International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69183

Voice Assistant for Blind

Sakshi Natuskar[1], Anisha Sable[2], Yashkumar Sharma[3], Simran Shende[4], Prof. Dr. P.Y. Itankar[5]
Department of Computer Engineering, Datta Meghe College of Engineering, Airoli, Navi Mumbai, India

Abstract: Navigating safely through everyday environments can be extremely challenging for people with visual impairments. To address this, we developed a smart Blind Voice Assistant that uses the YOLOv8 object detection model to identify surrounding objects in real time. The system can detect both common and potentially dangerous items, estimate their approximate distance in steps, and immediately inform the user through voice feedback. This approach allows users to become more aware of their surroundings and make safer decisions as they move around. By training the model with both standard COCO data and a custom dataset including items like knives, pens, and bags, we ensure the assistant is tuned to recognize important objects in daily life. Overall, the assistant aims to empower visually impaired individuals by offering real-time environmental awareness and guided navigation.
Keywords: Blind Assistant, Object Detection, YOLOv8, Voice Navigation, Step Estimation, Visually Impaired, Real-Time Guidance, Custom Dataset.

I. INTRODUCTION

For individuals who are visually impaired, even simple tasks like walking through a room or identifying nearby objects can pose significant challenges. In such scenarios, having real-time awareness of one's surroundings becomes essential for ensuring both safety and independence. While there are traditional tools like white canes and guide dogs, they come with limitations—either lacking detailed feedback or being costly and inaccessible to many. With advancements in artificial intelligence and computer vision, it is now possible to build smart systems that bridge this gap more efficiently and affordably.

This project introduces a Blind Voice Assistant that leverages the YOLOv8 object detection algorithm to recognize and announce objects in the user's environment. The assistant not only detects general objects but also identifies harmful items like knives or scissors and provides alerts based on their proximity, measured in steps. It combines the power of real-time object detection, distance estimation, and speech output to give users a better understanding of their surroundings. By training the model with both the COCO dataset and a custom dataset containing objects of daily relevance, the system becomes more tailored and reliable for practical use in everyday life.

II. LITERATURE SURVEY

Visual disability assistive technology has developed very much over the past few years, but despite this, major hurdles are encountered in making real-time and thorough support for wayfinding and detecting objects both in and outside. Google's TalkBack, one of the most widely used accessibility services of our times, is an Android feature that notifies the user about the contents of what's on the device screen with spoken descriptions, vibrations, and audio feedback. TalkBack enables users to comprehend what item they are choosing, touching, or engaging with, thus making mobile phones accessible to blind people. The system has significant limitations. In a 2019 paper published in the Journal of Emerging Technologies and Innovative Research (JETIR), TalkBack uses battery quickly and is inflexible when used. It keeps reading announcements even when the user tries to discontinue, which is disruptive and inconvenient. Additionally, TalkBack is confined to on-screen interaction and fails to take its capabilities to help users navigate environments or detect real-world objects, which are the core elements needed for independent mobility.

In order to overcome the shortcomings of such screen-based support tools, various researchers have come up with systems based on computer vision, machine learning, and sensor fusion that can offer more dynamic assistance. Thakkar et al. (2022) suggested a smartphone-based object detection system that employs image recognition to offer audio cues regarding the environment. While pioneering in mobile integration, the system does not possess real-time spatial awareness and is limited in dynamic environments. DeepNAVI, proposed by Kuriakose et al. (2023), improves this strategy by providing context-aware navigation with deep learning models for object classification and distance estimation. The major benefit of DeepNAVI is its offline mode, light architecture, and real-world usability for the blind without specific hardware or constant internet connectivity. Indoor navigation continues to be a problem, as GPS is not available indoors. Song et al. (2020) addressed this with a visual landmark-based navigation system using ORB (Oriented FAST and Rotated BRIEF) feature detection and pre-recorded environmental images.

The solution, effective as it is in controlled settings, is not able to dynamically adapt to new or changing environments. Verma et al. (2016) proposed a panoramic image-based mapping system in conjunction with dead-reckoning for indoor path planning. Despite scalability and simplicity in operation through a smartphone, requiring ahead-of-time image capture and marking restricts real-time uses.

Additionally, Naser et al. (2023) presented an overview of smartphone-based indoor localization systems, in which the cruciality of using IMU sensors (Inertial Measurement Unit sensors) and PDR (pedestrian dead reckoning) was mentioned. The review cited that variations in walking patterns and phone orientation affect localization precision extensively, indicating a requirement for context-sensitive and adaptive algorithms. At the same time, Shen et al. (2022) developed a wearable navigation system based on a real-time object detection algorithm (YOLOv3) and haptic feedback using shape memory alloy (SMA) actuators. This device exclusively provides multimodal communication (touch rather than sound), which is useful in loud city environments where audio cues might not be effective.

A more elaborate solution was introduced by Duh et al. (2020) with V-Eye, a wearable navigation assist that integrates monocular camera input with visual SLAM (Simultaneous Localization and Mapping). This system fills the gap between indoor and outdoor navigation by providing dynamic obstacle detection and spatial awareness, a stark improvement over TalkBack's on-screen attention. Litoriya et al. (2023) also enhanced object detection capacity by combining YOLO and SSD models in a visual assistant system for the blind. Their system enhances object recognition in complicated situations and offers audio alerts to navigate the user. While still relying on commonly used general-purpose datasets such as COCO and VOC, the system may lack objects of certain interest to blind people, like potholes, curbs, or suspending obstacles.

For a more straightforward use, Salunkhe et al. (2021) showed how a real-time object detector using a TensorFlow Lite-based Android app can employ a mobile camera and issue voice warnings. This method is a convenient and portable solution but has limited capabilities to sense motion or depth and performs poorly on dense scenes involving multiple overlapping objects. Another important contribution by Naser et al. (2022) highlighted the importance of multi-sensor fusion and adaptive algorithms in enhancing indoor positioning and navigation, advocating that future systems need to be adaptable enough to accommodate user behavior and environmental fluctuation. Together, these studies indicate an increasing trend away from static, screen-based systems such as TalkBack towards intelligent, real-time, and multimodal navigation tools. Yet none of the existing solutions fully combines object detection, dynamic spatial orientation, voice-controlled commands, and safety alarms in one low-cost, easy-to-use device. Most systems address one or two features—e.g., object recognition or indoor positioning—without tackling the entire range of needs experienced by blind users. Hence, there is an urgent need for a single platform that provides real-time awareness of the environment, voice and audio-based interaction, and context-aware navigation in known and unknown spaces. The system should try to bridge the loopholes of current solutions such as TalkBack, providing not only interaction with mobile devices but actual environmental awareness and independent mobility for the visually impaired.

III. PROPOSED SYSTEM

A. *Voice Assistant for Blind Overview*

Vision is essential for navigating and understanding the world, and its absence creates daily challenges for the visually impaired. This project introduces a smart voice assistant that uses real-time object detection and spatial feedback to enhance safety and independence. It combines technologies like YOLOv8 for object recognition, Places365 for scene understanding, and Tesseract for reading text. The system runs offline, using a webcam, a standard laptop, and pyttsx3 for voice output. A simple GUI ensures ease of use with minimal interaction. Overall, it offers a practical, all-in-one solution to help blind users engage confidently with their surroundings.

B. *Framework*

The Framework Components are:

1) *Computer Vision*: The live feed from the user's webcam is interpreted using OpenCV, which streams real-time frames to the detection module.
2) *YOLOv8 Detection*: YOLOv8 processes frames from the webcam, detects multiple objects in one go, and identifies their types and positions
3) *Hybrid Datase*t Integration: The model is compatible with both the default COCO dataset (80 common object categories) and a custom-trained dataset for visually impaired users, such as objects pen, wallet, spectacles, or local signs.
4) *Text-to-Speech (TTS)*: The processed data is then converted to speech using the pyttsx3 library, which is offline-based and has the assurance of working even in environments with no internet connectivity.

The Major Features of YOLOv8 in the System are:
1) *Speed:* Processes one frame per millisecond.
2) *Accuracy:* Employs enhanced anchor-free detection and bounding box regression.
3) *Flexibility:* Enables detection of general as well as custom-defined objects.

This analytical framework underpins the continuous execution of the assistant and keeps the user abreast of new, context-sensitive insights per second.

C. *Workflow of the Platform*

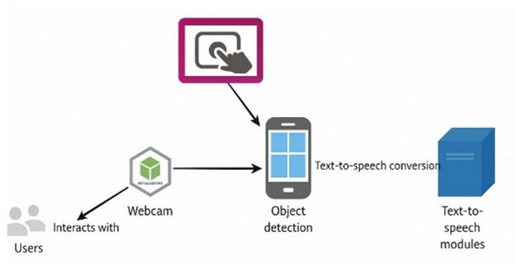

Fig. 1 Process Flow of Blind Assistant

The Voice Assistant workflow is built to offer a seamless experience to visual impaired individual. The main steps are:
1) *System Initialization:* User starts the system via button press or voice command. System loads the object detection models (YOLOv8) and initializes the TTS engine (pyttsx3).
2) *Frame Capture & Preprocessing:* Webcam captures continuous video frames. Frame is captured using OpenCV. Resized and normalised to match YOLOv8 input requirement.
3) *Object Detection (YOLOv8):* The pre-processed frame is fed into the YOLOv8 model. The model output: Object (e.g., "bottle", "person"), Bounding Boxes, and Confidence scores.
4) *Spatial Localization:* The frame is divided into zones: Left, Center and Right. Objects are mapped to these zones based on bounding box location. Estimated distance (steps) is calculated using object size and known distance metrics.
5) *Object Classification & Priority Handling:* Object classified using both COCO Datasets and Custom Dataset. Dangerous items (e.g., knife, fire) are flagged for priority alerts
6) *Decision Making & Message Generation:* Based on object type and location, context-aware messages are generated: "Person 3 steps ahead" or "Knife detected on your left — danger!"
7) *Voice Feedback Output:* Once Messages are passed to pyttsx3. Voice feedback is generated offline and played immediately. Feedback is synchronized with object detection for real-time awareness.
8) *GUI Interface (Optional):* Minimal GUI shows: Start/Stop buttons. Live or minimized video preview (for debugging or assistance).
9) *Continuous Loop:* Steps 2–8 repeat continuously, providing live feedback. The loop runs until: User says "exit" OR Clicks the Stop button.
10) *System Termination*: Webcam and model resources are released. Voice engine is closed.

D. *Core Technologies Used*

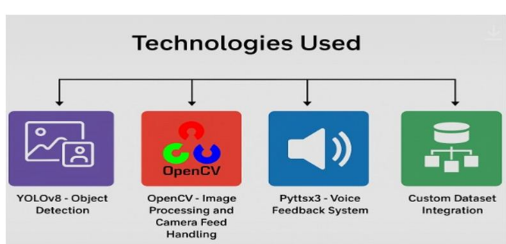

Fig. 2 Technologies Used

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

1) YOLOv8 (You Only Look Once version 8) – Object Detection

YOLOv8 is a powerful and real-time object detection algorithm that forms the core of the assistant. It detects and identifies objects in the camera's view with high accuracy. The model is trained using a combination of the COCO dataset and a custom dataset to ensure both general and specific objects (like pens, knives, bags) are recognized. YOLOv8 allows efficient processing, making it suitable for real-time applications with quick and reliable object classification and localization using bounding boxes and confidence scores.

2) OpenCV – Image Processing and Camera Feed Handling

OpenCV (Open Source Computer Vision Library) is used to capture real-time video from the webcam, preprocess the frames, and pass them to YOLOv8 for object detection. It enables drawing bounding boxes, calculating object positions, and estimating distances in the frame. OpenCV is essential for integrating visual processing into the system in an efficient and lightweight manner.Node.js: A JavaScript runtime for handling server-side logic, API endpoints, and data processing.

3) Pyttsx3 – Voice Feedback System

Pyttsx3 is an offline text-to-speech library that converts detected object data and navigation alerts into audible voice feedback for the user. It works without requiring internet access, making it a reliable and consistent tool for the visually impaired to hear the names and distances of detected objects.

4) Custom Dataset Integration

A custom dataset is created to train the YOLOv8 model for recognizing domain-specific and critical objects, especially harmful or essential items like knives, pens, bottles, and bags. This ensures the system provides safety-specific alerts in addition to general object identification.

5) Python – Programming Language

Python is the primary language used to integrate all modules including YOLOv8, OpenCV, pyttsx3, and the dataset. Its simplicity, vast libraries, and flexibility make it ideal for developing AI-based voice assistant systems.

IV. OUPUT

Fig. 3 Book & Pen identified output

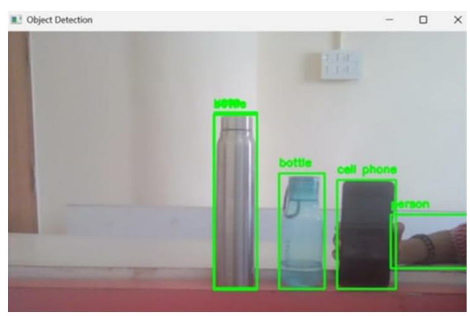

Fig. 4 Multiple object detected

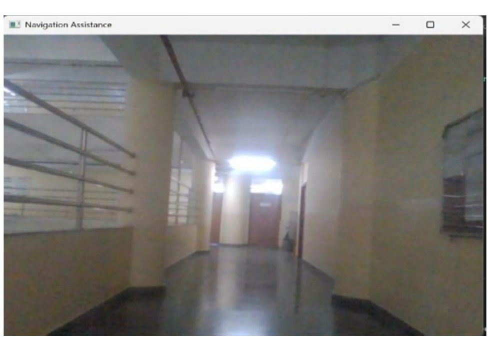

Fig. 5 Display of the Navigation Assistance

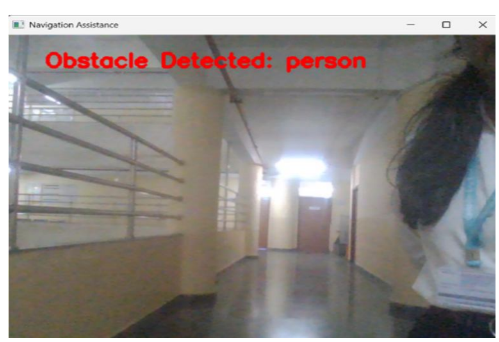

Fig. 6 Warning of Obstacle Detected in the Navigation Assistance

V. CONCLUSIONS

The "Voice Assistant for Blind" project seeks to create a more inclusive environment for the visually impaired by integrating computer vision, deep learning, and text-to-speech technologies. Utilizing YOLOv8 for real-time object detection and both default and custom datasets, the system provides customized auditory feedback to assist users with navigation in unfamiliar or dynamic environments. Its capability ensures consistent performance independent of internet connectivity, and a webcam-based visual sensor, optimized processing, and voice or button interface ensure it's both functional and easy to use. From detecting common objects in daily life to alerting one to possible threats, the assistant serves as an electronic companion that facilitates mobility, safety, and independence—testifying to the significant role that artificial intelligence can have in assistive technology.

REFERENCES

[1] Bineeth Kuriakose, Raju Shrestha, Frode Eika Sandnes, "DeepNAVI: A deep learning-based smartphone navigation assistant for people with visual impairments," *Expert Systems with Applications*, vol. 212, 2023, Article ID 118720.

[2] Shivam Verma, Rohit Omanwar, Sreejith V, Meera G S, "A Smartphone Based Indoor Navigation System," Proceedings of the IEEE International Conference on Mechatronics (ICM).

[3] D. Ravi Kumar, Hiren Kumar Thakkar, Suresh Merugu, Vinit Kumar Gunjan, Suneet K. Gupta, "Object Detection System for Visually Impaired Persons Using Smartphone," Lecture Notes in Electrical Engineering, vol. 761, Springer, 2022, pp. 1561–1572

[4] Ping-Jung Duh, Yu-Cheng Sung, Liang-Yu Fan Chiang, Yung-Ju Chang, Kuan-Wen Chen, "V-Eye: A Vision-based Navigation System for the Visually Impaired," IEEE Transactions on Multimedia, 2020.

[5] Fanghao Song, Zhongen Li, Brian Clark, Dustin Grooms, Chang Liu, "Camera-Based Indoor Navigation in Known Environments with ORB for People with Visual Impairment," *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, 2020.

[6] Akilesh Salunkhe, Manthan Raut, Shayantan Santra, Sumedha Bhagwat, "Android-based object recognition application for visually impaired," *ITM Web of Conferences*, vol. 40, 03001, 2021.

[7] Junjie Shen, Yiwen Chen, Hideyuki Sawada, "A Wearable Assistive Device for Blind Pedestrians Using Real-Time Object Detection and Tactile Presentation," *Sensors*, vol. 22, no. 12, 2022.

[8] Rana Sabah Naser, Meng Chun Lam, Faizan Qamar, B. B. Zaidan, "Smartphone-Based Indoor Localization Systems: A Systematic Literature Review," *Electronics*, vol. 12, no. 8, 2023.