ISSN 2319-5991 www.ijerst.com

Peer Reviewed,Referred & Indexed Journal

Research Paper

Image to Text Conversion with Voice Announcement using RaspberryPi

DHEEKONDA SWATHI[1], VOLLALA VAISHNAVI[2], SABAVATH CHAMANTHI[3], RATHLAVATH KALAVATHI[4], SALPALA CHANDU[5]

[1]Assistant Professor, Dept of ECE, TKR College of Engineering and Technology

[2,3,4,5] Student, Dept of ECE, TKR College of Engineering and Technology

ABSTRACT

This project presents the design and implementation of an image-to-text conversion system with voice announcement using a Raspberry Pi, aimed at providing an efficient and low-cost assistive solution. The system captures images through a camera module and applies image preprocessing techniques to enhance quality for accurate recognition. Optical Character Recognition (OCR) is utilized to extract textual information from the captured images, which is then converted into speech using a text-to-speech (TTS) engine. The generated audio output is delivered through a speaker, enabling users to hear the extracted content in real time. This system is particularly beneficial for applications involving Smart reading system, Document digitization, Translation systems. The integration of image processing, OCR, and speech synthesis technologies ensures a compact, portable, and user-friendly solution. However, the system's accuracy is influenced by factors such as lighting conditions, image clarity, and font variations. Future enhancements may include support for multiple languages, improved recognition accuracy using advanced algorithms, and real-time video processing capabilities, making the system more robust and versatile.

INTRODUCTION

In the modern digital era, extracting and interpreting information from images has become increasingly important due to the vast amount of textual data embedded in printed documents, photographs, scanned files, and handwritten notes. Such data is not readily accessible in editable or audible formats, creating challenges for automation and accessibility, especially for visually impaired individuals. The project "Image to Text Conversion with Voice Announcement" addresses this issue by integrating image processing, Optical Character Recognition (OCR), and speech synthesis technologies. It captures or accepts images containing text, processes them to extract meaningful information, and converts the extracted content into audible speech, thereby bridging the gap between visual and auditory communication.

Image detection and preprocessing form the foundation of this system. The system captures real-time images using a camera and identifies regions containing textual information. To enhance accuracy, preprocessing techniques such as grayscale conversion, noise reduction, contrast enhancement, and thresholding are applied. These steps simplify the image, reduce distortions, and improve text visibility under varying lighting conditions. Additional methods like edge detection and region segmentation help isolate text from complex backgrounds, ensuring that only relevant and clear text regions are forwarded for further processing. This structured preparation significantly

improves the quality and reliability of the input for text recognition.

Optical Character Recognition (OCR) serves as the core component for extracting text from images. In this project, the Tesseract OCR engine is used to analyze preprocessed images and convert detected characters into machine-readable text. The system identifies patterns, segments text into lines and words, and compares them with stored character models for accurate recognition. It also handles variations in font styles, sizes, and orientations, ensuring robust performance even in slightly distorted conditions. Post-processing techniques such as error correction and removal of unwanted symbols further refine the output, resulting in meaningful and usable text for subsequent stages.

The final stage of the system involves converting the extracted text into speech using Text-to-Speech (TTS) technology. Tools such as pyttsx3 or gTTS are employed to generate natural-sounding voice output from the recognized text. The TTS engine processes the text by breaking it into linguistic units and applying pronunciation and language rules to produce intelligible audio. This feature greatly enhances accessibility, allowing users—especially those with visual impairments—to listen to the content instead of reading it. Additionally, the system improves productivity by reducing manual data entry and enabling quick information retrieval, making it valuable in applications such as education, digital archiving, and real-time text recognition.

1. LITERATURE REVIEW

The conversion of image-based text into machine-readable and audible formats has been extensively studied in the fields of image processing and artificial intelligence, with significant contributions to the development of Optical Character Recognition (OCR) and Text-to-Speech (TTS) systems. Early OCR techniques relied on pattern matching and template-based methods, which performed well on structured and high-quality printed documents but faced limitations when dealing with variations in font styles, sizes, and image quality. With the advancement of machine learning and deep learning, modern OCR systems have achieved substantial improvements in accuracy and flexibility, enabling the recognition of complex patterns such as handwritten text and text in natural scenes. To further enhance performance, researchers have incorporated image preprocessing techniques including noise reduction, grayscale conversion, thresholding, and edge detection, along with segmentation methods that isolate characters and words for more precise recognition.

In parallel, TTS technology has evolved from basic rule-based systems to advanced models capable of generating natural and human-like speech using concatenative synthesis and neural network-based approaches. Modern TTS systems support multiple languages and voice styles, making them suitable for diverse applications. Several existing solutions integrate OCR and TTS to assist visually impaired users by capturing images, extracting text, and delivering real-time audio output. However, challenges such as low-resolution images, complex backgrounds, and multilingual text handling still persist. Recent advancements include the use of cloud-based OCR services and open-source frameworks, which provide improved accuracy, faster processing, and easier implementation. Despite these developments, further improvements are needed in terms of system accuracy, speed, and adaptability. This project builds upon existing research by integrating OCR and TTS into a unified system aimed at delivering reliable text extraction and clear voice output, thereby contributing to the advancement of accessible and intelligent technologies.

2. DESIGN METHODOLOGY

The design methodology of the proposed “Image-to-Text Conversion with Voice Announcement” system follows a structured and modular approach to ensure efficient implementation and reliable performance. Initially, system requirements are identified with a focus on accessibility, accuracy, cost-effectiveness, and real-time operation. Based on these requirements, appropriate hardware components such as Raspberry Pi, camera module, and audio output devices are selected, along with software tools including image processing libraries, an OCR engine, and a Text-to-Speech (TTS) engine.

The system is organized into functional modules, namely the Input Module, Image Preprocessing Module, OCR Processing Module, and TTS Output Module. The Input Module captures images containing textual information, which are then enhanced using preprocessing techniques such as grayscale conversion, noise reduction, and contrast adjustment to improve readability. The processed image is forwarded to the OCR module, where segmentation and character recognition techniques are applied to extract machine-readable text.

The extracted text undergoes basic post-processing to ensure accuracy and proper formatting before being passed to the TTS module. The TTS module converts the text into audible speech using a speech synthesis engine, and the generated output is delivered through speakers or headphones. Finally, the system is tested under various conditions, including different lighting environments and text formats, to evaluate performance and ensure reliability. This modular design approach enhances system flexibility, simplifies maintenance, and supports future enhancements.

3. PRINCIPLE OF OPERATION

The principle of the proposed "Image-to-Text Conversion with Voice Announcement" system is based on the integration of image processing, Optical Character Recognition (OCR), and Text-to-Speech (TTS) technologies to transform visual textual information into audible output. The system operates by first capturing an image containing text using a camera, which serves as the primary input. The captured image is then preprocessed using techniques such as grayscale conversion, noise reduction, and contrast enhancement to improve the clarity and readability of the text.

Following preprocessing, the OCR module analyzes the image to detect, segment, and recognize characters by comparing patterns with stored models, converting them into machine-readable text. This extracted text is then passed to the TTS module, where it is processed and synthesized into speech using predefined linguistic and acoustic models. The generated audio output is delivered through speakers or headphones, enabling users to listen to the content of the image.

Thus, the system fundamentally works on the principle of converting visual data into digital text and subsequently into speech, ensuring an efficient flow of information from image acquisition to audio output. This integrated approach enables effective communication between visual inputs and auditory outputs, making the system particularly useful for accessibility and real-time applications.

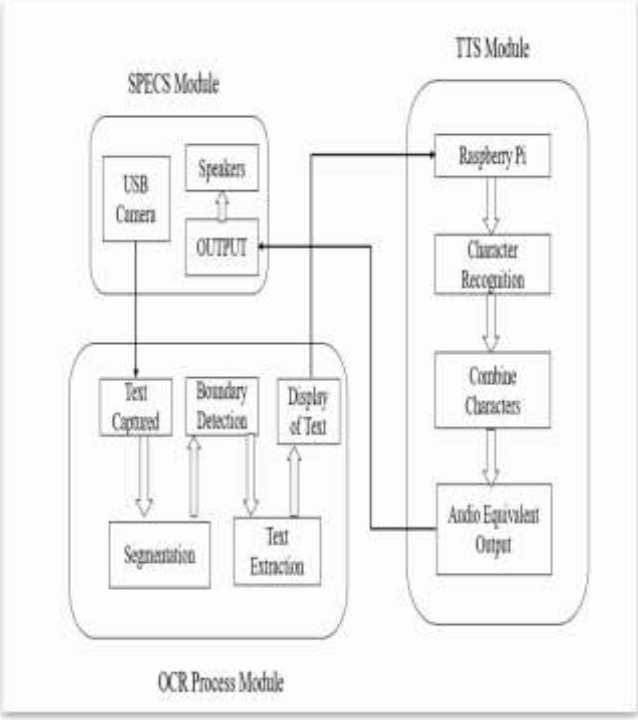

Fig .3.1 Block Diagram

3.1 Overview of the System

The proposed system, "Image-to-Text Conversion with Voice Announcement," converts text present in images into audible speech by integrating image processing, Optical Character Recognition (OCR), and Text-to-Speech (TTS) technologies. It captures images using a camera, enhances them through preprocessing techniques, extracts textual content using OCR, and converts the recognized text into speech output.

The system is designed with a modular architecture consisting of input, processing, and output stages, ensuring flexibility and ease of implementation. It provides a cost-effective, low-power, and user-friendly solution capable of

real-time operation. Overall, the system enhances accessibility and efficiency, making it highly useful for assistive applications and real-world information retrieval.

3.2 Input Module

The Input Module is the initial stage of the Image-to-Text Conversion with Voice Announcement system, responsible for collecting the data required for processing. The input is provided in the form of an image containing textual information, captured using a USB camera or Raspberry Pi camera module. This module acts as an interface between the real-world environment and the system.

Image Acquisition: The system captures images from the surrounding environment using the camera, ensuring that textual content is clearly visible. The captured image serves as the raw input for subsequent processing stages.

Importance of Image Quality: The quality of the captured image significantly affects system performance. Factors such as lighting conditions, focus, distance from the text, and background clarity influence the accuracy of text detection and recognition. Proper camera positioning and adequate illumination are essential for obtaining clear images.

Initial Image Processing: After capturing, the image is transferred to the processing unit in digital form. Basic operations such as resizing and format conversion may be applied to prepare the image for the OCR stage, improving efficiency and reducing recognition errors.

Role in System Workflow: The Input Module forms the foundation of the system by providing the necessary visual data. A high-quality input image leads to better text recognition accuracy and enhances the clarity of the final voice output, making this module a critical component of the overall system.

3.3 OCR Process Module

OCR Process Module Overview: The OCR (Optical Character Recognition) Process Module is the core component of the Image-to-Text Conversion with Voice Announcement system. It is responsible for analyzing the captured image and converting visual text into a machine-readable format, serving as the main processing unit of the system.

Image Segmentation: The process begins by dividing the input image into smaller sections such as lines, words, or individual characters. This segmentation simplifies analysis by isolating different text elements, enabling more accurate processing.

Boundary Detection: The module identifies the exact regions in the image where text is present, allowing the system to focus only on relevant areas while ignoring unnecessary background information.

Text Extraction: The OCR engine analyzes the segmented regions, recognizes characters based on their shapes and patterns, and converts them into digital text. The extracted text can then be stored or displayed for further use.

Role in System Workflow: The OCR Process Module bridges the gap between image input and speech output. Its accuracy and efficiency directly impact the overall system performance, as precise text recognition is essential for generating clear and meaningful audio output in the next stage.

3.4 TTS (Text-to-Speech) Module:

TTS Module Overview: The TTS (Text-to-Speech) Module is the final stage of the Image-to-Text Conversion with Voice Announcement system. It is responsible for converting the extracted digital text into audible speech, providing the final output in an easily understandable form.

Text Structuring: In this stage, the recognized text is processed and arranged into meaningful words and sentences. This ensures that the speech output follows the correct sequence and is clear, logical, and easy for users to understand.

Speech Synthesis Process: The structured text is passed to a text-to-speech engine such as pyttsx3, which converts the text into an audio signal using predefined voice models. The use of an offline speech engine makes the system reliable and suitable for real-time applications without requiring internet connectivity.

Audio Output Delivery: The generated speech is transmitted to output devices such as speakers or headphones, allowing the user to listen to the extracted text. Parameters such as speech speed and volume can be adjusted to improve clarity and user comfort.

Role in System Workflow: The TTS Module completes the overall system operation by transforming recognized text into voice output, enabling effective communication between the system and the user, particularly for accessibility-based applications.

4. Results and Discussion

4.1 OUT PUT

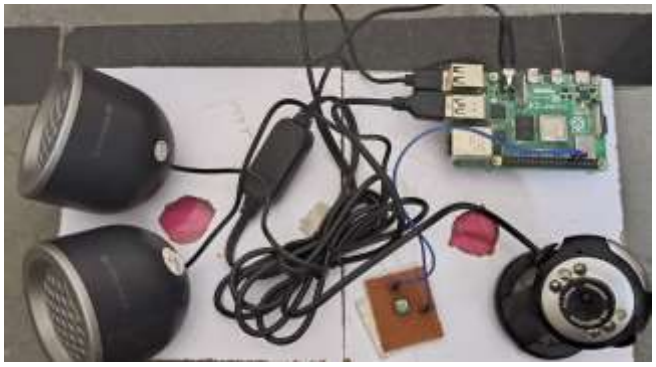

Figure 4.1 Out Put

The proposed "Image-to-Text Conversion with Voice Announcement" system was successfully implemented and tested under various conditions. The system effectively captured images, extracted text using OCR, and converted it into audible speech with satisfactory performance. Higher accuracy was observed for clear, well-lit images with standard fonts, while minor errors occurred in low-quality or complex backgrounds.

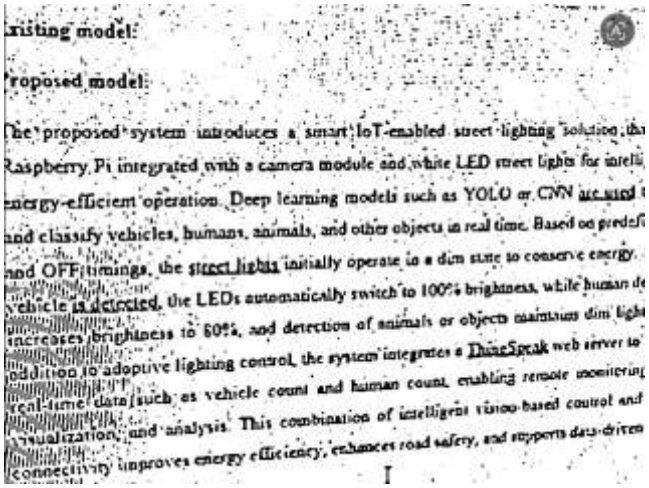

Figure 4.2 Captured Text from Image

The Text-to-Speech (TTS) module generated clear and understandable audio output, and the use of an offline engine ensured reliable real-time performance. Overall, the system demonstrated efficient operation with minimal delay and proved to be practical for assistive and real-world applications, although further improvements can enhance accuracy and speech quality.

ADVANTAGES

1. Improves Accessibility: Converts printed text into speech, enabling visually impaired individuals to access written information independently and enhancing usability.
2. Cost-Effective Solution: Utilizes Raspberry Pi and basic peripherals such as a webcam and speakers, making the system affordable and suitable for low-budget and educational applications.
3. Low Power Consumption: Operates on a low-power device, ensuring energy efficiency and making it ideal for continuous and portable use.
4. User-Friendly Operation: Incorporates a simple push-button interface for image capture, allowing easy operation even for non-technical users.
5. Modular and Flexible Design: Built using separate modules such as image acquisition, OCR, and text-to-speech, enabling easy maintenance, upgrades, and system modifications.
6. Real-Time Processing Capability: Processes images and generates speech quickly, making it suitable for real-time applications like reading books, labels, and signboards.
7. Wide Range of Applications: Applicable across multiple domains including assistive technology, education, automation, and smart devices, enhancing its overall practical utility.

APPLICATIONS

1. Assistive Technology for Visually Impaired: Enables visually challenged users to read printed text through voice output, improving independence and accessibility.
2. Reading Printed Documents: Can be used to read books, newspapers, notices, and official documents easily by converting text into speech.
3. Educational Support: Helps students access study materials through audio, making learning more inclusive and convenient.
4. Public Information Systems: Applicable in places such as railway stations and airports to read signboards, instructions, and announcements.
5. Retail and Product Label Reading: Assists users in reading product labels, price tags, and packaging details in stores.
6. Healthcare Applications: Useful for reading prescriptions, medicine labels, and medical instructions accurately.

CONCLUSION

The project successfully demonstrates an effective approach for converting image-based text into audible speech through the integration of both hardware and software modules. It offers a simple, cost-effective, and user-friendly solution, making it particularly suitable for assistive applications aimed at enhancing accessibility. The system meets its intended objectives with satisfactory performance, validating the feasibility of combining image processing, Optical Character Recognition (OCR), and Text-to-Speech (TTS) technologies into a unified framework.

Although certain limitations remain, the overall implementation proves to be practical and efficient for real-world usage scenarios. The results highlight the significant potential of integrating OCR and TTS technologies to develop intelligent systems capable of bridging the gap between visual and auditory information. This work contributes to the advancement of accessible technology and provides a strong foundation for future improvements and extensions in this domain.

FUTURE SCOPE

The proposed system can be further enhanced by incorporating advanced image processing techniques to improve the accuracy and robustness of Optical Character Recognition (OCR), particularly in challenging conditions such as low-resolution images and complex backgrounds. Expanding the system to support multiple languages would significantly increase its usability and applicability across diverse user groups and regions. Additionally, the Text-to-Speech (TTS) module can be upgraded with more sophisticated speech synthesis models to produce more natural, fluent, and human-like voice output, thereby improving the overall user experience.

Further improvements may include integration with mobile platforms and wireless technologies, enabling greater portability, real-time processing, and ease of access. The inclusion of intelligent features such as voice command functionality and automatic text detection can enhance system interactivity and reduce user effort. Collectively, these enhancements would contribute to the development of a more efficient, flexible, and user-friendly system, making it suitable for a wider range of real-world applications.

REFERENCES

1. Wydyanto, N., & Nayan, N. M., *"Converting Image Text to Speech Using Raspberry Pi,"* Journal of Information Systems Engineering and Management, 2025.
2. Dahiphale, S. V., & Nandedkar, S. J., *"Raspberry Pi & OCR-Based Image-Text to Speech Conversion,"* Journal of Electronics and Telecommunication System Engineering, 2024.
3. Sarak, M., Patil, S. S., & Mali, A. S., *"Image Text to Speech Conversion using OCR Technique in Raspberry Pi,"* International Journal of Engineering and Management Research, 2024.
4. Devi, P. L., et al., *"Raspberry Pi-Based Text-to-Speech Conversion for Blind Persons,"* International Conference on ARIIA, 2024.

5. Sisodia, P., & Rizvi, S. W. A., *"Optical Character Recognition Development Using Python,"* Journal of Informatics Electrical and Electronics Engineering, 2023.
6. Alsayaydeh, J., et al., *"Handwritten Text Recognition using Raspberry Pi with OpenCV and TensorFlow,"* International Journal of Electrical and Computer Engineering, 2025.
7. Tesseract OCR Official Documentation, Available at: https://tesseractocr.org/
8. Python Software Foundation, *Python Documentation*, Available at: https://www.python.org