Smart Glasses for Blind People

Dr. Sankit Kassa[1], Manisha Inchanalkar[2], Amisha Kamble[3], Mayuri Inchanalkar[4]
[1,2,3,4]*Department of Electronics Engineering, SNDT University*

Abstract **- Visual impairment is one of the biggest limitations for humanity, especially in this day and age when information is communicated a lot by text messages (electronic and paper based) rather than voice. The device we have proposed aims to help people with visual impairment. In this project, we developed a device that converts an image's text to speech. The basic framework is an embedded system that captures an image, extracts only the region of interest (i.e. region of the image that contains text) and converts that text to speech. It is implemented 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 removes the background. Two tools are used 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 heard through the raspberry pi's audio jack using speakers or earphones.**

Index Terms **- Raspberry Pi, OCR, TTS.**

I.INTRODUCTION

In our lives, there are many people who are suffering from different diseases or handicap. These people need some help to make their life easier and better. There are special schools and universities for people with special needs. There are different levels of needs and not all levels require special places and special schools. For instance, people with vision difficulties can study with normal students if they have an appropriate chance. Most blind people and people with vision difficulties did not study and that is because special schools for people with special needs not everywhere and most of them are private and expensive or they study at home acquiring basic knowledge from their parents. The main goal of "Smart Glasses" is to help blind people and people who have vision difficulties by introducing a new technology that makes them able to read the typed text. These glasses are provided with technology to scan any written text and convert it into audio text. The goal of "Smart Glasses" is helping those people in different life aspects. For example, these glasses effectively helpful in the education field.

Blind people and people with vision difficulties can be able to read, study and learn everything from any printed text images. "Smart Glasses" encourage blind people or people with vision difficulties to learn and succeed in many different fields. Visually impaired people report numerous difficulties with accessing printed text using existing technology, including problems with alignment, focus, accuracy, mobility and efficiency. These people need some help to make their life easier and better.

The main goal of "Smart Glasses" is to help blind people and project uses the methodology of a camera based "Smart Glasses" that can be used by people to read Text document. The design is motivated by preliminary studies with visually impaired people, and it is small-scale which enables a more manageable operation with little setup. these glasses effectively helpful in the education field. Blind people and people with vision difficulties can be able to read, study and learn everything from any printed text images. "Smart Glasses" encourage blind people or people with vision difficulties to learn and succeed in many different fields.

Figure 1: Design of "Smart Glasses"

II. WORKING PRINCIPLE

At the point when catch is clicked, this framework catches the archive picture put before the camera which is associated with ARM microcontroller through USB. After choosing the procedure catch the caught record picture experiences Optical Character. Recognition (OCR) Technology. OCR innovation permits the change of examined pictures of printed content or images into content or data that can be comprehended or altered utilizing a PC program. In our framework for OCR innovation we are utilizing TESSERACT library. Utilizing Text-to-discourse library the information will be changed over to sound. Camera goes about as primary vision in recognizing the picture of the set record, at that point picture is prepared inside and isolates mark from picture by utilizing open CV library lastly distinguishes the content which is articulated through voice. Presently the changed over content into sound yield is listened either by associating headsets by means of 3.5mm sound jack or by interfacing speakers by means of Bluetooth.

III. HARDWARE SPECIFICATIONS

A. Raspberry Pi

Raspberry Pi is an ease, Mastercard measured PC that connects to PC screen or TV and utilizations standard console and mouse.

Raspberry Pi is a credit card-sized computer. It needs to be connected with a keyboard, mouse, display, power supply, SD card and installed operating system. Raspberry Pi is a low-cost embedded system that can do a lot of significant tasks. It can be run as no-frills PC, a pocketable coding computer, a hub for homemade hardware and more. It includes GPOI (general purpose input/output) pins to control electronics components. It is also a great machine to attract children to learn more about how computers work and motivate them to improve their programming skills which help to create the next generation of developers.

In this project we are using raspberry Pi 4B, which is the latest version, having more powerful Quad core processor. It is available in 3 variations: 1GB, 2GB and 3GB based on RAM size. It provides 3 times better performance than the previous versions. In addition it provides 2 micro HDMI ports, Display Serial Interface (DSI) and Camera Serial Interface (CSI) and micro SD card slot for easy storage.

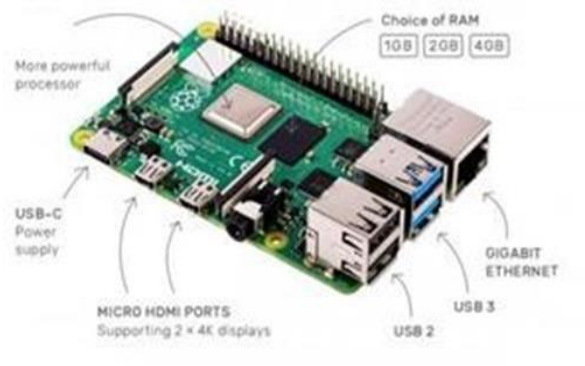

Figure 2: Raspberry Pi 4

B. Raspberry Pi Camera

The Raspberry Pi Camera Modules are official products from the Raspberry Pi Foundation. In this project, we are using Raspberry Pi rev1.3 camera module. The Raspberry Pi camera board plugs directly into the CSI (Camera Serial Interface) connector of Raspberry Pi. It features 5MP resolution with Omnivision sensor in a fixed focus module. It is connected with Raspberry Pi with a 15 pin ribbon cable. It is tiny in size and weighs about 3g. The purpose of the camera is to capture the image and store in Raspberry Pi which will be then converted into speech.

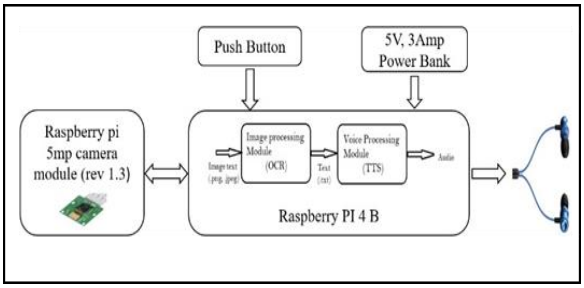

Figure 3: Block Diagram

IV. SOFTWARE SPECIFICATIONS

Working Framework:

- Operating System: Raspbian (Debian)
- Language: Python 2.7
- Platform: Tesseract, OpenCV (Linux-library)
- Library: OCR engine, TTS engine

A. OCR ENGINE

OCR is electronic identification and digital conversion of printed material or typed material; it can be used in handwritten materials and banners. Using OCR allows various devices to read or scan images

and convert those images into editable or text format. OCR typically includes three major procedures: scanning the image in OCR software, convert them into editable or readable 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. 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.

Using this technique, the computer tries to recognize the entire character and matches it to the matrix of characters stored in the software. As a result, this technique is also known as pattern matching or matrix matching. The drawback of this technique is that it relies on the input characters and the stored characters being of the same font and same scale.

B. Putty

PuTTY is Linux terminal emulator, serial console and network file transfer application that can connect between a Windows operating system and a remote device. PuTTY can be used to establish a secure command-line interface (CLI) between a Windows desktop and a remote Linux-base operating system. Connections can be secured using SSH (secure shell or secure socket shell). By logging in using the Raspberry Pi's username and password, the connected device can have access to various terminal commands which can be used to download and maintain packages, change configuration files, and create and edit directories and files — all from a remote location. Raspberry Pi are popular to use as small, standalone devices on a network. By using PuTTY, the need for these devices to have a keyboard, mouse and screen is reduced to only their setup phases. After an IP address has been obtained, PuTTY can be used from a remote computer to have full access via a terminal.

Requirements for using PuTTY:

1. The Raspberry Pi and the remote computer are connected to the same local network.
2. The Raspberry Pi has SSH enabled. Some Raspbian distributions have SSH enabled by default but can be configured using the Rasp-config tool.
3. The IP address of the Raspberry Pi is known.

C. VNC

Virtual Network Computing (VNC) is a graphical desktop- sharing system that uses the Remote Frame Buffer protocol to remotely control another computer. It transmits the keyboard and mouse events from one computer to another, relaying the graphical-screen updates back in the other direction, over a network. VNC Connect from Real VNC is included with Raspberry Pi OS. It consists of both VNC Server, which allows you to control your Raspberry Pi remotely, and VNC Viewer, which allows you to control desktop computers remotely from your Raspberry Pi should you want to.

D. Text to Speech synthesizer

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, for example, a number. Each symbol generated is assigned a phonetic. The back end then converts the phonetic into sound. In our project, we have used Festival TTS. Festival is the most widely used open source. TTS. It has a wide variety of voices and support English, Spanish and welsh language. We have used the English language

Raspbian OS follows the Advanced Linux Sound Architecture (ALSA) for managing audio devices. We have to install a few packages to test the sound device through ALSA. Using the apt-

Image acquisition:

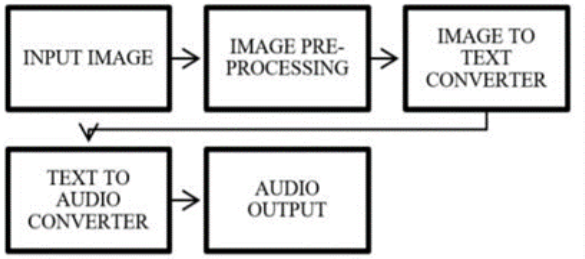

Figure 4

get utility, install the following packages (You will need admin privileges to install these packages) apt-get install alsa-utils apt-get install mpg321 apt-get install lame

Make sure that the Raspberry Pi is powered off while Connecting the USB sound card to one of the USB

ports of the Pi and power on the Pi. Once the Raspbian OS has booted, make sure that the audio hardware has been detected. With the sound device setup, it's time to test the application. Connect the speakers to the speaker jack on the sound card and play one of the test sound files provided by Raspbian OS.

V.METHODOLOGY

The working mainly consists of two methods:

a. Image Processing method and
b. Voice processing method

The image processing method includes the image capturing and image to text conversion. 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. The image processing method includes capturing of static image with the help of camera. The camera works as an eye for the raspberry-pi. The camera can be connected to the raspberry-pi with the help of Cable. We have used a raspberry pi camera to capture the image. After the successful connection the image is captured with the help of tesseract OCR software. We are using tesseract OCR which is raspberry pi compatible and can understand primarily English language. The teserract-ocr is library and is open source. The Tesseract-ocr is command line OCR which captures the image on the press of button. The image can be saved in .jpeg or .png format. With the help of tesseract OCR library of python, the captured image is converted to the text format in the raspberry pi and saved with the same name as an image. The converted text is provided to TTS system which converts the text to the voice format.

In this step, the inbuilt camera captures the images of the text. The quality of the image captured depends on the camera used. We are using the Raspberry Pi's camera which 5MP camera with a resolution of 2592x1944.

Image pre-processing:

This step consists of color to gray scale conversion, edge detection, noise removal, warping and cropping and thresholding. The image is converted to gray scale as many OpenCV functions require the input parameter as a gray scale image. Noise removal is done using bilateral filter. Canny edge detection is performed on the gray scale image for better detection of the contours. The warping and cropping of the image are performed according to the contours. This enables us to detect and extract only that region which contains text and removes the unwanted background. In the end, Thresholding is done so that the image looks like a scanned document. This is done to allow the OCR to efficiently convert the image to text.

Image to text conversion:

The above diagram (fig.3) shows the flow of Text-To-Speech. The first block is the image pre-processing modules and the OCR. It converts the pre-processed image, which is in .png form, to a .txt file. We are using the Tesseract OCR.

Text to speech conversion:

The second block is the voice processing module. It converts the .txt file to an audio output. Here, the text is converted to speech using a speech synthesizer called Festival TTS. The Raspberry Pi has an on-board audio jack, the on-board audio is generated by a PWM output.

VI.RESULT

The obtained output images after pre-processing are displayed below. Figure 6 shows the original image that was captured using the Pi Camera and Figure 13 displays the text obtained at the output of the OCR engine. Better results can be obtained if the camera used is a High-definition camera.

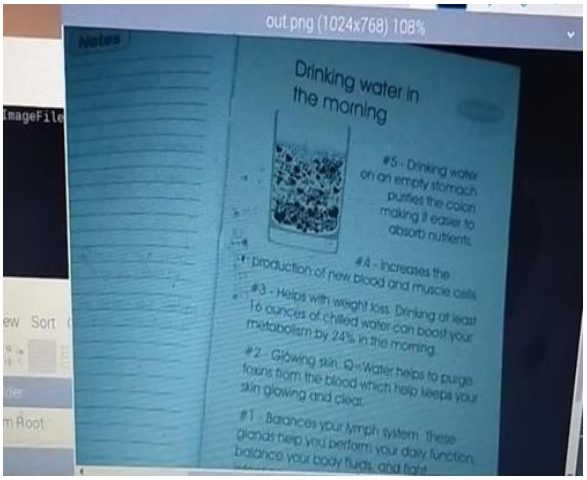

Figure 6 : Image by Raspberry Pi Cam Module

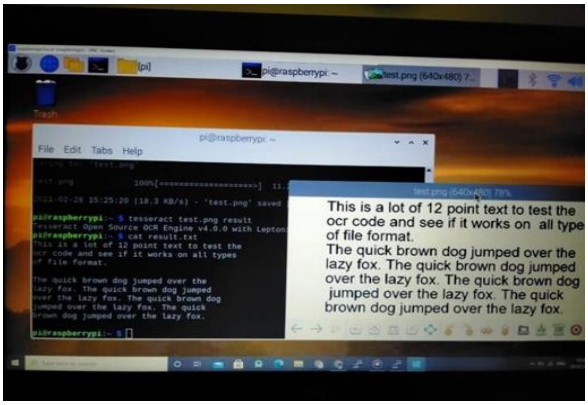

Figure 7: OCR Output

VII. CONCLUSION

In conclusion, these smart glasses are especially designed for visually impaired people. So that they can have more confident and independent life. This glass can help them move around safely as it can detect nearby object. It can also help them read text blocks so they can read any type of books, or anything written in their surroundings. Our prototype system is to read the printed text in the objects or products for assisting blind persons. The text characters are recognized using Optical Character Recognition, the text codes are transformed as speech for blind persons. Our future work will extend the text localization algorithm with furthermore features and we will address the human interface issues associated with text reading by the blind user. This portable device does not require internet connection and can be used independently by people.

REFERENCES

[1] D.Velmurugan, M.S.Sonam, S.Umamaheswari, S.Parthasarathy, K.R.Arun[2016]. A Smart Reader for Visually Impaired People Using Raspberry PI. International Journal of Engineering Science and Computing IJESC Volume 6 Issue No. 3.

[2] K Nirmala Kumari, Meghana Reddy J [2016]. Image Text to Speech Conversion Using OCR Technique in Raspberry Pi. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 5, Issue 5, May 2016.

[3] Alessandro Bissacco, Mark Cummins, Yuval Netzer, and Hartmut Neven. PhotoOCR: Reading Text in Uncontrolled Conditions. 12 2013.

[4] H. I. Koo and D. H. Kim. Scene Text Detection via Connected Component Clustering and Nontext Filtering, volume 22. 2013.

[5] Alexandre Trilla and Francesc Alías. (2013), "Sentence Based Sentiment Analysis for Expressive Text to- Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, Issue. 2. pp. 223-233.

[6] 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.

[7] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. Comput. Vision Pattern Recognit., 2004, vol. 2, pp. II-366–II-373. 2. S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Text extraction and document image segmentation using matched wavelets and MRF model," IEEE Trans Image Process., vol. 16, no. 8, pp. 2117– 2128, Aug. 2007.

[8] World Health Organization. (2009). 10 facts about blindness and visual impairment.

[9] Advance Data Reports from the National Health Interview Survey (2008).

[10] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011).