ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.11

An Automated Text Detection Algorithm for Unstructured Scene

Narra Teja Sri[1], G. Madhusudhana Rao[2], T.Pravallika[3], P.Jayababu[4]
M Tech Scholar[1], HOD[2], Assistant Professor[3, 4]
Department of ECE
Nannapaneni Venkat Rao College of Engineering and Technology, Tenali, India

Abstract:
This paper presents an automated scene text detection algorithm based on Stroke Width Transform (SWT), Maximally Extremely Regions (MSER) and candidate classification. Firstly, utilize the SWT and MSER to extract the candidate characters at the same time. Secondly, preliminary filtering the candidate connected components based on heuristic rules. Thirdly, using mutual verification and integration to class all candidates into two categories: strong candidates, weak candidates. If the weak candidate has similar properties with strong candidate, then the weak candidate is changed into strong candidate. Finally, the text area is aggregated into text lines by text line aggregation algorithm. The experiment results on public datasets show that the proposed method can detect text lines effectively.

Keywords: Text detection; connected component; Maximally Extremely Regions (MSER); Stroke Width Transform (SWT), candidate classification

I. INTRODUCTION

Scene text detection application in many ways, such as intelligent traffic systems, unmanned navigation, help the blind, the translation system, web filtering and content-based video retrieval. In summary, the study of natural scene text detection obvious plays an increasingly important role in people's daily lives. Although scenes text contains a lot of important semantic information, but due to a natural scene has a lot of non-text interference, which make the task more challenging. In recent years, text detection in natural scenes has been widely studied, and has made many achievements. In the past, both the sliding window method and the connected component analysis method are used in the text detection of the scene. The method of sliding window is used to detect the text in all the positions of the moving window, although it can get good results, but the computation is too large and there are many false alarms. On the other hand, the connected component based method fust separates the text from the background and then refme candidate by morphological constraints or machine learning methods. However, common constraints used to refine candidates some limitations are considered, resulting in low recall in practice [1]. In this paper, we present an efficient scene text detection method utilizing both the advantage of above two methods

II. RELATED WORK

There are a variety of text localization theories, though the most common approaches involve three key procedures [2]: character candidate detection, character classification and text merging. Scene text detection algorithrns can be roughly separated into two categories according to their character candidate detection technique: sliding window based methods and connected component (CC) based methods. In sliding window based methods, the text and the no text regions are distinguished by texture analysis. This method is usually used the sliding window to scan multi scale decomposition of the image, and then the characteristics of connectivity in different regions are analyzed, such as multi-scale wavelet decomposition coefficients [3], HOG characteristics [4], and different gradient edge features [5]. Then with different classifiers such as Support Vector Machine [6]. The key problem of this method is that when the resolution of the image is increased, the size of the window is increased, and the more complex the classification method is, the higher the computational complexity. The connected component based methods utilize the difference between the text and the background to extract the connected regions, and then uses the heuristic rules such as the aspect ratio, the size and the geometric features to filter the non-text connected regions. In recent years, more and more attention has been paid to those methods, which has gradually become the mainstream method in the field of text detection. Among this approach, SWT [8] and MSER [9] are most widely used basic detection algorithrns because of their efficiency and stability. What's more, it becomes the basis of the related text detection research. Neumann and Matas [lO] optimized maximally stable extrem al region (MSER) method to extract candidate connected region, and combined with the characteristics of morphological processing of text characters, and got a good result through the efficient search algorithrn. Huang et al. [11] proposed an algorithm which recall rate is improved by combining MSER with color clustering. Yin et al. [12, 13] proposed an MSER pruning algorithm to increase precision, and obtain the 1st place of the 2013 ICDAR Robust Reading task. It can be seen from the above that the MSER algorithm has been shown to be a state-of-the-art text detection method [14].

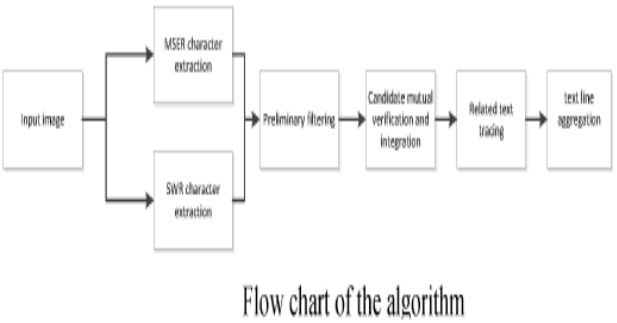


Figure.1. Flow Chart of the algorithm

Different from MSER-based research ideas, Epshtein et al. [8] proposed the SWT-based text detection method based on the texts in the same image have similar stroke widths. This method fIrst detect edge by canny detector, then the image is converted to stroke width map, and then pixels with a similar stroke width are grouped into connected region, then use morphological removing non-text region, fInally merged the candidate regions to text line. Employed SWT to extract candidate, and searched k-nearest neighbors of each CC and bu ild search path as a trace of text string in arbitrary directions to identify arbitrary orientations text strings.

However, machine learning, especially convolutional neural networks [16, 17] cannot be ignored. Recently, convolutional neural networks outperform heuristic rules because of parameters and thresholds automatic inference from training data. Detect candidate character pixels using convolutional features, then using edge and color features to extract connected components, achieved the state-of-the-art results in ICDAR 2015. Zhang et al. [18] utilized a Fully Convolutional Network (FCN) to predict the candidate text regions, and then candidate text lines separated into candidate characters by MSER algorithm, finally another FCN c1assifIer is used to predict the centroid of each character, in order to remove the false hypotheses. However, this requires a lot of hardware consumption and time consumption, and before the training of data marking requires manual operation, different training parameters will lead to different results. Analyzing the characteristics of the natural scene text, we found that the stroke width and gray contrast are important in the detection phase. In the text line, the contrast is relatively small, but the difference between the text and the background is significant.

Therefore, we consider the use of state-of-the-art algorithm MSER to detect text regions. But to some extent, the recall rate is not enough, Therefore, we use SWT and MSER to detect connected regions, and then through mutual verification and merging, the text is divided into strong candidate and weak candidate, thus improve the recall rate significantly. Then they are merged into the text line by fully connected algorithm. The low chart of the algorithm is shown in Figure1.

III. ALGORITHM DETAILS

A. Character Candidate Extraction MSER is an affme feature extraction algorithm proposed by Neumann [10] et al. In MSER algorithm, the image is converted into gray image firstly, and then the image is converted into aseries of binary images by using the continuous threshold range from 0 to 255. With the increase or decrease of gray threshold, there is a region of constant occurrence, and the variation of the two thresholds in the region is considered stable in a certain range. Mathematical defInitions are as folIows: The defmition of image I is the mapping of the region D on the gray S, I: D E Z2 ~ S. Among them, the S can meet the gray level full sequence structure. The relationship between adjacent pixels is defIned as: A cD x D. The region Q c D can be defmed as a subset of satisfy connected region criterion, which means for any pixel p, q E Q there are more than one path to connect p and q. The following formula will illustrate the connected region criterion.

$$pAa_1, a_1Aa_2, ..., a_nAq$$

Where ai E Q, i = 1,2, ... , n . The defmition of the boundary BQ is as follows: BQ is adjacent to at least one pixel in the Q, and BQ does not belong to the region Q.

$$\partial Q = \{q \in D - Q, \exists P \in Q, aAp\}$$

For Vp E Qand Vq E BQ, if I(p) > I(q)then Q is the maximum region. For a series of extremal regions, if change rate q(i) of the region is at the local minimum, it is considered to be the maximally stable extremal regions. q(i) defmed as follow:

$$q(i) = \frac{|Q_{i+\Delta} - Q_{i-\Delta}|}{|Q_i|}$$

After obtaining MSER, the connected component analysis is applied to detect the candidate characters, candidate detection results show in Figure2. We will also introduce the SWT method in the following paper.

Connected component extraction results of MSER

Figure.2. Connected Compont Extraction Results of MSER

The text is made up of strokes, which consist of two parallel edges. The SWT algorithm is based on the fact that texts in the same text line usually have similar stroke width. Therefore, the canny operator is firstly used to detect the edge of the image [19] in the SWT algorithm, and then get the corresponding edge response, fInally calculate the distance between the two parallel edges in certain conditions. This method can effectively detect the text line, and has strong fault tolerance ability. It detects stroke pixels by looking for the relative edge pixels q from the beginning of the pixel p along the gradient direction of the d pq' as shown in the Figure. Only when the gradient direction of the edge pixels is opposite to each other, the ray is considered to be effective. All through the ray path pixel has the same stroke width, which is the distance between two parallel edges. Stroke width calculation method is as folIows: Assume that the bottom left corner of the edge image is the origin of the co ordinate axis, then the coordinates of the edge pixel p is (x po' Y po) , the gradient direction is B p. Then from the starting point p along the gradient direction for ray representation:

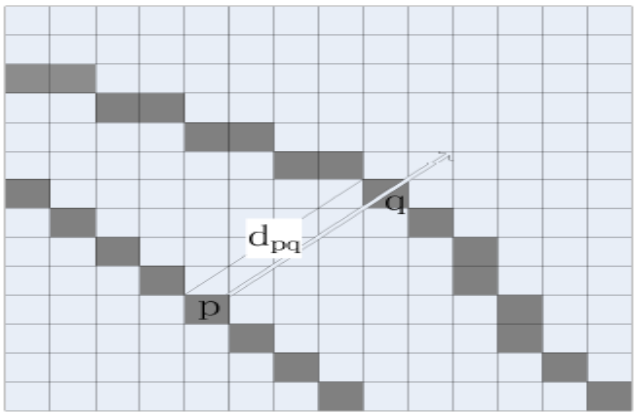
$$\lambda^b = \text{tg}\,\theta^b(\chi^b - \chi^{b0}) + \lambda^{b0} \quad , \quad \chi^b > \chi^{b0}$$

Search along the ray direction until you find the next edge pixel q, consider the relationship between B p and gradient direction Bq of pixel q, If the following constraints are satisfied:

$$\mathbf{abs}(\theta_p - \theta_q) \le \pi \pm \frac{\pi}{6}$$

In addition, if the distance is the minimum value, then d pq is the stroke width. If you do not fmd the matching pixel q from

the p point, or the gradient direction does not meet the criterions, the ray is discarded.



Stroke width schematic

Figure.3. Strocke width Schematie

The algorithm compares the stroke width between two adjacent pixels, and if there is a similar value two pixel will be aggregation. In order to ensure strokes can be aggregated together in this paper, the SWT ratio threshold is set to three. In order to ensure that the text can be accurately detected in the dark background and bright background, we should search from two opposite directions. Text detection results using the SWT method are shown in Figure4.

Candidate extraction results of SWT

Figure.4. Candidate Extaction Results Of SWT

B. Pre/iminary Filtering

Based on Heuristic Ru/es After the connected region is extracted, there are still a lot of regions that do not belong to the text. Therefore, some heuristic rules are needed to filter out the regions that do not contain text. Heuristic rules include some simple geometric features and statistical features of connected regions, which can be used for fast screening of candidate text regions. The range of a connected component is a minimal rectangle Rcc which containing all the pixels in the connected component. Where the Rcc size is m * n, the mean and standard deviation of stroke width is f.1 anda, respectively.

$$\mu = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{SWT}(i,j)$$

$$\sigma = \sqrt{\frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} (I_{SWT}(i,j) - \mu)^2}$$

For the initial screening of connected component, the following rules:

• The aspect ratio of the connected component, a and ß are aspect ratio constraint. In this paper a =6 and ß =116. Most of extremely large and extremely small aspect ratio will be filtered by this process.

$$\alpha < \frac{n}{m} < \beta$$

Connected region size constraint, most of extremely small connected component will be filtered by this process

$$n \times m > 40$$

Stroke width variance, var, l1ax is maximum stroke width variance constraint.

$$0 < \frac{\sigma}{\mu} < \mathrm{var}_{\max}$$

After this step, a large of part of obvious non-text candidates is filtered out.

C. Candidate Mutual Verification and Integration

After the previous operation, the remaining candidates are SWT and MSER candidate. They are based on different advanced text detection theory, there are a lot of overlapping parts, but also independent of each other. Therefore, we try to divide the all candidate into two categories by candidate mutual verification and integration. Assuming that the SWT method detects nSWT candidate regions, MSER method detects nMSER candidate regions. The candidate detected by the SWT method will be initially marked as SWT or MSER. Each region is represented as a set of pixels. Mutual verification and integration aims to get a strong text area and weak candidate. At this stage, the coincidence rate 0 between each SWT region and each MSER region is ca1culated, if the merge threshold 0T is satisfied, it is marked as a strong text area, else mark it as a weak text area. According to the experimental experience, 0T is set to 0.6. Coincidence rate 0 ca1culation method is as folIows.

$$\partial(i,j) = \frac{R_{SWT(i)} \bigcap R_{MSER(j)}}{R_{SWT(i)} \bigcup R_{MSER(j)}}$$

D. Related T ext Tracing

We classify strong text as the result because of their high confidence level. However, the weak text may be true to the text, potentially resulting in lower recall. As a result, weak text labels are converted to strong text when the weak text and strong text have similar properties. In order to achieve a high recall rate, we start tracking its adjacent weak text Rw from each strong text Rs. Where Rw shares similar properties with Rs' we change its mark into strong candidate and re execute the above operations. We use the following properties:

• The Rw and Rs should be close enough in space. The distance between them is less than twice of diagonal length of Rs.

• The Rw and Rs should be similar in size and scale. When the width and height difference of is Rw and Rs less than the half of Rs ' then them are considered belong to same text !ine.

• The ratio of theirs stroke width should be in the range of 0.5 to 2, since the width of the text strokes in the same text line is similar.

E. Text Line Aggregation

Through the previous steps, we can obtain credible reliable characters. The purpose of this stage is to aggregate the remaining candidate into text lines, fortunately, the main advantage of OUf algorithm is easy merging candidate to text !ine. We only need to use the similar conditions in the previous related text tracing to fmd the text line

IV. EXPERIMENT AL RESUL TS

In order to evaluate the effectiveness of the proposed method, the proposed algorithm is tested on the data base, which contains 229 training images, 233 test images. The method is implemented on the platform of software Matlab. The

algorithm is robust to text size, contrast, and sharpness. In order to further verify the effectiveness of the algorithm, we adopted the evaluation method proposed by Wolf and Jolion [20] to compare our method with other methods. The advantage of this algorithm is that it uses SWT, MSER and candidate mutual verification and integration that reduce the missing text and improve the recall rate. Moreover, the precession rate is greatly unproved by the initial filtering and text line aggregation.

Step 1: Load image

Step 2: Detect MSER Regions

Step 3: Use Canny Edge Detector to Further Segment the Text

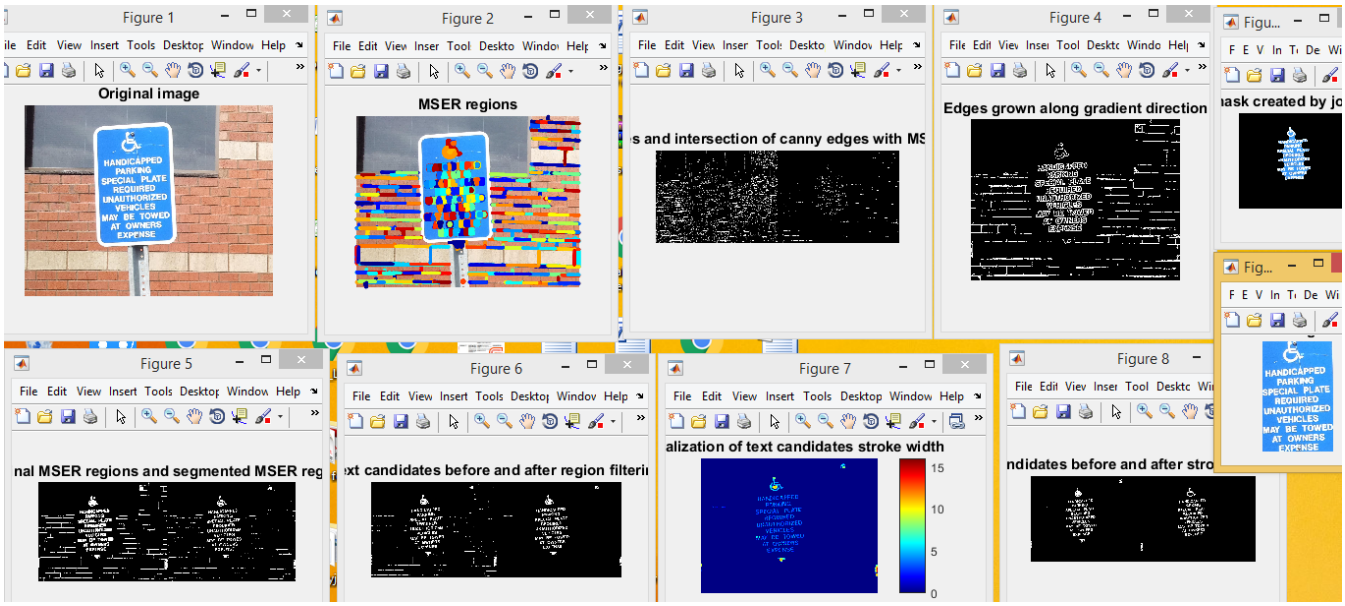
Step 4: Filter Character Candidates Using Connected Component Analysis

Step 5: Filter Character Candidates Using the Stroke Width Image

Step 6: Determine Bounding Boxes Enclosing Text Regions

Step 7: Perform Optical Character Recognition on Text Region

Step 8: Apply the Text Detection Process to Other Images



Some detection examples of the proposed method in the ICDAR 2013 dataset

Figure.5. some Detection Examples of the proposed method in the ICDAR

V. CONCLUSTON AND F UTURE WORK

In this work, an automated method based on MSER, SWT and a character classification method is presented. The experiments on public datasets demonstrate the availability and effectiveness of the proposed methods. Future work will aim to use a more intelligent approach to achieve a more robust result to adapt to various environments and languages.

VI. REFERENCES

[1]. H. Cho, M.Sung, and B Jun,"Canny Text Detector: Fast and Robust Scene Text Localization Aigorithm," Proc. International conference on Computer Vision and Pattern Recognition, 2016, pp. 3566-3573.

[2]. Q. Ye, Q. Huang, W. Gao, and D. Zhao, "Fast and robust text detection in images and video frames,".lmage and vision Computing, 23(6), pp. 565-576, 2005.

[3]. Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," IEEE Trans. Pattern Analysis Machine Intelligence, 37(7), pp. 1480- 1500, July 2015.

[4]. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 886-893.

[5]. P. Shivakumara, W. Huang, T. Phan Quy, and C. Tan Lim, "Accurate video text detection through classification of low and high contrast images, ".Pattern Recognit, 43(6), pp. 2165-2185,2010.

[6]. K.l. Kim, K. Jung, and J.H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE Trans. Pattern Anal. Mach. IntelI, 25 (12), pp. 1631- 1639, 2003.

[7]. 1.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch, "Adaboost for text detection in natural scene," Proc. ICDAR 2011 , pp. 429-434, 2011.

[8]. B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width trans form , "Proc. IEEE International Conference on Computer Vision and Pattern Recognition.2010,2963- 2970.

[9]. J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust widebaseline stereo from maximally stable extremal regions," Image and Vision Computing, 22(10):761- 767, 2004.

[10].1. Neuman, 1. Matas, "Real-time scene text localization and recognition," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 3538-3545. 30

[11]. X. Huang, T. Shen, R. Wang, and C. Gao, "Text Detection and Recognition in Natural Scene Images," Proc.of IEEE Conference on Estimation, Detection and Information Fusion, 20 15, pp. 44-49. [12] [13] [14]

[12]. X. Yin, W. Pei, J. Zhang, and H. Hao, "Multi-orientation scene text detection with adaptive clustering," IEEE Trans Pattern Analysis Machine Intelligence, 37(9), pp. 1930-1937, Sept 2015.

[13]. X. Yin, x.-c. Yin, H.-W. Hao, and K. Iqbal, "Effective text localization in natural scene images with mser, geometry based grouping and adaboost," International Conference on Pattern Recognition (ICPR), pp. 725- 728, Nov 2012.

[14]. K. Mikolajczyk, T. Tuytelaars, and C. Schmid, "A Comparison of Affine Region Detectors," International Journal of Computer Vision, 2005, vol 65, pp. 43-72.

[15]. Y. Zhang, J. Lai, P. C. Yuen ,"Text string detection for loosely [16] [17] [18] [19] [20] constructed characters with arbitrary orientations," Neuro computing, 168(20 15),970-978.

[16]. T. Wang, DJ. Wu, a Coates, and AY. Ng, "End-to-end text recognition with convolutional neural networks, "2012 21 st International Conference on Pattern Recognition, Tsukuba, 2012, pp. 3304-3308.

[17]. S. Zhu and R. Zanibbi, "A Text Detection System for Natural Scenes with Convolutional Feature Learning and Cascaded Classification" Proc. IEEE Conference on Computer Vision and Pattern Recognition,20 16,

[18]. pp. 625-632. Z. Zhang, C. Zhang, and W. Shen, "Multi-Oriented Text Detection with Fully Convolutional Networks," Proc. IEEE Conference on Computer Vision and Pattern Recognition, 20 16,

[19]. pp. 4159-4167. D. Karatzas, F. Shafait, S. Uchida, ICDAR 2013 robust reading competition [A].Proc. of IEEE International Conference on Document Analysis and Recognition. 2013,

[20]. 1484-1493. C. Wolf and J.-M. Jolion, "Object countlarea graphs for the evaluation of object detection and segmentation algorithms," Int. J. Document Anal. Recognit, 8 (4), 2006, pp. 280-296.