

Image Text to Speech Conversation and Face Detection using Raspberry PI

Pranali N. Humne¹ Dr. D. M. Bhalerao²

^{1,2}Department of Electronics and Telecommunication Engineering

^{1,2}SCOE, Pune, India

Abstract— Reading text image and text boards is hard for visually impaired persons. A person needs vision to access the information in text. The proposed system reads the text from image text and convert it into speech by using e-speak algorithm. The proposed system includes text extraction from scanned text image using tesseract optical character Recognition (OCR) and convert it into the speech using e-speak tool. The main aim of proposed system is to provide assistance to visually challenged persons. The faces can also be detected when persons move into frame using mode control. This paper proposes a camera based assistive text reading to provide an assistance to the visually impaired persons.

Key words: Raspberry pi; Tesseract OCR Engine; Google Speech API; Webcam

I. INTRODUCTION

In worldwide the visually impaired peoples are 314 million, in that 45 million are visual impairment which was released by “World Health Organization” in 10 facts regarding blindness. The estimation of The National Health Interview Survey 25.2 million adult Americans (over 8%) are blind or visually impaired. This number is increasing rapidly as the baby generation age

Reading is essential in today’s society. In day to day life text and sign is keeps an important role and makes our life easy. In recent technological developments in mobile phone, computers, camera make it easy to assist the blind persons by developing the camera based application that unite computer vision tools with other products such as Tesseract and Optical Character Recognition (OCR) [9] system.

In this proposed system open computer vision is used as library of functions is used for implementing image processing techniques. Text extraction from an image is method of conversation of image writings on paper, sign boards, printed books etc to text. The binary image is converted to text by Tesseract library in OCR [4] engine that detects the outline, slope, pitches, joint letters and white spaces. The e-speak algorithm is used to convert text file into voice output. The e-speak is text to speech system (TTS). Speech synthesizer is artificial production of human speech. This proposed idea aims to develop an efficient assistive text reading system using camera. The idea involves text extraction from capture text image then converted into speech. It also used to discover persons face in the frame. Raspberry pi is used to develop the prototype [6].

II. LITERATURE SURVEY

In this section, we present survey of some earlier work for supporting visually impaired persons with text to speech technology and face recognition.

The paper [1] proposed a camera based assistive text reading framework to help visually challenged persons to

read labels from handheld objects in day to day life [1]. In this paper Region of Interest (ROI) is define with the help of motion based method, for isolating the interested object from other surrounding objects or from unorganized backgrounds in the camera vision. To extract moving object region a mixture of Gaussians based background subtraction technique is used. Text localization and recognition are conducted, to get text details from the Region of Interest (ROI). In an Adaboost model the Novel Text Localization Algorithm is performed to find the distribution of edge pixels and the gradient features of stroke orientations. Thus the localized text region is converted into binary format and recognized by identification software or optical character recognition.

In [3] describes the camera based text reading system for blind person. In this prototype global or local thresholding are used to create the binary image which can be decided from fisher’s Discriminant Rate(FDR). In OTSU’s banalization method local histogram has two peaks when character present on frame and this consider as high value for FDR. When histogram has only one peak and value of FDR is small then it is quasi uniform frames. In complex area the histogram is scatter resulting in higher FDR values.

The proposed model is done image processing using Open Computer Vision [8] and Tesseract [9]. The web cam is used to capture the text image and person’s face.

III. PROPOSED METHOD

The proposed method is to help visually impaired person or blind person in reading label on product, text written on label. The idea includes text recognition and face detection from image taken by web cam and identify and recognizes the text using OCR. The e-speak algorithm is used to convert recognized text file to voice output with the help of Raspberry pi and webcam proposed idea is develop that work in real time.

The proposed model can operate in two modes as shown in Fig. 1. Using mode control switch mode are selected between text and face mode. The model capture image using webcam and check the frame for the text in the frame and also check the of face in the frame, and if present then notify through audio output message. When web cam capture image and check for character if character is found then user will informed that text is detected in the frame. Thus with help of text to speech software user will able to hear the voice output of the message if user wanted.

The web cam is used to capture the image then captured image is convert into gray and then it is filter to lower the noise level in the frame. The Gaussian filtered image is then converted into binary format. The portioned of image with no character are deleted. The new image without deleted portioned is then loaded into the Tesseract OCR[4] to execute text recognition. Thus output of Tesseract is text file

which is then given as input to the e-speak. The e-speak gives output in analog signal which is corresponding to the text file.

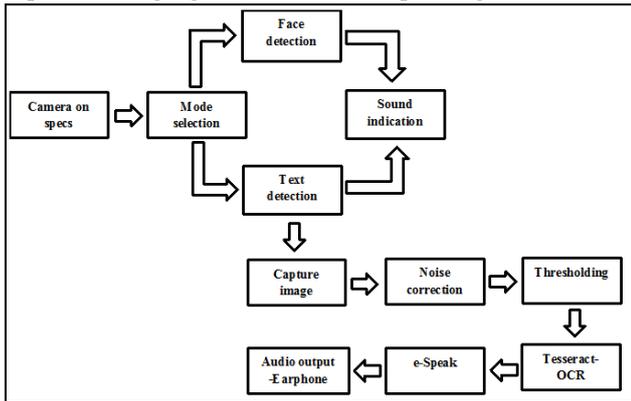


Fig. 1: System architecture

IV. SYSTEM HARDWARE DESIGN

The hardware consists of the following component: Raspberry pi 3 [model B] mounted with SD card, web camera, speakers, Internet connection via Ethernet or Wi-Fi, laptop.

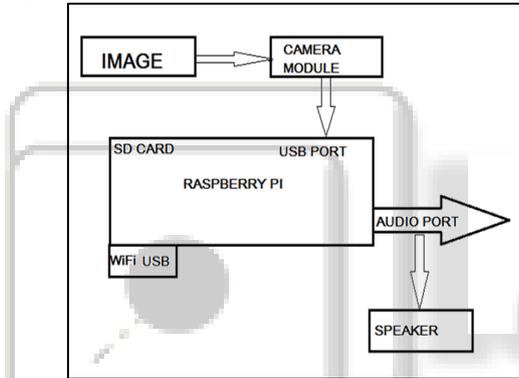


Fig. 2: System hardware design

A. Raspberry pi 3 Model B

The Raspberry pi [6] is low cost, credit-sized single-board computer developed in UK by Raspberry pi Foundation. An 8 GB SD card is loaded with Raspbian OS [6] and is put it into a slot for network connection.

The Raspberry pi [6] has a Bradcom BCM2835 system on a chip which includes an ARM1176JZE 700Mhz processor Video Core IV GPU and was originally shipped with 256 MB of RAM, later upgraded to 512 MB.

B. Camera

A compactable camera QHM 500LM [7] is used for image capturing. It has built-in mic with noise reduction, Interpolated to 25 Mega pixel, 10 level zoom: on live motion picture night Vision: 8 bright light.

V. SYSTEM SOFTWARE DESIGN

The System software design consists of various phases which help to produce the output. Python language are use to write the algorithm python language is a script language. The functions in algorithm are called from the open CV library [8]. Open CV is an open source computer vision library [8]. Tesseract input is binary image with optional polygonal text region defined [9].

The system software design consists various phases: Image to text conversation (OCR), Text to speech. First Camera capture the image and stores it as image file with .jpg extension. The optical character recognition (OCR) engine is used to convert it from image file to text file. OCR engine extracted the number and character and provided the text is printed. Using e-speak algorithm it is then converted into .flac file. The flac file is given as an input to a python program which gives a translated speech and text output using Google text to speech engine.

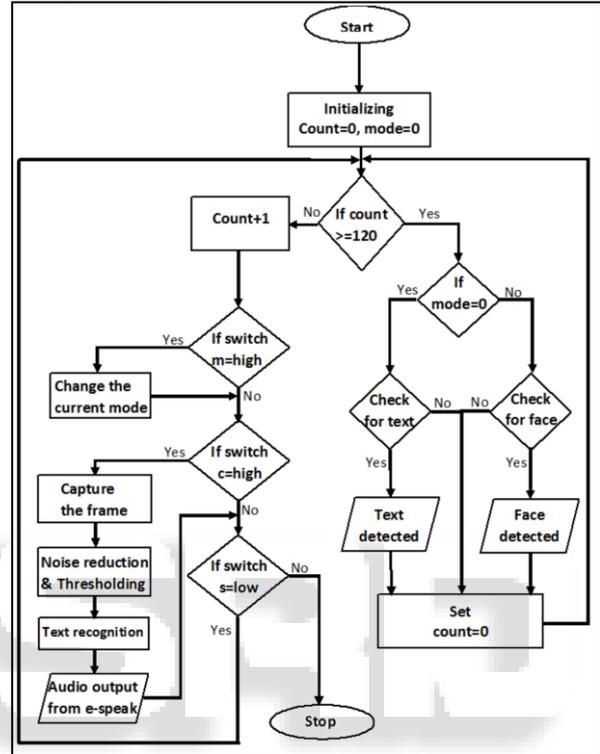


Fig. 3: Flow Chart

Flowchart of model is described in Fig. 3. First the initializes value of mode and count set to zero. Mode 'm' value is select text or face modes and count value is to hold the number of frames. When 'c' value raises up to is 120 frames the system search for face or text depending upon the mode 'm'. The system uses c and m switches. When c value is high it capture image and does the processing the switch m is used to control the mode of the system.

A. Image to Text converter

The web camera QHM 500LM [7] take the image which is to read and image file is stored in the folder. The tesseract[4] software is installed in the Raspberry pi[3]. Image to text converter software is to convert the .jpg file to .txt file by drawing out the text from image and saving it in the .txt file extension.

B. Text To Speech Synthesizer

A Text to Speech engine (TTS) engine is used to converts text written in image or text present in the image into a sound or audio representation. Phonemic representation is converted to waveform that can give a sound output. e-speak is a software which can easily installed. e-speak engine is used to convert text file into a audio file using extension .flac. flac stands for free lossless audio codec. Hence audio file is created.

VI. CONCLUSION

In this paper, we used various algorithm to extract text from image text and also detected the face of person. The model consists web cam to capture image and store it into file.

REFERENCES

- [1] Rajkumar N, Anand M.G, Barathiraja N, "Portable Camra Based Product Label Reading For Blind People" IJETT, Vol. 10 Number 11-Apr2014J.
- [2] Ms.Rupali, D Dharmale, Dr. P. V. Ingole, "Text Detection and Recognition with Speech Output for Visually Challenged person", Vol. 5 Issue 1, January 2016.
- [3] Ezaki, Nobuo, et al. "Improved text-detection methods for a camera-based text reading systems for visually impaired persons" Eighth International Conference on Document Analysis and Recognition (ICDAR'05) IEEE, 2005.
- [4] Rithika. H, B. Nithya santhoshi, "Image Text To Speech Conversation In Desired Language By Translating with Raspberry pi" International Conference on Computational Intelligence and computing Research IEEE, 2016
- [5] Mr Rajesh M, Ms Bindhu K. Rajan, Ajay Roy, Almaria Thomas K, "Text Recognition And Face Detection Aid For Visually Impaired Person Using Raspberry Pi" International Conferences on Circuits Power and Computing Technologies [ICCPCT], 2017
- [6] www.raspberrypi.org
- [7] <http://www.qhmpl.com/QHM500-8LM-%20Web-Camera.html>
- [8] <https://opencv.org/>
- [9] <https://opensource.google.com/projects/tesseract>