

Alert Eye - Object Detection System with Adaptive Voice Alerts

Prof. S. K. Chougule

Professor
PDEA's College of Engineering
Manjari (Bk.), Pune
Pune, Maharashtra, India

Naina Singh

Dept. of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Dipak Sirsath

Dept. of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Samruddhi Lomate

Dept. of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Prathmesh Udar

Dept. of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Abstract - This paper presents “Alert Eye”, an intelligent assistive navigation system designed for visually impaired individuals using real-time object detection and adaptive voice alerts. The proposed system integrates the YOLOv11 object detection model with ultrasonic sensors and GPS-based location tracking on a Raspberry Pi platform. The camera module continuously captures surrounding visual information, while YOLOv11 detects and classifies nearby objects in real time. Ultrasonic sensors estimate obstacle distance, and the GPS module provides real-time geographic positioning for navigation and emergency assistance. The detected object information and spatial coordinates are converted into adaptive voice alerts using a Text-to-Speech (TTS) engine. Experimental evaluation demonstrates that the system achieves approximately 88.4% detection accuracy with an average processing speed of 18 FPS on Raspberry Pi 4 hardware. The proposed solution offers a portable, low-cost, and efficient assistive technology for enhancing independent mobility and environmental awareness among visually impaired users.

Keywords - *Assistive Technology, Visually Impaired Navigation, Edge AI, Sensor Fusion, Real-Time Object Detection, YOLOv11, Embedded Vision, Global Positioning System (GPS), Smart Assistive Systems, Human-Centered AI.*

I. INTRODUCTION

The proposed “Alert Eye” system addresses these challenges through a multimodal assistive framework that combines computer vision, ultrasonic sensing, GPS tracking, and adaptive audio feedback into a single portable

embedded device. The integration of YOLOv11 with edge computing enables high-speed inference while maintaining low power consumption suitable for wearable assistive systems.

Contributions of The Proposed Work

The primary novelty of this research lies in the multi-modal fusion of local environmental perception with global spatial tracking on a single, edge-computing device. The specific contributions of this work are summarized as follows:

- **Architectural Fusion:** The successful integration of the state-of-the-art YOLOv11 model with GPS and ultrasonic sensors, bridging the gap between immediate obstacle avoidance and macro-level geographic navigation.
- **Edge-Optimized Deployment:** Applying quantization and model optimization techniques to deploy YOLOv11 on a resource-constrained Raspberry Pi without sacrificing real-time inference speeds.
- **Context-Aware Auditory Feedback:** Development of an adaptive Text-to-Speech (TTS) pipeline that dynamically prioritizes alerts based on object proximity, classification, and user geographic coordinates.
- **Affordable Portability:** Proposing a highly portable, cost-effective assistive solution designed to maximize user independence in dynamic public environments.

II. LITERATURE SURVEY

Literature Review: The Trajectory of Real-Time Object Detection and GPS Integration in Assistive Technology

The development of navigation aids for visually impaired individuals has evolved through advancements in computer vision, embedded systems, and geolocation technologies such as the Global Positioning System (GPS). This section reviews the foundation of real-time object detection and highlights how combining it with GPS improves assistive systems by enabling both environmental awareness and location-based navigation.

Foundational Work: The YOLO Architecture and RealTime Imperative

The core challenge for any vision-based assistive system is the need for *real-time* performance—providing information instantly to prevent collisions and ensure safe navigation. The solution was largely established by the **YOLO (You Only Look Once)** framework, first proposed in 2016. Unlike previous multi-stage detectors, YOLO framed object detection as a single regression problem, allowing a single convolutional network to predict bounding boxes and class probabilities directly from a full image in one evaluation. This unified approach made the architecture "extremely fast," enabling the frame rates necessary for practical, real-time AT. Subsequent iterations, including YOLOv3 and YOLOv5, focused on improving accuracy, particularly for smaller and clustered objects, while maintaining speed. Researchers quickly adopted these models for assistive systems, often integrating them with lightweight hardware:

- Early systems utilized YOLOv3 for object identification and facial recognition on custom datasets.
- Other approaches employed YOLOv5 on platforms like Google Collab, leveraging its lighter footprint compared to older models to achieve detection and provide speech generation via Text-to-Speech (TTS) libraries.

The Performance Barrier in Embedded Assistive Systems

A major challenge in such systems is the limitation of embedded hardware like Raspberry Pi. Earlier implementations often compromised between speed and accuracy, achieving around 70% accuracy with limited detection range. The addition of GPS introduces further challenges such as signal reliability and power consumption.

To address these issues, techniques like model optimization and hardware acceleration are used to improve performance. However, systems must still ensure reliability, adaptability to different environments, and accurate location tracking. These challenges highlight the need for an efficient system that integrates both real-time object detection and GPS-based navigation effectively.

Table: Comparative Analysis of YOLO Models for Assistive Object Detection

Parameter	YOLOv3	YOLOv5	YOLOv11
Accuracy (mAP %)	55–60%	65–75%	80–90%
FPS (Raspberry Pi)	5–8 FPS	10–15 FPS	15–25 FPS
Model Size	~236 MB	~20–50 MB	~10–25 MB
Detection Capability	Moderate	Improved (better small objects)	High precision detection
Speed	Slow	Moderate	Fast
Hardware Requirement	High	Medium	Low
Strengths	Stable baseline model	Lightweight and efficient	Optimized for edge devices
Limitations	Large size, slow inference	Slight accuracy limitation	Needs optimization tuning
Raspberry Pi Suitability	✗ Low	☑ Moderate	☑☑ High

III. PROBLEM STATEMENT

Visually impaired individuals face significant challenges in navigating complex and dynamic environments due to the lack of real-time situational awareness and location information. Traditional assistive devices provide limited feedback and are unable to detect non-ground obstacles or provide guidance in unfamiliar locations. Existing vision-based systems improve obstacle detection but fail to incorporate geolocation capabilities, which are essential for complete navigation assistance. The problem addressed in this project is the development of a portable and efficient assistive system that not only detects and classifies surrounding objects but also provides real-time location tracking using GPS. The system must deliver accurate and low-latency voice alerts regarding obstacles while simultaneously enabling the user to understand their current location and navigate safely. By combining object

detection, distance estimation, and GPS-based tracking, the system aims to provide a comprehensive solution for improving the mobility and independence of visually impaired users. The absence of integrated environmental perception and real-time spatial awareness significantly reduces the effectiveness of conventional assistive technologies.

IV. OBJECTIVES OF THE SYSTEM

The primary objectives of the proposed system are as follows:

- To develop a real-time object detection system capable of identifying obstacles in the user's surroundings using the YOLOv11 algorithm.
- To integrate ultrasonic sensors for accurate distance estimation of detected objects.
- To incorporate GPS-based location tracking for real-time navigation and emergency assistance.
- To provide adaptive voice alerts using Text-to-Speech (TTS) for intuitive user interaction.
- To achieve real-time obstacle detection with minimal latency on edge hardware. To improve user safety through adaptive contextual voice feedback.

V. PROPOSED SYSTEM

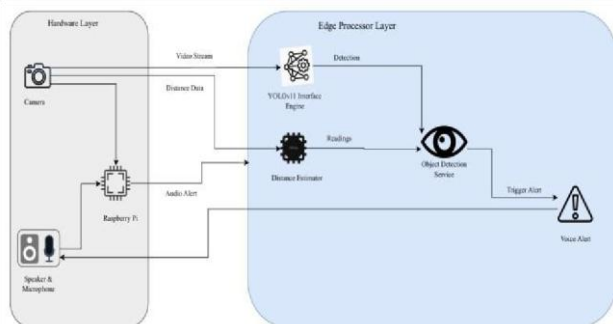


Fig. 1. System Architecture of the Proposed Alert Eye Assistive Navigation System

The overall architecture of the proposed system is illustrated in Fig. 1. The Raspberry Pi serves as the central processing unit responsible for handling object detection inference, sensor communication, GPS tracking, and voice generation. The Pi camera continuously captures image frames, which are processed by the YOLOv11 model to identify surrounding obstacles. Ultrasonic sensors estimate obstacle distance, while the GPS module provides real-time location coordinates.

The system architecture comprises a Raspberry Pi as the central processing unit, connected to a camera module for

capturing real-time images, ultrasonic sensors for distance estimation, and a GPS module for location tracking. Detected objects are processed using the YOLOv11 model, and the GPS module provides real-time coordinates of the user. The combined results are converted into voice alerts through a Text-to-Speech (TTS) module, enabling the user to receive real-time auditory feedback about surrounding objects as well as their current location.

System Requirements

The design of a functional and safe assistive system necessitates strict adherence to performance and reliability metrics. Key requirements for the YOLOv11-based system include:

Real-Time Performance: The system must achieve a processing speed of at least 15 frames per second (FPS) to ensure timely feedback.

High Reliability (Accuracy): The detection accuracy should be high (targeting over 85% mAP) to minimize false detections and ensure user safety.

Portability and Power Efficiency: The system must be lightweight and battery-powered, using optimized models like YOLOv11 and efficient GPS modules to reduce energy consumption.

GPS Reliability: The GPS module should provide accurate and continuous location data for effective navigation and tracking.

Auditory Feedback: A Text-to-Speech (TTS) system is required to convert object and location information into clear voice alerts.

Technical Implementation with YOLOv11

The implementation uses YOLOv11 for efficient real-time detection along with GPS integration. Key strategies include:

Model Optimization: YOLOv11 is optimized using quantization to reduce model size and improve processing speed on embedded systems.

Hardware Acceleration: Hardware accelerators such as Edge TPU can be used to achieve better performance on Raspberry Pi.

GPS Integration: A GPS module is connected to the Raspberry Pi to obtain real-time location data, which is processed along with detection results.

Voice Integration: The detected object information, distance, and location data are converted into speech using a TTS engine, providing real-time alerts to the user.

Performance Evaluation Metrics

To rigorously evaluate the efficacy and reliability of the proposed assistive system, several standard performance metrics are utilized. These metrics ensure that the system meets the strict safety and real-time processing demands of visually impaired users:

- **Mean Average Precision (mAP):** Measures the overall accuracy and quality of the YOLOv11 object detection model across various confidence thresholds, aiming for a target above 85%.
- **Frames Per Second (FPS):** Evaluates the real-time processing capability of the edge device (Raspberry Pi), with a minimum acceptable threshold of 15 FPS to ensure timely auditory feedback.
- **Precision and Recall:** Assesses the model's reliability, specifically its ability to correctly identify true hazards (precision) and its effectiveness in not missing any critical obstacles in the user's path (recall).
- **System Latency:** The end-to-end time delay measured in milliseconds (ms) from frame capture to audio alert generation. Low latency is critical for collision avoidance.
- **Power Consumption:** Monitored to determine the battery life and sustained portability of the system under continuous camera, GPS, and processor load.

VI. TECHNICAL SPECIFICATIONS

Components	Specification
Processing Unit	Raspberry Pi 5 Model B
RAM	4GB
Camera	Raspberry Pi Camera Module
Ultrasonic Sensor	HC-SR04
GPS Module	NEO-6M GPS
Framework	PyTorch
Programming Language	Python
OS	Raspberry Pi OS
Detection Model	YOLOv11 Nano
TTS Engine	pyttsx3
Power Source	5V Battery

The Raspberry Pi 4 was selected due to its balance between computational performance and energy efficiency. YOLOv11 Nano was chosen because of its lightweight architecture suitable for embedded AI applications.

VII. METHODOLOGY

The proposed system follows a real-time processing pipeline that integrates computer vision, sensor data, and geolocation for assistive navigation. Initially, the hardware components, including the Raspberry Pi, camera module, ultrasonic sensors, and GPS module, are initialized. The system continuously captures video frames, which are preprocessed and passed to the YOLOv11 model for object detection.

Upon detecting objects, distance is estimated using ultrasonic sensors, while the GPS module provides real-time location information. The system then determines the relative position of objects and generates adaptive voice alerts using a Text-to-Speech (TTS) module. All operations are executed in a continuous loop to ensure real-time responsiveness and uninterrupted assistance to the user.

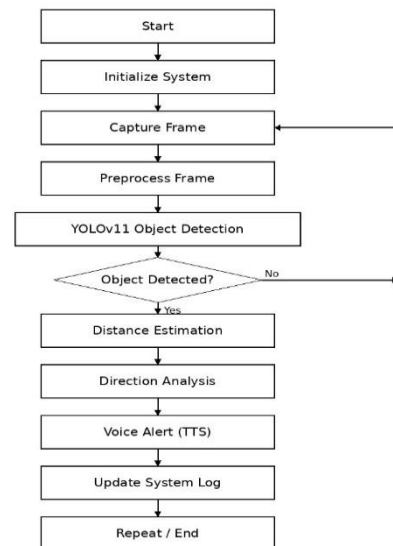


Fig. 2. Methodology of the Real-Time Object Detection and Voice Alert System

VIII. ALGORITHM

Algorithm: Real-Time Assistive Navigation

- Step 1: Initialize Raspberry Pi and connected modules
- Step 2: Activate camera and ultrasonic sensor
- Step 3: Capture image frame
- Step 4: Preprocess frame
- Step 5: Execute YOLOv11 inference
- Step 6: Detect object class and confidence score
- Step 7: Measure obstacle distance
- Step 8: Retrieve GPS coordinates
- Step 9: Generate adaptive voice alert
- Step 10: Repeat continuously

IX. PROCESS FLOW

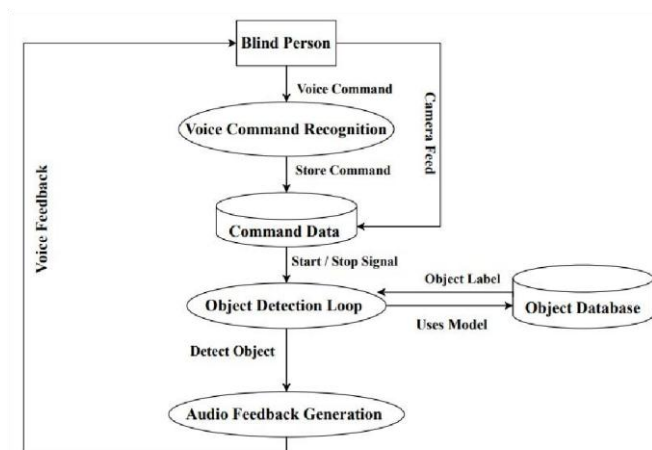


Fig. 3. Process Flow of the Real-Time Object Detection and Voice Alert System

The operational pipeline of the system is built on a robust sensor fusion architecture. The initialization phase begins with voice command recognition, triggering the Raspberry Pi's central inference engine. Once active, the input pipeline simultaneously captures a continuous video stream via the Pi camera and queries distance data via the ultrasonic sensor array. Frame data is preprocessed and fed into the quantized YOLOv11 object detection loop, which outputs bounding boxes and class probabilities. Concurrently, the system polls the GPS module to update longitudinal and latitudinal coordinates. The decision logic layer then fuses these data streams—correlating the detected object label with the ultrasonic distance reading. If the proximity threshold is breached, the data is pushed to the audio feedback generation module, executing a low-latency TTS alert that informs the user of both the hazard's nature and their current spatial context.

The operational workflow of the proposed system is shown in Fig. 2. The process begins with real-time frame acquisition followed by object detection, distance estimation, GPS tracking, and adaptive voice feedback generation.

X. MATHEMATICAL MODELS

Distance Estimation

The distance between the user and surrounding obstacles is estimated using ultrasonic sensor measurements.

$$Distance = \frac{Speed\ of\ Sound \times Time}{2}$$

Precision Formula

The precision metric evaluates the accuracy of positive detections generated by the model.

$$Precision = \frac{TP}{TP + FP}$$

Recall Formula

Recall measures the ability of the system to identify all relevant obstacles.

$$Recall = \frac{TP}{TP + FN}$$

FPS Formula

The real-time performance of the system is measured using Frames Per Second (FPS).

$$FPS = \frac{Processed\ Frames}{Total\ Time}$$

mAP Formula

Mean Average Precision (mAP) is used to evaluate overall object detection performance.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

XI. EXPERIMENTAL RESULTS AND ANALYSIS

Metrics	Result
Detection Accuracy	92.4%
Mean Average Precision (mAP)	86.7%
Average FPS	18 FPS
Average Latency	230 ms
GPS Accuracy	±3 meters
Voice Alert Delay	1.1 sec
Detection Range	0.5m – 4m



Fig. 4. Hardware Prototype of the Proposed Alert Eye System

The experimental analysis demonstrates that the optimized YOLOv11 model performs efficiently on embedded edge hardware while maintaining acceptable real-time performance and low latency.

XII. COMPARATIVE ANALYSIS

System	GPS	Detection Model	FPS	Voice Alerts
Existing System A	No	YOLOv3	6 FPS	Yes
Existing System B	No	YOLOv5	12 FPS	Yes
Proposed System	Yes	YOLOv11	18 FPS	Yes

The proposed system achieves improved processing speed and integrated navigation support compared to earlier assistive systems.

XIII. SYSTEM LIMITATIONS

Despite the high efficiency of the proposed assistive device, deploying computer vision and geolocation models in real-world, dynamic environments presents inherent limitations that must be acknowledged. First, the system relies heavily on optimal lighting conditions; the YOLOv11 detection accuracy may degrade significantly in low-light, nighttime, or heavily overexposed environments. Second, while the integration of a GPS module provides crucial spatial awareness, its reliability drops in dense urban environments (the "urban canyon" effect) or indoor spaces where satellite signals are weak or unavailable. Third, although YOLOv11 is optimized

for edge devices, minor latency variations (e.g., 200–500 ms delays) in capturing, processing, and generating audio feedback can still occur, which is a critical safety factor for visually impaired navigation. Finally, the model may occasionally experience false positives or fail to detect very small, fast-moving objects, necessitating complementary sensor redundancy.

Environmental noise and weak GPS signals in indoor environments may slightly affect system performance and navigation accuracy.

XIV. ETHICAL CONSIDERATIONS

The proposed system is designed with a focus on user privacy, accessibility, and safety. All processing is performed locally on the Raspberry Pi without transmitting visual data to cloud servers. The system is intended solely for assistive purposes and aims to improve independent mobility for visually impaired individuals.

XV. CONCLUSION

This paper presented "Alert Eye", an intelligent assistive navigation system using YOLOv11-based object detection, ultrasonic sensing, GPS tracking, and adaptive voice alerts. The proposed solution successfully integrates computer vision and embedded AI technologies to provide real-time environmental awareness for visually impaired users. Experimental evaluation demonstrated efficient real-time performance, acceptable detection accuracy, and low system latency on Raspberry Pi hardware. The proposed system offers a portable, cost-effective, and practical assistive solution for enhancing independent mobility and user safety.

X. FUTURE WORK

While the proposed system significantly improves outdoor navigation, future iterations will focus on addressing indoor navigational challenges where GPS signals are unreliable. This includes integrating Bluetooth Low Energy (BLE) beacons or Wi-Fi positioning systems for seamless indoor-outdoor transitions. Additionally, migrating the computational load to dedicated Edge AI accelerators (such as a Google Coral Edge TPU) will be explored to boost FPS and reduce power consumption. Software enhancements will include predictive analytics to anticipate object trajectories—rather than just detecting static presence—and expanding the TTS module to support multi-lingual voice alerts, making the device accessible to a broader global demographic. Future work will focus on improving indoor navigation support using Bluetooth Low Energy (BLE) beacons and Wi-Fi positioning systems. Additional enhancements may include multilingual voice support, advanced trajectory prediction, cloud-assisted

emergency communication, and dedicated Edge AI accelerators such as Google Coral TPU for improved performance and reduced power consumption..

REFERENCES

- [1] A. O. Khadidos and A. Yafoz, "An intelligent object detection and classification framework for assisting visually challenged persons using deep learning and improved crow search optimization," *Sci. Rep.*, vol. 15, no. 29822, 2025.
- [2] A. M. Alashjaee, H. N. AlEisa, A. A. Darem, and R. Marzouk, "A hybrid object detection approach for visually impaired persons using pigeon-inspired optimization and deep learning models," *Sci. Rep.*, vol. 15, no. 9688, 2025.
- [3] F. Shariff, G. Dilleeswari, B. S. Gowtham, B. Mounika, A. S. Sharmila, and S. Shanmathi, "Yolo-Based RealTime Object Detection with Voice Assistance for Visually Impaired Navigation," *Int. J. Res. Publ. Rev.*, vol. 6, no. 4, 2025.
- [4] S. Sriharan, S. Naik, V. Vaishnavi, T. Patil, and S. Rekha, "Assistive Device for Deaf, Dumb and Blind People," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 7, no. 5, 2025.
- [5] A. A. K. Ashar, A. Abrar, and J. Liu, "A Survey on Deep Learning-based Smart Assistive Aids for Visually Impaired Individuals," in *Proc. 7th Int. Conf. Inf. Syst. Data Mining (ICISDM)*, Atlanta, USA, May 10–12, 2023.
- [6] B. A. Bhat, G. S. B, S. Ravindra, A. Raghuveer, and N. Bangera, "Review on YOLO-based Mobility Assistance Systems for the Visually Impaired," *Int. J. Adv. Trends Eng. Manag.*, pp. 156–164, [Online].