Enhancing Accessibility and Autonomy for Visually Impaired Individuals Using AEyeD: An Assistive System

Nehal Khan
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai, India
nehalmicro29@gmail.com

M Diviya
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai, India
diviyalouis@gmail.com

Kaushik Kumbhat
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai, India
kaushikkumbhat54@gmail.com

Rohit Prasanna
Department of Computer Science and Engineering
Amrita School of Computing
Amrita Vishwa Vidyapeetham
Chennai, India
therohitprasanna@gmail.com

***Abstract*—In this technological era, visually impaired individuals still face significant challenges in their daily lives, and still seek other's dependence. The Artificial Eye Decipher (AEyeD), aimed at increasing user independence and safety, combines face and gesture recognition; detection of vehicles and stairs; and a text-to-speech conversion feature. AEyeD implements Convolutional Neural Networks (CNN) for face and gesture recognition, and object detection. The system recognizes approaching individuals, interprets gestures like handshakes, and identifies offered objects. Moreover, AEyeD identifies stairways and stationary automobiles, providing audio feedback via the 'pyttsx3' module for safer navigation. Image-based textual extraction and reading on demand are made possible by means of the 'pytesseract' module. With the help of test trials with patients suffering from Retinitis Pigmentosa (RP) in different scenarios on each task done by AEyeD, highlighting its potential to change the daily lives of visually impaired people by granting them the ability to navigate and engage with their environment independently.**

Keywords—AEyeD, visually impaired, RP, assistive system, user independence, face recognition, gesture recognition, object detection, navigation assistance, text-to-speech.

I. INTRODUCTION

In today's swiftly evolving technological landscape, innovations are emerging to better the independence and quality of life for a visually impaired person. Innovations have altered people's perceptions and comprehension regarding their space that is caused by visual impairment. This paper introduces AEyeD as state-of-the-art assistive systems devoted to enhancing the autonomy and situational awareness of visually impaired users.

AEyeD contains some of the modern computer vision and machine learning that enables a wide range of assistive functionalities. Real-time face and gesture recognition, along with object detection [8] using Convolutional Neural networks, are likely to make up the major component of the system, allowing people to know who is coming towards them, interpret day-to-day gestures such as handshakes, and detect objects presented to them.

Besides socialisation, other capacities of AEyeD that can be used to make the user safe and navigate include object detection around the surroundings of the user. It will show the existence and direction of staircases and the parked status of nearby vehicles, among others. This will be communicated to the user through a prompted auditory cue using the 'pyttsx3' text-to-speech module end. Moreover, the system comes with inbuilt OCR functionality employing 'pytesseract' to allow users to take advantage of everything that is written in the surroundings through transposing it to audible sound upon request and therefore increasing its users' information input.

Our research investigates how the AEyeD framework has been created, put into place, and tested in practice. We assess system performance in different scenarios and tasks through trials conducted with people suffering from Retinitis Pigmentosa. The findings demonstrate that AEyeD can substantially enhance the daily experiences of visually impaired people, providing them with aids for bolder and independent movement as well as communication.

II. LITRATURE SURVEY

Assistive technologies for the visually impaired have gained significant attention in recent years. Apart from the white cane conventionally used, it cannot perceive obstacles of any height or even see-through objects, and thus there is a need for technology of higher sophistication [1]. Several Electronic Travel Aids (ETAs) have been evolved to address the issues at hand. ETAs include sensor-based approaches; computer vision-based approaches, and various combinations in between the two as hybrid systems to help guide visually impaired people.

A. Sensore Based Approach

Ultrasonic sensors have widely been used in the domain of sensor-based solutions due to their cost-effectiveness and capability to detect obstacles at short ranges. For example, ultrasonic sensors are used along with wearable devices such as belts and smart canes to alert users through haptic feedback about the nearby obstacle [2][10]. Such devices are often ineffective in detecting clear objects or obstacles at high altitudes. In a bid to counter this, researchers employed LiDAR technology that gives better obstacle detection but fails to detect transparent objects [3].

B. Computer Vision-Based Approach

The latest advancements in AI and ML have recently been exploited to improve the obstacle recognition process by using computer vision-based approaches. Another such system was proposed and demonstrated the real-time delivery of obstacle detection and navigation support [6][11]. However, such kinds of systems require high computational power and energy consumption and are thus not suited for low-power wearable devices [4].

C. Hybrid Approach

These are hybrid systems combining sensor-based methods with computer vision techniques. For example, the combination of LiDAR and ultrasonic sensors in assistive devices results in improved outcomes in obstacle detection when compared to those solely based on vision, yet its computational resources were smaller in comparison to purely vision-based systems [5]. Some modern innovations in AR also offer some solutions to assist visually impaired people in their navigation through difficult environments both by providing contextual information as well as guidance through AR displays [6][13].

Despite such progressions, existing solutions have problems in terms of high costs, low adoption rates, and require further refinements to better satisfy the users. Most commercially accessible devices like the eSight and IrisVision prove to be out of reach for most people because it's expensive. Others like Orcam are also inaccessible because they sell it at a very high price [7]. However, to-date research into low-cost light and energy-saving solutions has not been done toward meeting the growing population of the visually impaired.

Summary Much more dramatic research will be needed to really make provision for a low-cost, reliable, and user-friendly device, considering some breaks in the envisioning of assistive technologies for the visually impaired. What the quality of life for people with impairments to their sight will require from future research studies is strong, low-cost solutions capable of giving real-time feedback with further integrated developments of AI, ML, and sensor technology [12].

III. METHEODOLOGY

A. Hardware Implementation

The hardware design of the AEyeD device as shown in Fig. 1 developed in the research integrates a camera, earphone, and a microprocessor, aimed at providing reliable real-time face and gesture recognition and object detection [8] for visually impaired users. The Raspberry Pi represents the central processing unit that governs the data that the camera gathered and initiates appropriate responses according to the input after it has gone through some process.

The camera, mounted on the user's glasses, serves as the primary input device for capturing the surroundings [9]. This visual data is transmitted to the Raspberry Pi, where it is processed according to the user-selected mode. The device operates in three primary modes: Rename Mode, Text Analysis Mode, and Recognition Mode.

Rename Mode allows users to register and learn the new faces that they encounter. Once the face is unknown to any user, this mode can be activated, and then the face can be renamed using a system-assigned unique ID number. The device, through the OS library, downloads the image of the identified person from the set folder. Then the user announces, aloud, the name to attach to the identified person. The machine performs the voice-to-text operation; it then stores the photograph in the "recognised" folder under that name. Then, the next time that person comes to the house, the machine identifies the person and reminds the owner that such an individual is at the house.

In Text Analysis Mode, the device can scan print or e-text. The system works as such by capturing the text the user is focused on and, through 'pytesseract', rendering it to give audio feedback. The system read the content aloud very effectively for the user. Practicality is enhanced while improving accessibility with this feature and, thus, makes the consumption of written content by visually impaired users easy and smooth.

In Recognition Mode, the primary function of the device is to recognise faces and provide audio feedback to the user. The system generates a unique ID number for an unknown face that approaches, which can be renamed by the user for future identifications. Along with face recognition, this mode also detects various objects or obstacles in front of the user and warns them about it on time, thereby enhancing safety. Moreover, Recognition Mode can recognise gestures too. It detects usual gestures like handshakes and can tell if someone is passing something or a gift to another person so as to inform them accordingly. Thus, it integrates face recognition with situational awareness, enhancing the user's ability to interact with their surroundings effectively.

In general, the AEyeD is conceived as an intuitive aid, portable, and functionally aimed at the sighted user with processing capability to deliver real-time situational awareness and interaction.

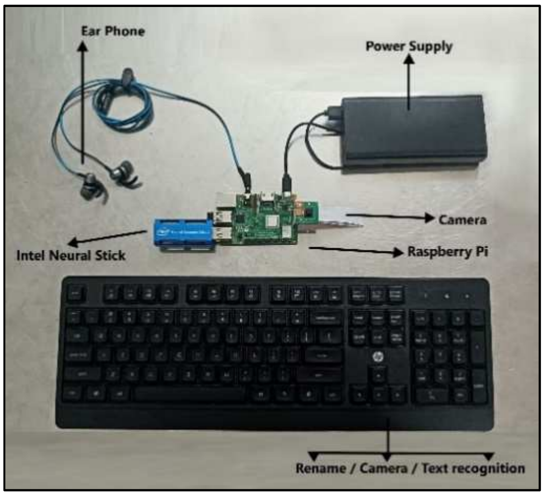

Fig. 1. AEyeD hardware design

B. Software Implementation

AEyeD uses a camera mounted through the wearer's glasses to capture live video using OpenCV. As shown in the Fig. 2

the image is fed into an object-detection model created using Keras to look out for predefined objects like cars or stairs and reports it to the user immediately upon recognition of such an object [8]. The user is notified voice with a module called 'pyttsx3, which is a text-to-speech module. If no such objects have been recognized, the image is converted to greyscale to increase further ease in processing.

To detect face, AEyeD used a pre-trained Haar Cascade Classifier with the file haarcascade_frontalface_default.xml, which only detects within particular limits of height and width. It crops the detected face temporarily before saving it. The FaceNet model was actually a Convolutional Neural Network whose output is numeric vectors of the cropped image. Comparing these vectors with any images in a folder has known faces. It declares the actual name if it finds one; otherwise, it stores it in some unknown folder.

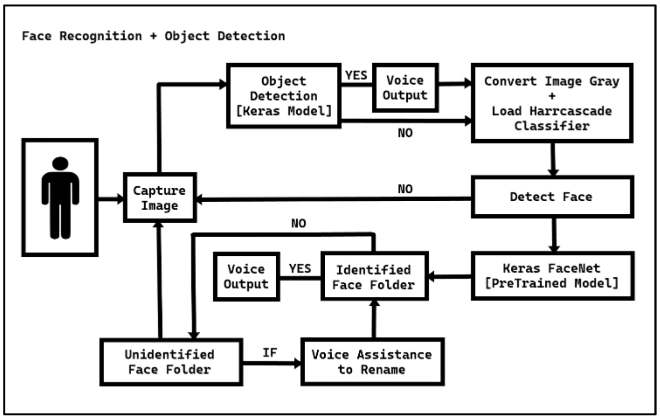

Fig. 2. AEyeD architecture diagram for face/object recognition

The face of any unknown user can be renamed using voice command, which tells the unique ID number given to the face and mentions a new name. But, in Text Recognition Mode, the user can toggle over two modes to image capture and scan for text within a given environment. If there is text, AEyeD will go about using the ‘pytesseract’ module to perform OCR of text-that is, converts text into speech and will read that out to the user.

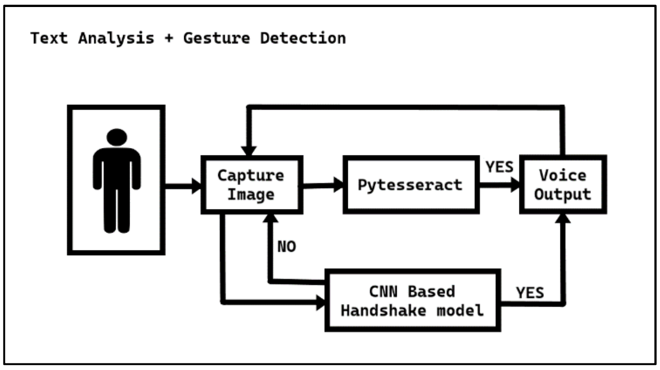

Fig. 3. AEyeD architecture diagram for text and gesture analysis

In the Text Recognition Mode as shown in Fig. 3 captures images to scan for text; on finding the text, it uses ‘pytesseract’ module to interpret and define the text to speak to the user.

IV. Results

Does visually impaired people always like to depend on others? To make them feel independent, they must get a feel of confidence in navigation and recognising things on their own. This could be possible with the help of AEyeD which enhances their social interaction capabilities by recognising the person who the user is going to talk to. This avoids confusions and any other embarrassment to the visually impaired user.

This paper includes a test that demonstrate the effective use of AEyeD on a person who is affected by Retinitis Pigmentosa which is a group of inherited eye diseases that cause progressive degeneration of the retina. The following contains the issues faced by the patient and solution for each ability of AEyeD.

A. Face Recognition

The patients who are affected by RP are generally light sensitive, causing tunnelling of vision. Due to narrowing the field of view, the user is unable to recognize faces clearly and in a wider spectrum.

Fig. 4. AEyeD captures face for face recognition

As shown in Fig. 4 AEyeD captures the field of view and recognize the known faces in it and conveys to the user. Thus, helping them to recognize people.

B. Vehicle/Object Detection

As the field of view is reduced, visually impaired people find it difficult to navigate while walking. They often bump onto parked vehicles or misstep on stairs.

Fig. 5. AEyeD captures vehicles for vehicle detection

As shown in Fig. 5 AEyeD helps them by detecting stairs, parked vehicles, their type, and where is has been parked. Thus, helping them in safer navigation.

C. Gesture Analysis

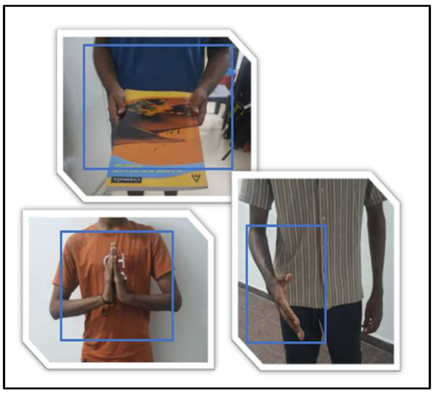

Fig. 6. AEyeD captures gestures for gesture detection

Due to reduced central vision, visually impaired people won't be able to recognize the main gestures of the person to whom their talking to, example handshake and object offering as shown in Fig. 6. AEyeD helps in analysis these gestures.

D. Text-to-Speech Conversion

Fig. 7. AEyeD captures text for audio response

To aid them in reading where texts are too small or distant AEyeD helps in capturing them and convert it to audio. Thus, helping them to read without depending on others as shown in Fig. 7.

E. Evaluation Metrics

The following graph shows the comparison of test cases passed by the patient affected by RP with no external help and with the help of AEyeD.

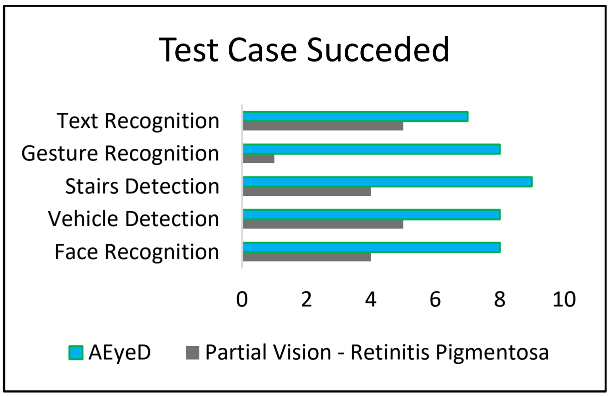

Fig. 8. Comparison graph between users and non-users of AEyeD

Based upon the test case data in graph shown in Fig. 8 the following value are attained to fill the confusion matrix

*a) **True Positive (TP):*** Occurs when the detected image is recognised correctly and classified into positive class. In this problem statement, when a face/object/gesture/text is recognised correctly then TP occurs.

*b) **True Negative (TN):*** Occurs when the detected image is recognised correctly and classified into negative class. In this problem statement, there is no TN as there is no need of negative class, example: This person is not person 1.

*c) **False Positive (FP):*** Occurs when the detected image is recognised incorrectly and classified into positive class. In this problem statement, when a face/object/gesture/text is recognised incorrectly but still recognised the type correctly then FP occurs, example: person giving handshake but recognised as high-five.

*d) **False Negative (FN):*** Occurs when the detected image is recognised incorrectly and classified into negative class. In this problem statement, when a face/object/gesture/text is recognised incorrectly and the type incorrectly then FN occurs, example: detecting face but actually there is no one.

Table 1. Confusion matrix of visually impaired using AEyeD

n = 50	Positive	Negative
Positive	40	5
Negative	5	0

Based upon sample group

Table 2. Confusion matrix of visually impaired without external help

n = 50	Positive	Negative
Positive	19	16
Negative	15	0

Based upon sample group

To evaluate the performance of the classification model, the following metrics are utilised on both AEyeD as show in Table 1 and without external help as shown in Table 2.

*a) **Precision:*** The ratio of correctly predicted positive instances to the total predicted positive instances are shown in Table 3.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Table 3. Precision values

Precision	
AEyeD	0.8889
Without external help	0.5429

*b) **Recall:*** The ratio of correctly predicted positive instances to the total actual positive instances are shown in Table 4.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Table 4. Recall values

Recall	
AEyeD	0.8889
Without external help	0.5588

c) F1 Score: The harmonic mean of Precision and Recall, providing a single measure of a model's performance, especially useful when there is an uneven class distribution are shown in Table 5.

$$\text{F1 Score } = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Table 5. F1 Score

F1 Score	
AEyeD	0.8889
Without external help	0.5507

On the basis of F1 Score which gives importance to False Positive and False Negative that is crucial to this study, The AEyeD system achieved an F1 score of 0.8889, indicating very good performance regarding the balance between precision and recall. On the other hand, without external help the user scored 0.5507, which indicates poor or unreliable classifications. This contrast clearly reinforces how much the AEyeD system can be reliable, and its overall effectiveness.

V. Discussion

The following tells the limitations of AEyeD and its room for future improvement.

A. Limitations

a) Wired : All components from camera to headphone all are wired.This can cause discomfort for the indiviuals.

b) Facial Recognisation: Here the model detects only a complete face, meaning model can't recognise a person from side view.

c) Spoof attacks: Use a facial photograph of a user to spoof the face recognition system

d) Dim lighting conditions: Here the system fails to detect face or object in very low light environment.

B. Future Improvement

a) GPS integration: For navigation assistance, GPS can be integrated in order to receive directional support. It will guide users as they walk through cities.[6]

b) Emergency Assistance: Incorporate an emergency assistance feature that can contact emergency services if the user is in distress.

c) Multiple language support: Incorporate different language support for user's convenience.

d) Image enhancement: Incorporate Low Light Enhancement (LLE) on the captured images for better detection and recognition.

e) Comfortable glasses for the future: Build spectacles for the user's comfort and convenience as shown in Fig. 9 [14]

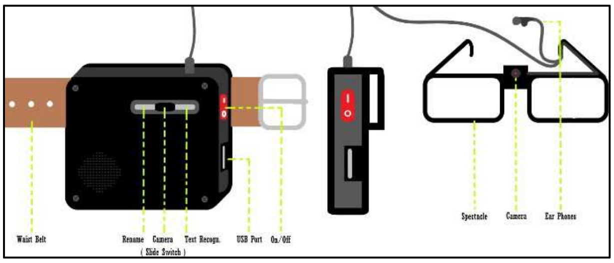

Fig 9. Future implementation of AEyeD

VI. Conclusion

It concludes that AEyeD significantly enhances the degree of independence and quality of life of the visually impaired through real-time face and gesture recognition, object detection, and text-to-speech conversion. As shown in tests with patients affected by Retinitis Pigmentosa, AEyeD can improve mobility and social interaction, providing a better situational awareness for the user and, at the same time, an enhanced communicative ability. The proposed computer-vision-enhanced system is implemented in full here. It can bridge gaps in accessibility, allowing visually impaired users to view their surroundings and lead more confident and independent lifestyles.

References

[1] S. Busaeed, R. Mehmood, I. Katib, and J. M. Corchado, "LidSonic for visually impaired: Green machine learning-based assistive smart glasses with smart app and Arduino," Electronics, vol. 11, no. 1076, pp. 1-34, Mar. 2022. doi: 10.3390/electronics11071076.

[2] Tudor, L., Barbu, D., & Neagoe, N. (2018). "Wearable Assistive Device for Blind People Navigation Based on a Fuzzy Controller," 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC), pp. 399-404. doi: 10.1109/ICSTCC.2018.8540709.

[3] R. Jafri, M. A. Khan, H. R. Arabnia, and S. A. H. Mubarak, "Computer vision-based object recognition for the visually impaired using Google Tango and deep learning," IEEE Access, vol. 7, pp. 113787-113796, 2019. doi: 10.1109/ACCESS.2019.2935444.

[4] J. Gutierrez-Gomez, F. Escobar-Molano, and J. P. Duarte-Guevara, "Human Gait Recognition Using Machine Learning Techniques and Convolutional Neural Networks," in 2019 IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton, FL, USA, 2019, pp. 432-437. doi: 10.1109/ICMLA.2019.00168.

[5] Bai, S., Chen, Z., and Tian, Y., "A Wearable Glasses System for Assisting Visually Impaired People to Avoid Obstacles," 2019 IEEE International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, pp. 4979-4985, May 2019. doi: 10.1109/ICRA.2019.8793667.

[6] A. Lee and G. Medioni, "RGB-D camera based navigation for the visually impaired," Computer Vision and Image Understanding, vol. 149, pp. 3-20, Aug. 2016. doi: 10.1016/j.cviu.2016.02.002.

[7] A. Kumar, "OrCam MyEye 2: Wearable Assistive Technology for the Visually Impaired," 2018 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sousse, Tunisia, 2018, pp. 1-5. doi: 10.1109/ATSIP.2018.8364487.

[8] N. N. Singh, P. Singh, and A. I. Choudhary, "Real-Time Object Detection and Classification System for Visually Impaired People Using Deep Learning," 2018 International Conference on Computer, Communication, and Signal Processing (ICCCSP), Chennai, India, 2018, pp. 1-6. doi: 10.1109/ICCCSP.2018.8452851.

[9] H. Y. Ha and S. D. Lee, "Wearable Infrared Sensor-Based Smart Glasses for Visual Assistance," IEEE Sensors Journal, vol. 19, no. 11, pp. 4055-4064, Jun. 2019. doi: 10.1109/JSEN.2019.2897435.

[10] S. U. A. Bukhari, H. K. Kang, and S. H. Hong, "Smart Cane Assisted Mobility for Visually Impaired Persons," IEEE Access, vol. 7, pp. 54206-54214, Apr. 2019. doi: 10.1109/ACCESS.2019.2909962.

[11] C. Bai, W. Wei, and Q. Huang, "A Navigation System for the Visually Impaired Using RGB-D Sensors and Mobile Devices," 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 2019, pp. 2291-2296. doi: 10.1109/ICMA.2019.8816248.

[12] M. Tanveer, A. Abouelenien, and M. N. Ghanem, "A Lightweight Assistive System for Navigation and Object Recognition," 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 2020, pp. 122-127. doi: 10.1109/IROS45743.2020.9341086.

[13] S. Patel, P. Kumar, and A. Khare, "A Deep Learning-Based Smart Assistant for Visually Impaired People Using Hand Gesture Recognition," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 207-213. doi: 10.1109/ICAIS50930.2021.9395784.

[14] R. A. Jadhav, V. A. Patil, and M. A. Patil, "Smart Glasses for Visually Impaired Using IoT and Deep Learning," 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2021, pp. 1-6. doi: 10.1109/ICCCI50826.2021.9402566.

[15] D. M. Chen and A. D. Aparna, "Assistive Vision Systems Using Edge Computing and Cloud Services," IEEE Transactions on Human-Machine Systems, vol. 50, no. 3, pp. 281-291, Jun. 2020. doi: 10.1109/THMS.2020.2984417.