



# Speech Assistance using OCR for Visually Impaired

Vipul Samala<sup>1</sup>, Faizal Shaikh<sup>2</sup>, Pratik Shukla<sup>3</sup>, Rahul Vishwakarma<sup>4</sup>, Gaurav Tawde<sup>5</sup>  
Student<sup>1,2,3,4</sup>, Assistant Professor<sup>5</sup>

Department of Electronics and Telecommunication  
Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, India

## Abstract:

Machine replication of human functions like reading is ancient dream. However, over the last few decades, machine learning has grown from a dream to reality. Today, there are already a few systems that have some promise for portable use, like portable bar code readers designed to help blind people identify different products in an extensive product database can enable users who are blind to access information about these products through speech and braille. But a big limitation is that it is very hard for blind users to find position of the bar code and to correctly point the bar code reader at the bar code. Speech is probably the most efficient medium for communication between humans. To extract the text from image we use optical character recognition technique (OCR). Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. The proposed method is a camera based assistive text reading to help blind person in reading the text present on the text labels, printed notes and products.

**Keywords:** Text to speech, Optical character recognition, Segmentation, Template matching, Raspberry pi.

## I. INTRODUCTION

Optical character recognition is a vital task in the field of pattern recognition. English character recognition has been extensively studied by many researchers but in case of Chinese, Japanese languages which are complicated. OCR (Optical Character Recognition) is an active field of research in Pattern Recognition.

In Character recognition, many experiments differ in many factors such as sample data, pre-processing techniques, feature representation, classifier structure and learning algorithm. Because of these variable factors there is no stereotypical recognition method as it is difficult to assess the performance of special classification or learning methods from the recognition accuracies reported [1].

The recognition starts after writing and printing is completed and hence is an offline process. Many studies have been performed in foreign scripts like Chinese, Japanese but for Indian scripts very few can be traced for character recognition. Due to huge character set of Indian languages and presence of vowel modifiers text recognition is difficult and even now no complete system is available.

Within the two types of character recognition, one is based on the type of text that is printed or handwritten and the other is based on data acquisition process which can be online or offline. Online systems provide better recognition as they are linked with timing information and can skip the initial step of locating the character as in offline systems. There is no control over the medium and instrument used in offline systems hence offline character recognition is more difficult and challenging.

[2] The purpose of our project is to design a system for the visually challenged persons to hear texts from the book which utilizes the benefits of the Optical Character Recognition technology which is used for character recognition by separating the text from the image. The text output thus

obtained will be synthesized and fed to a text to speech converter whose audio output will enable the user to listen to printed or handwritten text. In this way, a blind person can also read the books in the form of sound.

## II. LITERATURE REVIEW:

The history of OCR research, like that of speech recognition, is comparatively old in the field of pattern recognition. History shows many methods of character recognition have been proposed and evolved in three ages. With the development of computers, Russian Scientist Tyering attempted to provide the first character recognizers for the blind in mid 1940s. Image processing and pattern recognition techniques were combined with artificial intelligence in the early nineties to provide aid to the visually impaired [2].

Almost everyone took the subject of OCR in the early days of pattern recognition research and against what was the expectation of many people, after some initial easy progress, great difficulty in solving this problem surfaced.

Hence, people diversified their interests over a wide range of topics in the pattern recognition field. For example, image understanding and 3-D object recognition. Any computer based system that is text to speech synthesizer should be able to read any text aloud. Front end and back end are the two parts of Text to speech synthesis.

Raw text containing symbols like numbers and abbreviations are converted into equivalent written out words and the process is generally known as normalization. Phonetic transcriptions are assigned to each word by the front end and it also divides and marks the text into prosodic units and this process is known as text to phoneme conversion. The conversion of symbolic linguistic representation to sound is done by the back end [3]. The uniqueness of our project is converting the actual text into speech which has not been much developed according to our research.

### III. METHODOLOGY

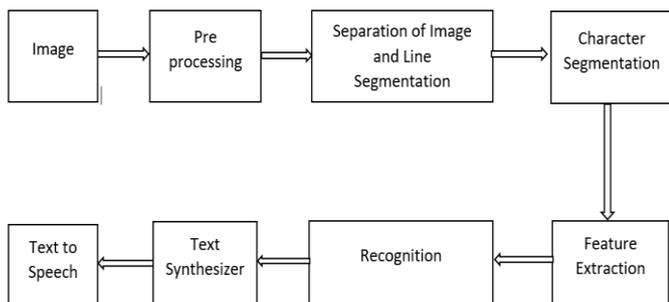


Figure.1. Block diagram

The hardware that we are using is Raspberry Pi processor which is similar to that of a mini computer and it consists of a port available for LAN (Local Area Network) and bluetooth. The camera is used to capture the plane which contains the characters. Raspberry Pi is used in our project as a module that will process the character recognition and will generate and gives the output to the text to speech synthesizer.

**Camera system:** -The camera that we are going to use is portable one that can be helpful to place to a spectacle so that the visually impaired person wearing the specs will be able to scan the printed book and image can be taken.

**Pre-Processing:-** The captured image will be further processed by the optical character recognition system. Then we binarize the image, this is done by converting the image into gray scale image and then converting the intensity values to either a high 1 value or low 0 values by thresholding. **Feature Extraction:-** The separated text will be further analyzed using feature extraction or statistics analysis and the retrieved characters will be compared to a standard set of alphanumeric characters stored in a database. The approach for feature extraction that has been adopted is template matching which performs a matching algorithm after effective segmentation of a specific character from a particular line. This is done using the Mat lab software. After complete character recognition, different words that are separated by spaces can be effectively distinguished with the help of image processing. **Speech synthesis** is the artificial production of human speech. A computer system used for this purpose is called a speech computer or speech synthesizer, and can be implemented in software or hardware products. **Text to Speech Converter:-** A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. Finally, the synthesized text will be converted into an audio output using text to speech conversion. The entire procedure will be carried out using Raspberry pi 3 models B and Matlab Simulink.

#### FLOW CHART:

##### Image Acquisition:

The image is captured using USB Webcam and the image is then pre-processed so as to simplify the recognition process.

##### Pre-processing Stage:

Preprocessing is an important step of applying a number of procedures for smoothing, enhancing, filtering etc. for making a digital image usable by subsequent algorithm in order to improve their readability for optical character recognition software. The various stages involved in the preprocessing are as following,

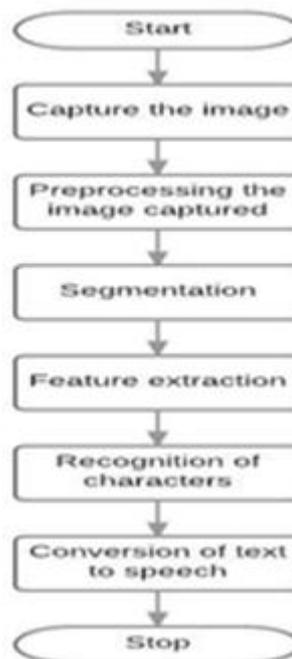


Figure.2. Flow diagram of Text to Speech system

- Binarization**

Renovation of a gray-scale image into a binary image is called as binarization or thresholding. There are two approaches for conversion of gray level image to binary form; i.e. global threshold and local or adaptive threshold. The purpose of binarization is to identify the extent of objects and also to concentrate on the shape analysis.

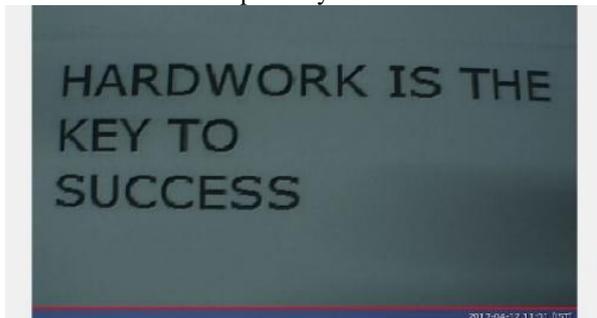


Figure.3. Image captured through webcam

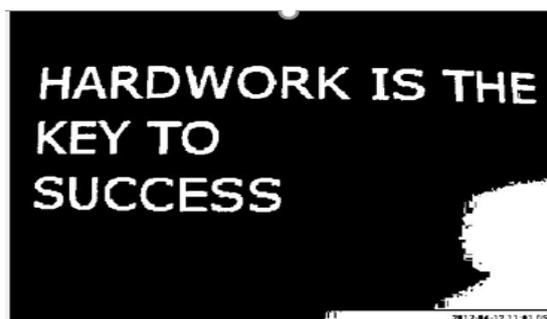


Figure.4. Pre Processed image

- Segmentation Stage**

Segmentation is one of the most important and essential process that decides the success rate of character recognition system. Segmentation is the process of partitioning an image/document into disjoint and homogeneous regions. This task is attained by finding the boundaries. There are several approaches for finding the character bounds. Threshold based segmentation, edge based segmentation and region based segmentation are some of the techniques used. In our case we have used threshold based segmentation. Thresholding maps a

grey-valued image to a binary image.



Figure.5. Segmented line

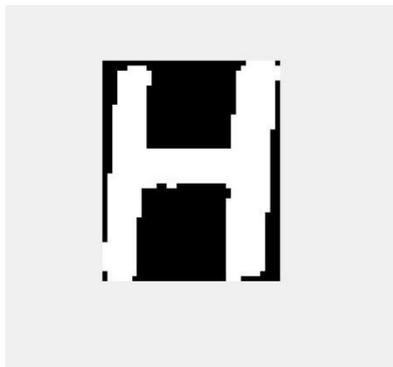


Figure.6. Segmented character

### Feature Extraction

This step is the heart of the OCR system. Feature extraction is the crucial phase in numeral identification as each numeral is unique in its own way, thus distinguishing itself from other numerals. Hence, it is very important to extract features a way that the recognition of different numerals becomes easier on the basis of the individual features of each numeral. Feature extraction is a set of procedures for extracting or measuring the most important and relevant shape information contained in the character or pattern. This step simplifies the process of classification.

### Template matching method

An appropriately placed input character is scanned vertically from top to bottom by a slit through which the reflected light on the printed input paper is transmitted to a photo detector. It is a simple calculation using only algebraic addition to obtain a value which is proportional to the area of the black portion within the slit which segments the input character. Then the sampled values are sent to the register to convert the analog values to digital ones. Template matching is done by taking the total sum of the differences between each sampled value and the corresponding template value, each of which is normalized.

### Structure analysis method

The principle underlying template matching is really only appropriate for the recognition of printed characters. However, we have another set of hand-printed handwritten characters is so large that it is difficult to create templates for them. A so called structure analysis method has been applied to handwritten character recognition. In the case of the structural analysis method, there is no mathematical principle. Rather it is still an open problem and there is no sign that it will be solved in the near future. Hence, our intuition has been the most reliable weapon in attacking this problem. The handwritten numeral recognition problem has been studied from syntactic recognition, structural classification, coding, etc. Recently, there is a trend of combining two or more recognition methods to obtain better recognition rate. Among the above methodologies, structural classification is the most frequently used, with feature extraction as its base decomposed images of handwritten numerals into a set of 15 branching features, which are of the following three categories: (1) straight lines: horizontal, vertical, positive slope, and negative

slope; (2) circles: plain circle and circle on the left, on the right, above, and below; and (3) open arcs: C-like, D-like, A-like, V-like, S-like, and Z-like.

### Density feature extraction method

Shape of individual numerals can be characterized by considering distribution of dark pixels over image regions. Relative densities of the pixels over these regions characterize the shape of the character at reduced resolution. Consequently, minor variations in individual character instances are masked out. But at the same time, stroke width, a determining factor for the density features, is a style-dependent phenomenon. Therefore, density features can be considered to encode style characteristics without being affected by variations in minute details of similarly formed characters. To extract local density feature, a  $3 \times 4$  grid is overlaid on the normalized numeral. The grid boundaries are co-incident with the bounding box of the numeral.

### Density feature for a single element of grid is calculated as:

$d(m, n) = (\text{Number of dark points in the element denoted by } m, n) / (\text{Total number of points in this grid element.})$  In this project we used to template matching technique for feature extraction and then converted the text recognized into speech using Simulink library of MATLAB which is a software that we used to for all image processing and recognition purpose. Below is a sample being tested for effective detection through our project?

HARDWORK Is THE  
KEY To  
sUcCEss j

Figure.7. Text recognized from the image

## IV. FUTURESCOPE

From a future perspective, the present system can be modified so that it is able to read and detect text in more than one language. This will help the people from several regional backgrounds in a great way. Further using neural network training algorithms, the proposed system would be able to detect handwritten characters as well. Also, the process of capturing the image can be automated by using techniques like eye blink detection which involves the use of IR sensors.

## V. REFERENCES:

- [1]. Cheng Lin Liu and Hiromichi Fujisawa, "Classification and learning for character recognition: Comparison of methods and remaining problems".
- [2]. Jangala. Sasi Kiran et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3), 2015, 2065-2069
- [3]. ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 1, January 2015