OCR Based Facilitator for the Visually Challenged

Shalini Sonth
Department of Computer Science and Engineering
M S Ramaiah Institute of Technology, Bangalore, India
shalinisonth@gmail.com

Jagadish S Kallimani
Department of Computer Science and Engineering
M S Ramaiah Institute of Technology, Bangalore, India
jagadish.k@msrit.edu

***Abstract*—This paper proposes a novel implementation of an Optical Character Recognition (OCR) based smart book reader for the visually challenged. There is a need for a portable text reader that is affordable, portable and readily available to the community. We propose a camera based framework built on the Raspberry Pi, integrated with Image processing algorithms, OCR and Text-to-Speech (TTS) synthesis module. The camera module is used to capture an image of the printed text, and the image is then subject to pre-processing before being fed into the OCR. The pre-processing stage includes binarization, de-noising, de-skewing, segmentation and feature extraction. This paper addresses the integration of a complete Text Read-out system designed for the visually challenged. The OCR used in this project is Google Tesseract and the TTS employed is Pico.**

Keywords— Optical Character Recognition (OCR), Image Processing, Text-to-Speech (TTS), Raspberry Pi, Text Extraction, English Text, Hindi Text, Assistive-text, Blind persons, OTSU Threshold, Tesseract OCR Engine.

I. INTRODUCTION

OCR is the electronic conversion of images into machine encoded text. It provides alphanumeric recognition of printed or handwritten characters. OCR has been an active topic of research in the recent past, and has wide applications in banking, healthcare, finance and education. OCR is a field of research in Computer Vision, Artificial Intelligence and Pattern Recognition.

According to the World Health Organization (WHO), around 285 million people around the world are estimated to be visually impaired, out of which 90% live in developing countries. Thus there is a pressing need to develop a book reader device that is affordable to the low income sections of the society.

This project addresses a complete text read out system built on an embedded framework. The proposed idea uses the principle of a camera based assistive device implemented on a Raspberry Pi 3 board. The integrated system consists of a camera module / web camera, Tesseract Optical Character Recognition Engine, Pico Text to Speech Engine, Speakers/ Headphones and Computer Vision software for image processing. The camera is an input device which feeds the required image for digitization. This image is processed using OpenCV libraries. The processed image serves as an input to the OCR, which digitizes the image and performs character recognition. Finally, the Text to Speech engine reads out the text to the user.

II. LITERATURE SURVEY

OCR has been an active subject of research since a decade. The rapid growth of digital libraries worldwide poses new challenges for document image analysis research and development.

A vision based text recognition system using Raspberry pi is presented in [1]. This paper proposes a complete optical recognition system. The image is first captured through the use of the Raspberry Pi camera module or through the use of a webcam. Image preprocessing techniques are applied where the unwanted noise is removed by applying threshold morphological transformations - morphological, dilation, black hat and discrete cosine. Text areas are extracted by drawing bounding boxes around the required text and the slant is corrected. The horizontal and vertical ratios are adjusted to eliminate unwanted high frequency components. After thresholding, contours of the image are produced by using special OpenCV functions. The result of image to text conversion is passed on to a speech engine, which is capable of converting text to speech using specialized libraries. The paper uses TTS Festival for this purpose.

A smart reader for the visually impaired using Raspberry Pi is discussed in [2]. The simulation of optical character recognition is done using Matlab. The captured image is sent for pre-processing where functions such as skew correction, noise removal and linearization are performed. The image is first brightened and then binarized. The image is then passed on to the segmentation phase where the image is decomposed into characters. This image is then sent to the TTS engine.

An optical character generation system for conversion of images to text is presented in [3]. The parsing consists of three phases: Character extraction, Recognition and post processing. During the recognition phase, the template with maximum

978-1-5386-2361-9/17/$31.00 ©2017 IEEE

correlation is declared as the character present in the image.

A Neural Network (NN) based technique for optical character recognition and handwritten character recognition is demonstrated in [4]. This paper also discusses other techniques like matrix matching, fuzzy logic, feature extraction and structural analysis. The neural network model used here is the Multi Level Perception model (supervised network). The outputs are given as a part of the training vector. The final output of the network is determined by activations from output layer.

A new algorithm for optical character recognition which takes advantage of the typographic uniformity of paragraphs or other layout components is proposed in [5]. The essence of adaptive OCR is to reduce multi-font classification by taking advantage of the normal occurrence of long strings of characters in the same typeface. Typeface homogeneity also helps increase the throughput.

A system that recognizes printed text of various fonts and sizes for the Roman alphabet is proposed in [6]. Shape extraction is performed directly on the graph of the run-length encoding of a binary image using a shape clustering approach. This is then fed into a Bayesian Classifier. Finally, layout and linguistic context are applied.

A technique where the binary pixels are processed using a syntactic approach that labels each logically significant block to the page according to the pre-compiled information is proposed in [7]. OCR is used to convert text blocks to ASCII forms. Image and text blocks are stored in X-Y tree data structure that can be accessed through a LAN or WAN.

A finger worn device to assist the visually impaired with reading printed books is presented in [8]. It

extracts text from a close-up camera view which enables continuous feedback.
An embedded optical character recognition on Tamil text images using Raspberry Pi is proposed in [9]. The Tamil text image is scanned and filtered, and is preprocessed. Pre-processing consists of character segmentation and image binarization. The features are extracted and the image is sent for post-processing which converts the image to text. This text is then fed to a speech engine which produces synthesized Tamil speech as output.

consists of a multimodal feedback via vibration motors, a new dual-material case design and a high-resolution mini video camera. Tesseract OCR and Flite Text to Speech has been employed. It is implemented in a novel tracking algorithm that

A survey of methods and strategies in character segmentation is discussed in [10]. The traditional approach partitions the image into sub images which are then classified independently. The second class of segmentation segments the image either explicitly or implicitly through the spatial features collected. The third class is in between the first two that employs dissection along with recombination.

A wearable text recognition tool that employs MSERs as the basis for real-time text detection is reported in [11]. The proposed method process our previous real time algorithm by exploiting hierarchical structure obtained from MSERs to yield more steady regions compared to the previous adaptive threshold method. It outperforms other published approaches computationally while maintaining equivalent text detection performance on the ICDAR dataset.

Methods such as - Candidate character detection, where the Canny edge detector acquires the edge map of the scene image, maximal stable external region, geometrical constraints of candidate characters and text string detection and discussed in [12].

III. System Design

A. Working Principle

The image of a printed text is captured through the Raspberry Pi camera module. This image is subject to pre-processing which includes correcting skew angles, sharpening of image, thresholding and segmentation. The processed image is sent to the TTS synthesizer.

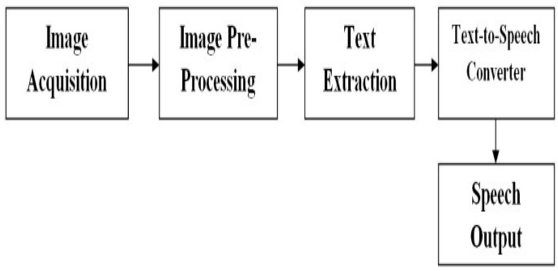

Figure 1: Working Principle

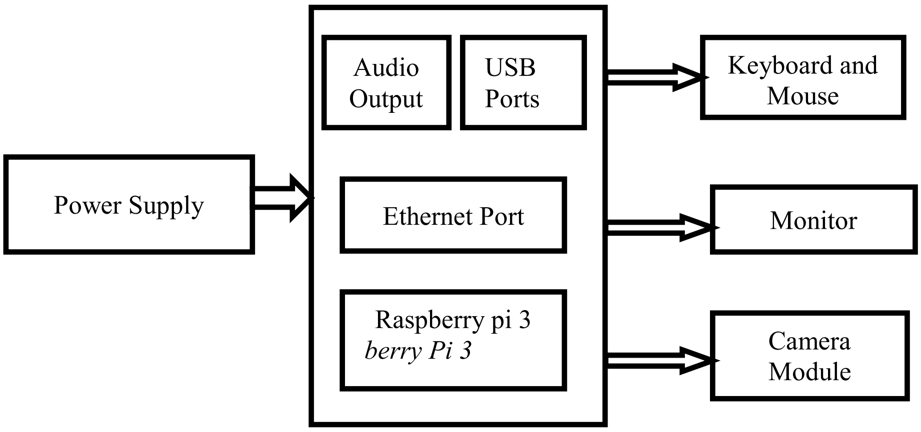

Figure 2: System Hardware Design

B. Raspberry Pi 3

The Raspberry Pi is a credit card sized single-board computer developed in the UK by the Raspberry Pi Foundation with the intention of stimulating the teaching of basic computer science in schools.
Specifications:

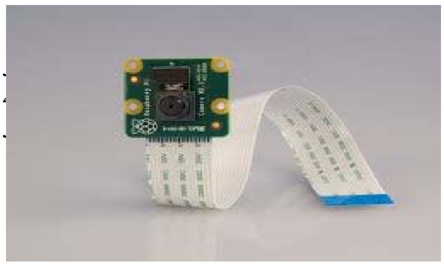

Figure 3: Raspberry Pi Camera Module

Monitor, Keyboard and speakers are connected to the GPIO pins of the Raspberry Pi. The Pi has 26 GPIO pins

- Camera interface (CSI), Display Interface(DSI)
- MicroSD card slot

C. Raspberry Pi Camera Module

Raspberry Pi NOIR Camera is used to capture images as shown in the Fig 3. The camera plugs directly into the Camera Serial Interface (CSI) connector on the Raspberry Pi. It's able to deliver clear 5MP resolution image, or 1080p HD video recording at 30fps.
Specifications:

- 5MP (2592×1944 pixels) Omni vision 5647 sensor in a fixed focus module Camera Module.
- 1.2 GHz 64 bit quad core ARMv8 CPU
- 802.11n Wireless LAN
- Bluetooth 4.1

D. Monitor, Keyboard, Speakers

The OpenCV library contains over500 functions that span many areas in vision, including factory product inspection, medical imaging, security, user interface, camera calibration, stereovision, and robotics. It also contains a generic Machine Learning Library (MLL).

E. System Software Design

- Operating System: Raspbian (Debian)
- Language: Python 2.7
- Platform: OpenCV (Linux-library)
- Library: OCR engine, TTS engine

Software released under the Apache license sponsored by Google. The TTS synthesizer used in this project is the Pico TTS.

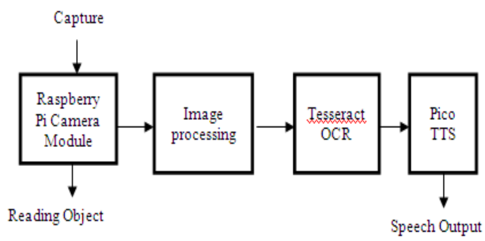

Figure 4: System Architecture

F. High Level System Architecture

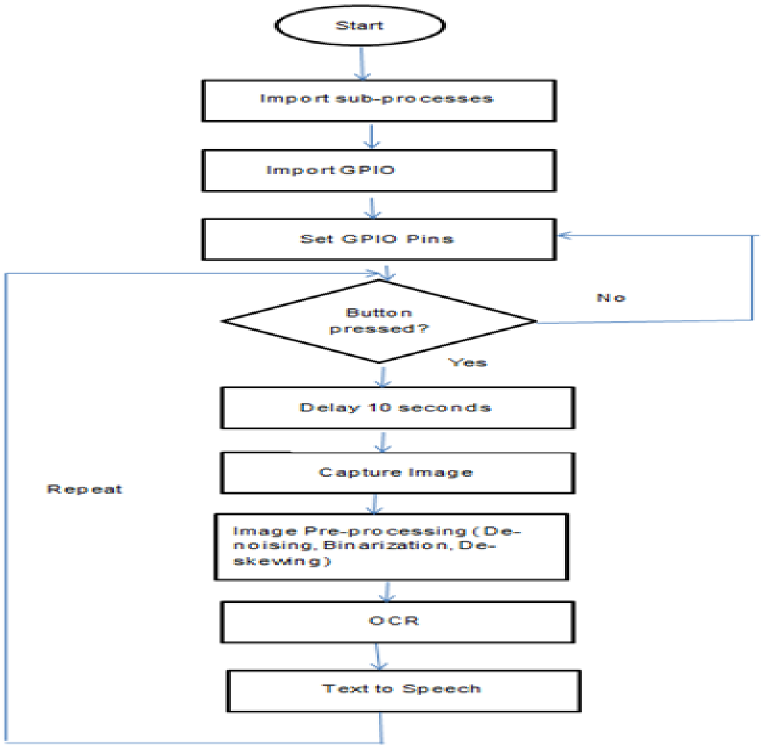

Figure 5: Detailed System Architecture

IV. TESTING AND COMPARISON

A. Defect Distribution Module Wise

Noise Removal Module

TABLE I. DEFECTS IN NOISE REMOVAL MODULE

Defect	**Defect Description**
Additional Noise in the Image	Works only for noisy images. However, it adds noise when a clear image is passed.

B. Segmentation Module

TABLE II. DEFECTS IN SEGMENTATION MODULE

Defect	Defect Description
Incorrect Segmentation	If the words within an image is too cluttered, Segmentation module fails to carry out segmentation with perfect accuracy

C. OCR Module

TABLE III. DEFECTS IN OCR MODULE

Defect	Defect Description
Italic Text	OCR fails to carry out the image to text conversion with a good accuracy when the text is italicized.
More Colors in the Image	If the image is composed of more than colors, OCR fails to perform image to text conversion with good accuracy.

D. TTS Module

TABLE IV. DEFECTS IN TTS MODULE

Defect	Defect Description
Misidentification of proper nouns	TTS fails to identify certain proper nouns which are not in dictionary.

V. RESULTS

A. *Image Processing Simulation Environment for English Language*

1) *Image Acquisition*

The image is acquired through the Raspberry Pi Camera Module. The following image has been captured from a magazine:

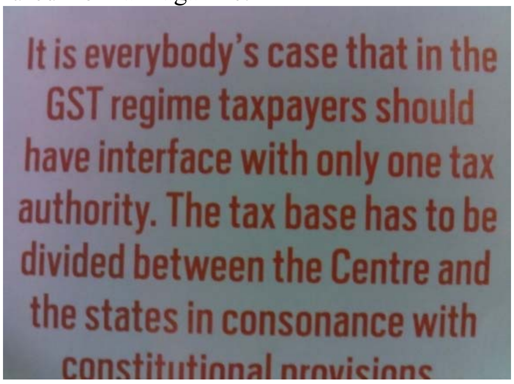

Figure 6: Image acquired from camera

2) *Binary Conversion*

The acquired image is converted to a binary format. Bit 0 indicates a black pixel and Bit 1 indicates a white pixel.

There are three threshold algorithms namely: Simple threshold, Adaptive Threshold and Otsu threshold. All the 3 algorithms work on a simple principle: if a pixel value is greater than the threshold, it is assigned one value (say white), else it is assigned another value (say black)

Algorithm:

a) Convert the image to grayscale and flip the foreground and background to ensure foreground is now white and the background is black.

b) Threshold the image setting all foreground pixels to 255 and background pixels to 0.

c) Thus, an RGB image is subject to binary conversion as shown below:

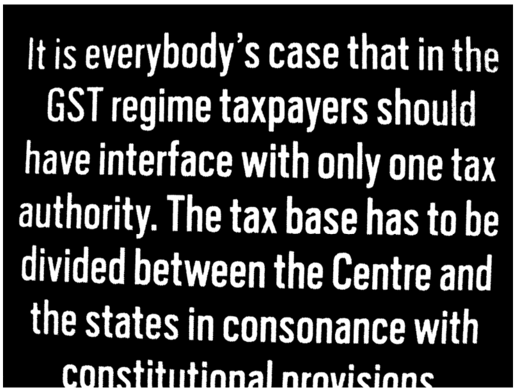

Figure 7: Binary conversion

3) *Skew Correction*

The image acquired may have orientation issues , an thus de-skewing is an important step in image processing.

Algorithm:

a) Detect the block of text in the image

b) Compute the angle of rotated text

c) Rotate the image to de-skew it.

An incorrectly oriented image is transformed as shown in the figure:

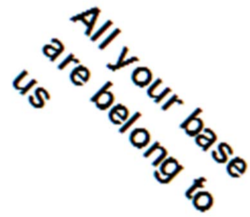

Figure 8: Disoriented image

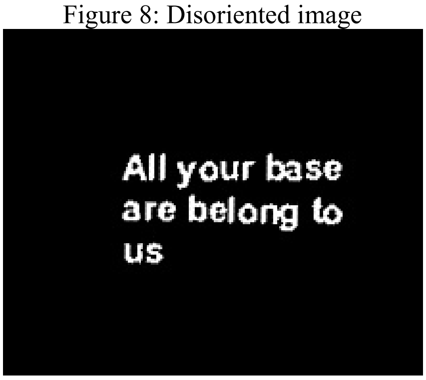

Figure 9: Binary converted de-skewed image

4) *De-noising*

The acquired image may contain noise. Noise is any random variable with zero mean.

Algorithm:

a) Apply global threshold.

b) Apply Otsu's threshold

c) Apply Gaussian Blur

The image in Figure 10 is de-noised as shown in Figure 11.

Figure 10: Noisy Image

Figure 11: De-noised image

5) Segmentation

Image segmentation partitions an image into multiple segments. The isolated blocks of characters are segmented and are labelled for identity. Connected-component labelling is used in computer vision to detect connected regions in binary digital images, although color images and data with higher dimensionality can also be processed.

Algorithm:

a) Apply Adaptive Threshold

b) Apply dilation and erosion to join the gaps

c) Find contours in the image.

d) For each contour, draw a bounding rectangle around it.

Figure 12: Original Image

Figure 13: Segmentation

B. Image Processing Simulation Environment for Hindi Language

The process is repeated for Hindi script images. Original Text Image is shown below in figure 14.

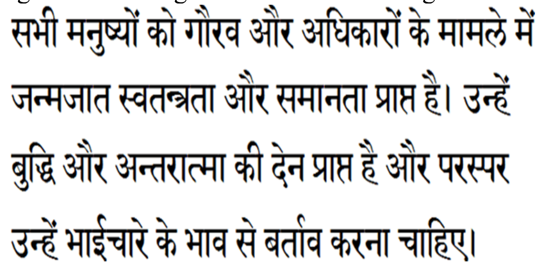

Figure 14: Acquired Hindi Image

Processed document is shown in figure 15 below.

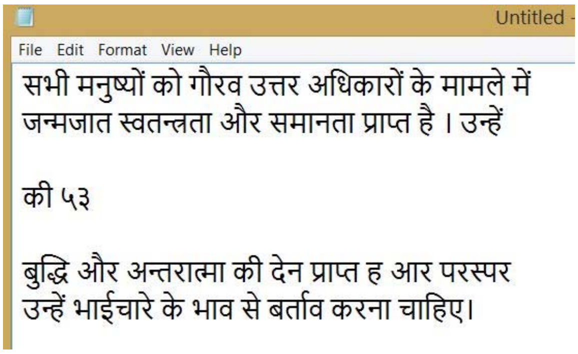

Figure 15: Processed Document by OCR Engine

VI. SCOPE AND FUTURE WORK

The scope of this project is promising. Optical Character Recognition finds its applications in Medicine, Online Retails, Education, and more. The proposed system works well for both English and Hindi scripts. Our aim for future work is to extend the same functionality for other Indian regional languages such as Tamil, Kannada, Telegu, etc. Optical character recognition finds its use in the following applications

- Captcha
- Optical Music Recognition
- Handwriting Detection
- Invoice Imaging
- Automatic Number Detection

VII.CONCLUSION

This paper proposes a novel implementation of a smart OCR based reader for the visually impaired. This project has been implemented on an embedded platform, and uses various technologies such as optical character recognition, image processing and text to speech engines. The major goal of this project was to provide an affordable hand held device to the under-represented sections of the society, i.e., the blind and the visually impaired.

REFERENCES

[1]. Nagaraja L. et al, Vision Based Text Recognition using Raspberry Pi, International Journal of Computer Applications, NCPSIA 2015.

[2]. Velumurugan D. et al, Hardware Implementation of Smart Reader for Visually Impaired using Raspberry Pi, IJAREEIE, Vol.5 , Issue 3, March 2016, pp 2055-2063.

[3]. Abhishek Verma et al, OCR- Optical Character Recognition, IJARSE,Vol. 5,Issue No 09, September 2016, pp 181-191.

[4]. N. Venkata Rao et al, Optical Character Technique Recognition Algorithms, Journal of Theoretical and Applied Information Technology, Vol.83, No. 2, 20th January 2016.

[5]. George Naggy, At the Frontiers of OCR, Invited Paper,pp 1093-1100.

[6]. Richard G. Casey et al, A survey of methods and strategies in Character Segmentation, IEEE Computer Society, July 1996.

[7]. George Naggy et al, A Prototype Document Image Analysis System for Technical Journals, University of Nebraska - Lincoln Digital Commons, at University of Nebraska – Lincoln, pp 10-22.

[8]. Roy Shilkrot et al, Finger Reader: A Wearable Device to Support Text Reading on the Go ,MIT Research Paper.

[9]. Capt Dr S Santhosh Baboo et al, Embedded Optical Braille Recognition on Tamil Braille System using Raspberry Pi, IJCTA, Vol.5, pp 1566-1574.

[10]. K. Nirmala Kumari et al, Image Text to Speech Conversion using OCR Technique in Raspberry Pi, IJAREEIE, Vol 5, Issue 5,May 2016.

[11]. Carlos Merino-Gracia et al, A Head-mounted Device for Recognizing Text in Natural Scenes, International Workshop on Camera-Based Document Analysis and Recognition, Springer Berlin Heidelberg, 2011, Page: 29-41.

[12]. Chucai Yi et al, Assistive Text Reading from Natural Scene for Blind Persons, Springer, Mobile Cloud Visual Media Computing. pp 219-241, edited book. ISBN: 978-3-319-24700-7 (Print) 978-3-319-24702-1 (Online).