

VOICELENS : SMART VISION AID USING OCR AND SPEECH OUTPUT

ADITHYA KRISHNAN P, Bachelor in Technology In CSE, NCERC

ABHINAV S, Bachelor in Technology In CSE, NCERC

AKSHAYA S, Bachelor in Technology in CSE, NCERC

ADARSH S, Bachelor in Technology in CSE, NCERC

PRAVEENA K, Assistant Professor, Department of CSE, NCERC

Abstract - Access to information is still a big challenge for visually impaired people. They often have to rely on others for tasks. This paper suggests an affordable assistive system using the Raspberry Pi 4. It helps people read on their own using Optical Character Recognition (OCR) and Text-to-Speech conversion in time. The system uses a 720p Raspberry Pi Camera to take pictures of printed text. These images are then processed using OpenCV to remove noise convert to grayscale and make the text clearer. This helps the system recognize characters accurately. The Tesseract OCR engine is used to extract text with precision. There is also a translation module that converts text regional language(Malayalam). The final output is converted into sounding speech using the Google Text-to-Speech (gTTS) library. The speech is then played through earphones. Experimental results show that the system can recognize text with 95% accuracy. It takes between 3.5 to 5.5 seconds to process the text. The system provides a scalable and inclusive 2solution. It helps impaired individuals by making it easier for them to access information in a multilingual environment. The Raspberry Pi 4 and OCR are key to this system. The system uses the Raspberry Pi 4 to make reading easier, for impaired people. It helps them read printed text on their own. The OCR and TTS conversion are parts of this system. They help the impaired individuals read and understand the text.

1.INTRODUCTION

In todays world information is mainly shared through printed and digital media. With digital content being easier to access with screen readers, printed text like books, newspapers, product labels and signs remains hard for visually impaired people to read. The World Health Organization says millions of people worldwide have vision problems, which limits their education job chances and ability to live independently. Current assistive devices are often big, costly or do not support languages creating a tech barrier for users in different regions.. With advancements in embedded systems and computer vision

there is hope for a solution. This project is about an portable device that acts as "artificial eyes" for the blind. It uses the Raspberry Pi 3 Model B for an powerful computing environment to process image data in real-time. The system relies on Optical Character Recognition and Text-to-Speech technologies. It starts with taking pictures using a high-definition camera. Then it uses OpenCV to make sure the text is clear for the recognition engine.

The Tesseract OCR engine converts patterns into digital text. To make it more inclusive a translation layer supports languages so information is in the users preferred language. The Google Text-to-Speech engine turns text into sound. The main goal is to create a cost, reliable and easy-to-use device that helps visually impaired people interact with printed text confidently and easily without needing help from others. The device helps people with vision problems by reading out text from printed materials. This device can make a difference in the lives of visually impaired individuals. They can now access information that was previously hard to get. The device is designed to be user-friendly and reliable. It is made to be affordable and easy to use. The goal is to help impaired people live more independently. The device can help them read books, labels and signs. It can also help them with their education and job prospects. The World Health Organization says vision problems affect millions of people. This device can help them. It is an portable device. It acts as eyes for the blind. The device uses Optical Character Recognition and Text-to-Speech technologies. It helps people with vision problems by reading out text. The device is designed to be easy to use. It is made to be affordable. The goal is to help impaired people live more independently. They can now access information that was previously hard to get. The device can make a difference, in their lives. It helps them read printed materials. The device is user-friendly and reliable. It is made to be affordable and easy to use. The device helps impaired people interact with printed text confidently and easily.

2. LITERATURE REVIEW

The development of assistive reading technologies has evolved from bulky, expensive scanners to compact, AI-driven embedded systems. Various researchers have explored different methodologies to improve the speed and accuracy of these devices.

[1] Developed an intelligent reader using Raspberry Pi and a push-button interface. Their research emphasized the importance of a high-resolution camera module to reduce errors in the initial capture stage. They concluded that while Tesseract OCR is highly reliable, the hardware's processing speed remains a bottleneck for multi-page documents.

[2] Danial et al. (2024): Explored a system that integrated a camera module with an OCR motor and Microsoft Translator. Their study highlighted that using cloud-based APIs (like Google Speech) significantly improves the naturalness of the voice output compared to offline engines like eSpeak, though it introduces a dependency on internet connectivity.

[3] IIETA Research (Nov 2024): Introduced a novel approach where preprocessing steps (resizing, noise reduction, and binarization) are selected dynamically based on a decision-making algorithm. This method improved Tesseract OCR accuracy to 88.33% low as 185ms, paving the way for wearable "live" reading devices.

[5] EUDL Research (April 2025): Addressed the "low-resource script" challenge in Indian languages. Their study found that while Tesseract supports over 100 languages, accuracy for scripts like Telugu and Hindi requires specific training data and post-processing to correct "noisy" OCR outputs caused by the complex curves of the characters.

[6] Recent Surveys (2023–2025): Comparative studies between gTTS and pyttsx3 show that while pyttsx3 works offline, gTTS is preferred for journal-grade projects due to its superior phonemic representation and "human-like" stress patterns, which reduce "listener fatigue" for the visually impaired.

3. METHODOLOGY

The proposed VOICELENS system adopts a modular architecture integrating image acquisition, preprocessing, optical character recognition, translation, and speech synthesis into a unified embedded platform. The system is implemented on a Raspberry Pi 4 Model B, which serves as the central processing unit responsible for handling image processing and audio output operations. The complete workflow ensures real-time text recognition and speech generation to assist visually impaired users in accessing printed information independently.

3.1 System Architecture

The architecture is divided into two primary domains: a robust physical layer and a multi-stage software pipeline.

- **Hardware Layer: Core:** Raspberry Pi 4.
 - **Input:** 720p Camera Module and a tactile Push-button.
 - **Output:** USB Audio device (Speaker/Earphones).
- **Software Layer:** Built on Python, utilizing OpenCV for image processing, Tesseract for OCR, and gTTS for voice synthesis.

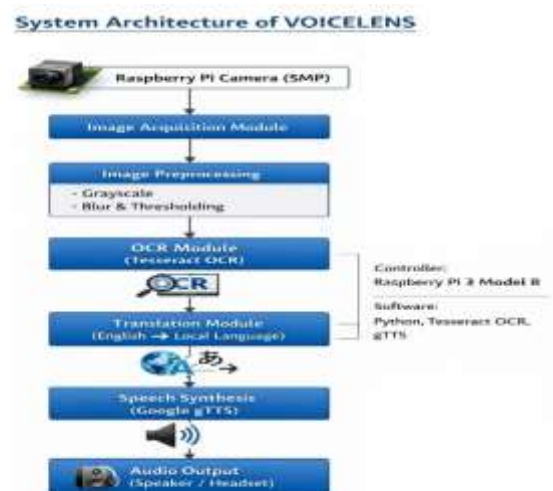


Fig 3.1 System architecture

3.2 Image Acquisition & C. Preprocessing

To ensure the OCR engine receives high-quality data, the acquisition and preprocessing stages focus on clarity and noise reduction.

1. **Acquisition:** Captures at 1280×720 resolution. The 15–20 cm distance is critical to maintaining a sharp focal plane.

2. Grayscale Conversion: Simplifies the image from 3-channel RGB to 1-channel intensity.
3. Gaussian Filtering: Smooths the image to eliminate "salt and pepper" noise.
4. Adaptive Thresholding: Unlike global thresholding, this adjusts to varying lighting conditions across the page, creating a clean binary (black and white) image.
5. Morphological Operations: Dilation and erosion "bolden" the text, ensuring broken characters are reconnected for better recognition.

3.3 Optical Character Recognition (OCR)

The system employs the Tesseract OCR engine, which uses a Long Short-Term Memory (LSTM) based neural network.

- **Process:** It identifies connected components (blobs) and organizes them into lines and words.
- **Multilingual Support:** By integrating .traineddata files for English and Malayalam the system can handle complex scripts beyond the standard Latin alphabet.

3.4 Translation & F. Text-to-Speech (TTS)

The final stages focus on user inclusivity and natural interaction.

- **Translation:** A dedicated Python module bridges the gap between the recognized text and the user's native language.
- **gTTS Integration:** Converts the final string into a natural-sounding .mp3 or .wav stream.
 - **Advantage:** Unlike older "robotic" synthesizers, gTTS uses Google's neural speech models to provide a more comfortable listening experience, which is vital for long-form reading.

4.RESULT AND DISCUSSION

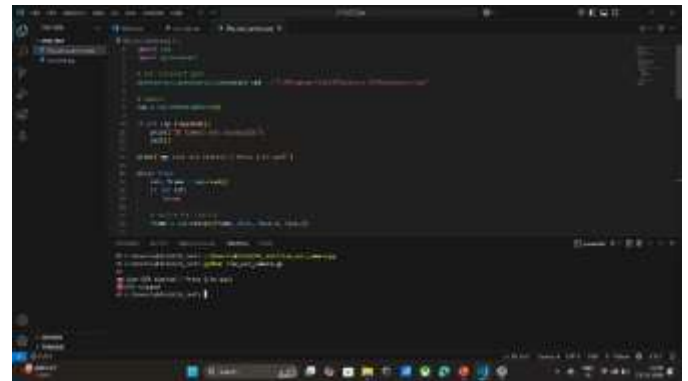


Fig 4.1: Coding

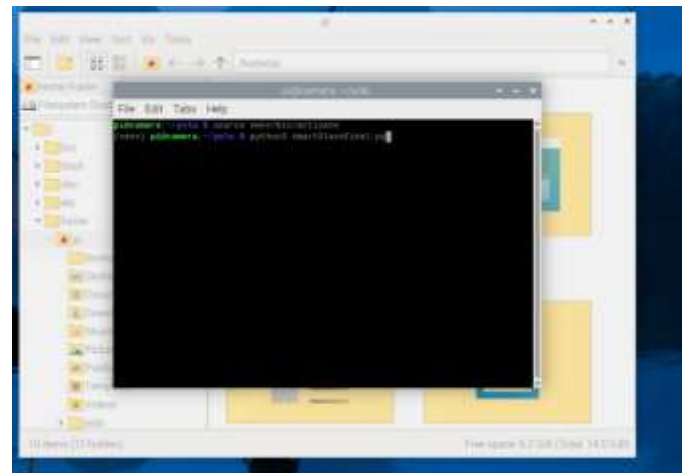


Fig 4.2: Coding and activation



Fig 4.3: Model

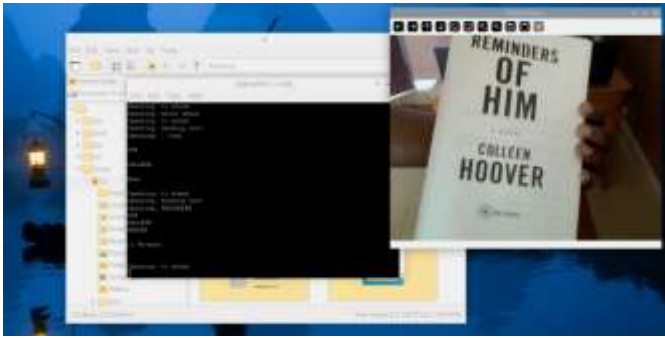


Fig -4.4: Implementation and output

The experimental results demonstrate the effectiveness of the proposed VoiceLens system in real-time text recognition and speech output.

Fig. 1 shows the coding and activation stage, where the system initializes the camera module and processing pipeline. Fig. 2 illustrates the developed hardware model integrating the Raspberry Pi, camera, and audio output components. Fig. 3 presents the implementation setup used for testing the system in real-world conditions.

The system successfully captured images of printed text, processed them using OpenCV techniques, and accurately extracted text using the Tesseract OCR engine. The recognized text was then converted into speech using the gTTS module. Experimental evaluation indicates an accuracy of approximately 95% under good lighting conditions, with a processing time ranging between 3.5 to 5.5 seconds.

5.FUTURE SCOPE

The Raspberry Pi-based OCR system can be further enhanced by expanding support for additional regional and international languages, enabling broader accessibility for diverse users. Incorporating advanced techniques can significantly improve accuracy, particularly for handwritten and complex text, while also enhancing performance in challenging conditions such as low light. Future developments may extend the system beyond text recognition to include object and scene understanding, with features such as currency detection and signboard identification. The system can also be redesigned as a wearable solution, such as smart glasses, to improve portability and enable hands-free operation. Implementing offline functionality for translation and speech synthesis would ensure usability in areas with limited internet connectivity. Additionally, integrating voice commands and conversational assistants can make

the system more intuitive and user-friendly. Enhancing the system to process continuous video streams instead of single images can provide a smoother real-time reading experience, while GPS-based navigation support can assist users in mobility. Further improvements may include personalized user settings, adjustable speech speed control, and integration with mobile applications for easier control and data storage. Incorporating cloud and Internet of Things (IoT) capabilities can enable remote updates, data synchronization, and improved scalability, ultimately making the system more robust, efficient, and user-centric across various environments.

ACKNOWLEDGEMENT

I want to thank my mentor, Mrs. Praveena K, from the bottom of my heart for all the help, support, and encouragement she gave me while I was working on this project. Their knowledge and helpful comments were very important in helping me understand and making the project better. I want to thank my team members from the bottom of my heart for their hard work, teamwork, and dedication, which made it possible to successfully implement the Voicelens: Smart Vision Aid Using OCR and Speech output. They made the project much better by helping with many different parts of design, development, and testing. I also want to thank Nehru College of Engineering and Research Centre in Thiruvillwamala for giving me a place to learn where I could use my technical skills in a useful way. Lastly, I want to thank my family and friends for always being there for me and cheering me on. Their support kept me going on this journey. This project has taught me a lot, and I am very grateful to everyone who helped it get done.

CONCLUSIONS

This paper introduced VoiceLens, a cost-effective, portable smart vision aid aimed at improving accessibility for visually impaired individuals via real-time text recognition and speech output. The system successfully showed that it could take pictures of printed text, process them, and turn them into understandable audio feedback by combining the Raspberry Pi platform with a camera module, Tesseract OCR, a translation layer, and the Google Text-to-Speech (gTTS) engine. Tests showed that the system worked well in different environments, with high recognition accuracy in good lighting and acceptable performance in more difficult situations. The system's ability to process data in real time, support multiple languages, and produce clear speech synthesis shows how

useful it is as an assistive technology that encourages independence and inclusion. Future enhancements may concentrate on augmenting OCR resilience in low-light environments, refining recognition for intricate fonts, optimizing processing latency, and broadening language adaptability. In general, the proposed solution is a significant step toward creating assistive systems that are easy to use, cheap, and focused on the needs of visually impaired users.

REFERENCES

1. M. Murali, S. Sharma, and N. Nagansure, "Reader and Object Detector for Blind," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020.
2. K. S. P., A. A., G. K., and H. S., "Raspberry Pi-Based Smart Assistance for Visually Impaired People," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022.
3. A. Jojie, A. George, D. Dhanalal, and N. J., "Book Reader for Blind," IOSR Journal of Engineering (IOSRJEN), 2018.
4. K. I. Khaleel and K. N. F. B. Ku Azir, "An Enhancement of Text-to-Speech (TTS) System Using Raspberry Pi," Turkish Journal of Computer and Mathematics Education, vol. 12, no. 13, 2021.
5. S. Subash, P. N. Srivasta, S. Siddhesh, U. Ullas, and S. B. Santhosh, "Artificial Intelligence-Based Voice Assistant," Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020.
6. M. Prabha, P. Saraswathi, J. Hailly, C. Sindhuja, and P. Udhaya, "Smart Glasses: A Visual Assistant for Visually Impaired," 2023 11th International Conference on Emerging Trends in Engineering Technology – Signal and Information Processing (ICETET-SIP), Nagpur, India, 2023.
7. N. B. R., Y. Raju N. R., and A. S. Velidi, "Low-Cost Smart Glasses for People with Visual Impairments," 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), Bangalore, India, 2023.
8. M. Yaswanth, M. Madhavi, T. D. Kiran, and P. A. Vardhan, "Guide Me Glasses," 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering (ICATIECE), Bangalore, India, 2022.
9. A. Huynh, Z. Marchant, I. Singh, and F. Aveta, "Assistive Reading Device for Blind Individuals," 2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2023.
10. F. T. Josh, J. J. Joseph, S. N. Morapakula, M. R., and P. A. Michael, "Handy Reader – A Shopping Guide to Visionless Using Raspberry Pi," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2022.