WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

An IoT-Based Smart Spectacle for Object Detection and Audio Guidance System for the Visually Impaired

Chandana R[1], Shreeya [2], U Supriya[3], Vamshi K T[4], Prof. Anusha U A[5]

[1]Department of Information Science and Engineering, Global Academy of Technology Bengaluru, India chandanar1306@gmail.com
[2]Department of Information Science and Engineering, Global Academy of Technology Bengaluru, India shreeyaacharya07@gmail.com
[3]Department of Information Science and Engineering, Global Academy of Technology Bengaluru, India supriyaub09@gmail.com
[4]Department of Information Science and Engineering, Global Academy of Technology Bengaluru, India vamshiktakur2005@gmail.com
[5]Assistant Professor, Department of Information Science and Engineering, Global Academy of Technology, Bengaluru, India anushaua@gat.ac.in

ABSTRACT—

Visual Impairment is a challenging task which hinders a person's ability to recognize things, faces, and environmental factors. In this paper, we present an AI-based assistive solution that will help in environment recognition for the blind using real-time object detection and audio. Our proposed framework will accept live video as input (captured using a camera), process it via deep learning, and detect objects using the YOLO (You Only Look Once) algorithm. Some additional modules in our framework will be face recognition using CNNs and currency detection using image processing.

Identification of the object will lead to provision of instant audio feedback through text to speech technology. The paper also gives an overview of the current research in the area of computer vision based assistive technologies along with an analysis of existing methods to identify their shortcomings and emphasizes on the need for development of an accurate, real-time and cost-effective assistive system.

Index Terms—Artificial Intelligence (AI), Internet of Things (IOT), Computer Vision, You Only Look Once (YOLO) Deep Learning, Real-Time Object Detection, Voice Assistance, Audio Guidance, Face Recognition, Optical Character Recognition (OCR), Currency Detection, Text-to-Speech (TTS), Visual Impairment Assistive Technology, Accessibility.

I. Introduction

Assistive technologies are becoming more and more relevant for enhancing the quality of life of disabled people. For visually-impaired people, comprehending and engaging with their environment is a difficult task. The present solutions such as white cane or guide dogs merely help navigate but fail to offer detailed information about objects, people, and the environment. As a result, blind and visually impaired people often find themselves unable to perform normal daily tasks independently and safely. Advances in Artificial Intelligence (AI), Machine Learning (ML) and Computer Vision have offered new possibilities for developing intelligent assistive technologies. In contrast to conventional tools, an AI-driven system is able to assess visual data gathered from cameras, recognize objects and offer useful information to the user immediately. With the help of these tools, visually-impaired individuals are able to enhance awareness of the current situation and respond accordingly.

Object recognition is one of the key areas in computer vision. For example, YOLO algorithm (You Only Look Once) is state-of-the-art, which means that it is capable of detecting various objects at once in a highly accurate and speedy way, making it fit for real-time applications. What is more, Convolutional Neural Networks (CNNs) have proved to be very efficient in face recognition.

Additionally, image processing methods are applied in the development of other assistive systems, like currency recognition for performing financial transactions independently. This paper introduces an assistive system that detects objects in real-time using artificial intelligence and provides assistive narration of the detected objects using text-to-speech technology. This system integrates three types of recognition – object detection, face recognition, and currency detection – in a single system to provide comprehensive assistance to the users. Additionally, recent studies on the recognition of virtual objects through 3D reconstructed shape have revealed that the structure of space can be perceived by the visually impaired using advanced techniques.

II. Literature Survey

Technological advancements in Artificial Intelligence (AI), Computer Vision, Deep Learning and Internet of Things(IoT) have significantly impacted the creation of assistive technologies for visually challenged people. Recent research has been concentrated on improving environmental awareness, object detection, navigation and interaction by implementing intelligent algorithms in wearable/portable devices. Models such as You Only Look Once

(YOLO), Convolutional Neural Networks (CNN),Vision Transformers (ViT) and OpenCV-based models have led to great improvements in real time object detection. Furthermore, advancements in speech recognition, Optical Character Recognition (OCR), Text-to-Speech(TTS), face recognition, spatial audio rendering and 3D reconstruction have contributed to the improvement in functionality of assistive technologies making it easier for users to interact with their environment independently.

This literature survey gives an overview of some key contributions of existing research and highlights the gaps in research that inspired the creation of the proposed system.

Object Detection Systems for Visually Impaired:

Object detection plays a critical role in today's assistive technologies for the visually impaired. Multiple research works have implemented the algorithms of deep learning including YOLOv4,YOLOv5, YOLOv7, and YOLOv8 for detecting and classifying multiple objects from real-time video recorded by cameras. The objects include common items like chairs, doors, bottles, vehicles, and pedestrians. The information detected through such a technology is then converted to speech by means of the technology of Text-to-Speech. The combination of fast object detection and audio guidance enables the users to identify the objects around them without vision. With recent improvements in YOLO architecture, there has been an improvement in object detection accuracy along with the decrease in the time of inference. But there is still an impact of poor lighting, object occlusion, motion blur, and crowded scenes on object detection accuracy.

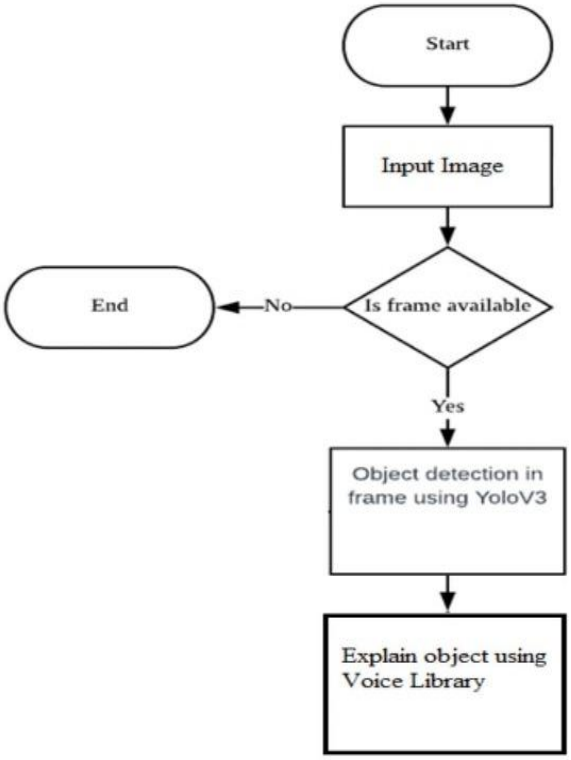


Fig 1:Flow diagram of YOLO

Intelligent Assistive Systems Using Computer Vision:

Computer vision has been considered a vital component of intelligent assistance systems since it helps machines perceive and interpret visual information just like humans do. In recent times, researches have focused on the combination of computer vision and Artificial Intelligence to enable object recognition, scene understanding, obstacle detection and environment analysis. The systems continuously acquire visual data, process images using deep learning algorithms, and provide audio-based contextual feedback to the user. People who are visually impaired can make use of cameras and processors built into wearable technologies like AI-powered smart glasses and vision systems to handle everyday tasks. These systems improve their awareness of the environment. But most of these existing technologies are basically image-based without any spatial or contextual description of the scene.

Voice Interaction and Audio Guidance Systems:

The development of voice interactions has become one of the most successful approaches to enabling blind users to communicate with assistive technologies. Current research involves the use of Automatic Speech Recognition (ASR), Natural Language Processing(NLP) and Text-to-Speech (TTS) technologies in order to design interactive assistance systems that work on voice commands. Users can give voice commands for object recognition, text

reading, people recognition, or navigation. The system will process the voice command, conduct the necessary visual analysis, and provide an audio response. Intelligent voice assistants can help to improve accessibility and reduce the necessity for tactile interaction with electronic gadgets. However, accuracy of speech recognition systems is influenced by ambient noise, microphone quality, language variations and delays in processing, which may lead to unreliability of the system outdoors.

Face Recognition and Human Identification Systems:

It has become indispensable in the intelligent assistive systems that help people with visual impairment to identify their families, friends and others surrounding them. In terms of existing research on this topic, researchers have been able to attain accurate results in the detection and recognition of human faces with the help of CNN, FaceNet, RetinaFace and OpenCV. The system makes comparisons with the facial features with the databases and provides identification in voice form to the users. Face recognition improves social interactions, personal security and self-confidence. Even if these intelligent systems achieve high accuracy in recognition of human faces under controlled conditions, their accuracy is hindered by varied facial expressions, lighting, occlusions and crowded scenes.

Optical Character Recognition and Currency Detection: Various assistive technologies make use of OCR (Optical Character Recognition) technology in order to help visually impaired people to read printed texts available in books, documents, product labels, medicinal packages, and signboards. OCR technology uses image processing to extract textual data and then using TTS (Text-to-Speech) technology converts this textual data into speech. Similarly, currencies can be recognized using various image processing and deep learning algorithms. This makes it possible for people to perform financial transactions independently. Although assistive, the performance of OCR is greatly affected by the quality of the image, text orientation, and the font of text and lighting conditions. The currency recognition process is hampered when currency is damaged, folded or only partially visible.

Navigation and Obstacle Detection Systems:

The primary objective of assistive technology for blind people is that of navigation assistance. There are emerging technologies which include object detection, ultrasonic sensing, GPS, spatial audio feedback, and IoT communication technology to help navigate through indoor and outdoor environments safely. The AI-based navigation system identifies obstacles and their distances and gives directions either through auditory or haptic feedback. Wearable devices are being embedded with smart glasses, IoT sensors, and embedded processor to constantly keep track of the surroundings of the user. These systems improve mobility and safety; however, navigation accuracy remains affected due to dynamic obstacles, environmental factors, weather conditions, and limitations of sensors. Indoor navigation further needs infrastructure such as maps.

3D Reconstruction and Spatial Understanding:

The recent developments in 3D Computer Vision have resulted in the emergence of new methods for constructing three-dimensional models from two-dimensional images. The research carried out in relation to Gaussian splatting, mesh generation and depth estimation revealed the fact that 3D reconstructions may facilitate visual understanding of object geometry and depth perception by people with visual impairment. The 3D reconstruction method gives a user a more accurate idea about the shape and orientation of an object in comparison with conventional object recognition systems. The use of spatial sound and haptic feedback adds more realism into the interaction process by helping people feel the position and structure of objects. However, these methods have a number of disadvantages such as high demand for computing power, special cameras and advanced graphics processing units.



Fig 2:Object Recogniti0n

Limitations of Existing Systems:

From the literature review, it is clear that a great deal of work has been done in the field of assistive technologies, however, there are still several open issues in the field of research. Most of the current assistive technologies are limited to individual tasks like object detection, navigation, OCR, face detection, etc. and don't give an overall solution for the problems. Deep learning models require a large amount of computational power as well as good hardware which makes it difficult for deployment on wearable devices which are less powerful. Poor lighting condition, object occlusion, environmental changes, and complex backgrounds affect the detection accuracy. Also, speech recognition errors due to noise in voice interaction systems and low performance of navigation systems in highly dynamic and unfamiliar environment affect the performance. Moreover, very little work has been done in the field of integration of object detection, face detection, OCR, currency detection, intelligent voice interaction, and IoT communication in one framework. Such limitations lead us towards the realization of an intelligent, compact, and integrated assistance system that would be capable of providing the blind with precise object detection, contextual awareness, voice interaction, and navigation assistance.

III.COMPARATIVE ANALYSIS

The following table, Table I, demonstrates the presence of object detection, navigation, voice assistance, OCR, face recognition, spatial audio, 3D reconstruction, IoT, and wearable systems in the studies. Object detection methods using YOLO can conduct real-time recognition but have issues with poor lighting, distances, and dynamics. Navigation and spatial assistance systems may improve mobility but will need additional sensors and hardware. Voice systems are good for hands-free operation but suffer from background noise and processing issues. Face recognition and scene understanding technologies may need more computational power. All the considered papers collectively give the appropriate solution for the individual or limited assistive functions. However, the above-mentioned comparison highlights the necessity to create more integrative technologies which could combine such complementary features as object detection, voice interaction, OCR, face recognition, currency recognition, navigation, and audio guidance in one wearable device. The proposed system is meant to solve this task.

A. Research Gap

The above-reviewed research works exhibit a high level of success in developing assistive technologies for visually impaired people employing artificial intelligence, computer vision, deep learning, IoT, wearables, and audio-based technology. The twenty reviewed articles are based on different requirements that include object detection, obstacle avoidance, navigation, voice assistance, face recognition, scene understanding, spatial perception, sign recognition, and wearable assistance. Nevertheless, most of the reviewed systems are aimed at the use of one or several assistive functions, but not an integrated solution. Both object-detection and navigation technologies provide valuable environmental data; however, in most cases, they are targeted at recognizing obstacles, objects, lanes or targets. Voice and NLP-based systems promote hands-free operation, but they might be impacted by noise, processing time and the environment.

Other technologies, like face recognition and scene understanding, enable more complex perception features, however, they might also consume a significant amount of computing power. Spatial audio and 3D reconstruction techniques make the spatial perception more effective; however, it could require special hardware. IoT and wearable technology can make the system portable and networked, however, their functionality is limited to some certain tasks, for example, obstacle detection or basic recognition. Another limitation mentioned in literature reviews is the absence of a comprehensive context support. All current systems are able to detect or recognize certain types of information and generate appropriate audio feedback. But they are not able to consistently merge information from different visual and interaction modules based on user needs.

These tasks of object detection, text recognition, face recognition, currency detection, and navigation are normally treated as separate tasks. In case of an application where a user is able to interact by using voice commands and can choose the appropriate information from these modules, then it becomes a better way of doing things. This will call for an integration mechanism for these multiple perception and assistance tasks.

In this regard, there exists a necessity for a comprehensive, lightweight and integrated assistive mechanism that encompasses object detection, voice communication, OCR-enabled text recognition, face recognition, currency detection, navigation assistance and audio assistance in addition to the integration of IoT technology. It is the purpose of this paper to address this issue and propose an integrated solution to these functions using smart spectacles.

TABLE 1

COMPARATIVE ANALYSIS OF EXISTING LITERATURE

#	**Technology / Method**	**Key Contribution**	**Limitation**
1	YOLO + OCR + IoT	Real-time object detection and audio navigation	High computational cost

#	Technology / Method	Key Contribution	Limitation
2	Improved YOLO + CNN	Robust outdoor object detection	Affected by poor lighting and occlusion
3	YOLOv8 + RGB Camera + TTS	Real-time obstacle detection and audio guidance	Mainly for indoor navigation
4	YOLOv5 + Speech Recognition + Maps API	Object detection with navigation	Affected by environmental noise
5	OpenCV + MobileNetV3 + CNN + TTS	Voice-guided object recognition	Lower accuracy in dynamic scenes
6	YOLO + OCR + NLP + Speech	Voice assistant for visually impaired users	Processing delay and lighting dependency
7	3D Gaussian Splatting + Mesh	3D spatial understanding	Requires specialized hardware
8	CNN + YOLO + OpenCV	Face and obstacle detection	High computational requirements
9	YOLOv8 + Speech + OpenCV	Voice-assisted object detection	Lower accuracy for distant objects
10	ViT + LSTM + Reinforcement Learning	Intelligent scene understanding	High computational complexity
11	YOLOv7 + Edge AI + Raspberry Pi + NLP	Real-time detection and navigation	Latency in crowded environments
12	EEG + Wavelet Scattering	Blind-road warning system	Complex implementation
13	YOLOv4 + COCO + pyttsx3	Lightweight detection with audio feedback	Less effective outdoors
14	RetinaFace + FaceNet + Deep Learning	Real-time face recognition	Sensitive to lighting
15	Semantic Segmentation + Deep Learning	Pedestrian and lane detection	Limited real-time performance
16	Spatial Audio Rendering	Audio-based environment perception	Requires audio calibration
17	RGB-D + Spatial Audio + TTS	Wearable target localization	Requires extra hardware

#	Technology / Method	Key Contribution	Limitation
18	Raspberry Pi + Object + Face Recognition	Assistive smart glasses	Limited battery and processing
19	IoT + Ultrasonic Sensors + Arduino	Obstacle detection and safe navigation	Limited beyond obstacle avoidance
20	Camera + Sign Detection + Audio	Indoor sign-based navigation	Limited to structured environments

B. Summary of Literature Survey

The literature review demonstrates that Artificial Intelligence (AI), Deep Learning (DL), Computer Vision, and Internet of Things (IoT) have greatly influenced the creation of advanced solutions for visual assistance for visually impaired people. The use of deep learning techniques such as YOLO, CNN, OpenCV, and Vision Transformers has significantly enhanced the accuracy of real-time object detection, and speech recognition along with Text-to-Speech technology has provided additional interaction through voice usage. Recent experiments have also involved face recognition, OCR-based text recognition, spatial sound rendering, and wearables to aid in navigation and environmental perception. According to the literature review, the combination of different technologies, such as object identification, voice interaction, navigation, and IoT connection is much more efficient than single-task performing solutions. Nevertheless, the majority of current systems are still based on single functionalities and suffer from poor performance under changing light, complex background, and dynamic external environment. High complexity and reliance on specialised hardware make it difficult to deploy many systems on wearable devices. In general, research highlights the importance of developing an efficient and integrated AI-powered assistive system featuring real-time object detection, intelligent voice interaction, OCR, face recognition, and navigation assistance. The integration of this system will increase environmental awareness, enhance autonomy of users and applicability in the real world while providing high accuracy of object detection and low computational cost, overcoming the limitations of current research.

IV. CONCLUSION

Impairment in vision has a major impact on the ability of an individual to identify objects in his/her environment, to move around safely, and to carry out activities independently. In recent times, the fast progress made in AI, Computer Vision, Deep Learning and IoT has enabled intelligent assistance technologies that increase environmental awareness through object detection and audio guidance in real time. The literature discussed in this paper revealed that deep learning models such as YOLO, CNN, ViT and OpenCV have made a remarkable improvement in the accuracy and speed of object detection. In the same way, technology such as face recognition, Optical Character Recognition (OCR), speech recognition, and Text-to-Speech (TTS) have increased the potential of assistive devices through their user-friendly interface support.

Till date, many achievements have been made in this aspect; however, the majority of the existing assistive devices are oriented towards performing particular tasks like object recognition, navigation, face recognition, and reading. These assistive devices depend on factors such as lighting condition and occlusions. Such deficiencies point to the need for the development of an all-encompassing and real-time as well as lightweight solution which can tackle the diverse problems faced by the visually impaired in their daily lives.

The real-time IoT voice based smart object detection and audio guidance system for blind people is designed to overcome the above shortcomings by incorporating all the below features into one system: real-time object detection, voice intelligent interaction, text detection using OCR, face recognition, currency detection, and audio guidance .

Combination of multiple assistive features in one intelligent interface will help in enhancing the accessibility aspect and ensuring user safety while providing a more practical solution for people with visual impairments. Combination of such technologies as face recognition, OCR, voice recognition and text to speech will increase the effectiveness of the assistive technologies. Improvements that will be made in future will include the use of multilingual voice recognition, cloud IoT, scene interpretation and edge computing AI optimization.

REFERENCES

[1] An IoT Smart Eyewear with Real-Time Object and Audio Recognition for Visually Impaired Users, IEEE,2026.

[2] Multi-Scene Dataset and Object Detector for Outside Blind Individual Identification, IEEE,2026.

[3] Real-Time Obstacle Detection and Audio Navigation for the Visually Impaired Using YOLO, IEEE, 2025.

[4] A Real-Time Object Detection and Navigation System for Visually Impaired Individuals, IEEE, 2025.

[5] Real-Time Object Recognition with Voice Guided Navigation for Visually Impaired Using OpenCV, IEEE, 2025.

[6] IoT and NLP-Based Voice Assistant for Visually Impaired People, IEEE, 2025.

[7] Visually Impaired People Recognise Virtual Objects Through 3D Reconstructed Shapes, IEEE,2025.

[8] Face and Obstacle Detection-Based Smart Glasses for Blind and Visually Impaired, IEEE,2025.

[9]Echo Vision: Real-Time Object Detection and Voice Assistance for the Visually Impaired, IEEE,2025.

[10]Human Scene Understanding Mechanism- Based Image Captioning for Blind Assistance, IEEE,2025.

[11]YOLOv7-Based Intelligent System for Real-Time Object Detection and Assistive Navigation in Smart Assistive Solutions Using NLP Feedback, IEEE,2025

[12] Real Environment Warning Model for Visually Impaired People in Trouble on the Blind Roads Based on Wavelet Scattering NETWORK, IEEE, 2024.

[13] Real-Time Guide for Visually Impaired, IEEE, 2024.

[14] Surveillance System for Real-Time High-Precision Recognition of Criminal Faces from Wild Videos, IEEE, 2023.

[15] Pedestrian Lane Detection for Assistive Navigation of Vision-Impaired People: Survey and Experimental Evaluation, IEEE, 2022.

[16] Intuitive Environmental Perception Assistance for Blind Amputees Using Spatial Audio Rendering, IEEE, 2022.

[17] Stereo Pilot: A Wearable Target Location System for Blind and Visually Impaired Using Spatial Audio Rendering, IEEE, 2022.

[18] Smart Assistive Sunglasses for the Visually Impaired, IEEE, 2022.

[19] IoT-Based Smart Shoe for the Blind, IEEE, 2022.

[20] 2021 Camera + Sign Detection + Audio Guidance Indoor sign recognition for navigation Restricted to structured indoor environment.