Middle-East Journal of Scientific Research 24 (11): 3429-3433, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.3429.3433

A Cost Effective Facile Reading Device for Visually Impaired People

[1]M.A. Fathima Tabassum, [2]S. Bhuvaneswari, [3]R.M. Deivanai and [4]J. Priyanka

[1]Assistant Professor, Department of Computer Science and Engineering,
Rajalakshmi Institute of Technology, Chennai, India
[2]UG Students, Department of Computer Science and Engineering,
Rajalakshmi Institute of Technology, Chennai, India

Abstract: The major challenge faced by visually impaired students in the educational environment is the abundant usage of text books and other visual materials.Majority of the visually impaired use Braille for reading documents and books. The students who have recently lost their vision over a number of years find it difficult to adapt to the braille technology. Therefore there is a need for automated devices that help people with low or no vision to access books or documents at an affordable cost. Very few researches focus on systems that assist visually impaired people in reading texts. This project aims to contemplate the image recognition system with text to speech synthesis and to manipulate various edge detection algorithms such as Roberts, Canny and Sobel edge detection algorithm. Their results are compared for better accuracy in extracting the complex texts at the edges. The image recognition process is done by raspberry pi,which is integrated with raspberry pi camera module via USB port,scans the input image and undergoes image processing to obtain the region of interest. Then Tesseract, an open sourced OCR engine, which converts the image into machine encoded text is used. The output combined is given to TTS synthesizer, which converts text to audio that can be stored in a databasethereby making it accessible for future use.

Key words: Visually impaired · Reading assistant · Optical Character Recognition (OCR) · Edge detection · Text-to-Speech

INTRODUCTION

According to the reports of recent survey, from the world's total population,most of the visually impaired people live in developing countries. Reading is the major problem faced by Blind people. The reading materials available for the blind people are in the form of braille so the people must know the braille language for learning. The disadvantage of the braille language, the books and documents have to be converted into raised dots but it is difficult for all the blind people to know the braille language to overcome this,a facile device is needed to help the blind in reading. The Raspberry pi is the main unit that scans the input image and undergoes image processing to obtain the region of interest. The edges of the image are detected by using canny edge detection, sobeledge detection and Kirsch edge detection algorithms and their results are compared for better accuracy. Then the Optical character recognition (OCR) [1], tesseract which is aopen sourced and highly portable software that converts the image into machine encoded text. The obtained output is combined into words are given to the text to speech (TTS) synthesizer, a computer based system that can read text aloud automatically for the written document.

Existing System: The most basic and widely used method is Braille. Apart from that the other technology used is Talking Computer Terminal, Computer Driven Braille Printer, Paperless Braille Machines, Optacon etc. These technologies use different techniques and methods allowing the person to read or convert document to Braille. Nowadays Computers are designed to interact by reading the books or documents. Synthesized voice is

Corresponding Autor: M.A. Fathima Tabassum, Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, India.

used to read the content by the computers. We also have devices that scan the documents and use interfaced screen to allow the blind to sense the scanned documents on the screen either in Braille or by using the shape of the letters itself with help of vibrating pegs. Various phone applications are also developed to help in reading. The "DRASHTI"-an android reading aid which identifies and interpret the text or characters displayed on the screen and this is been re-presented to the user in the form of voice.

Proposed System: The proposed system consists of two phases (i)OCR (ii)Text –To-Speech. The OCR phase includes the Raspberry pi, Pre-processing the text, Feature Extraction,Tesseract for converting into machine encoded text. The Text-to-speech phase uses synthesizer to convert the text into audio output.

Raspberry PI: Raspberry Pi's inception began in 2006 it was finally released on 19 February 2012 as two models: Model A and Model B. For the years,the work was done on Arduino Boards but with the launch of very cheap Raspberry Pi, it all changed.Raspberry Pi board costs only $25 and does the work of a computer costing hundreds of dollars.Though its purpose is not to replace computers,laptops,etc.Boot it up, and you have a got a fully funtional powerhouse.The Raspberry pi offers an another path by encouraging exprimentation by lowering the cost.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.The technical specifications and the Raspberry pi is shown in Fig 1(a)(b).

Fig. 1a: Raspberry pi module

Features	Model B
CPU	700 MHz
RAM	512 MB
Ethernet	10/100 Gbit
GPU	Broadcom video Core 4
Video input	HDMI and RCA composite
USB Ports	2
I/O Ports	26 pin P1 Header(General purpose Input Output)

Fig. 1b: Raspberry pi Module Model B specifications

Camera Module: The Raspberry Pi camera module has a small piece of translucent blue plastic film covering the lens and the 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.The camera module is used to capture the image and store it for future use.

Pre-Processing: The image has to be formatted according to the specification for the tessaract engine. The colour image are converted into gray scale image by removing the background for retrieving the text. Next, the text is extracted from the image by using feature extraction algorithm.

Feature Extraction: The edges are detected to obtain the required information from the image.for this various edge detection techniques are used.

Canny Edge Detection Technique

Canny edge detection algorithm extracts thin, clear edges. Apply Gaussian filter to smooth the image in order to remove the noise.

Step1: Find the intensity gradients of the image

Step2: Apply non-maximum suppression to get rid of spurious response to edge detection.

Step3: Apply double threshold to determine potential edges.

Step4: Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

Since all edge detection results are easily affected by image noise, it is essential to filter out the noise to prevent false detection caused by noise. To smooth the image, a Gaussian filter is applied to convolve with the image. This step will slightly smooth the image to reduce the effects of obvious noise on the edge detector. The equation for a Gaussian filter kernel of size $(2k+1)\times(2k+1)$ is given by:

$$H_{ij} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(i-k-1)^2+(j-k-1)^2}{2\sigma^2}\right)$$

Here is an example of a 5×5 Gaussian filter, used to create the image to the right, with σ = 1.4. (The asterisk denotes a convolution operation).

$$\mathbf{B} = \frac{1}{159}\begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} * \mathbf{A}.$$

Sobel Edge Detection Technique: The Sobel operator performs a 2-D spatial gradient measurement on an image and emphasizes regions of high spatial gradient that correspond to edges. Typically it is used to find the approximate absolute gradient magnitude at each point in an input greyscale image. Compared to other edge operator, Sobel has two main advantages, Since the introduction of the average factor,it has some smoothing effect to the random noise of the image. Because it is the differential of two rows or two columns,so the elements of the edge on both sides has been enhanced,so that the edge seems thick and bright.

Roberts Edge Detection Technique: According to Roberts, an edge detector should have the following properties: the produced edges should be well-defined, the background should contribute as little noise as possible and the intensity of edges should correspond as close as possible to what a human would perceive. With these criteria in mind and based on then prevailing psychophysical theory Roberts proposed the following equations:

$$y_{i,j} = \sqrt{x_{i,j}}$$

$$z_{i,j} = \sqrt{(y_{i,j} - y_{i+1,j+1})^2 + (y_{i+1,j} - y_{i,j+1})^2}$$

where x is the initial intensity value in the image, z is the computed derivative and i, j represent the location in the image.

The results of this operation will highlight changes in intensity in a diagonal direction. One of the most appealing aspects of this operation is its simplicity; the kernel is small and contains only integers. However with the speed of computers today this advantage is negligible and the Roberts cross suffers greatly from sensitivity to noise.

OCR Engine: Optical character recognition, usually abbreviated to OCR [2], 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. Tesseract was open sourced in 2005 and is highly portable. Tesseract is a free software optical character recognition engine for various operating systems. Tesseract [4] 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. The input text is converted into a binary image and then it outlines the character by finding the text in lines. After processing is completed, the content of the output is present in. txt file. Tesseract is a command based tool.

Text to Speech: A text to speech (TTS) synthesizer is a computer based system that can read text aloud automatically. 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.

Block Diagram: As shown in the figure 2,the scanned image is taken as input and converting the color image to Gray scale image by using image detection algorithms such Roberts, Sobel and Canny edge detection algorithms are used and their results are compared for better accuracy in extracting the complex texts at the edges. Then, the acquired text from the OCR is sent to the Text to speech synthesizer to get the audio as output

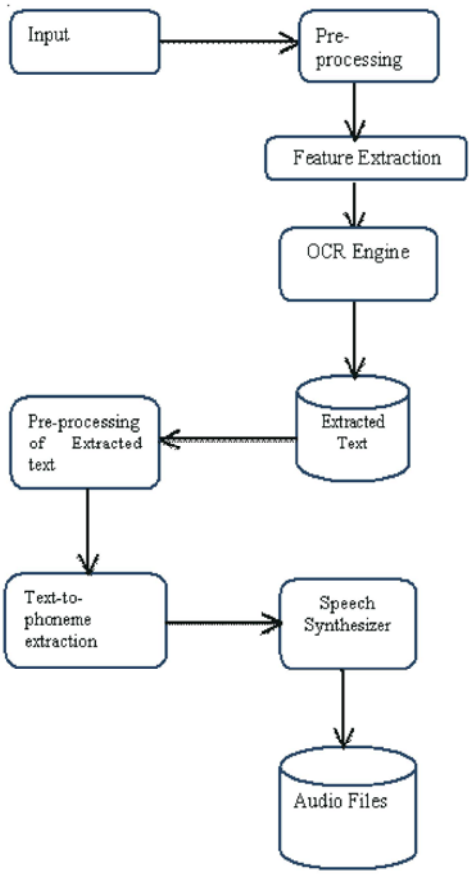

Fig. 2: Automated Facile Reading Device

Step1: The first part was booting the Raspberry Pi board by installing the Operating system Raspbian OS and installing the essential libraries and packages.

Step2: Next is the image acquisition system, in which we have interfaced a webcam, to capture the image of the text document.

Step 3: This image then goes through Pre-processing [3], in which we obtain the region of interest (ROI) where-in, we get separate sentences and then words are separated and segmented.

Step 4: This data is then given to Template Identification, where the characters are detected and we obtain the individual alphabets.

Step 5: This data is then given to the OCR Algorithm which converts the image data to text data. For OCR we will be writing a program for better outputs.

Step 6: The Algorithm scans the image, checks each alphabet or letter and gives a corresponding text output after verifying it with its own database.

Step 7: Text to speech converter where the text data is converted to an audio output and is played through the earphones connected to the audio jack.

Implementation: The setup is foldable and hence its portability is enhanced. It can be broken down into two parts and barely takes 5 seconds to be set up again. The two parts of the device are:

- The stand, onto which the RaspberryPi board is mounted with slot in the wooden board for the camera.
- A plain slate which has slots for inserting the paper.

Fig. 3: Input image

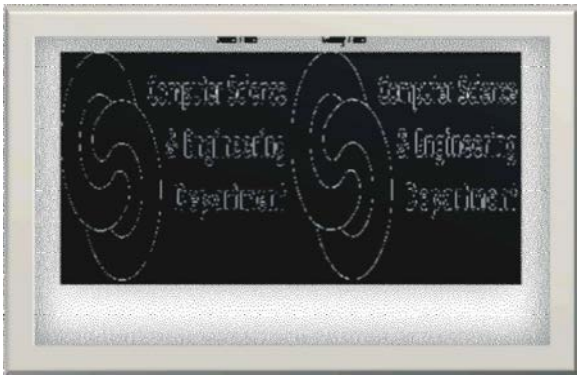

Fig. 4: Comparison of Sober and canny edge detection

CONCLUSION

The proposed system contemplates the image recognition system with text to speech synthesis and to manipulate various edge detection algorithms such as Roberts, Canny and Sobel edge detection algorithm. Their results are compared for better accuracy in extracting the complex texts at the edges.. Peak signal to noise ratio (PSNR) is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects

the fidelity of its representation. It is the logarithmic function of the peak value of the image and the mean square error. Its value must be high. It have been observed that that the Canny edge detector produces higher accuracy in detection of object edges with higher entropy, PSNR, MSE and execution time compared with Sobel, Roberts, Prewitt, Zero crossing and LOG. On the other hand Roberts edge detector has the minimum entropy with PSNR, MSE and execution time compared with others. The output combined isgiven to TTS synthesizer, which converts text to audio that can be stored in a database thereby making it accessible for future use.the (i) Number handling facility, (ii) In the case of long sentences, even the pauses (phrase breaks) can be made at the instances, where a human speaker will naturally pause for better intelligibility and clarity.This work may be extended in the future by focusing more on the Text –to-speech synthesis phase (phase 2) by enhancing.

REFERENCES

1. Ajantha Devi and Dr. S SanthoshBaboo, 2014. Optical Character Recognition on Tamil Text Image Using Raspberry Pi, International Journal of Computer Science Trends and Technology (IJCST) – Vol. 2 Issue 4, (Jul-Aug 2014).
2. Kanimozhi, V.M. and I. Muthumani, 2014. Optical Character Recognition for English and Tamil Script, International Journal of Computer Science and Information Technologies, 5(2): 1008-1010, pp: 1008-1010(2014).
3. Shinde, A.A.D., 2012. Text Pre-processing and Text Segmentation for OCR. International Journal of Computer Science Engineering and Technology, pp: 810- 812.
4. Smith, R., 2007. An Overview of the Tesseract OCR Engine. In proceedings of Document analysis and Recognition, IEEE Ninth International Conference, ICDAR 2007.
5. Babu Renga Rajan, S., *et al*, 2015. Product label reading from handheld objects, using portable camera for visually challenged persons, (IJCSIT) International Journal of Computer Science and Information Technologies, 6(6): 5173-5176.
6. Vikram Shirol, M. Abhijit, A. Savitri, R. Pooja and K. Amrutha, 2015. DRASHTI-An android reading aid, (IJCSIT) International Journal of Computer Science and Information Technologies, 6(4): 3434-3437.