Automated Book Reader for Blind

Akhil.B, Veneeth Raj, Sushrut Reddy,
S.Metilda Florence
Department of Information Technology
S.R.M University
Chennai, India
akhil161096@gmail.com

Dr. S. Mohan
College of Computer & Information Systems
Al Yamamah University
Kingdom of Saudi Arabia
s.mohan77@gmail.com

***Abstract*— Optical Character Recognition (OCR) is the machine replication of reading and has been the discipline of rigorous research for more than a decade. OCR can be described as the process of converting images of handwritten or printed text into machine encoded text using various mechanical and electronic techniques. It is a method of digitizing text from documents and images so that the text can be electronically searched and used in other processes. OCR plays a major role in various machine processes such as cognitive computing, text mining and machine translation. Further a visually impaired person cannot recognize text from images or scanned documents. This paper presents a simple, low cost and efficient approach to construct OCR for reading printed or handwritten text from a document or image. The proposed system can handle complex backgrounds and multiple patterns, extract text information and convert it into audio output and braille script. The audio output is provided using a text to speech engine. The braille output is provided using a tactile mechanism which is generally understood by the visually impaired. To achieve efficiency and less computational cost OCR is implemented on a Raspberry Pi. The system provides support for English and for regional languages.**

Keywords— Scanned documents, Braille script, text mining, tactile mechanism, complex background, Raspberry Pi

I. INTRODUCTION

The improvements in pattern recognition has escalated in recent times due to the emerging applications which are computationally demanding, apparent in Optical Character Recognition, Computer Vision, Shape Recognition, Data Mining for instance. The field of OCR forms a vital part of document scanners and is used in many software's such as handwriting recognition, banking security etc. The research in this particular field is ongoing for several decades and outcomes have been quite impressive with high recognition rates.

In the modern era a number of organizations are depending on OCR systems to reduce the human influence for improved efficiency and performance. Optical Character Recognition is a system that provides complete alphanumeric recognition of printed and handwritten characters. Documents are scanned using a camera are given to OCR systems which recognize the characters and converts them into ASCII values.

II. OBJECTIVE

In this paper, we propose a multilingual OCR system to recognize and extract printed or handwritten text from scanned documents or images and provide output in the form of audio and Braille tactile feedback. The system can recognize characters of different languages simultaneously. The Braille tactile feedback will aid the deaf in identifying characters

III. PROBLEM STATEMENT

Existing OCR systems are not very robust and work with only good quality images and scanned documents. The accuracy of these systems with handwritten documents is low. These systems have an issue when documents consist of text of different languages and both text and images. There are systems for detecting handwritten script but they do not provide support for regional languages.

IV. LITERATURE REVIEW

An OCR system requires various analyses and processing of the given input image. OCR is closely related to other disciplines such as digital image processing (DIP), pattern recognition (PR). Therefore, OCR systems majorly consist of three tasks. The first and second task is completely associated with the target language from which characters are recognized and extracted. The third task is concerned with the approach of comparing the results of the previous tasks with those of the preceding iterations [9].

An OCR system consists of the several stages such as image acquisition, pre-processing, segmentation, feature extraction, classification, post-processing etc. [1].

The first function is to practice the system to learn the features that exactly describe the characters in the target language. Consequently, the system will have the trivial understanding about the characteristics of the target language. The second function is to extract the features from the scanned documents or images. The third function is to collate the extracted features from the images from the previous task with those learned in the first task [2].

The types involved in the OCR are on-line and off-line. On-line domain involves the text that is typed during the run time, while off-line domain is concerned with both handwritten and

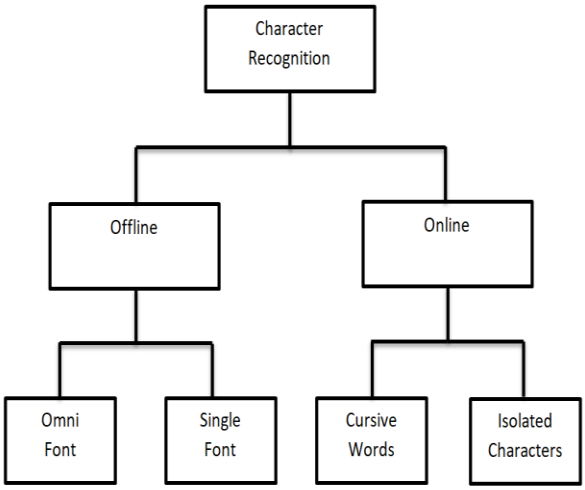

Fig. 1. OCR System Capabilities

printed text that is prepared before the run time. The accuracy of the off-line OCR relies on the number and type of fonts to be recognized [1]. An Omni-font system can recognize most of the fonts without the need to main databases which consist of specific font information. The important character of such OCR systems is feature extraction. However, no OCR system performs well on all the fonts of the contemporary computers. This paper deals with only offline recognition.

OCR systems can be designed to identify block characters or cursive words. The recognition of cursive words is a challenging task because each word could be written in different ways. Suitable segmentation algorithms must be used to segregate the words into characters which are recognized in the later stages [4].

V. Methodology

An OCR algorithm consist of the following stages: image acquisition, pre-processing, segmentation, feature extraction, classification and post-processing

A. Image Acquisition

The initial task of any OCR algorithm, is to capture the scanned documents or images and convert into digital form, there are two ways to acquire input: on-line and off-line. The on-line system provides higher a recognition performance where each character is represented as a vector of points. Off-line systems recognize after it has been completely printed or written on pages.

B. Pre Processing

Additional operation should be performed to enhance the images. These operations include thresholding and noise elimination. Thresholding involves mapping the pixel values of a matrix to either 0 or 1. One common practice is to pre-process the image to compensate the illumination issues. One of the important pre-processing step in this context is the removal of noise. A canny edge detector is applied which produces white pixels wherever there is text or edge in the image. This technique removes most of the noise and converts text into clumps of bright edges. The next step is to differentiate text and border. A rank filter is used for this purpose. The text mostly consists of white pixels but borders comprise of a thin line. The areas around the border will be black so a rank filter can be used to eliminate them. Contours can be used instead of a rank filters.

C. Segmentation

For extraction of features from the image, text in the image should be segmented into lines, words and then characters. The segmentation phase is the most challenging one. This phase determines the accuracy of the system. For this OCR System a novel text segmentation algorithm is used which combines the watershed algorithm and the region based level set method. When using a watershed algorithm one must begin with user-defined markers. The markers can be manually defined or they can be defined using methods such as thresholding. For watershed segmentation the markers should be correctly placed. Region based level set method based on Gaussian kernel function is adapted.

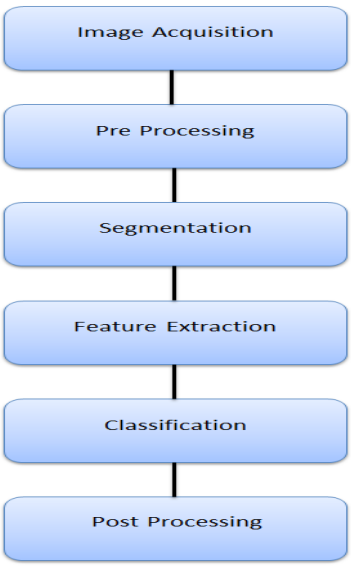

Fig. 2. Different phases in OCR System

D. Feature Extraction

After the image is segmented to various regions, these regions are translated into useful data which express certain parameters. Hence feature extraction involves image to parameter transformation.

E. Classification

This is the decision-making phase. Based on the extracted features, the classifier identifies the pattern that resembles the input features.

F. Post Processing

The primary objective of this phase is to improve the output of recognition process. This phase of the algorithm is system specific.

VI. Implmentation Of Proposed System

A. Software

1) Python: Python is one of the prominent high-level programming languages. Python is a interpreted language. It employs a design notion which underscores code readability and a syntax to specify or express concepts in few lines of code. The important features of Python include dynamic type system and automatic memory management. It also provides support for multiple programming paradigms including functional, object-oriented etc. It includes a comprehensive and large standard library. Python is pre-installed on Raspbian OS. Libraries like NumPy, Scipy needs to be installed which allow effective use of Python. The latest version of Python is 3.6 buy Python 2.6 will be used because of the bindings it provides with OpenCV.

2) OpenCV: OpenCV is released under BSD license and therefore it's available for both commercial and academic use. It has interfaces for C, C++, Java and Python and supports various operating systems like Windows, Linux, and Android. OpenCV was developed for achieving computational efficiency with a strong emphasis on real-time applications. It can take advantage of hardware acceleration of the underlying compute platform. Usage areas range from interactive art to mine inspection, advanced robotics, stitching maps on the web. OpenCV's application areas include 2D and 3D feature toolkits, facial recognition system, gesture recognition etc. OpenCV also includes a statistical machine learning library that consist various algorithms like Decision Tree, Artificial Neural Networks, Support Vector Machine etc. OpenCV have a variety of uses, but the significant ones for implementing OCR are segmentation and recognition features.

3) Text-to-Speech Synthesizer: A text-to-speech system converts language text into speech. These systems are composed of two components: a front-end and a back-end. The crucial task of the front-end is the conversion of text containing alphabets into words; this process is called text normalization or merging. The back-end component then converts the symbolic linguistic into appropriate sound.

B. Hardware

1) Raspberry Pi: The Raspberry Pi is a small credit card sized single-board computer. The hardware of the Raspberry Pi has evolved through various versions. Peripherals such as keyboard or mouse are not included with the Raspberry Pi. The Raspberry Pi is comparatively cheaper when compared to other development boards. The Pi provides easier hardware control when compared to other systems. The Raspberry Pi houses a Ethernet adapter which is internally to an additional USB port. The Pi 3 also comprises of an inbuilt Wi-Fi and Bluetooth module. Because of these improvements, a Raspberry Pi 3 is preferred.

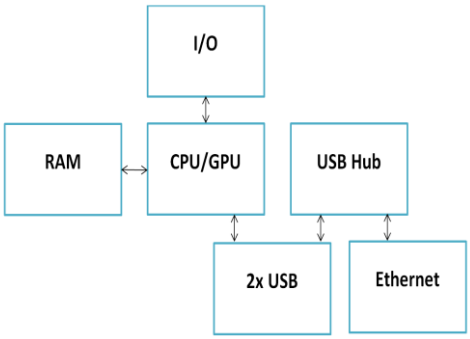

Fig. 3. Systems in Raspberry Pi

2) Camera Module: A camera module consists of an image sensor with a lens, control electronics and interface like Ethernet, CSI or raw low-voltage signaling. A Raspberry Pi 5 MP camera module is used for the project. This sensor is capable of providing 2592 x 1944 pixel static images. The module also supports 1080p30, 720p60 and 480p60/90 video.

3) Audio Amplifier: An audio amplifier is an electronic device that increases the strength of audio signal by increasing the amplitude of the audio signal. The audio amplifier amplifies the low power audio signal to a level which is suitable for earphones. In this system, we use an amplifier to strengthen the audio generated by the text-to-speech system to ensure the audio is audible through headphones/speaker.

4) Braille Terminal: A Braille terminal is a mechanical device for displaying Braille characters. These characters are represented using several round-tipped pins raised through holes on a flat surface. The text which is recognized is represented in Braille script which ensures a better understanding for visually challenged.

VII. Result and Discussion

When the design and implementation phase of the OCR system is completed, the results of the system are ready for assessment. In most OCR systems, the evaluation criteria depend upon the overall recognition percentage and the effect of computation load on the performance of the system. Hence the time taken for each stage of the OCR is calculated to assess the overall performance of the system. The changes in shape of the characters introduced by different fonts under test causes a small amount of miscalculation percentage. Further the feature extraction stage has contributed to the maximum time since it consists of three sub-stages; each having its own computational level [from fig.4.].

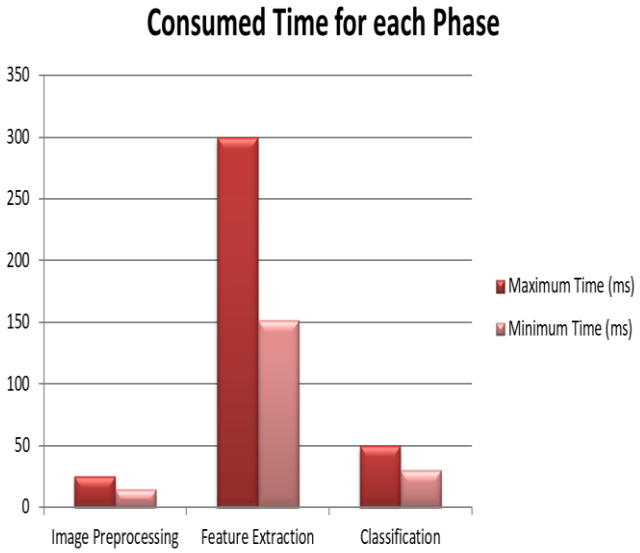

Fig. 4. The consumed time for each phase

The input and the results of the system are shown in Fig.5 and Fig.6 respectively. The pdf document consists of English and other language text. Random characters and digits were generated as output for the text in other language. The proposed algorithm exhibited an accuracy of 75% with printed characters and about 50% with handwritten text.

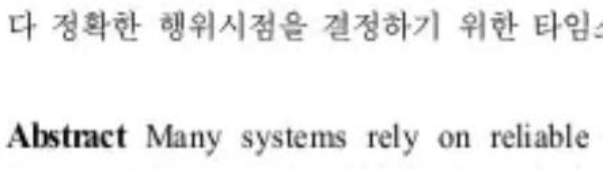

Fig. 5. A screenshot of the image provided as input to the OCR System

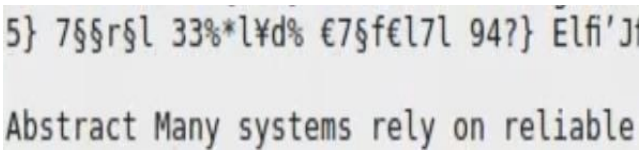

Fig. 6. A screenshot of the text file generated by the OCR System

IX. Conclusion and Future Enhancements

The segmentation algorithm in work exhibits high levels of accuracy in segmenting the text. Some problems were found especially whenever high distortion and noise was introduced in the scanning process. Post-processing stage which is based on the target language will improve the recognition results. Artificial intelligent techniques such as Hidden Markov Model (HMM) could be utilized to obtain better accuracy with handwritten characters.

References

[1] Shalin A.Chopra , Amit A.Ghadge, Onkar A. Padwal, Karan S.Punjabi, Prof. Gandhali S.Gurjar, "Optical Image Recognition", International Journal of Advanced Research in Computer and Communication Engineering,2014

[2] B.Vani ,M..Shyni Beaulah, R.Deepalakshmi, High accuracy optical character recognition algorithms using learning array, 2014

[3] Smart Mobile Software-Mobile OCR. Internet: http://www. smartmobilesoftware. com/mobile-ocr. Html ,April 13, 2013.

[4] Priya Sharma and Randhir Singh, Survey and Classification of Character Recognition System, International Journal of Engineering Trends and Technology-Volume 4 Issue 3-2013.

[5] "A Review on the Various Techniques used for Optical CharacterRecognition", Pranob K Charles, V.Harish, M.Swathi, CH. Deepthi, International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol. 2, Issue 1,Jan-Feb 2012

[6] Y A. Shahab, F. Shafait, and A. Dengel, "ICDAR 2011 robust reading competition: ICDAR Robust Reading Competition Challenge 2: Reading text in scene images," in Proc. Int. Conf. Document Anal. Recognit., 2011, pp. 1491–1496 .

[7] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2011). [Online]. Available: http://www.m.cs.osakafu-u.ac.jp/cbdar2011/

[8] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in Proc. Comput. Vision Pattern Recognit., 2010, pp. 2963–2970.

[9] α-Soft: An English Language OCR", 2010 Second International Conference on Computer Engineering and Applications. Junaid Tariq, Umar Nauman Muhammad Umair Naru.

[10] "Image processing Algorithms for Improved Character Recognition and Components Inspection", 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC 2009), Anima Majumder.