Comprehensive Literature Review on Raspberry Pi Based Text Reading System for Visually Impaired Persons

Leena Patil [1], V. D. Chaudhari [2], I. S. Jadhav[3], H. T. Ingale[4]

[1]PG (VLSI & Embedded System) student, [2,3,4] Asstt. Professor
E&TC Engg dept, GF'S Godavari College of Engineering, Jalgaon - 425001

Abstract- *Speech and text is main medium for human communication. a personal needs eyesight access the knowledge during a text. However people that have very poor eyesight can collect information from voice. The proposed system can help the visually impaired persons to read any printed text in vocal form. A specs inbuilt camera is employed to capture the text image from printed text and thus the captured image is analyzed using Tesseract-Opical Character Recognition (OCR). The detected text is then converted into speech employing a compact open source software synthesizer, speak. Finally the synthesized speech is produced by the headphone by provides an interface between camera, sensor & image processing results, while also performing functions to manage the peripheral units. The programing language is employed is python.*

Keywords- Raspberry pi 3, Optical character recognition, Open CV, Text to speech conversion, Python programming.
.

I - INTRODUCTION

Text to speech technology is the process where in the computer is made to speak. It uses the concepts of natural language processing. In Text reading applications, there are many different techniques available such as label reading, voice stick, brick pi reader and pen aiding but these methods can perform text to speech by creating datasets. In order to address this problem, finger reading technique has been developed, it eliminates the datasets created and stored previously and provide a previous response of reading any text given as input captured image. [4]The speech synthesizer converts the audio input into the text form and processes the text to further learning modules. Despite the advancement of technology that allows for storing information electronically, textual information still remains the most common mode of information exchange. Virtually people who could restore normal vision with eye glasses or contact lenses are around 20% from the survey of ABF (www.abf.com) who could lead their normal lives. Apart from them 90% of world`s visually impaired people who live in low, middle and even in most developed countries, cataract remains the leading cause of blindness.

In this system, we wanted the device to be able to detect the text from printed text image and read it efficiently. Inspired by the methodology utilized by Apps like "Cam Scanner". Optical character Recognition (OCR) is a conversion of scanned or printed text images, handwritten text into editable text for further processing.[1] In this paper, we have presented a robust approach for text extraction and convert it to speech. This device was tested on raspberry pi platform. The Raspberry pi is initially connected to the internet through Wi-Fi. The software is installed using command lines. The first setup is to transfer 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 got wind of is finished as in The digital camera is

International Journal of Innovations in Engineering and Science,   www.ijies.net

manually targeted towards the text. Then, to take a picture, press capture button. After that, picture is taken and processed by Raspberry pi to hear the spoken word of the text through the earphone or speaker plugged into Raspberry pi through its audio jack.

II - LITERATURE SURVEY

In A smart spec for the blind persons can developed text detection thereby create a voice output. This smart spec can assist the visually impaired persons to read any printed text in vocal form. A specs inbuilt camera is usages to capture the text image from the printed text and therefore the captured image is analyzed using Tesseract-Optical Character recognition (OCR). [1]The detected text is then converted into speech employing a compact open source software speech synthesizer, e-Speak. Finally output, the synthesized speech is generated via the headphone by TTS method. In this system Raspberry Pi is the chief goal for the implementation, as it provides an interface between camera, sensors, and image processing results, while also developing functions to manipulate outer units (Keyboard, USB etc.,).

In the design involves human face, object and textual recognition which make vision for visually challenged. The smart kit contains an eye fixed glass given camera, an earphone, a microphone and therefore the system where the processing is administered . The camera present at the nose head of eye glass take the printed image of the user as snapshots and transfers to the system where it get processed and produces the specified audio description as output. In This paper denote a camera based assistive text reading to assist visually impaired person in reading the text present on the captured image. When a person enter into the frame by the mode control the faces can also detected. The proposed idea involves text extraction from scanned image using Tesseract Optical Character Recognition (OCR) and converting the text to speech by e-Speak tool, a process which makes visually impaired persons to read the text.[6] this is often a prototype for blind people to acknowledge the products in world by extracting the text on image and converting it into speech. These method is designed by using Raspberry pi and portability is accomplished through a battery backup. Thus the user can handle the device anywhere and also to use at any time. previously stored faces are identified using camera and informed which can be implemented as a future technology. This technology assist millions of people in the world who have a significant loss of vision. In the basic goal is an embedded system that captures a picture , extracts only the region of interest (i.e. region of the image that contains text) and

change that text to speech. It is generated using a Raspberry Pi and a Raspberry Pi camera. The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and eliminates the background. Two tools are utilized convert the new image (which contains only the text) to speech. They are OCR (Optical Character Recognition) software and TTS (Text-to-Speech) engines. The audio output is listend via the raspberry pi's audio jack using speakers or earphones. In The aim of this syatem to develop a system that will voice control the home appliances and also provide security against intrusion when the home owner is not in home. This project is made up of Raspberry Pi 3 and Relay driver circuit.

2.1 Hardware Description
2.1.1 5MP Raspberry Pi 3 Model B Camera
The device consists of a Raspberry Pi 3B, speaker or earphones, Raspberry pi camera, power offer (230V AC) and a rectifier converts ac to pulsating dc. The voltage regulator regulates the voltage to a fixed level to power the Raspberry Pi. The camera should manually be pointed towards the text and an image is captured. This image is then processed by the Raspberry Pi and therefore the audio output is given through the speaker.

Fig .1- A 5 MP Raspberry Pi 3 Model B Camera

The raspberry pi has integrated peripheral devices like USB, ADC, Bluetooth, WiFi and SPI. Raspberry pi 3 B+ Linux operating system named Raspbian stretch. In this system the printed text is placed under camera view by the blind person to ensure the image of good quality and fewer distortions. Then an applicable blind assistive system, a localization algorithm might prefer higher recall by sacrificing some precision.

Hardware Specification
   5MP Raspberry Pi 3 Model B Camera

International Journal of Innovations in Engineering and Science,   www.ijies.net

1. Raspberry Pi 3 Model B
2. LCD Display
3. Capacitors
4. Transistors
5. Cables & Connectors
6. Diode
7. PCB

2.1.2 Raspberry Pi 3 Model B

Fig. 2 -   Raspberry Pi 3 Model B

The Raspberry Pi 3 Model B+ is the latest product in the Raspberry Pi 3 range, boasting a 64-bit quad core processor running at 1.4GHz, dual-band 2.4GHz and 5GHz wireless LAN, Bluetooth 4.2/BLE, faster Ethernet, and PoE capability via a separate PoE HAT The dual-band wireless LAN comes with modular compliance certification, allowing the board to be designed into end products with significantly reduced wireless LAN compliance testing, improving both cost and time to market.

Sofware Specification

Operating system:Raspbian (Stretch)
Language: Python3
Platform: Tesseract, OpenCV
 Library: OCR engine, TTS engine
The operating system under which the proposed project is executed is Raspbian which is derived from the Debian operating system. The algorithms square measure written victimization the python

language that may be a script language. The functions in algorithm are called from the OpenCV Library. OpenCV is an open source computer vision library, which is written under C and C++ and runs under Linux, Windows and Mac OS X. OpenCV was designed for process potency and with a powerful target period of time applications. OpenCV is written in optimized C and may profit of multi-core processors.

2.1.3 Optical character recognition (OCR)
Optical character recognition (OCR) refers to each

thetechnology and method of reading and changing typewritten, written or written characters into machine-encoded text or one thing that the pc will manipulate. It is a set of image recognition and is wide used as a type of knowledge entry with the input being some type of written document or knowledge record such as bank statements, sales invoices, passports, resumes and business cards. The document is either scanned or an image is taken associated it's up to the program to acknowledge the characters and provides an output within the type of a text document.

2.2. Working Principle
When capture button is clicked, this technique captures the document image placed before of the camera that is connected to Raspberry pi through USB. After selecting the process button the captured document image undergoes Optical Character Recognition(OCR) Technology. OCR technology permits the conversion of scanned pictures of printed text or symbols into text or info which will be understood or edited using a computer program. In our system for OCR technology, we have a tendency to ar mistreatment TESSERACT library. Using Text-to-speech library the info is going to be reborn to audio. The camera acts as the main vision in detection the image of the placed text document, the image is processed internally and separates label from the image by mistreatment open CV library and finally identifies the text that is pronounced through voice. Now the reborn text into audio output is listened either by connecting headsets via three.5mm audio jack or by connecting speakers via Bluetooth.
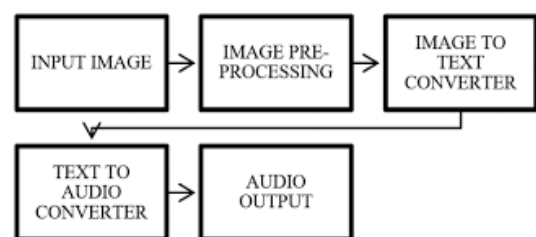
2.2.1 Methodology

Fig. 3-Bliock diagram of the process

Raspberry pi
Raspberry pi is a small computer which could be programmed. It works like a Linux based computer which could do all the normal

operations in PC. Raspberry Pi works in open source platform. Raspberry Pi 3 Model B 1GB is used in this system. This model comes with 40 GPIO pins and 4 USB ports which makes it more useful. Also it has camera interface and 3.5mm audio jack. USB ports available on this board are used to connect the camera with raspberry pi.

International Journal of Innovations in Engineering and Science,   www.ijies.net

Three GPIO pins are used, for capturing image, for mode control and for shutting down the system respectively. The board is operated in such a way that the code starts executing when it is powered ON. The audio output is available through the audio jack.

Image Capturing

The first step is the one in which the document is placed under the camera and the camera captures an image of the placed document. The quality of the image captured are going to be high thus on have quick and clear recognition because of the high-resolution camera.

Pre-Processing

The pre-processing stage consists of 3 steps: Skew Correction, Linearization, and Noise Removal. The captured image is checked for skewing. There are possibilities of the image getting skewed with either left or right orientation. Here the image is first brightened and binarized.

The perform for skew detection checks for AN angle of orientation between ±15 degrees and if detected then an easy image rotation is disbursed until the lines match with the true horizontal axis, that produces a skew corrected image. The noise introduced during capturing or due to the poor quality of the page has to be cleared before further process.

Image to Text Converter

The computer code values of the recognized characters area unit processed by Raspberry Pi board. Here every of the characters is matched with its corresponding guide and saved as normalized text transcription. This transcription is additional delivered to the audio output.

Text to speech

The scope of this module is initiated with the conclusion of the receding module of Character Recognition. The module performs the task of conversion of the transformed text to audible form. The Raspberry Pi has an on-board audio jack, the on-board audio is generated by a PWM output and is minimally filtered. A USB audio card will greatly improve the sound quality and volume. As the recognition method is completed, the character codes within the computer file square measure processed mistreatment Raspberry Pi device on that acknowledge a personality mistreatment Tesseract algorithmic rule and python programming, the audio output listens.

2.2.2 Design and Implementation

Setup Rapsbserry Pi:

For the raspberry pi to able to perform all the functions

properly we need to define all the input and output pins of devices connected to it and to which no. of pin of raspberry pi it is connected to

```python
pwmPin = 18 # Broadcom pin 18 (P1 pin 12)
ledPin = 23 # Broadcom pin 23 (P1 pin 16)
butPin = 17 # Broadcom pin 17 (P1 pin 11)

dc = 95 # duty cycle (0-100) for PWM pin

# Pin Setup:
GPIO.setmode(GPIO.BCM) # Broadcom pin-numbering scheme
GPIO.setup(ledPin, GPIO.OUT) # LED pin set as output
GPIO.setup(pwmPin, GPIO.OUT) # PWM pin set as output
pwm = GPIO.PWM(pwmPin, 50)  # Initialize PWM on pwmPin 100Hz frequency
GPIO.setup(butPin, GPIO.IN, pull_up_down=GPIO.PUD_UP) # Button pin set as input w/ pull-up

# Initial state for LEDs:
GPIO.output(ledPin, GPIO.LOW)
pwm.start(dc)

print("Here we go! Press CTRL+C to exit")
try:
    while 1:
        if GPIO.input(butPin): # button is released
            pwm.ChangeDutyCycle(dc)
            GPIO.output(ledPin, GPIO.LOW)
        else: # button is pressed:
            pwm.ChangeDutyCycle(100-dc)
            GPIO.output(ledPin, GPIO.HIGH)
            time.sleep(0.075)
            GPIO.output(ledPin, GPIO.LOW)
            time.sleep(0.075)
except KeyboardInterrupt: # If CTRL+C is pressed, exit cleanly:
    pwm.stop() # stop PWM
    GPIO.cleanup() # cleanup all GPIO
```

Conversion of image to text using OCR tool

Tesseract is an open source-OCR engine. It assumes that its input is a binary image with optional polygonal text region defined. The first step is a connected component analysis in which outline of the components is stored. By the inspection of the nesting of outlines, it is easy to detect inverse text and recognize it as early as black on white text. At this stage, outlines are gathered together, purely by nesting, into blobs. Blobs are organized into text lines, and the lines and regions are analyzed for fixed pitch or

International Journal of Innovations in Engineering and Science,   www.ijies.net

proportional text. Slope across the line is used to find text lines. Quality of recognized text is verified. If clarity is not enough the text is passed to associator.

```
DEFAULT_CHECK_COMMAND = "which"
WINDOWS_CHECK_COMMAND = "where"
TESSERACT_DATA_PATH_VAR =
'TESSDATA_PREFIX'

VALID_IMAGE_EXTENSIONS = [".jpg", ".jpeg", ".gif",
".png", ".tga", ".tif", ".bmp"]
import argparse
import logging
import os
import shutil
import subprocess
import sys
import tempfile

from constants import VALID_IMAGE_EXTENSIONS,
WINDOWS_CHECK_COMMAND,
DEFAULT_CHECK_COMMAND,
TESSERACT_DATA_PATH_VAR


def create_directory(path):
    """
    Create directory at given path if directory does not exist
    :param path:
    :return:
    """
    if not os.path.exists(path):
        os.makedirs(path)


def check_path(path):
    """
    Check if file path exists or not
    :param path:
```

Conversion of text to voice using E-Speak

Normal text to speech conversion is done using ESpeakwhich is a TTS system. The artificial production of human speech is known as speech synthesis. Speech computer or speech synthesizer is used for this purpose and can be implemented in software or hardware products. To create a completely "synthetic" voice output a synthesizer can be used to incorporate a model of the vocal tract and other human voice characteristics. A TTS (or"engine") is composed of two parts a front-end and a back-end. The front-end has two major tasks, the normalization and phonetic transcription of text. Normalization, preprocessing, or tokenization of text is the conversion of text containing symbols like abbreviations and numbers into equivalent written-out words. The front-end then assigns

2.3 Discussion on Results of Earlier Implemented Systems

2.3.1 Simulation Environment

The image to text and text to speech conversion is done by the OCR software installed in raspberry pi. The conversion which is done in OCR can be simulated in MATLAB. The conversion process in MATLAB includes the following processes. 1. Binary image conversion. 2. Complementation. 3. Segmentation and labeling. 4. Isolating the skeleton of character.

1. **Sample Image**

    The following image which is captured by the webcam contains the following word. This image is in the jpeg format which has to be converted into text.

2. **Binary Conversion**

    In this section sample image is converted into binary format. The image which was a 3D image initially is converted to 2D image .Binary 0 represents black color of the characters. Binary 1 represents white color of the characters

3. **Boundary Marking**

    The area of the text is bordered and the boundary for each character is isolated. The boundary for each character is programmed and it can vary from 0 to 255 bits of characters occupying memory in the database. Segmentation and labelling The isolated blocks of characters are segmented and are automatically labelled for identity. Image segmentation is the process of partitioning a digital image[8] into multiple segments (sets of pixels, also known as super pixels).
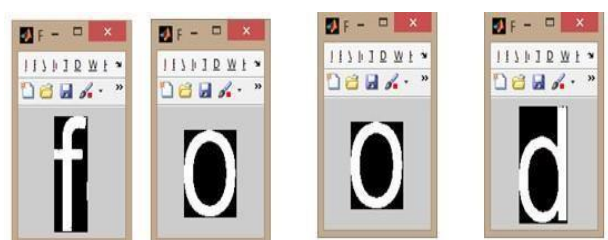

Fig. 4- Segmentation and labelling

International Journal of Innovations in Engineering and Science,   www.ijies.net

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics.

Connected-component labeling is used in vision to detect connected regions inry digital images, although color images and data with higher dimensionality can also be processed. When integrated into an image recognition system or human-computer interaction interface, connected component labeling can operate on a variety of information. Blob extraction is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, filtered, and tracked.

4. Forming Character Skeleton

Skeletonization is a process for reducing foreground regions in a binary image to a skeletal remnant that largely preserves the extent and connectivity of the original region while throwing away most of the original foreground pixels. To see how this works, imagine that the foreground regions in the input binary image are made of some uniform slow burning material.

Character Skeleton Light fires simultaneously at all points along the boundary of this region and watch the fire move into the interior. At points where the fire travelling from two different boundaries[9] meets itself, the fire will extinguish itself and the points at which this happens form the so called _quench line'
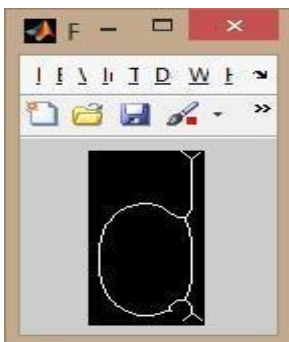

Fig. 5: Character Skeleton

2.3.2 Audio Output

The programming codes are run in MATLAB and corresponding output is generated. The output is in the form of audio. The audio is heard using headphone or speaker connected to the system. Each character of the word is spelled out first and then the entire word is read out.

5] Advantages And Disadvantages

- **Advantages:**
  1. As the main intension of the project is to help people who travels around the globe and to help the blind by assisting them in reading text.
  2. User friendly (easy to use).
  3. It is very compact to the user.
  4. High rate of translation:
  5. The speed of translation is much higher than the human translation it takes more than an hour for translating 10,000 words for human whereas, few seconds are enough for the device to translation
  6. Cost efficient

- **Disadvantages:**
  1. The main challenge of this project is portability.
  2. It is not provided with a battery backup hence face the issue in transportation.

III - CONCLUSION AND FUTURE WORK

The text to speech system is implemented with the assistance of Raspberry Pi board. The input image is processed efficiently by algorithm used and clear output is generated. This piece of kit is affordable and productive to blind people. The device is tested with different text and verified. This device is beneficial for the society and compact. With this application blindness problem can't be overcome completely, but it can help them a minimum of to some extent.

Visually impaired people have to face many issues while accessing the text which are not in braille script. To help them there are many technologies, one of them is text to speech converter, which converts text to speech hence they will hear the words. But usually voice which is employed are going to be inconvenient to concentrate and even the spellings won't be known to the user. This can be avoided by using character by character recognition but its time consuming. Hence the opposite technology which will be wont to help them is Braille displays. The servo motor Braille display over comes all the disadvantages of other sorts of displays used. Now days the worth of Braille displays are increasing, but we used servo motors which are available at lower costs. The remanufactured servo motors which has the driver circuits build in it thus it occupies more space. But the dimensions are often reduced by building own servo motors because the dimensions of

International Journal of Innovations in Engineering and Science,   www.ijies.net

the gears that required to rotate are small Matlab has been used to convert text in the image to editable text form in which programming is easy compared to other software's such as python, c# etc. Earlier using matlab we have created desktop application. This helps to obtain the output in the systems which don't have mat lab software. This can be improved by creating voice controlled desktop application.

REFERENCES

[1] *Ms.AthiraPanicker Smart Shopping assistant label reading system with voice output for blind using raspberry pi, Ms.Anupama Pandey, Ms.Vrunal Patil YTIET, University of Mumbai ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 5, Issue 10, Oct 2016 2553 www.ijarcet.org*

[2] *Agarwal,D.Kumarand A.Bhardwaj,"Ultrasonic stick for bling,"International journal of enegineering and computer science,ISSN:2319-7242 Vol.4,Issue 4,April 2015,pp.11375-11378.*

[3] *Ray Kurzweil (2008), "K-Reader mobile user guide", Knfb Reading Technology Inc. 2*

[4] *Ando. B. 'A smart multi sensor approach to assist blind people in specific urban navigation tasks', IEEE Trans. Neural Syst. Rehab. Eng., vol. 16, no. 6, pp. 592–594, Dec. 2008.*

[5] *Ando. B. and Baglio. S. 'A haptic solution to assist visually impaired in mobility tasks', IEEE Trans. Human–Mach.Syst., vol. 45, no.5, pp. 635–640, Oct. 2015.*

[6] *William A Ainsworth (Jun 1973), "A System for converting English text into Speech", IEEE Transactions on Audio and Electroacoustic, Volume 21, Issue 3,*

[7] *Xilin Chen, Jie Yang, Jing Zhang, Alex Waibel (Jan 2004), "Automatic detection and recognition of signs from natural sceneries transactions on image processing", Volume 13, Issue 1,*

[8] *J. Liang D. and DoermannH. (2005), „Camera-based analysis of text and documents: a survey," International Journal on Document Analysis and Recognition, Vol.7, No-6, pp. 83-200.*

[9] *Chucai Yi. YingLiTian.AriesArditi. (2014), „Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons", IEEE/ASME Transactions on Mechatronics, Vol. 3, No. 2, pp. 1-10.*

[10] *Deepa Jose V. and Sharan R. (2014), „A Novel Model for Speech to Text Conversion", International Refereed Journal of Engineering and Science (IRJES) Vol.3, Issue.1, pp. 39-41.*

[11] *Alías F. Sevillano X. Socoró J. C Gonzalvo X. (2008), „Towards high-quality next-generation text-to-speech synthesi", IEEE Trans. Audio, Speech, Language Process, Vol. 16, No. 7. pp. 1340-1354.*