READING ASSISTANT FOR VISUALLY IMPAIRED PEOPLE

Abhijith Shaji[1], Abhishek Aravindan[2], Nisham Rafeeque[3], Naveen K K[4]

[1,2,3,4] Dept. of Computer science and engineering, SDIT, Karnataka, India.

Abstract - *A Majority of the visually impaired use Braille for reading documents and books which are difficult to make and less readily available. This gives rise to the need for the development of devices that could bring relief to the agonizing tasks that the visually impaired has to go through. Due to digitization of books there are many excellent attempts at building a robust document analysis system in industries, academia and research labs, but this is only for those who are able to see. This project aims to study the image recognition technology with speech synthesis and to develop a cost effective, user friendly image to speech conversion system with help of Raspberry Pi. The project has a small inbuilt camera that scans the text printed on a paper, converts it to audio format using a synthesized voice for reading out the scanned text quickly translating books, documents and other materials for daily living, especially away from home or office. Not only does this save time and energy, but also makes life better for the visually impaired as it increases their independency.*

Key Words: **Visually Impaired People, Raspberry pi, pycam, Tesseract.**

1. INTRODUCTION

A key fact by WHO, 285 million people are estimated to be visually impaired worldwide: 39 million are blind and 246 have low vision. With all the various problems faced by blind people, the problem of reading is our field of interest. We are advancing in technology and science, but in spite of such development the techniques used by the blind are old fashioned. Most of the reading materials available for the blind are in the form of Braille. A person has to learn using Braille just for reading, and if a person is unable to learn Braille then he will be unable to read. Another disadvantage is that an error in understanding will result in reading wrong data. The last disadvantage is that the books, documents etc. have to be converted into a form of raised dots for the blind to read. The books and papers available for the blind in Braille format are quite less in comparison to the vast pool of books which are printed daily. Hence a device to help the blind in reading is a necessity. This document provides details about using raspberry pi as the main unit which has an inbuilt camera that is used to scan any written document and uses Optical character recognition (OCR) to convert the image into a digital text. We then use a text to audio system that will enable us to convert the digital text into a synthesized voice. We are using a raspberry pi which is a credit card-sized single-board computer. It is a complete computer in itself with a working operating system. The operating system can differ as per the use of the device. In our document we have used Raspbian operating system within raspberry pi and have written the code in python language. The camera in the pi is used for scanning any written document (books, bank statements, menus etc.).The scanned image undergoes image processing for obtaining the region of interest and segmentation of the letters from the word. The segmented letters undergo OCR. The output is combined to obtain the individual words as it was present originally in the document. The words obtained are given to a text to speech convertor that allows us to obtain the voice converted output according to the written document.

2. LIRTERATURE SURVEY

2.1 Raspberry pi

The engineers with a pragmatic approach are the biggest boon to a society. The application of ideas, theories, and new innovations is what drives them. For years the work was done on Arduino boards but with the launch of the very cheap Raspberry Pi it all changed. Raspberry Pi's inception began in 2006 it was finally released on 19 February 2012 as two models: Model A and Model B. After the sale of 3 million units in May 2014, the latest Model B+ was announced in July 2014. It contains many minor improvements based on the user suggestions without any increase in price. Raspberry Pi board costs only $35 and does the work of a computer costing hundreds of dollars. Though its purpose is not to replace computers, laptops etc. but to work in supplement with them. Boot it up, and you have a got a fully functional powerhouse. Grab a four-gigabyte SD card and flash it with the free Linux-based operating system on the Raspberry Pi Foundation's website. Put the SD card into the slot, apply power, and you've got a 700 megahertz workstation with hardware accelerated 3-D graphics—something that would have been state-of-the-art in 2001 and set you back several thousand dollars. The Raspberry Pi offers another path: encouraging experimentation by lowering the cost of accidentally breaking when you're trying to learn.

Technical Specifications

The following are specifications for Model B

- Broadcom BCM2835 SoC processor with 700 MHz ARM1176JZF-S cores.
- 512MB RAM.
- Videocore 4 GPU supports up to 1920x1200 Resolution.

- SD card slot.
- 10/100Mbps Ethernet port.
- 2 x USB 2.0 ports.
- HDMI, audio/video jack.
- GPIO header containing 40 pins.
- Micro USB power port providing 2A current supply.
- DSI and CSI ports.
- Dimensions: 85.6x56mm.

2.2 Camera Module

The Raspberry Pi camera module size is 25mm square, 5MP sensor much smaller than the Raspberry Pi computer, to which it connects by a flat flex cable (FFC, 1mm pitch, 15 conductor, type B.

Fig. 1 Raspberry Pi Camera Module

The Raspberry Pi camera module offers a unique new capability for optical instrumentation with critical capabilities as follows:

- 1080p video recording to SD flash memory cards.
- Simultaneous output of 1080p live video via HDMI, while recording.
- Sensor type: OmniVision OV5647 Colour CMOS QSXGA (5-megapixel)
- Sensor size: 3.67 x 2.74 mm
- Pixel Count: 2592 x 1944
- Pixel Size: 1.4 x 1.4 um
- Lens: f=3.6 mm, f/2.9
- Angle of View: 54 x 41 degrees
- Field of View: 2.0 x 1.33 m at 2 m
- Full-frame SLR lens equivalent: 35 mm
- Fixed Focus: 1 m to infinity
- Removable lens.
- Adapters for M12, C-mount, Canon EF, and Nikon F mount lens interchange.
- In-camera image mirroring.
- Higher-resolution still-image capture (2592 x 1944 native resolution, 5 megapixels)
- Low cost ($25 for the add-on camera module).

Open-source, modifiable software for certain aspects of camera control, image capture, and image processing Underlying Linux-based microcontroller with capabilities for HTTP service of images, Ethernet and WiFi connectivity, etc.

2.3 Image Processing

Books and papers have letters. Our aim is to extract these letters and convert them into digital form and then recite it accordingly. Image processing is used to obtain the letters. Image processing is basically a set of functions that is used upon an image format to deduce some information from it. The input is an image while the output can be an image or set of parameters obtained from the image. Once the image is being loaded, we can convert it into gray scale image. The image which we get is now in the form of pixels within a specific range. This range is used to determine the letters. In gray scale, the image has either white or black content; the white will mostly be the spacing between words or blank space.

2.4 Feature Extraction

In this stage we gather the essential features of the image called feature maps. One such method is to detect the edges in the image, as they will contain the required text. For this we can use various axes detecting techniques like: Sobel, Kirsch, Canny, Prewitt etc. The most accurate in finding the four directional axes: horizontal, vertical, right diagonal and left diagonal is the Kirsch detector. This technique uses the eight point neighborhood of each pixel. This way the feature maps along the respective directions are obtained. Then is compression of the Kirsch patterns. This is done to ensure that the neural network will learn without a large number of training samples.

2.5 Optical Character Recognition

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine encoded text. It is widely used as a form of data entry from some sort of original paper data source, whether documents, sales receipts, mail, or any number of printed records. It is crucial to the computerization of printed texts so that they can be electronically searched, stored more compactly, displayed on-line and used in machine processes such as machine translation, text-to-speech and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

Optical character recognition (OCR) systems allow desired manipulation of the scanned text as the output is coded with ASCII or some other character code from the paper based input text. For a specific language based on some alphabet, OCR techniques are either aimed at printed text or handwritten text.

History of OCR

1929 – Digit recognition machine

1953 – Alphanumeric recognition machine

1965 – US Mail sorting

1965 – British banking system

1976 – Kurzweil reading machine

1985 – Hardware-assisted PC software

1988 – Software-only PC software

1994-2000 – Industry consolidation

2.6 Tesseract

Developed on HP-UX at HP between 1985 and 1994 to run in a desktop scanner. It came neck and neck with Caere and XIS in the 1995 UNLV test. It never was used in an HP product. It was open sourced in 2005 and is highly portable. Tesseract is a free software optical character recognition engine for various operating systems. Tesseract is considered as one of the most accurate free software OCR engines currently available. It is available for Linux, Windows and Mac OS, however, due to limited resources only Windows and Ubuntu are rigorously tested by developers. An image with the text is given as input to the Tesseract engine that is command based tool. Then it is processed by Tesseract command. Tesseract command takes two arguments: First argument is image file name that contains text and second argument is output text file in which, extracted text is stored. The output file extension is given as .txt by Tesseract, so no need to specify the file extension while specifying the output file name as a second argument in Tesseract command. After processing is completed, the content of the output is present in .txt file. In simple images with or without color (gray scale), Tesseract provides results with 100% accuracy. But in the case of some complex images Tesseract provides better accuracy results if the images are in the gray scale mode as compared to color images. Although Tesseract is command-based tool but as it is open source and it is available in the form of Dynamic Link Library, it can be easily made available in graphics mode.

2.7 Text To Speech

A text to speech (TTS) synthesizer is a computer based system that can read text aloud automatically, regardless of whether the text is introduced by a computer input stream or a scanned input submitted to an Optical character recognition (OCR) engine. A speech synthesizer can be implemented by both hardware and software. Speech is often based on concatenation of natural speech i.e. units that are taken from natural speech put together to form a word or sentence.

Rhythm is an important factor that makes the synthesized speech of a TTS system more natural and the prosodic structure provides important information for the prosody generation model to produce effects in synthesized speech. Many TTS systems are developed based on the principle, corpus-based speech synthesis. It is very popular for its high quality and natural speech output.

Bell Labs Developed VOCODER, a clearly intelligible keyboard-operated electronic speech analyzer and synthesizer. In 1939, Homer Dudely developed VODER which was an improvement over VOCODER. The Pattern Playback was built by Dr. Franklin S. Cooper and his colleagues at Haskins Laboratories. First Electronic based TTS system was designed in 1968. Concatenation Technique was developed by 1970"s. Many computer operating systems have included speech synthesizers since the early 1980s. From 1990s, there was a progress in Unit Selection and Diphone Synthesis.

3. SYSTEM ARCHITECTURE

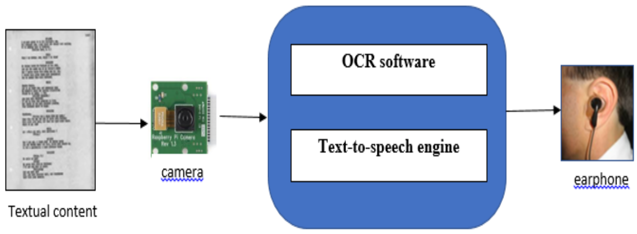

Figure3: System Architecture

The system architecture shows the interconnection of components that are included in the system. The base station here is Raspberry Pi. The camera is connected to the Raspberry pi through USB0.The earphone is connected to the Raspberry pi through USB1.Switch is connected on the device which could be controlled by the person who is using the device. Camera is connected through CSI camera connector of the Raspberry pi. The external SD card is inserted to Micro Card Connector.

The power supply is given through the power bank provided on the device. The OCR is used to get the text from the content to be read for the person. The text-to-speech synthesizer will convert the text to speech. An earphone will be used to hear the speech which was early converted by the text-to-speech synthesizer. The correct text that was captured by the camera will be heard by the person.

3.1 Raspberry Pi 3 Model B

The Raspberry Pi is a series of small single-board computers. The Raspberry Pi 3 is the third generation Raspberry Pi. It is recommended for use in schools, or for any general use, which is more useful for embedded projects, and projects which require very low power.

Here are the various components on the Raspberry Pi board:

ARM CPU/GPU – This is a boardcom BCM2835 System on a chip(SoC) that's made up of an ARM central processing unit(CPU) and a Video core 4 graphics processing unit(GPU). The CPU handles all the computations that make a computer work (taking input, doing calculations and producing output), and the GPU handles graphics output. GPIO – These are exposed general-purpose input/output connection points that will allow the real hardware hobbyists the opportunity to tinker. RCA – An RCA jack allows connection of analog TV's and other similar output devices. Audio out – This is a standard 3.55-millimeter jack for connection of audio output devices such as headphones or speakers.

There is no audio in. LED's – Light-emitting diodes, for all of your indicator light needs. USB – This is a common connection port for peripheral devices of all types (including your mouse and keyboard). Model A has one, and Model B has two. You can use a USB hub to expand the number of ports or plug your mouse into your keyboard if it has its own USB port. HDMI – This connector allows you to hook up a high -definition television or other compatible device using an HDMI cable. Power – This is a 5v Micro USB power connector into which you can plug your compatible power supply.

SD card slot – This is a full-sized SD card slot. An SD card with an operating system (OS) installed is required for booting the device. They are available for purchase from the manufacturers, but you can also download an OS and save it to the card yourself if you have a Linux machine and the wherewithal. Ethernet – This connector allows for wired network access and is only available on the Model B.

3.2 Pi Camera

The Raspberry Pi camera module size is 25mm square, 5MP sensor much smaller than the Raspberry Pi computer, to which it connects by a flat flex cable (FFC, 1mm pitch, 15 conductor, type B.

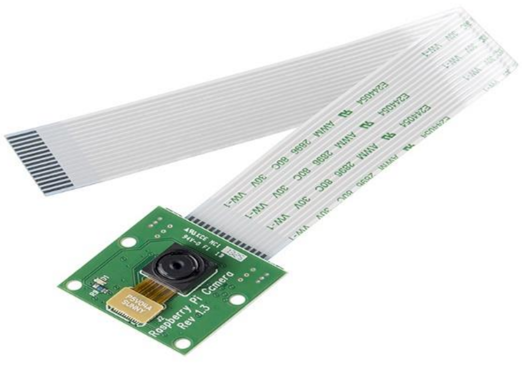

Figure3.2: Raspberry pi camera module

Books and papers have letters. Our aim is to extract these letters and convert them into digital form and then recite it accordingly. Image is used to obtain the letters. Image processing is basically a set of functions that is used upon an image format to deduce some information from it. The input is an image while the output can be an image or set of parameters obtained from the image. Once the image is being loaded, we can convert it into gray scale image. The image which we get is now in the form of pixels within a specific range. This range is used to determine the letters. In gray scale, the image has either white or black content; the white will mostly be the spacing between words or blank space.

3.3 Button

The switch is used to control the images taken which the reader has to read. When the person need to start the reading, the switch must be pressed, and the text get captured by the pi camera and the text will be converted using the text-to-speech synthesizer and whenever the person needs to stop the reading he must press the button again.

The switch is normally less cost and it is easily available. The switch is connected on device for easy usage.

3.4 Earphone

The earphone is used to hear the text which has been converted by the text-to-speech converter. Since it can easily be connected to the Raspberry Pi, usage of earphone is made very popular. Through the earphone the converted data can be heard very clearly by the visually impaired person using this device.

4. IMPLEMENTATION

4.1 Booting of Pi

The Operating System that is used is Raspbian. It's a derivative of the Linux OS. Specific applications of this software include added multimedia functionalities (essential libraries and packages to our project). It offers a Graphic User Interface which is intuitive to use along with the regular Linux command line interface. To install this: The image files of Raspbian OS are downloaded from the Raspberry Pi forum. This image file is extracted and is subsequently burned to the SD card using the software SDFormatter. The SD card is inserted to the Pi and it is given power supply. Default configuration settings are set.

4.2 Updating and Upgrading

Since the Raspberry Pi model in use is B model (the latest being B+), it required updating. Using apt commands we synchronizes the available software packages and the versions available and downloaded and installed the newer versions.

4.3 Reading Images

We first captured images using RaspiCam and since Standard Image Viewers do not offer image editing services, hence the open source software suite ImageMagik was installed for displaying converting and editing images. It offers support to the Tesseract software that we use for optical Character Recognition (OCR).

4.4 Optical Character Recognition

1. This is a major step of the project and is subdivided into following steps:

2. Installing the open source Tesseract-OCR engine

3. Since Tesseract on its own is a quirky command line tool and a bare back bone structure. It requires layout detection in the image. To achieve this, it is interfaced with ImageMagik.

4. The functionalities of ImageMagik's features are exploited and put to great use. The default output format is changed to churn out pictures which are of the following format: GrayScaleimage, generalized pixel distortion is corrected, text x-height is changed to 20 pixels.

5. Background steps that run when tesseract is called are: Image is opened in ImageMagik and gets converted to a gray scale image first to improve the contrast ratio, is then converted to a black and white image which is read and the identified characters (alphabets and digits in our case) are stored into a word matrix which is later stored as a notepad file containing the text present in the image.

6. The text that is stored in the notepad is in the same order of words and lines as it appears in the image file.

4.5 Text To Speech Converter

To make the Pi read out text, the open source software eSpeak was installed, it is useful for speech synthesizing or as it is commonly referred: text-to-phoneme translation. The eSpeak synthesizer creates voiced speech sounds such as vowels and sonorant constants by adding together sine waves to make the format peaks. It supports Speech Synthesis Markup Language (SSML).

To further improve eSpeak's functionality Festival multi language speech synthesis system was installed. It extends the eSpeak's already existing text to speech system with various APIs (Application programming Interface) and also boosts the languages supported to include voice packages for Hindi and Marathi. It also improves the words database already existing in eSpeak for the English language.

4.6 Combining Everything

A python script was written which when executed first instructs the Raspberry Pi camera to capture an image, then calls the tesseract OCR to process it, which in turn saves the file with a desired filename, After which the Festival speech package reads the saved text file.

4.7 Setup

We have worked on an unprecedented design to create a portable device for assisting the visually impaired. We glued the setup on a glass which can be easily worn by blind people.

The setup is looks like this.

Figure4.7: Overall Setup

5. CONCLUSION

The system enables the visually impaired to not feel at a disadvantage when it comes to reading text not written in braille. The image pre-processing part allows for the extraction of the required text region from the complex background and to give a good quality input to the OCR. The text, which is the output of the OCR is sent to the TTS engine which produces the speech output. To allow for portability of the device, a battery may be used to power up the system.

6. FUTURE ENHANCEMENTS

The future work can be developing devices that perform object detection and extracting text from videos instead of static images. It will be also concentrated on developing an efficient product that can convert the text in the image to speech with high accuracy.

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]. Silvio Ferreira, C´eline Thillou, Bernard Gosselin. From Picture to Speech: An Innovative Application for Embedded Environment. Faculté Polytechnique de Mons, Laboratoire de Théorie des Circuits et Traitement du Signal Bˆatiment Multitel - Initialis, 1, avenue Copernic, 7000, Mons, Belgium.

[4]. Nagaraja L, Nagarjun R S, Nishanth M Anand, Nithin D, Veena S Murthy [2015]. Vision based Text Recognition using Raspberry Pi. International Journal of Computer Applications (0975 – 8887) National Conference on Power Systems & Industrial Automation.

[5]. Poonam S. Shetake, S. A. Patil, P. M. Jadhav [2014] Review of text to speech conversion methods.

[6]. International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, Aug.-2014.

[7]. S. Venkateswarlu, D. B. K. Kamesh, J. K. R. Sastry, Radhika Rani [2016] Text to Speech Conversion. Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/102967, October 2016.

[8]. World Health Organization. 10 facts about blindness and visual impairment. 2015. Available from: http://www.who. int/features/factfiles/blindness/blindness_facts/en/

[9]. http://elinux.org/RPi_Text_to_Speech_(Speech_Synthesis)

[10]. https://en.wikipedia.org/wiki/Visual_impairment

[11]. https://en.wikipedia.org/wiki/Braille

[12].https://www.classycyborgs.org/braille-literacy-statistics-india/

[13]. www.raspberrypi.org

[14].http://www.zdnet.com/article/raspberry-pi-11-reasons-why-its-the-perfect-small-server/