Artificial Vision-Raspberry Pi Based Reader for Visually Impaired

Srushti Nemade[1], Richa Patil[2], Ishika Bijwe[3], Kishor Bhangale[4], Rahul Mapari[5]
[1,2,3]*Department of E&TC, PCCOER, Ravet, Pune, India*
E-mail: [1]srushti2000n@gmail.com , [2]srushti2000n@gmail.com, [3]ishikabijwe01@gmail.com

***Abstract*—Human communication is entirely dependent on speech and writing. As a result, visually challenged persons can collect information through voice. This article proposes an easy-to-use and comfortable solution based on the Raspberry Pi 3 for visually impaired people who want images or text transformed to their native language for simple comprehension. The Pi camera is used to capture images, which are then transformed into scan images. The scan picture is sent into the Tesseract OCR (optical character recognition) programmed, which converts the image into text.**

***Keywords*—Raspberry Pi camera, optical character recognition (OCR), tesseract , raspberry pi 3 , translator**

I. INTRODUCTION

The greater part of the areas on the planet are getting digitized at an extremely fast speed. However, there is a need to assemble items that can be utilized to take care of issues on everyday premises [1]. Some of these issues are, individuals with visual disability track down it hard to pursue, restricted education among individuals, individuals who have different learning types, and so on. Outwardly incapacitated individuals account different troubles to comprehend printed text involving existing innovations, for example, issues with arrangement, center, exactness, versatility and proficiency.

Remembering the issues looked by outwardly hindered people; we intend to propose a feasible, serviceable technique for change of picture Text to Speech in the client required dialects. The proposed model empowers the clients to hear out the substance, present in the picture, without holding back in the ideal language [2]

While it might appear to be nonsensical to determine every one of the issues immediately, a straightforward and vvy answer for the previously mentioned issues is a reasonable text to discourse changing over gadget with high accuracy. There are many gadgets made but those gadgets have no provision if the image is kept in a dark room.[1] So we have implemented a module such that it will sense the light in surroundings and the LED will glow.

This paper presents such a wise message understanding framework. It is built utilizing Raspberry Pi 3 B+ model. The item is amalgamated utilizing the accompanying subsystems: camera module, Tesseract OCR (Optical Character Recognition) engine, Python module-based TTS (Text to Speech) synthesizer.[3] The camera associated with the Raspberry Pi goes about as an information gadget which is utilized to take care of text picture. The caught picture put away in the Raspberry Pi is handled utilizing specific picture handling calculations and afterward is taken care of to OCR for character recognition. The OCR changes over the message present in the picture into binary coded text which is subsequently given for semantic checking. The analyzed text report is given to the TTS module which plays out the undertaking of changing text over to speech which is audible.

II.RELATED WORK

There are several technological devices available today to assist blind people. The first system that springs to mind when we think about introducing these technologyis 'Brail.' To represent a sign in this method, 6 raised dots in a 2X3 grid are used in each cell.[4] This symbol represents a linguistic character, punctuation, or any other indication meaning. The dots are detected with the finger, and information is derived from this pattern. The right column is numbered 1-3, while the remainder are numbered 4-6.[5] There will be 64 (26) various types of patterns in each cell.[6] Each dot pattern represents a letter from any language, including punctuation, symbols, and so on. If one cell is insufficient to represent a sign in any language, a multi-cell is utilized. Another unique cell is employed to allow a multi-cell pattern in order to make the pattern recognized. Braille comes in a variety of forms for the blind and visually handicapped. Grade-1 Braille is a one-on-one communication.[7] It does not allow abbreviations or words, whereas grade-2 Braille is the advanced form that allows abbreviations and words, saving space in printing. Grade-1 Braille is a one-on-one communication. It does not allow abbreviations or words, whereas grade-2 Braille is the advanced form that allows abbreviations and words, saving space in printing. In grade-3 type Braille, an unstandardized shorthand format is utilized, which is not used in publications.[3].This is shown in the Fig1

1. III. METHODOLOGY

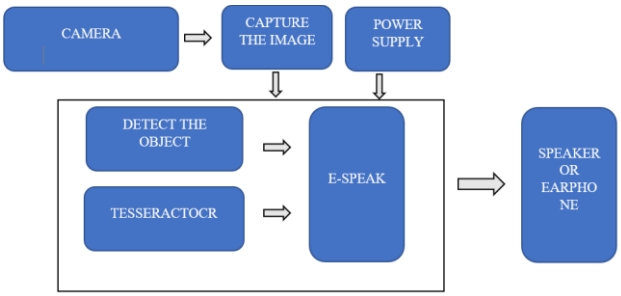

Fig 1. Block Diagram

The raspberry pi controller processes the data and with the help of OCR installed in raspberry pi the characters

This is an open access article

identification in the text data is carried out. OCR engine the extraction of the text in the image is done using optical character recognition (OCR).

OCR is a field of research in pattern recognition, artificial intelligence and computer vision. It is the conversion of the images of typed, handwritten or printed text into a digital text or computer format text.

Earlier OCR versions had to be trained in each character of a text with its specific font. Today, advanced OCRs are available that have a high degree of accuracy, support a wide variety of image formats, languages and fonts.

For our project, we have used Tesseract OCR. It is the most accurate open-source OCR engine and is powered by google. It can be used on the Linux, mac and windows platform.

The newest Tesseract version, 3.4 supports a hundred languages. However, images must undergo a number of pre-processing stages like noise removal, scaling etc. otherwise the output will be of low quality.

The raspberry pi then converts this data into audio form with the help of TTS (text to speech) converter's software The process of converting text to speech by a computer is called speech synthesis. A text to speech system (TTS) is used to perform speech synthesis. A TTS is composed of two parts: front end and back end. The front end converts the text to a symbol. The back end then converts the phonetic into sound.

IV. EXPERIMENTAL RESULTS

1.HARDWARE SETUP

A. RASPBERRY PI 3 B+

With an upgraded processor boasting impressive new packaging and improved networking capabilities, the Raspberry 3B+ sits head and shoulders above its predecessor the Raspberry Pi 3.

B.WEB CAMERA

A minimal expense, 5 Mega Pixel camera is utilized in the framework, it is utilized to catch Images from which the text or content is separated for converting to speech.

C. EARPHONE/SPEAKER

The earphone is connected to the Raspberry Pi audio jack. You can also use speakers for audio output.

D.LED LIGHT AND SENSOR

When the image is kept in a dark room the light sensor will sense the light and glow the LED accordingly.LDR sensor is used here to detect the presence of light in surroundings.

2.SOFTWARE SETUP

A. OPENCV

OpenCV (Open-SourceComputer Vision Library) is an open-source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

B. TESSERACT OCR

The captured image is first enhanced, and character recognition is done either by online or offline methods. In the offline method, Tesseract library and Python programming are used. Here the text files are processed by various libraries like OpenCV, NumPy. In the online method, Google Cloud Vision is used.

C. RASPBIAN OS

Raspbian is a free operating system based on Debian optimized for the Raspberry Pi hardware. An operating system is the set of basic programs and utilities that make your Raspberry Pi run. The Pi's official Raspbian OS is loaded with software to teach users about programming, including the drag-and-drop coding tool Scratch, and various utilities for writing and debugging using the Python programming language.

D. E-SPEAK

In this method the extracted text is converted into speech using the speech synthesizer. For this process we used an e-speak TTS engine or google speech synthesizer. The output of the speech synthesizer is in the form of sound or audio format. The flowchart is shown in Fig2.

FLOWCHART

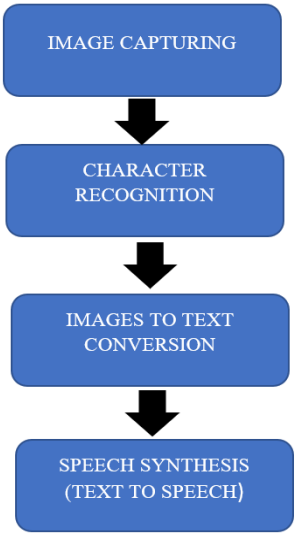

Fig2. Flow Chart of Raspberry pi-based reader for blind

The above diagram shows the flow of the process involved. In the first step the image is captured. In the next step the image is converted in the text format using the optical character recognition method. The extracted text file is then converted to speech.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

Input: Fig 3 shows the input screen shot.

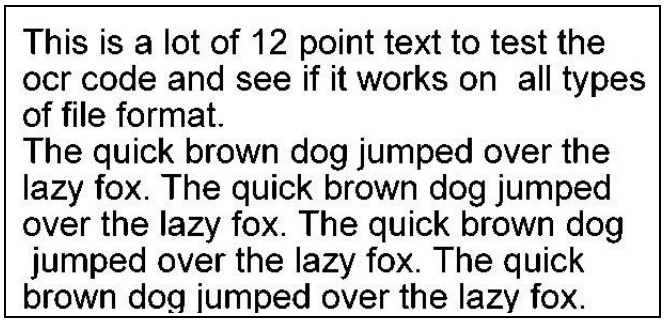
This is a lot of 12 point text to test the ocr code and see if it works on all types of file format.
The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox. The quick brown dog jumped over the lazy fox.

Fig 3. Input Text

Image to text file conversion:

Fig3 shows the input we provided and Fig4 is the image to text conversion of the provided image.

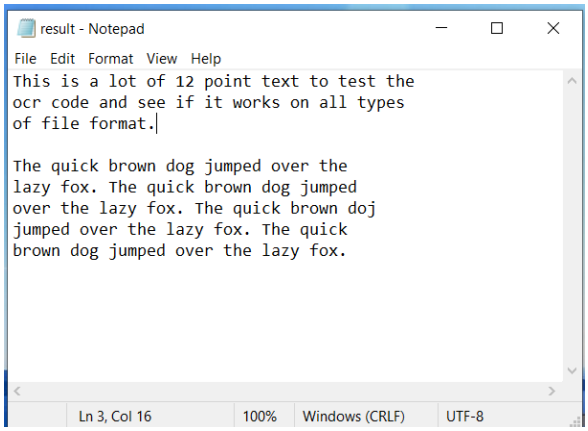

Fig 4. Output screenshot

VI. CONCLUSION AND FUTURE SCOPE

The design and implementation of a functional model in this Smart Reader for Blind People provides smart reading assistance for the Blind and Visually Impaired. The gadget provides a smart way to recognize images. The image is converted to text and text is converted to speech. This helps the blind person to understand what is written on the page.

The machine may be made smarter in the sense that when a user requests that a specific web page be read, if the page N.mp3 file already exists, the system can skip the image-to-speech conversion and instead play the mp3 file immediately. If the user needs to browse to a certain page number, the gadget will activate the page turning mechanism to take them there. Recently, various deep learning based techniques [8-11] have been presented for distinct application which can be applied for text recognition in future.

VII. REFERENCES

[1] L. George, R. Salim, R. Baby, and A. Thomas, "Raspberry PI Based Reader for Blind," *J. Opt. Commun. Electron.*, vol. 5, 2019, doi: 10.5281/zenodo.3570357.

[2] A. Prakash Kotkar and P. V Baviskar, "RASPBERRY PI BASED READER FOR BLIND." [Online]. Available: http://aegaeum.com/

[3] S. Kurlekar SITCOE, I. A. Onkar Deshpande SITCOE, I. V Akash Kamble SITCOE, I. A. AniketOmanna SITCOE, and I. B. Dinesh Patil SITCOE, "Reading Device for Blind People using Python, OCR and GTTS," 2018. [Online]. Available: www.ijsea.com49

[4] Institute of Electrical and Electronics Engineers and Hindustan Institute of Technology, *Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) : 02-04, July 2020*.

[5] P. Manage, V. Ambe, P. Gokhale, V. Patil, R. M. Kulkarni, and P. R. Kalburgi Math, "An intelligent text reader based on python," in *Proceedings of the 3rd International Conference on Intelligent Sustainable Systems, ICISS 2020*, Dec. 2020, pp. 51–54. doi: 10.1109/ICISS49785.2020.9315996.

[6] J. James, M. Aldo, and A. Andrew, "Image Text to Speech Conversion Using OCR Technique in Raspberry PI," 2018. [Online]. Available: http://www.ijser.org

[7] Dayananda Sagar College of Engineering, Institute of Electrical and Electronics Engineers. Bangalore Section, and Institute of Electrical and Electronics Engineers, *2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020) : conference proceedings : 5-7 March, 2020*.

[8] Bhangale, KishorBarasu, and MohanaprasadKothandaraman. "Survey of Deep Learning Paradigms for Speech Processing." *Wireless Personal Communications* (2022): 1-37.

[9] Bhangale, KishorBarasu, and K. Mohanaprasad. "A review on speech processing using machine learning paradigm." *International Journal of Speech Technology* 24, no. 2 (2021): 367-388.

[10] Bhangale, Kishor B., and K. Mohanaprasad. "Content based image retrieval using collaborative color, texture and shape features." *Int. J. Innovative Technol. Exploring Eng.(IJITEE)* 9, no. 3 (2020).

[10] Bhangale, Kishor, Piyush Ingle, RajaniKanase, and DivyashriDesale. "Multi-view Multi-pose Robust Face Recognition Based on VGGNet."In *International Conference on Image Processing and Capsule Networks*, pp. 414-421.Springer, Cham, 2021.

[11] Bhangale, Kishor, and K. Mohanaprasad. "Speech Emotion Recognition Using Mel Frequency Log Spectrogram and Deep Convolutional Neural Network." In *Futuristic Communication and Network Technologies*, pp. 241-250. Springer, Singapore, 2022.