Image Text To Speech Conversion with Raspberry-Pi Using OCR

Shivkanya V Dahiphale[1], Prof. S. J. Nandedkar[2]

[1] PG Student, Electronics & Telecommunication Engineering, Aurangabad, India
[2]Professor, Electronics & Telecommunication Engineering, Aurangabad, India

---

Abstract – In this paper, we have proposed an image text to speech converter using Raspberry Pi. It is very difficult to read text from text images or text-boards. Visual impairment is one of the greatest handicaps of humanity, especially in today's world where information is communicated through text messages rather than voice. We have tried to extract and convert text from an image, i.e., capture an image that only contains text and convert it into speech. This is done using Raspberry Pi and Optical Character Recognition (OCR). The captured image undergoes several image processing steps to find only the part of the image which contains the text. Various tools are used to convert a new image (which only contains text) into speech. These tools include OCR software, TTS (Text to Speech), and the audio output can be heard through speakers or earphones.

***Key Words*: OCR, Text Translator, TTS, Raspberry Pi, Visually impaired person, Text Extraction**

1. INTRODUCTION

Every year, the number of visually challenged persons are increasing due to eye diseases, age related causes, traffic accidents and other causes. As reading is one of the most important tasks in the daily routine (text is present everywhere) of humankind, visually impaired people face many difficulties. Speech gives support to the visually challenged persons for reading out the text. The focus of this research is that the visually challenged person can get information about text into audio format. This paper have presented design for a camera based reading system that extract text from image and identify the text characters and strings from the captured image and finally text will be converted into audio. The captured image goes through a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. The OCR and TTS process the image. Text Recognition (OCR) has become one of the most popular uses of technology in text recognition and AI. Optical Character recognition (OCR), is the process of converting scanned images of machine printed or handwritten text (numerals, letters, and symbols), into a computer format text [1].

1.1 Summary of Literature Review

In this section, we present some previous research done to assist visually challenged people with text-to-speech technology. This literature review is used to study different image-text-to-speech conversion techniques. By using these research papers, the survey is done for the invention of new techniques. This paper is going to be a solution for the conversion of images into sound. The system includes the head-mounted video camera. Its design is portable and low-power. In this work, images are converted into text, and then that text is converted using MATLAB coding, which is commonly used for image processing techniques. So, there is scope to increase the database of the proposed system [2]. The device consists of a portable device for assisting the visually impaired. The setup is foldable, and hence its portability is enhanced. It can be broken down into two parts, and it barely takes 5 seconds to be set up again. The two parts of the device are 1) the stand, onto which the Raspberry Pi board is mounted, and 2) a slot in the wooden board for the camera. 2) A plain slate that has slots for inserting the paper [3]. This paper presents an approach for text extraction and the conversion of it to speech. The OCR (optical character recognition) converts the text images into machine-encoded text and saves it in a text file. Tesseract is the OCR engine that is used for extracting the English text from the image and storing it in a text file. The text-to-speech engine converts text-to-speech output [4]. Rama Mohan Babu, p. Srimaiyee, a. Srikrishna used characters in a text of different shapes and structures. Text extraction may employ binarization or directly process the original image; it consists of a survey of existing techniques for page layout and analysis. Mathematical morphology is a geometrical-based approach to analyzing images. For the extraction of geometrical structures and representing shapes in many applications, it provides powerful tools. Morphological Feature Extraction (MPE) has been proven to be a powerful tool for character detection and document analysis, particularly when using dedicated hardware. They proposed an algorithm for text extraction based on morphological operations [5]. OCR technology allows a machine to automatically recognize a character through an optical mechanism. OCR is the method of translating images that contain text into machine-editable form. If we read a page in a language other than our own, we may recognize the various characters, but we may be unable to recognize words. However, on the same page, we are usually able to interpret numerical statements-the symbols for numbers are universally used [6]. Bhushan Sonawane, Kiran Patil, Nikhil Pathak, and Ram Gamane used the Microsoft Office Document Imaging OCR technology to extract text tokens, prototypes, and templates. Then they performed the following processes:

A. Extracted Text from Image: Webcam-captured image will be processed by Microsoft Office Document Imaging used in "Third Eye: An Image Explorer." Text will be extracted from the image & kept in a separate text file with the same name as the image file.

B. Text Analysis & Text Detection: Text analysis is mainly concerned with the analysis of extracted text from an image that is in a text file. Organize and maintain them into a list of words. This list contains abbreviations, numbers, and acronyms & converts them into a full line when needed. Text detection is the process of identifying precisely where it is located in that page image.

C. Text Transformation: It is a normalization of text into pronounceable form. It pronounces line-by-line words and takes a pause when space is detected between words. It reads the text according to the punctuation rules, accent marks, & stop words, much like human beings [7]. In this work, the image is converted into text, and then that text is converted into speech by MATLAB. E-text is converted into speech. With this approach, text from a Word document, web page, or e-book can be read and can generate synthesized speech through a computer's speakers. For image-to-text conversion, the first image is converted into a grey image. A gray image is converted into a binary image by thresholding, and then it is converted into text by using MATLAB. The Microsoft Win 32 SAPI library has been used to build speech-enabled applications that retrieve the voice and audio output information available for computers. In this work, one character can be converted into text at once [8]. First, they create a database of images. Images from the database are given to the system as input. When these images are taken as input from the system, the system checks for similar kinds of images in the database to identify the objects in the image. Once objects are detected, it identifies the image, and the system gives text output. This generated text then undergoes text-to-speech synthesis, and they get the speech output [9].

V. Phutak, R. Kamble, S. Gore, M. Alave, and R. R. Kulkarni (2019). "Text-to-Speech Conversion using Raspberry Pi". The authors used the Google text-to-speech API for the conversion of text to speech, and they implemented the system on a Raspberry Pi [10]. The system used image processing techniques to detect and recognize characters and then convert them into speech using a text-to-speech engine. The authors carried out testing on some visually impaired people for the demonstration of the system's effectiveness [19].

The system used optical character recognition (OCR) to extract and recognize the text from the image and then convert it into speech using a text-to-speech engine. Language translation functionality has also been added by the authors to the system [11].

The system uses OCR for text recognition from the image and then converts that text into speech using a TTS engine. The voice command functionality has also been added to the system by the authors [12].

This paper presents a blur detection algorithm using the Laplacian operator and OpenCV. The authors compared their algorithm with other blur detection techniques and claimed that their algorithm provides better results than the existing ones [17].

The authors have done a comparison of various blur detection techniques and proved their system's effectiveness in the detection of images in different kinds of blur [18].

Darshan Badhe, P. M. Ghate proposed a Marathi text-to-speech system using MATLAB. They created a database for text and sound for this proposed system. Marathi text is broken into words and then mapped to English transliteration. After mapping, the audio file from the database is matched and then played. They implemented prosody generation, which is to add emotions to sound signals depending on various factors such as age, gender, etc. [13].

2. HARDWARE DESCRIPTION

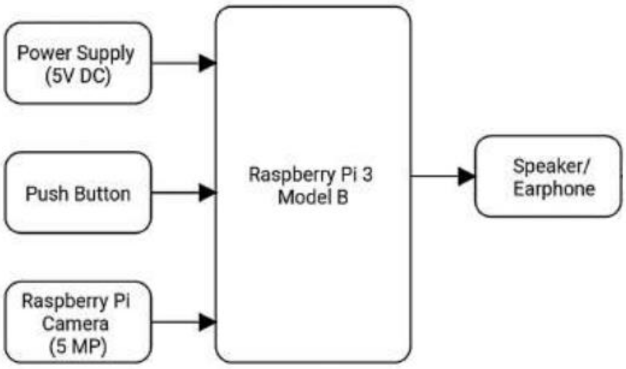

Figure 1: System Hardware Design

The system uses the Raspberry pi's input/output pins to interface with the camera and speaker. The 5 MP camera module is connected to the Raspberry Pi's Camera Serial Interface (CSI) port, and the speaker/headphone is connected to the audio output port. The 5V power supply is given to the Raspberry Pi's micro-USB port [10] [11] [12]. The hardware components used are relatively cost-effective and easily available, making the system easy to access and affordable for visually impaired persons. The hardware for the image to speech conversion system includes the following components:

Raspberry Pi 3 B: The Raspberry Pi 3 B serves as the primary processing unit for the system, functioning as a compact computer in a small form factor. It is a single- board computer developed by the Raspberry Pi Foundation. It has a quad-core 1.4 GHz processor, 1GB RAM and includes onboard Wi-Fi and Bluetooth connectivity. The Raspberry Pi 3 B is the third generation Raspberry Pi and is widely used in various projects due to its affordable cost and high processing power [11] [12].

Raspberry Pi Camera 5 MP: With a 5-megapixel sensor, the Raspberry Pi Camera 5 MP is a versatile camera module ideal for capturing images and converting them as needed. It can record 1080p video at 30 frames per second, providing high-quality footage. The camera easily connects to the Raspberry Pi board via a ribbon cable, making it simple to incorporate into various projects for different applications.[11] [12].

3. IMAGE-TEXT TO SPEECH CONVERSION PROCESS

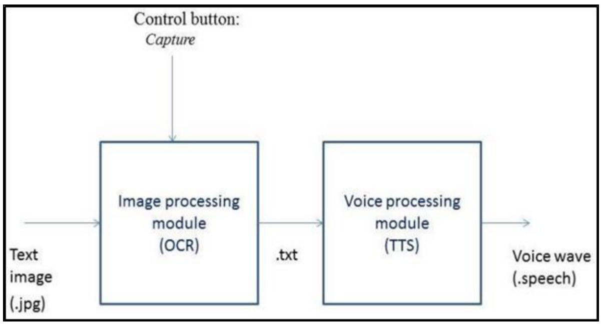

Figure 2: Text to Speech Conversion Process

Text-to-speech device consists of two main modules, the image processing module and voice processing modules (Fig.2). The image processing module utilizes a camera to capture images and then converts them into the text. The voice processing module subsequently transforms the text into sound, adjusting it with particular physical attributes to ensure the sound is comprehensible and can be listened using earphones/speakers.

4. SOFTWARE DESCRIPTION

We use Python programming language and Raspberry Pi OS as the operating system for this project [12] [14]. The system seizes an image and subsequently employs a variety of image processing techniques. Then the system uses an OCR engine to extract the text from the captured images and a TTS engine to synthesize the speech output [10]. For this project, the following libraries have been employed:

- **cv2:** OpenCV is a computer vision library that provides a range of algorithms for image processing, object detection and more. The capturing and processing of images from the Raspberry Pi camera is facilitated by this library. [14].

- **pytesseract:** Pytesseract is a Python wrapper for Tesseract-OCR, an optical character recognition (OCR) engine that can recognize text from images. The library is utilized for extracting text from images that have been captured. [14].

- **numpy:** The Python library NumPy is specifically designed for numerical computing tasks that involve multi-dimensional arrays and matrices. It is commonly used for mathematical operations like image normalization and scaling.

- **subprocess**: Subprocess is a Python module that is utilized for creating new processes and running external commands. It is used in this software to run the Tesseract-OCR command-line tool.

- **eSpeak:** Espeak serves as a concise open-source text-to-speech synthesizer that is compatible with multiple platforms. It is used in this software to convert the extracted text into speech [10].

Image Processing

The initial step involves capturing an image by utilizing the camera of the Raspberry Pi [15]. We then apply a Gaussian blur filter using OpenCV to remove any noise from the image [19] [16]. We use the Laplacian method is employed to identify the edges within the image and assess the variance of the Laplacian in order to ascertain the presence of blurriness in the image [17] [18]. If the variance is below a certain threshold, we consider the image to be blurry and discard it.

Optical Character Recognition

Once we have verified that the image is not blurry, we proceed to utilize pytesseract for extracting the text from the image. We pass the image to pytesseract, which uses OCR to recognize the characters in the image and convert them into a machine- readable format [14] [16].

Text-to-Speech

we make use of a TTS engine to convert the extracted text into spoken language. We use the sub process library to execute the espeak command to generate the speech output [10]. We play the audio file using the speaker and amplifier connected to the Raspberry Pi [11] [20].

5. IMPLIMENTATION

The implementation of the system is as follows:

1) Attach the Raspberry Pi 5 MP Camera Board Module to the Raspberry Pi 3 B, ensuring that it is connected to a power supply and network. Then install the necessary software including the Raspberry Pi [14] [17].

2) Initiate the task by creating a brand new Python script file on the Raspberry Pi board and proceed to duplicate the code within it. The program will use OpenCV [16] [17] to capture an image from the camera.

3) The program then uses OpenCV to process the image by converting it to grayscale.

4) After capturing the image, the program will use Tesseract OCR to extract the text from the image [15]. If text is found, the program will convert the text to speech and play the speech through the speaker or earphones [14].

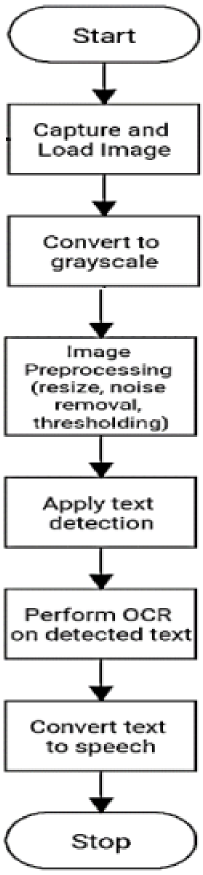

Figure 3: Flowchart

6. CONCLUSION

Thus, this paper is presented as a solution for the conversion of images into sound. To propose the methodology, number of parameters are considered by the researchers like simple and robust system, recognition time & accuracy, automatic detection and minimum time.

REFERENCES

[1] https://www.scribd.com/document/325331905/kh

[2] Raja Venkatesan.T, M.Karthigaa, P.Ranjith, C.Arunkumar, M.Gowtham, "Intelligent Transalate System for Visually Challenged People" International Journal for Scientific Research & Development (IJSRD), ISSN (online): 2321-0613, Vol. 3, Issue 12, 2016.

[3] Anusha Bhargava, Karthik V. Nath, Pritish Sachdeva, Monil Samel (April 2015), "Reading Assistant for the Visually Impaired" International Journal of Current Engineering and Technology (IJCET), E-ISSN 2277 – 4106, P-ISSN 2347 – 5161, Vol.5, No.2.

[4] K Nirmala Kumari, Meghana Reddy J (May 2016), "Image Text to Speech Conversion Using OCR Technique in Raspberry Pi" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering(IJAREEIE), ISSN (Print): 2320 – 3765, ISSN (Online): 2278 – 8875, Vol. 5, Issue 5

[5] G. RAM A MOHAN BABU, P. SRIMAIYEE, A. SRIKRISHNA, "TEXT EXTRACTION FROM ETROGENOUS IMAGESUSING MATHEMATICAL MORPHOLOGY", Journal of Theoretical and Applied Information Technology 2005-2010.

[6] M. Nagamani, S.Manoj Kumar, S.Uday Bhaskar, " Image to Speech Conversion System for Telugu Language", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 6, November 2013.

[7] Bhushan Sonawane,Kiran Patil, Nikhil Pathak, Ram Gamane, "Third Eye : An Image Explorer", International Journal of Emerging Technology [5]and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 4, April 2013)

[8] Chaw Su Thu Thu, Theingi Zin, "Implementation of Text to Speech Conversion", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 3 Issue 3, March – 2014.

[9] Mrunmayee Patil, Ramesh Kagalkar, "An Automatic Approach for Translating Simple Images into Text Descriptions and Speech for Visually Impaired People", International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 3, May 2015

[10] V.Phutak, R.Kamble, S.Gore, M.Alave, & R.R.Kulkarni (2019). Text-to-Speech Conversion using Raspberry Pi. International Journal of Innovative Science and Research Technology, 4(2), 291-295. ISSN No: 2456-2165. Retrieved from https://ijisrt.com/assets/upload/files/IJISRT19FB152.pdf

[11] H.Rithika and B.Nithya Santhoshi, "Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi," Department of Information Technology, MNM Jain Engineering College, Chennai, India,2017. DOI: 10.1109/ICCIC.2016.7919526

[12] S.Sarkar, G.Pansare, B.Patel, A.Gupta, A.Chauhan, R.Yadav and N.Battula, "Smart Reader for Visually Impaired Using Raspberry Pi," in International Conference on Innovations in Mechanical Sciences (ICIMS'21), IOP Conf. Series: Materials Science and Engineering 1132 (2021) 012032, IOP Publishing, 2021.Retrieved from https://iopscience.iop.org/article/10.1088/1757-899X/1132/1/012032/pdf

[13] Darshan Badhe, P. M. Ghate. “Marathi Text to Speech Synthesis Using Matlab”, IJCSN International Journal of Computer Science and Network, Volume 4, Issue 4, August 2015 ISSN (Online) : 2277-5420.

[14] A.G.Hagargund, S.V.Thota, M.Bera and E.F.Shaik (2017), Image to Speech Conversion for Visually Impaired. International Journal of Latest Research in Engineering and Technology (IJLRET), 03(06), 09- 15. ISSN: 2454-5031. Retrieved from http://www.ijlret.com/Papers/Vol-3-issue-6/2-B2017160.pdf

[15] A.Siby, A.P.Emmanuel, C.Lawrance, J.M.Jayan, & K.Sebastian (2020). Text to Speech Conversion for Visually Impaired People. International Journal of Innovative Science and Research Technology, ISSN No: 2456-2165, IJISRT20APR1045, 1253-1260. https://ijisrt.com/text-to-speech-conversion-for visually-impaired-people

[16] S.C.Madre and S.B.Gundre, "OCR Based Image Text To Speech Conversion Using MATLAB," in Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018). DOI: 10.1109/ICCONS.2018.8663023

[17] R.Bansal, G.Raj and T.Choudhury, "Blur Image Detection using Laplacian Operator and Open-CV," in Proceedings of the 5th International Conference on System Modeling & Advancement in Research Trends (SMART-2016), IEEE Conference ID: 39669, Nov. 25-27, 2016, College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, India. DOI: 10.1109/SYSMART.2016.7894491

[18] R.A.Pagaduan, M.C.R.Aragon and R.P.Medina. iBlurDetect: Image Blur Detection Techniques Assessment and Evaluation Study. Technological Institute of the Philippines, Information Technology Department Quezon City, Philippines. https://www.scitepress.org/Papers/2020/103077/103077.pdf

[19] A.A.Panchal, S.Varde, & M.S.Panse, (2016, May 20-21). Character Detection and Recognition System for Visually Impaired People. Paper presented at the IEEE International Conference On Recent Trends In Electronics Information Communication Technology, India. DOI:10.1109/RTEICT.2016.7808080

[20] S.M.Qaisar, R.Khan and N.Hammad, "Scene to Text Conversion and Pronunciation for Visually Impaired People," 2019 Advances in Science and Engineering Technology International Conferences (ASET), Electrical and Computer Engineering Department, Effat University, Jeddah, KSA. DOI: 10.1109/ICASET.2019.8714269