Document Reader System

Atole Swapnali D.[1] Mote Kajal B.[2] Kamble Akash G.[3] Sudrik Rahul D.[4]
[1,2,3,4]SVPMS COE Malegaon BK, India

Abstract— In today's post, we will learn how to recognize text in images using an open source tool called Tesseract and OpenCV. The method of extracting text from images is also called Optical Character Recognition (OCR) or sometimes simply text recognition. In this paper an assistive system has been proposed which is useful for visually impaired or also normal person. It is the system which reads texual information present on papers and produce corresponding voice using OCR(Optical Character Recognition)and TTS(Text-To-Speech) system. Optical Character Recognition (OCR) is a system that provides a full alphanumeric recognition of printed or handwritten characters by simply scanning the text image. OCR system interprets the printed or handwritten characters image and converts it into corresponding editable text document. The text image is divided into regions by isolating each line, then individual characters with spaces. After character extraction, the texture and topological features like corner points, features of different regions, ratio of character area and convex area of all characters of text image are calculated. Previously features of each uppercase and lowercase letter, digit, and symbols are stored as a template. Based on the texture and topological features, the system recognizes the exact character using feature matching between the extracted character and the template of all characters as a measure of similarity.

Keywords: Tresaract, Optical Character Recognition, Feature Extraction (OCR), Text To Speech (TTS), Feature Matching, Text Extraction, Character Extraction

I. INTRODUCTION

An estimated 253 million people live with vision impairment: 36 million are blind and 217 million have moderate to severe vision impairment. However, it is Braille; thus, the library of a visually impaired person is estimated that the number of people with vision impairment could triple due to population growth and ageing. For example, by 2050 there could be 115 million people who are blind, up from 38.5 million in 2020. So, the rate of visually impaired people is increasing in huge margin each year, so they face difficulties in their life, one of the problems is reading. The current day scenario of reading for blind people is with the help of braille. Braille is a code-a system of dots that represent letters of an alphabet. All books are not written in Braille; thus, the library of a visually impaired person is limited to countable number of books a in limited to countable number of books.

Accessing text documents is troublesome for visually impaired (VI) people in many scenarios, such as reading text on the go and accessing text in less than ideal conditions (i.e. low lighting, columned text, unique page orientations, etc.) Interviews we conducted with VI users revealed that available technologies, such as screen readers, desktop scanners, smartphone applications, eBook readers, and embossers, are commonly under-utilized due to slow processing speeds or poor accuracy. Technological barriers inhibit VI people's abilities to gain more independence, a characteristic widely identified as important by our interviewees. In this project, we present our work towards creating a document reader device that could overcome some issues that current technologies pose to VI users. The contribution is twofold: First, we present results of focus group sessions with VI users that uncovered salient problems with current text reading solutions and the user's characters of future assistive devices and their capabilities. The results serve as grounds for our design choices. Second, we present the concept of extra acting the text from the text image captured through the camera mounted on it. This extracted text then processed by the text-to-speech to generate the audible speech. The digital document reader is a device that enables a more manageable eyes free operation. It includes methods of text-reading for the visually impaired based on key features: adaptation for non-perfect imaging, type of text, User Interface suitable for VI and the evaluation method.

II. PROPOSED SYSTEM

The proposed system can be broadly divided into two modules: OCR module and TTS module. Optical character recognition (OCR) takes a text image as input and gives editable text document as output. The OCR system primarily involves four steps: Preprocessing, Features extraction, Features training, and Feature matching. Flowchart of the OCR is shown in Figure 1. Here, two data sets are considered, one for training dataset and another for test dataset. Preprocessing and feature extraction is done in both cases. Features extracted from test data is compared with features extracted from training data to get the desired output.

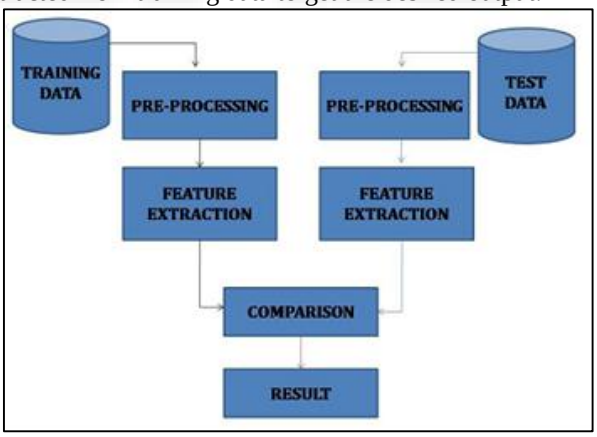

Fig. 1: Flowchart of OCR system

Feature extraction technique is applied for all individual extracted characters. The character image is divided into four regions. Sum of the pixels value of the whole image and sum of pixels value in each of the sub-regions are calculated. Then their ratios are calculated as the features value of f1, f2, f3, f4 respectively.

f1= Sum of the pixels value of 1st quadrant /
Sum of the pixels value of the whole image
f2= Sum of the pixels value of 2nd quadrant /
Sum of the pixels value of the whole image
f3= Sum of the pixels value of 3rd quadrant /

Sum of the pixels value of the whole image
f4= Sum of the pixels value of 4th quadrant /
Sum of the pixels value of the whole image
To get better accuracy, features f5, f6, f7, f8, f9, and f10 are calculated using f1, f2, f3, and f4.
f5=f1+f2
f6=f2+f3
f7=f3+f4
f8=f1+f4
f9=f2+f4
f10=f1+f3

Using Harris corner method, numbers of corner points are calculated from character image. Feature f11 is considered as the number of corner points of a character. Total area of extracted character image is calculated using the actual number of pixels in the character image. Convex area of the character is calculated using the number of pixels in convex hull that can contain the character region. Feature f12 is ratio of convex area to total area.
f12= Convex Area / Total Area

Total twelve features f1 to f12 are extracted for all individual extracted characters.

III. SYSTEM ARCHITECTURE

This system combines the concept of Optical Character Recognition and Text To Speech synthesizer. In this system the first stage involves that the auto-focus camera which capture the image of printed text and converting it into text document using optical character recognition. The second stage involves Natural Language Processing(NLP) and Digital Signal Processing(DSP) for converting the text into speech using text-to-speech synthesizer.

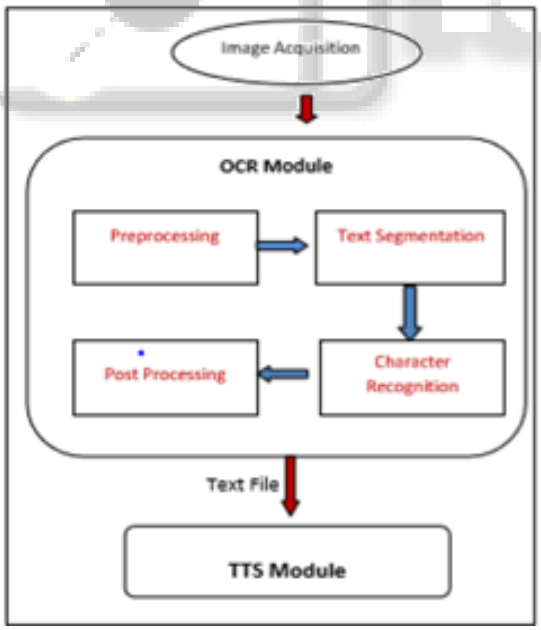

Fig. 2: System Architecture

The proposed system can be broadly divided into two modules: OCR module and TTS module as shown in figure. Image of printed text is captured using auto-focus camera. Then image preprocessing is done on the image to reduce the noise and skew present in image. The image is binarize and segmented into text and non text regions. Individual characters are isolated and normalized in order to facilitate feature extraction process and also improve their classification accuracy. Post-processing is done to group the various characters together so as to formed meaningful word and numbers. Then text file is given as input to the TTS module which convert the text file to speech signal.

IV. ALGORITHM

1) STEP 1: The input text image is converted into binary image.
2) STEP 2: The binary image is complimented so that the letters constitute by binary 1 (one) and background constitute by binary 0 (zero).
3) STEP 3: All text lines are separated from the binary image. This is done by finding the sum of all values in a row. When the sum is 0, a new line is identified and separation is done. The sum of all rows in between two lines should be zero.
4) STEP 4: For each line, the characters are to be extracted. This is done by finding the sum of all pixels value in a column. When sum is zero, a new character is identified and separation is done.
5) STEP 5: Total 12 features value f1 to f12 are extracted for each character.
6) STEP 6: The features value are matched with the trained features set to recognize the exact character.
7) STEP 7: Then the output of OCR module is given to the TTS module which produces the sound from recognized text.

V. RESULTS

In this paper we have presented a system for selection and separation of text and its conversion to speech from scanned document image to build an assistive system aiming to support people with learning disability. It works well for images having variations in textfonts, colors, sizes as well as the low contrast between the text and often complicated background.

Fig. 3: Project Image with Raspberry Pi and Camera

VI. CONCLUSION

In this paper we have presented a system which consists both OCR and TTS module. We have presented a scheme for selection and separation of text and its conversion to speech from a scanned document image to build an assistive system aiming to support people with visual impairment.

REFERENCES

[1] V. Wu, R. Manmatha and E. M. Riseman, "Finding Text in Images", In Proc. of Second ACM International Conference on Digital Libraries, Philadelphia, PA, pp. 23-26, 1997.
[2] H. Li and D. Doermann, "Automatic Text Tracking In Digital Videos", In Proc. of IEEE 1998 Workshop on Multimedia Signal Processing, Redondo Beach, California, USA, pp. 21-26, 1998.
[3] K. Jain and B. Yu, "Automatic Text Location in Images and Video Frames", In Proc. of International Conference of Pattern Recognition (ICPR), Brisbane, pp. 1497-1499, 1998.
[4] P. K. Kim, "Automatic Text Location in Complex Color Images Using Local Color Quantization", IEEE TENCON, Vol. 1, pp. 629-632, 1999.
[5] L. Agnihotri and N. Dimitrova, "Text Detection for Video Analysis", In Proc. of the International Conference on Multimedia Computing and Systems, Florence, Italy, pp. 109-113, 1999.
[6] Garcia and X. Apostolidis, "Text Detection and Segmentation in Complex Color Images", In Proc. Of International Conference on Acoustics, Speech and Signal Processing (ICASSP2000), Istanbul, Vol. 4, pp. 2326- 2330, 2000.
[7] W. Pan, J. Jin, G. Shi, Q. R. Wang, "A System for Automatic Chinese Business Card Recognition", ICDAR, pp. 577-581, 2001.
[8] M. Cai, J. Song and M. R. Lyu, "A New Approach for Video Text Detection", In Proc. of International Conference On Image Processing, Rochester, New York, USA, pp. 117-120, 2002.
[9] X. Luo, J. Li and L. Zhen, "Design and implementation of a card reader based on build-in camera", International Conference on Pattern Recognition, pp. 417-420, 2004.
[10] X. Luo, L. Zhen, G. Peng, J. Li and B. Xiao, "Camera based mixed-lingual card reader for mobile device", International Conference on Document Analysis and Recognition, pp. 665-669, 2005.
[11] M. Koga, R. Mine, T. Kameyama, T. Takahashi, M. Yamazaki and T. Yamaguchi, "Camera-based Kanji OCR for Mobile-phones: Practical Issues", Proceedings of the Eighth International Conference on Document Analysis and Recognition, pp. 635-639, 2005.
[12] K. S. Bae, K. K. Kim, Y. G. Chung and W. P. Yu, "Character Recognition System for Cellular Phone with Camera", Proceedings of the 29th Annual International Computer Software and Applications Conference, vol. 1, pp. 539-544, 2005.
[13] M. Laine and O. S. Nevalainen, "A standalone OCR system for mobile camera-phones", Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium, pp.1-5, Sept. 2006.
[14] H. Shen and J. Coughlan, "Reading LCD/LED Displays with a Camera Cell Phone", Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, 2006.
[15] F. Mollah, S. Basu, M. Nasipuri and D. K. Basu, "Text/Graphics Separation for Business Card Images for Mobile Devices", Proc. of the Eighth IAPR International Workshop on Graphics Recognition (GREC`09), pp. 263-270, July, 2009, France.
[16] F. Mollah, S. Basu, N. Das, R. Sarkar, M. Nasipuri, M. Kundu, "A Fast Skew Correction Technique for Camera Captured Business Card Images", Proc. of IEEE INDICON-2009, pp. 629-632, 18-20 December, Gandhinagar, Gujrat.
[17] F. Mollah, S. Basu, M. Nasipuri, "Segmentation of Camera Captured Business Card Images for Mobile Devices", International Journal of Computer Science and Applications, 1(1), pp. 33-37, June 2010.

