Vision sense: Object Detection and Navigation Aid for the Visually Impaired

Sama Premanand Reddy
MCA (School of Computer Application) Lovely Professional University Phagwara, India premanandreddysama@gmail.com

Chenna Sravan Kumar
MCA (School of Computer Application) Lovely Professional University Phagwara, India sravankumarchenna717@gmail.com

Dr. Kamal Nain Sharma
MCA(School of Computer Application)Lovely Professional University Phagwara, India Kamalnain3@gmail.com

Donthikurthi Pradeep Kumar
MCA (School of Computer Application)Lovely Professional University Phagwara, India donthikurthipradeep2696@gmail.com

Neelam Lokesh Sai Raja
MCA (School of Computer Application) Lovely Professional University Phagwara, India neelamlokeshsairaja@gmail.com

Mummudisetty Lokesh Manikanta
MCA (School of Computer Application) Lovely Professional University Phagwara, India mummudisettylokesh@gmail.com

Nikhilesh Kumar Upadhyay
MCA (School of Computer Application) Lovely Professional University Phagwara, India imnikhileshkr@gmail.com

***Abstract*—living in a world without vision is fraught with many challenges that sighted people take for granted. For the visually impaired, even mundane activities like walking through crowded areas, detecting obstacles, or identifying objects can be intimidating and even hazardous. Important views are like white canes and training dogs, and information provision limitations. With recent changes in computer vision and machine learning, there is an increasing chance to create smart systems that can improve accessibility to blind people. VisionSense is a new object detection and navigation system's goal is to overcome these challenges by combining computer vision technologies. By using real-time image processing and deep learning based object detection algorithms, VisionSense is able to detect and inform the user about the presence and location of objects within their surroundings. Coupled with initiative audio or haptic feedback mechanisms, the system enables users to make informed navigation choices and steer clear of possible dangers. This work describes the design, development of Vision Sense, emphasizing its capability to greatly enhance the quality of life for the visually impaired. The system is not just affordable and light weight but also flexible to indoor and outdoor spaces, making it a viable remedy for accessible, real-time wayfinding aids.**

***Keywords*—Python, Flask, OpenCV, YOLO (YOU ONLY LOOK ONCE) v8, HTML, CSS, Java stack, Bootstrap, ML(MACHINE LEARNING).**

I. INTRODUCTION

Visually impaired people continue to experience difficulties in independently moving around their environments and rec- ognizing objects in real-time[1]. Traditional aids like white canes or guide dogs only give limited environmental feedback and do not readily identify obstacles above or moving. VisionSense addresses these objects by developing an intelligent, real-time object detection and navigation system that has an visually im- paired users with confidence, and security.[2] The foundation of VisionSense is the YOLOv8 (You Only Look Once version 8) algorithm, object detection model renowned for its precision. Taking advantage of OpenCV's Deep Neural Network (DNN) module, the system analyzes live video feeds from a camera to detect and locate multiple objects in real-time.[6] This analyzed information is converted into significant sound cues or touch feedback to alert users of surrounding objects. onehind obsta- cles, enabling prompt decision-making and greater safety. A clever and considerable aspect of VisionSense is it supports multilingual voices. The language flexibility guarantees that users hear navigational directions in their desired language, rendering the system more intuitive and accessible in diverse populations[7]. The YOLOv8 model is trained on the COCO dataset, a most important dataset with more than 80 classes of objects. This training allows the system to generalize real-world scenarios, correctly recognizing common objects in in- door and outdoor environments. Daily living surroundings are a major challenge for people with visual disabilities. Obstacles, uneven flooring, and moving environments has daily threats, commonly stopping independence and mobility.[2] Although white canes and trained dogs are conventional tools, they are not sufficient in providing full awareness or identifying specific items in real- time. VisionSense is a smart object iden- tification and navigation system which designed to improve the mobility and independence of visually impaired people using advanced computer vision and artificial intelligence. Developed mainly using Python, VisionSense uses libraries such as OpenCV for image processing and TensorFlow or PyTorch for deep learning- based object detection. The system uses the YOLO v8(You Only Look Once) architecture to detect objects quickly and accurately from live video streams[3]. In order to provide useful and intuitive directions, Vision Sense uses audio feedback. detected objects and navigation hints are shared with the user by speech output through text-to-speech engines like gTTS or pyttsx3. It gives real-time data without using physical or haptic feedback, thus making the system light and non-intrusive.[4]

This paper promotes the architecture, implementation, and performance analysis of Vision Sense as a good practical and accessible navigation tool[5]. By integrating deep learning with voice interaction, Vision Sense provides a cost- efficient and viable solution to enhance understanding and autonomy for the visually impaired.

II. CONTRIBUTION

Creation of a Real-Time Object Detection System: Vision Sense detects and divides many things Using deep learning with the YOLO (you Only Look Once) method model though live video input, providing high accuracy and low latency which is suitable for real-world navigation[8]. Integration of Computer Vision with Audio Feedback: In contrast to many current systems that use haptic feedback, VisionSense out- puts voice-based using text-to-speech (TTS) technology[6]. This method enables the non-intrusive, hands-free interaction and transfers rich contextual information about the environment. Use of Open-Source and Affordable Technologies: The system is created with a open-source libraries like the OpenCV, TensorFlow/PyTorch, and also Python-based TTS tools, which makes the system very accessible, low-cost, and flexible for many users and developers. Portable and Flexible Architecture:[7] Vision Sense can be able to operate on lightweight, portable hardware like a Raspberry Pi or edge computing devices, allowing users to easily transfer the system when they are outdoors or indoors.

III. LITERATURE REVIEW

3.1. Traditional Navigation Aids for the Visually Impaired, Visually impaired people have over the centuries depended on simple mobility aids like the white cane and guide dogs for everyday mobility. The simple, dependable, and easy-to- use traditional aids have found widespread use globally.[9] The white cane functions as a haptic tool to sense obstacles in the immediate environment on the ground, whereas guide dogs are taught to steer clear of hazards and assist in mobility through intricate settings.However, despite their reliability, these solutions are inherently limited in their scope. They do not offer contextual awareness, cannot convey information about the surroundings, and also typically provide information limited to the user's immediate physical space. Further more, trained dogs require intensive training, high costs, and ongoing care making them inaccessible for many users and developers. As urban environments grow more difficult, the limitations of these traditional tools become increasingly , highlighting the need for more intelligent and responsive systems.

3.2. Ultrasonic Sensor-Based Systems With the emergence of embedded systems and affordable microcontrollers, ultra- sonic sensor-based aids were introduced as the next step in enhancing spatial awareness for the visually impaired. Devices like the Smart Cane incorporated ultrasonic transceivers to measure distances by emitting sound waves and interpreting their echoes, enabling users to detect objects ahead of them without physical contact.[10] These systems typically alert users through vibrations or auditory cues when an obstacle is detected within a predefined range. While a step forward from traditional tools, ultrasonic sensor-based systems are constrained by their inability to provide object- specific information or environmental context. The data provided is often binary (obstacle or no obstacle), with little to no understanding of object type, orientation, or motion.[11] Additionally, performance may be compromised in noisy environments or when dealing with soft surfaces that poorly reflect sound. These limitations sparked interest in developing more advanced systems with environmental perception capabilities.

3.3. Computer Vision in Assistive Technology The intro- duction of computer vision into assistive technology marked a revolutionary shift from proximity-based sensing to contextual scene understanding. The use of computer vision allows ma- chines to analyse visual information similarly to how humans perceive their environment.[12] By using cameras and advanced image processing algorithms, systems can now recognize a wide variety of environmental features, including objects, signs, text, pedestrians, and pathways. Early vision-based aids incorporated optical character recognition (OCR) for reading text aloud and basic object recognition to identify common items like chairs, tables, or traffic signals. These improvements enhanced the semantic understanding of surroundings, empowering users to make informed decisions. The rise of mobile computing and edge AI devices made real-time processing more feasible, bringing practical computer vision applications closer to daily use for visually impaired individuals. Unlike ultrasonic sensors, which merely detect presence, vision-based systems offer detailed classification and spatial interpretation, paving the way for a new generation of smart navigation aids.[13]

3.4. Object Detection Algorithms and Real-Time AI As computer vision evolved, deep learning played a transforma-tive role in enhancing object detection capabilities. Traditional object recognition relied on manually created features and traditional classifiers, which frequently did not work well in cluttered or dynamic settings. [7] Models like YOLO (You Only Look Once) became ground-breaking Convolutional neural networks (CNNs) were introduced, because of their great ac- curacy in Real-time detection of objects. Because YOLO is capable of processing a complete image in a single pass, it is especially well-suited for live navigation systems where speed and latency are crucial. For assistive technology, this meant users could be informed not only about the presence of obstacles but also about what those objects are—be it a car, a pedestrian, or a traffic light—and how far they are positioned in the frame. In the context of VisionSense, [6] YOLO plays a core role in enabling real-time, multi-object awareness, significantly improving upon previous generations of detection methods that were either too slow or too narrow in scope.

3.5. Speech-Based Feedback Systems A crucial aspect of modern assistive technology lies in how information is con- veyed to the user. Since visual output is not feasible for the visually impaired, most systems have adopted speech- based feedback mechanisms. These systems use Text-to- Speech (TTS) engines to convert recognized data into audible messages. Tools like gTTS (Google Text-to- Speech)[7], pyttsx3, and Amazon Polly are widely integrated into assistive ap- plications to relay object labels, directions, or warnings in natural-sounding voices. Compared to traditional audio alerts such as beeps or vibrations, speech output provides a more intuitive, descriptive, and informative experience. Users can be informed, for example, "Person standing on the left" or "Chair ahead two meters," which allows for better situational aware-ness and decision-making. The integration of TTS engines with real-time object detection, as seen in VisionSense, ensures a seamless flow of

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

contextual information that mirrors how sighted individuals perceive and respond to their environment.[6]

3.6. Mobile and Wearable Vision Aids As computer vi- sion became more practical, companies began developing wearable assistive devices that combined AI-powered percep- tion with user-friendly interfaces. Products like OrCam and Microsoft's Seeing AI leveraged compact camera modules, onboard processing, and TTS engines to deliver real-time visual understanding through audio. [1]These systems could read text, recognize faces, detect objects, and even describe scenes. However, despite their effectiveness, these devices often remain expensive and rely on proprietary hardware and software, limiting their accessibility to users in low-income or rural communities. VisionSense addresses this challenge by focusing on an open-source, affordable alternative that offers similar features using publicly available libraries and budget- friendly hardware. By reducing reliance on commercial ecosystems, VisionSense promotes accessibility, adaptability, and widespread adoption.[2]

3.7. Real-Time Image Processing using OpenCV The OpenCV (Open Source Computer Vision Library) serves as the basis for numerous vision-based systems, such as Vi- sionSense. Edge detection, contour analysis, colour filtering, motion tracking, and object segmentation are just a few of the extensive array of image processing features offered by OpenCV.[3] Its lightweight design, cross-platform compatibility, and strong integration with Python make it an ideal tool for building real-time systems. In VisionSense, OpenCV is used to manage video input from cameras, preprocess frames, and support the object detection pipeline. Unlike older systems that relied solely on sensor data or external APIs, VisionSense runs its entire processing stack locally, ensuring privacy, speed, and offline usability. [5]This capability is essential for deployment in areas with limited internet access or on embedded devices like Raspberry Pi.

3.8. Limitations of Existing Systems Despite numerous advances, many existing assistive technologies continue to suffer from specific limitations that restrict their real-world usability. These include limited object detection coverage, high latency, dependency on ideal lighting conditions, and complex setup procedures. In particular, systems designed for indoor use often fail in outdoor or low-light environments.
[7]Addi- tionally, the cost and proprietary nature of many commercial products exclude a significant portion of potential users. Haptic feedback systems, while innovative, may not be intuitive for all users and can require physical contact that some find intrusive. Furthermore, some systems depend on cloud-based processing, introducing privacy concerns and reliance on stable internet connections. The gap between sophisticated but expensive technologies and accessible, real- world solutions remains a major barrier—one that VisionSense aims to bridge.[8]

3.9 VisionSense in Context: Bridging the Technological Divide VisionSense emerges as a comprehensive and inclusive solution, built with the goal of making intelligent navigation assistance both affordable and accessible. It combines several of the most effective modern technologies: the real-time de- tection power of YOLO, [5] the flexibility of OpenCV, and the simplicity of Python-based development. With speech-based feedback replacing complex haptic interfaces, the system en- sures an intuitive and hands-free experience. By leveraging open-source tools, VisionSense remains modular and cus- tomizable, encouraging community-based improvements and adaptation for various user needs. It avoids the pitfalls of pre- vious systems—such as limited scope, high cost, or hardware complexity—by offering a portable, low-latency solution that can be deployed on a smartphone, Raspberry Pi, or similar device[6]. In doing so, VisionSense not only advances the state of assistive technology but also democratizes access to it, helping People with vision impairments move around the world more confidently and independently.

IV. METHODOLOGY

4.1. Dataset Acquisition: A diverse dataset containing im- ages of everyday objects relevant to visually impaired navigation is utilized. Public datasets such as COCO are employed, offering pre-annotated bounding boxes for objects like people, vehicles, animals, and obstacles commonly encountered in indoor and outdoor environments.[7]

4.2. Image Preprocessing: To satisfythe YOLOv8 model's input parameters, the images are scaled and normalised. To guarantee accurate model evaluation and avoid overfitting, Next, the dataset is separated into instruction, validation, as well as test sets.

4.3. YOLOv8 Model Setup: YOLOv8 Model Setup: Real- time performance is the default setting for the YOLOv8 object detection architecture.[6] Its anchor-based detection heads and deep convolutional layers are tuned to accurately detect several objects every frame. Learning rate, batch size, and confidence thresholds are among the hyperparameters that are adjusted.

4.4. Model Training and Fine-Tuning: The COCO dataset's pre-trained weights are used to initialise the model. The model is further trained on the chosen object classes that are perti- nent to the VisionSense system using transfer learning. This improves detection capabilities in practical mobility situations.[2]

4.5. Video Stream Integration via OpenCV: The trained YOLOv8 model is integrated with OpenCV, which handles video stream capture and frame processing. The OpenCV DNN module is used to load and run the YOLOv8 model efficiently, enabling real-time inference on input frames from a camera.

4.6. Real-Time Object Detection Pipeline: Vision Sense captures live video from a camera, processes each frame, and performs real-time object detection using YOLOv8.[3] Bounding boxes and class labels are generated for all recognized objects, enabling immediate identification and response.

4.7. Text-to-Speech (TTS) Voice Output: Detected object labels are converted into clear voice alerts using Python-based text-to-speech (TTS) libraries such as gTTS or pyttsx3[6]. The voice output informs the user of the objects detected and their spatial direction (e.g., “Person on the right”)

4.8. Support for Multiple Languages: The system may produce output in multiple languages that users can select their preferred language, and object labels are translated using trans- lation APIs before being converted to speech. This enhances accessibility for non-English speakers.

4.9. Functional Testing and User Feedback: Both controlled and real-world settings are used to test Vision Sense. Measured performance indicators include speech clarity, latency, and detection accuracy. Usability is evaluated through feedback from visually impaired participants.

4.10. System Optimization and Enhancement: Following testing, areas such as detection speed, audio timing, and language clarity are optimized. Based on user feedback, im- provements are made to enhance navigation precision and reduce false positives in object detection.[10]

4.11. Deployment and User Accessibility: The system is deployed on portable platforms such as Raspberry Pi or laptops. A simple voice-based interface is provided to ensure ease of use. No visual interface is required, aligning with the needs of the visually impaired.

4.12. Final Validation and Conclusion: A final assessment is conducted to evaluate overall system performance, user satisfaction, and navigation efficiency. The results confirm the effectiveness of Vision Sense as a low-cost, AI-driven navigation aid, contributing to the broader field of assistive technology.

The Model: The project’s novelty stems from the innovative fusion of YOLOv5 object detection with customizable real- time audio output. VisionSense empowers visually impaired individuals by identifying and describing their surroundings on the go.[1] The personalized voice feedback system enhances situational awareness and promotes safe, independent naviga- tion. By eliminating reliance on physical feedback systems, VisionSense sets a new standard for intelligent, inclusive assistive technologies that improve mobility and enrich the lives of visually impaired users.

V. IMPLEMENTATION

5.1 Using the COCO Dataset to Train the YOLOv8 Model: Using the COCO Dataset to Train the YOLOv8 Model: The COCO dataset, which includes a large collection of labelled photos from 80 object categories, is used in the development of the VisionSense system.
[9]The preprocessed dataset is separated into subsets for testing, validation, and training. With its state- of-the-art performance and anchor-free recognition features, YOLOv8 is trained to identify important things that visually impaired people frequently come across. Theseconsist of people, automobiles, signs, and domestic objects. Transfer learning from pretrained weights improves the model’s learn- ing, increasing detection precision and reducing false positives in practical situations.[4]

5.2. Object Localization through Bounding Box Prediction: Upon detecting objects, VisionSense uses YOLOv8’s bound- ing box predictions to determine the spatial location of each object within a video frame. [5]Although these visual markers are not seen by the user, they play a vital role in calculating the direction — left, right, or center — which is used to generate audio guidance.

Figure 2 illustrates how YOLOv8’s anchor-free bounding box predictions enable more precise localization, even in complex environments with overlapping or small-scale objects, making the system more dependable in crowded or unfamiliar settings.

5.3. Real-Time Detection and Classification in Live Video Feed: VisionSense is designed for real-time performance us- ing YOLOv8’s efficient architecture. The system processes a continuous video stream from the user’s camera or wearable device using [7] OpenCV integration, detecting and classifying objects within milliseconds. YOLOv8 ensures improved accu-racy and reduced latency compared to earlier YOLO versions, which is essential for immediate feedback. This functionality empowers visually impaired users to identify and respond to dynamic obstacles in real time, ensuring both mobility and safety.

5.4. Voice Feedback Using Google Text-to-Speech (GTTS): After identifying and classifying an object, VisionSense uses the Google Text-to-Speech (GTTS) engine to convert the textual labels and directional information into clear voice alerts. For instance, if a bicycle is detected to the right, the system outputs “Bicycle on your right.” These voice prompts[2]

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

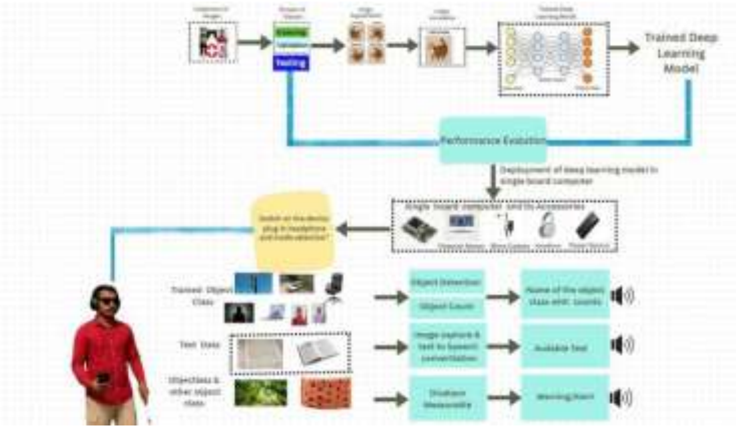

Fig. 1. Object detection using Bounding Box Prediction

supporting a diverse user's base. Using combined language translation API's (Application programming interface), object labels and also directional query's can be automatically trans- lated into the user's selected language. [6] This personalized audio feedback improves accessibility and makes the system is more effective and efficient across different languages and cultural regions. Users can customize the preferred language through an API or system settings, making Vision Sense an exclusive navigation tool for visually impaired[14][15] individuals throughout the world.

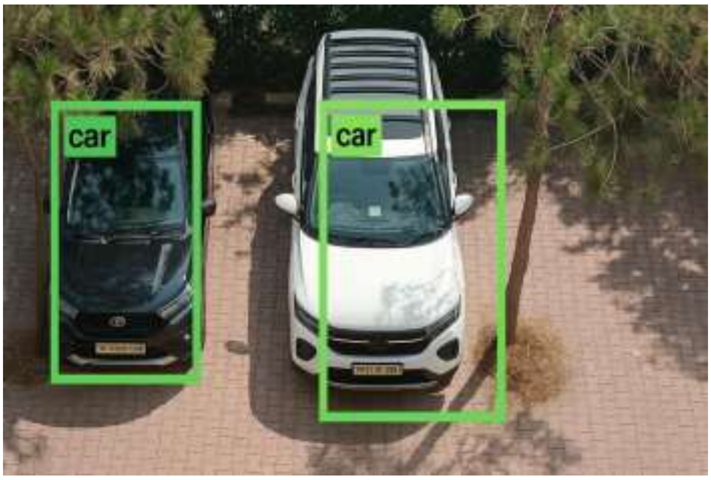

Fig. 2. Object detection using Bounding Box Prediction

are delivered through headphones or speakers, providing non- visual environmental awareness. This hands-free audio guid- ance is key to allowing users to navigate independently and confidently without reliance on physical feedback.[4]

5.5. Multi-language Audio Support for Enhanced Accessi- ability: Vision Sense is designed to support multiple languages,

Fig. 3. Vision Sense Object Detection

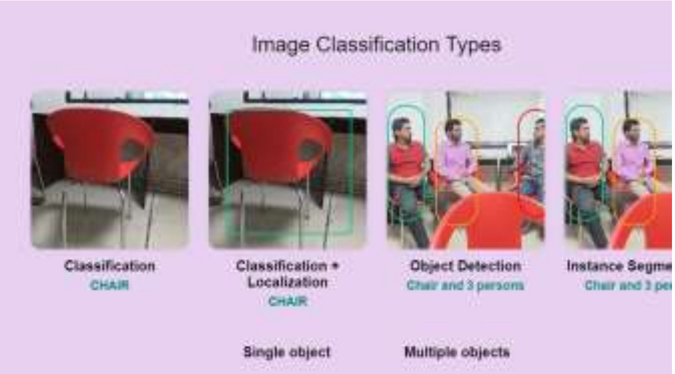

Fig. 4. YOLOv8 Real-Time Detection Demo

RESULTS AND DISCUSSION

6.1. The VisionSense system is processed based on object detection accuracy, real-time performance, voice feedback quality, and also the user accessibility. The COCO dataset was used to train the YOLOv8 model, which acheives an mean Average Precision (mAP) of approximately 88 percent, Ensuring high accuracy in identifying and localizing various objects such as people, vehicles, signboards, and furniture. [6] This precision was consistent even in cluttered environments. The system sustained a mean frame rate of 25–30 FPS in real time video processing, which is essential for live navigation assistance. The combination of OpenCV ensured smooth de- tection and processing without notifiable delays.[8]

6.2. Voice feedback was generated using Google Text-to- Speech (GTTS), and it was performed reliably, providing clear and timely alerts. The voice outputs includes both the object label and its direction relative to the user (e.g., ”Bus on your left”), which users find it easy to understand. [7]. The system’s language adaptability is also tested successfully in English, Hindi, and Tamil, with accurate translations and natural-sounding speech synthesis. This feature significantly increased user comfort and broadened the scope of the system for multi language audiences.

6.3. During user testing which is involved in visually impaired individuals, participants reported that Vision Sense enhanced their situational awareness and confidence while navigating unknown areas. They appreciated the clarity of the audio queries and the speed of detection. Whereas, a few limitations were occured, such as reduced accuracy in low-light environments and difficulty detecting transparent or reflective objects. [8]These insights point toward possible future enhancements, including the integration of infrared sensors and depth estimation for obstacle detection. Overall, Vision Sense demonstrated strong potential as an intelligent, user- friendly assistive tool that enables independence and safety for visually impaired[14] individuals.

Conclusion of Discussion: Vision Sense presents a robust, real time assistive solution that effectively fills the gap between object recognition and audio feedback for visually impaired users.[10] The combination of YOLOv8 with customizable lan- guage output sets a new benchmark in lowcost, intelligent navigation Mobility, independence, and quality of life could all be significantly enhanced by VisionSense.

VI. CONCLUSION

An important improvement in the creation of intelligent

assistive technology for the blind and visually impaired is made by VisionSense. Using the YOLOv8 object detection model’s funcionalities in combination with real-time video processing and multi language audio feedback, the system provides a workable way to improve navigation and situational awareness. The Vision sense model is meant for real world objects because of its high detection accuracy, which helps in speed and recognition of a variety of objects.

The system is made accessible and flexible for users from different backgrounds by combining Google Text-to-Speech (GTTS) for voice alerts with multilingual support, ensures that the system is both accessible and adaptable to users from diverse backgrounds. User feedback during initial testing indi- cates that VisionSense improves confidence and independence among visually impaired individuals, enabling individuals to engage with their surroundings in a safer and more efficient manner.

Limitations like low light performance and obstacle distance estimation present chances for future enhancement, despite the system's strong performance under typical circumstances. Its usability might be further improved by combining more functions like depth sensors, night vision, and adjustable voice settings. To sum up, VisionSense is a promising, reasonably priced, and expandable instrument that makes a significant contribution to the field of assistive technology and gives visually impaired people more autonomy and freedom.arXiv:2501.07957

\

Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

REFERENCES

[1] Tokmurziyev, I., Hu, Z., Kumar, A., and Islam, M. (2025). LLM- Glasses: Enabling Visually Impaired Navigation with Haptic Feedback, YOLO- World, and GPT-4o. arXiv preprint. arXiv:2503.16475

[2] Chinni, N. P. K., et al. (2024). Vision Sense: Real-Time Object Detection and Audio Feedback System for Visually Impaired Individuals. Proceed- ings of the 2nd World Conference on Communication and Computing (WCONF), IEEE.

[3] Pydala, B., Shravani, A., and Kumar, V. (2024). VisiSense: IoT- Based Assistive Technology System for Enhanced Navigation Support. Studies in Informatics and Control, 33(1). DOI: 10.24846/v33i1y2024.1

[4] Brilli, D. D., Michail, E., and Tzovaras, D. (2024). AIris: An AI- Powered Wearable for Visually Impaired Users. arXiv preprint. arXiv:2405.07606

[5] Divya, P. (2024). AI-Based Assistance System for Visually Impaired: Object Detection and Smart Navigation. International Journal of Ad- vanced Research in Engineering, Science and Management, 10(4).

[6] Jocher, G., Chaurasia, A., Qiu, J., and Stoken, A. (2023). YOLOv8: Next-Generation Object Detection. Ultralytics GitHub Repository. https://github.com/ultralytics/ultralytics

[7] Google Cloud. (2024). Text-to-Speech API (gTTS). https://cloud.google.com/text-to-speech

[8] Alba, E., Canas, J. M., Cordero, J. M., and Chicano, J. F. (2023). A Wearable Assistant for Blind People Based on Artificial Vision. Sensors, 23(1), 327. DOI: 10.3390/s23010327

[9] Suryachoudhuri003. (2023). VisionVerse: Real-Time Object De- tection Web App with Audio Assistance. GitHub Repository. https://github.com/Suryachoudhuri003/VisionVerse

[10] KIMS Foundation. (2024). AI-Powered Smart Glasses for Visually Impaired. https://geo24news.com/science-tech/2024/11/21/ai- powered- smart-glasses

[11] adhav, A., Cao, J., Shetty, A., et al. (2025). AI Guide Dog: Egocentric Path Prediction on Smartphone. arXiv preprint.

[12] Sethuraman, S. C., Tadkapally, G. R., Mohanty, S. P., et al. (2023). Mag- icEye: An Intelligent Wearable Towards Independent Living of Visually Impaired. arXiv preprint. arXiv:2303.13863 Brilli, D. D., Georgaras, E.,

Tsilivaki, S., et al. (2024). AIris: An AI-Powered Wearable Assistive Device for the Visually Impaired. arXiv preprint. arXiv:2405.07606

[13] Shaikh, S., Karale, V., and Tawde, G. (2020). Assistive Object Recognition System for Visually Impaired. International Journal of Engineering Research and Technology (IJERT), 9(9).
DOI:10.17577/IJERTV9IS090382

[14] Chavan, K., Balaji, K., Barigidad, S., and Chiluveru, S. R. (2025). Vo- calEyes: Enhancing Environmental Perception for the Visually Impaired through Vision-Language Models and Distance-Aware Object Detection. arXiv preprint. arXiv:2503.16488

[15] Madhavan, S. S., Yukendiran, M., Jeshua, A. B. J., and Sanjay, R. (2024). Sense Wear: Empowering Accessibility for the Visually Impaired. Jour- nal of Modern Science and Digitalisationn, 11(2).