ISSN No:-2456-2165

Text to Speech Conversion for Visually Impaired People

Anjaly Siby, Anisha P Emmanuel, Chikku Lawrance, Jain Mariya Jayan
Prof. Kishore Sebastian, Assistant Professor
Department of Computer Science and Engineering
St Joseph's College of Engineering and Technology Palai, Kerala, India

Abstract:- Visually impaired persons are those who have lost the ability of vision partially or permanently (blinds). This leads to difficulties in performing the daily activities such as reading, socializing, walking etc. Text and speech are the main communication medium for humans [1]. The number of visually impaired persons increases day by day due to different eye diseases like Cataract, Refractive error, Glaucoma, childhood blindness, age related macular degeneration, Diabetics etc. Here we set forth a camera based text reading mechanism that converts the text present on the paper using an auto focusing camera into speech which the person can listen to by using a pair of headphones. The proposed concept involves extracting text from the image captured using Tesseract Optical Character Recognition (OCR) and converting the text to speech. Here, we divide texts to Syllables so as to convert it to speech, where the sound of each syllable is pre-recorded.

I. INTRODUCTION

One of the greatest difficulties faced by a sightless person is the disability to read. Text is present everywhere ranging from bulletin to billboards to digital sections etc. Blind people face a lot of difficulties. There have been developments on mobile phones and computers that assist a blind person by combining computer vision tools with other existing expedient products such as Optical Character Recognition (OCR) system.

The proffered system assists blind people by capturing the text and then by reading it to them. Extracting the text present is enacted with OCR. It is a tactic for transformation of images of writings on a label, printed books etc. OCR replaces binary images with texts and also detects white spaces. It also parses the integrity of the recognized text.

Optical Character Recognition is the mechanical or electronic transformation of images of typed, handwritten or printed text into machine-encoded text, or from subtitle text superimposed on an image [5]. In order to take a picture, a delay of around 2 seconds occurs and is then processed by Raspberry Pi. The text is splitted into syllables and the corresponding sounds are produced which can be heard through the audio jack.

This paper strives to build an effectual camera based convenient text reading device which is intuitive.

II. EXISTING SYSTEMS

In the paper 'Image to Speech Conversion for Visually Impaired', the authors Asha G. Hagargund, Sharsha Vanria Thota, Mitadru Bera, Eram Fatima Shaik[2] explain about the device which can detect efficiently from any complex background. The basic framework is an embedded system that captures an image, extracts the image that contains text and then converts that text to corresponding speech. A series of image pre-processing steps are used to find the text and remove the background. Here, one of the disadvantages is that when edge detection is performed some of the alphabets can be miss-detected and this leads to false output generation by the OCR.

In the paper 'Text to Speech Conversion using Raspberry - PI' by Vinaya Phutak, Richa Kamble, Sharmila Gore, Minal Alave, R.R.Kulkarni[3] explains about the device which helps the blind people to interact with computer effectively through vocal interface. It consists of two parts, mainly, image processing module and voice processing module. OCR is used to extract text which is then being used for further processing. The image is taken by the webcam and is processed by the Raspberry Pi. The image is converted to grayscale. TTS is used in order to read English alphabets and numbers. The TTS algorithm used here is not known.

In the paper 'Text to Speech for the Visually Impaired' Mrs.Shilpa Reddy K, Mounika S.K, Pooja K, Sahana N[4] explains about an RPi-based system with a high resolution webcam. The system uses OTSU algorithm, for image thresholding, MODI algorithm, for image transposition and SAPI libraries. The image can also be uploaded by the admin to the cloud.

ISSN No:-2456-2165

III. PROPOSED METHOD

The main intention is to help amaurotic persons in reading the text as a camera based convenient text reader. The concept involves text recognition by OCR from an image captured by a camera mounted on a spectacle. Conversion of the text file to a voice format is done by dividing the text into syllables, which the user can listen to via headphones. Portability of the system is attained by providing a battery backup, thus enabling the user to carry the device everywhere, at any time.

The entire project is divided into two sections. Former is the conversion of text using OCR and the latter is the conversion of text to speech.

Figure 1 shows the system design.

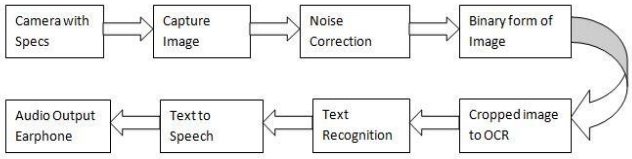

Fig 1:- System design

The work plan is as follows: The camera mounted on the Spectacle captures the image of the text in front of it. The captured image is then changed to Grayscale. It is then filtered to reduce the noise on the image using a Gaussian filter. Hence noise correction on the image is done. The filtered image is then transformed to a binary form, which is then cropped to remove the image portions with no characters. The cropped image is then loaded to the Tesseract OCR in order to perform text identification. The output of the OCR, text file, is converted to syllables and their corresponding sounds are obtained, which can be heard using headphones.

IV. HARDWARE IMPLEMENTATION

The hardware required for this system includes:

- Raspberry Pi
- Camera on Spectacle
- Headphones
- Powerbank

- *Raspberry Pi*

Raspberry Pi is a linux based programmable computer that could do all the normal operations in PC. Raspberry Pi 4 Model B with a 1.5 GHz 64-bit quad core ARM Cortex-A72 processor is used in this system. There are two USB 2.0 ports, two USB 3.0 ports. The Pi 4 is also powered via USB-C port. The model comes with a standard 40 pin GPIO header. The board is operated in such a way that the code starts to execute as the power is turned ON. the output is obtained through the audio jack.

- *Camera on Spectacle*

A compatible camera mounted on the camera is used to capture the image. 8.0 Megapixel Auto Focus USB 2.0 Camera is used in this project. It provides a very high quality image with a resolution of 1920×1080 with a 30fps frame. It is low power consuming. The USB powered camera is connected to the Raspberry Pi.

- *Power Bank*

Power bank of 10,000 mAh capacity is used to achieve portability. It can provide backup for at least 10 hours.

V. SOFTWARE IMPLEMENTATION

Raspberry Pi works on the platform Raspbian, which is a Debian based operating system. Python 3 is used to write the entire code. Libraries like OpenCV, Pillow, Time, PyAudio etc are used. OpenCV is a library function which is mainly focussed on real time computer vision. Pillow, Python Imaging Library, was used to support opening, manipulation and saving different image file formats. PyAudio supports Python bindings for PortAudio, the cross-platform audio I/O library. Tesseract is an optical character recognition engine with open-source code and is the most popular and qualitative OCR-library. OCR uses artificial intelligence for text search and its recognition of images. The open-source digital audio editor and recording application software, Audacity, is used for recording datasets to create a text to speech engine. The CMU Pronouncing Dictionary, created by the Speech Group at Carnegie Mellon University is an open-source pronouncing dictionary, originally for use in speech recognition research was used for developing the datasets.

- *Image to Text Conversion*

The conversion of the image captured by the camera to the corresponding text is carried out using PyTesseract**,** an optical character recognition (OCR) tool for python.

At first, the capture duration of the camera is initialized in milliseconds. The start time of the camera is set to a variable. When the elapsed time is less than the capture time, the image is captured and the camera is released.

Figure 2 shows the image taken by the camera.

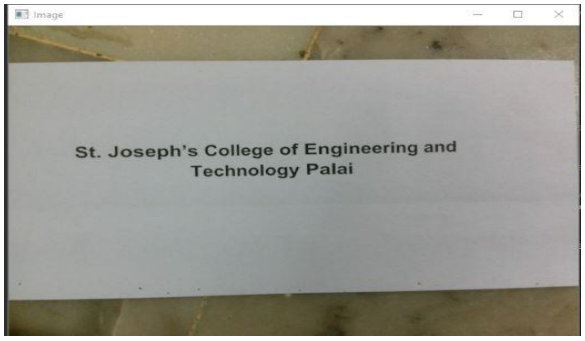

Fig 2:- Image taken by camera

ISSN No:-2456-2165

Now, the image is preprocessed so as to convert it to the corresponding text. For this, an adaptive gaussian threshold is used. Adaptive thresholding is useful in separating desirable foreground image objects from the background based on the intensity differences in the pixel of each particular region. Image is enhanced, smoothening and sharpening, using the method Blur. Figure 3 shows the image after applying threshold and blur.

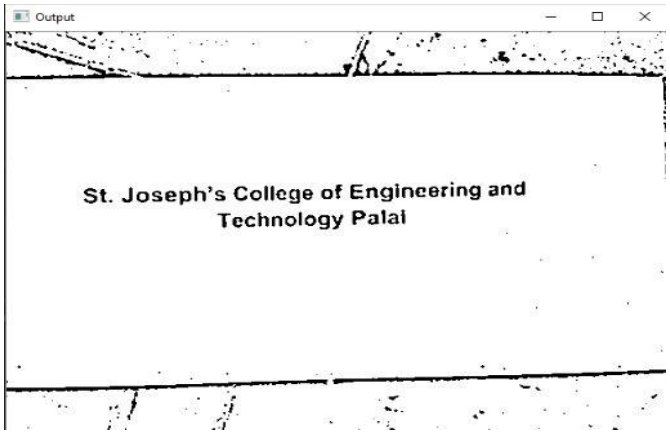

Fig 3:- Image after preprocessing

The text present in the image should be converted to string. Python-Tesseract is used for this process. Figure 4 shows the result of this.

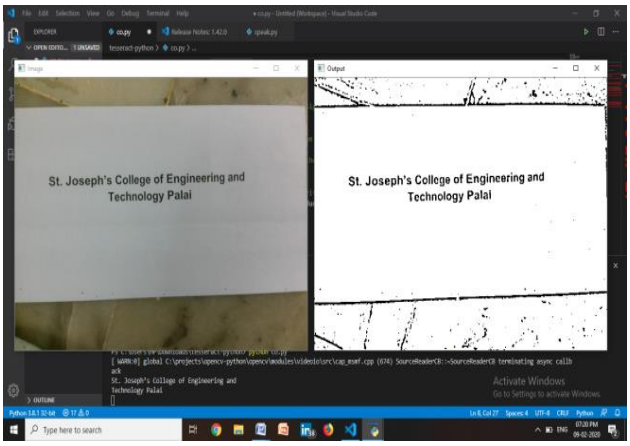

Fig 4:- Conversion of image to text

The entire process of image to text conversion using pytesseract is shown by means of a flowchart in Figure 5.

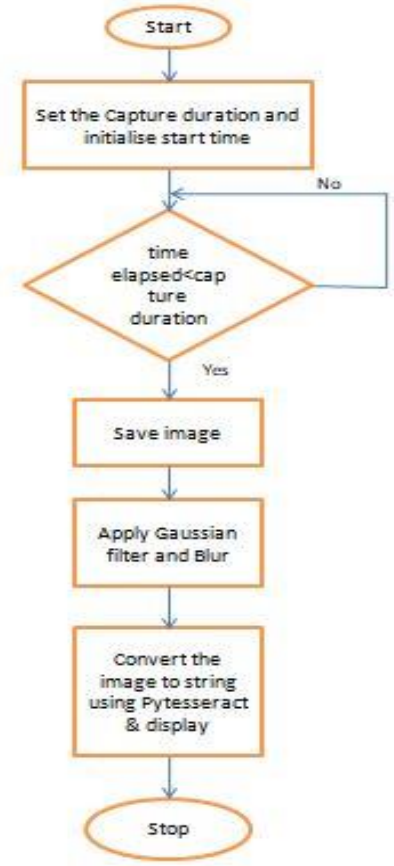

Fig 5:- Flowchart

➢ *Text to Speech Conversion*

The CMU Pronouncing Dictionary, created by the Speech Group at Carnegie Mellon University is an open-source pronouncing dictionary, originally for use in speech recognition research was used for developing the datasets. The open-source digital audio editor and recording application software, Audacity, is used for recording datasets to create a text to speech engine. The data sets are saved with .wav extension. The input of the TTS engine is a text file, which is converted to speech and the output can be obtained through headphones. Figure 6 shows the block diagram of the conversion.

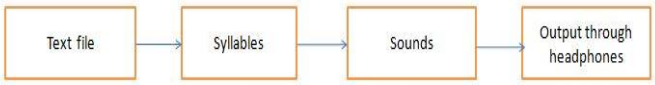

Fig 6:- Text to Speech Block diagram

The input from the TTS engine, text file, is converted to the corresponding syllables. A syllable is a unit of organization for a sequence of speech sounds. The corresponding sound of each syllable is collected and are produced, which is audible to the user via headphones.

ISSN No:-2456-2165

VI. RESULTS

Figure 4 shows the result that has been obtained. Furtherly, the text is converted to speech by dividing it into syllables and producing the corresponding sounds through the earphone plugged into the audio jack.

VII. CONCLUSION

The suggested approach helps the amaurotic people to read effectual. It is also a convenient, economical and viable approach.

ACKNOWLEDGEMENT

The project is being funded by Kerala State Council for Science, Technology and Environment(KSCSTE).

REFERENCES

[1]. M. Rajesh ; Bindhu K. Rajan ; Ajay Roy ; K. Almaria Thomas ; Ancy Thomas ; T. Bincy Tharakan "Text recognition and face detection aid for visually impaired person using Raspberry PI" 20-21 April 2017
[2]. Asha G. Hagargund, Sharsha Vanria Thota, Mitadru Bera, Eram Fatima Shaik "Image to Speech Conversion for Visually Impaired" Volume 03 - Issue 06 || June 2017 || PP. 09-15
[3]. Vinaya Phutak, Richa Kamble, Sharmila Gore, Minal Alave, R.R.Kulkarni "Text to Speech Conversion using Raspberry - PI" Volume 4, Issue 2, February – 2019
[4]. Mrs.Shilpa Reddy K , Mounika S.K,Pooja K, Sahana N "Text to Speech for the Visually Impaired" International Research Journal of Computer Science (IRJCS) ,Issue 05, Volume 4 -May 2017
[5]. Optical Character Recognition https://en.wikipedia.org/wiki/Optical_character_recognition