Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Converting Image Text to Speech Using Raspberry Pi

Wydyanto[1], Norshita Mat Nayan[2]
[1] *Faculty Science Technology, Universitas Bina Darma, Palembang, South Sumatra 30264 Indonesia wydyanto@binadarma.ac.id*
[2] *Institut Visual Informatics, Universiti Kebangsaan Malaysia,43600 UKM Bangi, Selanggor, Malaysia, norshitaivi@ukm.edu.my*

ARTICLE INFO

Received: 18 Dec 2024
Revised: 10 Feb 2025
Accepted: 28 Feb 2025

ABSTRACT

This project aims to develop a system that can convert image text into speech using Raspberry Pi. This system will use optical character recognition (OCR) technology to extract text from images, which will then be converted into speech using text-to-speech software. (TTS). (TTS). This will provide a valuable tool for individuals with visual impairments or those who have difficulty reading text in images. This research explores a Raspberry Pi-based device that can translate English into 53 dialects, using a camera module, OCR motor, Google Speech API, and Microsoft Translator. This feature is accessible to visually impaired individuals and those who do not speak English. Raspberry Pi, with the Tesseract OCR engine, Google Voice API, Microsoft Translator, and camera board, enables real-time text translation on images, making it more accessible and inclusive for users with visual impairments or language barriers.

Keywords: Raspberry Pi; Tesseract OCR engine; Google Speech API; Microsoft translator; Raspberry Pi camera board.

INTRODUCTION

The Raspberry Pi Foundation developed the Raspberry Pi to promote basic computer science education by providing a low-cost computing platform. Python is the official programming language of the Raspberry Pi and is popular for various applications (Pajankar, 2022). By utilizing the Raspberry Pi platform and a combination of OCR technology, voice recognition, and translation services, a user-friendly and accessible text-to-speech scanner for visually impaired individuals can be created. The Raspberry Pi 3, Pi camera, and earphone can be used to capture text images and convert them to speech through OCR and TTS technologies. Our device will leverage the processing power and connectivity options of the Raspberry Pi to quickly and accurately interpret text from images or documents"The device described in the source uses OCR technology to convert scanned images of printed text into text that can be understood or edited using a computer program. It then converts the text into speech using a Text-to-speech library, allowing blind users to listen to the extracted text. The device can be used independently and does not require an internet connection (Danial *et al.*, 2011). The extracted text will then be processed through a voice recognition system that can identify languages and dialects, allowing the device to accurately translate and pronounce the text in the user's chosen language. "The use of technology, such as computer-supported environments and multimedia interactive activities, has shown potential in enhancing accessibility and communication for individuals struggling with reading or understanding text in English. However, Web-based environments may introduce cognitive barriers and distractions that can overwhelm readers, especially those with reading disabilities. New literacies are required to navigate interactive features and multimedia representations on the Internet effectively, which can be challenging for some readers. Teachers need to be aware of these cognitive challenges posed by Internet environments to support struggling readers effectively (Coiro, 2003).

The Raspberry Pi is a Raspberry Pi-based single-board PC that uses Raspbian for its operating system. It uses a camera module to capture images and convert them into machine-encoded content. The OCR motor, Tesseract, and content to speech motors are used for converting the images into speech. eSpeak is a speech synthesizer for English speech, while Google content to speech and Microsoft interpreter are used for interpreting the content into different languages. The model is run on a Secure Socket Shell (SSH) using Putty, a free and open-source terminal emulator. concept of Optical Character Recognition (OCR) and Text to Speech Synthesiser (TTS) in Raspberry pi. This kind of system helps visually impaired people to interact with computers effectively through vocal interface. Text Extraction

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

from colored images is a challenging task in computer vision. Text-to-Speech is a device that scans and reads letters and numbers in English found in images using OCR techniques and converts them into speech. This paper explains the design, implementation, and experimental results of the device. This device consists of two modules, an image processing module and a sound processing module. This device is developed based on the Raspberry Pi v2 with a processor speed of 900 MHz. (Reshmi, Salagar and Veni, 2019) Most visually impaired individuals use Braille to read documents and books that are difficult to produce and less available. This creates a need for the development of devices that can alleviate the difficult and torturous tasks that visually impaired individuals have to endure. Because of the digitization of books, there have been many extraordinary efforts in building strong document analysis systems in industry, academia, and research laboratories, but this is only for those who are able to see. This project aims to study image recognition technology with voice synthesis and develop costs.

II. RELATED WORK

The use of Raspberry Pi for home automation has been explored in previous research, where it was connected to various IoT devices such as temperature sensors, lighting systems, and music players via a secure VPN [1]. Additionally, a smart home automation system based on IoT and Edge-Computing paradigm was proposed, using Raspberry Pi as a central controlling unit for interconnecting devices and sensors in a home . (Venkatraman, Overmars and Thong, 2021) & (Venkatraman, Overmars and Thong, 2021) . "The study on real-time object detection on Raspberry Pi 4 involved deploying a quantized pretrained SSD model to evaluate throughput for object recognition. The inference time obtained was 185 ms with an input size of 300x300, showing improvement over previous models. Transfer learning and fine-tuning with Tensor Flow were used to train a custom object detection model on images scraped from the web, particularly focusing on people in winter landscapes. The custom model performed better at detecting people in snow, indicating the effectiveness of web scraping for model refinement (Panduman *et al.*, 2024). (Simon, Williem and Park, 2015). In addition, there have been projects that utilize Raspberry Pi to create interactive art installations and data visualization tools. These studies demonstrate the flexibility and potential of Raspberry Pi as a platform for innovation and experimentation. (Devi and Baboo, 2014) Text to translate: Context: Text to translate: In this research, the focus will be on using Raspberry Pi to build an intelligent conversion system from text to speech that integrates various sensors and devices to automate and monitor tasks for visually impaired individuals. By leveraging the capabilities of the Raspberry Pi, cost-effective and customizable solutions can be created to enhance comfort and efficiency in daily life (Danial *et al.*, 2011). Through research, it will explore the possibility of integrating voice recognition, motion detection, and remote access features to create a seamless and smart home environment. This project will showcase the potential of Raspberry Pi as a versatile tool for creating innovative solutions in the field of home automation. By using this technology, it is hoped that it can provide opportunities for blind people to be more independent and feel safe in carrying out daily activities. In addition, the use of Raspberry Pi can also help reduce the cost of building home automation systems, which tend to be expensive. Thus, this project is expected to provide significant benefits to the communities in need, and serve as a real example of how technology can be used to improve the overall quality of life.

proposing an optical character recognition technique using Intro Sort. The main feature of this proposed technique is to perform image segmentation using intro sort. This reduces the comparison time to match pixels from an image. This reflects a reduction in OCR time. The intro sort algorithm starts with quick sort, and when the recursion depth exceeds a certain level, the algorithm switches to heap sort, based on the number of pixels being sorted. This approach also has the advantage of recognizing license plates and text.

document in a very short time. This approach is capable of recognizing objects with high accuracy and efficiency. extract characters with different font sizes. This technique also works well on noisy images. In addition, it is resistant to various types of image distortion. (Zhang, Sugano and Bulling, 2018), Text to translate: Context: (Seema Barate1, Chaitrali Kamthe2, Shweta Phadtare3, 2016)Text to translate: According to the World Health Organization (WHO), it is estimated that 285 million people worldwide have visual impairments, with 90 percent living in developing countries and 45 million people blind globally. Although there are many existing solutions for helping individuals with dyslexia, none of them provide a reading experience equivalent to that of the sighted population. Specifically, there is a need for affordable and easily accessible portable text readers for the blind community.

The most common solution currently is to use screen reader software or specialized reading aids. However, many of these solutions are still expensive and difficult to access for those who need them. Therefore, the development of texts for translation has become an urgent need for the blind community worldwide. With the advancement of increasingly sophisticated technology, it is hoped that more affordable and accessible solutions will emerge for those who need them.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The addition of those specifically empowered in the IT revolution is both a social responsibility and a computational test in today's rapidly evolving digital world. This work suggests a smart reader for the blind using a Raspberry Pi. This paper provides a lecture on the application of a complete Text-to-Speech system designed for the visually impaired. This system consists of a webcam connected to a Raspberry Pi that receives printed text pages. OCR package Towards a System for People with Disabilities The vision in this research will propose a system that reads text found in natural scenes with the aim of providing assistance to people with visual impairments. This paper explains the system design and evaluates several character extraction methods (Thakkar AShah P, 2017). Automatic text recognition from natural images is gaining more attention due to its potential applications in imaging, robotics, and intelligent transportation systems. Camera-based document analysis has become a real possibility with the increasing resolution and availability of digital cameras. In this paper, an innovative, efficient, and cost-effective real-time technique is introduced that allows users to hear the content of text images instead of reading them "The paper discusses the development of a system that enables editors to correct errors in captions created by Automatic Speech Recognition, emphasizing the importance of real-time stenography transcription in higher education (Jundale, 2016) It also mentions the transLectures project, which aimed to implement automatic transcription and translation systems for video lectures to provide subtitles for non-native speakers and the deaf and hard-of-hearing (Juang, Tsai and Fan, 2015) This combines.

III. PROPOSED SYSTEM

In the research we conducted, we will focus on how to separate highlights from images containing content and convert them into sound. The filtered content page will be used to prepare a temporary classifier, while the images taken by the camera will convert the content of those images into sound. The proposed method will involve the use of image processing techniques and pattern recognition to identify and separate highlights from the images. In addition, we will also develop an algorithm that can convert image content into text, which will then be transformed into speech using text-to-speech technology. Thus, users will be able to listen to the information contained in the image without having to see it visually.

This research will also plan to conduct more writing research on the restoration of archival images, combining and enhancing them while preserving memory that requires little time. We might find the ideal compromise between training speed and coordination accuracy, and to demonstrate this idea on a Raspberry Pi. The camera will take photos of the content and naturally replay the corresponding timestamps from the recording.

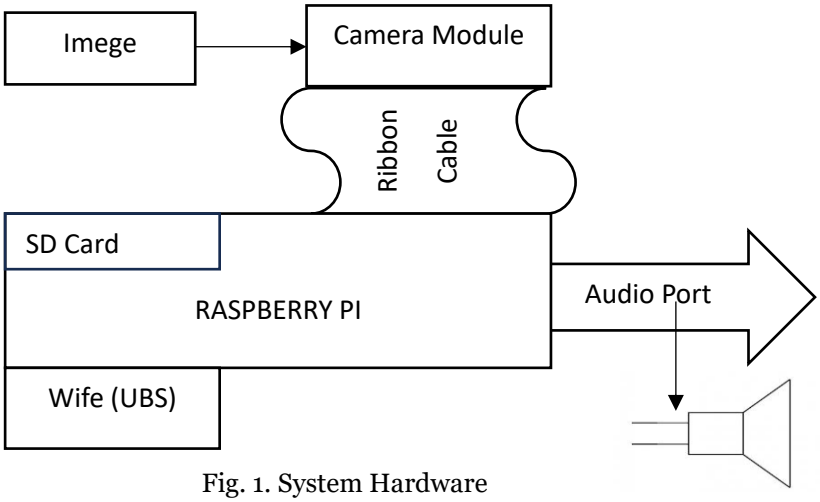

Fig. 1. System Hardware

V. PROPOSED SYSTEM

We will centre on separating highlights from pictures of a content and change over them in to sound. Filtered pages of content will prepare the classifier while pictures taken by camera will change over that picture content into sound. We plan to accomplish more writing research on archive picture recovery, join and enhance them by keeping memory require little in the meantime. We will probably find the ideal exchange off between preparing speed and coordinating exactness, and to make a showing of this idea on raspberry pi. The camera would take a photo of content and naturally play back the applicable time stamp of its relating book recording.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fig. 2. Ilustration Result Hardware

Tool Specifications:

1. Contexto: Laptop: Texto a traducir: Laptop:

a) Function: The laptop is used as the main device for programming and configuring the Raspberry Pi. The laptop runs software that interacts with the Raspberry Pi to perform OCR (Optical Character Recognition) and Text-to-Speech processes.

b) Operating System: The operating system displayed on the laptop screen is Linux-based, showing the use of the terminal to run code.

2. Raspberry Pi:

a) Function: The Raspberry Pi serves as the main computing device that handles the process of text recognition from images and converts it into speech. Raspberry Pi is connected to other input/output devices such as speakers.

b) Model: Most likely Raspberry Pi 3 or Raspberry Pi 4, which support network connections and USB devices.

3. Contexto: \nTexto a traducir: Orador:

Function: The speaker functions to produce sound output after the text-to-speech conversion process is complete. The text extracted from the image through OCR is processed by the Raspberry Pi, and then the result is converted to audio through this speaker.

4. Router (behind the scenes):

Function: A router may be used to connect a Raspberry Pi to the internet or LAN, allowing for remote access or additional online resources, such as APIs for translation or TTS processing.

How the Tool Works:

1. Image Input: This system accepts images containing text. The image can be obtained through a camera or taken from another source. (misalnya, file gambar yang sudah tersimpan). (misalnya, file gambar yang sudah tersimpan).

2. OCR (Optical Character Recognition) process: The Raspberry Pi, connected to the laptop, runs OCR software to extract text from the given image. OCR software like Tesseract can be used to detect and recognize text in images.

3. Translation Process (If Any): If needed, the OCR output text can be translated into another language. Raspberry Pi can connect to online translation APIs, such as the Google Translate API, to translate the extracted text.

4. Text-to-Speech (TTS) Conversion: After the text is recognized or translated, the system uses TTS (Text-to-Speech) software to convert the text into speech. Raspberry Pi processes this text and generates audio output.

5. Output Audio: The audio resulting from the TTS conversion is sent to the speaker connected to the Raspberry Pi. The speaker then produces sound that can be heard by the user.

This system is very suitable for applications such as reading aids for the visually impaired, text-to-speech conversion in various languages, or direct translation of text into speech.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

System Software Design

In this research, the system software design is the core of the solution development that combines OCR (Optical Character Recognition) technology, TTS (Text-to-Speech), and translation capabilities using Raspberry Pi as the main computing platform. This system is designed to enable text extraction from images, translate the text if necessary, and convert it into speech. There are several key software components integrated into this research, all of which run on Raspberry Pi.

1. Arsitektur Sistem

The system software architecture is designed in several modules that function sequentially. The modules are:

a) Image Input Module: Capturing the image to be processed.

b) OCR Module: Extracting text from images.

c) Translation Module (optional): Translate the text into the desired language.

d) Text-to-Speech Module: Converts text (in the original language or translation) into audio.

e) Output Module: Output sound results through the speaker.

This architecture follows pipeline processing where each module functions sequentially and processes input from the previous module. Raspberry Pi acts as the main processor that manages the entire process.

2. Main Components of Software

a. Optical Character Recognition Module (OCR) (OCR)

1. Description: The OCR module is the main component used to extract text from images. The technology commonly used in research like this is Tesseract OCR, which is an open-source library and supports various languages.

2. How It Works:

a. The image is taken as input and forwarded to the OCR module.

b. OCR then scans the image to search for text and converts the visual patterns of characters into a digital representation. (teks). (teks).

c. The result of the OCR is text ready to be further processed by the system.

3. Challenge: The main challenge in the OCR module is the accuracy in recognizing text, especially when the image has low quality or there is visual noise.

b. Translation Module

1. Description: The optional translation module is used if the system is set to convert text from one language to another before the TTS process.

2. How it Works:

a) After the text is extracted by OCR, it is sent to the translation module.

b) The Raspberry Pi can use online translation APIs, such as the Google Translate API, to automatically translate the text.

c) If the system is offline, the module can use local translation libraries, although support is usually limited compared to online APIs.

d) Challenges: Translation speed and accuracy can be affected by network connectivity, API latency, and the accuracy of the translation in the context of the sentence.

c. Text-to-Speech (TTS) Module

1. Description: This module converts text to speech. The Raspberry Pi uses TTS software, such as espeak or Google Text-to-Speech, to read the processed text.

2. How it Works:

a. The text generated by the OCR module (or translated text from the translation module) is converted to audio by the TTS module.

b. The system generates a digital sound file that is played through a connected speaker.

b) Challenge: The challenge here lies in the naturalness of the voice produced by the TTS software. Most default TTS modules produce a somewhat robotic voice, so choosing the right TTS engine is crucial.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

d. Integration with Raspberry Pi

1. Description: Raspberry Pi acts as the brain of the system. Raspberry Pi has several advantages, such as its small size, energy efficiency, and support for various relevant libraries such as Tesseract and espeak.

2. How it Works:

a) The developed software is run on the Raspberry Pi.

b) Raspberry Pi interacts with various components, such as the camera (to take pictures) and speakers (to output sound).

c) For the translation process, Raspberry Pi can be connected to the internet via Wi-Fi or Ethernet to access the translation API if needed.

3. System Workflow

Here is the software workflow from start to finish:

1. Image Input:

An image containing text is taken through a camera connected to the Raspberry Pi or uploaded from local storage.

2. OCR Process:

The image is sent to the OCR module for scanning and text recognition. OCR processes the image and extracts the text contained in it.

3. Translation (If Required):

Once the text is recognized, the system can translate the text using the translation module if the user chooses to translate.

4. Text-to-Speech Conversion:

Once the text is recognized or translated, the system sends the text to the TTS module to be converted into speech.

5. Speech Output:

The audio generated by the TTS module is sent to the speaker, and the text is read aloud.

4. Design Considerations

In designing this software, several important aspects were considered:

a) Memory and Computational Efficiency: Since Raspberry Pi has limited resources compared to desktop computers, efficiency is very important. The use of lightweight libraries such as Tesseract and espeak is prioritized.

b) Internet Connectivity: For the translation feature, a stable internet connection is required. However, the system is also designed to still function for text recognition and TTS even in offline mode.

c) Simple User Interface: Since this is an IoT device-based system, the user interface design on the laptop or Raspberry Pi is kept to a minimum to make it easier to operate.

5. System Benefits

a) Portability: The small and lightweight Raspberry Pi makes the system easy to carry and install in various locations.

b) Cost Effective: The system uses affordable hardware, making it ideal for large-scale deployment at low cost.

c) Language Flexibility: With translation and TTS modules supporting multiple languages, the system can be used in a variety of multilingual situations.

With this design, your research shows great potential in the development of technology that can integrate text recognition, translation, and text-to-speech conversion. The system can also be further enhanced and customized to suit the specific needs of real-world applications, such as reading aids or trasation.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

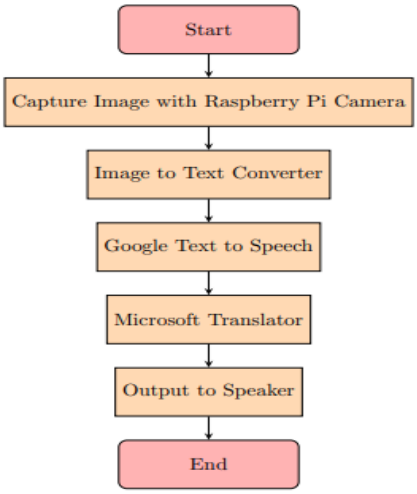

Fig. 3. Flowchart of Image to Speech Conversion Process

V. CONCLUSION

This research successfully implemented a system for converting text in images to speech through a translation process using Raspberry Pi devices. The designed system utilizes OCR (Optical Character Recognition) techniques to extract text from images, and then uses Text-to-Speech (TTS) to convert the text into speech. The research results show that the Raspberry Pi, with its simple yet efficient configuration, is capable of running this process with good accuracy and quick response time, despite the device's limited computational resources. The use of Raspberry Pi allows this system to be portable, energy-efficient, and affordable, making it ideal for applications in various contexts such as reading aids for the visually impaired or real-time translation on signboards. In addition, this system can also be integrated with language translation modules, making it a versatile solution for processing text and voice across languages. Nevertheless, this research opens up opportunities for further development in terms of improving OCR accuracy on low-quality image texts and integrating more natural TTS algorithms. Thus, the results of this research can make a significant contribution to the development of user-friendly and widely accessible IoT-based assistive technology.

REFERENCES

[1] Coiro, J. (2003) 'Reading comprehension on the Internet: Expanding our understanding of reading comprehension to encompass new literacies', *Reading Online*, 2003(February 2003), pp. 1–16.

[2] Danial, M. N. *et al.* (2011) 'Image Segmentation and Text Extraction: Application To The Extraction Of Textual Information In Science Images', *International Seminar on Application of Science Mathematics 2011*, pp. 1–8.

[3] Devi, V. A. and Baboo, S. S. (2014) 'Embedded Optical Character Recognition On Tamil Text Image Using Raspberry Pi', 2(4), pp. 127–131.

[4] Juang, J., Tsai, Y. and Fan, Y. (2015) 'Visual Recognition and Its Application to Robot Arm Control', pp. 851–880. doi: 10.3390/app5040851.

[5] Jundale, T. (2016) 'Research Survey on Skew Detection of Devanagari Script Research Survey on Skew Detection of Devanagari Script', (April), pp. 41–44.

[6] Pajankar, A. (2022) 'Raspberry Pi Image Processing Programming', *Raspberry Pi Image Processing Programming*. doi: 10.1007/978-1-4842-8270-0.

[7] Panduman, Y. Y. F. *et al.* (2024) 'A Survey of AI Techniques in IoT Applications with Use Case Investigations in the Smart Environmental Monitoring and Analytics in Real-Time IoT Platform', *Information (Switzerland)*, 15(3). doi: 10.3390/info15030153.

[8] Reshmi, S., Salagar, R. D. and Veni, S. S. (2019) 'Text Detection In Image Based On The Morphology Method', 7(3), pp. 468–471.

Research Article

[9] Seema Barate1, Chaitrali Kamthe2, Shweta Phadtare3, R. J. (2016) 'Implementasi Ekstraksi Karakter Teks dari Gambar Tulisan Tangan yang Dipotret ke Teknik Pencocokan Template Konversi Teks', *MATEC Web of Conferences*.

[10] Simon, C., Williem and Park, I. K. (2015) 'Correcting geometric and photometric distortion of document images on a smartphone', *Journal of Electronic Imaging*, 24(1), p. 013038. doi: 10.1117/1.jei.24.1.013038.

[11] Thakkar AShah P (2017) 'Review on Tesseract OCR Engine and PerformanceKateryna Zinchenko', *International Journal of Innovative and Emerging Research in Engineering*, 4(12), pp. 4–6.

[12] Venkatraman, S., Overmars, A. and Thong, M. (2021) 'Smart home automation—use cases of a secure and integrated voice-control system', *Systems*, 9(4). doi: 10.3390/systems9040077.

[13] Zhang, X., Sugano, Y. and Bulling, A. (2018) 'Revisiting data normalization for appearance-based gaze estimation', *Eye Tracking Research and Applications Symposium (ETRA)*. doi: 10.1145/3204493.3204548.