Scientific Journal of Impact Factor (SJIF): 3.134

E-ISSN (O): 2348-4470
P-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 12,December -2015

Scene Text Recognition by using EE-MSER and Optical Character Recognition for Natural Images

Miss. Beena Dhamecha[1],Prof. Vaishali J. Kalariya[2]

Department of Computer Engineering, R.K.University, Rajkot, Gujrat, India[1]
Department of Computer Engineering, R.K.University, Rajkot, Gujrat, India[2]

Abstract –*Text in Natural Images contains a valuable information about scene, convey the information about what is actually depicted in the images. Text extraction is challenging task because of text might be in different style, different font, illumination condition, poor quality, surface deformation, complex background etc. In this paper proposes a two method one is EE-MSER for Text Detection and OCR for Text Recognition.In EE-MSER,by using geometric and stroke width information the letter candidates are filtered to exclude non-text region. OCR method takes the detected text portion and gives the recognized text. Experiments on ICDAR 2003 Robust Reading character dataset and CHAR74k dataset shows that proposed technique perform well compared with other techniques.*

Keywords-*text detection; text recognition; complex background; edge-enhanced maximally stable extremal regions(EE-MSER),Optical Character Recognition.*

I. INTRODUCTION

Text in images carry high-level semantic information of scene. Images are increasing on webs and in databases. It is a pressing task to develop effective methods to manage and retrieve these resources by their content. It is a difficult task to detect and segment text from scene/captured images due to main reasons like: different types of text patterns like size,font style,orientations,colors, background outlier similar to the text characters. After text detection and segmentation, text recognition system is applied to convert image into readable text, but it performs poorly when there is a text on the complex background. Text recognition is important for a lot of applications like automatic sign reading, navigation, language translation, license plate reading, content based image search etc. So it is necessary to understand scene text than ever.

With the rapid growth in digital technologies and gadgets which are made with megapixel cameras and other devices are like PDA, mobiles etc. are responsible for increasing the attention for information retrieval and it leads to a new research task.Text, in the images contain valuable information and provide a cues about images. So it is very important for a human as well as computer to understand the scenes.

II. LITERATURE REVIEW

2.1 Text Detection

Number of methods has been proposed for text detection in the past. Normally Text detection can be classified into following categories: Region-based method, Edge-based, texture-based,and connected-component based method. Whereas Text Recognition can be classified into two categories such as Traditional Optical character recognition (OCR) and Object recognition based method.

In region-based method, scan the images at multiple scales and uses the text/non-text classifier to find the potential text areas. Generally a feature vector extracted from local area is fed into a classifier. Because of text region have different properties from non-text ones, it can detect and localize text accurately even when are noisy. For region-based methods, the speed is very slow, and it is sensitive to text alignment orientation.

Coates et al. [1] proposed to learn features automatically from unlabeled data using unsupervised features learning and then train liner SVM to classify whether a sliding window is text or non-text ones.

In connected-component based method, it directly segments the candidate text components by edge detection, color clustering to get the CCs. The non-text components are then removed using heuristic rules or classifiers.In this number of segmented candidate components is small, so that computational cost is low and the located candidate text components are directly used for text recognitions. CC-based methods cannot segment text component properly without prior knowledge of text position and scale.

Epshtein et al. [2] used the CCs in a SWT images to form text lines. Shivakumara et al. [3] proposed to extract CCs by performing K-means clustering in the fourier - Laplacian domain, and use text straightness and edge density to discard false positives. Chen et al. [4] proposed edge-enhanced MSER as basic text candidates and geometric filtering and SWT is used to discard the non-text ones.

Edge-based method focus on the 'high contrast between the text and the background' and edges og the text boundary are identified and merged. Liu et al. [5] extract statistical features from Sobel edge in four directions and use K-means classifier to classify text or non-text cluster. This method is robust for complex background .but still it fails to detect low contrast and small font size texts. It is also expensive. Wong et al. [6] compute maximum gradient difference to identify the line segments then extended to neighboring top and bottom rows to form candidate text regions. It has low false positive rate, but it uses many classifiers and sensitive to threshold values.

In texture-based methods, it considers text as a special textures. It apply Fast Fourier Transform, DCT, wavelet decomposition, and Gabor filter for feature extraction. Ye et al. [7] calculate the wavelet energy features at different scale and perform the thresholding to find out the candidate text pixels then it is merged into text lines.

2.2 Text Recognition

In Traditional Optical Character Recognition (OCR) based method, different binarization methods have been applied to get the binary images, that directly fed to the off-the-self OCR[8]. Text in scene images differs from the scanned document in terms of size, font, illumination condition, resolutions etc.The loss of information in binarization is not recoverable

Other method Object Recognition based method, it directly extract features from original images and uses various different classifier to recognize the text [8]. In this method, do not do the binarization and segmentation but uses multi-scale sliding window technique to get the candidate character detection result. A special structure information is not used in the sliding window technique, so it will produce many false positives. Because of this, it is depended on post-processing methods like pictorial structure or CRF model.

In Otsu's method, it is based on histogram and used a global thresholding[9]. Using k-means clustering,text detection and binarization method is worked for Korean sign board. But in complex background and lighting, it is difficult for finding best value for k. Different methods have been suggested for text extraction.

Cai et al. [10] proposed a text detection method which is character features such as edge density, edge strength. First apply a coloe edge detection algorithm in YUV colorand filter out non-text edges by using a low threshold. Then, By applying a local thresholding techniques to keep low-contrast text and simplify the background. At last, to localize text regions projection profiles are analyzed.

Kim[11] proposed a method in which LCQ is performed for each color separately. Each color is assumed as a text color without knowing whether it is real text color or not.to reduce processing time, image is converted into a 256-color image before color quantization done. To find candidate text line, the candidate components which are extracted for each color are combined when show text region features. Disadvantage of this method is high processing time.

Jain and Yu[12] by bit dropping first perform a color reduction and color quantization, then after multi-valued images decomposition algorithm is applied to decompose the input image into multiple background and foreground images. Then to localize text candidate,connected component analysis combined with projection profile features are performed on each of the images.it extract only horizontal text for large sizes.

Neumann and Matas [13] first detect characters as a MSERs and the apply text recognition using segmentation which is obtained by the MSER detector.

Wang et al. [14] to train the text detection and character recognition modular, it used a CNN method, and built an end-to-end system with non-maximal suppression (NMS). And search with help of lexicon.Speed is relatively slow.

In this paper,proposed a novel CC-based EE-MSER method for text detection.it combines the complimentary properties of canny edges and MSER.Than by using the distance transform generate the stroke width transform images to obtain more accurate result.The stroke width information and geometric information are applied to perform filtering and pairing of CCs.At the final extract the detection region and remove non-text region. Now on the detected region applied an OCR method to recognize the characters that first apply preprocessing task like normalization and smoothing process and then extract the features and classify the characters according to the classes and after that apply the postprocessing task like grouping the character into the same string and make a word and at the end recognize the detected text and give is as an output.It is a good method to different illumination and local geometric transformation.

III. PROPOSED WORK

3.1 Text Detection

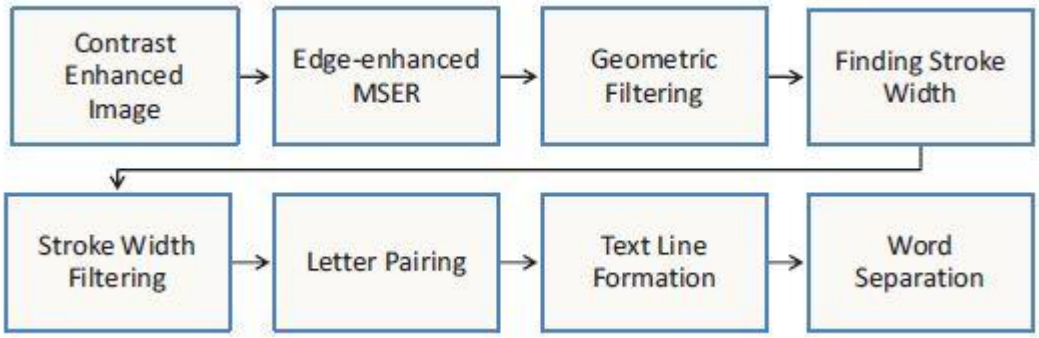


Fig 1.1 Flowchart of EE-MSER

This is a flowchart of text detection method shown in fig 1.1.For the input image intensities are adjustedlinearly to enhance the contrast.Then extract the MSER regions from images and enhanced by using canny edge. In next stepResulting CCs are filtered using geometric information.The stroke width transformation is computedfor object and object with high variation are rejected. Text CC are grouped pairwise and form the text lines.At the end words are separated.

3.1.1 Edge enhanced MSER[15]

MSER is identified as a one of the best region detector.It is robust against lighting condition, view point, etc. but it is a sensitive to blurred images. So that by applying a simple MSER to images, small letters cannot be detected. That's why there is some modification is required in MSER and it is combined with canny edge detection technique. Here remove the MSER pixel outside of the boundary form by the canny edges and text letters are separated.

With the extraction of EE-MSER,get the binary image where foreground CCs are considered as a letter candidates. Geometric filtering is applied to eliminate the non-text regions form the images.First too small or too large object is removed, and some threshold value is decided to make sure that not to remove the letters like "i" and"l".

To label each foreground pixel with the distance to its near background pixel, Euclidean distance transform is applied. The ridge value of the distance map correspond to half the width of stroke width.Then,transform the stroke width information from the ridge to the background of the object.The rejection criterion is std/mean>0.5 decided.

A letters which is belonging to the same text lines have a similar stroke width and character height.If CCs are much distant then two CCs should not be paired.If a line contains three-or more text object,it is declared as a text line.At the end,text line are divided into individual words by using word spacing and character spacing information.

3.2 Text Recognition

OCR is used for the recognize the text from the images.It is used for both the offline handwritten text or printed /scanned text or on-line text which is recognized by the computer.It's dependent on the given document/images as an input to the OCR system.

3.2.1 Optical Character Recognition

In segmented portion of the text image contain certain amount of noise.Text have a different kind of font style,size illumination condition,etc.In Preprocessing smoothing is one kind of task which does the filling and thinning.Filling discard the gaps,holes,etc. and thinning decreases the width of the line. It also includes the normalization as a preprocessing task to obtain a uniform size,orientation, etc.

Feature extraction is used to capture the characteristics of the characters,symbol,and numbers.There are many feature extraction methods are available.Then apply the classification process to classify the characters according to the classes in which it belongs.There are two types of classifications one which is used when the character description can be numerically available in feature vector.Another is extract the pattern characteristics from the character which is not easily quantified.For example 'L' and 'T' both have a one horizontal and one vertical stroke, so we need to distinguish them as a different characters. Many classifiers are used like k-means, BP-algorithm, Baye's classifier,neural network, SVM classifier.

After that Post processing task is applied .The characters which are classified individually so we should have to group together into a same string which is belongs to them and form a word and numbers.Fonts have a fixed pitch so that the process of grouping is easy as the position of each characters is known. Distance between each character is fixed and distance between words is larger than the distance between two characters so grouping is possible.In this task it also do the error detection and error correction to the characters.

Scene text can be recognized by training a classifier .By using LIBLINEAR [17], a linear SVM classifier is trained up to 18500 character images, which is much faster.OCR recognize the text and gives the output.

IV. EXPERIMENTAL RESULT

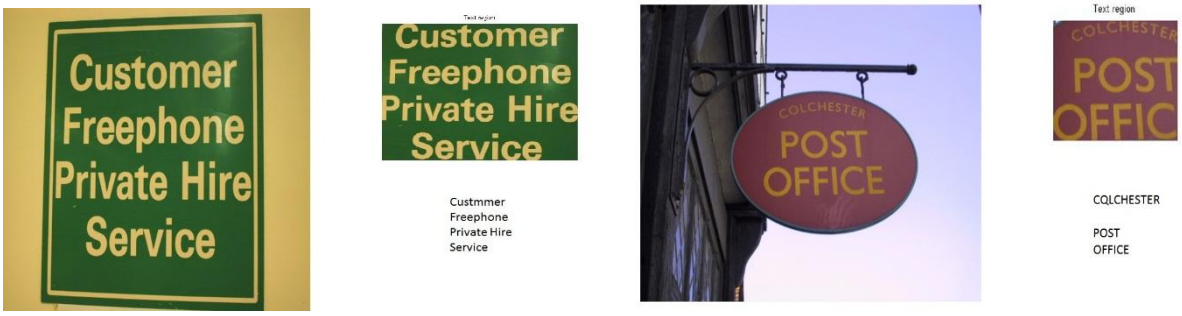
This proposed recognition technique is tested on public dataset including CHAR74k[18], ICDAR 2003[19] dataset, case sensitive and insensitive datasets. Here it works for only recognition of an English character consists of 62 classes (0-9,A-Z,a-z).

The ICDAR 2003 robust reading competitions provides different kind of images for text detection, word recognition and character recognition. It contains a 6113 characters for training and 5379 character for testing. The characters are collected from various kind of scene like board, book covers, brand logo, road sign etc. Here CHAR74k data set is used as a training dataset.

Methods	ICDAR 2003	Char74K
ABBYY FineReader 10[20]	26.6%	15.4%
GB+NN[21]	41.0%	-
HOG+NN[22]	51.5%	-
MSER[13]	67.0%	-
HOG+SVM[23]	-	61.9%
Proposed OCR + EE-MSER	**78.30%**	**63.06%**

Figure 4.1: Character Recognition accuracy on ICDAR 2003 and CHAR74k dataset

In this experiment, first phase MSER detection method is modified with the edge detection method and segment the text area from non-text area. After that apply the preprocessing and post processing task to recognize the text most accurately then OCR text recognition method is used for the recognize the segmented text.



V. CONCLUSION

Text Detection and Text Recognition from a natural images like document, digital camera based and web, email is challenging due to the random text appearances and complex backgrounds. In this paper, proposed an EE-MSER for text detection and OCR is for text recognition to improve the performance and recognize accurately.in EE-MSER, simple MSER and canny edge detection method are combine to extract the text region. OCR performs the Preprocessing and

processing task and recognize the segmented text. It is very robust to various illumination conditions and geometric transformation. It is evaluated on both ICDAR 2003 and CHAR74k dataset and gives us a better result.

In future examine the other techniques and other features descriptors method and try to improve the performance of text recognition process which consists of a complex background.

REFERENCES

[1] A. Coates, B. Carpenter, C. Case, S.Satheesh, B.Suresh, T. Wang, D.Wu, A.Ng, Text detection and characterrecognition in scene images with unsupervised feature learning, in International Conference on Document Analysis and recognition (ICDAR), IEEE, 2011, pp.440-445.
[2] B. Epshtein, E. Ofek, Y. Wexler, Detecting text in natural scenes with stroke width transform, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE,2010,pp,2963-2970.
[3]P.Shivakumara, T.Phan, C.Tan, A Laplacian approach to multi-oriented text detection in video, IEEE Trans. Pattern Anal. Mach. Intel. 33 (2) (2011) 412-419.
[4]H.Chen, S.S. Tsai, G.Schroth, D.M. Chen, R.Grzeszczuk, B.Girod, Robust text detection in natural images with edge-enhanced maximally stable extremal regions, in: 18th IEEE International Conference on Image Processing (ICIP), IEEE, 2011, pp. 2609-2612.
[5] C. Liu, C. Wang, and R. Dai, "Text Detection in images Based on Unsupervised Classification of Edge-Based features", Proc. Eighth Int'l Conf. Document Analysis and Recognition, pp.117-120,2002.
[6] E.K. Wong and M. Chen, "A New Robust Algorithm for Video Text Extraction," Pattern Recognition, vol. 36, pp. 1397-1406, 2003.
[7]Q. Ye, Q. Hung, W. Gao, and D.Zhao, "Fast and robust Text Detection in Images and Video frames," Image and vision Computing, vol.23, pp. 565-576, 2005.
[8] Cunzhaoshi,Chunheng Wang, Baihua Xiao, Yang Zhang, Song Gao,Zhong Zhang "Scene text Recognition using Part based tree structured Character Detection,"pattern recognition in 2013 Elsevier Ltd.
[9] N. Otsu, A threshold selection method from gray-level histograms. Automatica, 11:285-296, 1975.
[10] 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.
[11] P.-K. Kim." Automatic Text Location in Complex Color Images Using Local Color Quantization.IEEE TENCON, vol. 1, pp. 629-632, 1999.
[12] A.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.
[13] L. Neumann, J.Matas, A method for text localization and recognition in real world images, in: ACCV, Springer, Queenstown, New Zealand, 2010, pp. 770-783.
[14]T.Wang, D.J. Wu, A. Coates, A.Y. Ng, "End to end text recognition with convolutional neural networks," in: 21st International Conference on Pattern Recognition (ICPR), IEEE, 2012, pp. 3304-3308.
[15]H.Chen,S.S.Tsai,G.Schroth,D.M.Chen,R.Grzeszczuk,B.Girod,"Robust text detection in natural images with edge-enhanced maximally stable extremal regions",in:18th IEEE international Conference on Image Processing(ICIP),IEEE,2011,pp,2609-2612.
[16]N.Dalal and B.Triggs,"Histogram of Oriented Gradients for Human detection",in Computer Vision and pattern Recognition,2005.CVPR 2005,IEEE Computer Society Conference on,vol.1.IEEE,2005,pp.886-893.
[17]C.J.Hsiek,K.W.Chang,C.J.Lin,S.S.Keerthi,S.Sundarajan,"A dual coordinate descent method for large-scale linear svm", in Proceedings of the 25th international conference on Machine learning, vol.951,no.08,2008,pp.408-415.
[18] Z. Xiangxin and D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in Proc. IEEE CVPR, Jun. 2011, pp. 28792886.
[19] JEROD J. WEINMAN, "unified detection and recognition for reading text in scene images"University of Massachusetts Amherst 2008.
[20] AbbyyFineReader 10, http://www.abbyy.com/
[21] T. E. de Campos, B. R. Babu, and M. Varma, " Character recognition in natural images." In VISAPP(2), 2009,pp.273-280.
[22] K.Wang and S.Belingie," Word spotting in the wild", Computer Vision-ECCV 2010, pp.591-604,2010.
[23] A. Mishra, K. Alahari, and C.Jawahar, "Top-down and bottom-up cues for scene text recognition," in Computer Vision and pattern Recognition(CVPR), 2010 IEEE Conference on. IEEE,2012,pp.2687-2694.

