IJRCCT International Journal of Research in Computer and Communication Technology

Text Detection On Scene Images Using MSER

Kethineni Venkateswarlu[1], Sreerama Murthy Velaga[2]
Department of Computer Science & Engineering,
GMR Institute of Technology, Rajam, Andhra Pradesh, India,
[1] PG –M.Tech, Dept. of CSE, venkat0534@gmail.com
[2] Associate Professor, Dept. of CSE, sreeramamurthy.v@gmrit.org

Abstract

Text detection and recognition is one of the difficult tasks in the computer vision community and there is a lot of research going on in recent years. This paper focuses on the problem of text detection and recognition from scene images. We propose a novelapproach to recognize the text from the scene images. First, we detect Maximally Stable Extremal Regions**(**MSER) from the input image. Then the image containing MSER regionis fed as input to the canny edge detector, which produces edges over text region and helps us to remove the remaining part of the image by applying filtering technique. Finally text region image is given to the Optical Character Recognition (OCR).The OCR produces the actual text presented in the input image.

Key Words: MSER, OCR, CANNY Edge Detection.

1. Introduction

Now-a-days with the rapid growth of technology there are many camera based applications are available in portable devices like cell phones, tabs etc. Everyone is able to capture the images easily, but whenever we want to read the text presented in those images are very difficult. This is the main problem in computer vision community [11, 14].Since so many years, the text detection play very important role in human life it can be helpful in the language translation and navigation. More over text extraction plays a crucial role for blind people when they want to read the text presented in the scene images. By these ways the text recognization and detection can play vital role in humans every day and in future it can be part of so many computer applications.

In this paper, we focus on a special case of scene text detection problem. Section 2 describes the overview of text detection and recognition. Section 3 deals with extracting Maximally Stable Extremal Regions (MSER). Section 4 describes extracting text regions and removing other parts of the image by employing canny edge detector with filtering techniques. Experimental results are discussed in Section 5.

2. Overview of Text Recognition and Detection

For the text recognition and detection process we perform the different steps all the steps are clearly showed in the figure 1.

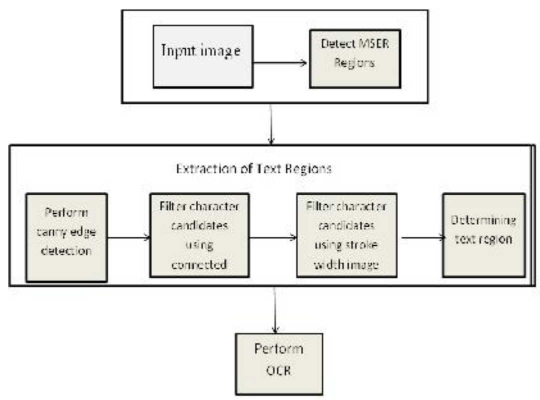

Figure 1: overview of text recognition and detection.

Initially we have to take the input image and next we have to perform the MSER technique to detect the text regions. The text regions consists the some unwanted information to remove that information we have to apply canny edge detection filtering technique, use stroke width method to filter the character candidates finally we get the text region that region gives to the OCR.

3. Extracting Maximally Stable Extremal Regions (MSER)

MSER is a method for blob detection in the images; it is a stable connected component of some gray level sets of the image. MSER depends on the threshold of the image, if we give them some threshold value the pixels below that threshold value are 'white' and all those above or equal are black. In our project we choose the minimum threshold value 0.9, MSER detect the objects and all the objects can be filled with different colors in this process some of the regions include the extra background pixels [6-8].

ISSN (Online) 2278- 5841
ISSN (Print) 2320- 5156

Those are removed in the canny edge detection process.

Implementation of MSER

- First of all sweep threshold of intensity from black to white performing a simple luminance thresholding of the image.
- Then extract the connected components ('Extremal Regions')
- Find a threshold when an extremal region is maximally stable.
- Finally we get the regions descriptors as features of MSER.

Image I is a mapping $I: D \subset Z^2 \to S$ these are the extremal regions defined on the image.

1. $S$is totally ordered that means it is reflective, anti-symmetric and transitive binary relation <=exists. In this paper we consider $S=\{0,1,..225\}$
2. An adjacency relation $A \subset D * D$ is defined in this paper we used 4-neighbourhoods are used i.e. $p,q \in D$ are adjacent $(pAq) iff \, \mathcal{E}_{i=1}^{d} |p_i - q_i| \leq 1$.

3. **Extraction of Text Regions**

In this section we discuss about Extraction of text regions for this we have employed canny edge detection, region filtering and finally stroke width technique to extract text regions from MSER.

Canny edge detection is well known method to detect the wide range of edges in the images [13]. We used canny to detect the edges of text region in the image, in our input image we have text and some other useless objects like people, trees, cars…etc. [9-10, 13].We want to detect the only text from that image the MSER identifies the regions and by using canny edge detector we detect the edges of text regions only so we can eliminate the other things easily.

Region filtering is used to identify the properties of different regions presented in the input image using the pixel values, by using those properties we can separate the image into sub images and we get the text region image.

Stroke width is useful discriminator for text in images, is the variation in stroke width within each text candidate. Most languages have the similar stroke width for the characters, so it is useful to eliminate regions where the stroke width exhibits a large variation.

Optical Character Recognition (OCR):

OCR is the most successful method to identify the text presented in the image.

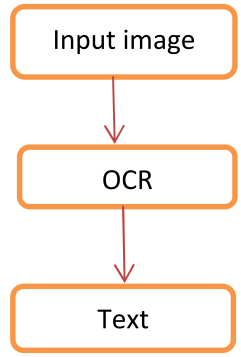

Figure2: Optical Character Recognition

The entire OCR is mainly classified into two categories: traditional optical character recognition and object recognition based. For traditional OCR based methods [1, 12, 2], various binarization methods have been proposed to get the binary image which is directly fed into the off-the-shelf OCR engine. On the other hand, object recognition based methods assume that scene character recognition is quite similar to object recognition with a high degree of intraclass variation for scene character recognition; these methods [3-5, 11] directly extract features from original images and use various classifiers to recognize the character.

The figure2 shows the overview of the OCR process. After completion of MSER and CANNY edge detection process the text region in the image can be given to the OCR, the OCR read the entire text presented in that image and display the output text.

4. EXPERIMENTAL RESULTS

In this section we discuss about various case studies. The input image should at least contain one word in it.

Test Case 1: In this test case we consider an image as shown in figure 3 which contains some text in foreground and in background there is some theme present. Here the text is of different size.

Figure3: original input image.

The input image is given to the MSER module, and MSER detects the various objects presented in image which is shown in the figure 4.

ISSN (Online) 2278- 5841
ISSN (Print) 2320- 5156

Figure4: MSER regions

After finding the MSER regions each and every region can be filled with different colors from 0 to 255.Subsequently, all MSER regions are fed to the canny edge detector. The main use of the canny is to remove non-text regions from the input image.

Figure5: Canny edges and intersection of canny edges with MSER regions.

After the canny edge detection we have to separate the letters from the background and many of the non-text regions have been separated from text, and we have to remove gradient grown edge pixels, this is shown in the Figure 6 and Figure 7.Now we have to perform the filtering, the purpose of filtering is to remove some of theconnected components by usingtheir region properties. Figure 7 clearly shows after filtering the input image and before filtering it depends on the threshold.

Figure6: Original MSER regions and segmented MSER regions.

Figure7: Text candidates before and after region filtering

In the most cases some of the characters have the similar stroke width or thickness so we have to remove the regions where the stroke width exhibits with more variation.Figure 8 shows the visualization of stroke width but most non text regions show a large variation in stroke width that can be filtered using the coefficient of stroke

Figure8: Visualization of text candidate's stroke width

Figure9: Text candidates before and after stroke width filtering

Figure10: Image region under mask created by joining individual characters.

width variation showed in figure 9.After applying the stroke width we have to compute the bounding box of the text region, actually there is a bounding box in the Figure 10, it is invisible because the background of the image and bounding box color are same. The bounding box showed in the test case 2 clearly. We will first merge the individual characters into a single connected component as shown in the Figure 10.

Finally we get the image with the only text region;this text region is given to the OCR to produce corresponding text.

Figure11: Text region

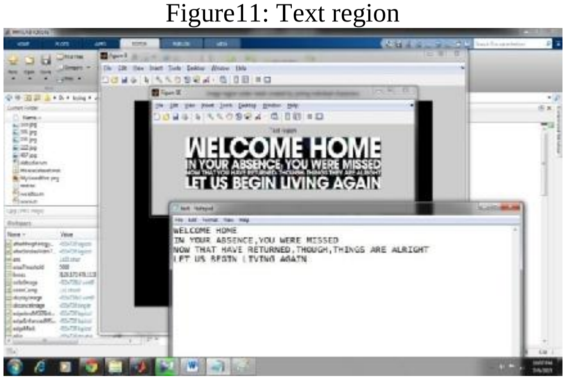

Figure 12: OCR results

The Figure12 shows the OCR results for the text region in a separate notepad file.

Test Case 2: In this test case we have chosen text with plain background.

Figure (13.A), (13.B)

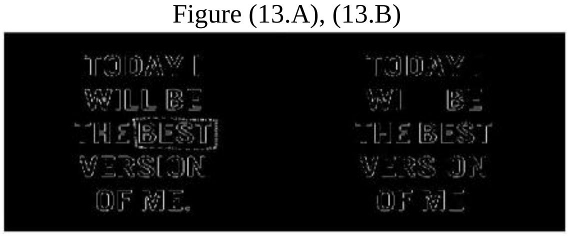

Figure (13.C)

Figure (13.D):

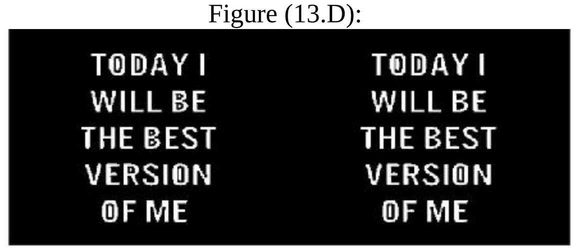

Figure (13.E):

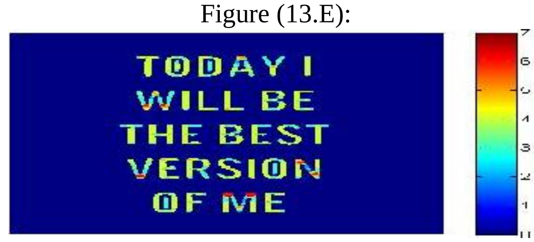

Figure (13.F)

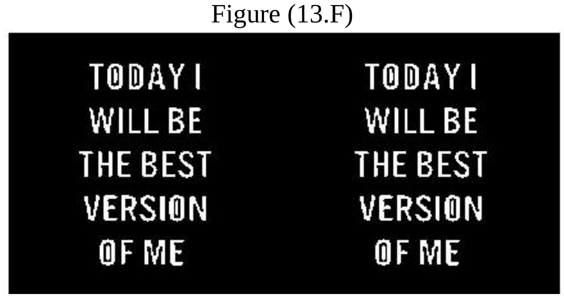

Figure (13.G)

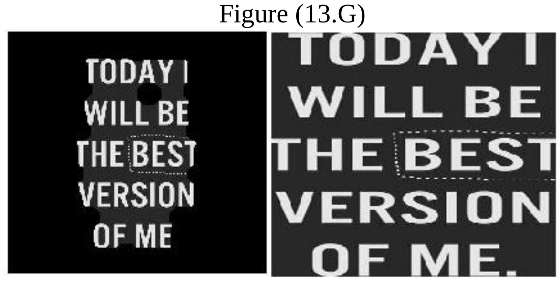

Figure (13.H), (13.I)

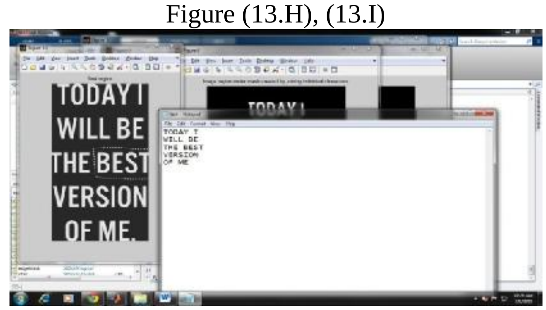

Figure (13.J)

ISSN (Online) 2278- 5841
ISSN (Print) 2320- 5156

Figure 13(a) Original Image (b) MSER regions (c) Canny edges and intersection of canny edges with MSER regions (d) Original MSER regions segmented MSER regions (e)Text candidates before and after region filtering (f) Visualization of Text candidate's stroke width (g)Text candidates before and after stroke width filtering (h) Image region under mask created by joining individual characters (i) Text Region (J) OCR results.

Conclusion

In this paper we proposed new methodology to recognize the text presented in scene images. Firstly our algorithm detects the MSER regions and that regions filled with multiple colors, next we use canny edge detection method for them to detect the text region edges we compare the original MSER region and segmented MSER regions, by using the mask we join individual characters in the image finally the text region image we will get without any noisy objects in the image and text image can be given to the OCR, the OCR checks the text in the image and gives the results.

REFERENCES

[1] X.chen and A.yuille. Detecting and reading text in natural scenes. In proceedings of *IEEE Conference on computer viosn and pattren recognigiation(CVPR)* volume 2, pages II-366. IEEE, 2004

[2] L.Neumann and J.Matas. Real-time scene text localization and recognition. In *proceedings, Eighth International conference on Document Analyisis and Recognition (CVPR)*, 2012.

[3] T.de campos, B.Babu and M. varma Character recognition in natural images. In *VISAP*, 2009.

[4]. K.Wang and S.Belongie. word spotting in the wild. *Computer vision-ECCV,* pages 591-604, 2010.

[5]. K.Wang, B.Babenko and S.Belongie. End-to-end scene text recognition. In *international conference on computer vision(ICVV), 2011.*

[6]. Per-Erik Forssen and David G.Lowe Shape Descriptors For Maximally Stable Extermal Regions.

[7]. P.E Forssen. Maximally Stable colour regions for recognition and matching. In *IEEE coference on computer vision and pattren Recognition,* Minneapolis, USA, June 2007. IEEE Computer Society.

[8]. J. Matas, O. Chum, M.Urban and T. Pajdla. Robust wide baseline stero from maximally stable extremal regions. In *13th BMVC,* pages 384-393 September 2002.

[9]. JOHN CANNY, member IEEE. A Computational Approach to Edge Detection.*IEEE Transactions on pattren analysis and Machine Intelligence Novmber 1986.*

[10]. J.F. Canny, "Finding edges and lines in image" M.I.T. Artificial Intell. Lab Cambridge, MA, Rep AI-TR-720, 1983.

[11]. Isard, M. ; Microsoft Res., Mountain View, CA USA. "PAMPAS: real valued graphical models for computer vision". *In Computer Vison and Pattern Recognition, 2003 IEEE Computer Socity Conference.*

[12] M.Yokobayashi and T.wakahara. Segmentation and recognition of characters in scene images using selective binarization in color space and get correlation. In *proceedings, Eighth International conference on Document Analysis and Recognition (IDAR)*, pages 167-171, IEEE, 2005

[13]. M.Gennert, "Detecting half-edges and verticesin images". In *IEEE conf. comput. Vision and pattren recoginition*, Mimai Beach, FL. June 24-26, 1986.

[14]. Bo Zang, Dept.of computer science,Tsinghua univ., China. "Computer vision vs. human vison". In *Cognitive informatics (ICCI), 2010 9th IEEE International Conference.* July 2010.

[15]. Saleh Ali K. Al-Omari, Putra sumari, Sadik A. Al-Taweel and Anas J.A Husain. Digital Recognition using Neural Networks. Journal of computer sciences, 2009.