International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 03 - Issue 06 || June 2017 || PP. 09-15

Image to Speech Conversion for Visually Impaired

Asha G. Hagargund[1], Sharsha Vanria Thota[2], Mitadru Bera[3], Eram Fatima Shaik[4]

[1]*Assistant Professor, Department of Electronics and Communication Engineering, BMSIT&M, Bangalore*
Affiliated to Visvesvaraya Technological University, Belgaum, India
[2]*Dept. of Electronics and Communication Engineering, BMSIT&M, Bengaluru*
Affiliated to Visvesvaraya Technological University, Belgaum, India
[3]*Dept. of Electronics and Communication Engineering, BMSIT&M, Bengaluru*
Affiliated to Visvesvaraya Technological University, Belgaum, India
[4]*Dept. of Electronics and Communication Engineering, BMSIT&M, Bengaluru*
Affiliated to Visvesvaraya Technological University, Belgaum, India

Abstract: Visual impairment is one of the biggest limitation 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.

Keywords: Embedded system, OCR, pre-processing, Raspberry Pi, TTS

1. Introduction

In our planet of 7.4 billion humans, 285 million are visually impaired out of whom 39 million people are completely blind, i.e. have no vision at all, and 246 million have mild or severe visual impairment (WHO, 2011). It has been predicted that by the year 2020, these numbers will rise to 75 million blind and 200 million people with visual impairment [7]. As reading is of prime importance in the daily routine (text being present everywhere from newspapers, commercial products, sign-boards, digital screens etc.) of mankind, visually impaired people face a lot of difficulties. Our device assists the visually impaired by reading out the text to them.

There have been numerous advances in this area to help visually impaired to read without much difficulties. The existing technologies use a similar approach as mentioned in this paper, but they have certain drawbacks. Firstly, the input images taken in previous works have no complex background, i.e. the test inputs are printed on a plain white sheet. It is easy to convert such images to text without pre-processing, but such an approach will not be useful in a real-time system [1][2][3]. Also, in methods that use segmentation of characters for recognition, the characters will be read out as individual letter and not a complete word. This gives an undesirable audio output to the user. For our project, we wanted the device to be able to detect the text from any complex background and read it efficiently. Inspired by the methodology used by Apps such as "CamScanner", we assumed that in any complex background, the text will most likely be enclosed in a box eg billboards, screens etc. By being able to detect a region enclosing four points, we assume that this is the required region containing the text. This is done using warping and cropping. The new image obtained then undergoes edge detection and a boundary is then drawn over the letters. This gives it more definition. The image is then processed by the OCR and TTS to give audio ouput.

1.1. BASIC BLOCK DIAGRAM

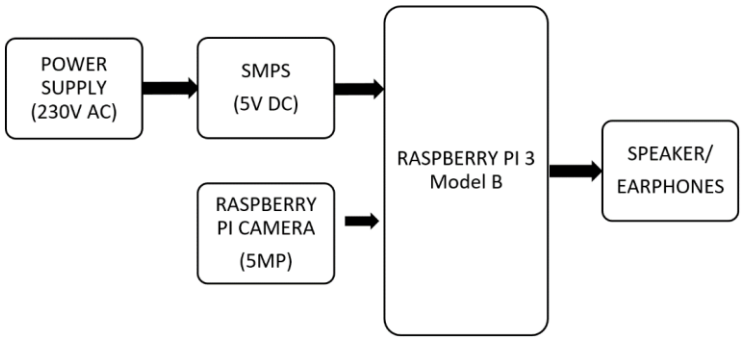

Fig 1.

The device consists of a Raspberry Pi 3B, speaker or earphones, Raspberry pi camera, power supply (230V AC) and a switched mode power supply (SMPS). The SMPS converts the 230V AC power supply to 5V DC to power the Raspberry Pi. The camera must manually be pointed towards the text and a picture is captured. This picture is then processed by the Raspberry Pi and the audio output is heard through the speaker.

1.2. OCR ENGINE

The extraction of the text in the image is done using optical character recognition (OCR). OCR is a field of research in pattern recognition, artificial intelligence and computer vision. It is the conversion of the images of typed, handwritten or printed text into a digital text or computer format text. Earlier OCR versions had to be trained in each character of a text with its specific font. Today, advanced OCRs are available that have a high degree of accuracy, support a wide variety of image formats, languages and fonts. For our project, we have used Tesseract OCR. It is the most accurate open source OCR engine and is powered by google. It can be used on the Linux, mac and windows platform. The newest Tesseract version, 3.4 supports a hundred languages. However, images must undergo a number of pre-processing stages like noise removal, scaling etc. otherwise the output will be of low quality.

1.3. TTS SOFTWARE

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.

2. Motivation

Our device is designed for people with mild or moderate visual impairment by providing the capability to listen to the text. It can also act as a learning aid for people suffering from dyslexia or other learning disabilities that involve difficulty in reading or interpreting words and letters. We wish to enable these people to be independent and self-reliant as they will no longer need assistance to understand printed text. Such people will always have access to information hence they will never feel at a disadvantage. The impact of the development and introduction of our system into the technological world will be a revolutionary boon to modern civilization.

3. Literature Survey

Visual impairment or vision loss is defined as the decreased ability to see clearly and cannot be fixed using glasses. Blindness is the term used for complete vision loss. The common causes of vision loss are uncorrected refractive errors, cataracts and glaucoma. People with visual impairment face a number of difficulties in normal daily activities like walking, driving and reading.[9]

3.1. BRAILLE

Braille is writing and reading system used people who have visual impairment. Braille language is written on embossed paper. The braille characters are small rectangular blocks called cells that contain bumps called raised dots. The visually impaired person feels the arrangement of the raised dots which conveys the information. [10]

Braille literacy statistics of India: One out of every three blind people in the world is an Indian. It is estimated that nearly 15 million Indians are blind and out of that 2 million are children. Only 5% of the children receive education. Although braille readers, keyboards and monitors exist, they are not accessible to the rural communities and braille material is not easily and abundantly available. [11]

3.2. RASPBERRY PI

The raspberry Pi is a small, low cost CPU which can be used with a monitor, keyboard and mouse to become an efficient, full-fledged computer [12]. The reason we chose Raspberry Pi micro-computer for our project is that, firstly, it is an easily available, low-cost device. RPi uses software which are either free or open source, which also makes it cost-effective. The Raspberry Pi uses an SD card for storage and its small size also gives us the advantages of portability.
[13]

As a part of the software development, the Open CV (Open source Computer Vision) libraries are utilized for image processing. Each function and data structure was designed with the Image Processing coder in mind. [14]

3.3. EXISTING SYSTEMS AND THEIR LIMITATIONS

- One of the biggest advantages of barcode readers is portability. Hence, they can be used by the visually impaired in identifying different products. An extensive database is created which contains all the information about the product. The user simply scans the bar code and the product details are listed through e-braille readers. The disadvantage with this product is that the user might not be able to point the bar code reader in the correct direction. [2]
- Another approach is optical enhancement solutions such as an optical zooming device that expands the braille character. However, not all visually impaired people need to know braille language. [4]
- Some methods aim at converting text to speech. This is accomplished using a scanner, speakers and a computer. This method is efficient only with simple scanned documents. It cannot extract text from an image with a complex background. [4]

4. System Specifications

4.1.1. SOFTWARE SPECIFICATIONS

Raspbian is a free operating system, based on Debian, optimized for the Raspberry Pi hardware. Raspbian Jessie is used as the version is RPi's main operating system in our project. Our code is written in Python language (version 2.7.13) and the functions are called from OpenCV. OpenCV, which stands for Open Source Computer Vision, is a library of functions that are used for real-time applications like image processing, and many others [14]. Currently, OpenCV supports a wide variety of programming languages like C++, Python, Java etc. and is available on different platforms including Windows, Linux, OS X, Android, iOS etc. [15]. The version used for our project is opencv-3.0.0. OpenCV's application areas include Facial recognition system, Gesture recognition, Human–computer interaction (HCI), Mobile robotics, Motion understanding, Object identification, Segmentation and recognition, Motion tracking, Augmented reality and many more. For performing OCR and TTS operations we install Tesseract OCR and Festival software. Tesseract is an open source Optical Character Recognition (OCR) Engine, available under the Apache 2.0 license. It can be used directly, or (for programmers) using an API to extract typed, handwritten or printed text from images. It supports a wide variety of languages. The package is generally called 'tesseract' or 'tesseract-ocr'.

Festival TTS was developed by the "The Centre for Speech Technology Research",UK. It is an open source software that has a framework for building efficient speech synthesis systems. It is multi-lingual (supports British English, American English and Spanish). As Festival is a part of the package manager for Raspberry Pi, it is easy to install.

4.1.2. HARDWARE SPECIFICATIONS

Raspberry pi is a device that contains several important functions on a single chip. It is a system on a chip(SoC). The Raspberry Pi 3 uses Broadcom BCM2837 SoC Multimedia processor. The Raspberry Pi's CPU is the 4x ARM Cortex-A53, 1.2GHz processor. It has internal memory 1GB LPDDR RAM (900Mhz) and external memory can be extended to 64 GB. In Raspberry Pi 3, the two main new features are wireless internet connection 802.11n and Bluetooth 4.1 classic. It has 40 GPIO pins. [16] The Raspberry pi camera is 5MP and

has a resolution of 2592x1944. The Raspberry Pi has a 3.5mm audio port so earphones or speaker can easily be connected to it to hear audio.

5. Methodology

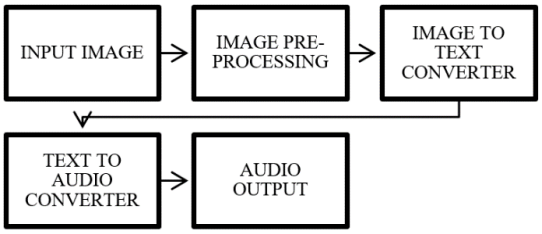

Fig 2.

Image acquisition: 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.

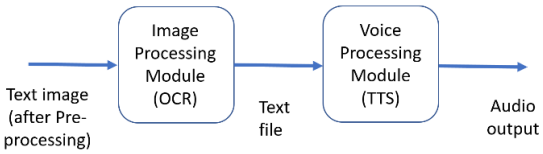

Fig 3.

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.

6. Results

The obtained output images after pre-processing are displayed below. Figure 4 shows the original image that was captured using the Pi Camera. Figure 5 to Figure 11 display the pre-processing done in each stage. And finally Figure 11 represents the image which is given as input to the OCR. Figure 12 displays the text obtained at the output of the OCR engine. It is evident that the result is not completely accurate. This is because of the less resolution of the camera used. Better results can be obtained if the camera used is a High definition camera.

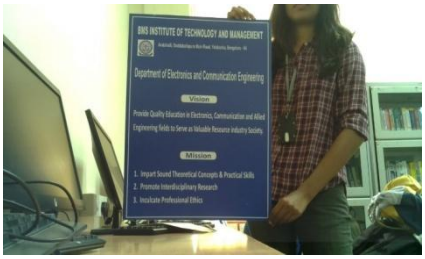

Fig 4: Original image captured from the camera

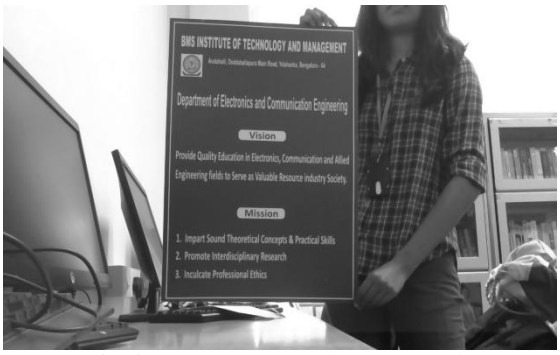

Fig 5: Image converted to gray scale

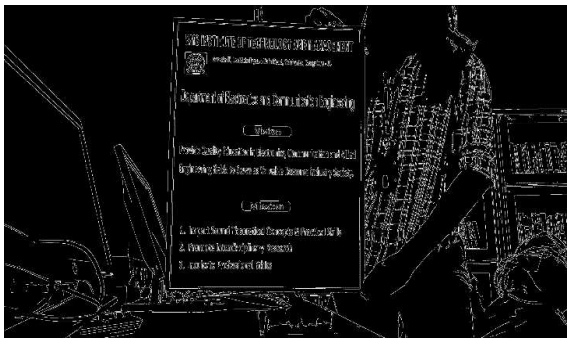
Fig 6: Performing edge detection

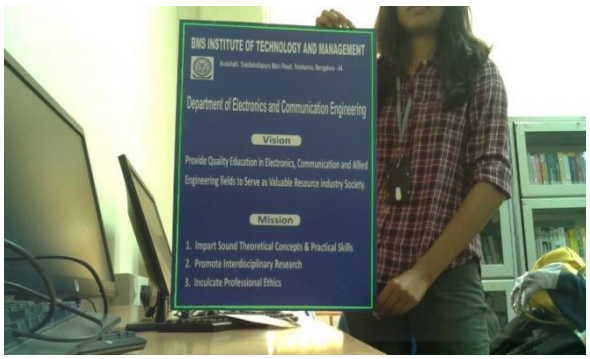

Fig 7: Contour detection

Fig 8: Warped and cropped image

Fig 9: Sharpening the image

Fig 10: Convert to grayscale before thresholding

Fig 11: Thresholding

```
BMS INSTITUTE OF TECHNOLOGY AND MANAGEMENT
(ï¬rst Aualahalii, Doddahallapura Main Road, Yelahanlca.
Bengaluru - 64

Department of Electronics and Communication Engineering
Provide Quality Education in Electronics, Communication and
Allied
Engineering fields to Serve as Valuable Resource industry Society.
1. Import Sound Theoretical Concepts & Practical Skills
2. Promote Interdisciplinary Research
3. Inculcate Professional Ethics
```

Fig 12: Tesseract output

7. 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. The future work can be developing devices that perform object detection and extracting text from videos instead of static images.

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/

[15]. http://aishack.in/tutorials/opencv/

[16]. http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_setup/py_intro/py_intro.html

[17]. http://hackaday.com/2016/02/28/introducing-the-raspberry-pi-3/