International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 03 | Mar-2018 www.irjet.net p-ISSN: 2395-0072

A Survey on MSER Based Scene Text Detection

Deepthy Joshy[1], Dr. Anishin Raj M. M[2]

[1]PG Student, Dept. of CSE, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India
[2]Assoc.Professor, Dept. of CSE, Viswajyothi College of Engineering and Technology, Muvattupuzha, Kerala, India

---

Abstract - *Scene text detection is a field where numerous researches are being held, due to its wide scope and applicability. But complex image backgrounds and adverse imaging conditions generally reduce the performance. Several techniques already exist to overcome these difficulties in producing efficient results. This paper focus on one of the major technique called MSER (Maximally Stable Extremal Region) based scene text detection. The survey conducted regarding how MSER technique is used for performing this task efficiently and which all techniques can support and enhance MSER outcomes is discussed in detail. The pros and cons of MSER is also enlisted here in this paper.*

Key Words: MSER, Text detection, Character filtering, Recognition, Scene image.

1. INTRODUCTION

Images hold large amount of semantic information which is very much useful in wide range of applications such as robotic navigation, fine grained classification, logo retrieval, car number plate recognition, product recognition and as support for visually impaired persons. Analysis of textual contents of an image can provide exact information about the image, when the visual cues are not sufficient to make a right judgement. Scene text detection techniques play a major role in this task of content based image analysis and information retrieval.

But due to the various imaging conditions such as irregular backgrounds, varying font style, size and colour of the characters, effects of lighting conditions such as shadows and occlusions make text detection and recognition from natural images, a complex task. So the direct application of OCR (Optical Character Recognition) systems became impossible on such images for text recognition. Therefore, additional methods are employed for detecting the text regions and after performing some filtering, it is given to the OCR text recognition module, thus making the task easier. One of the major techniques that is widely being used and has made great contribution to scene text detection is Maximally Stable Extremal Region.

2. RESEARCH WORKS

Houssem, Mohamed and Adel [1] proposed a method that separates out complex backgrounds by combining three different strategies of edge candidate detection and effective selection of text candidates done using stroke width transform and some heuristic filtering. The edge candidate detection is enhanced by adjusting the images into H, S and V channels of HSV color space, use of fractal processing for image intensity transformation, background filtering using Otsu, edge extraction by projecting both vertically and horizontally by Sobel operator and finally edge merging of edges of each of the channels. MSER based candidate detection includes MSER based mask creation in HSV color space and in grey level, text candidates are the intersection of the merged MSER masks of HSV channels with that of MSER mask in grey level. Then text candidates are selected and they are given for classification into text and non-text into an SVM classifier trained using HOG (Histogram of Gradients) and CNN (Convolutional Neural Network) features.

Leibin and Jizheng [2] proposed an algorithm where SWT (Stroke Width Transform) and MSER are used together for text detection. Connected component analysis is used to obtain candidate characters from MSER. SWT algorithm makes use of canny edge to find edges of an image and then from these responses calculate distances between two parallel edges. This helps to detect text line with similar stroke width. After this a preliminary filtering is performed based on some heuristics rules related to the geometric and statistical features of connected regions to avoid non-text regions. Stroke width variation, aspect ratios are some of those feature. Coincidence rate between each MSER and SWT region is calculated and that above a merge threshold is considered as a strong candidate and below this threshold value is considered as weak candidate. If any weak candidates have properties similar to strong candidate, then it is moved to strong text label. All the reliable characters are found out through all these steps. Finally the remaining characters are connected by a text line aggregation step.

A different approach to text localization is proposed by Jin Ma, Weiqiang, Ke Lu and Jianshe [3] that focuses on the pruning of MSER and linkage trees. MSERs are first extracted in R-G-B and L-A-B channels. For each channel, MSER possesses two polarities depending on whether it is darker or brighter compared to its surroundings based on which an MSER tree is formed. Nodes with maximal score are selected into the set of candidates where all its ancestors and off springs are pruned out. Each node of a MSER tree is considered as a sub tree that contains text components with high probability of belonging to the same text line. The sub trees whose root text elements having positive linkage and high confidence are merged together to form linkage trees. The linkage tree pruning is done via an SVM (Support Vector Machine) classifier trained with a set of features calculate the confidence score, based on which all non-root nodes are eliminated. All remaining root nodes refer to candidate text

lines. Also a filtering is done using CNN feature and intensity contrast.

A method that is useful in fine grained classification of business places and logo retrieval is proposed by Sezer, Ran , Theo and Arnold [4] that combine the textual cues along with visual cues. MSER and also a text saliency map created from background subtraction and binarization are the major methods used here for text detection. MSER and text saliency map are complementary to each other. Each one helps to overcome the disadvantages of the other. Character candidates are also computed in different color spaces using a variety of invariant properties. Character filtering for removing non-character regions uses features such as aspect ratio, solidity, size and contrast. Each of the character detection algorithms and color spaces produced word box proposals are combined and provided for word recognition into the convolutional neural network classifier.

Chayut and Karn [5] proposed a region based text detection technique that can be applied for multiple languages. MSER is used here for performing this task. The grouping of MSER components is done based on a new rule for connectivity. A chain of constraints and double-threshold scheme is used for classifying these MSER groups. A short circuit based evaluation is used for complexity reduction of the extraction process. Based on all these steps text regions are classified into three classes as those of high confidence, low confidence and non-text regions. This method works very efficiently in detecting text without having any language, camera view, text alignment constraints.

Text localization is very important for natural scene image and Xuwang , Xu-Cheng, Hong-Wei and Khalid [6] proposed an effective approach to this task. Here MSER is used for extraction of letter candidates. All non-letter candidates are eliminated with the help of some geometric information. An adjacency relation is defined between these letter candidates. All candidate regions are connected components of an undirected graph. This is constructed by grouping similar letter candidates by the help of disjoint set. An AdaBoost classifier is used for recognizing the text regions and it works on the basis of some important region features such as stroke width, number of letter candidates, horizontal or vertical variances, aspect ratio, color and geometry.

3. MSER TECHNIQUE

Maximally stable extremal region (MSER) is a major technique used for text detection. It is used for finding text candidate regions. This method belongs to connected component family. Connected component based analysis is used to detect character candidates by analysis and group them to form connected regions of text. It deals with the arrangement of the intensity values of an image. MSER is designed in such a manner that the text regions are uniformly coloured, i.e., of stable intensity values and they contrast with that of the surrounding.

Table 1: MSER Advantages and Disadvantages

Pros	Cons
Low computational cost	Low performance in blurry images
Donot miss text	Sensitive to character sizes.
Best region detector	Requires tuning for varying styles
Robust against light changes	Orientation effects high
Covariance to adjacency preserving	Low performance in high contrast
Invariance to affine transformation	Poor results on strong illumination
Multi-scale	Shadows and reflections affect performance

4. CONCLUSION

Text detection from natural scene images is a research field with numerous scopes and applications. Numerous methodologies exist in this arena to make the job more efficient and to overcome the difficulties such as lighting effects, orientations, varying text size and style and complex background. MSER is one such technique which has found its own role in scene text detection. Several advancements and enhancements are being applied to improve the results of MSER. Some of the latest works that employed MSER for text detection from scene images are discussed and a detailed understanding about the method and its pros and cons are also presented in this paper.

REFERENCES

[1] Houssem Turki, Mohamed Ben Halima, Adel M. Alimi, "Text Detection based on MSER and CNN Features", 14th IAPR International Conference on Document Analysis and Recognition, 2017.

[2] Leibin Guan, Jizheng Chu, "Natural Scene Text Detection Based on SWT, MSER and Candidate Classification", 2nd International Conference on Image, Vision and Computing, 2017.

[3] Jin Ma, Weiqiang Wang, Ke Lu, Jianshe Zhou, "Scene Text Detection Based on Pruning Strategy of MSER-Trees and Linkage-Trees", Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2017.

[4] Sezer Karaoglu, Ran Tao, Theo Gevers and Arnold W. M. Smeulders, "Words Matter: Scene Text for Image Classification and Retrieval", IEEE Transactions on Multimedia , 2016.

[5] Chayut Wiwatcharakoses, Karn Patanukhom, "MSER Based Text Localization for Multi-language Using Double-Threshold Scheme", International Conference on Industrial Networks and Intelligent Systems (INISCom), 2015.

[6] Xuwang Yin, Xu-Cheng Yin, Hong-Wei Hao, Khalid Iqbal, "Effective Text Localization in Natural Scene Images with MSER, Geometry-based Grouping and AdaBoost", 21st International Conference on Pattern Recognition, 2012.