2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-09

A SURVEY ON SIGHTSENSE: AI POWERED REAL TIME VISUAL INTERPRETATION SYSTEM

***Dr. Abhishek Kulkarni, Shivathej S. Gowda, Sreepriya B. M., Sudhiksha Prabhakar, Sunaina Ravi**

J. P Nagar, Bengaluru – 560078.

Article Received: 01 April 2026
Article Revised: 21 April 2026
Published on: 11 May 2026

***Corresponding Author: Dr. Abhishek Kulkarni**
JP Nagar, Bengaluru – 560078.

DOI: https://doi-doi.org/101555/ijrpa.2664

ABSTRACT

Visual impairment significantly affects an individual's ability to navigate and interact with their surroundings, creating challenges in performing everyday activities independently. Recent advancements in **Artificial Intelligence (AI)** and **Computer Vision** have enabled the development of intelligent assistive systems that enhance environmental awareness for visually impaired individuals. This survey paper presents a comprehensive review of **AI-powered assistive systems** that utilize real-time object detection and speech synthesis to provide auditory feedback. The paper explores various techniques and technologies, including deep learning models such as **YOLO (You Only Look Once)** for object detection, along with tools like **OpenCV**, **PyTorch**, and **Text-to-Speech (TTS)** systems. It analyzes different system architectures, implementation approaches, and hardware platforms such as **Raspberry Pi** used for building portable solutions. Additionally, the survey highlights key challenges, including accuracy in complex environments, real-time processing constraints, power efficiency, and user adaptability. Furthermore, this paper identifies existing research gaps and discusses future directions, such as improving model efficiency, integrating context-aware intelligence, and enhancing user experience through adaptive feedback systems. The objective of this survey is to provide a clear understanding of current advancements and to guide the development of efficient, affordable, and user-friendly assistive technologies for visually impaired individuals.

KEYWORDS: Artificial Intelligence (AI), Assistive Technology, language Visual Impairment, Object Detection, YOLO, Computer Vision, Real-Time Systems, Text- to-Speech (TTS), Deep Learning, OpenCV, PyTorch, Raspberry Pi, Wearable Devices, Human-

Computer Interaction, Smart Navigation.

1. INTRODUCTION

Visual impairment is a significant global challenge that affects millions of people, limiting their ability to navigate, interact, and live independently. Everyday activities such as walking, identifying objects, and understanding surroundings become difficult without proper assistance. Traditional aids like white canes and guide dogs provide basic support, but they often fail to deliver detailed information about the environment. With the rapid advancement of **Artificial Intelligence (AI)**, **Computer Vision**, and **Deep Learning**, new possibilities have emerged to enhance assistive technologies. AI-powered systems can analyze visual data from cameras in real time, detect objects, and convert this information into meaningful audio feedback. This allows visually impaired individuals to better understand their surroundings and make safer decisions. Recent developments in object detection models, especially **YOLO (You Only Look Once)**, have enabled fast and accurate real- time detection, making them suitable for assistive applications. Combined with technologies such as **OpenCV**, **PyTorch**, and **Text-to-Speech (TTS)** systems, these solutions can provide continuous guidance about nearby objects, obstacles, and environmental context. This survey paper focuses on **AI-powered assistive systems for visually impaired people**, particularly those that use real-time object detection and speech synthesis. It reviews existing techniques, technologies, and system architectures, highlights their advantages and limitations, and identifies research gaps. The goal is to provide a comprehensive understanding of current advancements and explore future directions for building efficient, affordable, and user-friendly assistive solutions.

1.1 Background

Visual impairment affects millions of individuals worldwide, limiting their ability to perceive and interpret their surroundings. For many visually impaired people, daily activities such as walking, recognizing objects, reading signs, and avoiding obstacles become challenging tasks. Traditionally, assistive tools like white canes and guide dogs have been widely used to support mobility and navigation. While these tools are reliable and simple, they provide only limited information about the environment and cannot fully describe complex surroundings. With the advancement of **Artificial Intelligence (AI)** and **Computer Vision**, assistive technologies have evolved significantly. Modern systems use cameras and sensors to capture visual data and process it using deep learning algorithms. Among these, object detection models such as **YOLO (You Only Look Once)** have gained popularity due to their ability to

perform fast and accurate detection in real time. These models can identify multiple objects in a scene and provide immediate feedback, making them suitable for assistive applications.
In addition, the integration of **Text-to-Speech (TTS)** technology enables these systems to convert detected information into audio output, allowing users to receive real-time guidance. Platforms like **OpenCV** and **PyTorch** have further simplified the development of such systems by providing powerful tools for image processing and model implementation. Moreover, the use of compact hardware devices such as **Raspberry Pi** has made it possible to create portable and affordable assistive solutions. Despite these advancements, existing systems still face challenges such as processing delays, limited accuracy in complex environments, and lack of personalization. Therefore, there is a growing need to develop efficient, reliable, and user- friendly AI-powered assistive systems that can provide real-time support and improve the independence of visually impaired individuals.

1.2 Motivation

Visual impairment continues to pose serious challenges to independent living, especially in tasks such as navigation, object recognition, and environment awareness. Although traditional assistive tools like white canes provide basic support, they lack the ability to deliver detailed, real- time information about surroundings. This limitation creates a strong need for intelligent systems that can enhance safety, confidence, and independence for visually impaired individuals. Recent advancements in **Artificial Intelligence (AI)**, **Deep Learning**, and **Computer Vision** have opened new opportunities to address these challenges. Modern object detection models, particularly **YOLO (You Only Look Once)**, enable fast and accurate identification of multiple objects in real time. When combined with **Text-to-Speech (TTS)** technology, these systems can convert visual information into audio feedback, allowing users to understand their environment without relying on vision. Another key motivation is the increasing availability of low-cost hardware platforms such as **Raspberry Pi**, which makes it possible to develop portable and affordable assistive devices. This ensures that such technologies are not limited to high-end users but can be accessible to a wider population, especially in developing regions. Despite these technological advancements, many existing solutions are either too expensive, lack real time performance, or do not provide user-friendly interaction. Additionally, challenges such as accuracy in dynamic environments, power consumption, and system responsiveness still need to be addressed. Therefore, the motivation of this survey is to explore and analyze current AI-powered assistive systems, identify their strengths and limitations, and highlight opportunities for

developing more efficient, adaptive, and user-centric solutions. The ultimate goal is to contribute toward building smarter assistive technologies that significantly improve the quality of life for visually impaired individuals.

1.3 Contribution of this Survey

This survey paper provides a comprehensive overview of existing research and technologies related to **AI-powered assistive systems for visually impaired individuals**, with a focus on real- time object detection and audio feedback mechanisms. The key contributions of this survey are summarized as follows:

Review of Object Detection Techniques:

This survey examines state-of-the-art deep learning models used in assistive systems, including **YOLO (You Only Look Once)** and other convolutional neural network (CNN)-based approaches. It evaluates their performance in terms of speed, accuracy, and suitability for real-time applications.

Analysis of System Architectures:

The paper explores different system designs that integrate components such as cameras, embedded devices (e.g., **Raspberry Pi**), image processing frameworks like **OpenCV**, and speech synthesis technologies. It highlights how these components work together to create practical assistive solutions.

Evaluation of Assistive Approaches:

Various assistive methods, including wearable devices, mobile-based applications, and real-time navigation systems, are analyzed. The survey compares their advantages, limitations, and usability in real-world scenarios.

Identification of Research Gaps:

The survey identifies key challenges in current systems, such as limited accuracy in complex environments, high computational requirements, lack of personalization, and latency issues. It emphasizes the need for more adaptive and efficient models.

Future Research Directions: Based on the analysis, the paper suggests future improvements, including context-aware assistance, energy-efficient models, enhanced user interaction, and integration with IoT and smart devices for better performance and scalability.

2. LITERATURE REVIEW

Recent research in **AI-powered assistive systems for visually impaired individuals** has focused on integrating computer vision, deep learning, and multimodal feedback to enhance environmental awareness and navigation. Various approaches have been proposed, each addressing specific aspects of assistance such as object detection, scene understanding, navigation, and user interaction. One of the most widely used techniques is **real-time object detection using YOLO models**. Systems based on YOLOv3, YOLOv5, and YOLOv8 combined with platforms like Raspberry Pi and OpenCV provide fast and efficient detection of objects, making them suitable for real-time applications. These systems offer immediate audio feedback through Text-to-Speech (TTS), improving user independence. However, they often struggle in complex environments with cluttered scenes and lack depth perception for accurate distance estimation. To enhance scene understanding, **vision- language models such as BLIP and LLaVA** have been introduced. These models generate rich semantic descriptions and can answer contextual questions about the environment. While they significantly improve user awareness, their high computational requirements make them unsuitable for deployment on low-power edge devices, limiting their real-time usability. Lightweight models such as **MobileNet combined with SSD** have been proposed for edge computing platforms like Jetson Nano. These approaches offer faster inference and reduced computational cost, but they compromise on detection accuracy and often lack integration with speech systems, which is essential for assistive applications. In addition to object detection, **depth sensing technologies such as Intel RealSense cameras** have been explored to improve navigation by providing distance estimation. These systems help users avoid obstacles more effectively, but they increase system cost and power consumption, making them less accessible for widespread use.

Reference	Algorithms /Approach	Advantages	Research Gap
[1]	YOLOv3 + Raspberry Pi + pyttsx3 TTS	Real-time object detection with instant audio feedback; low cost and portable	Limited accuracy in cluttered scenes; no scene captioning or depth estimation
[2]	BLIP + Vision-Language Model for scene captioning	Rich semantic descriptions of scenes; handles complex environments	High computational cost; not suitable for edge deployment without optimization
[3]	MobileNetV2 + SSD on Jetson Nano	Lightweight model suitable for edge devices; good detection speed	Lower accuracy than YOLO; limited object classes; no TTS integration

[4]	Depth sensing (Intel RealSense) + CNN for obstacle detection	Provides distance estimation; effective for navigation and collision avoidance	Expensive hardware; high power consumption; limited object identification
[5]	YOLOv5 + OpenCV + Google TTS	Improved detection accuracy; natural-sounding voice output; faster inference	Requires internet for Google TTS; not fully offline; limited personalization
[6]	Faster R-CNN + haptic feedback wearable	Combines visual and tactile feedback; effective for indoor navigation	Slow inference speed; complex hardware setup ;not lightweight
[7]	OCR (Tesseract) + TTS for text reading	Enables reading of printed text, signs, and labels in real time	Limited to text; no object or obstacle detection; struggles with handwriting
[8]	YOLOv8 + PyTorch + quantization for Raspberry Pi 4	State-of-the-art accuracy with optimized edge deployment; offline operation	Quantization reduces accuracy slightly; limited battery life on Pi
[9]	L La VA (Large Language and Vision Assistant) for contextual scene Q&A;	Allows user to ask questions about the scene; highly informative responses	Very high compute requirements; not feasible on Raspberry Pi without off loading
[10]	OpenCV + face recognition + object detection for smart glasses	Identifies known persons and objects; social interaction support	Privacy concerns; requires large labeled dataset of known individuals
[11]	Federated learning for privacy preserving model training across devices	Protects user data; enables personalization without central data collection	Complex setup; communication overhead; requires many participating devices
[12]	CNN + GPS + indoor mapping for navigation assistance	Provides route guidance alongside object detection; useful for outdoor navigation	GPS unreliable indoors; high power consumption; complex integration
[13]	Transformer-based image captioning (CLIP + GPT) for rich scene descriptions	Highly descriptive natural language output; adaptable to diverse scenes	Cannot run on low- power edge devices; latency too high for real time use

3. CONCLUSION AND FUTURE WORK

The rapid advancement of **Artificial Intelligence (AI)** and **Computer Vision** has significantly improved assistive technologies for visually impaired individuals. This survey reviewed various AI-powered systems that utilize real-time object detection, speech synthesis, and embedded platforms to enhance environmental awareness and navigation. Techniques such as **YOLO-based object detection**, **CNN models**, and **vision- language models** have demonstrated promising results in identifying objects and providing meaningful feedback to users. From the literature, it is evident that existing systems offer valuable features such as real-time detection, portability, and multimodal feedback. However, most solutions are limited in one or more aspects, including high computational requirements, reduced accuracy in complex environments, lack of personalization, and dependency on internet connectivity. Additionally, many systems focus on specific tasks like text reading or navigation, rather than providing a comprehensive assistive solution. The analysis highlights that there is still a gap in developing an integrated system that combines **high accuracy, low latency, affordability, and user-friendly interaction**. Systems deployed on edge devices such as **Raspberry Pi** show potential but require further optimization to balance performance and efficiency.

Future research can focus on the following directions:

Improved Model Efficiency

Develop lightweight and optimized deep learning models that can run efficiently on low-power edge devices without compromising accuracy.

Context-Aware Assistance:

Integrate contextual understanding (such as location, user activity, and surroundings) to provide more meaningful and adaptive feedback.

Multimodal Interaction:

Combine audio, haptic, and visual feedback to enhance user experience and reliability, especially in noisy or complex environments.

Offline and Privacy-Preserving Systems:

Design systems that function without internet dependency and ensure user data privacy through techniques like federated learning.

Enhanced Navigation Support:

Integrate indoor and outdoor navigation systems with obstacle detection for complete mobility assistance.

User-Centric Design:

Focus on personalization and ease of use to ensure acceptance and long-term usability among visually impaired individuals.

In conclusion, AI-powered assistive systems hold great potential to transform the lives of visually impaired people by enabling safer navigation and greater independence. Continued research and development in this area can lead to more efficient, affordable, and accessible solutions that address real world challenges effectively.

4. REFERENCES

1. J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.
2. J. Li *et al.*, "BLIP: Bootstrapping Language- Image Pre-training for Unified Vision-Language Understanding and Generation," in *Proceedings of ICML*, 2022.
3. A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.
4. Intel Corporation, "Intel RealSense Depth Camera D435: Technical Specifications and Applications," Intel White Paper, 2019.
5. G. Jocher *et al.*, "YOLOv5 by Ultralytics," GitHub Repository, 2022. [Online]. Available: https://github.com/ultralytics/yolov5
6. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *NeurIPS*, 2015.
7. R. Smith, "An Overview of the Tesseract OCR Engine," in *ICDAR*, 2007.
8. G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," GitHub Repository, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
9. H. Liu *et al.*, "LLaVA: Visual Instruction Tuning," in *NeurIPS*, 2023.
10. L. Ran *et al.*, "RS-GAN: Exploring Reciprocal Spatial Relations for Indoor Navigation," in *CVPR*, 2020.
11. A. Radford *et al.*, "Learning Transferable Visual Models From Natural Language Supervision," in *ICML*, 2021.
12. N. Kosaka and S. Omachi, "Currency Recognition Using CNN," *IEICE Transactions on*

Information and Systems, 2010.

13. N. Chaudhary *et al.*, "Clothing Assistance for the Visually Impaired Using Color Detection and CNN," *International Journal of Computer Applications*, 2021.
14. A. M. Brock *et al.*, "Multimodal Assistive Technology for Blind Users: Audio, Haptic, and Visual Feedback Integration," in *ACM ASSETS*, 2015.
15. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *CVPR*, 2015.
16. B. McMahan *et al.*, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *AISTATS*, 2017.