



Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming
and Machine Learning

GOD-EYE: AN AI-POWERED DEPTH-BASED ASSISTIVE VISION SYSTEM FOR THE VISUALLY IMPAIRED

Rohit Khandelwal

Student of Computer Studies and Emerging Technology,
TransStadia University, Ahmedabad

Email: baljeet.khandelwal316@gmail.com

Manthan Khopkar

Assistant Professor, School of Computer Studies and Emerging Technology,
TransStadia University, Ahmedabad

Email: manthankhopkar@gmail.com

DECLARATION: I AS AN AUTHOR OF THIS PAPER /ARTICLE, HERE BY DECLARE THAT THE PAPER SUBMITTED BY ME FOR PUBLICATION IN THE JOURNAL IS COMPLETELY MY OWN GENUINE PAPER. IF ANY ISSUE REGARDING COPYRIGHT/PATENT/OTHER REAL AUTHOR ARISES, THE PUBLISHER WILL NOT BE LEGALLY RESPONSIBLE. IF ANY OF SUCH MATTERS OCCUR PUBLISHER MAY REMOVE MY CONTENT FROM THE JOURNAL WEBSITE. FOR THE REASON OF CONTENT AMENDMENT /OR ANY TECHNICAL ISSUE WITH NO VISIBILITY ON WEBSITE /UPDATES, I HAVE RESUBMITTED THIS PAPER FOR THE PUBLICATION.FOR ANY PUBLICATION MATTERS OR ANY INFORMATION INTENTIONALLY HIDDEN BY ME OR OTHERWISE, I SHALL BE LEGALLY RESPONSIBLE. (COMPLETE DECLARATION OF THE AUTHOR AT THE LAST PAGE OF THIS PAPER/ARTICLE

Abstract

Visual impairment greatly impacts the ability of an individual to see the objects, distances, and possible hazards around them, thus affecting their independent mobility. Despite the availability of several assistive technologies, most of the current assistive technologies are either not capable of accurately identifying the objects around them or cannot accurately perceive the depth of the objects around them, thus not offering the necessary navigation guidance. In this regard, this paper presents GOD-EYE, an AI-based depth perception assistive vision system that is meant to serve as the digital eyes of the visually impaired individuals. The proposed GOD-EYE system utilizes the combination of real-time object detection and monocular depth estimation, thus allowing the system to identify the objects around them, their distances, their directions, and their navigation guidance. The GOD-EYE system is developed with the ability to address the major challenges that most of the current assistive technologies are facing, including inaccurate distance estimation, scene understanding, and navigation guidance.

Keywords: Computer Vision, Assistive Technology, Depth Estimation, Object Detection, Visually Impaired, Artificial Intelligence



1. Introduction

Millions of people around the world have trouble seeing, which makes it hard for them to get around and interact with the world around them. White canes and guide dogs are common mobility aids, but they don't always help people become aware of their surroundings, especially in situations that are complicated or changing. AI-based assistive systems have become promising ways to help blind people see better thanks to improvements in computer vision and deep learning [6], [7].

Recent advances in object detection and depth estimation have made it easier for machines to understand visual scenes. Object detection models like YOLO have shown that they can quickly and accurately find multiple objects in real time [1], [10]. Monocular depth estimation methods like MiDaS also let you guess the depth of an object using just one RGB image, without the need for special hardware [2].

But most current assistive vision systems only focus on recognizing objects and don't give accurate distance estimates or useful navigation help.

This paper presents GOD-EYE, a depth-based assistive vision system that integrates object detection, depth perception, and intelligent voice feedback to navigate visually impaired users safely and effectively.

2. Problem Statement

Even though there are many types of assistive technology, people who are blind or have low vision still have a hard time getting around safely and on their own. White canes and other traditional tools only give you a little bit of tactile feedback and can't find things that are out of reach.

A lot of current vision-based systems are mostly good at finding objects, but they don't give accurate distance estimates, which is important for figuring out how close an obstacle is. Users can't properly judge the risks in their environment without depth information. Also, some solutions depend on visual or text feedback, which isn't good for people who are blind.



Another big problem is that there is no directional guidance. Most systems don't make it clear if obstacles are on the left, right, or directly ahead, which makes it hard to decide how to get around.

Some assistive technologies also need to be connected to the internet or have special hardware, which makes them less useful for everyday use.

So, we need a real-time, depth-aware, and voice-guided assistive system that can work well with just one camera and give visually impaired users useful, ongoing navigation help.

3. Literature Review

Numerous research initiatives have investigated assistive vision systems for individuals with visual impairments. The first methods mostly used convolutional neural networks (CNNs) to find objects, which let systems find common barriers like people, cars, and furniture [6]. But these systems often only used bounding-box detection, which made it hard for them to figure out how far apart objects were and how they were related to each other in space.

Recent improvements in object detection algorithms, especially the YOLO (You Only Look Once) family, have made real-time detection much better. YOLO-based models can quickly and accurately find many objects at once, which makes them good for real-time assistive applications [1], [9]. Faster R-CNN and other similar architectures have also shown to be very good at recognizing objects.

Another important part of assistive navigation systems is figuring out how deep something is. To figure out how deep something is, traditional methods use stereo cameras or LiDAR sensors. However, these methods make the hardware more complicated and expensive. Recent studies have developed monocular depth estimation methods employing deep neural networks to deduce depth from a solitary RGB image. Models like MiDaS have shown that they can make relative depth maps work well in a wide range of settings [2].

Researchers have also investigated wearable assistive technologies for people with vision problems, in addition to object detection and depth estimation. These systems usually have



cameras, sensors, and audio feedback systems that work together to make the environment aware [3], [4]. Many of these solutions work well, but they need special hardware or complicated calibration steps, which makes them hard to get and hard to scale.

Other research has looked at semantic scene understanding and segmentation methods to help people be more aware of their surroundings [5], [7]. These methods help systems find structural parts like walls, floors, and pathways, which makes navigation help even better. Datasets like Pascal VOC have also helped set standards for how well object detection works in real-world situations [9].

Even with these improvements, most current systems still handle object detection and depth estimation separately. Not much has been done to combine both features into a single, real-time assistive navigation system with smart audio feedback. The GOD-EYE system fills this gap by bringing together object detection, monocular depth estimation, and voice-based navigation into a single, lightweight, and affordable framework that can be used in the real world.

4. System Architecture

The GOD-EYE system is built in a modular way so that it can help with navigation and sensing the environment in real time. There are several parts of the architecture that work together to get images, find objects, estimate depth, make navigation logic, and make audio feedback.

The first step in the system is to get real-time video frames from one RGB camera. A deep learning-based object detection model like YOLO processes these frames and finds objects that are relevant to the environment [1]. At the same time, a monocular depth estimation model like MiDaS makes a depth map that shows how far apart things are in the scene [2].

Then, the detected objects are mapped with depth values that show how far away they are from the user. This information is analysed by the navigation module in order to find out which obstacles are present in relation to your position and whether they're categorized as left, centre



or right. Creating alerts from the system based on how urgent they were to heed depending on how close objects were to meeting each other, and their potential threat level.

Lastly, the navigation instructions are delivered via a Text to Speech (TTS) engine, giving audio feedback to the user on an ongoing basis. This architecture enables the system to achieve real-time operation with a single camera without relying on any specialized sensors.

5. Methodology

The modular computer vision pipeline and proposed GOD-EYE system include the following steps:

1. **Image Acquisition:** Capturing streaming video frames using 1 camera.
2. **Object Detection:** Deep learning-based object detection models like YOLOv8 [1] can detect many objects present in the scene.
3. **Depth Estimation:** Monocular depth estimation models like MiDaS [2] create a depth map for each frame.
4. **Fusion Logic:** The mapped depth values associated with detected objects are used to compute the change in distance and direction.
5. **Navigation Intelligence:** The potential hazards are assessed and prioritized based on their closeness and their type.
6. **Audio Feedback:** Continuous and clear voice direction via text-to-speech conversion.

The system also creates an understanding of the environment by fusing object detection with depth estimation, and streamlining practical, real-time scene comprehension and navigation assistance, without the need for specialized sensors. The system also helps users to navigate and avoid obstacles by analysing objects based on their distance to the user and alerting them when the obstacles are the closest.

6. Implementation

The construction of the GOD-EYE system was accomplished using the Python programming language in conjunction with Libraries of Computer Vision and Deep Learning for the purpose



of facilitating real-time understanding of the environment and for aiding in navigation. The system adopts a modular design that consolidates audio processing modules with modules for object detection, depth analysis, and navigational system logic.

The system utilizes OpenCV for the purpose of real-time video frame gleaning and preprocessing from a standard RGB camera. For object detection, the system utilizes the Rapidly evolving Capable Neural Network (YOLOv8) model, which rapidly and accurately detects multiple objects in each frame [1], [10].

The model detects people, vehicles, and obstructions, and provides the bounding box coordinates for each detected object. The system utilizes the MiDaS model [2] for the purpose of generating a single RGB image-based relative depth map to a model for Pseudo 3D visualization of the image.

The depth maps generated by the MiDaS model enables the system to evaluate the approximate distance to a detected object, thus eliminating the need for supplemental hardware. The system utilizes a fusion mechanism to map bounding boxes to depth values. Through this mechanism, the system evaluates the type and distance of objects in the environment.

The obstacle navigation module segments the frame into left, centre, and right areas and identifies the direction of obstacles. For the given object and its corresponding estimated distance, the system focuses on nearby potential threats and creates relevant navigation suggestions.

The last step involves the use of a text-to-speech engine that translates the navigation details into audio, giving blind users constant real-time voice assistance. The system is intended to work with a single camera, making it ideal for real-world use.

7. System Workflow

The GOD-EYE system consists of several workflows, including video capture and audio-based navigation instructions, which happen simultaneously to provide real-time support to blind users.



The first step begins with the capture of video frames by the one and only RGB camera. Each frame is processed, and the results are sent to the object detection module, where others, such as bounding boxes and class scores, are compiled by a deep learning model.

The object detection module then sends the data to the depth estimation module, which provides a depth map of the scene. The depth map indicates how far or close an object is from the camera. Once a depth map is generated, the system can locate the region of the object. The navigation decision module then analyses the available positive and negative depth and positional attributes.

For deciding which way to direct the user, a frame is divided into several vertical sections, such as left, centre, and right. Objects that are as close as possible to each user are highly prioritized, especially if they are right in the centre of the frame region. Based on this analysis, the system generates navigation instructions that include object identification, direction, and proximity. For example, the system may alert the user with messages such as “Person ahead, move slightly left” or “Obstacle on right, maintain distance.”

Lastly, the text-to-speech engine synthesizes the instructions and delivers them via an audio interface. This ongoing feedback loop enables the user to maintain situational awareness and safely navigate through changing environments. The entire workflow functions in a loop to ensure that fresh frames are processed and guidance in frames are updated without delays.

The system's real-time processing capability serves for practicality and effectiveness in daily navigation tasks.

8. Discussion

By combining these two technologies, the GOD-EYE system can not only detect objects but can also infer how close are they to the user. This is indispensable for safe navigation.

The deep learning models used in our system can simultaneously detect different objects and adapt to varying environments. Furthermore, because only one camera is employed, we have



controlled hardware costs more easily than other approaches such as solutions that need more special sensors.

However, there are still many obstacles to be overcome. For example, poor lighting, impurities resulting from the motion of the camera, and scenes with heavy crowd all exert an influence on the precision of the detection or on how accurate the depths are estimated. In addition, high-speed objects may cause real-time decision delay.

Yet despite these constraints, the system shows great potential for practical application. With further optimization and enhancement of AI models, the system's performance and reliability will be greatly improved.

9. Proposed Solution

The proposed GOD-EYE system is an intelligent assistive vision solution built for people with a visual impairment. With the aid of its digital eyes, users can recognize face familiar persons and read signs in street/road name written Chinese. The system combines real-time object detection, depth estimation, and voice-based navigation to provide a compelling view of your environment.

The system can detect many objects in indoor or outdoor installations such as distance between them with monocular depth estimation. By telling whether obstacles are at the left, right or ahead and providing some guidance that way, he receives information to keep him from bumping into things.

A continuous voice-based feedback system is the result of one key feature in the system. This allows users to receive real-time navigational information without depending on visible interfaces. The system works offline and needs only one camera, making it very practical for daily use.

By combining these features, GOD-EYE offers a complete navigation solution that improves safety, independence, and situational awareness for visually impaired individuals.



10. Experimental Evaluation

To evaluate the performance of the GOD-EYE system, several experiments were conducted using real-time video inputs in both indoor and outdoor settings. The evaluation focused on object detection accuracy, depth estimation reliability, and real-time navigation performance.

The YOLOv8 object detection model showed high accuracy in detecting common objects like people, chairs, doors, vehicles, and obstacles. The model achieved real-time speeds that are suitable for assistive navigation applications.

Depth estimation with the MiDaS model produced reliable relative depth maps, which helped the system estimate how close objects were. By combining detection results with depth maps, the system effectively determined the approximate distance and direction of obstacles.

The navigation module focused on nearby obstacles and generated timely audio warnings. During testing, the system provided continuous voice guidance with minimal delay, allowing users to react quickly to potential dangers.

Overall, the experimental results suggest that the GOD-EYE system can perform well in real-time environments and offer valuable navigation help for visually impaired users.

11. Expected Results

The GOD-EYE system is expected to provide real-time detection of multiple objects in different indoor and outdoor settings. By combining object detection with depth estimation, the system can figure out the approximate distance of detected objects and identify their direction relative to the user.

The system is designed to offer continuous audio guidance, notifying users about nearby obstacles and giving navigation instructions. Preliminary tests show that the system can successfully detect common objects like people, chairs, doors, and vehicles while estimating their relative distance using monocular depth estimation.



Using AI-based perception allows the system to improve situational awareness for visually impaired users and support safe navigation in changing environments.

Module	Technology Used	Purpose
Object Detection	YOLOv8	Detect obstacles
Depth Estimation	MiDaS	Estimate distance
Navigation Logic	Python Algorithms	Determine direction
Speech Output	pyttsx3	Voice guidance

12. Conclusion

The research paper introduced an AI-based navigation assistance system which helps blind users to navigate different environments while enhancing their spatial awareness. The system uses real-time object detection technology together with monocular depth estimation to create a complete environmental picture.

The system combines computer vision methods with smart audio feedback to help users recognize objects and measure their distance while getting real-time navigation advice. The proposed system functions with basic RGB camera equipment because it requires only one camera which makes it more affordable than other solutions that need stereo cameras and LiDAR sensors.

The system achieves effective navigation support through its combined use of object detection and depth estimation technology. The system demonstrates its capabilities through experimental testing which shows it can detect multiple objects while measuring their distance and providing voice instructions with minimal delay.



The GOD-EYE system demonstrates how AI-powered assistive technology can help blind users achieve greater independence and safety while improving their overall living experience. The proposed approach provides a practical, scalable, and cost-effective solution for real-world deployment.

13. Future Work

The upcoming upgrades to GOD-EYE system will create better operational performance and user experience. The system needs to enhance its scene recognition abilities because it currently handles basic outdoor environments while failing to identify advanced elements such as stairs and traffic signals and crosswalks and various road surfaces.

The system can also be extended through the integration of additional technologies which include GPS-based navigation for outdoor mobility and route guidance purposes.

The accuracy and reliability of the system will improve through the implementation of sensor fusion techniques which combine camera input with smartphone sensors and wearable devices. The first goal requires researchers to develop methods for better depth estimation while they need to enhance systems performance for usage on mobile devices.

The system will improve its performance in difficult conditions through the development of advanced deep learning models which handle low-light and crowded situations. The research will develop a natural language interaction system which allows users to control the system through voice commands using conversational speech.

The system requires this feature to become easier for users to operate because it enhances their understanding of system functions. The GOD-EYE system will develop into an intelligent assistive companion which provides complete navigation assistance to visually impaired users after these enhancements.

REFERENCES

[1] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,



2016. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf
- [2] R. Ranftl et al., “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9178977>
- [3] A. Aladrén et al., “Visual-based Navigation Assistance for the Visually Impaired,” *IEEE Transactions on Intelligent Transportation Systems*, 2016. <https://webdiis.unizar.es/~glopez/papers/AladrenISJ2016.pdf>
- [4] N. Silberman et al., “Indoor Segmentation and Support Inference from RGBD Images,” *European Conference on Computer Vision (ECCV)*, 2012. https://cs.nyu.edu/~fergus/datasets/indoor_seg_support.pdf
- [5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf
- [6] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE CVPR*, 2015. https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf
- [7] M. Everingham et al., “The Pascal Visual Object Classes Challenge,” *International Journal of Computer Vision*, 2010. https://www.researchgate.net/publication/220659463_The_Pascal_Visual_Object_Classes_VOC_challenge
- [8] D. Hoiem, A. Efros, and M. Hebert, “Recovering Surface Layout from an Image,” *International Journal of Computer Vision*, 2007. https://dhoiem.cs.illinois.edu/publications/hoiem_ijcv2007SurfaceLayout.pdf
- [9] A. Bochkovskiy, C. Wang, and H. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv*, 2020. https://www.researchgate.net/publication/340883401_YOLOv4_Optimal_Speed_and_Accuracy_of_Object_Detection
- [10] R. Girshick et al., “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” *IEEE CVPR*, 2014. https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf



Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming
and Machine Learning

Author's Declaration

As an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my hole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible, shuffled, or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who find trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher, then my paper may be rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds any complication or error or anything hidden or implemented otherwise, my paper may be removed from the website, or the watermark of remark/actuality may be mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Rohit Khandelwal
Manthan Khopkar
