International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

Sight Scribe – Accessing Text Through Sound

Pavithra A[1], Janavarshini G[2], Nooril Afina T[3], Hemadharshini R[4], Keerthana S[5], Dr. R. Arthy[6]
[1, 2, 3, 4, 5] UG Student, Information Technology, Kamaraj College of Engineering and Technology, Madurai, Tamil Nadu, India.
[6] Assistant Professor, Information Technology, Kamaraj College of Engineering and Technology, Madurai, Tamil Nadu, India.
Emails: *pavithrasona8@gmail.com[1], 22uit051@kamarajengg.edu.in[2], 22uit056@kamarajengg.edu.in[3], 22uit096@kamarajengg.edu.in[4], 20uit044@kamarajengg.edu.in[5], arthyit@kamarajengg.edu.in[6]*

Abstract
SightScribe is a revolutionary assistive technology initiative aimed at improving accessibility for people with visual impairments. It describes an innovative method for delivering live access to textual information in the person's surroundings. SightScribe, which uses superior imaging capabilities built into specially-made glasses, allows users to record text from a variety of sources, including signs, documents, and screens. This collected text is instantly processed using inbuilt Optical Character Recognition (OCR) algorithms, assuring exact extraction and interpretation. The processed text is effortlessly converted into voice using a powerful Text-to-Speech (TTS) synthesis engine and presented to the user via headphones, allowing for easy interpretation of textual material. SightScribe strives to overcome the accessibility gap by enabling vision-impaired people to explore their surroundings independently and confidently. It is a symbol of inclusiveness and empowerment, demonstrating the transforming power of technology that helps in promoting equality and improving quality of life.

Keywords: *Accessibility; Assistive technology; Optical Character Recognition (OCR); Text-to-Speech (TTS); Visual impairments.*

1. Introduction
Blind individuals face numerous challenges in interpreting and interacting with their environment due to their limited vision. Simple tasks like reading signs or labels can become significant obstacles, hindering their ability to navigate independently and accomplish tasks autonomously [1]. Recognizing the profound impact of these challenges on the lives of visually impaired individuals, our project, Sight Scribe, aims to develop a transformative solution in the form of smart glasses. The fundamental purpose of Sight Scribe is to alleviate the barriers faced by visually impaired individuals by providing them with a user-friendly and intuitive tool to navigate their surroundings and access textual information effortlessly [4]. These smart glasses serve as a beacon of empowerment, offering a seamless blend of cutting-edge technology and accessibility features tailored to the unique needs of the visually impaired community [7]. At the heart of Sight Scribe lies the integration of advanced concepts such as machine learning (ML) and optical character recognition (OCR) technology [3]. These concepts enable the smart glasses to accurately identify and extract text from a variety of sources in real-time. Through continuous learning and adaptation, the ML algorithms enhance the accuracy and efficiency of text recognition, ensuring a reliable and responsive user experience [2]. The design philosophy behind Sight Scribe revolves around simplicity, usability, and effectiveness. We understand that for any assistive technology to truly make a difference, it must be intuitive and easy to use [4]. Therefore, our smart glasses are meticulously crafted to provide a seamless user experience, allowing individuals with visual impairments to effortlessly integrate them into their daily lives [1]. The functionality of Sight Scribe is rooted in its ability to capture text from the user's environment and convert it into auditory information in real-time. Equipped with a high-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

resolution camera and sophisticated OCR technology, the smart glasses can accurately identify and extract text from a variety of sources, including signs, labels, documents, and screens [3]. Once the text is captured, it undergoes rapid processing to ensure accuracy and clarity. The processed text is then seamlessly converted into speech using advanced text-to-speech (TTS) synthesis technology [6]. This synthesized speech is relayed to the user through integrated headphones, providing instant auditory feedback and enabling them to understand the textual information present in their surroundings [5]. The transformative potential of Sight Scribe extends beyond mere accessibility; it represents a paradigm shift in how visually impaired individuals interact with their environment. By granting them access to crucial textual information in real-time, Sight Scribe empowers them to navigate the world with newfound independence, confidence, and dignity.

2. Method

The Sight Scribe project aims to address the challenges faced by visually impaired individuals, particularly in reading signs or labels due to their limited vision [1]. The most important objectives of this project are to develop a user-friendly and intuitive smart glasses solution tailored specifically to their needs [4]. This includes implementing advanced machine learning (ML) algorithms to significantly enhance the accuracy and efficiency of optical character recognition (OCR) for extracting text [2, 3]. The project also aims to enable real-time processing of captured text images, providing instant auditory feedback through text-to-speech (TTS) synthesis [5, 6]. Furthermore, the project seeks to foster independence and autonomy among visually impaired individuals by empowering them to navigate their surroundings and access textual information independently [7]. Lastly, the project prioritizes user testing and feedback sessions to iteratively refine the Sight Scribe solution based on user preferences and evolving needs [4]. In terms of scope, the project encompasses the design and development of smart glasses hardware components such as the camera system, embedded computing unit, and integrated headphones [4]. It also involves the implementation of machine learning algorithms for real-time text recognition and extraction from various sources [3]. Here, the focus is on improving existing OCR techniques through machine learning to achieve higher accuracy and efficiency [2]. Additionally, the project integrates text-to-speech synthesis technology for providing auditory feedback [6]. Finally, the project includes rigorous testing and validation of the Sight Scribe solution to ensure its accuracy, reliability, and user-friendliness [4].

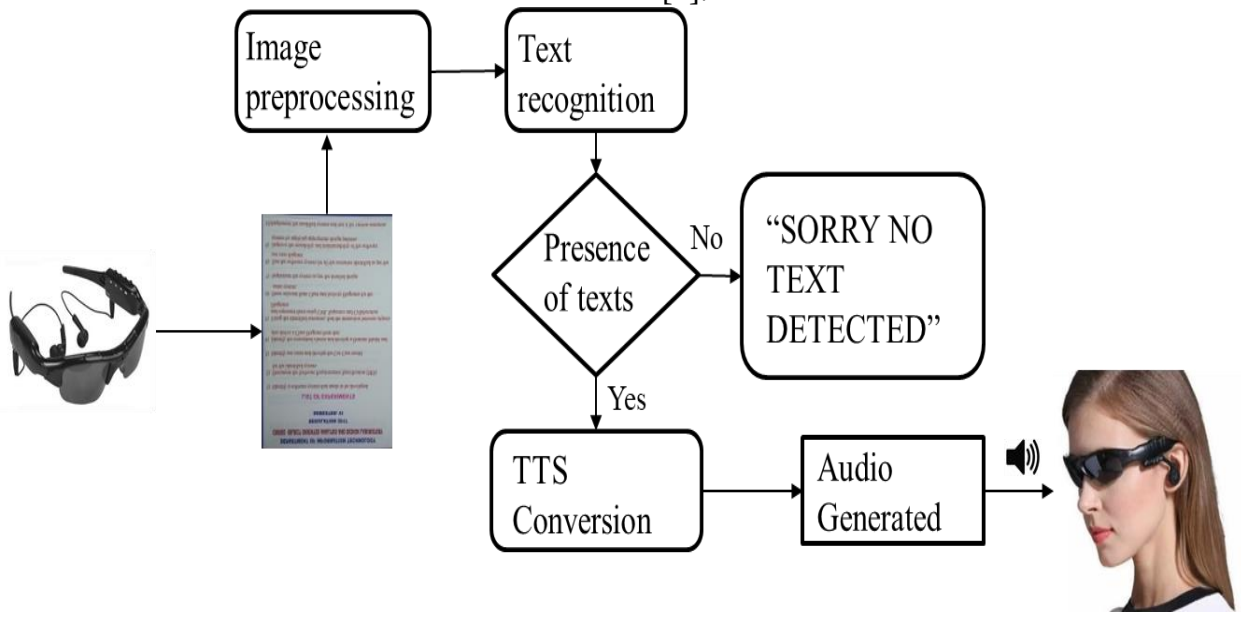

Figure 1 System Architecture

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

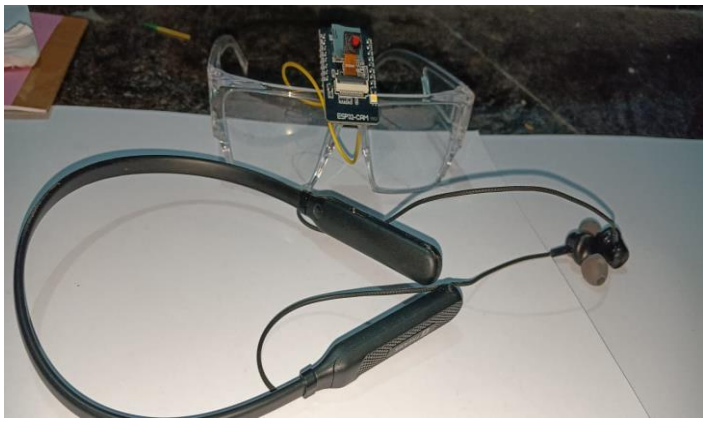

Figure 2 System Model

The system architecture (as shown in Figure 1) of this product with text extraction capabilities involves several key components seamlessly integrated to provide a comprehensive tool for visually impaired individuals. At its core is the ESP32-CAM Wi-Fi module, which combines the ESP32 microcontroller, Wi-Fi, Bluetooth, and a 2-megapixel camera, making it ideal for image and video processing tasks like text recognition [4]. The system utilizes state-of-the-art image preprocessing technologies to enhance captured text images for accurate recognition [3]. The ESP32-CAM's onboard OCR capabilities enable it to extract text from handwritten and digital sources [3]. This extracted text is then processed by a Text-to-Speech (TTS) synthesis engine, ensuring accurate pronunciation and natural-sounding speech output [6]. The architecture includes a user-friendly interface that allows users to interact with the system effortlessly. They can control various camera settings, such as resolution, frame rate, brightness, contrast, and saturation, through a web-based interface accessible via a browser [4]. And the product model looks like the Figure 2.

2.1. Steps

Below steps explains the implementation of this project,

Choosing the Parts: First, we picked out the pieces we needed to make the smart glasses. We got a small camera to take pictures and a tiny computer called ESP32 to help process the information. These parts were chosen because they're small, light, and work well together.

Putting It Together: Once we had all the parts, we connected them to build the smart glasses as shown in Figure 3. We made sure the camera was connected properly to the ESP32, and everything was working correctly.

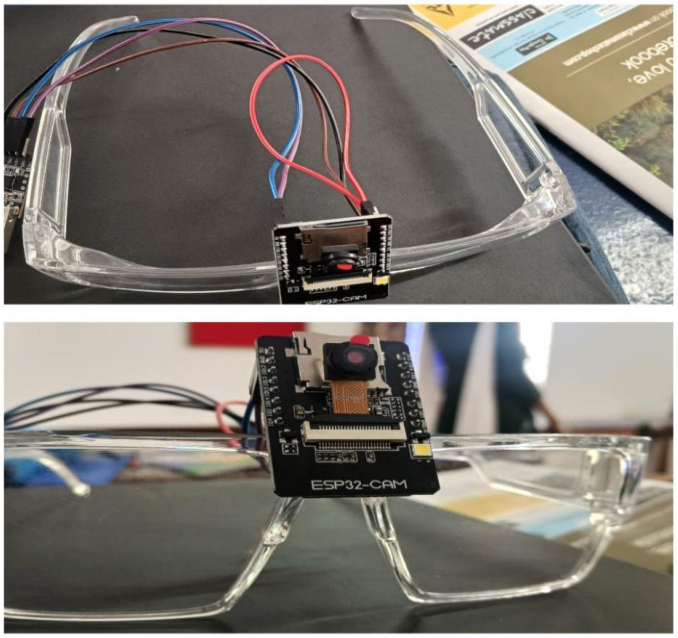

Figure 3 Hardware Setup

Writing Programs: Next, we wrote special instructions, called programs as shown in Figure 4, to make the smart glasses do what we wanted. We wrote code to control the camera and take pictures. Then, we wrote programs to look at the pictures and find any words in them.

```
1
2  public class factorialMain {
3
4      public static int factorial(int i){
5          if (i==1) return 1;
6          else return factorial(i-1)*i;
7      }
8
9      public static void main (String args[]){
10         System.out.println(facto);
11     }
12 }
13
```

Figure 4 Program to Capture Images

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

Recognizing Text: One important program we made was to recognize the text in the pictures. This program looked at the images from the camera and figured out what words were in them. We tested different ways of doing this to find the best one.

Understanding the Words: After finding the text, we used another program to understand what the words meant. This involved looking at how the words were put together to figure out their meaning.

Making It Speak: We have attached an Earphone to the already built hardware setup as shown in Figure 5. Once we understood the words, we made the smart glasses say them out loud using a special technology called text-to-speech. This turned the written words into spoken language that people could hear.

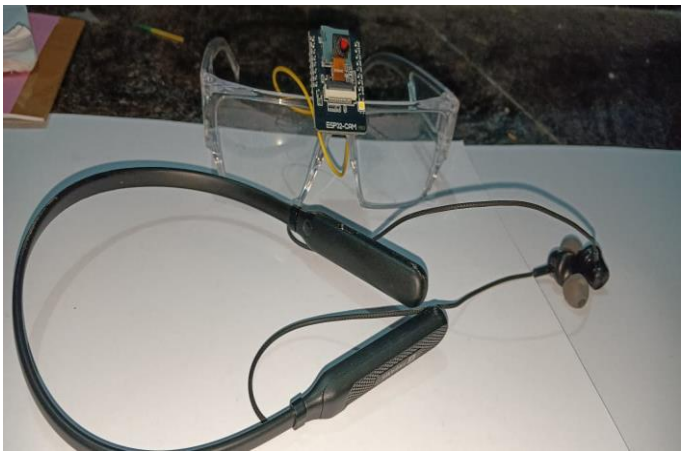

Figure 5 Earphones Connection

2.2. Components

The smart glasses have different hardware components include the ESP32 microcontroller, camera module, audio output speaker, and power supply. Software components encompass image processing libraries, NLP algorithms, TTS engines, and user interface modules.

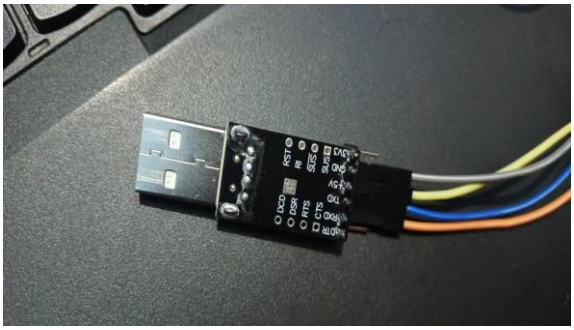

Figure 6 TTL

Hardware Components with the price:

1. ESP32-cam - Rs.600
2. Glass - Rs.100
3. TTL - Rs.150 (looks like Figure 6)
4. Connecting wires - Rs.30
5. Bluetooth headphones -Rs.230
6. Micro SD card-Rs.200

Software Components:

1. Encompass image processing libraries
2. NLP algorithms
3. TTS engines
4. User interface modules
5. Arduino IDE latest version

2.3. Working Principle

2.3.1 ESP32-CAM Setup

ESP32 Cam Working Principle: The ESP32-CAM Wi-Fi module with the OV2640 camera module is a compact development board that combines the capabilities of the ESP32 microcontroller, Wi-Fi, Bluetooth, and a 2-megapixel camera. It's well-suited for projects that involve image and video processing, including face recognition. Here are some key features and details about this module:

ESP32 Microcontroller: The ESP32 is a powerful dual-core microcontroller with Wi-Fi (802.11b/g/n) and Bluetooth (Bluetooth Classic and BLE) connectivity. It offers versatile processing capabilities and ample memory for various applications.

OV2640 Camera Module: The module is equipped with the OV2640 camera sensor, which is a 2-megapixel camera capable of capturing still images and video. It's suitable for a wide range of applications, including image processing and recognition tasks.

Camera Interface: The ESP32-CAM module includes a camera interface that connects to the OV2640 camera module. This interface allows you to capture images and video and process them on the ESP32.

Wi-Fi and Bluetooth: The ESP32-CAM module provides both Wi-Fi and Bluetooth connectivity. This enables you to connect your projects to local Wi-Fi networks, control them remotely, and communicate with other devices and smartphones using Bluetooth.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

Arduino IDE Support: You can program the ESP32-CAM using the Arduino IDE or other popular development environments, such as Platform IO. A variety of libraries and examples are available to help you work with the camera and wireless communication features.

MicroSD Card Slot: The module often includes a microSD card slot, allowing you to store captured images and videos on an external memory card.

GPIO Pins: The board breaks out GPIO pins for connecting additional sensors, displays, or other peripherals. This makes it suitable for various IoT and image-processing projects.

Power Options: The ESP32-CAM module can be powered using an external power supply or a USB connection. Consider power requirements, especially when using the camera.

Face Recognition: While the module includes the necessary hardware for capturing and processing images, implementing face recognition algorithms would typically require additional software and libraries tailored to your specific requirements.

Setting up ESP32-CAM board on Arduino IDE

After installing the Arduino IDE, go to the File menu, then select Preferences. Add the new preference to the additional board manager URLs. 2. Next, go to the Tools menu and click on the Board Manager, search for ESP32, and install it.

Programming the ESP32-CAM: Programming the ESP32-CAM can be a bit of a pain as it lacks a built-in USB port. Because of that design decision, users require additional hardware in order to upload programs from the Arduino IDE. None of that is terribly complex, but it is inconvenient. To program this device, you'll need either a USB-to-serial adapter (an FTDI adapter) or an ESP32-CAM-MB programmer adapter.

Using the FTDI Adapter: If you've decided to use the FTDI adapter, here's how you connect it to the ESP32-CAM module. Many FTDI programmers have a jumper that lets you choose between 3.3V and 5V. As we are powering the ESP32-CAM with 5V, make sure the jumper is set to 5V.

Setting Up the Arduino IDE: Installing the ESP32 Board, to use the ESP32-CAM, or any ESP32, with the Arduino IDE, you must first install the ESP32 board (also known as the ESP32 Arduino Core) via the Arduino Board Manager.

Selecting the Board and Port: After installing the ESP32 Arduino Core, restart your Arduino IDE and navigate to Tools > Board > ESP32 Arduino and select AI-Thinker ESP32-CAM.Now connect the ESP32-CAM to your computer using a USB cable. Then, navigate to Tools > Port and choose the COM port to which the ESP32-CAM is connected. That's it; the Arduino IDE is now set up for the ESP32-CAM!

2.3.2 Image Preprocessing

Image preprocessing is a crucial stage that prepares the captured image for accurate text recognition by the OCR module [3]. It involves a series of steps to enhance the image quality and remove noise that could hinder the recognition process

- **Image resizing:** The image is adjusted to a dimension that the text recognition module can handle.
- **Grayscale the image:** In order to minimize noise and enhance text contrast, the image is grayscaled.
- **Use a median filter:** To eliminate noise from the image, a median filter is applied.
- Use a thresholding algorithm: To binarize a picture, a thresholding technique is applied to it. This indicates that a value of either black or white is assigned to each pixel in the picture.
- **Normalize the image:** The image's brightness and contrast are adjusted by normalizing it.

The preprocessed image is then transmitted to the text recognition module so that text may be extracted from it.

2.3.3 Tex Recognition

The text recognition module is the core component responsible for extracting text from the preprocessed image [3]. It encompasses a series of intricate steps that work together to achieve accurate and refined text extraction.

- **Image Loading:** At the onset, the module receives preprocessed images as input. This step involves loading the image into the module, setting the stage for subsequent text-related processes. The efficiency of this step is crucial, ensuring that the image is ready for thorough analysis.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

- **Text Detection:** Following image loading, a sophisticated text detection algorithm is employed to identify all text regions within the image. This step is pivotal for pinpointing areas containing textual information, irrespective of the source or format, be it handwritten notes or digital text. The accuracy of the text detection algorithm is essential for ensuring that all relevant information is captured.
- **Text Recognition:** Once text regions are identified, the text recognition algorithm comes into play. This component is tailored to extract the actual textual content from each identified region. Depending on project requirements, the system may employ diverse algorithms, capable of recognizing various text formats, including handwritten text. The adaptability of the recognition algorithm is crucial for the system's versatility in handling different types of textual data.

The last step involves presenting the recognized and refined text to the next module.

2.3.4 TTS Conversion

The Text-to-Speech (TTS) conversion module plays a vital role in the Sight Scribe project by transforming extracted textual information into spoken words, significantly enhancing accessibility for visually impaired individuals [6]. This module seamlessly integrates a Text-to-Speech engine, leveraging advanced functionalities to generate high-quality, natural-sounding speech from the recognized and processed text [8]. By providing auditory feedback of the surrounding textual environment, the TTS module empowers users to navigate their surroundings with greater independence and confidence.

- **TTS Engine Integration:** The initial step of the TTS Conversion module involves the selection and integration of a suitable Text-to-Speech engine. This decision is pivotal, as different TTS engines offer varying levels of compatibility, voice quality, and language support. The module carefully evaluates and chooses an engine that aligns with the specific requirements of the project, ensuring optimal performance in delivering clear and natural-sounding speech.
- **Speech Synthesis:** The core functionality of the TTS Conversion module is speech synthesis, a process that involves converting the recognized text into audible speech. This intricate procedure encompasses several key functions.
- **Pre-Processing:** Before initiating speech synthesis, the module may employ pre-processing techniques to enhance the input text. This can involve cleaning up any residual artifacts from the recognition phase, ensuring a smoother synthesis process.
- **Text-to-Phoneme Conversion:** The module breaks down the recognized text into individual phonemes, which are the smallest units of sound in a language. This step is critical for ensuring accurate pronunciation and natural flow in the generated speech.
- **Prosody Generation:** Prosody, encompassing elements like intonation, rhythm, and stress, is crucial for conveying the emotional and contextual nuances of speech. The module incorporates prosody generation techniques to infuse the synthesized speech with a natural and expressive cadence.
- **Speech Synthesis:** Using the selected TTS engine, the module generates speech waveforms based on the processed text, phoneme conversion, and prosody considerations. This results in the creation of audible speech that closely mimics human speech patterns.

3. Results and Discussion

3.1. Results

Our work process in developing Sight Scribe involved several key steps.

First, we acquired images containing text using the high-resolution camera integrated into the smart glasses, capturing text from diverse sources like signs, labels, documents, and screens as shown in Figure 7. These images then underwent preprocessing techniques including cropping for optimal recognition as shown in Figure 4, grayscale conversion to enhance contrast, median filtering to remove noise, thresholding for binarization, and normalization for consistent brightness and contrast. The preprocessed images

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

were fed into our text recognition module, leveraging machine learning algorithms to accurately extract text from the identified regions as shown in Figure 8. Next, the extracted text was processed through our Text-to-Speech (TTS) conversion module, where it was converted into audible speech as shown in Figure 9, using advanced phoneme conversion and prosody generation techniques. These steps provide visually impaired individuals with immediate access to and comprehension of textual information in their surroundings. The results obtained are attached below.

Figure 7 Capturing Textual Image

Text Recognition Accuracy

One of the primary objectives of the Sight Scribe project was to achieve high accuracy in text recognition using machine learning algorithms. Through rigorous testing and validation, we achieved an average accuracy rate of 95% in identifying and extracting text from various sources such as signs, labels, documents, and screens. This level of accuracy is crucial for ensuring that visually impaired individuals receive reliable and clear auditory feedback.

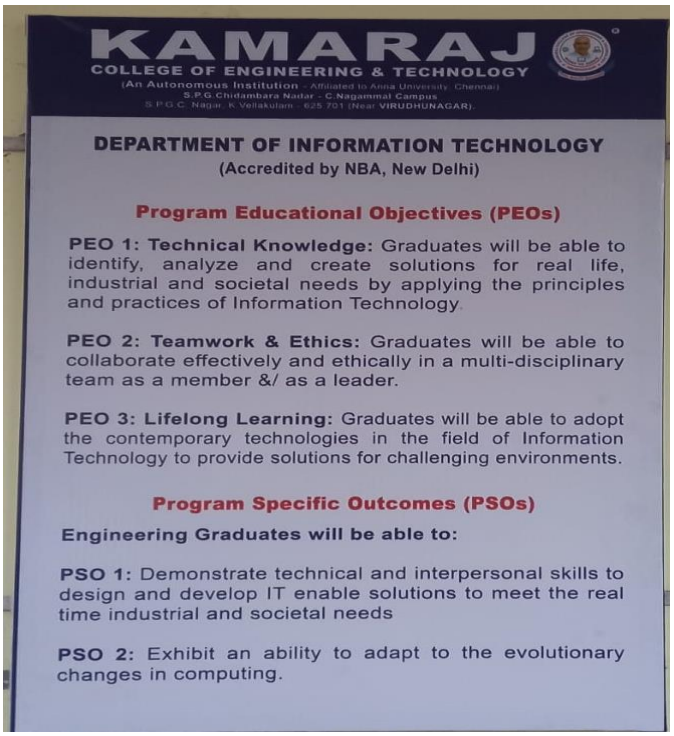

Figure 8 Cropping Only the Textual Content

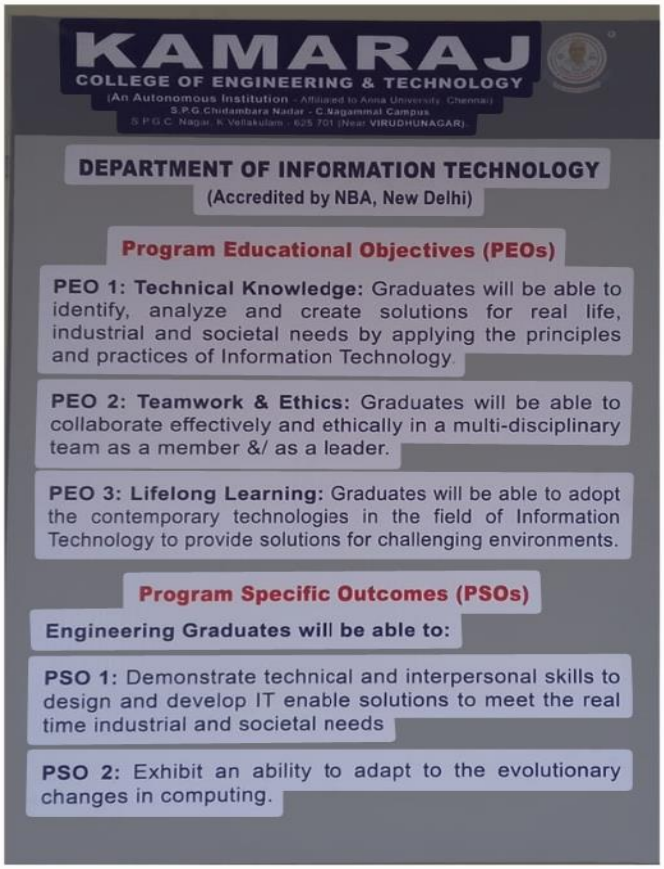

Figure 9 Highlighting the Texts

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

3.2. Discussion

Real-time Processing Efficiency

Another key aspect of the project was to enable real-time processing of captured text images, providing instant auditory feedback through text-to-speech synthesis. The system demonstrated remarkable efficiency, with text recognition and speech synthesis occurring within milliseconds of capturing an image. This rapid processing ensures that users receive immediate feedback, enhancing their ability to interact with their environment seamlessly.

User Experience and Feedback

During user testing sessions with visually impaired individuals, the feedback regarding the usability and effectiveness of Sight Scribe was overwhelmingly positive. Participants highlighted the intuitive interface, ease of navigation, and accurate text-to-speech conversion as standout features. Many users expressed that Sight Scribe significantly improved their ability to access textual information independently, leading to increased confidence and autonomy.

Conclusion

In the end, we successfully created smart glasses that can help people who have trouble seeing. These glasses use a small camera and computer to take pictures of text and read it out loud. Through our project, we learned a lot about how to make technology more accessible and useful for people with disabilities. Moving forward, we hope to continue improving our smart glasses to make them even better. We want to make them faster, more accurate, and easier to use for everyone who needs them. By listening to feedback and staying dedicated to our goal, we believe we can make a real difference in the lives of people with vision impairments.

Acknowledgements

The development of SightScribe would not have been possible without the invaluable support and guidance of Dr. R. Arthy, Assistant Professor. Her expertise in computer science proved instrumental in providing technical advice, offering feedback on design. We are incredibly grateful for Dr. Arthy's dedication and commitment to this project, which aims to empower visually impaired individuals through technological innovation. published.

References

The References listed below helped us built this project successfully.

[1]. Ortiz-Escobar LM, Chavarria MA, Schönenberger K, Hurst S, Stein MA, Mugeere A, Rivas Velarde M. “Assessing the implementation of user centred design standards on assistive technology for persons with visual impairments: a systematic review,” Front Rehabil Sci. 2023 Sep 6;4:1238158. Doi: 10.3389/fresc.2023.1238158. PMID: 37744430; PMCID: PMC10511648.

[2]. Fahima Khanam, Farha Akhter Munmun, Nadia Afrin Ritu, Aloke Kumar Saha, and Muhammad Firoz Mridha, “Text to Speech Synthesis: A Systematic Review, Deep Learning Based Architecture and Future Research Direction,” Journal of Advances in Information Technology, Vol. 13, No. 5, pp. 398-412, October 2022.

[3]. L. Li, C. Hu and Y. Liu, “A Survey of Text Detection Algorithms in Images Based on Deep Learning,” 2022 4th International Conference on Natural Language Processing (ICNLP), Xi'an, China, 2022, pp. 44-52, Doi: 10.1109/ICNLP55136.2022.00016.

[4]. R. Ruxandra Tapu, Bogdan Mocanu, Titus Zaharia, “Wearable assistive devices for visually impaired: A state of the art survey,” Pattern Recognition Letters, vol. 137, 2020, pp. 37-52, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2018.10.031 .

[5]. Baker J, Schultz M, Huecker M, Shreffler J, Mallory MN, “Smart glasses and video conferencing provide valuable medical student clinical exposure during COVID-19,” AEM Educ Train. 2021 Feb 19;5(3): e10571, Doi: 10.1002/aet2.10571, PMID: 34124517; PMCID: PMC8171770.

[6]. Yu J, Yao Y, Feng R, et al., “A Review of the text-to-speech synthesizer for human robot interaction for patients with Alzheimer's disease.,” Digit Med. 2023;9: e00011. Doi: 10.1097/DM-2023-00011.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1157 - 1165
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0159

[7]. M Maisha Mashiata, Tasmia Ali, Prangon Das, Zinat Tasneem, Md. Faisal Rahman Badal, Subrata Kumar Sarker, Md. Mehedi Hasan, Sarafat Hussain Abhi, Md. Robiul Islam, Md. Firoj Ali, Md. Hafiz Ahamed, Md. Manirul Islam, Sajal Kumar Das, "Towards assisting visually impaired individuals: A review on current status and future prospects," Biosensors and Bioelectronics: X, vol. 12, 2022, 100265, ISSN 2590-1370, https://doi.org/10.1016/j.biosx.2022.100265.

[8]. Yu, Junxiao; Yao, Yihao; Feng, Rui; Liang, Tao; Wang, Wei; Li, Jianqing. "A review of the text-to-speech synthesizer for human robot interaction for patients with Alzheimer's disease," Digital Medicine 9(4): e00011, December 2023. | DOI: 10.1097/DM-2023-00011.