SMART READER FOR BLIND PEOPLE

Dr. A. Ravi¹, Sk. Khasimbee², T. Asha³, T. Naga Sai Joshna⁴, P. Gnana Jyothirmai⁵

¹Professor and HOD, ²³⁴⁵UG Students, Department of Electronics and Communication Engineering, Potti Sriramulu Chalavadi Mallikarjunarao College of Engineering and Technology, Vijayawada, Andhra Pradesh, India.

ABSTRACT

Most of the people are estimated to be blind and visually impaired worldwide. There are some challenges faced by the blind people while reading. Previously, the solution for blind and visually Impaired people for reading is Braille. But there is a general lack of training in the use of Braille instruction and also lack of Braille machine and their materials. To overcome these, the smart reader for blind people using Raspberry pi method is useful. In this project - A smart reader for Blind presented the integration of a complete text read out system with page turning mechanism and dictionary query feature. Finally, the text is read out through the speaker or headphone. The system maintains the synchronization with the page turning mechanism and the dictionary feature proving interactive session got the user.

Keywords - Visually impaired, Braille, Raspberry Pi, Webcam, Speaker, Page turning mechanism.

[1] INTRODUCTION

A lot of information is available on the web and in books. Every single person can use this information in some form or the other to enhance their knowledge and to apply the same elsewhere. People born with sight are very fortunate enough to absorb this information by ourselves. However, there are people who are born without sight or lost their sight during their course of life. It is beyond the bounds of possibility for these people to draw in information from the media, especially books. As the technology has advanced, it has become fairly easy for such type of people to grasp in information from the web by using e-readers or any type of computers which have the capability to read out the text and it make the person understand without the ability of sight.

Nevertheless, it is still out of reach for people with such disabilities to read information present in printed books. Although this problem has been solved, notably by Louis Braille - creator of the famous Braille script, this does have some limitations. One of which includes different and expensive printers to print in that particular script. These expensive printers will definitely result in producing expensive books. Moreover, the library for such type of books are incredibly small. If the person wants to read a book which isn’t present in the library for blind, they’ll have to be dependent on other people who can read with ease. And more often than not, this cannot be the case always.

Therefore, in order for the person to read any printed book with ease we have developed a system which is cost efficient and easy to use.

This system is developed in such a way that, all that the user has to do is place a book in the system and the system does all the remaining tasks, like reading out the text, turning the page after the text in the present page has been done reading. It also helps in making the user
understand the meaning of any word which he or she may not have understood completely with the help of dictionary where the user interrupts the system and asks the meaning of the word. The user also has the ability to pause or end the reading according to his or her wish. This is done by the press of the designated buttons.

[2] RELATED WORK

Nowadays, there are some technical systems to help the blind persons. To introduce these systems, the first one which comes in our mind is ‘Braille’. In this system, it uses 6 raised dots in 2X3 fashion per cell to represent a symbol. A character of any language, a punctuation, or any other indicator meaning is represented by this symbol. With the help of finger, the dots are sensed and from this pattern information is gained. Numbering of right column is from 1-3 and the rest is from 4-6. There will be 64 (26) possible different kinds of patterns from per cell. But different language is not represented from our native ones. This is the representation or encoding from our known native language such as English, Spanish, Arabic, Bengali and so on. A letter of any language including punctuation, symbols and so on are represented by each dot pattern. Multi-cell is used if one cell is not sufficient for representing a symbol of any language. To make multi-cell pattern recognizable we have to use another special cell to allow a multi-cell pattern. Blind and visual impaired uses different kinds of Braille. Only one letter, number, punctuation sign, or special Braille composition sign per cell representation comes under grade-1 type of Braille. It is a one-to-one conversation. Grade-1 type braille it doesn’t permits abbreviations or words but grade-2 braille is the advanced version that can permit abbreviations and words and thus saves spaces in printing. An unstandardized shorthand format is used in grade-3 type braille which is not used in publications [1,2].

Another technology is **Screen Reader** that is used to help blind and also visually impaired people. It helps to read the text aloud that is displayed on the monitor of an image that is captured through image sensor which is connected to mobiles, computers, tablets or other devices. A text to speech synthesizer helps to produce speech of corresponding text, it gets information from the screen and the speech synthesizer recognizes. A text to speech engine software is embedded with the screen reader in this technology. [3,4]

Another technology which helps BVI is **Finger Reader**. The user wears this device to his finger. The plain printed text is accessed by BVI through finger reader. The user who put this device to their finger can be able to scan the text and get corresponding audio. It also has haptic sense of the layout. These senses alert the reader if he moves away from the line what he is scanning, it helps user to maintain scanning in straight line. [5]

Mostly used device which helps BVI is **Braille Notetaker** that replaces standard QWERTY keyboard to give input to computer. A text to speech synthesizer is also used and sometimes Braille displays used for output to BVI from the screen. [6]

**Braille Printer** is a hardware that can print hardcopy of Braille, for that it uses braille translation software program that converts electronic text from computer device or tablet or mobile to braille. Heavy weight papers are used to print the braille on both sides, but it is slower and more expensive process.[7]

**Screen Magnifier** is a graphical output of a computer which represents the text on the screen in enlarged form. Thus, in result it improves the visibility which helps people who have bad visibility. [8]

An **E-Book Reader or Audio Book Reader** which is available for pc or mobile, that can read the text louder from their screen to BVI, which helps to read the e-book that appeared
on the screen with a touch screen gesture and it performs this job by using text to speech function. [9]

Blind Reader can be able to read the electronic documents. Through this a reader can clearly understand and go through the materials effectively by using sense of touch. Blind reader is an application. It is developed for android devices and it is designed and developed in a way. So that a blind people can easily access the available options by the application. The person has to move his fingertip horizontally from left to right on the android mobile screen and the words that are touched are read out through speaker which is inbuilt with the mobile device. In this application the document is considered as the collection of words so that the system can understand the touched word accurately. For visually impaired people the total information on the screen is not clear will not be visible by them.[10]

[3] PROPOSED METHODOLOGY

A SMART READER is a method in which When the book is placed in the system, the text is read aloud with the help of implementation of python programs which help in capturing the image of the page with the help of a camera and converting the image into text (the pictures in the page are ignored automatically) and this text is then converted into an audio file and that is then played through a speaker or a headphone. The system provides various options for the user like PAUSE, PLAY, RESUME, STOP, DICTIONARY, MECHANISM etc. According to user wish the action is provided. So, the system is more interactive than being autonomous for itself.

If the user wishes to pause and resume the system, he or she could do so with a press of a button. If the user does not understand a meaning of a particular word, he or she can pause the system and press another button. When the button is pressed the system listens to the word the user says into the microphone. It then converts this speech into text and searches the web for the meaning of the given word. The meaning is pulled from the web and then places this string in a text document. This text is then converted into speech and read aloud to the user either on a speaker or on headphones, whichever the user prefers.

To go for the next page user opts for Mechanism, the mechanism is used turn the page successfully. Hence, the system gives the user a feel of autonomous page reading and interactive session with Dictionary query and page turning mechanism.

[4] BLOCK DIAGRAM

The Block diagram of proposed method is shown in below figure. The framework is Raspberry Pi. The block diagram mainly consists of three units.

They are,

1. Input unit
2. Processing unit
3. Output unit
1. Input Unit:

The Input unit of system consists of **USB camera, Microphone, Interrupt Buttons/NUM Keys.**

**USB camera** is the main input device that feeds the processing unit with appropriate quality image. It is USB plug-n-play device that can be programmed through python OpenCV library. It supports USB 2.0 protocol with mini USB port. It’s usually available in market as 2MP, 5MP, 8MP, 13MP, 25MP resolution camera. The more the MP, better the quality of an image, and more the accuracy of the converted text, and ultimately the speech. The camera is placed at certain height right above the page with the help of a host stand with a clip to clutch the camera. The stand is calibrated such that the camera captures crisp and clear image.

**Microphone** is an audio input device which converts sound signals into electrical signals. It does this transducing action with different technologies embedded in it viz Condenser, Dynamic, Piezoelectric, Fibre-optic, Laser, MEMS etc. Even it can be wired or wireless. For this design we make use of analog 3.5mm jack supported microphone of condenser type and wired connectivity.

**Interrupt button/Push Button** is a mechanical switch which closes the switch when it is pressed and opens it on release. This action of opening and closing is the logic for reading input. Push buttons are most commonly used as interrupt buttons in Robotics, Embedded System applications. Interfacing push button for input pin of the raspberry pi. These NUM Keys/Buttons are required for the events/Instructions i.e., Play, Pause, Resume, Stop.

2. Processing unit:

The **Processing Unit** is the heart and brain for any tasks to be accomplished. Our design is best suited with **Raspberry Pi 3 Model B board** as processing platform, it acts as both processor and controller. It supports Debian based OS and hence give the feel of portable pocket computer. Also, it provides programmable GPIOs and peripherals like SPI, I2C, UART, USB, SD card slot etc. and hence act as a microcontroller.
Raspberry Pi 3 Model B is a powerful credit-card sized single board computer that can be used for many applications and projects. It brings more powerful processing and 10x faster computing than the first-generation Raspberry Pi. Additionally, it supports wireless LAN’s connectivity making it the ideal solution for powerful connected designs.

Pi 3 is connected to a display monitor supporting HDMI or VGA port. If the monitor does not support HDMI-input port then it is resolved using HDMI to VGA converter.

Pi is powered up by plugging in the micro USB cable, then it automatically starts to boot. Once all BIOS start-up actions are done, the pi is ready to display its GUI monitor. Also, we can access its command window through which all the applications/modules can be downloaded and installed. The shell commands help a lot in easy install of all the python modules, packages etc. We can install all the software modules, python modules and image processing engines etc. in the Raspberry Pi.

3. Output unit:

The Output unit consists of *Page turning mechanism, Speaker/Headphone.*

*Page turning mechanism* is basically an Electro-mechanical system that turns the page of the book when user requests. It mainly consists of Roller wheel, Lifter arm, Turner arm. With the help of these the page turning mechanism is done successfully.

*Speaker/Headphone* is an output device. It converts electrical signals into sound signals. Speaker is used for listening to the speech output.

[5] SYSTEM OVERVIEW

According to the proposed method, the smart reader for blind people is used to read any text document by hearing. The system architecture/system flow is shown in below.

**SYSTEM FLOW**

- Capture Image of Text document
- Image to Text conversion
- Text to Speech Conversion
- Dictionary queries and Instructions
- Page Turning Mechanism

In Smart Reader, initially it captures the image of text document. *OpenCV* is used for capturing the image of the books page and apply its in-built functions for pre-processing like de-skewing, noise removing, binarizing etc. so that image will be in best quality for converting it to text by OCR module. Throughout this we will be running around the word Image Processing. So, it’s better to define it now.
Image processing is the method of applying some effects to the digital image like cropping, zooming, gray scaling, thresholding etc. to achieve some reformation in the image as per the requirement. Image processing done on digital image stored on computer is called Digital Image Processing (DIP). It has lot of scope in the present trending concepts like pattern recognition, face recognition, scene recognition in AI and Machine Learning domains. In the design, the DIP techniques to produce the captured image in suitable quality that tesseract-ocr requires. So that, it can reproduce good and accurate text. This method of processing the image before it is fed to image to text conversion by OCR engine is called pre-processing. The in-built functions for pre-processing are Binarization, De-Skewing, Image smoothing.

**Binarization** is a process of converting multi-coloured image into monochrome image, i.e. It converts an RGB image into a black and white image. This is the first step in the pipeline of pre-processing. Because making the image into monochrome would reduce the size of image, remove the noise that might be present in the RGB space, and makes the way easier for OCR engine to recognize the characters faster and lead to accurate results. In Binarization a pixel image is converted into a binary image, by writing on to all the pixels the uniform value. This is done by converting to gray scale first and then applying thresholding. Sophisticated algorithms are there in OpenCV that perform binarization.

**De-Skewing** is a process of straightening the image which is captured crookedly. The image has slants and skews. After removing the skewness, the image is properly aligned both horizontally and vertically. This process of removal is known as De-skewing. OpenCV has inbuilt algorithms and functions which can help for achieving good de-skewing results.

**Image smoothing** process is done to remove unwanted noise present in an image. It does image smoothing by applying special filters (2D convolution, linear filters, Gaussian filter etc.). This process is also known as image blurring. Selection of appropriate filter, with specific parameters will lead to appropriate smoothened image. The image smoothing is achieved by successive thresholding and Gaussian blurring an image. Having applied pre-processing techniques for a raw image, to get suitable image.

Now image processing module is ready, so there should be a module that converts **image to text** using OCR technology. There are many APIs available that are compatible with python which does this conversion from image to text. In this design, tesseract-ocr is used to extract text from the image, as it is more accurate, robust, easy to program. Thus, using this API we have got the converted text document from the image file. For more accuracy, the Google Cloud Vision API is used.

In **text to audio conversion**, the text is read out loud with the help of an API called Google Text to Speech. This API takes any text form over 100 languages and sends it over to the servers to decode it. Google Text to Speech is an API created by Google for reading out text on the screen.

**Dictionary query** is a feature which is used by the user at his or her discretion using PyDictionary. As the name suggests this feature provides a meaning for a word when the user seeks for it. The user starts of by pressing a dedicated button for dictionary in the system. The system pauses whatever process is going on and asks the user to say any word to which the user has not understood the meaning of. The user is given a microphone into which he says the desired word. Once the system gets the input as an audio file, this audio file will be converted to text. This process is done by a library in python language called **speech recognition**. This text is again sent over to Web Note which is an online API for dictionary. The meaning is later downloaded to the raspberry pi and again stored in a text document. We then use the same Google text to speech API to convert this text into speech. The meaning for the word in which the user wants to know is to be hear audibly.
Page Turning Mechanism  Which is basically an Electro-Mechanical System that turns the page of the book when user requests. The page turning Mechanism is the paramount to the Education and empowerment of people with a disability by allowing them to independently perform this task using a suitable actuation device. It mainly consists of Roller wheel, Lifter arm, Turner arm.

Roller can be any wheel which can lift a page with a suitable friction. Lifter arm is a mechanical arm used to lift the roller wheel to the correct height. It is a scale shaped steel or iron rod of good strength fixed to a servo motor. Turner arm is also a mechanical arm which is placed on a plane parallel to the page, it helps in turning the lifted page and makes the page turning mechanism successful.

[6] RESULT

Image is successfully captured by the system; it saves the image of first page and the captured image is shown below.

Fig (a): Captured Image

Text is converted by the system, even though there exist some errors, Google Cloud Vision has perfectly converted the image to text.
"There once was a little boy who had a bad temper. His father gave him a bag of nails and told him that every time he lost his temper, he must hammer a nail into the back of the fence. The first day, the boy had driven fifty nails into the fence. Over the next few weeks, as he learned to control his anger, the number of nails hammered daily gradually dwindled down. He discovered it was easier to hold his temper than to drive those nails into the fence. Finally, the day came when the boy didn’t lose his temper at all. He told his father about it and the father suggested that the boy now pull out one nail for each day that he was able to hold his temper. The days passed and the boy was finally able to tell his father that all the nails were gone. The father took his son by the hand and led him to the fence. He said, "You have done well, my son, but look at the holes in the fence. The fence will never be the same. When you say things in anger, they leave a scar just like this one. You can put a knife in a man and draw it out. It won’t matter how many times you say I’m sorry. The wound is still there."

In the next step, google TTS engine successfully converted text into speech. The speech cannot be displayed so not attached here.

The next stage is the Interactive Dictionary query session. The results for dictionary query being made are shown in below.
User ask the meaning of “Hammer” and “Temper”, system fetches the meaning and plays through speaker.

Once the user is satisfied with the meaning, he or she stops dictionary by pressing 0. The system resumes the speech and once the page is done, user opts for mechanism by pressing 7.

The page turning mechanism successfully turns the page as shown.

Fig (d): Page Turning

[7] FUTURE SCOPE

The system can be extended to support languages other than English, and to support for different speech rates. The system can be made smarter in the sense – whenever user wants particular page to be read, if the page_N.mp3 is already existing then system can skip the conversion of image to speech, rather should directly play the mp3 file.
The system can provide the bookmark feature. If the user feels any sentences or paragraphs important, he/she can save them and keep them for ease of access. If the user wants to navigate to the particular page number, the system should actuate page turning mechanism to go to that particular page.

**[8] CONCLUSION**

In this Smart Reader for Blind people, the design and development of a working model provides a smart reading assistance for the Blind and Visually Impaired people. The system provides autonomous page turning mechanism and interactive dictionary querying features, ultimately giving a feel of comfort for the Blind and visually impaired people.

The system not only supports Blind and visually impaired people, even a person who wants luxury and comfortability, or an aged person can afford this. Once installed and configured can act as a perfect personal device for the user. The system even finds small scale applications in Schools, Libraries etc. We hope this encourages the reader and brings in them an inspiration to work on the advancements or a similar project that helps the society.

**REFERENCES**


