Image Text to Speech Conversion using Raspberry Pi & OCR Techniques

Nikhil Mishra
M.Tech Student
Department of Electronics & Telecommunication Engineering
AGPCET Nagpur

Abstract— Image Text to Speech Conversion can be used to read labels and text on banners of different sizes. To capture the image of the object we propose a camera or a wi-fi/ Bluetooth based device. These devices will capture the image with the help of raspberry pi and camera modules associated with it. In the extracted ROI, text finding and recognition are conducted to acquire text information. To automatically understand the text from the image we propose a tesseract optical character recognition(ocr) module which will understand the image. Text obtained from the image is then localized and converted to text format. The recognized text codes are output to users in speech. Performance of the proposed text recognition methods are proposed on normal paper with large font text and banners of large size. Experimental results demonstrate that our algorithm achieves the state of the arts. The proof or our concept is also evaluated on normal images and banners. We explore user image capturing issues and assess robustness of the algorithm in extracting and reading text from different objects with complex backgrounds.

Key words: OCR, Image Text To Speech, Raspberry-Pi, Wi-Fi Based Camera Module, Image Capturing, Tesseract OCR, Pi Camera

I. INTRODUCTION

Optical Character Recognition is major term which covers all types of character recognition [6]. The Character recognition has become a major field which has been used significantly by human beings to create digital libraries, restore lost data, understand different languages etc. OCR is electronic identification and digital conversion of printed material or typed material, it can also be used in handwritten materials and banners. Using OCR allows computer or various devices to read or scan standalone images and convert those standalone images to editable and search-able data. OCR typically includes three major procedures: scanning and or opening a document in OCR System or software and convert them into editable, readable and understandable data by machines, and then saving the output of OCR software created document in format of our choice.

Optical Character Recognition can be used widely in healthcare applications to aid blind people. We have proposed a system which is used to help blind people to read books, understand different text on banners, pictures and large ad boards. OCR includes mainly three components the camera to capture the images, the programmable system to convert the captured camera into whatever format we want and finally the output system to show the output of OCR.

We are using raspberry pi a compact size and low weight system for creating an OCR system. The raspberry pi is connected with the help of ethernet cable to the computer so that we can get the display. We have proposed capturing of image through raspberry pi camera or through android camera which can be connected with ethernet, Bluetooth or wi-fi. The webcam or raspberry pi camera capture the images large font size image and store it into raspberry pi system. We are using tesseract OCR library provided by python to convert the captured image to the text. The image text will be saved in .txt file or as per the user requirement. The camera and tesseract takes 5-7 minutes to capture and convert the image to text.

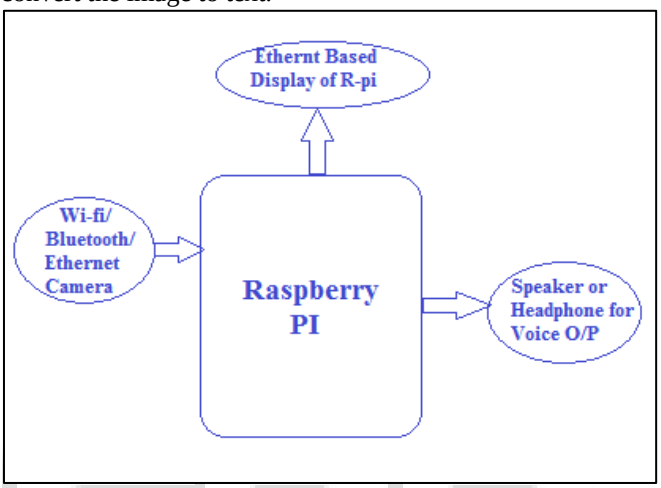

Fig. 1: Raspberry Pi Based OCR System Block Diagram

The display of raspberry pi is shown on the computer display with the help of putty software and ethernet cable. The camera is connected to the raspberry pi system which works as an eye for OCR system. The camera usually takes 5-7 minutes to capture and process the images. With the help of tesseract OCR library of python, the captured image is converted to the text format in the raspberry pi hard disk location. The converted text is provided to TTS system which converts the text to the voice format.

II. METHODS

Raspberry Pi OCR consists of two main methods, the Image Processing method and voice processing method. The image processing method includes the image capturing and image to text conversion. Voice processing method includes captured image to speech conversion. The speech conversion can be optimized i.e. the voice can be changed to male or female voice so that it can be understood by the specific user. We can also control the echo for the voice generated from voice processing method.

A. Image processing method

The image processing is done with the optical character recognition method. The optical character recognition is a method that captures or scans the images and has an ability to convert the image into readable or text format which can be processed further. The image captured with OCR can be of any resolution but we have preferred 720p resolution for the images of .jpg or .jpeg format.

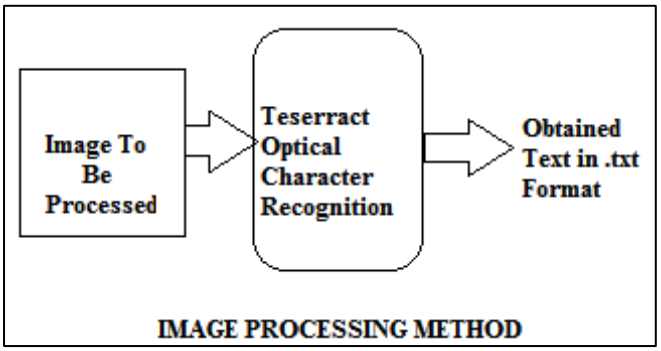

Fig. 2: Image Processing Method Block Diagram

The image processing method includes capturing of static image with the help of camera. The camera works as an eye for the raspberry-pi[4]. The camera can be connected to the raspberry-pi with the help of LAN Cable, Wi-fi or Bluetooth. We have primarily used a raspberry pi camera to capture the image. After the successful connection the image is captured with the help of tesseract OCR software. There are various OCR tools such as gocr, ocrad, ocrfeeder, ocropus, teserract-ocr and cuneiform. We are using tesseract OCR which is raspberry pi compatible and can understand primarily English language. The teserract-ocr is library provided by the Linux community which is open source. The Tesseract-ocr is command line OCR which captures the image on the press of button i.e. on execution of instruction. The image can be saved in .jpeg or .png format. The captured image is converted to the .txt format and saved with the same name as an image.

B. TESERRACT OCR Limitations

The teserract ocr library installed in the raspberry pi with the help of import function in python. The teserract OCR can produce outputs in various formats such as txt, pdf, hosc, tsv and pdf with text layer only. The teserract cannot handle any color 24 or 34 bit. It can read only 1-bit binary image or 8-bit greyscale fixed in 2.04. The minimum character size tjat can be read by Teserract OCR Engine is 20 pixel uppercase letters. Accuracy of Teserract OCR decrease with font size of 14pt.

C. Voice Processing Method

In this method our main aim to convert obtained text to speech with the help of coding in raspberry pi. The text to speech module is installed in raspberry pi. Various methods and synthesizers can be used in raspberry pi to convert obtained text to speech. Those modules are pyttsx text to speech, NS Speech Synthesizer, SAPI 5 for Windows XP, Vista and Windows 7, espeak on any platform that can host share library such as ubuntu and Fedora Linux and Google Text to Speech. We are using pyttsx or festival for converting obtained text to the speech.

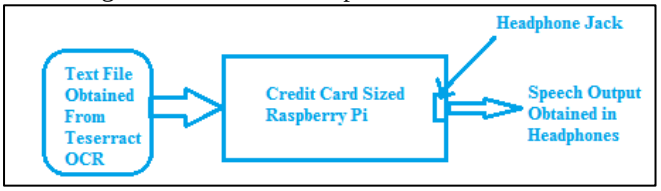

Fig. 3: Voice Processing Method

We will provide obtained text file to the text to speech synthesizer with the help of text file input command. We have installed pyttsx in raspberry pi with the help of command sudo apt install pyttsx and festival with command sudo apt-get install festival. These commands will install pyttsx for text to speech conversion in raspberry pi and festival for the same. We can also use Google Text to Speech or pico software for text to speech conversion.

III. DESIGN IMPLEMENTATION

The Testing was done on the Raspberry Pi which is credit card sized and easily portable with following specifications:

- Raspberry Pi 3 Model B Quad-Core 64 Bit
- 1GB of LPDDR2-900 SDRAM
- 5MP Camera
- Headphones for Voice
- Broadcom BCM2837 SoC
- SanDisk 16GB microSD Card Bootable

Steps Followed:

A. Importing Libraries:

We have imported various libraries, functions and sub processes which are required by the raspberry pi.

B. Main Program

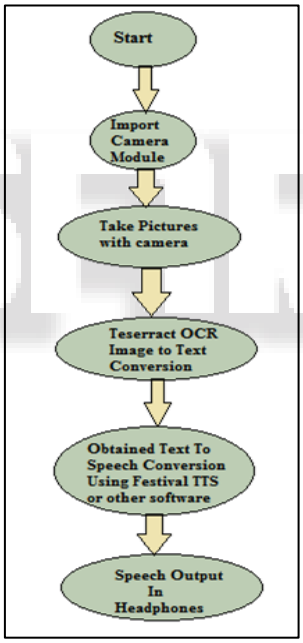

Fig. 4: Process Flow Block Diagram

Fig. 4 Shows process flow block diagram which includes importing GPIO pins, Importing pyttsx, festival, tesseract ocr and all the other required software's.

IV. RESULT

The output observed for image text to speech conversion is

- Image is Captured with The Camera
- The captured image is converted to the text and saved at same location of image
- It takes approximately 7-8 sec to convert the text
- The converted text is processed with the Festival or pyttsx
- The speech is obtained as an output in Headphones
- It recognizes numbers as well as letter in English
- Range of reading Distance is 30cm.

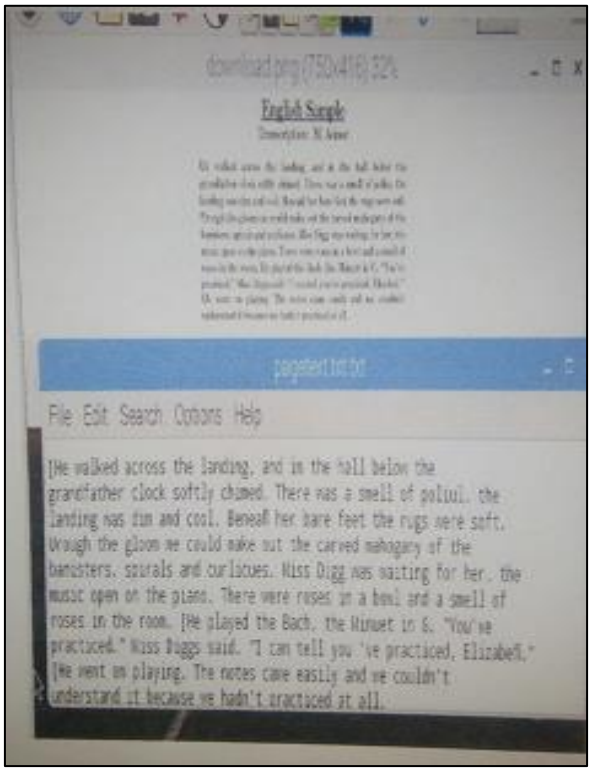

Fig. 5: Converted Image to Text Result

Fig. 6: Project Image with Camera and Headphones

Figure 5 and 6 shows the results and image of image text to speech conversion module. In figure 5 the upper section shows the image which is captured by the camera and the lower section shows text file which is obtained as an output of teserract-ocr.

V. CONCLUSION

Image Text To speech device is completely portable and can be connected to eyeglasses which will capture the images rapidly and will convert the text to speech and can be used by blind people to understand banners and texts. The Image processing time is 3 mins and output speech has tolerance of less than 2%.

REFERENCES

[1] R. Mithe, S. Indalkar and N. Divekar. " Optical Character Recognition" International Journal of Recent Technology and Engineering (IJRTE)", ISSN: 2277-3878,Volume-2, Issue-1, March 2013.
[2] Kanimozhi.V.M, Muthumani.I (2014), "Optical Character Recognition for English and Tamil Script", International Journalof Computer Science and Information Technologies, Vol. 5 (2) ,2014, 1008-1010, pp 1008 – 1010.
[3] Raspberry pi cook book by simon monk.
[4] T.Rubesh Kumar, C.Purnima, " Assistive System For Product Label Detection With Voice Output For Blind Users ", IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 6, Dec-Jan, 2014, pp 1- 4.
[5] GeorgiosVamvakas, Basilis Gatos, Stavros J. Perantonis, "Handwritten character recognition through two-stage foreground sub-sampling", Pattern Recognition, Volume 43, Issue 8, August 2010
[6] ARCHANA A. SHINDE, D. "Text Pre-processing and Text Segmentation for OCR. International Journal of Computer Science Engineering and Technology", pp. 810-812, 2012.
[7] SEETHALAKSHMI R., SREERANJANI T.R., BALACHANDAR T., Abnikant Singh, Markandey Singh, Ritwaj Ratan, Sarvesh Kumar (2005), "Optical Character Recognition for printed Tamil text using Unicode", Journal of Zhejiang University SCIENCE ISSN 1009-3095, pp1297 – 1305.

