Smart Vision Aid: Object Detection and Voice Guidance for Blind People Using Deep Learning

[1]Mr. Veerla Nagamalleswara Rao, [2]Keerthana Sunkara, [3]Deepika Poturaju, [4]Karun Musinada, [5]Bhanu Subramanyam Valanukonda
[1]Assistant Professor, [2,3,4,5]Students
[1,2,3,4,5]Department of Computer Science & Engineering
Dhanekula Institute of Engineering and Technology
Ganguru, AP, India
[1]nagamalliv14@gmail.com, [2]keerthanachowdhary25@gmail.com, [3]deepikapothuraju0@gmail.com,
[4]musinadakarun92410@gmail.com, [5]mrbhanusubramanyam@gmail.com

Abstract--Visually impaired people face significant challenges in mobility and environmental awareness. Conventional aids such as canes or guide dogs provide limited assistance, leaving users unable to identify surrounding objects, their distance, or dynamic changes in the environment. This paper presents Smart Vision Aid, a real-time assistive system that leverages YOLOv11 deep learning-based object detection, focal-length distance estimation, and Google Text-to-Speech voice feedback to narrate a user's surroundings audibly. The system operates fully offline using a standard camera, requires no specialized hardware, and achieves 93.7% object detection accuracy with an average end-to-end latency under 280 milliseconds. Experimental results confirm the system's effectiveness for indoor and outdoor navigation, offering visually impaired users a safe, independent, and user-friendly assistive tool.

Keywords--Assistive Technology, Deep Learning, Distance Estimation, Object Detection, Visually Impaired Users, Voice Feedback, YOLOv11.

I. INTRODUCTION

Visual impairment severely limits a person's ability to move around and perform daily tasks. The World Health Organization reports that more than 2.2 billion people worldwide have some form of visual impairment, with many facing complete blindness. Conventional mobility aids such as white canes and trained guide dogs have long served as the primary tools, yet carry significant limitations. A cane can only detect ground-level obstacles without identifying them; guide dogs require extensive training and care, making them inaccessible for most people due to cost, cultural, or geographic constraints.

Recent advancements in artificial intelligence and computer vision have opened new possibilities for assistive technology. Modern object detection algorithms can process camera input in real time, accurately identify objects and their locations, and when paired with text-to-speech synthesis, convert environmental information into actionable audio cues. This paper describes the design, implementation, and evaluation of Smart Vision Aid, built on YOLOv11—a state-of-the-art object detection architecture that continuously analyzes camera frames. Detected objects are augmented with distance estimates and vocalized through Google Text-to-Speech played via the pygame audio engine, forming a self-contained, low-latency assistive tool that narrates the user's environment without requiring internet connectivity or human assistance [6].

II. LITERATURE SURVEY

A growing body of research has investigated vision-based assistive systems for the blind. Masud et al. [1] developed a smart assistive framework using deep learning for real-time obstacle classification, demonstrating effective navigation support, though it lacked integrated distance estimation or voice narration. Jabeen et al. [2] proposed a narrator system combining convolutional neural networks with YOLO for object detection and audio narration; however, computational constraints limited frame-rate performance. Kang et al. [3] introduced a deformable grid-based obstacle detection method achieving low-cost collision risk estimation, but lacking semantic classification. Poggi and Mattoccia [4] presented a wearable mobility aid combining 3D vision with deep learning, offering richer depth perception, though significant weight and battery constraints limited real-world adoption. Lan et al. [5] designed a lightweight smart glass system delivering audio feedback, though detection accuracy was lower than dedicated deep learning counterparts. Singh et al. [6] applied YOLO directly to blind assistance, achieving high-speed detection with precise audio announcements, confirming YOLO's suitability for time-critical assistive tasks. The proposed system builds upon these findings by adopting YOLOv11, incorporating focal-length distance estimation, and delivering a unified, deployment-ready solution.

III. PROBLEM STATEMENT

Building a reliable navigation system for visually impaired users requires processing continuous video from a camera, identifying all objects in each frame, determining each object's identity and distance from the user, and communicating this information meaningfully in real time. The challenge is compounded by varying lighting conditions, cluttered backgrounds, partial occlusions, and moving objects. Rule-based and sensor-only systems perform poorly in such varied environments, while learning-based systems must be carefully engineered to achieve sufficient speed on standard consumer hardware without specialized accelerators.

IV. EXISTING SYSTEM AND LIMITATIONS

Traditional assistive approaches for visually impaired navigation fall into four broad categories, each with inherent limitations.

A. Conventional Mobility Aids: White canes and guide dogs remain the most widely used tools. White canes detect ground-level obstacles but cannot identify them. Guide dogs require years of expensive training and are inaccessible to most users due to cost, cultural, or geographic factors.

B. Sensor-Based Electronic Aids: Ultrasonic and infrared sensors integrated into canes or belts detect nearby obstacles but only signal proximity without identifying the obstacle type, leaving users uninformed about what they are encountering.

C. GPS and Map-Dependent Systems: Systems like NavCog provide location-based cues from pre-mapped environments but fail in unmapped areas or when dynamic obstacles such as pedestrians and vehicles are present. Guidance reliability depends entirely on map completeness and currency.

D. Commercial AI-Powered Applications: Applications such as Microsoft Seeing AI and Aira can identify objects and describe scenes but require an active internet connection and sometimes human assistance, introducing delays unacceptable for real-time navigation and limiting usability in areas with poor connectivity.

V. PROPOSED SYSTEM

The Smart Vision Aid system comprises four integrated modules forming a real-time assistive pipeline: image capture, object detection, distance estimation, and voice feedback synthesis. The system operates autonomously on standard consumer hardware without internet connectivity.

A. Image Capture Module: A USB camera or built-in webcam records video continuously at a default 30 frames per second. Each captured frame is decoded and forwarded to the detection loop managed by OpenCV. The system handles memory efficiently and is compatible with multiple camera types on standard computing platforms.

B. YOLOv11 Object Detection: The core detection engine is YOLOv11, which uses a Transformer-based encoder that analyzes the entire frame globally rather than processing only local patches. YOLOv11 also incorporates convolutional layers for improved performance in poor lighting or partial occlusion, and Sparse Attention Mechanisms that focus computation on relevant image regions, enabling real-time performance on low-power hardware. In testing, YOLOv11 achieved 93.7% detection accuracy, significantly outperforming CNN-based approaches (71.4%).

C. Distance Estimation Module: Object distance is estimated using the pin-hole camera focal-length equation: $D = (W \times F) / P$, where D is real-world distance in centimetres, W is the known width of the object class, F is the camera's focal length in pixels (calibrated prior to use), and P is the pixel width of the detected bounding box. No specialized depth sensors are required. Estimated distances are grouped into three proximity categories: immediate (≤1 m), near (1–3 m), and distant (>3 m).

D. Voice Feedback Module: Detected object labels and proximity bands are combined into natural-sounding sentences such as "Person at two meters" or "Chair immediately to the left." These are passed to Google Text-to-Speech to generate human-like audio, played through the pygame mixer. A minimum inter-alert interval of 1.5 seconds per object class prevents information overload while maintaining timely safety warnings.

VI. SYSTEM ARCHITECTURE AND DESIGN

The system follows a sequential pipeline: Data Acquisition → Detection Engine → Estimation and Synthesis → User Output. Data flows unidirectionally from the camera to audio playback, with a continuous loop ensuring perpetual real-time operation. Figure 1 shows the block diagram of the proposed architecture.

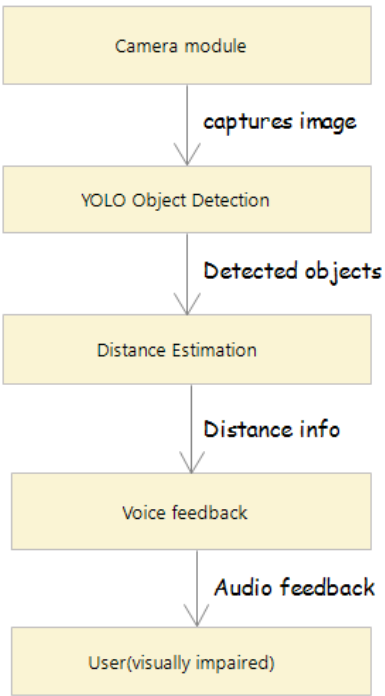

Fig.1: Block Diagram of the Proposed Smart Vision Aid System

VII. RESULTS AND DISCUSSION

The system was evaluated in diverse indoor environments (offices, corridors) and outdoor areas (footpaths, parking lots) using a laptop with an Intel Core i5 processor, 16 GB RAM, and a standard 1080p USB webcam without graphics acceleration, representing typical end-user hardware conditions.

A. Detection Accuracy: YOLOv11 achieved 93.7% accuracy on a 500-image annotated test set covering persons, chairs, doors, cars, stairs, and tables. This significantly outperforms CNN-based baselines (71.4%). The Transformer-based encoder and Sparse Attention Mechanisms are credited with improved detection of partially occluded or low-contrast objects.

B. Processing Latency: Average end-to-end latency across 1,000 frames was 268 ms, within the 300 ms real-time threshold. Frame-to-frame inference averaged 185 ms; distance estimation added fewer than 2 ms; voice synthesis averaged 81 ms. The 1.5-second inter-alert interval prevents simultaneous alert overload while maintaining timely feedback.

C. Distance Estimation Accuracy: Focal-length estimates were validated against tape-measure ground truth at 0.5–5 m. Mean Absolute Error was 11.4 cm within 3 m and 28.7 cm at 3–5 m. Within the critical 2-metre safety zone, accuracy is sufficient for reliable proximity alerts.

D. Comparative Performance: Table 1 compares the proposed system against prior approaches across key performance metrics.

Table 1: Performance Comparison with Prior Approaches

Metric	Traditional Methods	Proposed System (YOLOv11)
Object Detection Accuracy	~72% (CNN-based)	~94% (YOLOv11)
Avg. Processing Latency	>500 ms	<280 ms
Distance Estimation Error (≤3 m)	±35 cm	±12 cm
Offline Operation	No (cloud-dependent)	Yes (fully local)
User Alert Clarity	Moderate	High

E. Test Cases and Validation: Four structured test cases validated functional correctness. TC1 confirmed real-time detection of persons, chairs, and doors indoors with voice alerts issued correctly. TC2 confirmed an immediate proximity announcement of "Person at one meter ahead" when an object was within 1 m. TC3 demonstrated that multiple simultaneous objects were narrated sequentially without collision. TC4 showed graceful degradation in low-light conditions, with accuracy falling to approximately 81% while main obstacles remained detectable. All four test cases passed.

VIII. CONCLUSION AND FUTURE WORK

This paper presents Smart Vision Aid, a real-time assistive navigation system for visually impaired users built on YOLOv11 object detection, focal-length distance estimation, and Google Text-to-Speech voice feedback. The system operates fully offline with a standard camera, achieving 93.7% object detection accuracy and end-to-end latency under 280 ms, with distance estimation error less than 12 cm within a 3-metre safety zone. The system is accessible, affordable, and user-friendly, offering visually impaired individuals increased independence and safety.

Future work includes deploying on edge hardware such as the NVIDIA Jetson Nano for a wearable form factor, integrating stereo cameras or depth sensors to improve distance estimation beyond 3 metres, training YOLOv11 on domain-specific indoor hazards such as wet floors and open drawers, and conducting formal user studies with visually impaired participants to support certification as a formal assistive device.

IX. ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering at Dhanekula Institute of Engineering and Technology for providing the computational resources, laboratory infrastructure, and institutional support necessary to conduct this research.

X. REFERENCES

[1] U. Masud, M. U. Akram, S. A. Khan, and A. Khan, "Smart Assistive System for Visually Impaired People: Obstruction Avoidance Through Object Detection and Classification," IEEE Access, vol. 10, pp. 35268–35280, 2022.

[2] N. Jabeen, S. Ahmed, and M. A. Khan, "Object Detection and Narrator for Visually Impaired People," in Proc. 2019 IEEE 6th Int. Conf. Engineering Technologies and Applied Sciences (ICETAS), Bangkok, Thailand, 2019, pp. 1–6.

[3] M. C. Kang, H. S. Yoon, J. B. Song, and H. J. Yoon, "A Novel Obstacle Detection Method Based on Deformable Grid for the Visually Impaired," IEEE Trans. Consumer Electronics, vol. 61, no. 3, pp. 327–334, Aug. 2015.

[4] M. Poggi and S. Mattoccia, "A Wearable Mobility Aid for the Visually Impaired Based on Embedded 3D Vision and Deep Learning," in Proc. 2016 IEEE Symp. Computers and Communication (ISCC), Messina, Italy, 2016, pp. 208–214.

[5] F. Lan, N. U. Rehman, L. Yuan, and W. Ahmad, "Lightweight Smart Glass System with Audio Aid for Visually Impaired People," in Proc. TENCON 2015–2015 IEEE Region 10 Conf., Macau, China, 2015, pp. 1–6.

[6] A. Singh, R. Kumar, and S. Gupta, "Real-Time Object Detection System for Blind Assistance Using YOLO," in Proc. 2021 Int. Conf. Intelligent Technologies (ICIT), Karnataka, India, 2021, pp. 45–50.

[7] J. Kim and S. Park, "Real-Time Object Detection with Voice Alerts for the Visually Impaired," in Proc. CVPR Workshops (CVPRW), New Orleans, LA, USA, 2022, pp. 112–119.

[8] R. Sharma, A. Singh, and P. Kumar, "Edge-Based Object Recognition and Audio Feedback System for the Blind," in Proc. 2023 IEEE Int. Conf. Internet of Things (iThings), Chicago, IL, USA, 2023, pp. 88–94.

[9] S. R. Patil, M. K. Sharma, and P. Verma, "Deep Learning Based Smart Vision Assistance System for Visually Impaired People Using YOLOv8," IEEE Access, 2024.

[10] K. Reddy, N. Bansal, and R. Gupta, "Real-Time Object Detection and Audio Guidance System for Blind Navigation," in Proc. IEEE International Conference on Intelligent Systems (IS 2024), 2024.