International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Visassist

Ms. Swati Hadke[1], Devyani Belsare[2], Ashwini Sarode[3], Ayurshi Moudekar[4], Dewyani Date[5], Mangesh Ghayal[6], Rishiket Deshmukh[7]

[1]Associate Professor, Department of Electronics Telecommunication Engineering, G H Raisoni Collage of Engineering & Management, Nagpur, Maharashtra
[2]Student, Department of Electronics & Telecommunication Engineering, G H Raisoni Collage of Engineering & Management, Nagpur, Maharashtra

Abstract
VisAssist is an innovative, portable, and low-cost system for enhancing independence and mobility in the visually impaired. The integrated hardware-software architecture will ensure real-time environmental awareness and detection of obstacles. A Raspberry Pi micro-controller acts as a central processing and control unit, while ultrasonic sensors and a camera module are interfaced to it for data acquisition and environmental analysis. These ultrasonic sensors estimate object distances from the individual, and the camera provides relevant information for contextual understanding of those objects. As soon as an obstacle is detected within the threshold range, immediate audio feedback is provided through earphones or a speaker linked with a TTS engine. The design ensures that the system can work exclusively offline without any dependency on internet connectivity and has been tested for both indoor and outdoor environments. Compactness, lightweight, and energy efficiency further support its suitability for daily use. With embedded system design, sensor fusion, and speech processing, VisAssist enhances user safety, situational awareness, and self-reliance, hence adding a new dimension to assistive technology for visually impaired people.

1. Introduction

Visual impairment is one of the main obstacles that hinder the lives of millions of people across the globe, and they find it very difficult to navigate through their daily activities and recognize objects. A lot of the time, the blind and low vision people are helping others which results in less freedom for them. The usual aids like walking sticks or guide canes only give a very rough idea of the surroundings and offer no specific details about the ambiance. The Visual Assistive System deals with these problems by detecting obstructions and giving audio directions to the user instantly. A Raspberry Pi board with a camera and ultrasonic sensors will be used in the system for detecting objects in the way. As soon as an obstruction is identified, an audio warning is sounded through Text-to-Speech output. This not only provides the user with an understanding of the environment but also allows the user to move safely without the help of others. The system is made portable, friendly for operation, and cost-effective which are the factors that make it ideal for daily application by the sightless. The Visual Assistive System, through the combination of straightforward electronic components and efficient feedback, offers improved safety, mobility, and confidence thus allowing a more independent lifestyle for the visually impaired.

2. Literature review

1. Ratnesh Litoriya, Kailash Chandra Bandhu, Sanket Gupta, Ishika Rajawat, Hany Jagwani, Chirayu Yadav, Hany Jagwani, and Chirayu Yadav, "Implementing Visual Assistant using YOLO and SSD for Visually-Impaired Persons," Journal of Automation, Mobile Robotics and Intelligent Systems, This study Developed a low-cost real-time object detection system on Raspberry Pi 3 using YOLO and OpenCV, providing accurate recognition and audio feedback for visually impaired navigation.
2. Wei Wang ,Bin Jing ,Xiaoru Yu , Yan Sun, Liping Yang, Chunliang Wang, "YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance," Sensors journal, This paper Proposed an obstacle detection model named YOLO-OD, introducing a Feature Weighting
3. Mrs. Rashmi BN, Dr. R. Guru, Dr. Anusuya MA, "Small Object Detection for Indoor Assistance to the Blind using YOLO NAS Small and Super Gradients," arXiv preprint, The authors addressed the challenge of recognizing small indoor objects in complex environments. Their proposed YOLO NAS Small architecture improved accuracy and reduced computational overhead, making it efficient for embedded assistive systems deployed indoors.
4. Usman masud, Tareq saeed, Hunida m. Malaikah, Fezan ul Islam and Ghulam Abbas, "Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification," IEEE, This paper introduced low-cost, portable, offline, efficient (91% accuracy), and designed with Raspberry Pi with Arduino, ultrasonic sensors, and camera, combining classical (Viola–Jones) and modern (TensorFlow) object detection for assisting visually impaired users.
5. Abijith R. Nair & Sunitha S. Nair, "A Real-Time Object Detection System for Assistive Navigation in the Visually Impaired," International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), Developed a real-time object detection system for assistive navigation, addressing dataset diversity, low-power model optimization, and multi-modal feedback for fast, practical deployment.

3. Problem Statement

Individuals with visual impairments face severe problems while perceiving and interacting with their surroundings. The inability to detect and recognize obstacles in proximity often leads to mobility challenges, reduced independence, and a higher likelihood of accidents. In spite of several assistive technologies, most are either costly, limited in real-time functionality, or dependent on external connectivity. Furthermore, most traditional tools, like the walking cane or basic ultrasonic devices, detect only distance-based alerts without identifying the nature of the obstacles. Therefore, there has been an essential demand for a low-cost, portable, and intelligent assistive device able to accurately detect, recognize, and communicate information about nearby objects in real time, thus contributing to safe navigation for visually impaired subjects.

4. Proposed Solution

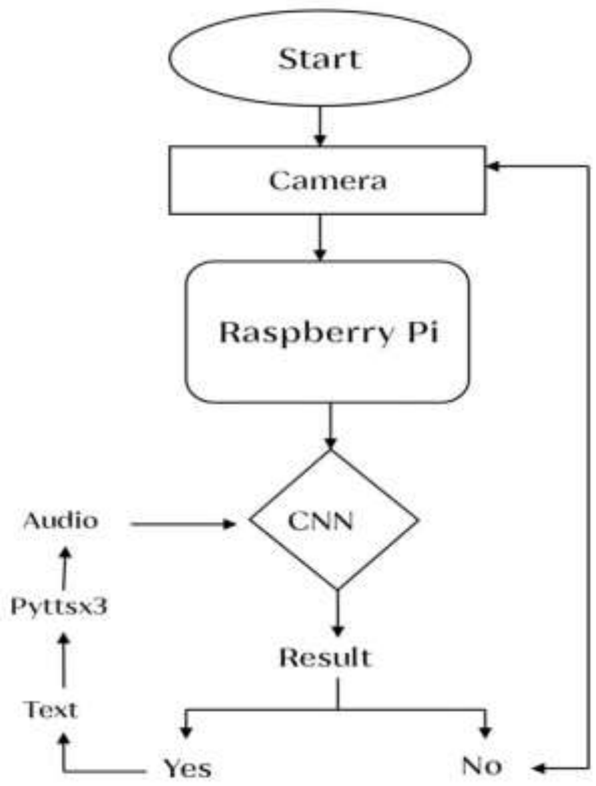
This system project proposes a Visual Assistive System that uses object detection and text-to-speech technologies to provide real-time environmental awareness. The system will use a Raspberry Pi as the processing unit and a camera module that captures live video input. It will be using YOLOv11, a deep learning model trained on the COCO dataset, for detecting and classifying multiple objects in real time. Using the pyttsx3 developed Text-to-Speech engine, the names of detected objects are converted into speech, giving immediate audio feedback through headphones for any blind user to perceive their surroundings with auditory signals.

The proposed design emphasizes low cost, portability, and offline functionality that guarantees users will have independent navigation without dependency on internet connectivity. This blend of computer vision and speech output promotes special awareness, safety, confidence, and inclusion in an accessible assistive technology solution.

The features of the proposed system are listed below:

a. Utilizes YOLOv11, a deep learning model trained on the COCO dataset, to detect and classify 80+ common objects in real time.
b. Integrates pyttsx3 Text-to-Speech engine to convert detected object names into audible voice output.
c. Provides audio alerts through Bluetooth earphones, enabling users to understand their surroundings via sound.
d. Ensuring no dependency on internet connectivity.
e. Employs a Raspberry Pi 4 as the main controller for image acquisition, object detection, and audio processing.
f. Captures live video input from the user's environment to be processed in real time.

5. Flow Chart



6.Block Diagram

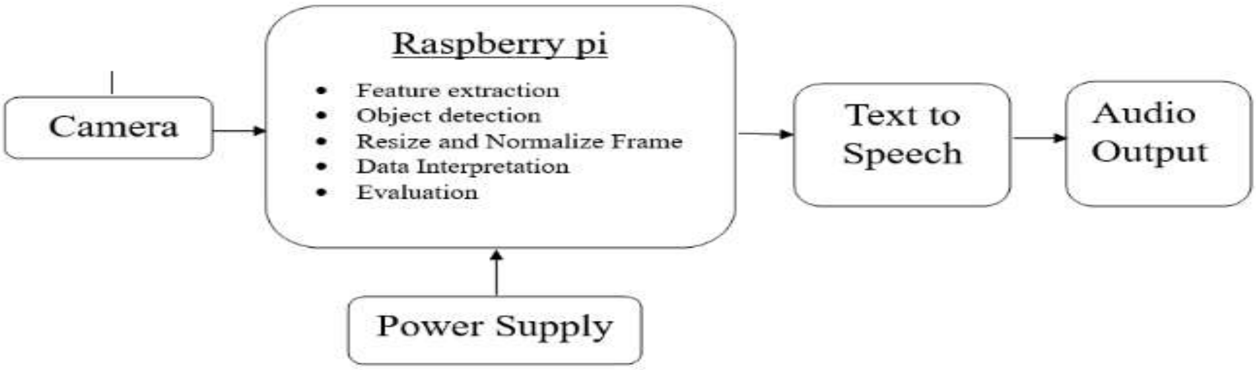


Fig. Block Diagram of Proposed Syste

7. Working

For the best performance in our system, we designed a system to help visually impaired people. The proposed Visassist system uses interconnected subsystems: a stabilized power delivery module, an image acquisition module, core processing, and output generation. The regulated power supply ensures continuous, noise-free power for the Raspberry Pi.

The Camera Input Block uses the image acquisition interface to capture real-time video frames from the environment through the OpenCV library and sends them to the Central Processing Unit. This is the first visual data processed using OpenCV. OpenCV is used to activate the camera, analyze each frame, and ensure high-quality image input.

The Camera Input Block uses the image acquisition interface to capture real-time video frames from the environment through the OpenCV library and sends them to the Central Processing Unit. This is the first visual data proceThe Core Processing Module runs on the Raspberry Pi and serves as the main Central Processing Unit. It performs feature extraction, object detection, frame resizing, normalization, data interpretation, and evaluation. Initially, raw frames are converted into a suitable format using the OpenCV library for resizing and the NumPy library to normalize pixel values (0-1 range) for improved accuracy. Feature extraction identifies and extracts features, such as edges, corners, shapes, and textures, by using convolutional layers in a CNN based on the YOLO architecture CNN acts as the main computational framework for object detection and recognition, trained to develop a distinct visual understanding. The CNN relies on the YOLO deep learning architecture from ultralytics to provide the YOLOv11 model for real-time object recognition and can identify multiple object classes in a single frame with high accuracy. Torch and torchvision support deep learning computations with GPU/CPU acceleration,. ensuring faster and optimized processing of image data. Both libraries are part of the PyTorch framework. Data interpretation organizes structured data through labels and positional information. Libraries such as Pandas and Scipy assist in interpreting the results and evaluating each detected object with meaningful descriptions.

Finally, the output generated from the evaluated textual representation is sent to the text-to-speech (TTS) module. Symbolic text is transformed into parametric acoustic signals. The Pyttsx3 library is used to synthesize the symbolic textual results, providing clear and natural audio feedback.

8. Hardware: -

Raspberry Pi 4 model B: It should serve as the main processing unit for the system and supports Python programming, deep learning libraries like YOLOv11, and has sufficient computational power for real-time image processing, is compact, and energy-efficient.

.

Fig. Raspberry Pi 4 model B

Web Camera: Captures real-time video frames from the user's surroundings.It Offers high-resolution input for accurate object detection at low light levels, hence guaranteed performance both indoors and outdoors..

Fig. Web Camera

Bluetooth Earphone: .Provides audio feedback to the visually impaired user and provides wireless communication, free movement, and clear audio output to the user's ear for comfort and ease.

Fig. Bluetooth Earphone

Power Bank: It provides stable and continuous power to the Raspberry Pi and camera. Guarantees noise-free voltage supply and avoids system fluctuations in long operations.

Fig. Power Bank

SD Card (32 GB): Stores operating system, object detection model, and code files.Sufficient storage and fast read/write speeds ensure quick system performance.

Fig. SD Card

9. Result Analysis

The VisAssist system was designed and prototyped successfully as an assistive smart system for the visually impaired, and experimentally tested as such. The system combines deep learning and computer vision methodologies in an embedded computing platform to deliver real-time environmental perception and audio feedback to the user.

The power supply subsystem, which was regulated, provided the noise-immunity stable 5 V/3 A regulated power supply to the Raspberry Pi 4 Model B, the processing unit. The stable voltage regulation kept transient fluctuations and electromagnetic interferences at bay, providing a reliable operation of the hardware and calculation unbroken under intense image processing.

The acquisition subsystem, based on a Zebronics HD 720p webcam connected through OpenCV, efficiently captured live video streams with zero latency All the frames were passed through pre-processing operations such as normalization (via NumPy) and spatial resizing for data homogenization and limiting computational load during model inference.

In the central processing module, a pipelined data structure was used to execute sequential operations such as feature extraction, object localization, and classification. The YOLOv11 model, whose implementation was facilitated through the Ultralytics framework, was the primary deep neural network utilized for object detection. The one-stage CNN architecture successfully predicted a single bounding box and some class probabilities per frame for real-time multi-object detection. The network was extremely successful when it detected more than 80 classes from the COCO dataset, with a very high precision and recall rate at the time of evaluation.

The step of feature extraction in the CNN effectively extracted low- and mid-level spatial features such as edges, contours, and textures to have accurate contextual scene understanding. Use of GPU-based computation and PyTorch and Torchvision libraries also improved inference speed and system performance to enable smooth real-time execution with negligible loss of frames.

The processed results were then translated and formatted into human-readable text labels for object names and confidence values that were detected. These were then written out as audio feedback through the Text-to-Speech (TTS) module using the Pyttsx3 library. The TTS subsystem produced natural and comprehensible speech, providing the user with continuous situational awareness of objects and obstacles surrounding them.

Experimental trials proved that the VissAssist system would detect obstacles and known objects with high accuracy in a distance from 2 cm to 400 cm under reliable detection consistency in various lighting conditions and environments. Comparative observations showed the developed prototype to have high responsiveness, low power consumption, and real-time processing, proving it suitable as a low-cost, portable, and efficient assistive device.

Fig. Result

10. Conclusion

The Visual Assistive System presented in this project brings about an innovative and feasible solution for the possibly lifting of the mobility and independence of persons with visual impairment. The system, through the use of sophisticated technology such as Raspberry Pi, camera, and YOLOv11 object detection algorithm, can identify and track more than 80 common objects in real-time. The informed objects are converted into speech output using Text-to-Speech (TTS) technology allowing the user to recognize his/her surroundings through auditory feedback. The system is designed to be handy, cost-effective, and efficient in comparison with existing costly devices. With continuous testing and tuning, the project convincingly illustrates the integration of artificial intelligence and embedded systems for the purpose of enhancing accessibility. In a nutshell, the project helps to shape a more user-friendly, safer and inclusive environment for the visually impaired population, thus highlighting the role of technology in the improvement of life quality

11. Future Scope

The Technology may shoulder the development of the Visual Assistive System into a more intelligent, dependable, and user-friendly technology. The incorporation of GPS and navigation in the system can indeed provide the visually impaired with independent outdoor movement. The system can inform a user of his position in real-time. Cloud-based AI will be an important contributor to the enhancement of the object detection accuracy and reduction of response time, thus boosting the fastness and the efficiency of the system. A vocal interface can also be introduced for the users to have their conversations in natural ways without the need for manual operations. The hardware can be re-configured into miniature and portable forms like smart spectacles or light headsets providing convenience and comfort during daily usage. Connectivity to a mobile app can be another beneficial addition, as it would entail live location sharing, alert notifications, and navigation tracking of the caregiver or family. With the above-mentioned innovations, the Visual Assistive System can be an all-inclusive, stand-alone assistive system that not only offers mobility but also creates for the user a world that is safe, aware, and confident. In short, such a system would assist blind people in living a more independent, more connected life, thus bridging the gap between technology and accessibility.

12. References:

1. Litoriya, Kailash Chandra Bandhu, Sanket Gupta, Ishika Rajawat, Hany Jagwani, Chirayu Yadav. Implementing Visual Assistant using YOLO and SSD for Visually-Impaired Persons. Journal of Automation, Mobile Robotics and Intelligent Systems, Vol. 17, Issue 4/2023, pp. 79–87, 2024. DOI:10.14313/JAMRIS/42023/33
2. Changchun University Researchers. YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance. Sensors, Vol. 24(23), Article no. 7621, 2024. DOI: 10.3390/s24237621
3. Rashmi BN, R. Guru, Anusuya MA. Small Object Detection for Indoor Assistance to the Blind using YOLO NAS Small and Super Gradients. arXiv preprint, 2024. arXiv:2409.07469
4. Pramod Kumar Sagar, Anant Mishra, Akhilendra Mishra. Object Detection for Blind Person. International Journal of Modern Developments in Engineering and Science (IJMDES), Vol. 4, No. 5, pp. 15–18, 2025. Link
5. Abijith R. Nair & Sunitha S. Nair. A Real-Time Object Detection System for Assistive Navigation in the Visually Impaired. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering (IJIREEICE), 2025. DOI: 10.17148/IJIREEICE.2025.13807
6. Usman masud, Tareq saeed, Hunida m. Malaikah, Fezan ul Islam and Ghulam Abbas. Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object, IEEE, 2022 DOI: 10.1109/ACCESS.2022.3146320

7. Swati C.Hadke, Ravi Mishra, Shot boundary detection in video using dual-stage optimized VGGNet based feature fusion and classification. Multimed Tools Appl 83, 34417–34444 (2024). https://doi.org/10.1007/s11042-023-16982-7
8. Swati C.Hadke, Ravi Mishra, An attention driven long short term memory based multi-attribute feature learning for shot boundary detection. Knowledge-Based Systems, Volume 317, 23 May 2025, 113379. https://doi.org/10.1016/j.knosys.2025.113379