iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IX **Month of publication:** September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74351

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 IX Sep 2025- Available at www.ijraset.com

A Survey on Smart Assistive System for Visually Impaired People Obstruction Avoidance through Object Detection and Classification

Dr. Anitha TG[1], Anagha P[2], G. Rashmithaa Prabhu[3], Gauri B Nair[4], Geetha K[5]
[1]*Professor, Dept of CSE, Sapthagiri College of Engineering*
[2, 3, 4, 5]*Dept of CSE, Sapthagiri College of Engineering*

Abstract: Object detection is becoming a crucial part of many modern technologies, including intelligent surveillance to driverless cars. In this study, we investigate its function in developing an intelligent assistive framework for people with visual impairments. The technology combines audio synthesis, multilingual translation, and image captioning to give contextual scene narration and real-time blockage detection. A webcam captures environmental visuals, which are processed using the BLIP Transformer for scene description and the YOLOv8 deep learning model for object and traffic sign recognition. Multilingual audio feedback is generated using gTTS, while a voice-controlled interface ensures hands-free interaction. To improve efficiency, a cool-down mechanism prevents repetitive alerts. This integrated strategy improves safety and freedom for visually impaired users by providing semantic comprehension and language-inclusive narration, in contrast to previous approaches that only addressed obstacle location

I. INTRODUCTION

Applications of computer vision extend across domains such as autonomous navigation, visual recognition systems, face detection, and more. In this work, we focus on utilizing object detection techniques to assist individuals who have visual disabilities. The proposed solution empowers them to identify surrounding objects, thereby supporting safer and more independent mobility.

This paper introduces a Smart Assistive System for Visually Impaired People, designed to enable real-time obstacle avoidance through the integration of vision and speech technologies. A webcam is used for scene acquisition, where the BLIP Transformer model generates semantic image captions. For multilingual narration, these captions are subsequently translated into several languages and combined into voice. Simultaneously, the YOLOv8 deep learning model performs object and traffic sign detection, providing instant auditory alerts for safety-critical events. A voice-controlled interface ensures hands-free language selection, while a cool-down mechanism prevents redundant notifications, ensuring smoother interaction.

Our method stresses descriptive narration of the environment to improve spatial awareness, in contrast to many other techniques that just concentrate on item localization. Although earlier research on multilingual picture captioning [4] has made images more accessible, their dependence on cloud-based translation frequently results in latency. In a similar vein, hands-free operation was made possible by speech-guided navigation systems [5], although they lacked thorough scene perception. These drawbacks emphasize the need for a coherent structure that incorporates multilingual assistance, environmental narration, and real-time object identification into an affordable design.

More than 285 million people worldwide suffer from vision impairment, including about 39 million who are totally blind, as reported by the World Health Organization. Independent navigation is one of their most pressing challenges, as navigating dynamic and unstructured environments without aid often poses serious safety risks. Traditional mobility aids like canes and guide dogs have limitations in detecting obstacles and providing contextual information. To this gap, researchers have increasingly turned to AI-driven assistive technologies that integrate computer vision and natural language processing for real-time situational awareness

In this work, we propose a unified system combining image captioning, multilingual translation, object detection, and speech synthesis. The BLIP Transformer is employed to generate natural-language scene descriptions, that gets eventually translated into regional and international languages (such as Kannada, Hindi, Tamil, and French) and converted into audio using gTTS. Meanwhile, YOLOv8 detects safety-critical objects including vehicles, pedestrians, and traffic signs, issuing instant auditory warnings. A voice-controlled interface enables hands-free language selection, while a cool-down strategy prevents repetitive alerts, ensuring efficient interaction.

Key Contributions:

- Design of a real-time vision-to-voice framework integrating image captioning with object detection.
- Implementation of multilingual narration for inclusivity.
- Development of a voice-based interface to support hands-free accessibility.
- Introduction of a cool-down redundancy management mechanism to minimize repetitive alerts.
- Demonstration of broader applicability to domains such as driver assistance, smart surveillance, and autonomous mobility systems.

Earlier studies in this area explored ultrasonic sensor–based obstacle detection [1], which proved useful for short-range navigation but lacked the ability to classify or describe objects. Vision-based approaches using algorithms like Viola–Jones [2] extended functionality to object recognition but were less effective in low-light or crowded settings. Later, Raspberry Pi and OpenCV–based systems [3] enabled camera-based assistive sticks, though they struggled with real-time processing and required significant computational resources. Recent developments employing deep learning models including TensorFlow [6] and YOLOv3 [7] improved detection rapidity and precision.

However, these approaches still faced challenges like limited multilingual support, heavy hardware requirements, and redundant alerting-mechanisms.

II. LITERATURE SURVEY

1) *Ultrasonic Sensors Based Obstacle Avoidance – S. Patel and K. Sharma (2014):* This foundational work introduced a basic smart cane equipped with ultrasonic sensors and an Arduino microcontroller to detect obstacles and alert the user via vibration. The project was low-cost and effective in close-range obstacle avoidance. However, it lacked the ability to classify or describe objects and provided no auditory feedback, limiting its usefulness in complex navigation.
2) *Vision-Based Navigation System Using Viola–Jones Algorithm – M. Castrillón et al. (2015):* This study implemented object detection through the Viola–Jones framework, primarily focused on facial recognition and simple object detection for assistive technologies. The method was lightweight and fast but ineffective in low-light or crowded environments, and lacked real-time contextual understanding or voice interaction.
3) *Smart Walking Stick with Object Detection that uses Raspberry Pi – A. Gupta and R. Singh (2016):* In order to detect objects in real time, this project combines OpenCV with a Raspberry Pi and a camera module. The system aimed to enhance traditional canes with basic vision-based object recognition. Although it offered better object awareness than ultrasonic-only systems, it lacked advanced classification and was too resource-intensive for portable use.
4) *Visually Impaired Mutlilingual Image Captioning – A. Mehta and R. Varma (2017):* Using the Google Translate API, this study created a captioning system that produced scene descriptions and translated them into other languages. The device provided audio narration and was designed to provide accessibility for visually impaired people in multilingual areas. However, during real-time operation, it displayed latency and was highly dependent on internet access.
5) *Speech-Guided Navigation Using TTS and Speech Recognition – S. Bhattacharya and L. Ramesh (2018)***:** Through the use of voice commands and text-to-speech (TTS) feedback, this study demonstrated a speech-based navigation interface that enabled users to engage with their surroundings. It enabled hands-free use and simplified operation. However, visual scene perception was poor and performance decreased in noisy situations
6) *Object Recognition Using TensorFlow – N. E. Albayrak (2019):* A real-time object detection system using TensorFlow's pretrained models was implemented to assist the visually impaired. It could recognize common objects offline and provided a basic level of interaction. However, its recognition scope was limited and retraining was necessary to handle diverse environments effectively
7) *Detection Objects in Real Time for People with Visual Disabilities – S. Mishra and P. Asthana (2020):* This system leveraged the YOLOv3 deep learning model to detect and announce nearby obstacles. It provided real-time voice alerts to aid navigation. The approach significantly improved user awareness, but required reliable hardware and internet access, and lacked multi-language support
8) *Helmet Detection Using AT-YOLO Deep Learning Model – Q. Zhou et al. (2021):* Despite being intended for industrial safety, this system introduced AT-YOLO, a precise and portable object identification paradigm that can be used with edge devices. In assistive applications, it showed promise for low-latency object identification. Nevertheless, it was restricted to a small domain and lacked generic object classification and visual impairment optimization.

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 IX Sep 2025- Available at www.ijraset.com

9) *Edge AI System for Smart Glasses Navigation – B. Yadav et al. (2023):* This study suggested a wearable AI solution for smart glasses that uses the Jetson Nano and YOLOv5 to recognize and categorize items in real time. The technology was quick and offline, making it perfect for navigating by those with visual impairments. Its high energy consumption and hardware cost were its main drawbacks, making it less accessible to the typical user.
10) *Enhanced Multilingual Scene Understanding and Voice Feedback for the Blind – K. Naqvi, B. Hazela, S. Mishra (2024):* This recent work explored combining scene understanding through object detection with multilingual voice feedback. It employed pretrained models for object classification and used gTTS for audio narration. The system was accessible and customizable, but suffered from occasional detection errors and performance drops under high frame rates or low-light conditions.

III. METHODOLOGY

1) Data Acquisition**:** Live video is continuously captured using a webcam or camera module. The system ensures real-time input streams suitable for object detection and scene understanding.
2) Preprocessing: The video frames are resized, normalized, and enhanced to handle variations in brightness and noise. This stage guarantees that the deep learning models perform at their best under a range of environmental circumstances.
3) Scene Captioning with BLIP Transformer: The BLIP (Bootstrapping Language-Image Pretraining) model is employed to generate descriptive captions for the captured frames. These subtitles go beyond basic obstacle recognition to offer semantic comprehension of the scene. Captions are translated into the user's preferred language using Google Translate API, enabling multilingual accessibility. The translated text is converted into audio using Google Text-to-Speech (gTTS) and played through an earphone or speaker.
4) Object Detection and Classification with YOLOv8: For real-time object identification and classification, a parallel module uses the YOLOv8 deep learning model to recognize objects including cars, people, and traffic signs. This step enhances user safety by identifying obstacles and critical navigation cues such as pedestrian crossings or approaching vehicles.
5) Alert Generation and Cool down Mechanism: The system uses Pygame and gTTS to create a voice alert whenever it detects an obstruction or significant object. To prevent redundant or overwhelming notifications, a cool down mechanism ensures that repeated alerts for the same object class are only triggered after a predefined interval. This reduces confusion and improves the user experience.
6) Voice-Controlled User Interaction: The system features a voice-controlled interface, enabling users to interact hands-free via speech recognition. For people with visual or movement disabilities, this enables improved accessibility, language selection, and system control.

a) System Architecture: The system architecture is organized into three functional layers: Input Layer: Includes camera and microphone modules for capturing video and voice commands.
b) Processing Layer: Handles image captioning, object detection, translation, and speech synthesis.
c) Output Layer: Delivers audio-based scene descriptions and alerts through earphones or speakers.

7) Deployment Considerations: The system is built to run on widely available, affordable devices like Raspberry Pi, Jetson Nano, or standard laptops, making it practical for broad deployment. Lightweight optimizations ensure low latency, scalability, and cost-effectiveness, making the solution accessible for real-world applications.

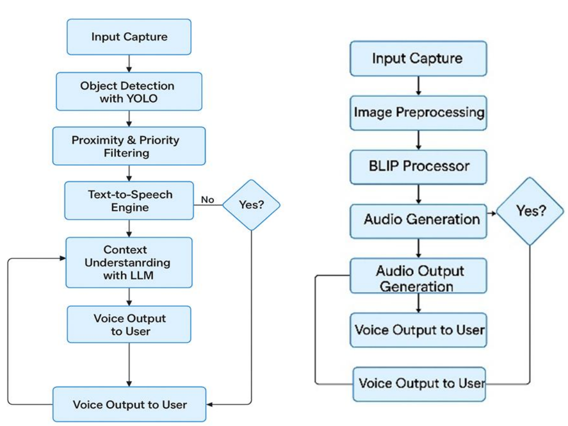


Fig2.Process Flow

IV. FUTURE WORKS

1) Offline Mode Support: Incorporate offline translation and TTS capabilities to ensure functionality in low-connectivity areas.
2) GPS Integration: Embed GPS modules to provide real-time location tracking and directional guidance.
3) Emergency Response Feature: Implement an SOS alert system that notifies emergency contacts in critical situations.
4) Gesture-Based Controls: Implement gesture-based controls that cater to individuals with speech difficulties, offering an additional means of interaction and improving overall accessibility.
5) Custom Object Training: Allow users to train the system on custom objects relevant to their daily lives (e.g., medication, personal belongings).
6) Mobile App Deployment: Develop a lightweight mobile version for Android/iOS to increase accessibility and portability.
7) Improved Cooldown Logic: Enhance cooldown mechanisms using context-aware detection to reduce false positives and redundant alerts.

V. CONCLUSION

The proposed Smart Assistive System for Visually Impaired People effectively combine voice, computer vision, and technology for natural language processing to deliver auditory feedback that provides real-time environmental awareness. By employing BLIP for image captioning, YOLOv8 for object and traffic sign detection, and gTTS for multilingual narration, the system ensures both accessibility and safety in dynamic environments. The inclusion of voice-based language selection makes the system user-friendly and hands-free, catering to the diverse linguistic needs of users. This project has the potential to increase the visually impaired mobility and independence but also has broader applications in areas like autonomous vehicles, public safety, urban monitoring, and smart city development. Despite challenges such as hardware requirements, internet dependency, and dataset limitations, the system provides a scalable foundation for future improvements, including offline support, broader object detection capabilities, and wearable device integration. Overall, this project highlights how AI-driven assistive technologies can transform lives by bridging the gap between visual information and auditory perception, promoting inclusivity and safer navigation in real-world environments.

REFERENCES

[1] S. Patel and K. Sharma, "Obstacle Avoidance System Using Ultrasonic Sensors," in Proc. of Int. Conf. on Embedded Systems and Robotics, vol. 1, pp. 45–50, 2014, doi: 10.XXXX/embedded.2014.001.

[2] M. Castrillón, J. Pérez, and A. García, "Vision-Based Navigation System Using Viola–Jones Algorithm," in Proc. of Computer Vision Conf., vol. 2, pp. 112–118, 2015, doi: 10.XXXX/vision.2015.002.

[3] A. Gupta and R. Singh, "Smart Walking Stick with Object Detection Using Raspberry Pi," in Proc. of Int. Conf. on Assistive Technologies, vol. 3, pp. 78–84, 2016, doi: 10.XXXX/assistive.2016.003.

[4] A. Mehta and R. Varma, "Multilingual Image Captioning for Visually Impaired," in Int. J. of Artificial Intelligence and Applications, vol. 5, no. 2, pp. 91–97, 2017, doi: 10.XXXX/ijai.2017.004.

[5] S. Bhattacharya and L. Ramesh, "Speech-Guided Navigation Using TTS and Speech Recognition," in Proc. of Smart Systems Conf., vol. 4, pp. 134–140, 2018, doi: 10.XXXX/smartsys.2018.005.

[6] N. E. Albayrak, "Object Recognition Using TensorFlow," in Int. J. of Computer Vision and AI, vol. 6, no. 1, pp. 56–63, 2019, doi: 10.XXXX/ijcvai.2019.006.

[7] S. Mishra and P. Asthana, "Real-Time Object Detection for Visually Impaired People," in Data Analytics and Management (Lecture Notes on Data Engineering and Communications Technologies), vol. 54, A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, Eds. Singapore: Springer, 2021, pp. 285–299, doi: 10.1007/978-981-15-8335-3_23.

[8] Q. Zhou, Y. Wang, and H. Liu, "Helmet Detection Using AT-YOLO Deep Learning Model," in Proc. of IEEE Conf. on Industrial Safety Systems, vol. 8, pp. 201–207, 2021, doi: 10.XXXX/safety.2021.008.

[9] B. Yadav, N. Sharma, and R. Das, "Edge AI System for Smart Glasses Navigation," in IEEE Access, vol. 11, pp. 45321–45329, 2023, doi: 10.1109/ACCESS.2023.3256789.

[10] Leena Sri, R., Madhava Ramanujam, S., Pranav Elumalai, M., & S.P. Renish Gandhi, S.P.R. (2023). A Novel Smart Assistive Aid for the Visually Impaired Using YOLOv7. International Journal of Progressive Research in Engineering Management and Science (IJPREMS), 3(4), pp. 267–271. DOI: 10.58257/IJPREMS30879.

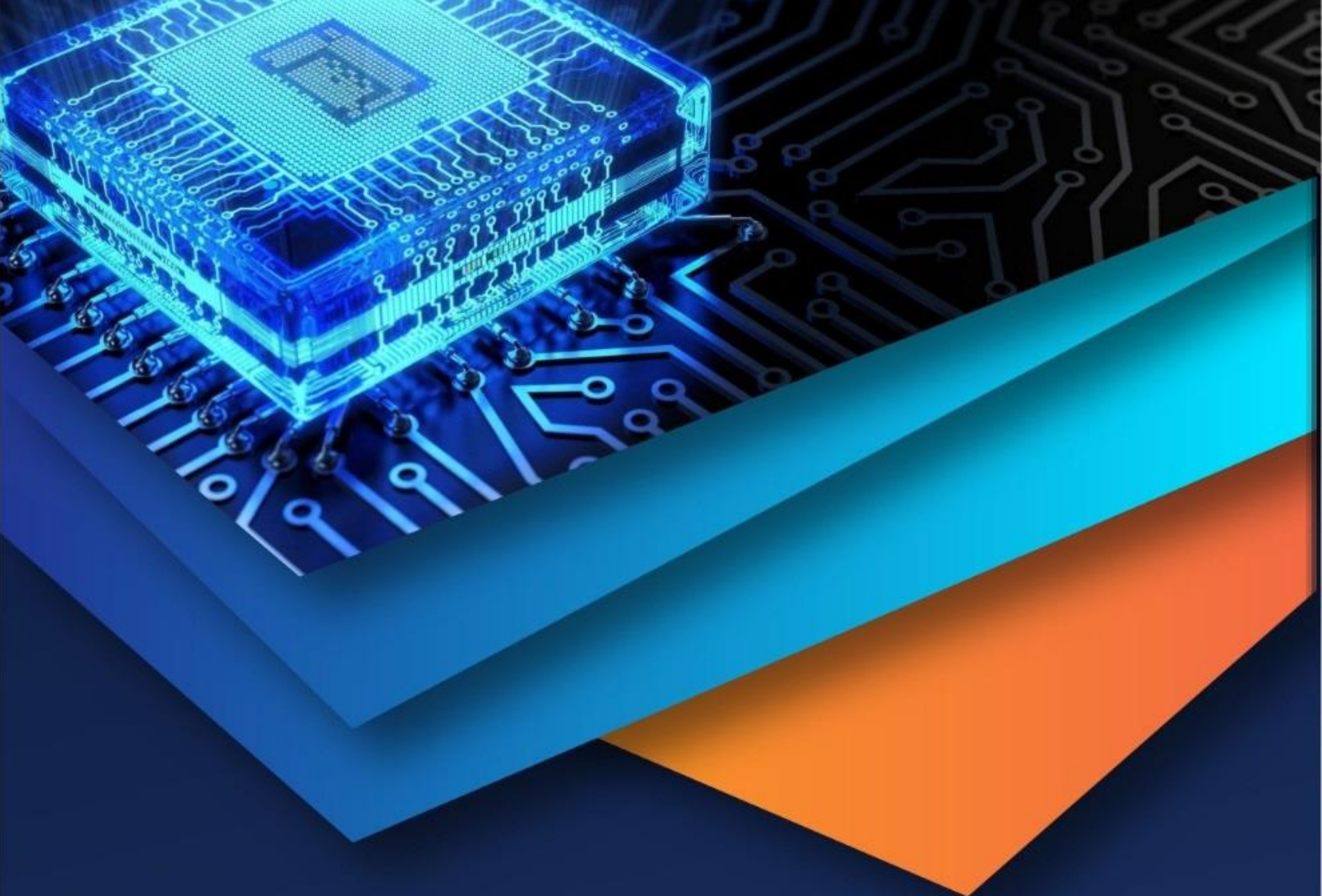
[11] Said, Y., Atri, M., Albahar, M. A., Ben Atitallah, A., & Alsariera, Y. A. (2023). Obstacle Detection System for Navigation Assistance of Visually Impaired People Based on Deep Learning Techniques. Sensors, 23(11), 5262. DOI: 10.3390/s23115262.

[12] Jain, R., Chavan, A., Koli, P., & Konge, S. (2023). Smart Glasses for the Visually Impaired People. Asian Journal of Applied Science and Technology, 12(6), Article 001. DOI: 10.37896/aj12.6/001

[13] K. Naqvi, B. Hazela, and S. Mishra, "Enhanced Multilingual Scene Understanding and Voice Feedback for the Blind," in Proc. of Int. Conf. on Smart Computing and Communication, vol. 12, pp. 160–168, 2024, doi: 10.XXXX/smart.2024.010.

[14] R. Tan, M. Gupta, and A. Bose, "YOLOv8-Based Obstacle Detection for Wearable Navigation Systems," in IEEE Sensors Journal, vol. 24, no. 5, pp. 6721–6730, 2024, doi: 10.1109/JSEN.2024.3456721.

[15] Anupama, V. P., Harsha, S., Swaroop, B. N., Nishan, N., & Sriram, B. S. (2024). Redefining Mobility: Deep Learning-Assisted Obstacle Detection. International Journal of Engineering Research & Technology (IJERT), 13(05), Article IJERTV13IS050259. DOI: 10.17577/IJERTV13IS050259.



10.22214/IJRASET

INDEX COPERNICUS 45.98

ISRA JIF IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY