Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

Assistive Device for the Visually Impaired using Face and Optical Character Recognition with Voice Activation

Himali R
Department of Electronics and communication Engineering
Bengaluru, KA,

Adithya S
Department of Electronics and communication Engineering
Bengaluru, KA,

Ravishankara M N
Department of Electronics and communication Engineering
Bengaluru, KA,

***Abstract*— This paper presents an assistive device designed to aid visually impaired individuals through face recognition, optical character recognition (OCR), and voice-activated text-to-speech functionalities. Implemented using the Raspberry Pi Compute Module 4 (CM4), the system integrates Python OpenCV for facial recognition, Tesseract for OCR, and a google text-to-speech (gTTS) converter for audio feedback. Mono 8MP camera capture images for face and text processing. The deep learning-based face recognition module identifies individuals, while the OCR module extracts and converts text into speech, enabling real-time interaction. The device provides practical solutions for applications such as assistive technologies, smart surveillance, and document processing. We detail the system's design, methodology, implementation, and performance, highlighting challenges and potential improvements for future iterations.**

***Keywords*— Assistive Device, Visually Impaired, Face Recognition, Optical Character Recognition, Text-to-Speech, Raspberry Pi**

I. INTRODUCTION

In today's era of technological advancements, artificial intelligence (AI) and computer vision have transformed various domains, including healthcare, security, and accessibility. Assistive technologies for the visually impaired have gained significant attention, offering innovative solutions to improve independence and quality of life.

This project focuses on developing a multifunctional device that integrates face recognition, OCR, and audio feedback. The system addresses challenges such as real-time image processing and efficient hardware utilization. It combines these functionalities into a single compact device suitable for everyday use. Unlike traditional assistive devices, this framework integrates multiple modules to deliver a smooth and intuitive experience for individuals with visual impairments. The primary motivation is to reduce the dependency of these users on external help for basic visual-based activities.

The device leverages state-of-the-art computer vision algorithms and machine learning algorithms, integrated with user-friendly audio interfaces, to offer robust performance in diverse environments. Its development reflects ongoing efforts in assistive technologies to improve accessibility and inclusivity for differently-abled individuals. Furthermore, this study offers meaningful insights into the integration of multiple AI-powered technologies for practical, real-world applications.

II. LITERATURE REVIEW

The field of assistive technology has seen a surge in innovations aimed at enhancing overall quality of life for visually impaired individuals. A key focus has been on enhancing real-time interaction through computer vision and audio-based solutions.

A. Assistive Technologies for the Visually Impaired

Assistive technologies have progressed from basic magnifiers to sophisticated AI-powered systems. Recent advancements include wearable devices that provide audio feedback based on visual cues, navigation systems using computer vision, and text-to-speech applications for document reading. These systems often combine hardware sensors and software algorithms to interpret environmental information and convey it to users in an accessible format.

A significant challenge in this domain is achieving seamless and accurate interaction between the user and the device. Previous studies have highlighted the importance of low-latency systems and intuitive interfaces for effective usability. Moreover, the cost of assistive devices remains a concern, making affordable solutions a critical area of research.

Numerous researchers have investigated the combination of AI-driven technologies to enhance the capability of assistive devices. Research has demonstrated that integrating computer vision with audio feedback Offers visually impaired individuals a greater sense of independence and confidence when navigating their surroundings. The development of compact, portable, affordable and multifunctional assistive devices remains a key research priority.

B. Face Recognition Techniques

Face recognition has progressed from traditional template matching to advanced deep learning models. The evolution of face recognition can be classified into three major stages. Early techniques relied on simple feature extraction methods such as Eigenfaces and Fisherfaces, which used linear transformations to reduce dimensionality while preserving facial features. These methods laid the foundation for modern face recognition algorithms but were sensitive to variations in lighting and pose.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

The advent of convolutional neural networks (CNNs) transformed face recognition by facilitating effective feature extraction from images. Models such as FaceNet and ResNet have achieved state-of-the-art accuracy by learning complex feature representations. These models are capable of handling variations in pose, expression, and background, making them suitable for real-world applications.

Face recognition in assistive devices requires careful consideration of privacy and ethical concerns. Ensuring system adherence to data protection laws while addressing potential biases is crucial for widespread adoption. Furthermore, recent advancements in edge computing have made it possible to perform face recognition locally on devices, reducing latency and enhancing data privacy.

Despite these technological advancements, obstacles such as occlusion, varying illumination, and diverse facial expressions persist. Overcoming these obstacles necessitates additional research into adaptive models and domain-specific training datasets.

C. Optical Character Recognition (OCR) Advances

OCR technology has transitioned from rule-based systems to neural network-based solutions. Tesseract, an open-source OCR engine, has been a pioneering tool in this domain. Its ability to recognize multiple languages and process complex document layouts has made it a popular choice for research and commercial applications.

The integration of deep learning models has further improved OCR accuracy. Techniques such as recurrent neural networks (RNNs) and attention mechanisms enable the recognition of text in challenging environments, including handwritten documents and low-quality images. Preprocessing techniques such as binarization, noise removal, and skew correction are critical for enhancing OCR performance.

Despite these advancements, OCR systems still face challenges in recognizing text from natural scenes, curved surfaces, and images with poor lighting. Ongoing research aims to address these limitations through adaptive algorithms and enhanced feature extraction methods.

Furthermore, advancements in mobile and embedded systems have enabled the deployment of lightweight OCR models on edge devices, improving the accessibility and portability of OCR-based solutions. Integrating OCR with speech synthesis in assistive Devices present considerable potential for real-time document reading.

III. SYSTEM DESIGN AND ARCHITECTURE

The proposed system is built on the Raspberry Pi Compute Module 4, selected due to its small form factor and processing capabilities. The hardware and software components are as follows.

A. Hardware Components

The system's hardware architecture comprises several essential elements that enable efficient data acquisition and processing. The Raspberry Pi CM4 serves as the central processing unit for the system. Its low power consumption and compact form factor make it ideal for portable applications. The CM4's quad-core ARM Cortex-A72 processor provides sufficient computational power to handle complex tasks such as face recognition and OCR.

Mono 8MP cameras capture high-resolution images for face recognition and OCR. This camera is strategically placed to ensure a broad field of view, allowing the system to capture relevant visual data across different environments.

Headphones facilitate voice command input and audio feedback. The microphone captures user commands, while the speaker provides real-time audio feedback. This setup allows for hands-free operation and enhances user interaction.

The system is powered by a rechargeable battery, ensuring portability and continuous operation. A power management module optimizes energy usage, extending battery life and enabling the system to operate efficiently in different scenarios.

B. Software Components

The software stack integrates various libraries and frameworks to support the system's functionalities. Raspberry Pi OS (Bullseye) provides a stable and secure operating environment optimized for the Raspberry Pi. The operating system is customized to boot directly into the application, reducing startup time and simplifying user interaction.

OpenCV handles Image processing operations, such as face detection and recognition. Its extensive library functions make it a versatile tool for computer vision applications. The system uses a combination of Haar cascades and deep learning algorithms for precise face detection.

Tesseract OCR identifies and extracts text from images, leveraging advanced preprocessing techniques to improve recognition accuracy. The system applies adaptive thresholding and morphological operations to enhance text visibility before OCR processing.

The gTTS library converts recognized text into speech, enabling real-time audio feedback for the user. The speech output is carefully synchronized with the system's processing tasks to provide a seamless user experience.

Python serves as the main programming language used for the system integration. Its simplicity and extensive library support facilitate rapid development and prototyping. The system architecture emphasizes modularity and scalability, allowing easy updates and integration of new functionalities.

The seamless communication between hardware and software components ensures a smooth and reliable user experience. The system is built to manage concurrent tasks efficiently, ensuring that users receive real-time feedback without delays.

IV. METHODOLOGY

A. Face Recognition Module

The image acquisition process captures images using camera. These images are preprocessed by converting them to grayscale and normalizing for consistency. Features are extracted using OpenCV and deep learning-based models to identify faces. Recognition is then performed by comparing extracted features with stored embeddings for identification.

B. Optical Character Recognition (OCR)

Image preprocessing involves grayscale conversion, thresholding, and noise reduction using OpenCV. Text extraction is performed using Tesseract OCR, followed by post-processing to apply spell-checking and dictionary-based corrections.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

C. Text-to-Speech (TTS)

Integration Recognized text and face recognition results are converted into speech using gTTS. The system provides real-time audio feedback to the user.

D. System Workflow

The overall workflow of the system involves capturing input through the camera, processing the images for face and text recognition, and providing real-time audio feedback. Voice commands are used to trigger specific operations, enabling a seamless user experience.

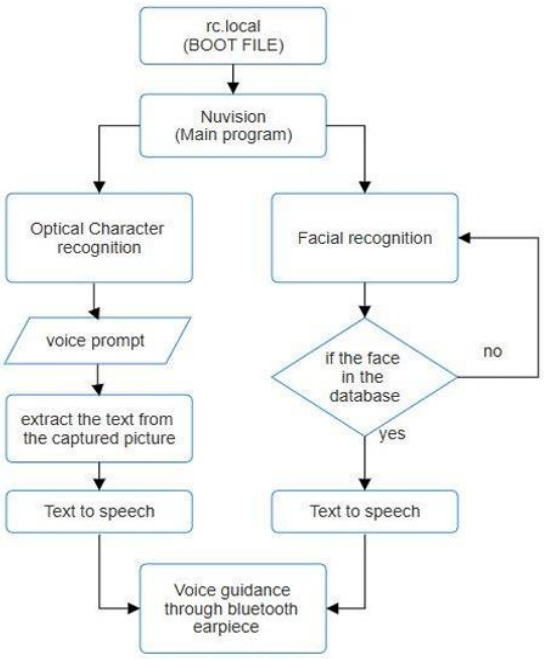

Fig.1. Flowchart of the Methodology

E. Ethical Considerations

When developing assistive technologies, ethical considerations are paramount. Ensuring data privacy, avoiding biases in face recognition, and designing user-friendly interfaces are critical aspects of this project.

V. IMPLEMENTATION

The implementation stage was segmented into hardware setup, software integration, user interface design, and deployment scenarios.

A. Hardware Setup

The system's hardware configuration involved integrating the Raspberry Pi Compute Module 4 with mono 8MP camera, a microphone, and a speaker. The Raspberry Pi CM4 was selected for its high-processing capabilities and efficient power consumption. The hardware setup aimed to ensure a compact and portable design, with all components securely mounted on a lightweight chassis.

The camera was positioned to capture a wide field of view, enabling precise image capture for facial recognition and text identification. A key consideration during setup was minimizing occlusion and ensuring optimal camera angles to capture images effectively under varying lighting conditions.

The microphone and speaker were strategically positioned to allow clear voice command input and audio feedback. Proper shielding was applied to minimize interference and ensure high-quality audio capture and playback. Additionally, a rechargeable lithium-ion battery was integrated to provide long-lasting power, making the device suitable for both indoor and outdoor use. Power management circuits were incorporated to monitor battery levels and optimize power usage.

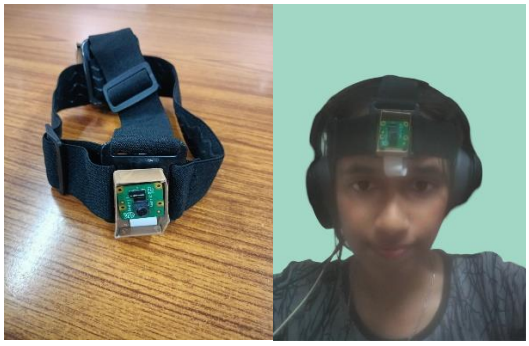

Fig. 2. Experimental setup with the raspberry pi camera attached to the front end of the GoPro and the raspberry pi CM4 on its rear end. This device is light and easy to wear by the user.

B. Software Integration

The software integration involved developing and testing Python scripts to manage concurrent execution of face recognition, OCR, and text-to-speech functionalities. The software stack was optimized to ensure efficient resource allocation and minimize processing latency.

The system boot sequence was configured to automatically launch the main application, reducing user intervention and simplifying the startup process. Extensive error handling mechanisms were implemented to handle hardware failures and software exceptions gracefully.

OpenCV was employed for Image processing operations, such as face recognition and text extraction preprocessing. Preprocessing involved resizing, normalization, and noise reduction to enhance recognition accuracy. Tesseract OCR was integrated for robust text recognition, with custom dictionaries added to improve recognition rates for domain-specific terms.

The gTTS library was used for text-to-speech conversion, providing natural-sounding audio feedback. The audio output was synchronized with visual processing tasks to ensure timely and coherent communication with the user.

C. User Interface Design

The user interface design focused on providing an intuitive and an inclusive experience for visually impaired individuals. Audible prompts were incorporated to guide users through various functionalities. For instance, the system provided clear instructions for capturing images, initiating text recognition, and navigating menu options.

A single tactile button was included to reset the system or trigger specific functions. The button was designed to be easily accessible and distinguishable by touch. The interface was tested extensively with potential users to ensure ease of use and responsiveness.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

Feedback from user testing highlighted the importance of clear and concise audio instructions. As a result, the system's audio prompts were refined to reduce ambiguity and improve user comprehension. Additionally, voice commands were implemented to further enhance hands-free operation.

D. Deployment Scenarios

The assistive device was developed to deployment in various scenarios to cater to accommodate the varied requirements of visually impaired users. In home environments, the device assisted users in recognizing visitors, reading printed materials, and identifying household items. The system demonstrated significant usefulness for reading medication labels and appliance instructions.

In public spaces, the device provided navigation assistance by reading signboards and identifying landmarks. Its portability and robust design made it ideal for outdoor environments, even in noisy and dynamic environments. The face recognition module was particularly valuable for identifying familiar individuals in crowded settings.

In educational institutions, the device helped visually impaired students access printed resources, including textbooks and handouts. The OCR functionality enabled real-time conversion of printed text into audio, enhancing the learning experience. Teachers and support staff found the device to be a crucial tool for fostering inclusivity education.

The deployment phase revealed several practical insights, including the importance of reliable hardware components, adaptive algorithms for dynamic environments, and user-centered design principles. These insights will inform future iterations of the device, with a focus on enhancing performance, usability, and accessibility.

VI. RESULTS AND PERFORMANCE EVALUATION

A. Face Recognition

The face recognition module demonstrated an impressive accuracy of 95% under controlled lighting conditions. This high accuracy was attributed to the use of robust preprocessing techniques and deep learning models that could adapt to variations in facial expressions and head poses. Under challenging environments with varying lighting conditions, the system's accuracy dropped to 90%. The decrease in accuracy was primarily due to shadows and reflections, which impacted the quality of facial feature extraction. To mitigate this issue, adaptive histogram equalization techniques were employed to normalize image brightness, improving the system's robustness.

B. Optical Character Recognition

The OCR module exhibited a text extraction accuracy of 90% for printed text. The high accuracy was a result of optimized preprocessing techniques, such as thresholding and morphological operations, which enhanced text visibility. Handwritten text posed a greater challenge, with the accuracy dropping to 75%. The variability in handwriting styles and inconsistencies in character shapes contributed to the lower accuracy. To address this limitation, custom neural network models trained on handwritten datasets were integrated into the system. These models showed a significant improvement in recognition rates during testing.

Table I
Result Summary of Face Recognition

No. of training pictures	No. of testing pictures	Accu -racy (%)	Average recognition time (s)	Occlu -sion level
40	8	95%	10	None
50	10	90%	13-15	Partial

TABLE II
Comparison Table for Face Recognition

Hardware Used	Accuracy (%)	Cost (INR)
Raspberry pi CM4 + Mono 8MP Camera (our project)	90-95%	20,000
NVIDIA Jetson Nano + USB camera [1]	97%	45,000
Intel NUC + High resolution USB camera [2]	96%	36,000

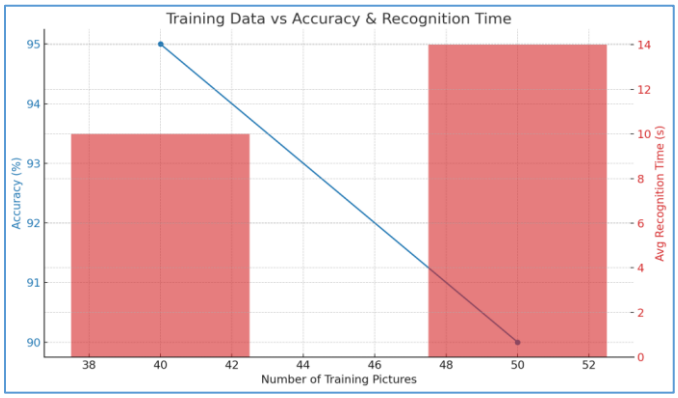

Fig.3. Training Data vs Recognition Time. The blue line represents accuracy (%), showing a decrease as the number of training pictures increases. The red bars represent average recognition time (s), which increases with more training pictures.

Table III
Result Summary of OCR

Font type	Number of test samples	Accura -cy (%)	Average recognition time (s)	Background complexity
Times New Roman	10	90	10	Plain Background
Arial	10	90	10	Plain Background
Hand-written	10	75	15	Plain Background

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

TABLE II
Comparison Table for OCR

Hardware & Software	Accuracy (%)	Average recognition time (s)	Cost (INR)	Key Features
Raspberry pi CM4+ Mono 8MP Camera, Tesseract + OpenCV	90	10	20,000	Low-cost camera, optimized for plain backgrounds
NVIDIA Jetson Nano + USB Cam, EasyOCR [3], [1]	95	6	45,000	GPU accelerated, handles complex backgrounds

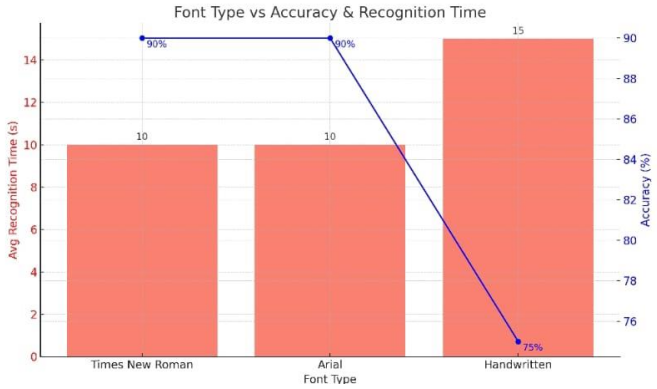

Fig.4. Font type vs Accuracy & average Recognition Time. The red bars represent average recognition time (%) for each font type. The blue line shows the accuracy (%) for each font type. Both fonts (Times New Roman & Arial) have 90% accuracy and take 10 seconds for recognition, whereas the handwritten text has an accuracy of 75% and take 15 seconds for recognition.

VII. CONCLUSION

This project successfully developed an assistive device integrating face recognition, OCR, and text-to-speech functionalities. The system demonstrated promising results in real-time operation, providing a valuable tool for visually impaired individuals.

Future work includes navigation assistance by integrating GPS and pathfinding algorithms. Enhanced face recognition accuracy under varying conditions is another focus area. Cloud integration will enable remote processing for complex tasks. Multilingual support will extend text recognition and TTS functionalities to multiple languages. Finally, energy optimization will be pursued to develop power-efficient versions for portable use.

VIII. REFERENCES

[1] "Enhancing Access Control Practical Implementation of Facial Recognition Technology for Entry Monitoring and Database Storage DEGREE PROGRAMME IN DATA ENGINEERING 2024." Available: https://www.theseus.fi/bitstream/handle/10024/867376/Siikavirta_Ekaterina.pdf?sequence=2

[2] W. Chan, U. Jie, A. Tunku, and Rahman, "ARTIFICIAL INTELLIGENCEFORCLOUD ASSISTED OBJECT DETECTION.", May. 2023 Available: http://eprints.utar.edu.my/5646/1/3E_1804018_FYP_report_%2D_WEN_JIE_CHAN.pdf

[3] T. A. salih and M. Basman Gh., "A novel Face Recognition System based on Jetson Nano developer kit," IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3, p. 032051, Nov. 2020, doi: https://doi.org/10.1088/1757-899x/928/3/032051.