AIoT Smart Eyewear with Real-time Object and Audio Recognition for Visually Impaired Users

Pritam Nanda[1], Soumya Ranjan Samal[2], Shuvabrata Bandopadhaya[3], Debi Prasad Pradhan[2], Antoni Ivanov[4,*] and Vladmir Poulkov[4]

[1]*Department of Computer Science & Engineering, Silicon University, Odisha, India*
[2]*Silicon University, Odisha, India*
[3]*Banasthali Vidyapith, Rajasthan, India, Gurugram, India*
[4]*Faculty of Telecommunications, Technical University of Sofia, Sofia, Bulgaria*
E-mail: astivanov@tu-sofia.bg
[*]*Corresponding Author*

Received 18 August 2025; Accepted 07 November 2025

Abstract

The goal of the proposed smart eyewear tool is to assist visually impaired (VI) or blind individuals by providing object detection and real-time audio descriptions of their surrounding environment. This will help these users in navigating their surroundings safely and independently. This is achieved through the integration of artificial intelligence (AI) and Internet of Things (IoT) technologies into a compact, wearable system. The proposed work presents a real-time, artificial intelligence of things (AIoT) based assistive system designed by using an ESP32-CAM module for real-time object detection, integrated with a TF-Luna LiDAR sensor for distance measurement, and a Bluetooth-enabled neckband for audio feedback. The system uses YOLO for object detection and Roboflow for dataset preparation and training. It also includes features such as night-time navigation and an emergency alert for enhanced safety. Experimental results showed high performance in

Journal of Mobile Multimedia, Vol. 22_1, 41–62.
doi: 10.13052/jmm1550-4646.2212

various lighting conditions, with object detection accuracy of 94.93%, a mean absolute error (MAE) of 0.34 cm, and a root mean square error (RMSE) of 0.44 cm in distance estimation. The SOS alert system responded with 100% accuracy in emergency situations.

Keywords: Assistive system, Roboflow, Tesseract OCR, TF-Luna LiDAR, visually impaired, VIuNI, YOLO.

1 Introduction

Nowadays, engineering and internet technologies play a crucial role in enhancing our daily lives in many ways. These advancements can be effectively utilized to assist people with disabilities in overcoming their challenges and adapting to the changing environment around them. Focusing specifically on visually impaired (VI) individuals/blinds, the World Health Organization (WHO) reports that over 2.2 billion people worldwide are affected across various age groups. Notably, around 90% of blinds reside in developing countries [1]. This includes individuals who are blind, those with moderate to severe distance impairment, mild distance impairment, and presbyopia (in the remainder of this article, the term VI is used to refer to all these categories). These individuals face a lot of difficulties in their daily lives in various aspects, especially during movement on a road with heavy traffic. This requires special attention towards these VI people to assist them as their life quality and independence are reduced, often leading to depression and social isolation. With the emergence of various technologies and gadgets designed to assist these individuals, many of these solutions are either highly expensive or with limited functionalities [2–4]. These assistive devices have been used for VI people to overcome various physical, and social accessibility barriers, enabling them to live independently in society [5, 6].

In the context of assisting VI individuals, a navigation system is essential, as it enables obstacle avoidance and provides guidance toward their destination and developing such systems that can effectively guide these users to move in both indoor and outdoor environments is required. AI vision and audio recognition for VI individuals' navigation independence (VIuNI) is an innovative smart eyewear system designed to empower VI individuals with enhanced mobility, safety, and confidence in their daily lives. Leveraging the combined capabilities of AI and the IoT, this smart assistive system provides real-time audio feedback to help users effectively perceive and understand their environment. This work aims to bridge the accessibility gap by enabling visually impaired individuals to interact with their environment more naturally and independently. With the use of this device, it becomes easier to navigate crowded streets and identify people or objects. The proposed solution is developed with a focus on accessibility and usability, aiming to improve the quality of life and support independent mobility of VI individuals. By utilizing low-cost, compact, and lightweight components, the system offers a scalable and affordable assistive technology suitable for widespread deployment.

The organization of this article is as follows: Section 2 reviews several relevant IoT-based approaches along with their limitations and highlights the key contributions of this work. Section 3 states the objectives and problem formulation of the proposed system. Section 4 presents the system architecture and describes the process-flow. Section 5 presents the model execution and discusses the results. Finally, Section 6 concludes the paper and outlines potential directions for future work.

2 Related Work, Key Contributions and Limitations

Due to the emergence of global technologies, particularly in the fields of AIoT and machine learning, significant advancements have been made in developing intelligent systems and applications for people with visual disabilities. Over the past few decades, several studies have been carried out to support VI individuals through various methodologies and techniques focused on addressing the challenges they face in daily life [7–10]. Most of the research is focused on assistive systems to help VI individuals, which can be broadly classified into two types: communication technology and navigation aids. These systems encompass a wide range of technologies that support navigation, communication, health, and skill development, which supports an active and independent social life [3, 11]. A more detailed study on VI individuals and the various general-purpose assistive technology solutions, such as blind mobility aids, Braille systems, computer and internet access, and communication technologies, are discussed in [12]. Some of these techniques use camera vision and signal processing for object detection and identification [13]. Based on the technology and techniques used, these assistive systems are classified into various categories. A schematic representation of these categories is shown in Figure 1 [4, 9, 12, 13].

A comprehensive review of object detection, with technical advancements and key technologies, is presented in [14]. It offers an in-depth analysis of detection speed-up techniques and focuses on the evolution of

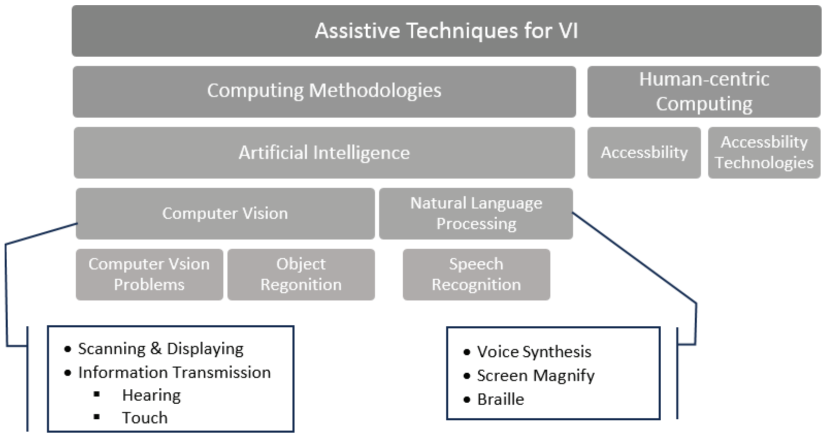


Figure 1 Classification of assistive techniques for the VI individual.

these methodologies. Additionally, it also discusses essential components like datasets, evaluation metrics, etc. An AI-powered smart vision kit is proposed in [15]. This system leverages AI to integrate computer vision (CV) with advanced AI techniques. The system employs the common objects in context (COCO) algorithm to effectively establish the visual interface. Additionally, it utilizes OpenCV and TensorFlow frameworks for efficient obstacle detection. Ashveena et al. [16] proposes a support system for VI individuals, integrating image recognition and text navigation to provide auditory guidance. The system detects indoor obstacles using ultrasonic sensors and processes images via a Raspberry Pi. A camera captures the object, which is analyzed using the single shot detection (SSD) algorithm, converted to text via optical character recognition (OCR), and delivered through a text-to-speech (TTS) system. In [17], the authors developed a prototype using online image processing services such as Microsoft Cognitive Services, Google Text-to-Speech (gTTS), a Raspberry Pi, and a Pi camera module. The goal of the prototype is to provide real-time visual narration for visually impaired and blind users. An intelligent arm-mounted navigation device that combines an RGB-D camera and an inertial measurement unit (IMU) for real-time obstacle detection and recognition is developed in [18]. The system employs advanced point cloud processing using the DBSCAN clustering algorithm enhanced with a KD-tree structure, and applies YOLO-based object recognition models trained on custom datasets. Users receive haptic feedback for directional guidance and distance awareness, along with a voice message about the obstacle's identity and its proximity. Similarly, by using an RGB-D camera, a wearable navigation system for VI individuals aimed at indoor navigation is proposed in [19, 20]. In [19], the authors have proposed a system that uses a head-mounted RGB-D camera and SLAM (simultaneous localization and mapping) to generate 2D maps of the surroundings. An ultrasonic sensor detects obstacles, and an optimal path is planned from the user's current location. A Raspberry Pi 3 B+ is used for data processing. Users receive real-time voice guidance through earphones. In [20], the authors advance the concept by integrating semantic visual SLAM with a high-performance mobile computing platform to provide not only navigation but also detailed understanding of the environment. Using an RGB-D camera with structured light and control center, the system extracts semantic information and delivers it to the user via voice feedback. Experimental results confirm that this system operates in real time with high accuracy, significantly enhancing spatial awareness for VI users. A few researchers have focused on real-time, mobile camera-based navigation systems designed for visually impaired individuals to navigate familiar or indoor environments. Unlike other systems, these solutions operate entirely on mobile devices without the need for external sensors or additional infrastructure [21–23]. A comparative analysis of commercially available AIoT-enabled assistive eyewear for VI users, as well as the proposed solution, is presented in Table 1.

3 Problem Formulation and Contributions

Many existing assistive solutions are priced beyond the financial reach of VI users. In India, a large portion of the VI population relies on modest means of income such as begging, singing or selling handmade items on the street. Setting high prices for these assistive technologies prevents these tools from reaching the intended users and fails to serve their actual purpose. To address this, the proposed model utilizes online CV services as a cost-effective alternative. These services are capable of analyzing images, generating interpretations, and providing textual descriptions, thereby helping reduce the overall development cost. This proposed technique is an innovative smart eyewear system designed to empower VI individuals by enhancing their confidence and independence in navigating their surroundings. This solution leverages the integration of AI and IoT technologies to deliver real-time audio descriptions of the environment.

The key contributions of this work are summarized as follows:

- A novel AI-based smart eyewear system for enhancement of the navigation independence and situational awareness of visually impaired

Table 1 A comparison between commercial AIoT-enabled assistive eyewear for VI Users, and the proposed solution

Product Name	Technology Used	Approx. Cost in USD	Cost Effectiveness	Pros & Cons	Target User Group
OrCam MyEye	• Wear clip camera, AI • OCR, TTS • Gesture control • Cloud access	3000–4500	Low–medium	• High accuracy • Expensive, heavy	Blind and low-vision users who need robust text and face recognition
Ally Solo glasses/AI Vision glasses	• Camera, AI, Bluetooth • Cloud access • Open-ear speaker • Companion app	350–750	High	• Extended battery life • Low-cost, lightweight • Companion to smartphone • Aira integrated	Blind and low-vision users
eSight 4	• High resolution display goggles • Video processing	4500–6000	Low	• Magnification contrast enhancement • Very expensive • Not suitable for users with no light perception	Low vision users (macular degeneration) with usable residual sight
IrisVision	• OCR, voice control, magnification, • Companion app	2500–3500	Medium	• Affordable • Magnification & media features • Useful for home/VR/reading • Less suitable for fully blind users	Low vision users (macular degeneration) who need magnification
NuEyes Pro Series	• Augmented reality (AR)/mixed reality (MR) enabled • TTS, OCR, magnification	3000–6000	Medium–low	• Strong AR features • Good magnification & assistive controls • Highly expensive	Low vision users
Proposed: AIoT-Based Smart Assistive Eyewear	• Light-weight on-device camera, AI, Bluetooth • TFLuna LiDAR sensor • TTS, OCR	300–400	Medium–low	• Extended battery life • Low-cost, lightweight • High accuracy	Blind and low-vision users

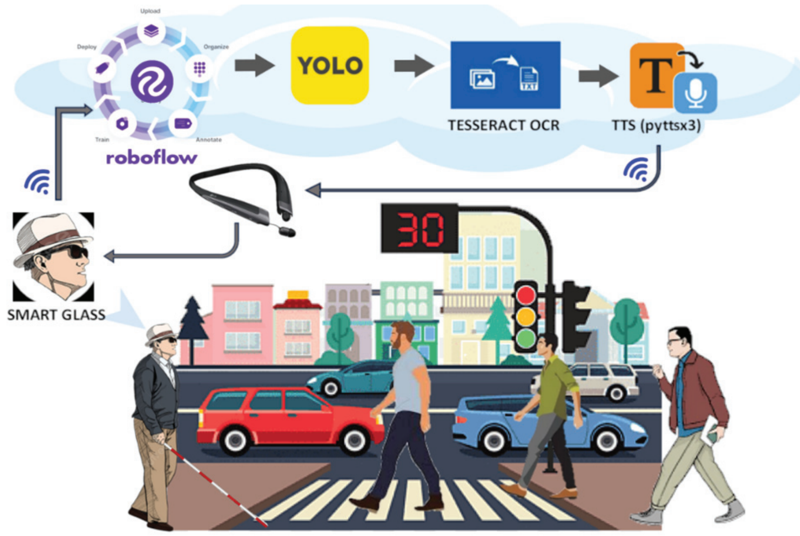
individuals. The system leverages the integration of AI and IoT technologies to provide real-time audio feedback describing the user's surroundings.

- The device is capable of detecting and identifying multiple classes of objects, including human presence and other immobile objects, thereby assisting in both navigation and everyday tasks.
- Real-time audio descriptions are delivered through a Bluetooth-enabled neckband, ensuring a hands-free and unobtrusive user experience.
- An emergency SOS alert feature is integrated into the system to enhance user safety in critical situations.

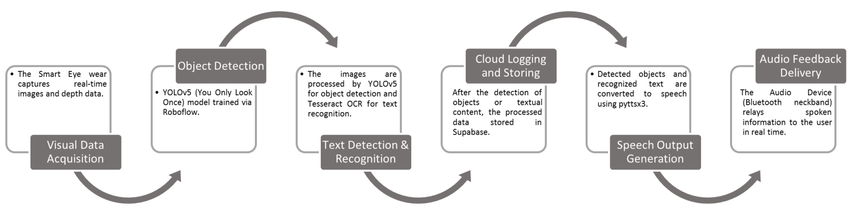
4 Proposed System Architecture and Process Flow

The proposed system, as shown in Figure 2(a), consists of an integrated network of hardware and software components designed to work collaboratively to provide real-time navigation and object recognition for VI users. The proposed smart eyewear acts as the core component, which is made up of lightweight and comfortable wearable device. It captures live visual data, takes sensor input, and sends information to other parts of the system for further processing. It encompasses an ESP32-CAM module, and a TF-Luna LiDAR sensor (with an 8-meter range), enabling efficient data acquisition and preliminary control operations. This innovative system, focused on AI vision and audio recognition, is specifically developed to enhance navigation independence for VI users. By utilizing the integration of AI and IoT technologies, this smart eyewear delivers real-time audio descriptions of the surrounding environment. The assistive feedback, communicated via a Bluetooth-enabled audio device, enables users to navigate independently within both known and unfamiliar surroundings. Furthermore, considering the power efficiency as key factor for such type of assistive wear for VI users as they depend upon uninterrupted operation, this prototype uses a 1000 mAh lithium-polymer battery, which allows this prototype to work for 2–3 hours of continuous operation. The complete workflow is illustrated in Figure 2(b).

Figure 3 shows a block diagram of the use case and the proposed tool, which incorporates an ESP32-CAM module, a TF-Luna LiDAR sensor, Roboflow, YOLO, Tesseract OCR, and a text-to-speech engine (pyttsx3). The ESP32-CAM is responsible for capturing real-time images of the surrounding environment, from the perspective of the VI user. The captured image undergoes augmentation using Roboflow. Then, it compares the captured images with pre-trained datasets, where the objects are marked with



(a) Proposed system model.



(b) Process flow.

Figure 2 System model overview.

bounding boxes and labels to improve detection accuracy. Which are then stored in the Supabase cloud database and can be accessed through a web application, allowing remote monitoring and analysis via mobile devices. The system uses the YOLO algorithm, which allows for fast and accurate real-time object detection. The LiDAR sensor can measure distance up to 8 meters (capable of measuring distances of up to 100 meters), which can assists in depth perception. The entire system operates over a Wi-Fi network, ensuring seamless communication between components. For reading text in the images, the system uses Tesseract OCR, which helps identify and convert the text into a format that the computer can understand and process. Finally, the detected objects or text are converted into speech using a text-to-speech

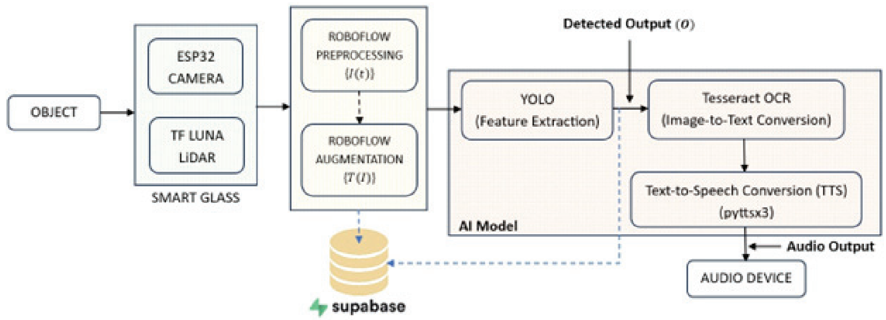


Figure 3 Block diagram.

engine. The audio output is then sent to the user via a Bluetooth-enabled neckband, providing real-time auditory feedback through a mobile device.

4.1 Mathematical Modeling for the Proposed Architecture

The proposed system encompasses integrating hardware (ESP32-CAM, TF-Luna LiDAR), cloud services (Supabase, Roboflow), and AI models (YOLO for object detection, Tesseract for OCR, pyttsx3 for TTS). Each component plays a distinct role in the system process flow. The processing logic is described analytically as follows:

1. Image acquisition: The camera (ESP32 CAM) captures real-time images at time t,
$$I(t) = f(x, y, c), \quad I(t) \in \mathbb{R}^{H \times W \times 3}, \tag{1}$$
where (x, y) denotes the pixel location in the image and x and y represent the horizontal and vertical coordinates, respectively. The variable $c \in R, G, B$ indicates the color channel, red, green, or blue. The set $\mathbb{R}$ represents the set of real numbers. H denotes the height of the image in pixels (i.e., the number of rows), and W denotes the width of the image in pixels (i.e., the number of columns).
2. Dataset preparation using Roboflow: Dataset preparation is a critical step for CV tasks such as image classification, object detection, and semantic segmentation. In this work, Roboflow was used as the main platform to handle the dataset preprocessing, data augmentation, and producing the final dataset for model training. The initial image dataset can be prepared as follows:
$$\text{Let, } \Xi_{Captured} = I_1, I_2, I_3 \ldots I_N. \tag{2}$$

After data augmentation the augmented image can be formed as,

$$I_i' = T_{avg}(I_i), \quad \Xi_{avg} = [\{I_i', (B_{ij}', C_{ij}')_{j=1}^{m_i}\}]_{i=1}^{N}, \tag{3}$$

where T_{avg} is the average number of augmentation transformations (e.g., rotation, flipping, brightness change) applied to each image in the dataset during augmentation and Ξ_{avg} denotes the average number of augmented samples generated from each original image in the dataset. C_{ij} denotes the associated class label and B_{ij} denotes the bounding box associated with the jth object in the ith image. Following data augmentation, $B_{ij}' = f_{aug}(B_{ij})$ represents the transformed bounding box coordinates resulting from applied augmentations such as rotation, scaling, translation, or flipping and C_{ij} denotes the associated class label. Here, m_i is the number of objects in image i, and N is the total number of augmented images generated.

3. Model training with YOLO via Roboflow: The model was trained using the YOLO object detection framework through the Roboflow platform. Let, Ψ represent the set of trainable weights in the YOLO model. The goal of training is to find the optimal model parameters Ψ^*, which minimize the total detection loss L_{YOLO} over the annotated dataset. The optimal values of the model parameters that minimize the loss function over the training dataset is as follows:

$$\Psi^* = \underset{\Psi}{\text{argmin}} L_{YOLO}(\Psi; \Xi_{avg}), \tag{4}$$

where the loss function incorporates multiple components, including localization loss, confidence loss, and classification loss, each contributing to the overall optimization objective of the model.

4. Real-time object detection: During inference, when a new input image $I(t)$ is passed through the trained YOLO model. The detection function is defined as from Equations (3) and (4):

$$\Xi_{YOLO} = \{I(t), \Psi^*\} \rightarrow \{Bi, Ci, Si\}_{i=1}^{n}. \tag{5}$$

The output consists of n triplets (B_i, C_i, S_i), indicating the spatial coordinates, classes, and corresponding confidence score of the detected objects in the image.

5. Distance measurement using TF-Luna LiDAR: To enhance object detection with depth information, each object identified by the YOLO model is matched with a corresponding distance measurement from the TF-Luna LiDAR sensor. The distance d_i to the center of a detected bounding

box B_i is written as,

$$d_i = L(B_i), \quad 0 < d_i \le 8 \text{ m}, \tag{6}$$

where $L(B_i)$ denotes the LiDAR-based distance measurement corresponding to the spatial region defined by bounding box B_i. The maximum effective range of the TF-Luna sensor is 8 meters under optimal conditions.

The final object detection output is written as in Equation (6),

$$O_i = (B_i,\ C_i,\ S_i,\ d_i) \tag{7}$$

In cases where the camera is unavailable or visual confidence is unreliable; a fallback mechanism is employed. The detection confidence is updated based on a function $\alpha(d_i)$ derived from LiDAR distance by using Equation (7),

$$\widetilde{S} = \begin{cases} S_i, & \text{if the camera is in active mode,} \\ \alpha(d_i), & \text{if fallback to LiDAR Sensor.} \end{cases} \tag{8}$$

Here, $\alpha(d_i) = \exp(-\beta d_i)$ is a monotonically decreasing confidence function based on object distance where $\beta > 0$ is a tunable decay parameter.

6. OCR using Tesseract: The region of the input image $I(t)$ that contains textual information can be written as $I_{text} \subset I(t)$. This region is processed through Tesseract OCR to produce a sequence of recognized text tokens as shown below,

$$T = O(I_{text}) = \{t_1,\ t_2,\ t_3, \ldots, t_k\} \tag{9}$$

where $O(\cdot)$ denotes the OCR operation, and t_i represents the ith recognized character. To quantitatively evaluate the accuracy of OCR output, the similarity score Φ is used, which measures how closely the predicted text T matches the ground truth T^*. The similarity can be represented as:

$$\Phi(T, T^*) = 1 - \frac{\mathbb{E}(|T|, |T^*|)}{\max(|T|, |T^*|)} \tag{10}$$

where $\mathbb{E}(T, T^*)$ denotes the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform T into T^*, and $(|T|, |T^*|)$ represents the lengths of the respective text sequences.

7. Text-to-speech using pyttsx3: Finally, the detected objects and the recognized text are combined into a spoken message using the pyttsx3 TTS engine. The complete message M can be generated by joining the object descriptions and the text detected from the image as given below

$$
\begin{aligned}
M = & \sum_{i=1}^{n} \text{“Object Found”} + C_i + \text{ “at” } + d_i \text{ “meters away”)} \\
& + \sum_{j=1}^{k} (\text{“Textread:” } + t_j).
\end{aligned}
\tag{11}
$$

This message M as in Equation (8) is then passed to the TTS engine Γ, producing an audio signal $A(t)$,

$$A\,(t) = \Gamma(M) \tag{12}$$

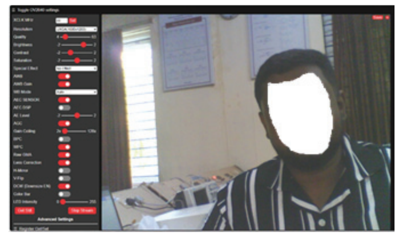
The resulting audio as expressed in Equation (9), is transmitted to the user in real time via a Bluetooth-enabled neckband.

5 Practical Realization and Result Discussions

The proposed smart eyewear tool has been evaluated for its practical utility in assisting VI individuals with their daily tasks. The system is implemented on an ESP32-CAM microcontroller integrated with a TF-Luna LiDAR sensor, and employs a suite of lightweight yet efficient AI models for real-time perception and interaction. Specifically, it utilizes YOLO for object detection, Tesseract OCR for text recognition, and custom-trained models via Roboflow for facial and currency identification. Additionally, the system incorporates an emergency response module to ensure user safety in critical situations. The complete prototype is shown in Figure 4(a) and an example of the feed of the camera module is given in Figure 4(b).

To enhance obstacle detection accuracy and depth perception, the proposed system incorporates a LiDAR sensor (TF-Luna) that provides real-time distance measurements. The LiDAR module continuously emits infrared pulses and calculates the time-of-flight (ToF) for each reflected signal, thereby generating precise distance data for objects in the immediate path of the user. This distance information is sampled at regular intervals and synchronized with visual input from the ESP32-CAM. The data stream helps to identify obstacles. The LiDAR feed is processed on-device using a lightweight distance-thresholding algorithm, triggering auditory alerts when

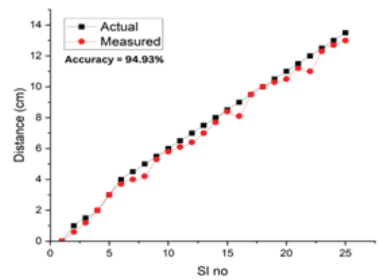
(a) Prototype smart eyewear.



(b) Camera module feed of the prototype.

Figure 4 Overview of the smart eyewear.

(a) Lidar feed example.



(b) Lidar distance accuracy curve.

Figure 5 Integration and performance of the Lidar.

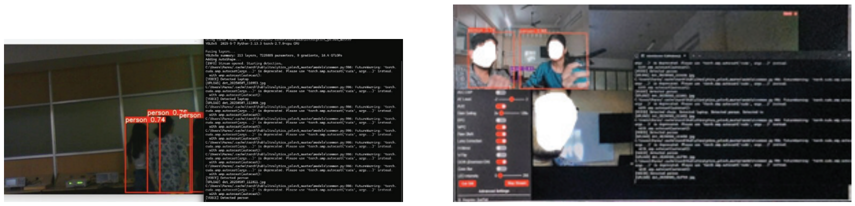
detected distances fall below a predefined safety threshold. The integration of LiDAR sensor with the smart eyewear tool is illustrated in Figure 5(a).

Figure 5(b) shows the accuracy of the LiDAR-based distance measurement module was validated by comparing the actual distances with the measured distances obtained from the TF-Luna LiDAR sensor. The graph above illustrates the comparison across 25 sample points, where the actual distances (black squares) and the measured distances (red circles) are plotted. The measured values closely follow the actual values, indicating a high degree of correlation and reliability. The system achieved a measurement accuracy of 94.93%, with a root mean square error (RMSE) of approximately 0.44 cm and a mean absolute error (MAE) of 0.34 cm, as shown in Table 2. These values confirm the reliability of the LiDAR sensor for near-field obstacle detection tasks, which are critical for real-time navigation assistance in visually impaired user applications. Minor deviations observed are within acceptable limits for wearable assistive systems, especially considering environmental noise and surface reflectivity.

Table 2 Actual vs. measured distance with error analysis

	Actual (cm)	Measured (cm)	Absolute Error (cm)	Squared Error (cm^2)
1	0.0	0.0	0.00	0.000
2	0.5	0.6	0.40	0.160
3	1.5	1.2	0.30	0.090
4	2.0	2.0	0.00	0.000
5	3.0	3.0	0.00	0.000
6	4.0	3.7	0.30	0.090
7	4.5	4.0	0.50	0.250
8	5.0	4.2	0.80	0.640
9	5.5	5.3	0.20	0.040
10	6.0	5.8	0.20	0.040
11	6.5	6.1	0.40	0.160
12	7.0	6.4	0.60	0.360
13	7.5	7.0	0.50	0.250
14	8.0	7.7	0.30	0.090
15	8.5	8.4	0.10	0.010
16	9.0	8.1	0.90	0.810
17	9.5	9.5	0.00	0.000
18	10.0	10.0	0.00	0.000
19	10.5	10.3	0.20	0.040
20	11.0	10.5	0.50	0.250
21	11.5	11.2	0.30	0.090
22	12.0	11.0	1.00	1.000
23	12.5	12.3	0.20	0.040
24	13.0	12.7	0.30	0.090
25	13.5	13.0	0.50	0.250
Mean absolute error (MAE)			**0.34 cm**	–
Root mean square error (RMSE)			–	**0.44 cm**

The object detection module, was evaluated under both bright-light and low-light environments to assess its robustness in real-world scenarios, as shown in Figures 6(a) and 6(b), respectively. In bright-light conditions, the proposed tool consistently achieved high accuracy, with clear object boundaries and minimal false detections due to optimal image contrast and detail visibility. Whereas under low-light conditions, the performance slightly decreased due to reduced illumination and increased image noise. Furthermore, objects with reflective or high-contrast surfaces were detected more easily than those with dark or uniform textures. The evaluation confirmed that the system remains functionally reliable across varying lighting conditions, ensuring continuous support for VI users during both day and night navigation.



(a) Detecting objects in low-light. (b) Detecting objects in bright-light.

Figure 6 Usage of the proposed eyewear tool.

6 Conclusions and Future Work

This work presents a real-time, AI-based smart assistive system designed to support VI individuals in navigating their surroundings safely and independently. The proposed model includes several features, including real-time object detection, distance measurement using a TF-Luna LiDAR sensor, night-time navigation, and an emergency SOS alert. All modules were implemented on a light weight, compact, and inexpensive wearable device. A Bluetooth-enabled neckband provides real-time audio feedback, helping users understand their surroundings through voice messages. To reduce implementation costs the system utilizes the lightweight YOLO object detection algorithm in combination with Roboflow for dataset preparation, and model training. Experimental results confirm the system's effectiveness in different lighting conditions, showing high accuracy in detecting both moving and stationary objects, including humans, along with reliable distance measurement and quick emergency response. The system achieved a high accuracy of 94.93%, with an MAE of 0.34 cm and an RMSE of 0.44 cm, in identifying commonly encountered objects in both low-light and bright-light conditions. These results validate the system's reliability in real-time obstacle distance estimation. The SOS feature successfully triggered alerts with 100% accuracy during simulated emergency situations. These results confirm the robustness and effectiveness of the proposed system in enhancing the mobility, safety, and independence of VI users. As a future work, improving the system's efficiency through hardware optimization and a deep learning model can be implemented to achieve faster processing on edge devices. Furthermore, additional features such as gesture/face recognition, multilingual support, GPS-based navigation, and integration with wearable AR glasses can also be added in the future to make the system easier and more effective to use. The priority is the incorporation of GPS to enable reliable outdoor navigation and seamless transition between indoor and outdoor environments.

Subsequently, AR-based assistance will provide spatial guidance through contextual overlays, such as directional cues and landmark highlights. This phased roadmap ensures systematic development while maintaining system usability and performance and offers a clear direction for advancing the navigation capabilities of the proposed framework. Testing the system with larger number of users in real-world environments will help in confirming its reliability and effectiveness.

Acknowledgments

This research is financed by the Operational Programme: Research, Innovation and Digitalisation for Smart Transformation Programme, through the Ministry of Education and Science of the Republic of Bulgaria, Executive Agency "Programme Education", project No. BG16RFPR002-1.009-0001 Enabling research excellence network in integrated sensing and communications for 6G (ENCORE-6G).

References

[1] WHO, "Visual impairment and blindness," August 2023.

[2] M. Mashiata, B. Mondal, K. Deb, P. P. Paul, and M. R. Islam, "Towards assisting visually impaired individuals: A review on current status and future prospects," *Biosensors and Bioelectronics: X*, vol. 12, p. 100265, 2022.

[3] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state-of-the-art survey," *Pattern Recognition Letters*, vol. 137, pp. 37–52, 2020.

[4] P. Kathiria, P. Parmar, D. Kothari, A. Patel, and P. Solanki, "Assistive systems for visually impaired people: A survey on current requirements and advancements," *Neurocomputing*, vol. 606, p. 128284, 2024.

[5] W. Elmannai and K. Elleithy, "A highly accurate and reliable data fusion framework for guiding the visually impaired," *IEEE Access*, vol. 6, pp. 33029–33054, 2018.

[6] M. Hersh and M. Johnson, "On modelling assistive technology systems–part i: Modelling framework," *Technology and Disability*, vol. 20, pp. 193–215, 2008.

[7] Z. Muhsin, R. Qahwaji, F. Ghanchi, S. Al-Alwani, and R. Al-Mufarji, "Review of substitutive assistive tools and technologies for people with

visual impairments: recent advancements and prospects," *Journal of Multimodal User Interfaces*, vol. 18, pp. 135–156, 2024.

[8] I. P. Singh, A. Misra, A. Maurya, and A. Verma, "Technical review of smart assistive devices for visually impaired people," in *2023 4th International Conference for Emerging Technology (INCET)*, (Belgaum, India), pp. 1–12, 2023.

[9] A. Lavric, C. Beguni, E. Zadobrischi, A.-M. Căilean, and S.-A. Avătămăniței, "A comprehensive survey on emerging assistive technologies for visually impaired persons: Lighting the path with visible light communications and artificial intelligence innovations," *Sensors*, vol. 24, no. 15, p. 4834, 2024.

[10] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet of Things*, vol. 11, p. 100188, 2020.

[11] R. Velázquez, "Wearable assistive devices for the blind," in *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment* (A. Lay-Ekuakille and S. Mukhopadhyay, eds.), vol. 75 of *Lecture Notes in Electrical Engineering*, Berlin, Heidelberg: Springer, 2010.

[12] J. Brabyn, K. Seelman, and S. Panchang, "Aids for people who are blind or visually impaired," in *An introduction to rehabilitation engineering* (R. Cooper, H. Ohnabe, and D. Hobson, eds.), pp. 287–313, Taylor & Francis, 2007.

[13] M. H. Abidi, A. N. Siddiquee, H. Alkhalefah, and V. Srivastava, "A comprehensive review of navigation systems for visually impaired individuals," *Heliyon*, vol. 10, no. 11, p. e31825, 2024.

[14] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[15] K. Iqbal, S. S. Ali, Z. Sajid, M. Samad, L. Mubarak, and S. Ali, "Advancement in smart vision systems: A computer vision-based assistive system for visually impaired individuals," *VAWKUM Transactions on Computer Sciences*, vol. 13, no. 1, pp. 244–257, 2025.

[16] A. A., B. D. J., S. P. Mary, and D. U. Nandini, "Portable camera based identification system for visually impaired people," in *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, (Tirunelveli, India), pp. 1444–1450, 2023.

[17] H. Baskaran, R. L. M. Leng, F. A. Rahim, and M. E. Rusli, "Smart vision: Assistive device for the visually impaired community using online computer vision service," in *2019 IEEE 4th International*

Conference on Computer and Communication Systems (ICCCS), (Singapore), pp. 730–734, 2019.
[18] Z. Li, F. Han, and K. Zheng, "An rgb-d camera-based wearable device for visually impaired people: Enhanced navigation with reduced social stigma," *Electronics*, vol. 14, no. 11, p. 2168, 2025.
[19] H. Hakim and A. Fadhil, "Indoor wearable navigation system using 2d slam based on rgb-d camera for visually impaired people," in *Proceedings of First International Conference on Mathematical Modeling and Computational Science* (S. Peng, R. Hao, and S. Pal, eds.), vol. 1292 of *Advances in Intelligent Systems and Computing*, Singapore: Springer, 2021.
[20] Z. Chen, X. Liu, M. Kojima, Q. Huang, and T. Arai, "A wearable navigation device for visually impaired people based on the real-time semantic visual slam system," *Sensors*, vol. 21, no. 4, p. 1536, 2021.
[21] F. Song, Z. Li, B. Clark, D. Grooms, and C. Liu, "Camera-based indoor navigation in known environments with orb for people with visual impairment," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*, (Seattle, WA, USA), pp. 1–8, 2020.
[22] S. u. Rahman, S. Ullah, and S. Ullah, "A mobile camera based navigation system for visually impaired people," in *Proceedings of the 7th International Conference on Communications and Broadband Networking*, (USA), Association for Computing Machinery, 2019.
[23] G. M. B. Catedrilla, "Mobile-based navigation assistant for visually impaired person with real-time obstacle detection using yolo-based deep learning algorithm," in *2022 5th Asia Conference on Machine Learning and Computing (ACMLC)*, (Bangkok, Thailand), pp. 63–67, 2022.

Biographies

Pritam Nanda is an Assistant Professor in the Faculty of Engineering and Technology at Sri Sri University, India. He is currently pursuing his Ph.D.

at Silicon University, India. He received his B.Tech. degree in electronics and communication engineering from Biju Patnaik University of Technology (BPUT), India, in 2015, and subsequently completed his M.Tech. degree in computer science and engineering from BPUT, India, in 2024. His academic and research interests include cloud computing, Internet of Things (IoT), data analytics, artificial intelligence, and machine learning, with a particular focus on their applications in healthcare and smart systems. He is actively engaged in teaching, research, student mentoring, and academic coordination, and his professional activities reflect a strong commitment to innovation, institutional development, and technology-driven education.

Soumya Ranjan Samal received his Ph.D. degree in communication networks, faculty of telecommunications from Technical University of Sofia at Sofia, Bulgaria. He received his B.Tech. degree in electronics & instrumentation engineering from Biju Patnaik University of Technology, India in 2004. Soumya then went on to pursue his M.Eng. in computer science & engineering from the Utkal University of Bhubaneswar, India in 2009. He, as an Associate Professor in Silicon University, India, has acquired solid experience of about 18 years of teaching in communication engineering. Soumya also worked as a Project Engineer in Indian Institute of Technology, Bombay, India in 2005. His research areas of interest include, Interference management in 5G cellular networks, green communication movement to develop energy efficient solutions through antenna parameters and IoT.

Shuvabrata Bandopadhaya is currently working as an associate professor in the department of electronics at the School of Physical Sciences, Banasthali Vidyapith, Rajasthan. He received his M.Tech. and Ph.D. degrees in communication systems specialisation from KIIT University, Bhubaneswar, India. He has nearly 20 years of experience in teaching and research at various reputed institutes and universities in India. His areas of research interest include wireless communication and networks, Internet of Things, and AI.

Debi Prasad Pradhan is currently working as a Technical Assistant at the IoT Lab and Industrial Control Lab at Silicon University, Bhubaneswar. He holds an M.Tech. degree in electronics and communication (2018) from CET Bhubaneswar, Odisha, and a B.Tech. in electronics and communication (2015). He has over 15 years of academic and mentoring experience at Silicon University, India. His research interests include AI-integrated wireless systems for IoT, optical sensor network design, and emerging technologies in industrial control, particularly PLC and DCS systems for Industry 5.0.

Antoni Ivanov received his Ph.D. degree in communication networks and systems from the Technical University of Sofia (TUS), Bulgaria. He holds a master's degree in innovative communication technologies and entrepreneurship from TUS, and Aalborg University, Denmark in 2016. He is currently a postdoctoral researcher at the "Teleinfrastructure Lab", Faculty of Telecommunications, TUS. His research interests include cognitive radio networks, adaptive algorithms for dynamic spectrum access, deep learning-based solutions for cognitive radio applications, volumetric spectrum occupancy assessment, and graph signal processing for resource allocation in current and future wireless networks.

Vladimir Poulkov received his M.Sc. and Ph.D. degrees from the Technical University of Sofia (TUS), Sofia, Bulgaria. He has more than 30 years of teaching, research, and industrial experience in the field of telecommunications. He has successfully managed numerous industrial, engineering, R&D and educational projects. He has been Dean of the Faculty of the Telecommunications at TUS and Vice Chairman of the General Assembly of the European Telecommunications Standardization Institute (ETSI). Currently, he is the Head of the "Teleinfrastructure" R&D Laboratory at TUS and

Chairman of Cluster for Digital Transformation and Innovation, Bulgaria. He is Fellow of the European Alliance for Innovation and a Senior IEEE Member. He has authored many scientific publications and is tutoring B.Sc., M.Sc., and Ph.D. courses in the field of information transmission theory and wireless access networks.