Text Extraction from Natural Images Using Maximally Stable Extremal Regions

[1] yash chugh, [2]Ranjeeta kaushik
[1] M.Tech student, [2]Associate professor
[1] Department of Computer Science
[1]N.C College of Engineering and Technology, panipat , India

***Abstract*- Detecting text in natural images is an important prerequisite. A large number of techniques have been proposed to address this problem and the purpose of this paper is to classify and review these techniques. In this paper a novel text detection algorithm as connected component based text detection algorithm is used for detection and recognition of text from the natural images using edge enhanced Maximally Stable Extremal Regions and stroke Width transform. Basic idea is to extract the connected component with MSER algorithms and then these candidates are filtered using CC based analysis and stroke width filtering. The text detected is then recognized using optical character recognizer.**

***Index Terms*- text detection, scene text, MSER, connected-component, stroke, OCR**

I. INTRODUCTION

Now a days the most information is available either on paper or in the form of photographs or videos. Large information is stored in images. The existing technology is restricted to extracting text against spotless backgrounds. Thus, there is a requirement for a system to extract text from general backgrounds. There are various applications in which text extraction is useful. The role of text detection is to find the image regions containing only text that can be directly highlighted to the user or feed into an optical character reader module for recognition. TEXT in images contains most important information and is browbeaten in many content-based image[1] and video applications, such as video information retrieval, content-based web image search and information retrieval systems in indoor and outdoor environments, visually impaired people, mobile based text analysis and recognition, and automatic robot navigation.Various scene text recognition and detection have received a large amount consideration for the last decades.

Among them, text detection and recognition in camera based images have been considered as very important problems in computer vision community [2]. It is because the text data is easily recognized by machines and can be used in a variety of applications. Due to multifarious background, variations of size, font, orientation and color. Text detection in natural scene images has to be robustly detected before being recognized and retrieved. As an essential prerequisite for text-based image search, text within images has to be robustly located. However, Text detection has been considered in many recent studies and numerous methods are reported in the literature [3]

II. Text In Images

Content-based image indexing refers to the process of attaching labels to images based on their content. Image content can be divided into two main categories: *semantic content* and *perceptual content* [4]. A number of studies on the use of relatively low-level perceptual content [5] for image and video indexing have already been reported Perceptual content includes attributes such as color, texture, shape, intensity and their temporal changes, whereas semantic content means events, objects, and their relation between them. Studies on semantic image content in the form of vehicle, text, face, and human action have also attracted some recent interest [6].

It is well known that scene text is more difficult to detect and very little work has been done in this area. In contrast to caption text, scene text can have any orientation and may be damaged by the point of view projection. Moreover, it is often affected by variations in scene and camera parameters such as illumination, focus, motion, etc. Page layout analysis usually deals with document images Readers may refer to papers on document segmentation analysis [7]for more examples of document images. Although images acquired by CD covers, scanning book covers or other multi-colored documents have similar characteristics as the document images.

Figure 1: Document Images

Figure 2: Scene Text

III. Properties Of Text In Images

Text in images can exhibit many variations with respect to the following properties:

1. Color: The characters in a text line tend to have the same or similar colors. This property makes it possible to use a connected component-based approach for text detection. Most of the research reported till date has concentrated on finding 'text strings of a single color (monochrome). However, video images and other complex color documents can contain 'text strings with more than two colors (polychrome) for effective visualization, i.e., different colors within one word.
2. Geometry:
   - Size: Although the text size can vary a lot, assumptions can be made depending on the application domain.
   - Inter-character distance: characters in a text line have a uniform distance between them.
   - Alignment: The characters in the caption text appear in clusters and usually lie horizontally, although sometimes they can appear as non-planar texts as a result of special effects. This does not apply to scene text, which can have various perspective distortions. Scene text can be aligned in any direction and can have geometric distortions
3. Edge: Most caption and scene texts are designed to be easily read, thereby resulting in strong edges at the boundaries of text and background.
4. Motion: The same characters usually exist in consecutive frames in a video with or without movement. This property is used in text tracking and enhancement. Caption text usually moves in a uniform way: horizontally or vertically. Scene text can have arbitrary motion due to camera or object movement.
5. Compression: Many digital images are recorded, transferred, and processed in a compressed format. Thus, a faster TIE system can be achieved if one can extract text without decompression.

IV. RELATED WORK

Automatic detection and recognition of text in images done using different techniques proposed. Generally, Text detection mechanism can be divided into two categories: connected component (CC)-based methods and Edge-based methods. Edge-based methods [8] approved a sliding window scheme, which is basically a brute force scheme which requires a lot of local decisions. Therefore, the region-based methods have attentive on an efficient binary classification (text versus non-text) of a small image area. In other words, they have focused on to determine whether a given patch is a part of a text region or not. Limitations of Edge-based methods are high computational complexity and the difficulty to select the best features for scene text detection.
On the other hand CC- based methods are simple and efficient text detection approach. Connected component methods generated separate CC regions. Connected component based methods use a bottom-up approach by grouping small components into sequentially larger components until all regions are identified in the image. A geometrical analysis is needed to join the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions in the images.

V. METHODOLOGIES

The input image to the proposed system has a complex background with text built in it. The first stage is Image Pre -processing, which serves to remove the noise from the input image and generates a clear binary image. Text segmentation is the next stage, where we differentiate each character from the entire word by circumscribing them into boxes and saving them each separately. The final stage is text recognition, where the segmented characters [9] are compared to the stored character matrices and as a result, the closest match for each character is displayed as a separate picture.

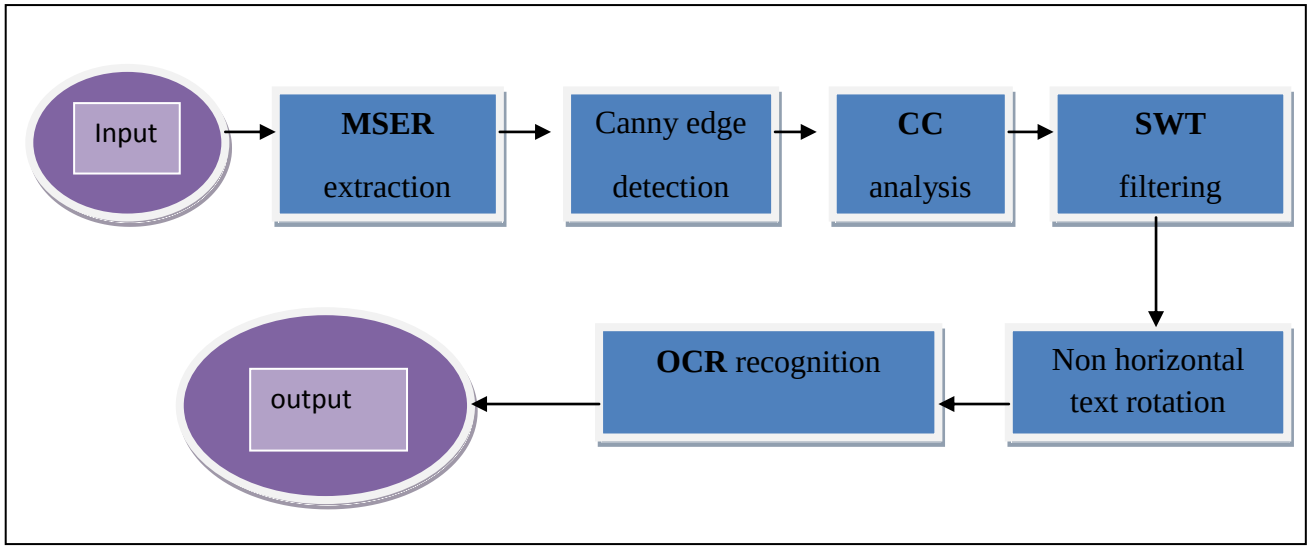

Figure 3: flowchart of our text detection algorithm

The flowchart of our text detection algorithm is shown in Fig.3. At the input of the system, the image intensities are linearly adjusted to enhance the contrast. Subsequently, MSER regions are efficiently extracted from the image [10] and enhanced using Canny edges obtained from the original gray-scale image As a next step, the resulting CCs are filtered using geometric constraints on properties like aspect ratio and number of holes. The stroke width information is robustly computed using a distance transform and objects with high variation in stroke width are rejected. Text candidates are grouped pair wise and form text lines. Finally, words within a text line are separated, giving segmented word patches at the output of our system.

In the method propose two level text/ nontext filtering using a novel image operator stroke width transform and cc analysis. The proposed method is work on non-horizontal text detection by rotating the text to make it as a horizontal text. Then the text regions are localized using parallelograms. Detected text regions are recognized using OCR recognizer [11]. For the generation of candidates, we extract CCs in images. This step includes Detect MSER region, canny edge detection and CC analysis.

Maximally Stable Extremal Regions (MSER
)[12] have become one of the commonly used region detector because of their high repeatability and partly because they are somewhat complementary to many other commonly used detectors . only the MSER algorithm could provide the stable binary results and also help us find most of the text components. And then as shown in above image the next step is to use the canny edge operator as Edge detection is the name for a set of mathematical methods which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. And since written text is typically placed on clear background, it tends to produce high response to edge detection image the Stroke width is defined as the length of a straight line from a text edge pixel to another along its gradient direction. The basic motivation of our stroke width extraction algorithm is that stroke width almost remains the same in a single character; however, there is significant change in stroke width in non-text regions as a result of their irregularity. And last sep is to use the optical character recognition in which the segmentation of text from a muddled scene can greatly improve OCR results. Since the algorithm already produced a well segmented text region, we can use the binary text mask to improve the accuracy of the recognition results.

Input image

Output image after text extraction

Figure 4: result of the experiment

VI. CONCLUSION

In this proposed system a new text detection method that effectively detects and recognizes texts of different directions in natural images. In this paper CC-based text detection algorithm is proposed to overcome the difficulties of grouping the characters and remove false positives of text in images. The system is stable and robust. All the system parameters remain the same throughout all the experiments. For future work we will focus on learning based methods for text extraction from complex backgrounds and

text normalization for OCR recognition. We also attempt to improve the efficiency and transplant the algorithms into a navigation system prepared for the way finding of visually impaired people.

REFERENCES

[1] J. Ohya, A. Shio, and S. Akamatsu, Recognizing Characters in Scene Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (2) (1994) 214-224.
[2] K. Jung, "Text information extraction in images and video: A survey," Pattern Recognition, vol. 37, no. 5, pp. 977–997, May 2004.
[3] J. Liang, D. Doermann, and H. P. Li, "Camera-based analysis of text and documents: a survey," *IJDAR*, vol. 7, no. 2-3, pp. 84–104, 2005.
[4] Antani, S., Kasturi, R., and Jain, R., "A Survey on the Use of Pattern Recognition Methods for Abstraction, Indexing, and Retrieval of Images and Video", *Pattern Recognition* 35 945-965,2002.
[5] Lyu, M.R., Song, J., Cai, M.: "A comprehensive method for multilingual video text detection, localization, and extraction", *IEEE Trans. Circuits Syst. Video Technol.* **15**(2), 243–255 ,2005.
[6] Leon, M., Vilaplana, V., Gasull, A. and Marques, F., "Caption text extraction for indexing purposes using a hierarchical region-based image model," *IEEE ICIP 2009, El Cairo, Egypt*, 2009.
[7] Xiaoqing Liu and Jagath Samarabandu, " Multiscale Edge-Based Text Extraction From Complex Images" , *Multimedia and Expo, 2006 IEEE International Conference,2006.*
[8]Leon, M., Vilaplana, V., Gasull, A. and Marques, F., "Caption text extraction for indexing purposes using a hierarchical region-based image model," IEEE ICIP 2009, El Cairo, Egypt, 2009.
[9] Hyung Il Koo, Member, IEEE, and Duck Hoon Kim, Member, IEEE "Scene Text Detection via Connected Component Clustering and Nontext Filtering" in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 6, JUNE 2013
[10] M. Basavanna , P. Shivakumara, S. K. Srivatsa, G. Hemantha Kumar,_Multi-Oriented Text Detection In Scene Images,_ in International Journal of Pattern Recognition and Artificial Intelligence , Vol. 26, No. 7 Dec 2012.
[11] Yao Li and Huchuan Lu "Scene Text Detection via Stroke Width" in 21st International Conference on Pattern Recognition (ICPR 2012) November 11-15, 2012. Tsukuba, Japan.
[12] . J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in Proc. Brit. Mach. Vis. Conf., 2002, pp. 384–393.