EYE MATE AI: A Multi-Modal Vision Assistance System for Visually Impaired and Elderly People

Aisha Shaikh[1], Saba Jalgaonkar[2], Aaliya Shaikh[3] , Humera Shaikh[4], Zeba Syed[5]

[1]*Aisha Shaikh Hardware & software*
[2]*Saba Jalgaonkar Hardware & Software*
[3]*Aaliya Shaikh Hardware Coding*
[4]*Humera Shaikh Hardware Coding*
[5]*Zeba Syed, Dept. of computer Engineering, Abdul Razzak Kalsekar Polytechnic, Maharashtra, India*

Abstract - Visually impaired individuals encounter numerous challenges while attempting to read any material in print or in handwritten form during their everyday living. This can make it difficult for a visually impaired person to be able to read books, documents, medicine labels, sign boards, or notes and may necessitate making use of another individual for assistance, which diminishes their ability to be independent and typically creates additional difficulty in accomplishing their daily activities. Presently known methods for helping individuals who are blind or visually impaired include braille books and audio recordings, however, these methods may not always be available for use at the time they are needed. Other assistive methods of reading are also expensive and difficult to use. Therefore, an affordable and simple method for assisting individuals who are blind or visually impaired to read text is necessary.

Key Words: Assistive Technology, OCR, Image Processing, ESP32-CAM, Text-to-Speech, Computer Vision, Visually Impaired, Real-Time Text Recognition, OpenCV, Embedded System

1. INTRODUCTION

Many people with vision impairments have challenges reading printed or hand-written text throughout the day. Books, documents, medicine packaging, and signage can all be hard to read without the assistance of another person. As a result of not being able to read independently, those with vision impairments have to rely on others for support, which hinders their independence and makes it that much harder to accomplish daily tasks independently. Braille books and audiobook recordings can assist individuals with print disabilities, however, neither option is always available in a timely fashion. Electronic assistive devices to help people with print disabilities read can also be prohibitively expensive or difficult to use, so therefore not every person with a print disability can use these devices.

Also, some tools may not provide immediate responses, thereby making them less useful for many people every day.

EyeMate solves this problem by providing visually impaired individuals and others with print disabilities with a simple, intelligent assistive decision support system for identifying and reading text (in real-time). The EyeMate captures an image of the text using a camera module and sends the image to a computer for processing. The computer will use both Image Processing and OCR to convert the text into speech (voice output). The EyeMate can read both printed and clearly hand-written text, and therefore would provide audio output in real-time to assist visually impaired users when they need assistance reading during their everyday lives.

1.1 Challenges Faced by Visually Impaired People

All over the world, printed information is available on books, posters, product labels, packaging, and official documentation. Visually impaired individuals cannot independently access information until someone helps them. They face continual dependence upon others and a lack of confidence due to this.

Braille and other tactile systems have made it possible for many to use printed text, but not all printed materials exist in braille. Audiobooks have limited use as they are not able to provide information in real-time. As a result, visually impaired consumers continue to experience significant challenges in their daily lives.

There is a need for an easy and quick way for visually impaired consumers to read and receive printed materials around them, using information from their environment.

1.2 Limitations of Existing Assistive Devices

Many tools exist to help those who are blind or have low vision; however, they come with several drawbacks. First, different tools can be prohibitively expensive for some users. Second, other tools may not be suitable for someone who has not previously had any training/experience in operating them. Third, some can only read print and cannot interpret handwritten materials (e.g., business cards).

This is why any new technology must be developed that allows for real-time and low-cost detection of both printed and handwritten documents within one product that speaks back through a voice synthesizer. This is what EyeMate is able to accomplish.

2. System Architecture

.EyeMate assists individuals with visual impairments in reading printed and handwritten words clearly. It allows for easy communication between the camera module, processing unit, and audio output devices. The ultimate purpose of EyeMate is to scan the environment and convert the contents of the scanned materials into audio format, allowing the visually impaired user to hear information.

There are three main components that comprise the EyeMate system: (1) capturing an image, (2) processing the text within the image, and (3) outputting audio. The EyeMate system captures an image of either printed or written text with the ESP32 camera module. Once the image has been captured, it is wirelessly transmitted to a computer for processing.

The processing of the captured image uses OpenCV to enhance the image quality through greyscale conversion and removing noise. After the image has been pre-processed, OCR is employed to identify the text within the captured image. The EyeMate system can recognize both printed text and importantly it also recognizes clear handwriting.

Once the text has been detected, the text-to-speech function of the EyeMate system converts the text into an audio file which is played back over a loudspeaker, thus enabling the visually impaired user to hear the text as it is read aloud.

2.1 Hardware and Processing Interaction

The relationship between the computer system and the hardware device is straightforward with the EyeMate System. The ESP32 CAM camera takes the picture of the environment around the user and sends it via wireless data transmission to the computer processor for processing. When the user points their camera toward an object, the ESP32 CAM camera will take a photograph of the object and then send the image wirelessly to the computer. The computer functions as the main processor where image processing and text recognition takes place. Image processing will be done by the OpenCV library, which allows for improved image quality so that text can be easily detected.

Once the image has been processed, the OCR system will read the detected text from the image and send the text to the text-to-speech conversion system, where the text will be converted to speech and output through a speaker. The hardware and software interaction provides a smooth operation of the EyeMate System in real-time for users with visual impairments.

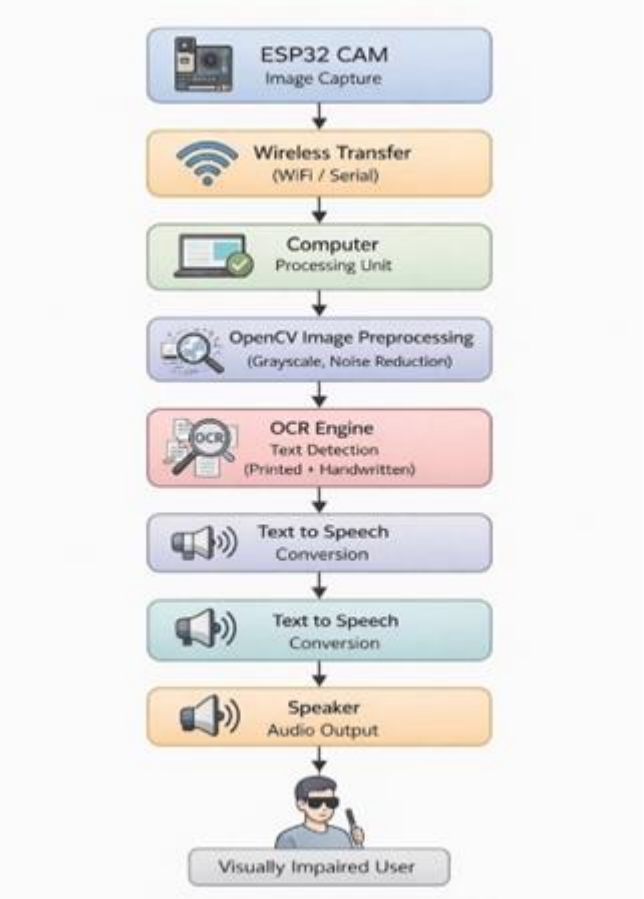

Fig -1: Working of Eye Mate

2.2 Image Processing and Text Recognition Layer

The EyeMate System consists of an Image Processing and Text Recognition Layer. At this layer, captured images are processed and converted to readable text. The image is processed and made available for text detection. The image captured may not be in good quality so the first step is at image preprocessing before detecting the text.

The first step in processing the image is to process the image using OpenCV techniques to change the original image into a grayscale image while removing i.e. unwanted noise, so the text will appear clearer. This will help in improving the overall accuracy of text detection. Once the image is preprocessed, it is utilized by the OCR Engine for recognition of characters.

After the OCR system does its job of recognizing the character in the image, the OCR converts the recognized characters to an electronic digital text file in either printed or clear handwriting. Once the text is retrieved, then it is passed to the text-to-speech module.

Finally, the text-to-speech module converts the detected text to an audio output or sound. The output audio will be played through the speaker so the user who is visually impaired can listen to the text in real-time.

3. Core modules

The Main (Core) modules are the key systems of EyeMate that allow for determining text and converting it to speech. They function collectively to collect image data, process the text, and produce aural output for those who are blind or vision impaired. The EyeMate system provides a precise and easy-to-use means of accessing wheel chair via sight by each module completing a distinctive duty.

For instance, the Image Capture module uses the ESP32 CAM to produce the image of printed or handwritten text around the EyeMate user. Next, the Image Processing module refines the captured image using OpenCV so the printed or handwritten text is displayed clearly. Finally, the OCR module reviews the text contained in the refined image and converts the text from the image to a digital text format. Next, the Text to Speech (TTS) module converts the text to be read to audio. Finally, the Audio Output module reproduces the audio version of text through the speaker to allow the user to hear their text.

The combination of the Image Capture, Image Processing, OCR, TTS, and Audio Output modules provide a quick, easy and effective way for EyeMate users to read printed or handwritten text instantaneously.

3.1 Image Capture Module

The Image Capture Module collects visual data in the surrounding environment. The ESP32 CAM can be used to collect images of printed or handwritten text using the ESP32 CAM as the image capture device. By aiming the camera at a book, label, or document, the user captures an image and wirelessly transfers it to a PC. The module is critical because the quality of the captured image is directly proportional to the accuracy of the text detection.

3.2 Image Processing Module

This module helps increase clarity of captured images in preparation for text detection through image pre-processing with OpenCV by converting the color image into a greyscale image, and removing unwanted image artifacts. The result is an improved text detection rate for each image as the text becomes more readable. Overall, the improvement of the performance of OCR systems and devices is a goal of image preprocessing.

3.3 OCR Text Recognition Module

Various features of OCR modules, such as recognition of both printed and clearly written text, can be useful in a variety of contexts, including character recognition and digitization of images. Detection accuracy is closely related to the quality of the input image and the preprocessing done prior to analysis. Once text has been detected, output from OCR is sent to an audio conversion module.

3.4 Text to Speech Module

Using special software called a text-to-speech engine, the text written out on the screen will get converted into an audio file that can be played back. Consumers that are either unable to see or have trouble reading will now have a way to hear what others have written, and the ability to experience what visually-impaired people have been able to do for decades. It also makes this system very handy for use with applications where speed is critical.

3.5 Audio Output Module

The last component of this program is the Audio Output Module. This module plays back the speech output from the program (audio) using either a set of speakers or headphones, allowing users to clearly hear whatever text has been detected by the system. This allows individuals who are visually impaired to receive information on their own without assistance and enables them to use the system effectively in their daily activities.

4, Methodology

The methodology of EyeMate describes how the system functions to determine what is printed text and convert it to speech for people with disabilities. The main function of EyeMate is to facilitate ease of reading by using a Cameras to take a snapshot of text in the environment it is located and processed in four steps; capturing the image, preprocessing (improving the quality) of the image, using OCR to determine text present within the image, and the final output from the text-to-speech to audio.

The process begins by imaging the printed or written text using a camera. Next, the computer takes the captured image and processes it to increase the quality of the image. After the image has been processed to improve its quality, OCR will run to determine the text. The text that has been identified by the computer is converted to audio for the user to listen (i.e., the user can read the data independently). This step-by-step process is designed to provide real-time reading to assist those who cannot see to read independently.

4.1 Image Capture

The captured image from an ESP32 CAM will be a still image taken with a camera as input from a user who will use this unit to take a picture of printed or handwritten text. Users will point the camera at an image text, a book, a text-printed or handwritten piece of paper, a label, a sign, or whatever is required, and the image captured will be sent via a wireless connection to a computer. Image quality must be good in order for the device to detect the text with High Accuracy.

4.2 Image Preprocessing

Preprocessing has been completed when the image has been received. The OpenCV library performs two main preprocessing functions: creating a grayscale copy of the image and reducing any noise present in the image. These two tasks result in a more viewable and readable version of the text; thus, this preprocessing increases the accuracy of the Optical Character Recognition (OCR) engine.

4.3 Text Recognition using OCR

At this stage, the utilization of Optical Character Recognition (OCR) is employed to identify characters from an image. The OCR reads the characters and converts them into digital text. The system recognizes both clear printed characters as well as clear handwriting. The recognized characters will be saved for the next step

4.4 Text to Speech Conversion

Once the text is recognized by the system, the software will utilize text-to-speech capabilities to generate sound from the recognized text allowing a visually impaired user to hear the information rather than read it.

4.5 Audio Output and User Interaction

The EyeMate process is completed by turning the identified textual information into audio, which is then delivered to the user. Once the optical character recognition (OCR) part of the EyeMate has correctly identified the text from the image it has taken. The identified text is sent to the text-to-speech processor so that it can be synthesized into clear voice output and sent out to either a speaker or headphones connected to the device.

The ability to generate and deliver the audible version of the identify text allows the visually impaired user to listen to the text rather than having to read the text, thus making the EyeMate useful in everyday situations, such as reading books, reading documents, reading labels and reading signs or billboards. The eyeMate delivers this output in real-time, allowing the user to receive the output without delay.
The overall interaction of the user with the device is simple. The user simply points the camera at the text and the EyeMate automatically captures, processes, and generates the audible output, which makes the EyeMate user-friendly and usable daily. The audio output portion of EyeMate finalizes the operation of EyeMate and allows visually impaired users greater independence.

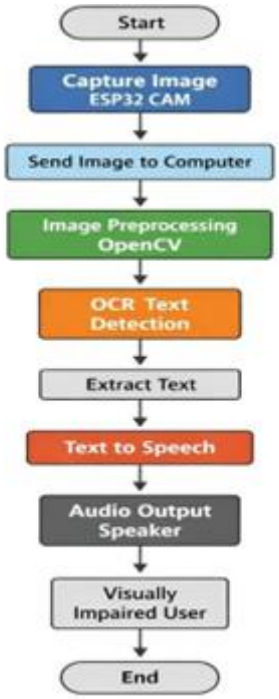

Fig -2: Workflow of Eye Mate

5. RESULTS AND SYSTEM PERFORMANCE

The EyeMate testing system was used to assess how well it could detect and convert text into speech. It comprised an ESP32 camera for capturing images; OpenCV to process images; Optical Character Recognition software (or OCR) to detect printed or clear handwritten text; and a Text-to-Speech Engine (or TTS) to convert the text into audio.

Through testing, the EyeMate System successfully identified both printed and clear handwritten text and read the information aloud to the user from books, labels, or documents that were available to the user.

Before capturing images, they were preprocessed by improving their quality resulting in improved clarity of image-recognized characters by the OCR engine. The converted audio would be of good enough quality such that users would find the audio readily understandable.

The EyeMate Automated Handheld System allowed for near real-time utilization, meaning the user could capture text, and receive audio output, with minimal delay. The testing results indicated that the EyeMate Automated System will provide visually impaired people with the means to read text independently. Additionally, the EyeMate Automated System is relatively simple and low- cost to use and can

therefore be utilized in many applicable situations on a daily basis.

5.1 System Accuracy

The legitimacy of the EyeMate system relies heavily on both the quality of photos taken of documents and the environment in which those documents were produced. Testing results produced extremely accurate results for printed materials while providing reasonably accurate results for clearly written documents created by hand. Pre-processing of images using OpenCV helped with noise reduction and improving visibility of text, leading to increased levels of accuracy when performing OCR processing.

The OCR engine had very high levels of accuracy across the entire range of detected characters when the document was properly photographed. Under appropriate lighting conditions, the system produced accurate results and created speech output that matched each character recovered. The text-to-speech module provided clear audible representations across the entire range of detected input characters without experiencing substantial errors. As noted above, the overall performance of the system under reasonable operating conditions was reliable.

5.2 User Experience and Performance

The EyeMate System is packed with features to ensure users with Poor Vision can use it effortlessly. Users simply have to point the camera at the text to enact the process. The EyeMate will automatically take an image of the text, process the image and verbalize the information on the image.

The overall process for users to receive the real-time verbal output of text is quick. On top of that, The EyeMate can be used to read Work-Books, Medicine Label's, Document's, and most Sign-Board's. This Independence will allow someone with Poor Vision to do activities more easily by themself.

5.3 Why EyeMate is Different from Other Systems

Unlike current text reading systems, EyeMate is a real-time assistive device for visually impaired users as it will work with both printed and handwritten text. All currently available solutions only provide an ability to read digital text. In addition to the above this allows EyeMate to be very useful in normal day-to-day situations when reading books, medicine labels, notices, and signs exactly as they are located in today's world using a camera.

EyeMate also uses an ESP32 CAM which is a great advantage in terms of reducing cost and portability. Many assistive devices currently available in the market are priced too highly, and difficult to use. The EyeMate system is simple to use, and also is affordable. All images captured through the use of OpenCV are enhanced for additional clarity which increases the overall accuracy of text detection. Therefore, the OCR engine can recognize text with a significantly higher degree of accuracy than the standard text reading applications.

EyeMate scans text continuously, and then converts that scanned text to speech almost immediately. Therefore, the combination of camera, image processing, OCR, and text to speech results in a completely automated EyeMate. EyeMate has a low price point, real time performance and can read both handwritten and printed text. This makes it easier to use and more practical than many other current systems available to read text.

6. FUTURE SCOPE

The EyeMate technology has a bright future ahead of it; there is great potential for the EyeMate solution to continue evolving into a system that is more accurate, faster and easier to use for those individuals who have a vision impairment.

Currently, the EyeMate can read text that is printed or in legible handwriting; however, future generations will enable the EyeMate to read many different handwriting styles and types of fonts.

Another area of development that will enhance the EyeMate is the portability of the system. The EyeMate can also be made into a fully portable product so users have the ability to carry it and utilize it in their everyday lives.

Another exciting opportunity for the EyeMate system is that its future versions will support many different language options, which will allow users to read in a variety of regional languages. Additionally, the quality of speech generated by the EyeMate can continue to improve based upon more "natural sounding" speech output. By utilizing improved image processing and improved optical character recognition (OCR) models, the EyeMate system will provide users with the ability to read in a much more accurate manner, even in low light.

In summary, future development initiatives for the EyeMate will focus on enhancing the EyeMate to provide users with the ability to read printed text in an independent manner

6.1 Expansion to Multiple Languages

One significant future enhancement to the EyeMate System involves providing support for multiple languages. Currently, the EyeMate System relies primarily on English text; however, it is anticipated that the EyeMate System will be trained in various regional languages at some point in the future. This functionality is essential to help users to

read books, documents, and other labels which contain text in several languages.

Adding multilanguage support will also improve access to the EyeMate System for many different people, no matter which region they live in. By utilizing advanced Optical Character Recognition models and language processing techniques, the EyeMate System will be able to accurately recognize letters from multiple alphabets and convert them into speech. As a result, the EyeMate System will be even more useful to many different people.

6.2 Portable Device and Mobile Integration

Future advancements include creating a truly portable system which will replace the computer used for processing today with a smaller, integrated device. Users will be able to carry the new device with them everywhere and use it on-the-go.

Current systems can connect to mobile devices, allowing images captured by the system to be processed using mobile-app technology. With mobile integration, users will experience faster result times and improved audio playback. Future enhancements may include wireless headsets, improved camera technology, and upgraded processor speeds. All of these features will provide EyeMate users with a much more functional and user-friendly system to use in everyday life.

6.3 Advanced AI and Accuracy Improvement

To improve the Eyemates system for the future, advanced AI models should be used instead of just the current Optical Character Recognition (OCR) being used now. The use of these types of deep learning models will help to further enhance the accuracy and speed of the system. These deep learning models will also assist in detecting readable text even if the image is taken in poor lighting, blurry or has some other type of obstructions.

Using various OpenCV techniques, the image processing aspect of the system may also be improved in order to obtain a clearer image for subsequent OCR processing. By making the image clearer, it will improve how the OCR engine can read the characters on the page or image. Future versions of the system may be programmed with automatic language identification so that users would not have to change the settings to access text in other languages from the original file.

Through enhanced AI models and rapid processing time, Eyemates can provide a much more dependable and efficient system, as well as offer additional assistance for visually impaired users in day-to-day situations.

CONCLUSION

In conclusion, EyeMate is an intelligent assistive system that helps visually impaired people read text (printed or handwritten) easily. The EyeMate system uses an ESP32-CAM module to take a picture of the target text. The image is then processed with OpenCV to identify regions of the image that may contain text. Next, the recognizable text is converted from image data to audio using Optical Character Recognition (OCR) technology and Text-to-Speech (TTS). The EyeMate system provides a means for a user that is visually impaired to understand the information contained in printed text.

In addition, unlike traditional text reading software solutions, EyeMate provides a real-time reading solution and can read text from a variety of sources including books, labels, documents, and signs directly without the need for assistance from another person. Furthermore, the EyeMate system is low-cost, portable, and easy to operate; therefore, it would be ideal for its use in everyday life. By using image preprocessing technology to process images before performing text recognition, the accuracy of the text recognition increases, and by providing an audio output of the recognized text, users have a clear auditory source of the information provided by the printed text.

The project further demonstrates that assistive technology can provide support to visually impaired individuals so they can live independently. Future enhancements of the EyeMate system could include improved OCR models, the ability to support multiple languages, and a fully portable version of EyeMate. Overall, EyeMate represents a simple but beneficial assistive technology solution which helps the visually impaired to have greater access to printed text in a more confident and efficient manner.

In summary, EyeMate is an intelligent assistant created to make reading printed and handwritten text easier for those who are blind or visually impaired, allowing them to complete tasks at their own pace and independently. The system utilizes ESP32-CAM to capture images, Opensource Computer Vision (OpenCV) for image preprocessing, Optical Character Recognition (OCR) for text recognition, and Text to Speech technology to convert text into audio output. By combining these technologies, EyeMate provides users with a way to comprehend printed material without needing to see it visually, thereby allowing them to perform daily tasks with greater ease.

REFERENCES

R. Smith, "An Overview of the Tesseract OCR Engine," Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR), 2007.
OpenCV Organization,
OpenCV: Open-Source Computer Vision Library, https://opencv.org
Espressif Systems,
ESP32-CAM AI-Thinker Module Datasheet, Espressif Systems, 2020.
J. Redmon et al.,
Image Processing and Computer Vision Techniques, IEEE Computer Vision Publications.
Google Developers,
Text-to-Speech API Documentation,
https://developers.google.com
P. Viola and M. Jones,
Rapid Object Detection using a Boosted Cascade of Simple Features,
IEEE Conference on Computer Vision, 2001.
World Health Organization (WHO),
Assistive Technology for Visually Impaired People,
https://www.who.int
Python Software Foundation,
Python Programming Language Documentation,
https://www.python.org
Gonzalez, R. C., & Woods, R. E.,
Digital Image Processing, Pearson Education, 2018.