ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

Image Text to Speech Conversion Using OCR Technique in Raspberry Pi

K Nirmala Kumari[1], Meghana Reddy J[2]

Associate Professor, Dept. of ECE, BIT College, Bangalore, Karnataka, India[1]

PG Student [VLSI and E.S], Dept. of ECE, BIT College, Bangalore, Karnataka, India[2]

ABSTRACT: In this paper an innovative, efficient and real-time cost beneficial technique that enables user to hear the contents of text images instead of reading through them as been introduced. It combines the 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 from colour images is a challenging task in computer vision. Text-to-Speech is a device that scans and reads English alphabets and numbers that are in the image using OCR technique and changing it to voices. This paper describes the design, implementation and experimental results of the device. This device consists of two modules, image processing module and voice processing module. The device was developed based on Raspberry Pi v2 with 900 MHz processor speed.

KEYWORDS: Text Extraction, OCR, Text-To-Speech, Image Processing, Voice Processing

I.INTRODUCTION

Optical character Recognition (OCR) is a conversion of scanned or printed text images [1], handwritten text into editable text for further processing. In this paper, we have presented a robust approach for text extraction and convert it to speech. Testing of device was done on raspberry pi platform. The Raspi is initially connected to the internet through VLAN. The software is installed using command lines. The first setup is to download the installation script, second command is to convert it to executable form and the last command starts the script which does the rest of the installation work. Device set up is done as in Fig.1 .The webcam is manually focused towards the text. Then, to take a picture, press pushbutton switch. Pushbutton switch set up is done as in fig.2 .A delay of around 7 seconds is provided, which helps to focus the webcam, if it is accidently defocused. After delay, picture is taken and processed by Raspi to hear the spoken words of the text through the earphone or speaker plugged into Raspi through its audio jack.

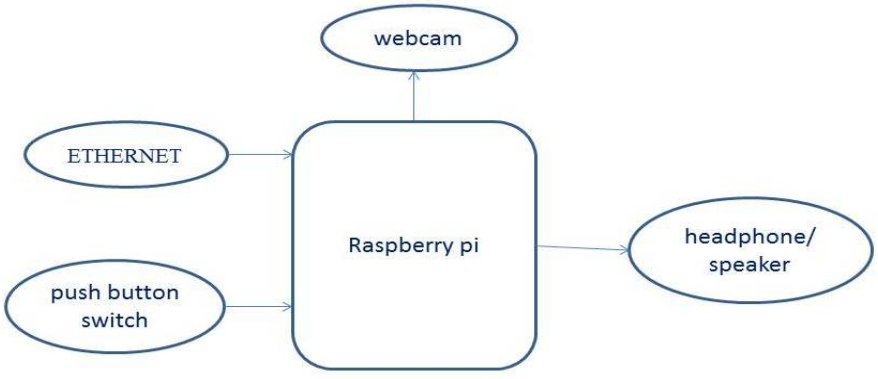

Fig.1 Basic Block diagram of image text-to-speech conversion

Fig.1 shows the hardware attached to raspberry pi board.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

A. Pushbutton Switch Set up

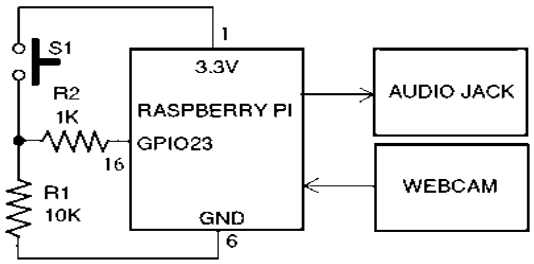

Fig.2 Circuit connection to Raspi board

In Fig.2, When the GPIO pin is set as input, it is floating and has no defined voltage level, to detect whether the input is high or low, simple resistive circuit is needed so that it is always connected and reads either high or low voltage. One of the terminals of switch S1 is connected to ground (GPIO pin 6) through pull-down resistor R1 of 10kiloohm. The other terminal is connected to 3.3V of GPIO pin 1.When S1 is pressed, a high voltage is read on GPIO pin 16. When S1 is released, GPIO pin 16 is connected to ground through R1, hence a low voltage is read by GPIO pin 16.When pushbutton S1 is pressed, the webcam takes a picture of the text (after delay).

II.METHODS

Text-to-speech device consists of two main modules, the image processing module and voice processing modules (Fig.3). Image processing module captures image using camera, converting the image into text. Voice processing module changes the text into sound and processes it with specific physical characteristics so that the sound can be understood.Fig 3 shows the block diagram of Text-To-Speech device, 1st block is image processing module, where OCR converts .jpg to .txt form. 2nd is voice processing module which converts .txt to speech

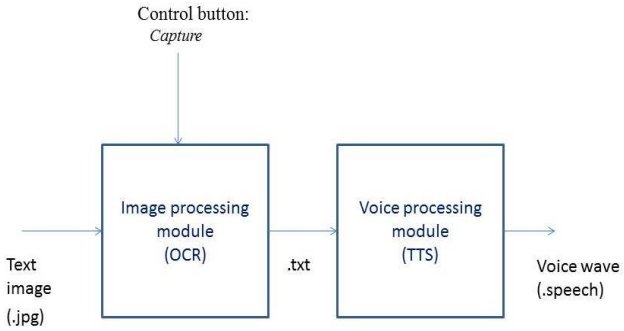

Fig.3 Block Diagram of Text-To-Speech Device

Fig 3 shows the block diagram of Text-To-Speech device, 1st block is image processing module, where OCR converts .jpg to .txt form. 2nd is voice processing module which converts .txt to speech.

A. Image Processing Module Using Optical Character Recognition

OCR is important element in this module. OCR or Optical Character Recognition is a technology that automatically recognize the character through the optical mechanism, this technology imitate the ability of the human senses of sight, where the camera becomes a replacement for eye and image processing is done in the computer engine as a substitute for the human brain [2]. Tesseract OCR is a type of OCR engine with matrix matching [3]. The selection of Tesseract engine is because of its flexibility and extensibility of machines and the fact that many communities are active researchers to develop this OCR engine and also because Tesseract OCR can support 149 languages. In this project we

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

are identifying English alphabets. Before feeding the image to the OCR, it is converted to a binary image to increase the recognition accuracy. Image binary conversion is done by using Imagemagick software, which is another open source tool for image manipulation. The output of OCR is the text, which is stored in a file (speech.txt) .Machines still have defects such as distortion at the edges and dim light effect, so it is still difficult for most OCR engines to get high accuracy text [4]. It needs some supporting and condition in order to get the minimal defect.

Tesseract OCR Implementation

The input image captured by the Logitech (C270) web camera has a size of 3 MPI (720 X 340 pixels). Based on the specifications of the Tesseract OCR engine, the minimum character size that can be read is 20 pixels uppercase letters. Tesseract OCR accuracy will decrease with the font size of 14pt.

Software Design

Software processes the input image and converted into text format. The software implementation is showed in Fig. 4

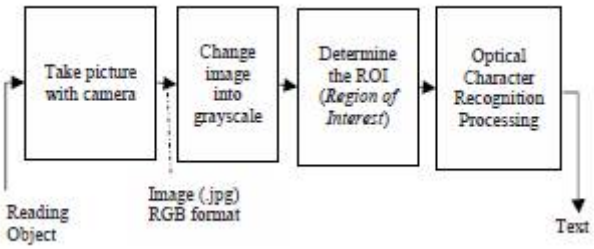

Fig.4 Software Design of Image Processing Module

The image is taken by the user via GPIO pin (23) that is connected to the button, using interrupt function. Furthermore, the picture is taken by using raspistillprogram with sharpness mode to sharpen the image. The resulting image has a .jpg format with a resolution of 720 x 340 pixels.

B. The Voice Processing Module

In this module text is converted to speech. The output of OCR is the text, which is stored in a file (speech.txt). Here, Festival software is used to convert the text to speech. Festival is an open source texttospeech (TTS) system, which is available in many languages. In this project, English TTS system is used for reading the text.

III. DESIGN IMPLEMENTATION

The testing was done using Raspberry Pi platform with the following specifications:

- SBU Raspberry Pi 2 900 MHz Quad Code ARM Cortex-A7
- Logitech 3MP Camera Module
- Bootable SanDisk Ultra 8GB microSD Card

Steps Followed:

1. Import and Initialization:

Import subprocess, time and RPi.GPIO and initialize GPIO pin 23 as input.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

2. Main Program:

The main program provides functions to retrieve and process the input image, convert it into a sound signal. Picture will be taken as soon as push button switch is pressed then this Captured image is thresholded before feeding it to OCR to increase the accuracy. Overall flow of program is done as in flowchart fig 5.

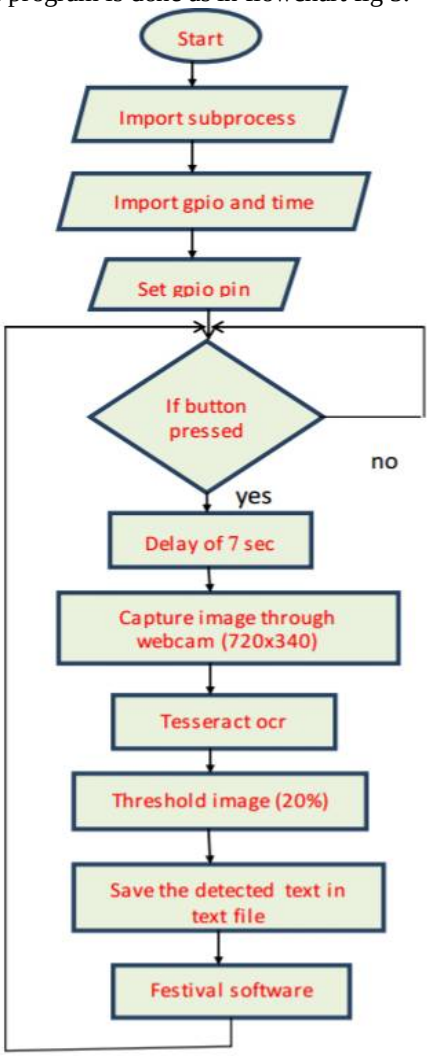

Fig.5 Flow chart of text-to-speech device program

Fig.5 shows the flow chart, all the necessary gpio, time and subprocess is imported in the python code, assign gpio pin to push button switch, if button is pressed after delay image is captured and sent to ocr, else it will be looping till button is pressed.

IV.RESULT

Observed outcome of project:

- Text is extracted from the image and converted to audio.
- It recognizes both capital as well as small letters.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

- It recognizes numbers as well.
- Range of reading distance was 38-42cm.
- Character font size should be minimum 12pt.
- Maximum tilt of the text line is 4-5 degree from the vertical.

Fig.6 a) captured image with color background b) threshold image c) output

In Fig 6, a) shows the image captured through camera which as color background, then this image is thresholded (20%) where image is converted to black and white image as in (b). final extracted text through ocr is shown in (c) .

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 5, May 2016

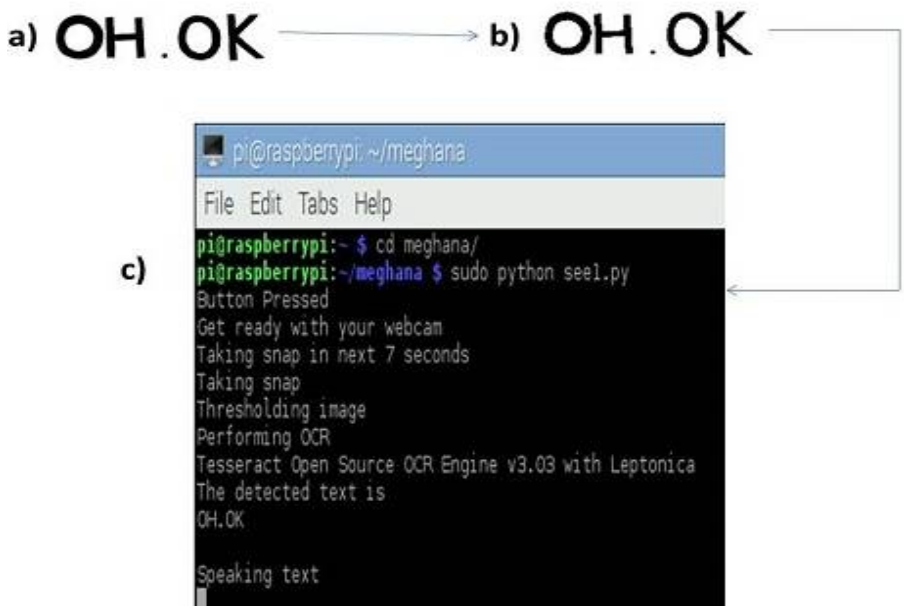

Fig.7 a) captured image with white background b) threshold image c) output

In Fig 7, a) shows the image captured through camera with white background, then this image is thresholded (20%) where image is converted to black and white image as in (b). final extracted text through OCR is shown in (c) .

V. CONCLUSION

Text-to-Speech device can change the text image input into sound with a performance that is high enough and a readability tolerance of less than 2%, with the average time processing less than three minutes for A4 paper size. This portable device, does not require internet connection, and can be used independently by people. Through this method, we can make editing process of books or web pages easier.

REFERENCES

[1] ARCHANA A. SHINDE, D. **"**Text Pre-processing and Text Segmentation for OCR. International Journal of Computer Science Engineering and Technology", pp. 810-812, 2012.
[2] 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.
[3] R. Smith. !An Overview of the Tesseract OCR Engine", USA: Google Inc.
[4] H. Shah and A. Shah. "Optical Character Recognition of Gujarati Numerical", International Conference on Signals, Systems &Automation, pp. 49-53,2009.
[5] Raspberry pi cook book by simon monk.
[6] Text Localization and Extraction in Images Using Mathematical Morphology and OCR Techniques,2013.