# Survey on Scene Text Recognition by using EE-MSER and OCR for Natural Images

Prof. Bere Sachin S.
Assistant Professor, Computer Engg. Department
Dattakala Group of Institution, Faculty of Engg.
Bhigwan, India.

Prof. Rajpure Amol S.
Assistant Professor, Computer Engg. Department
Dattakala Group of Institution, Faculty of Engg.
Bhigwan, India.

***Abstract*— Content in Natural Images contains a significant data about scene, pass on the data about what is really portrayed in the pictures. Content extraction is testing errand as a result of content may be in various style, distinctive text style, light condition, low quality, surface disfigurement, complex foundation and so on. In this paper proposes a two technique one is EE-MSER for Text Detection and OCR for Text Recognition. In EE-MSER, by utilizing geometric and stroke width data the letter applicants are sifted to prohibit non-content area. OCR strategy takes the distinguished content part and gives the perceived content. Investigations on ICDAR 2003 Robust Reading character dataset and CHAR74k dataset demonstrates that proposed system perform very much contrasted and different strategies.**

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

## I. INTRODUCTION

Content in pictures convey abnormal state semantic data of scene. Pictures are expanding on networks and in databases. It is a squeezing errand to create viable techniques to oversee and recover these assets by their substance. It is a troublesome errand to identify and section content from scene/caught pictures because of principle reasons like: distinctive kinds of content examples like size,font style,orientations,colors, foundation exception like the content characters. After content discovery and division, content acknowledgment framework is connected to change over picture into intelligible content, yet it performs inadequately when there is a content on the mind boggling foundation. Content acknowledgment is imperative for a ton of utilizations like programmed sign perusing, route, language interpretation, tag perusing, content based picture seek and so on. So it is important to comprehend scene content than any time in recent memory.

With the quick development in advanced innovations and contraptions which are made with megapixel cameras and different gadgets resemble PDA, mobiles and so forth are in charge of expanding the consideration for data recovery and it prompts another examination task. Text, in the pictures contain significant data and give a signals about pictures. So it is essential for a human just as PC to comprehend the scenes.

## II. LITERATURE SURVEY

### 2.1 Text Detection

Number of strategies has been proposed for content recognition previously. Ordinarily Text location can be grouped into following classes: Region-based strategy, Edge-based, surface based, and associated segment based technique. While Text Recognition can be ordered into two classifications, for example, Traditional Optical character acknowledgment (OCR) and Object acknowledgment based technique. In district based strategy, examine the pictures at numerous scales and uses the content/non-content classifier to locate the potential content territories. By and large a component vector extricated from neighborhood sustained into a classifier. Due to content district have diverse properties from non-content ones, it can identify and limit message precisely notwithstanding when are loud. For district based strategies, the speed is extremely moderate, and it is delicate to content arrangement introduction.

Coates et al. [1] proposed to gain includes naturally from unlabeled information utilizing unsupervised highlights learning and after that train liner SVM to group whether a sliding window is content or non-content ones.

In associated segment based technique, it straightforwardly sections the hopeful content segments by edge recognition, shading grouping to get the CCs. The non-content segments are then expelled utilizing heuristic guidelines or classifiers. In this number of fragmented hopeful parts is little, so computational expense is low and the found competitor content segments are straightforwardly utilized for content acknowledgments. CC-based techniques can't portion content segment legitimately without earlier information of content position and scale.
Epshtein et al. [2] utilized the CCs in a SWT pictures to frame content lines. Shivakumara et al. [3] proposed to separate CCs by performing K-implies grouping in the fourier-Laplacian space, and use content straightness and edge thickness to dispose of false positives. Chen et al. [4] proposed edge-improved MSER as fundamental content competitors and geometric sifting and SWT is utilized to dispose of the non-content ones.

Edge-put together technique center with respect to the 'high complexity between the content and the foundation' and edges of the content limit are distinguished and blended. Liu et al. [5] separate measurable highlights from Sobel edge in four ways and use K-implies classifier to characterize content or non-content group. This strategy is strong for complex foundation .yet at the same time it neglects to recognize low differentiation and little text dimension writings. It is likewise costly. Wong et al. [6] figure most extreme angle contrast to recognize the line fragments at that point reached out to neighboring best and base

-

lines to shape hopeful content locales. It has low false positive rate, however it utilizes numerous classifiers and touchy to limit esteems.

In surface based strategies, it thinks about content as exceptional surfaces. It applies Fast Fourier Transform, DCT, wavelet disintegration, and Gabor channel for highlight extraction. Ye et al. [7] ascertain the wavelet vitality highlights at various scales and play out the thresholding to discover the hopeful content pixels then it is converged into content lines.

2.2 Text Recognition

In Traditional Optical Character Recognition (OCR) based strategy, distinctive binarization techniques have been connected to get the paired pictures, that specifically bolstered to the off-the-self OCR[8]. Content in scene pictures contrasts from the checked record regarding size, textual style, brightening condition, goals etc.The loss of data in binarization isn't recoverable

Other technique Object Recognition based strategy, it straightforwardly extricate highlights from unique pictures and uses different diverse classifier to perceive the content [8]. In this strategy, don't do the binarization and division however utilizes multi-scale sliding window procedure to get the applicant character discovery result. A unique structure data isn't utilized in the sliding window procedure, so it will create numerous bogus positives. Along these lines, it is relied upon post-preparing strategies like pictorial structure or CRF show.

In Otsu's strategy, it depends on histogram and utilized a worldwide thresholding[9]. Utilizing k-implies clustering,text recognition and binarization strategy is worked for Korean sign board. Be that as it may, in complex foundation and lighting, it is troublesome for discovering best an incentive for k. Diverse strategies have been recommended for content extraction.

Cai et al. [10] proposed a content discovery strategy which is character highlights, for example, edge thickness, edge quality. First apply a coloe edge recognition calculation in YUV colorand sift through non-content edges by utilizing a low limit. At that point, By applying a nearby thresholding procedures to keep low-differentiate message and improve the foundation. Finally, to restrict content areas projection profiles are broke down.

Kim[11] proposed a strategy in which LCQ is performed for each shading independently. Each shading is expected as content shading without knowing whether it is genuine content shading or not. to decrease preparing time, picture is changed over into a 256-shading picture before shading quantization done. To discover applicant content line, the competitor parts which are extricated for each shading are consolidated when show content area highlights. Hindrance of this strategy is high preparing time.

Jain and Yu[12] by bit dropping first play out a shading decrease and shading quantization, at that point after multi-esteemed pictures disintegration calculation is connected to deteriorate the info picture into numerous foundation and frontal area pictures. At that point to restrict content candidate, connected part examination joined with projection profile highlights are performed on each of the images. it extricate just level content for extensive sizes.

Neumann and Matas [13] first identify characters as a MSERs and the apply content acknowledgment utilizing division which is acquired by the MSER finder.

Wang et al. [14] to prepare the content discovery and character acknowledgment secluded, it utilized a CNN strategy, and fabricated a start to finish framework with non-maximal concealment (NMS). Furthermore, seek with assistance of lexicon. Speed is generally moderate.

In this paper, proposed a novel CC-based EE-MSER strategy for content detection. it joins the complimentary properties of vigilant edges and MSER. Than by utilizing the separation change create the stroke width change pictures to get increasingly precise result. The stroke width data and geometric data are connected to perform sifting and blending of CCs. At the last concentrate the identification district and evacuate non-content locale. Presently on the distinguished district connected an OCR technique to perceive the characters that initially apply preprocessing undertaking like standardization and smoothing procedure and afterward remove the highlights and order the characters as per the classes and after that apply the post processing errand like gathering the character into a similar string and make a word and toward the end perceive the identified content and give is as an output. It is a decent strategy to various enlightenment and neighborhood geometric change.

## III. PROPOSED METHODOLOGY

3.1 Text Detection

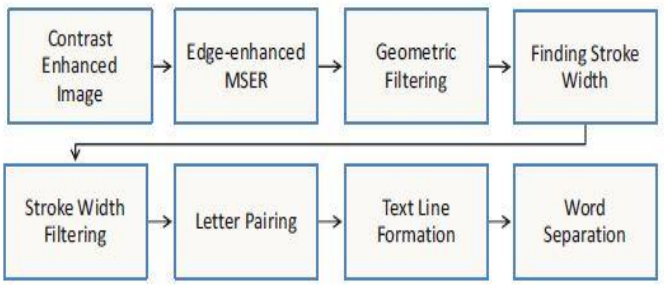



Fig 1.1 Flowchart of EE-MSER

This is a flowchart of content identification strategy appeared in fig 1.1.For the information picture powers are adjusted linearly to improve the contrast. Then separate the MSER areas from pictures and upgraded by utilizing vigilant edge. In next step Resulting CCs are separated utilizing geometric information. The stroke width change is computed for article and item with high variety are rejected. Content CC are assembled pair wise and structure the content lines. At the end words are isolated.

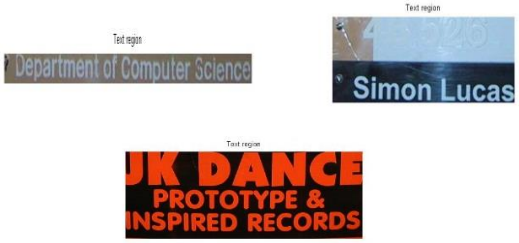
3.1.1 Edge enhanced MSER [15]

MSER is distinguished as a standout amongst the best area detector. It is hearty against lighting condition, see point, and so forth yet it is a delicate to obscured pictures. So that by applying a straightforward MSER to pictures, little letters can't be distinguished. That is the reason there is some change is required in MSER and it is joined with vigilant edge location strategy. Here expel the MSER pixel outside of the limit structure by the watchful edges and content letters are isolated.

With the extraction of EE-MSER, get the twofold picture

where closer view CCs are considered as a letter hopefuls. Geometric sifting is connected to take out the non-content locales structure the images. First excessively little or too expansive article is expelled, and some edge esteem is chosen to ensure that not to evacuate the letters like "I" and"1". To name each frontal area pixel with the separation to its close foundation pixel, Euclidean separation change is connected. The edge estimation of the separation map relate to a large portion of the width of stroke width. Then, transform the stroke width data from the edge to the foundation of the object. The dismissal rule is sexually transmitted disease/mean>0.5 chose.

A letters which are having a place with similar content lines have a comparative stroke width and character height. If CCs are much far off then two CCs ought not to be paired. If a line contains at least three content object, it is proclaimed as a content line. At the end, text line are separated into individual words by utilizing word dividing and character dispersing data.




3.2 Text Recognition

OCR is utilized for the perceive the content from the images. It is utilized for both the disconnected transcribed content or printed/checked content or on-line content which is perceived by the computer. It's reliant on the given record/pictures as a contribution to the OCR framework.

3.2.1 Optical Character Recognition

In sectioned bit of the content picture contain certain measure of noise. Text has an alternate sort of text style, size light condition, etc. In Preprocessing smoothing is one sort of undertaking which does the filling and thinning. Filling dispose of the gaps, holes, etc. also, diminishing reductions the width of the line. It likewise incorporates the standardization as a preprocessing undertaking to get a uniform size, orientation, and so on.

Highlight extraction is utilized to catch the qualities of the characters, symbol, and numbers. There are many component extraction strategies are available. Then apply the grouping procedure to order the characters as per the classes in which it belongs. There are two sorts of characterizations one which is utilized when the character depiction can be numerically accessible in highlight vector. Another is separate the example attributes from the character which isn't actually quantified. For model 'L' and 'T' both have a one flat and one vertical stroke, so we have to recognize them as alternate characters. Numerous classifiers are utilized like k-implies, BP-calculation, Baye's classifier, neural organize, SVM classifier.

After that Post preparing undertaking is connected .The characters which are ordered independently so we ought to need to gather together into an equivalent string which is has a place with them and structures a word and numbers. Fonts have a fixed pitch so the way toward gathering is simple as the situation of every character is known. Separation between each character is fixed and remove between words is bigger than the separation between two characters so gathering is possible. In this undertaking it likewise does the mistake location and blunder amendment to the characters.

Scene content can be perceived via preparing a classifier .By utilizing LIBLINEAR [17], a direct SVM classifier is prepared up to 18500 character pictures, which is much faster.OCR perceive the content and gives the yield.

## IV. 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.

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.