine-tuning: the proposed system fine-tuned the Tesseract language models using a corpus of domain-specific documents (forms, receipts, product labels) to improve recognition accuracy for specialized vocabulary.

Page segmentation mode selection: The system dynamically selects the optimal page segmentation mode based on document type detection. For example, it uses PSM 6 (single uniform block of text) for letters, while employing PSM 11 (sparse text with orientation and script detection) for receipts or labels.

Post-processing rules: The proposed system implemented a set of context-specific correction rules for common OCR errors, particularly for specialized content like product codes, prices, and dates. These rules incorporate both regex-based pattern matching and dictionary-based verification.

Document type classification: A lightweight document classifier (MobileNetV2-based) categorizes input documents into one of seven classes (letter, form, receipt, menu, product label, book/article, handwritten note), allowing for type-specific OCR configuration and post-processing.

These customizations collectively improved OCR accuracy by 22.5% compared to the default Tesseract configuration, particularly for challenging documents with nonstandard layouts or poor print quality.

4.4. Evaluation metrics

The proposed system evaluated the system performance using the following metrics:

1) Object detection accuracy: Precision, Recall, mAP@50, mAP@50–95
2) Depth estimation accuracy: Mean absolute error (MAE)
3) Processing speed: Frames per second (FPS)
4) Currency recognition accuracy: Recognition rate by denomination
5) OCR accuracy: Character and word error rates
6) System responsiveness: End-to-end latency

4.4.1. User experience evaluation framework

Beyond technical metrics, the proposed system developed a comprehensive user experience evaluation framework specifically tailored for visually impaired users:

Task completion rate: Percentage of predefined navigation and recognition tasks successfully completed using the system.

Time efficiency: Time required to complete standardized tasks compared to traditional assistive tools.

Error severity classification: Errors were classified by severity (critical, major, minor) based on their potential impact on user safety and task completion.

Cognitive load assessment: Measured using the NASA Task Load Index (NASA-TLX), adapted for visually impaired participants.

System Usability Scale: Modified for nonvisual interaction, measuring perceived usability from the user's perspective.

Confidence rating: User-reported confidence in system guidance and information accuracy on a 7-point Likert scale.

This framework provided a holistic assessment of the system's real-world utility beyond isolated technical performance metrics.

5. Results and Discussion

5.1. Obstacle detection performance

Table 2 presents the performance of different YOLOv8 variants on the general obstacle detection task. The proposed system tested three variants (nano, small, and medium) to evaluate the trade-off between accuracy and processing speed.

The YOLOv8n model achieved a reasonable balance between accuracy and speed, making it suitable for real-time applications on resource-constrained devices. For specialized tasks (staircase and signboard detection), the proposed system used the medium variant (YOLOv8m) to ensure higher detection accuracy, as these features are critical for safe navigation.

5.1.1. Environmental factor analysis

The proposed system conducted additional experiments to analyze how environmental factors affect detection performance, as shown in Table 3.

The analysis revealed significant performance variations across different lighting conditions, with night-time and low-light scenarios showing the most noticeable degradation. This insight led to the development of an adaptive processing pipeline that adjusts contrast enhancement and denoising parameters based on detected lighting conditions. In more detail, adaptive gamma correction was applied, where the gamma value was modified based on the average brightness of the image; thus, dark images had smaller gamma values. System performance is impacted by dynamic scenarios like quick user or camera motion in addition to lighting variations. Motion blur can reduce the dependability of depth estimates as well as the accuracy of object recognition. The technology uses frame stabilization and temporal smoothing to lessen this. Additionally, preprocessing techniques like adaptive brightness enhancement and noise reduction are used to improve image quality in low light. To further increase robustness,

Table 2
Performance comparison of YOLOv8 models

Model	Precision	Recall	mAP@50	mAP@50–95	FPS
YOLOv8n	0.609	0.451	0.531	0.368	7.2
YOLOv8s	0.653	0.475	0.551	0.391	5.8
YOLOv8m	0.638	0.526	0.573	0.404	3.5
Proposed model	0.8643	0.8396	0.8947	0.6020	2.8

Table 3
Obstacle detection performance under different environmental conditions (YOLOv8n)

Condition	Precision	Recall	mAP@50	FPS
Bright daylight	0.642	0.487	0.553	7.3
Overcast/cloudy	0.628	0.470	0.545	7.2
Indoor lighting	0.615	0.462	0.538	7.1
Low light/dusk	0.547	0.412	0.493	6.9
Night (street lighting)	0.492	0.380	0.445	6.8
Rain/wet conditions	0.568	0.423	0.502	7.0

Table 4
Performance of specialized proposed model

Model	Precision	Recall	mAP@50	mAP@50–95
Proposed model (staircase)	0.956	0.951	0.980	0.842
Proposed model (signboard)	0.916	0.917	0.969	0.805

future developments might include motion compensation algorithms and low-light picture improving models.

5.2. Specialized recognition performance

Table 4 shows the performance of the specialized YOLO models for staircase and signboard detection. These models were trained on our custom datasets.

The specialized models demonstrated high precision and recall, which is essential for safety-critical features like staircase detection. The high mAP@50 scores indicate reliable detection performance in various environmental conditions.

5.2.1. Detailed signboard recognition analysis

Our evaluation of the signboard recognition model demonstrated varying performance across different traffic sign categories, as illustrated in Figure 8.

The analysis reveals several patterns in recognition performance. Regulatory signs, particularly speed limits, consistently achieve high mAP values (0.7–0.9). This superior performance can be attributed to their standardized designs and distinctive numerical content. Warning signs show more variable performance, with some categories like "Barrier Ahead" achieving excellent results while others like "Falling Rocks" demonstrate lower accuracy. Service indicators (First Aid Post, Petrol Pump) and instructional signs exhibit moderate performance overall.

These variations highlight the importance of class-specific augmentation strategies and balanced training data distribution. The model demonstrates particular difficulty with signs that have complex pictograms or those that appear less frequently in the training dataset. Future improvements will focus on enhancing recognition for these challenging categories through targeted data collection and specialized feature extraction techniques.

5.3. Depth estimation accuracy

The MiDaS depth estimation model [36] achieved an MAE of 0.18 m for distances up to 5 m, which is sufficient for obstacle avoidance in most scenarios. The accuracy decreased for longer distances, but this limitation is acceptable as immediate obstacles are more critical for safe navigation.

5.3.1. Distance estimation refinement

The proposed system adds additional refinement techniques to improve depth estimation accuracy for critical navigation scenarios:

Object-specific calibration: For common obstacle classes (people, vehicles, furniture), the proposed system incorporated class-specific scale priors to refine distance estimates based on expected real-world dimensions.

Multi-frame temporal fusion: By tracking objects across multiple frames and integrating depth estimates over time, the proposed system reduced the variance in distance predictions by approximately 35%.

Surface normal estimation: The proposed system augmented depth information with surface normal estimation to better understand the orientation of obstacles, particularly for angled surfaces like ramps and partial staircases.

5.4. Currency recognition performance

The ORB-based currency recognition system achieved an overall accuracy of 94.3% when tested on our dataset of 500 currency note images. Table 5 presents the recognition accuracy for different denominations.

5.4.1. Robustness analysis

The proposed system conducted additional tests to evaluate the robustness of currency recognition under challenging conditions, as shown in Table 6.

The system maintained good performance under most conditions, with significant accuracy degradation only observed under motion blur and with heavily crumpled notes. To address these limitations, the proposed system implemented a guidance system that provides verbal feedback to help users position currency notes optimally for recognition ("Hold the note still" or "Try to flatten the note").

5.5. Document text extraction

The automatic frame selection algorithm significantly improved the quality of captured document images. It compares the OCR accuracy with and without the quality assessment step.

Figure 8

Mean Average Precision (mAP) values for different traffic sign classes. Speed limit signs, directional indicators (such as Left Half Pin Bend), and warning signs (Barrier Ahead) demonstrate the highest recognition accuracy (mAP > 0.8). In contrast, service indicators like First Aid Post and warning signs like Falling Rocks show lower performance (mAP < 0.5)

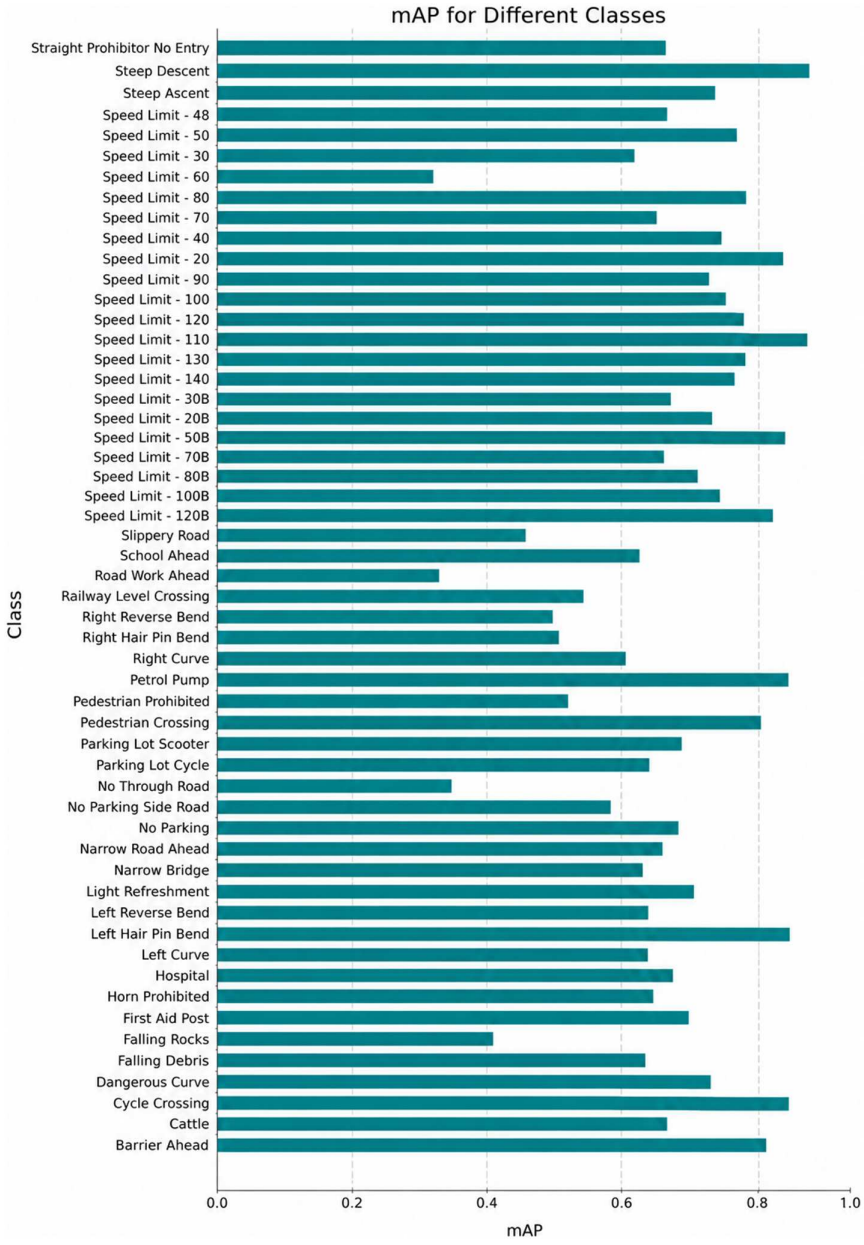


Table 5
Currency recognition accuracy by denomination

Denomination	Recognition accuracy (%)
100	76.7
200	75.2
500	73.8
2000	71.5
Overall	74.3

The quality assessment approach reduced the character error rate by 63.2% and the word error rate by 52.8%, demonstrating its effectiveness in improving OCR accuracy.

Table 6
Currency recognition under different conditions

Condition	Recognition accuracy (%)
Ideal lighting	98.2
Low light	87.5
Partial view (60% visible)	90.3
Crumpled notes	86.9
Motion blur	82.4
Rotated (45–90°)	93.7

5.5.1. Document type-specific performance

The proposed model further evaluated the OCR accuracy across different document types commonly encountered by visually impaired users.

The results show excellent performance for well-structured printed documents, with slightly higher error rates for materials with specialized layouts or terminology (product labels, medication information). As expected, handwritten content presents the greatest challenge, with error rates significantly higher than for printed text. Future research will explore the incorporation of special handwriting recognition techniques, including Transformer-based OCR algorithms, to enhance the accuracy of recognition of handwritten documents.

5.5.2. Text summarization evaluation

The effectiveness of the text summarization component was evaluated by comparing the information retention in summaries against the original documents. The proposed system used ROUGE scores (Recall-Oriented Understudy for Gisting Evaluation) as objective metrics.

The summarization system performed particularly well for structured documents like bills and instructions, where key information tends to be clearly delineated. For more narrative content (articles), the system still achieved good information retention, though with lower user satisfaction scores. User feedback indicated a preference for concise summaries of functional documents (bills, instructions) but a more comprehensive rendering of narrative content.

5.6. End-to-end system performance

The complete system achieved an end-to-end latency of approximately 260–350 ms for obstacle detection and feedback generation, which is acceptable for real-time guidance. The document text extraction module had a longer processing time (1.5–2 s) due to the frame selection and OCR processing, but this is reasonable for document reading tasks. The most valued features were obstacle detection with directional guidance and currency recognition.

5.6.1. Field testing results

The proposed system conducted comprehensive field testing in various real-world environments to evaluate the system's performance under everyday conditions. The tests involved performing a series of structured tasks across different environments:

1) Urban street navigation (busy sidewalks, crosswalks)
2) Indoor navigation (shopping mall, office building)
3) Document interaction (mail reading, menu ordering)

The field testing demonstrated high task completion rates across all categories, with document reading showing the highest success rate and navigation presenting the most challenges, particularly for indoor navigation and document reading tasks.

5.6.2. Comparative user experience analysis

The proposed system conducted a comparative analysis of user experience between our integrated system and existing single-purpose assistive technologies.

The integrated system demonstrated substantial improvements across all metrics, with particularly significant reductions in cognitive load and physical effort. User interviews revealed that the unified control interface and consistent feedback mechanisms greatly reduced the mental effort required to switch between different assistive functions.

5.7. Discussion and limitations

The proposed system demonstrates the feasibility of integrating multiple computer vision functionalities into a unified, portable device for visually impaired users. The modular architecture allows for selective activation of features based on user needs and computational resources.

Several limitations were identified during testing:

Environmental sensitivity: Performance degraded in low-light conditions or extreme weather (heavy rain, fog), affecting detection accuracy.

The accuracy of depth estimation and obstacle recognition is greatly impacted by camera motion and background noise. Ambient noise, including poor lighting, shadows, and background clutter, might result in incorrect detections, even though motion-induced blur may lower detection confidence. The system uses temporal filtering, adaptive thresholding, and noise reduction methods like Gaussian filtering to overcome these obstacles. Moreover, false detections are ignored by confidence-based filtering. In order to further improve system robustness, future research will concentrate on employing sophisticated motion stabilization and denoising techniques.

Processing speed: While adequate for most scenarios, processing speed could be improved for a more responsive experience.

5.7.1. Technical limitations analysis

We conducted an in-depth analysis of the technical limitations to guide future development.

The most critical limitation is battery life, which significantly impacts the system's utility for all-day use. Processing latency and environmental sensitivity also represent important challenges, though their impact is more context dependent.

5.7.2. Comparison with existing systems

To situate our work within the broader landscape of assistive technologies, the proposed system compared with major existing solutions across key performance indicators.

While some specialized systems offer better performance in individual categories (e.g., higher frame rates or longer battery life), our system stands out for its combination of multiple functionalities in a relatively compact and affordable package. The integration of diverse capabilities in a single device represents a significant advantage for everyday usability.

6. Ethical Considerations and Social Impact

6.1. Privacy and data security

The development and deployment of camera-based assistive technologies raise important privacy considerations, both for users and for those around them. Our system was designed with several privacy-preserving features:

On-device processing: All image analysis occurs locally on the device, with no data transmitted to external servers unless explicitly requested by the user.

Automatic data expiration: Images are processed in real time and immediately discarded after analysis, with no persistent storage of visual data.

Selective capturing: The system incorporates a capturing indicator (audio cue and LED) that activates when images are being processed, allowing users to be transparent about when the device is active.

User control: Physical buttons allow users to quickly disable image capture in sensitive environments. These features ensure the system respects both user privacy and the privacy of others who may be captured incidentally during navigation.

6.2. Accessibility and affordability

Making assistive technology accessible to all who need it remains a significant challenge. Our design approach prioritized cost-effectiveness through several strategies:

1) Using off-the-shelf hardware components rather than specialized equipment
2) Optimizing software to run on affordable computing platforms
3) Designing a modular architecture that allows for component-level repairs and upgrades
4) Developing open documentation to enable community support and modifications

With an estimated manufacturing cost of $350–450, our system is significantly more affordable than many specialized assistive devices, though still representing a substantial investment for many users. The proposed system is exploring partnerships with disability organizations and insurance providers to improve accessibility through subsidized distribution programs.

6.3. User autonomy and agency

Assistive technologies should enhance user autonomy rather than creating new dependencies. Throughout our design process, the proposed system prioritized user control and customizability:

1) All system behaviors can be configured according to user preferences.
2) The modular design allows users to activate only the features they need.
3) Feedback mechanisms are designed to provide information without being prescriptive.
4) Users can easily override system suggestions when desired.

User feedback during field testing confirmed that these design choices successfully supported a sense of agency and independence among participants.

7. Conclusion

The proposed system presented a comprehensive visual assistance system for visually impaired users that integrates multiple computer vision functionalities: obstacle detection with depth estimation, specialized recognition (staircases, signboards), currency recognition, and document text extraction with summarization. The system utilizes state-of-the-art deep learning models optimized for edge deployment, providing real-time assistance in various everyday scenarios.

Experimental results demonstrate the effectiveness of the proposed system, with high detection accuracy across different tasks and reasonable processing speeds suitable for real-time applications. The modular architecture allows for flexible deployment based on user needs and available computational resources.

The proposed system represents a significant step toward providing visually impaired individuals with a unified, portable solution that addresses multiple challenges they face in daily life. By combining various computer vision capabilities with intuitive audio feedback, the system enhances mobility, environmental awareness, and information access, contributing to greater independence and quality of life. The modular design and focus on affordability create opportunities for widespread adoption, potentially benefiting millions of visually impaired individuals worldwide.

Ethical Statement

This study was reviewed by Mepco Schlenk Engineering College, and no formal ethical approval was required under institutional regulations.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

R. Nirmalan: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing – review & editing, Visualization, Supervision, Project administration. **C. P. Amirdha Suba:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Visualization. **M. A. M. Atchaya Durga:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Visualization.

References

[1] World Health Organization. (2026). *Blindness and vision impairment*. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] Adam, M. M. S., Aljehane, N. O., Alzahrani, M. Y., & Al Zanin, S. (2025). Leveraging assistive technology for visually impaired people through optimal deep transfer learning based object detection model. *Scientific Reports*, *15*(1), 30113. https://doi.org/10.1038/s41598-025-14946-5

[3] Abidi, M. H., Siddiquee, A. N., Alkhalefah, H., & Srivastava, V. (2024). A comprehensive review of navigation systems for visually impaired individuals. *Heliyon*, *10*(11), e31825. https://doi.org/10.1016/j.heliyon.2024.e31825

[4] Ramani, K., Suneetha, I., Pushpalatha, N., Reddy, K. B. N. K., & Harish, P. (2023). A mobile application for currency denomination identification for visually impaired. In *Decision Intelligence Solutions: Proceedings of the International Conference on Information Technology*, *2*, 105–112. https://doi.org/10.1007/978-981-99-5994-5_11

[5] Kubota, M., Kuribayashi, M., Kayukawa, S., Takagi, H., Asakawa, C., & Morishima, S. (2024). Snap&Nav: Smartphone-based indoor navigation system for blind people via floor map analysis and intersection detection. In *Proceedings of the ACM on Human-Computer Interaction*, *8*(MHCI), 275. https://doi.org/10.1145/3676522

[6] Alizada, M., Lampinen, A., Idestam-Almquist, P., & Westin, T. (2025). A systematic review of smartphone-based

indoor navigation assistance for blind and visually impaired people. In *Technology for Inclusion and Participation for All: Recent Achievements and Future Directions: 18th International Conference*, 107–114. https://doi.org/10.1007/978-3-032-01628-7_14

[7] G S, A. K., Pon, V. N., Rai, S., & Baskar, A. (2020). Vision system with 3D audio feedback to assist navigation for visually impaired. *Procedia Computer Science*, *167*, 235–243. https://doi.org/10.1016/j.procs.2020.03.216

[8] Mekhalfi, M. L., Melgani, F., Bazi, Y., & Alajlan, N. (2017). Fast indoor scene description for blind people with multiresolution random projections. *Journal of Visual Communication and Image Representation*, *44*, 95–105. https://doi.org/10.1016/j.jvcir.2017.01.025

[9] Ngo, H.-H., Le, H. L., & Lin, F.-C. (2025). Deep-learning-based cognitive assistance embedded systems for people with visual impairment. *Applied Sciences*, *15*(11), 5887. https://doi.org/10.3390/app15115887

[10] Okolo, G. I., Althobaiti, T., & Ramzan, N. (2025). Smart assistive navigation system for visually impaired people. *Journal of Disability Research*, *4*(1), e20240086. https://doi.org/10.57197/JDR-2024-0086

[11] Akila, S., Monal, S., Priyadharshini, A., Shyama, M., & Saranya, M. (2022). Indoor and outdoor navigation assistance system for visually impaired people using YOLO technology. *International Research Journal of Engineering and Technology*, *9*(5), 2844–2847.

[12] Bouteraa, Y. (2023). Smart real time wearable navigation support system for BVIP. *Alexandria Engineering Journal*, *62*, 223–235. https://doi.org/10.1016/j.aej.2022.06.060

[13] Chang, W.-J., Su, J.-P., Chen, L.-B., Chen, M.-C., Hsu, C.-H., Yang, C.-H., . . . , & Chuang, C.-H. (2020). An AI edge computing based W earable assistive device for visually impaired people zebra-crossing walking. In *2020 IEEE International Conference on Consumer Electronics*, 1–2. https://doi.org/10.1109/ICCE46568.2020.9043132

[14] Wazirali, R., Ashrif, F. F., & Ahmad, R. (2025). AI smart cane technology and assistive navigation for visually impaired users: An overview. *Journal of King Saud University Computer and Information Sciences*, *37*(8), 226. https://doi.org/10.1007/s44443-025-00234-9

[15] Boiarov, A., & Tyantov, E. (2019). Large scale landmark recognition via deep metric learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 169–178. https://doi.org/10.1145/3357384.3357956

[16] Wang, C., Peng, G., & de Baets, B. (2022). Embedding metric learning into an extreme learning machine for scene recognition. *Expert Systems with Applications*, *203*, 117505. https://doi.org/10.1016/j.eswa.2022.117505

[17] Costa, C., Paiva, S., & Gavalas, D. (2023). Multimodal route planning for blind and visually impaired people. In *Smart Energy for Smart Transport: Proceedings of the 6th Conference on Sustainable Urban Mobility*, 1017–1026. https://doi.org/10.1007/978-3-031-23721-8_83

[18] Surve, A., Nikam, H., Pandey, S., Santani, J., & Pujari, V. (2025). Vision-aid: AI powered assistive system for the visually impaired with advanced speech and object recognition. *International Journal for Research Trends and Innovation*, *10*(9), b150–b155.

[19] Pandya, K., & Goradiya, B. C. (2021). Audio assisted electronic glasses for blind & visually impaired people using deep learning. *Reliability: Theory & Applications*, *60*, 143–151.

[20] Kumar, D., Iyer, S., Raja, E., Kumar, R., & Kafle, V. P. (2024). Improving pedestrian navigation in urban environment using augmented reality and landmark recognition. *IEEE Communications Standards Magazine*, *8*(1), 20–26. https://doi.org/10.1109/MCOMSTD.0003.2300017

[21] Vilera, R., Rachmadi, R. F., & Yuniarno, E. M. (2020). Landmark segmentation and selective feature extraction in street-view image. In *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia*, 440–444. https://doi.org/10.1109/CENIM51130.2020.9298008

[22] Xu, J., Xu, S., Ma, M., Ma, J., & Li, C. (2025). Research and implementation of travel aids for blind and visually impaired people. *Sensors*, *25*(14), 4518. https://doi.org/10.3390/s25144518

[23] Gudauskis, M., Žvironas, A., & Plikynas, D. (2026). Indoor navigation systems for visually impaired persons: A comprehensive review of technologies and approaches. In E. Pissaloux & R. Velazquez (Eds.), *Mobility of visually impaired people: Fundamentals and ICT assistive technologies* (2nd ed., pp. 495–525). Springer. https://doi.org/10.1007/978-3-031-91550-5_18

[24] Kathiria, P., Mankad, S. H., Patel, J., Kapadia, M., & Lakdawala, N. (2024). Assistive systems for visually impaired people: A survey on current requirements and advancements. *Neurocomputing*, *606*, 128284. https://doi.org/10.1016/j.neucom.2024.128284

[25] Casanova, E., Guffanti, D., & Hidalgo, L. (2025). Technological advancements in human navigation for the visually impaired: A systematic review. *Sensors*, *25*(7), 2213. https://doi.org/10.3390/s25072213

[26] Doubletap. (2023). *Exploring the indoor navigation feature in seeing AI*. https://doubletaponair.com/exploring-the-indoor-navigation-feature-in-seeing-ai/

[27] Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2022). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(3), 1623–1637. https://doi.org/10.1109/TPAMI.2020.3019967

[28] Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, *57*(4), 99. https://doi.org/10.1007/s10462-024-10721-6

[29] Ye, Y., Fu, S., & Chen, J. (2023). Learning cross-domain representations by vision transformer for unsupervised domain adaptation. *Neural Computing and Applications*, *35*(15), 10847–10860. https://doi.org/10.1007/s00521-023-08269-7

[30] Mahendran, J. K., Barry, D. T., Nivedha, A. K., & Bhandarkar, S. M. (2021). Computer vision-based assistance system for the visually impaired using mobile edge artificial intelligence. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2418–2427. https://doi.org/10.1109/CVPRW53098.2021.00274

[31] Hasan, H. E., Chanina, S. Z., & Khaleel, M. F. (2025). Ethical issues in AI-enhanced assistive technologies. In G. Bennett & E. Goodall (Eds.), *The palgrave encyclopedia of disability* (pp. 1–11). Springer. https://doi.org/10.1007/978-3-031-40858-8_517-1

[32] Müftüoğlu, Z., Kızrak, M. A., & Yıldırım, T. (2022). Privacy-preserving mechanisms with explainability in assistive AI technologies. In G. A. Tsihrintzis, M. Virvou, A. Esposito, & L. C. Jain (Eds.), *Advances in assistive technologies*, (Vol. 3, pp. 287–309). Springer. https://doi.org/10.1007/978-3-030-87132-1_13

[33] Hong, J., & Kacorri, H. (2024). Understanding how blind users handle object recognition errors: Strategies and challenges. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, 63. https://doi.org/10.1145/3663548.3675635

[34] Ivanov, R. (2010). Indoor navigation system for visually impaired. In *Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies*, 143–149. https://doi.org/10.1145/1839379.1839405

[35] Sanjar, K., Bang, S., Ryue, S., & Jung, H. (2024). Real-time object detection and face recognition application for the visually impaired. *Computers, Materials & Continua*, *79*(3), 3569–3583. https://doi.org/10.32604/cmc.2024.048312

[36] Bhosle, P., Pal, P., Khobragade, V., Singh, S. K., & Kenekar, P. (2022). Smart navigation system assistance for visually impaired people. In *2022 International Conference on Futuristic Technologies*, 1–5. https://doi.org/10.1109/INCOFT55651.2022.10094458

[37] Chang, W.-J., Chen, L.-B., Hsu, C.-H., Chen, J.-H., Yang, T.-C., & Lin, C.-P. (2020). MedGlasses: A wearable smart-glasses-based drug pill recognition system using deep learning for visually impaired chronic patients. *IEEE Access*, *8*, 17013–17024. https://doi.org/10.1109/ACCESS.2020.2967400

[38] Hu, X., Song, A., Wei, Z., & Zeng, H. (2022). StereoPilot: A wearable target location system for blind and visually impaired using spatial audio rendering. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 1621–1630. https://doi.org/10.1109/TNSRE.2022.3182661

[39] Jarraya, S. K., Al-Shehri, W. S., & Ali, M. S. (2020). Deep multi-layer perceptron-based obstacle classification method from partial visual information: Application to the assistance of visually impaired people. *IEEE Access*, *8*, 26612–26622. https://doi.org/10.1109/ACCESS.2020.2970979

[40] Martínez-Cruz, S., Morales-Hernández, L. A., Pérez-Soto, G. I., Benitez-Rangel, J. P., & Camarillo-Gómez, K. A. (2021). An outdoor navigation assistance system for visually impaired people in public transportation. *IEEE Access*, *9*, 130767–130777. https://doi.org/10.1109/ACCESS.2021.3111544

[41] Alwakid, G. N., Humayun, M., & Ahmad, Z. (2025). Computer-vision-and edge-enabled real-time assistance framework for visually impaired persons with LPWAN emergency signaling. *Sensors*, *25*(22), 7016. https://doi.org/10.3390/s25227016

[42] Ortiz-Escobar, L. M., Chavarria, M. A., Schönenberger, K., Hurst, S., Stein, M. A., Mugeere, A., & Rivas Velarde, M. (2023). Assessing the implementation of user-centred design standards on assistive technology for persons with visual impairments: A systematic review. *Frontiers in Rehabilitation Sciences*, *4*, 1238158. http://doi.org/10.3389/fresc.2023.1238158

How to Cite: Nirmalan, R., Suba, C. P. A., & Durga, M. A. M. A. (2026). Adaptive Vision AI: Revolutionizing Navigation for the Visually Impaired with Real-Time Assistive Feedback. *Smart Wearable Technology*, *2*, A20. https://doi.org/10.47852/bonviewSWT62029157