Character Detection and Recognition System for Visually Impaired People

Akhilesh A. Panchal, Shrugal Varde, M.S. Panse

Abstract— **Nowadays, increasing use of digital technology, availability of economical image capturing devices like mobile phones, digital cameras, etc. and need of powerful technology to aid blind or visually impaired people, attracting researchers to the problem of recognizing text in images. Detecting text from scene image is more difficult as compared to that from printed documents. Lots of research has been done on detecting scene text to overcome certain challenges like perspective distortion, aspect ratio, font size, etc. Speed, complexity, cost and accuracy are important parameters must be taken into consideration while designing such systems. Computer vision is one of the emerging technologies that can be used to aid visually impaired people for navigation (both indoor and outdoor), accessing printed material, etc. This paper describes an approach to extract and recognize text from scene images effectively using computer vision technology and to convert recognized text into speech so that it can be incorporated with hardware to develop Electronic travel aid for visually impaired people in future.**

Keywords— Computer Vision, Image Acquisition, Image Enhancement, Character detection and extraction, Pattern recognition, Text to Speech conversion.

I. INTRODUCTION

Visual / vision impairment is a decreased ability of visual system that is not fixable by usual means, such as glasses or contact lenses. It is used to describe any vision loss whether it is total or partial blindness. It affects almost every daily activity like walking, driving, reading and recognizing objects, places. Every year, the number of visually impaired (VI) people is increasing due to diabetic retinopathy, age related macular degeneration, corneal clouding, childhood blindness, number of eye infections, traffic accidents and other causes. The VI community is very diverse in terms of degree of vision loss, age, and abilities [1]. Research in assistive technology for VI people drives the development of useful hardware and software tools. To satisfy the need of VI community, it is necessary to develop device with smart algorithm to extract information from surroundings. Computer vision and mobile computing are powerful tools with great potential to enable a range of assistive technologies to aid growing population of visually impaired people. Development in this field duplicates the abilities of human vision by electronically perceiving and understanding an image. Image processing, image analysis and machine vision are most closely related fields to computer vision i.e. the basic techniques that are used and developed in these fields

Akhilesh A. Panchal, Department of Electrical Engineering , Veermata Jijabai Technological Institute, Mumbai, India, akhileshp31@gmail.com)

Shrugal Varde, Department of Electrical Engineering , Veermata Jijabai Technological Institute, Mumbai, India, shrugalv@gmail.com)

M.S. Panse, Department of Electrical Engineering , Veermata Jijabai Technological Institute Mumbai, India

are more or less identical. Combining techniques in these fields can be applied in various applications like Artificial Intelligence (AI), Pattern/Object Recognition, Imaging, Machine Learning, Signal Processing, Neurobiology, etc. Among these applications, pattern recognition and image processing can be used to aid VI people in daily tasks by extracting information from outside environment. This article focuses mainly on the use of computer vision systems and image processing techniques to support VI people in terms of mobility, wayfinding and printed information access [1]. Text data present in images plays a vital role to understand them. This textual information exists everywhere like in Books, Milestones, Shop names, street boards, etc. Taking this into consideration, this paper attempts to detect and recognize text from surrounding & converting into speech to guide VI people in day-to-day activities.

The remainder of this paper is organized as follows. Overview of Literature Survey is presented in Section II. Section III describes System block diagram. Section IV presents experimental results and discussion and Section V discusses conclusion and future scope of the paper.

II. LITERATURE SURVEY

In order to read text from image, it is essential to know the properties of image as well as text in it. Text can be of two types, Artificial/Graphic text and Scene Text, depending on source image whether it is artificial i.e. machine generated or real scene image captured from camera [2]. Fig. 1 shows the examples of artificial text and Scene text.

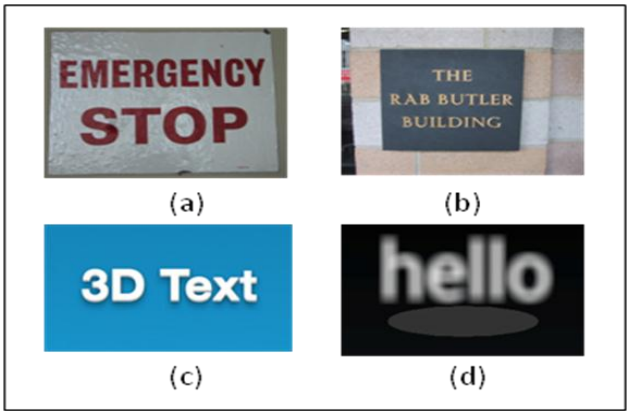


Fig.1: (a), (b) Scene text & (c), (d) Artificial text

From earlier study, it is found that text extraction from artificial or document image is easier than Scene image because Scene image may be mixed with noise and blur. Scene text present more difficult challenges such as uneven lighting, Blurring, Degradation due to noise, font style and size, multilingual environment of image giving rise to false text detection and misses unlike artificial text [2]. Hence, reading text from scene images is a challenging task and is still an active research today [3]. In case of visually

978-1-5090-0774-5/16/$31.00 © 2016 IEEE

impaired person, it becomes worse because it cannot be assumed that the acquired image contains only characters. So, system must find out first whether text is present or not in an image and then detect & extract text from it which is followed by recognition.

Text detection is a preliminary step in automatic text recognition. Many image processing techniques have been proposed based on edge detection, binarization, spatial-frequency image analysis, mathematical morphology operations, etc. [4]. Generally, text detection methods can be classified as either edge-based, connected-component based and texture-based methods [5]. Out of which, Edge based approaches use different masks like Sobel, Roberts, Prewitt, Canny to detect edges from image. Hough transform is widely used technique for Edge detection. This approach requires text to have a reasonably high contrast with respect to the background in order to detect the edges. So these methods encounter problems with complex backgrounds and produce many false positives resulting less accuracy to satisfy the requirement of our system to aid VI people where high accuracy is desired. It will be effective only when characters are surrounded by connected edges and the inner ink area is not broken. Texture based approach is based on variation of gray levels in pixels of image. As there are different texture types like Smooth, Grainy or Coarse, text extraction becomes more complex using this approach. Texture features including Fourier Transform, Discrete Cosine Transform, Wavelet and HOG have been used to localize text. Though texture features are effective for detecting dense characters, they may not detect sparse characters, i.e., signs in scene images which lack significant texture properties [2]. Connected Component based approach uses graph algorithm, where text is considered as combination of connected components. It is uniquely labeled based on heuristics about features, i.e., color similarity and spatial layout. This approach does not work well for all video images because it assumes that text pixels in the same region have similar colors or grayscale intensities [6]. But it is easy and fast.

As text detection is of fundamental importance in image understanding and content based retrieval, all these techniques have to be used in combination to design and develop algorithm for our project requirement.

III. System block diagram

As shown in fig. 2, our system is primarily divided into three stages: 1.Acquisition 2.Processing 3.Text to Speech conversion (T2S).

1. Acquisition stage:

It acquires high resolution video from camera. This video will be broken into different frames. Each frame will act as separate image. In this paper, standard images from ICDAR competition are used for testing purpose. In sophisticated system, video from camera is used as input to image processing unit. The challenges occurred due to camera fixation/ position problems like blurring/degradation due to motion of user, perspective distortion due to different angles of the object formed with camera lens axis. So, care must be taken while fixing camera of the system. Acquisition must be proper to produce desired output.

2. Image Processing:

Second stage is Image Processing unit which is a heart of the system. It is subdivided into three secondary stages: A. Pre-Processing, B. Processing C. Post-Processing.

A. Pre-Processing:

As acquired color image consist of 3 planes (Red, Green, and Blue); it is difficult to process it in quick time. So, it is first converted into Grayscale image. After color to grayscale conversion, pre-processing stage uses some enhancement techniques to eliminate challenges created by noise, blurring effect and uneven lighting. It is considered that acquired image might be mixed with noise like Salt and pepper noise, Impulse noise etc. or it can be blurred due to motion of camera. Salt and pepper noise can corrupt the image, where the noisy pixels can take only the maximum and minimum values in the dynamic range i.e. black dot on white background (pepper) and white dot on black background (salt) which degrades the text extraction performance of system [7]. To remove such type of noise, standard median filter (SMF), which is a non-linear filter used due to its good de-noising power and computational efficiency. Removal of this type of noise in a system is shown in Figure 3. However, when noise level is more than 50%, edge details will not be preserved by the median filter. So, it is recommended filtering process should preserve the edge details without losing the high frequency components of the image edges.

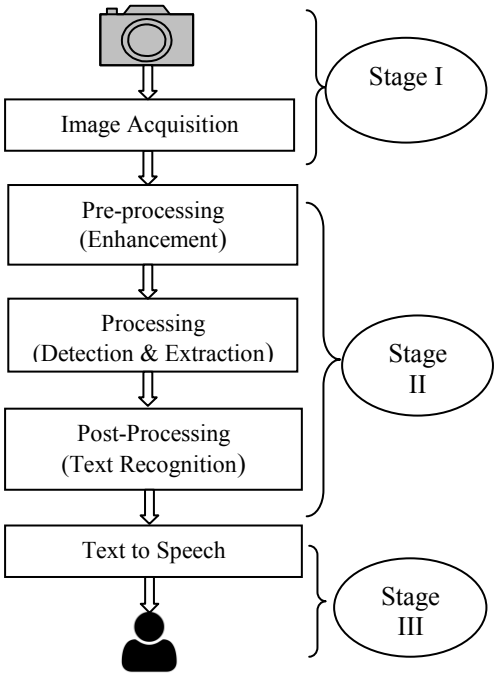


Fig. 2: General block diagram of system

Fig. 3: Salt and Pepper noise removal

Fig. 4: Contrast Enhancement

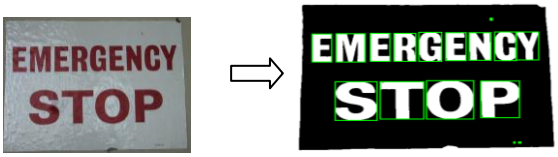
(a) (b) (c)

Fig. 5: De-blurring of an image (a) Blurred image; (b) Binarized image without filtering; (c) Binarization after De-blurring

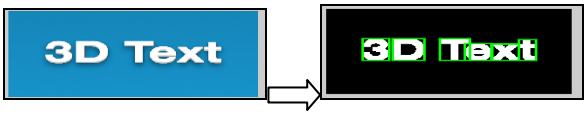
Sometimes, text extraction becomes troublesome for image captured in dark or uneven lighting. So, application of contrast enhancement is necessary. Histogram Equalization method is used for Contrast enhancement. Figure 4 shows enhanced image using histogram equalization. Another challenge i.e. Blurring can be generally removed using de-blurring techniques like Lucy Richardson algorithm, Blind de-convolution algorithm, Wiener de blurring techniques [8]. Wiener filter is selected which is a natural extension of the inverse filter when noises are present. Figure 5 shows how de-blurring using Wiener filter is effective on text embedded blurred image. From figure, it is observed that binarization after wiener filtering on blurred image produces better result than without applying wiener filtering.

B. Processing:

Enhanced/ Pre-processed image from previous block is forwarded to Processing stage where text detection and extraction is done. Before processing, it is binarized with adaptive thresholding. As a result of literature survey, we used combination of connected component (CC) & region based approach on this Binarized (black and white) image. Applying CC analysis using MATLAB software, areas having text similar patterns with white pixels on dark background, are detected. Using feature extraction algorithm, these detected areas are extracted on separate windows. Figure 6 shows text detection and extraction from Scene as well as Document image. Scene image requires preprocessing whereas Binarized Document image is directly fed to Text detection block. From figure 6, it is observed that false detections are present due to some amount of noise in scene image including the Real text of image. This false detection reduces accuracy of the system. It also increases processing time and hence, decreases speed of the algorithm.



(a) Scene image



(b) Artificial image

Fig. 6: Text Detection and Extraction

C. Post-Processing:

Next step is to recognize text from detected & extracted patterns in previous stage. From previous study, it is known that Recognition requires more sophisticated algorithm. Many approaches have been developed to recognize text from an image such as linear discriminate analysis (LDA) for single character recognition, Support vector machine(SVM) & Conditional random field (CRF) [9] for distorted/ multiple character recognition, Stroke Width Transform(SWT) for recognizing text appearing in different angles to overcome perspective distortion problem [10], etc. Feature learning is one of statistical pattern recognition algorithms are widely applied to computer vision problems and many results obtained with feature learning systems show high performance in recognition tasks [3]. Also this algorithm achieves better results compared to other recognition methods. Hence, Feature Learning method is used for Text Recognition. Feature vector is the simplest form of knowledge representation. For character recognition, Feature vector are number of strokes, number of loops, width to height ratio of character. Figure 7 shows architecture of the recognition system.

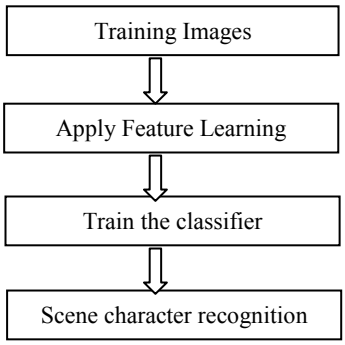


Figure 7: Feature learning Recognition system

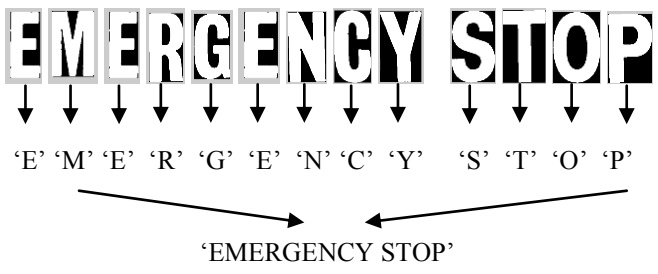


Figure 8: Text Recognition system output

This recognition algorithm uses above mentioned parameters to train the classifier. Each character has certain feature vector which is classified in training. So, this trained data is already stored in database. Each extracted character coming from previous block passes through this classifier where it extracts feature vectors. By matching feature vectors with available trained data, each character gets recognized and stored in a file. Using CRF model, these recognized characters joined together to form word. So, new database will be collection of recognized words. This block also removes false detections obtained from previous block. Hence, it improves accuracy of the system.

Figure 8 shows output of text recognition system. It consists of two steps: Character recognition followed by Word formation. This combination of algorithms is easy to develop and accurate.

3. Text to Speech (T2S) conversion

This block converts the recognize texts into speech. Output is voice which enables blind person to listen to scene instead of read & act accordingly. As there is lots of research has been done on T2S system, many algorithm and softwares are available for it. LabVIEW software is used to design a VI which converts text from a file to Speech using inbuilt speech synthesizer function. Speech Synthesizer Nodes like Property, invoke etc. from .NET object on LabVIEW used for T2S. As it uses Graphical programming, it is simple to design a code. Hence, while designing prototype of aid for VI people, it is a good choice to use LabVIEW software. Figure 9 shows block diagram of LabVIEW code generated for T2S conversion.

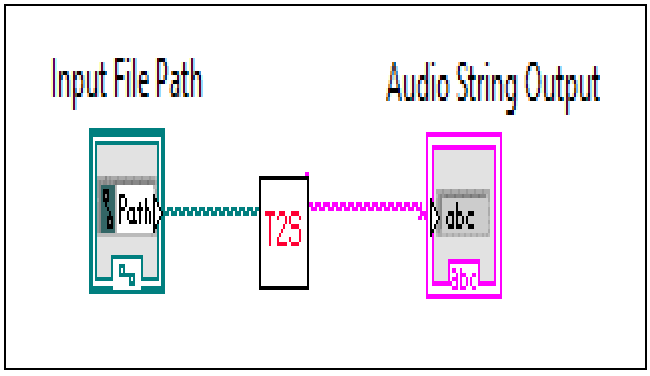
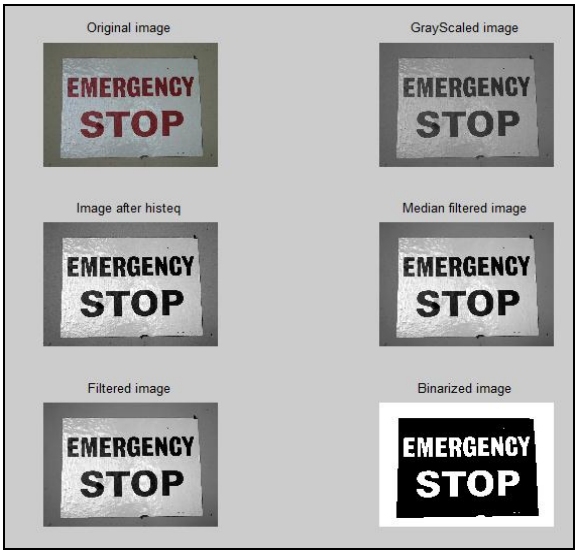


Fig. 9: Block diagram of T2S converter

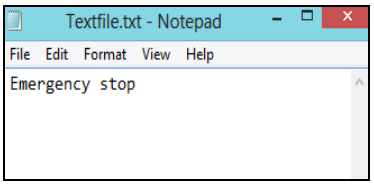
IV. Experimental Results

All the modules mentioned in system block diagram are joined together to form a complete system to aid VI people. Figure 10 shown below depicts the output of each block. First block figure 10(a) represents output of preprocessing stage. After preprocessing, text is detected and extracted from image shown in second block as shown in figure 10(b). Pattern recognition algorithm developed on MATLAB software recognizes text from extracted text and save it in a text file (Figure 10(c)). Figure 10(d) shows front panel of LabVIEW where T2S conversion is performed to produce speech output. This system works extremely well for artificial text in image as well as produces decent results for Scene text. From the result obtained it is found that this system has produces good accuracy & it is fast.

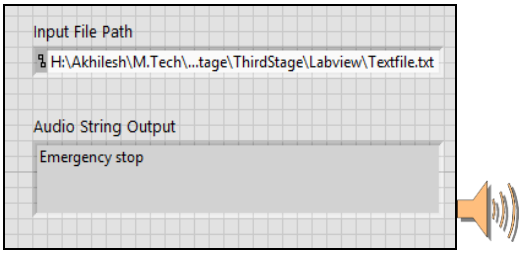


(a) Pre-Processing

(b)Text detection and Extraction



(c) Recognized text saved in text file



(d) Front panel of LabVIEW

Fig. 10: Result of Overall system

V. Conclusion And Future Scope

In this paper, design and implementation of prototype of text detection and recognition system has been discussed. After combining different techniques for Text detection and extraction, it is found that system works faster and better than using single technique for overall system. Text detection followed by recognition using supervised pattern recognition algorithm not only improves accuracy but also increases speed of the system. After successful recognition, text is converted into audio output. Using this system, Travel aid can be developed to assist visually impaired people during navigation in known or unknown, indoor or outdoor environments.

REFERENCES

[1] Roberto Manduchi and James Coughlan, "Computer Vision Without Sight",*Communications of the ACM*, pp. 96-104,Vol.55, no.1,Jan.2012.

[2] Qixiang Ye and David Doermann, "Text Detection and Recognition in Imagery: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.1480-1500, Vol.37, No.7, July 2015.

[3] Ho Vu, Duong and Quoc Ngoc, Ly, "A Feature Learning Method for Scene Text Recognition", *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT),* pp. 176 - 180, Ho Chi Minh City, 2012.

[4] Nobuo Ezaki, Marius Bulacu, and Lambert Schomaker ,"Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons", *17th Int. Conf. on Pattern Recognition (ICPR 2004), IEEE Computer Society*, pp. 683-686, vol. II, , Cambridge, UK, August 2004.

[5] Andrej Ikica, Peter Peer, "An improved edge profile based method for text detection in images of natural scenes", *IEEE EUROCON-International Conference on Computer as a Tool*, pp.1-4, Lisbon, 2011.

[6] M Sharmila Kumari, Akshatha, "Local Features based Text Detection Techniques in Document Images," *International Conference on Information and Communication Technologies (ICICT),* pp. 6-11, 2014.

[7] Madhu S. Nair, K. Revathy, and Rao Tatavarti, "Removal of Salt-and Pepper Noise in Images: A New Decision-Based Algorithm", *Proceedings of the International Multi-Conference of Engineers and Computer Scientists IMECS*, Volume I, Hong Kong, March 2008.

[8] Sonia George, Noopa Jagdeesh, "A Survey on Text Detection and Recognition from Blurred Images", *International Journal of Advanced Research Trends in Engineering and Technology(IJARTET)*, Vol. II, Special Issue X, pp. 1180-1184, March 2015.

[9] Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu, "Text Localization in Natural Scene Images based on Conditional Random Field", *10th International Conference on Document Analysis and Recognition*, pp.6-10, 2009.

[10] Linlin Li, Chew Lim Tan ,"Character Recognition under Severe Perspective Distortion", *IEEE 19th International Conference on Pattern Recognition*, pp 1-4, Dec. 2008.

[11] R.C.Gonzalez, R.E.Woods, S.L.Eddins, 'Digital Image Processing using MATLAB', *Pearson Education India*,2009.