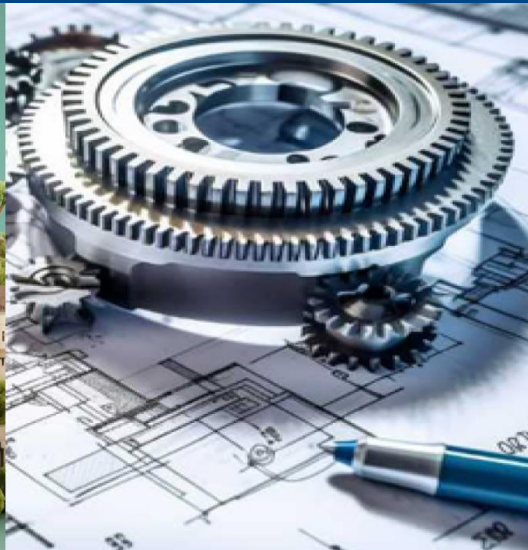
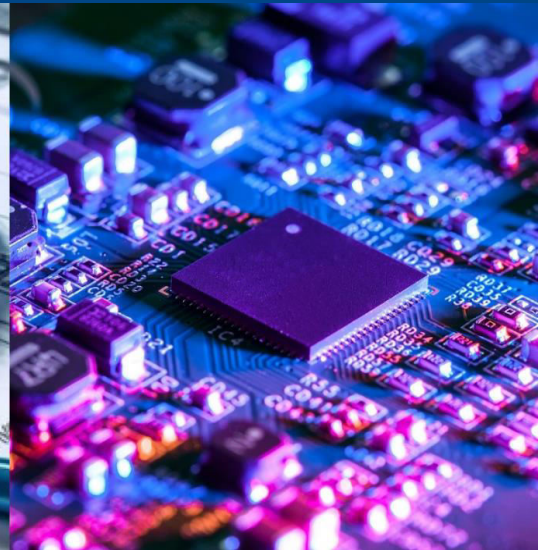
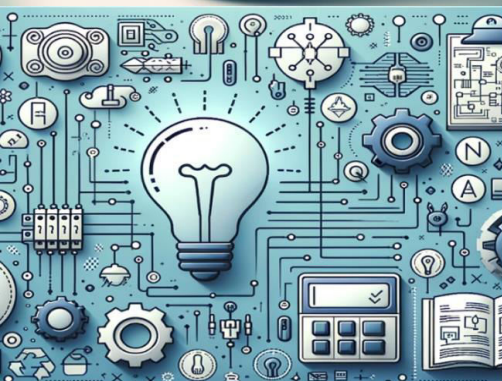




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.864

Volume 9, Issue 10, October 2026



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Image Text to Speech Conversion using Optical Character Recognition Technique in Raspberry Pi

Mangesh Sarak, Prof A. S. Mali

Student, Department of Electronics and Telecommunication Engineering, Tatyasaheb Kore Institute of Engineering and Technology (An Autonomous Institute), Warananagar, Kolhapur, Maharashtra, India

Professor, Department of Electronics and Telecommunication Engineering, Tatyasaheb Kore Institute of Engineering and Technology (An Autonomous Institute), Warananagar, Kolhapur, Maharashtra, India

ABSTRACT: Optical Character Recognition (OCR) is an important application of artificial intelligence and computer vision that enables computers to identify and interpret text from images. In this project, a Raspberry Pi is used as the main processing platform to capture images containing printed or handwritten text and convert the recognized content into speech. This approach can be particularly useful for visually impaired individuals by providing an audible representation of written information.

The proposed system uses image-processing techniques to identify the region containing text and prepare the image for recognition. Image preprocessing and text localization are performed to improve the quality of the input before applying OCR. Techniques such as grayscale conversion, noise reduction, thresholding, edge detection, and text-region extraction can be used to separate characters from the background. The processed text regions are then supplied to the Tesseract OCR engine, which recognizes the characters and converts them into digital text.

Python programs running on the Raspberry Pi control the image acquisition, preprocessing, OCR recognition, and text-to-speech stages. Once the characters are recognized, the resulting text is passed to a speech-synthesis module. The system then produces an audible output through a connected speaker, allowing the user to listen to the recognized information.

The overall system therefore provides a low-cost and portable solution for converting text from images into speech. It demonstrates how Raspberry Pi, Python, computer vision, OCR, and text-to-speech technologies can be integrated into an assistive system for people who have difficulty reading conventional printed or handwritten material.

I. INTRODUCTION

In today's digital world, access to information is essential for everyone, including visually impaired individuals and people who face difficulties in reading printed text. Advances in embedded systems, image processing, and artificial intelligence have made it possible to develop assistive technologies that can convert visual information into audible speech. One such technology is the integration of Optical Character Recognition (OCR) and Text-to-Speech (TTS) systems on compact computing platforms like the Raspberry Pi.

Optical Character Recognition (OCR) is a technique used to extract textual information from images, scanned documents, and photographs. Once the text is recognized and converted into machine-readable format, a Text-to-Speech (TTS) engine can transform the extracted text into spoken words. By combining these technologies, users can listen to the contents of printed documents, signboards, books, and other text-based materials without the need for manual reading.

The Raspberry Pi serves as an ideal platform for implementing this system due to its low cost, portability, energy efficiency, and ability to interface with cameras and audio devices. In this project, a camera captures an image



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

containing text, the OCR module processes the image to recognize the characters, and the TTS module converts the recognized text into speech output. This solution provides an effective and affordable assistive tool for visually impaired users and can also be applied in educational, industrial, and automation environments.

Visual impairment is a significant challenge that affects the ability of individuals to access written information independently. Since textual information is present in almost every aspect of daily life—including newspapers, product labels, road signs, advertisements, documents, and electronic displays—people with severe visual impairment may face considerable difficulty in reading and interpreting such content.

Recent developments in computer vision, optical character recognition (OCR), and speech-processing technologies have provided new possibilities for assistive applications. An OCR-based system can capture an image containing text, identify the text within the image, convert it into machine-readable characters, and subsequently transform the recognized text into speech. Such a system can allow visually impaired users to obtain information without depending entirely on assistance from another person.

Several existing systems have been developed for this purpose; however, many experimental approaches work effectively only when the text is presented on a simple, high-contrast background, such as black characters printed on a white sheet. Images containing complex backgrounds, uneven illumination, different text orientations, advertisements, signboards, or other visual elements can significantly reduce recognition accuracy. Therefore, preprocessing and accurate identification of the text-containing region are important for developing a practical real-time system [1–3].

Another challenge is ensuring that the recognized characters are processed as meaningful words rather than being treated as isolated letters. Reading individual characters separately can result in unnatural or confusing audio output. To overcome these limitations, the proposed system focuses on identifying the region containing the required text before performing OCR.

The system uses image-processing techniques to locate the relevant text area within an input image. A rectangular or quadrilateral region containing the text can be identified and isolated from the surrounding background. The selected region is then cropped and geometrically transformed to obtain a clearer representation of the text. Further preprocessing, including edge detection and image enhancement, can be applied to improve the visibility of the characters and facilitate accurate recognition.

After preprocessing, the enhanced image is supplied to the OCR engine to convert the visual characters into digital text. The recognized text is subsequently passed to a Text-to-Speech (TTS) module, which generates an audible representation through a speaker. The complete process therefore integrates image acquisition, text-region detection, image preprocessing, OCR, and speech synthesis to provide an assistive solution for accessing written information.

The proposed approach is intended to improve the usability of OCR-based systems in real-world environments where text may occur on non-uniform or visually complex backgrounds. By combining these technologies on a compact computing platform such as Raspberry Pi, the system can provide a portable and potentially cost-effective method for converting printed information into speech.

The main objective of this project is to develop a real-time image-to-speech conversion system that accurately recognizes text from images and delivers clear audio output. The system aims to enhance accessibility, improve user independence, and demonstrate the practical application of OCR and speech synthesis technologies using Raspberry Pi.

II. PROBLEM STATEMENT

1. **Milestone 1** – Warping and cropping the image.
2. **Milestone 2** – Edge detection and drawing a boundary over the letters.
3. **Milestone 3** - Processing the image by the OCR and TTS to give audio output.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Objectives

- To capture images containing printed text using a camera connected to the Raspberry Pi.
- To preprocess the captured images and enhance their quality for accurate text recognition.
- To implement an OCR system for extracting text from images and converting it into machine-readable format.
- To integrate a Text-to-Speech (TTS) engine that converts the recognized text into clear and understandable speech.
- To provide an assistive solution for visually impaired individuals by enabling them to access printed information through audio output.
- To develop a low-cost, portable, and user-friendly embedded system using Raspberry Pi.
- To improve the accuracy and efficiency of text recognition and speech generation in real-time applications.
- To demonstrate the practical application of image processing, OCR, and speech synthesis technologies in embedded systems.
- To enable users to read documents, books, labels, signboards, and other printed materials without manual reading.
- To create a reliable and scalable platform that can be further enhanced with support for multiple languages and advanced image recognition features.
- Communication Tool: To provide an effective means of communication for individuals with visual impairment.
- Education: To support visually impaired children by providing accessible and effective educational opportunities.
- Learning Tool: To develop an interactive learning aid that enables users to study independently in a supportive environment.
- Economical: To provide an engaging and enjoyable environment for learning Braille writing through a simple and cost-effective approach.

III. SYSTEM DESIGN

Introduction to System Design

The proposed system is an embedded assistive solution that converts printed or handwritten text from images into audible speech using Optical Character Recognition (OCR) and Text-to-Speech (TTS) technologies on a Raspberry Pi platform. The system is designed to be portable, low-cost, and efficient, making it suitable for real-time applications, especially for visually impaired users. The architecture integrates hardware components (camera, Raspberry Pi, audio output) with software modules (image processing, OCR engine, and speech synthesis).

Block Diagram:

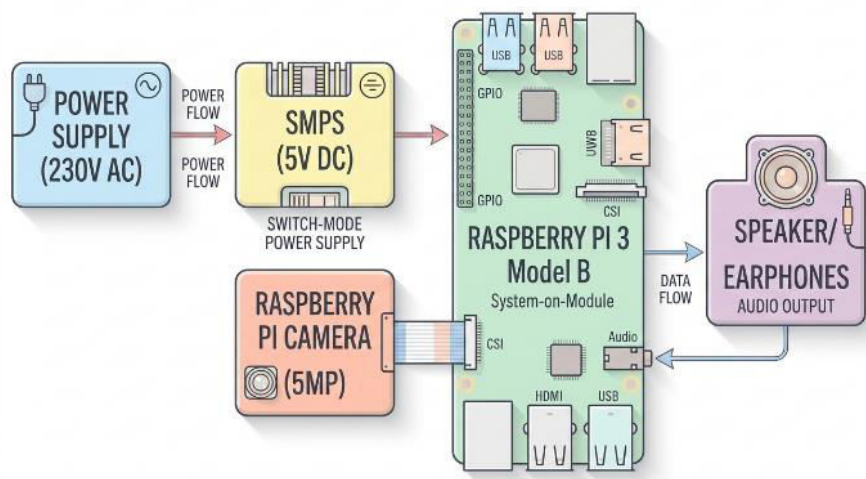


Fig No. 01: Block Diagram of proposed system.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Overall System Architecture

The system processes data step-by-step, with each module passing its output to the next stage.

Flow of System:

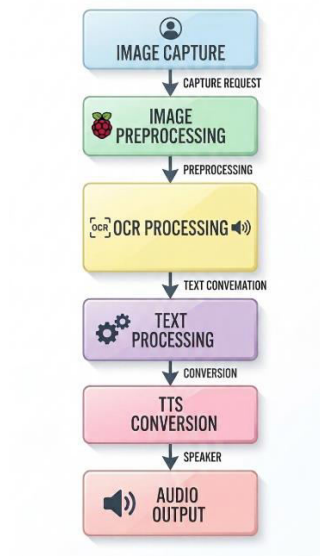


Fig No. 02: Flow of System

Each stage is optimized to improve accuracy, reduce noise, and ensure smooth conversion from image to speech.

IV. RESULTS AND DISCUSSION

• Step to start Image Text-to-Speech Conversion Using OCR Technique in Raspberry Pi

The following steps describe the complete procedure for implementing an Image Text-to-Speech Conversion System using Optical Character Recognition (OCR) on Raspberry Pi.

Step 1: Assemble the Hardware

Connect all the required hardware components.

Required Components

- Raspberry Pi 4/5
- Raspberry Pi Camera Module v2.1
- MicroSD Card (32 GB or higher)
- USB Type-C Power Supply
- USB Multimedia Speaker
- HDMI Monitor (Optional)
- Keyboard and Mouse (Optional)

Step 2: Install Raspberry Pi OS

1. Download **Raspberry Pi OS**.
2. Flash the operating system onto the microSD card using Raspberry Pi Imager.
3. Insert the microSD card into the Raspberry Pi.
4. Connect the power supply.
5. Complete the initial system setup:
 - Language
 - Time Zone



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Wi-Fi
- Password
- 6. Update the operating system.

Step 3: Enable Camera Interface

Enable the Raspberry Pi Camera.

```
sudo raspi-config
```

Navigate to:

Interface Options



Camera



Enable

Restart Raspberry Pi.

```
sudo reboot
```

Step 4: Update the Raspberry Pi

```
sudo apt update
```

```
sudo apt upgrade -y
```

Step 5: Install Required Software Packages

Install Python and OCR libraries.

```
sudo apt install python3-pip
```

```
sudo apt install tesseract-ocr
```

```
sudo apt install libtesseract-dev
```

```
sudo apt install espeak
```

```
sudo apt install python3-opencv
```

Install Python packages.

```
pip3 install pytesseract
```

```
pip3 install Pillow
```

```
pip3 install gtts
```

```
pip3 install playsound
```

Step 6: Test the Camera

Capture a test image.

```
libcamera-still -o test.jpg
```

Verify that **test.jpg** has been saved.

Step 7: Capture Image

The camera captures an image containing printed text.

Example:

Book

Newspaper

Medicine Label

Notice Board

Document



Captured Image

Step 8: Image Pre-processing

Improve image quality before OCR.

Operations include:

- RGB to Grayscale
- Noise Removal



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Thresholding
- Image Sharpening
- Contrast Enhancement

OpenCV is commonly used for these preprocessing tasks.

Step 9: Text Extraction Using OCR

Use Tesseract OCR to recognize text.

Example Python code:

```
import pytesseract
from PIL import Image
```

```
image = Image.open("test.jpg")
```

```
text = pytesseract.image_to_string(image)
```

```
print(text)
```

Output:

WELCOME TO OCR PROJECT

Step 10: Convert Text into Speech

Convert recognized text into voice.

Example using gTTS:

```
from gtts import gTTS
```

```
import os
```

```
tts = gTTS(text)
```

```
tts.save("voice.mp3")
```

```
os.system("mpg321 voice.mp3")
```

Or using eSpeak:

```
import os
```

```
os.system('espeak "Welcome to OCR Project"')
```

Step 11: Audio Output

The generated speech is sent to:

USB Speaker

↓

User hears the recognized text.

Step 12: Display Output (Optional)

If an HDMI display is connected:

- Show captured image
- Show extracted text
- Show processing status
- Display user interface

Step 13: Repeat Process

The application continues the following cycle:



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

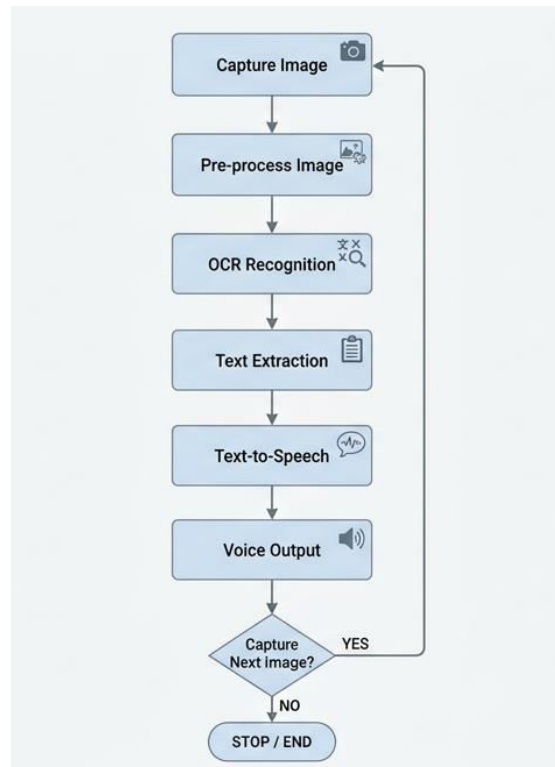


Fig No. 19: Repeat Process

Overall System Flowchart

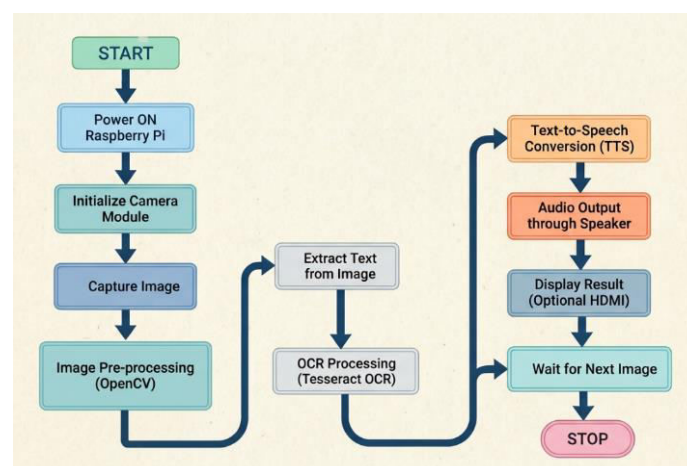


Fig No. 20: Overall System Flowchart

Working Principle

1. The Raspberry Pi powers on and initializes the camera module.
2. The camera captures an image containing printed text.
3. OpenCV preprocesses the image to improve OCR accuracy.
4. Tesseract OCR extracts text from the processed image.
5. The extracted text is passed to a Text-to-Speech (TTS) engine such as **eSpeak NG** or **gTTS**.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

6. The TTS engine synthesizes speech from the extracted text.
7. The generated speech is played through the USB speaker.
8. If an HDMI display is connected, the captured image and recognized text are displayed.
9. The system repeats the process for the next image.

The **Image Text-to-Speech Conversion Using OCR Technique on Raspberry Pi** integrates image acquisition, image preprocessing, optical character recognition, and speech synthesis into a single embedded system. By combining the Raspberry Pi Camera Module, Tesseract OCR, OpenCV, and a Text-to-Speech engine, the system can capture printed text, recognize it accurately, and convert it into audible speech. This solution is especially valuable as an assistive technology for visually impaired users, while also finding applications in education, automation, document digitization, and smart embedded systems.

Software Screenshots

Chhatrapati Shivaji Maharaj (1630–1680) was a legendary warrior-king and the founder of the Maratha Empire. Born at Shivneri Fort, he successfully carved an independent kingdom from the Bijapur Sultanate and defied the Mughal Empire. Revered as the "Father of the Indian Navy", he is celebrated for his brilliant guerrilla warfare tactics and just administration.

Chhatrapati Shivaji Maharaj is one of the most revered figures in Indian history. Born on February 19, 1630, in the rugged hills of the Sahyadri ranges, he grew up in a politically fractured region dominated by regional sultanates and the expanding Mughal Empire. Guided by his mother, Jijabai, and trained in statecraft and martial arts, Shivaji developed a deep sense of patriotism and a vision for *Hindavi Swarajya*.

Chhatrapati Shivaji Maharaj (1630–1680) was a legendary warrior-king and the founder of the Maratha Empire. Born at Shivneri Fort, he successfully carved an independent kingdom from the Bijapur Sultanate and defied the Mughal Empire. Revered as the "Father of the Indian Navy", he is celebrated for his brilliant guerrilla warfare tactics and just administration.

Chhatrapati Shivaji Maharaj is one of the most revered figures in Indian history. Born on February 19, 1630, in the rugged hills of the Sahyadri ranges, he grew up in a politically fractured region dominated by regional sultanates and the expanding Mughal Empire. Guided by his mother, Jijabai, and trained in statecraft and martial arts, Shivaji developed a deep sense of patriotism and a vision for *Hindavi Swarajya*.

Chhatrapati Shivaji Maharaj (1630–1680) was a legendary warrior-king and the founder of the Maratha Empire. Born at Shivneri Fort, he successfully carved an independent kingdom from the Bijapur Sultanate and defied the Mughal Empire. Revered as the "Father of the Indian Navy", he is celebrated for his brilliant guerrilla warfare tactics and just administration.

Chhatrapati Shivaji Maharaj is one of the most revered figures in Indian history. Born on February 19, 1630, in the rugged hills of the Sahyadri ranges, he grew up in a politically fractured region dominated by regional sultanates and the expanding Mughal Empire. Guided by his mother, Jijabai, and trained in statecraft and martial arts, Shivaji developed a deep sense of patriotism and a vision for *Hindavi Swarajya*.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. DISCUSSION

➤ Innovation

The Image Text-to-Speech Conversion system combines image processing, Optical Character Recognition (OCR), and Text-to-Speech (TTS) technologies on a compact Raspberry Pi platform. The innovative aspect of this project lies in its ability to automatically capture text from images and convert it into spoken words in real time without requiring a computer. By integrating a camera, OCR engine, and speech synthesis module into a single portable device, the system provides an efficient and affordable solution for accessing printed information.

The use of Raspberry Pi makes the system lightweight, energy-efficient, and cost-effective compared to traditional text-reading devices. The project can also be extended to support multiple languages, wireless connectivity, cloud services, and artificial intelligence-based text recognition, making it suitable for future smart assistive applications.

➤ Usefulness

1. **Assistance for Visually Impaired Individuals:** Enables users to hear printed text, increasing independence and accessibility.
2. **Reading Printed Documents:** Helps in reading books, newspapers, magazines, notices, labels, and official documents.
3. **Portable and Low-Cost Solution:** Raspberry Pi provides an economical and compact platform suitable for everyday use.
4. **Educational Support:** Assists students with visual disabilities or reading difficulties by converting study materials into speech.
5. **Real-Time Information Access:** Provides immediate audio output from captured images, reducing the need for manual reading.
6. **User-Friendly Operation:** Requires minimal technical knowledge and can be operated easily by people of different age groups.
7. **Wide Application Areas:** Useful in libraries, schools, public places, offices, healthcare facilities, and smart assistive devices.
8. **Scalability and Future Enhancement:** Can be upgraded with multilingual speech support, object detection, document scanning, and mobile connectivity.

VI. CONCLUSION AND PUBLICATIONS

Conclusion

The **Image Text-to-Speech Conversion Using Optical Character Recognition (OCR) Technique in Raspberry Pi** project successfully demonstrates an effective and economical solution for converting printed text into audible speech. The system integrates image acquisition, image preprocessing, text extraction using OCR, and speech synthesis to provide real-time audio output from textual information present in images.

The use of a Raspberry Pi platform makes the system compact, portable, energy-efficient, and cost-effective. By employing OpenCV for image enhancement and Tesseract OCR for text recognition, the system achieves accurate extraction of text from captured images. The extracted text is then converted into natural speech using a Text-to-Speech engine, enabling users to listen to the content without reading it visually.

This project is particularly beneficial for visually impaired individuals, elderly users, and people with reading difficulties, as it improves accessibility to printed documents, books, notices, labels, and other textual materials. The proposed system offers a practical assistive technology solution that can operate with minimal human intervention and low hardware requirements.

Overall, the project achieves its objective of converting image-based text into speech efficiently and accurately. Future enhancements such as multilingual OCR support, handwritten text recognition, real-time video text reading, cloud-based processing, and artificial intelligence integration can further improve the performance, usability, and applicability of the system in real-world environments.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

In conclusion, the Image Text-to-Speech Conversion Using OCR Technique in Raspberry Pi provides a reliable, affordable, and user-friendly assistive solution that enhances accessibility and promotes independent access to information for visually impaired users.

Publications

In 2024, **Mangesh Sarak, Prof. S. S. Patil, and Prof. Abhijit S. Mali** proposed "Image Text to Speech Conversion Using Optical Character Recognition Technique in Raspberry Pi". The researchers integrated image acquisition, OCR processing, and speech synthesis into a single platform, demonstrating the effectiveness of Raspberry Pi in developing assistive reading systems.

Bibliography Statement

The development of the **Image Text-to-Speech Conversion Using Optical Character Recognition (OCR) Technique in Raspberry Pi** project was supported by literature related to digital image processing, optical character recognition, computer vision, embedded systems, Raspberry Pi programming, and text-to-speech technologies. The above books, research papers, journals, and official software documentation were consulted to understand OCR algorithms, image preprocessing techniques, speech synthesis methods, and Raspberry Pi implementation procedures. These references provided the theoretical foundation and practical guidance required for the successful design and implementation of the proposed system.

REFERENCES

- [1] R. Smith, "An Overview of the Tesseract OCR Engine," in Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, 2007, pp. 629–633.
- [2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 2005, pp. 886–893.
- [3] S. Mori, H. Nishida, and H. Yamada, "Optical Character Recognition," John Wiley & Sons, New York, NY, USA, 1999.
- [4] U. Pal and B. B. Chaudhuri, "Indian Script Character Recognition: A Survey," Pattern Recognition, vol. 37, no. 9, pp. 1887–1899, Sept. 2004.
- [5] A. K. Jain, Y. Zhong, and S. Lakshmanan, "Object Matching Using Deformable Templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 3, pp. 267–278, Mar. 1996.
- [6] The different image processing techniques to text from natural images Priyanka Muchhadiya
- [7] PISEE: RASPBERRY PI-BASED IMAGE TO SPEECH SYSTEM FOR THE VISUALLY IMPAIRED WITH BLUR DETECTION Sanjay Dutta*1, Sonu Dutta*2, Om Gupta*3, Shraddha Lone*4, Prof. Suvarna Phule*5.
- [8] Raspberry pi based Intelligent Reader for Visually Impaired Persons Prof. Vaibhav V. Mainkar, Miss. Tejashree U. Bagayatkar, Mr. Siddhesh K. Shetye, Mr. Hrushikesh R. Tamhankar, Mr. Rahul G. Jadhav, Mr. Rahul S. Tendolkar
- [9] Raspberry Pi Based Reader for Blind People Prashant Chougule, Bharati Vidyapeeth Deemed University, International Research Journal of Engineering and Technology (IRJET)
- [10] Text Recognition from Images: A Study, Sahana K Adyanthaya, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181. Published by, www.ijert.org NCCDS - 2020 Conference Proceedings
- [11] E - Braille: A Study Aid For Visual Impaired, Chaitanya Jambotkar Jain College of Engineering & Research.
- [12] Raspberry Pi-A Small, Powerful, Cost Effective and Efficient Form Factor Computer: A Review, Anand Nayyar Duy Tan University.
- [13] A Review Paper on Raspberry Pi and its Applications Hirak Ghael BK Birla Institute of Engineering and Technology, Pilani.
- [14] Implementation of Image to Text Conversion using Android App Ishita Pal, Mohammadraza Rajani, Anusha Poojary, Priyanka Prasad.
- [15] Text Detection and Recognition in Imagery: A Survey Qixiang Ye, Member, IEEE, David Doermann, Fellow, IEEE.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com