NOVATEUR PUBLICATIONS
International Journal of Research Publications in Engineering and Technology [IJRPET]
ISSN: 2454-7875
VOLUME 3, ISSUE 6, Jun.-2017

ROBUSTDETECTION OF TEXT IN NATURAL SCENE IMAGES

SALONI S. AGRAWAL
Electronics and Telecommunication Khurana Sawant Institute of Engineering and Technology Hingoli, India
saloniagrawal0910@gmail.com

PROF. G. UBALE,
Electronics and Telecommunication Khurana Sawant Institute of Engineering and Technology Hingoli, India

ABSTRACT:

Detection of text and identification in natural scene images has applications in computer vision systems such as license plate detection, automatic street sign translation, image retrieval and help for visually challenged people. The text images has complex background, blur, occluded text, different font-styles, and noises in image and variation in illumination. Hence, scene text recognition puts forth challenges in computer vision. Hence, a potent method based on Maximally Stable External Regions (MSER) has been used as described in this paper. Here, the text characters are clustered, separating them from high probable non-text characters with the help of text categorizer. The algorithm is then verified by testing it on images based on the predefined rules.

KEYWORDS: MSER; edge detection; thresholding; image processing.

I. INTRODUCTION:

Detection of text means obtaining user required regions in the image and recognition of those regions with the help of the user. Various techniques are implemented to remove the non-text regions from the interested areas with highest probability of finding the text. Hence, the steps involved in detection of text and especially in natural scene images need extra tools in obtaining actual image before applying the text recognition algorithms. Challenges appear in the natural images due to complex background, variation in font type, size, color, texture and orientation. Hence, a potent detection of text must be used in the images before retrieval and recognition of text. For such required potent detection of text, renown MSER approach has been implemented.

The detection of text methods of this decade mostly start with a common pipeline with identifying individual components corresponding to the characters and then group or merge them to words. Typically, components are extracted as maximally stable external regions [1] or stroke width components [2] and classified as character or non-character. The method used in [3] and [4] are reliance on unsupervised image processing methods which were used with certain assumptions on the quality of text present. The methods assumed text to be contained in high contrast regions, generating edges of high intensity and single colored ensuring the respond uniformly to the image processing methods employed. These methods, generally classified as texture based, had a pipeline which started with a segmentation step, responsible for generation of an image with text areas having higher intensities. It was followed by a detection step which grouped the high intensity regions. The final step was a verification step which classified the candidate regions as text or non-text.

In last few decades, several detections of text in scene images methods were proposed by Neumann et al. which were highly successful due to MSERs [5]. The concept of MSERs i.e. connected components on grayscale image was exploited to generate candidate character regions potently. The method discussed by Pan et al. [6]describes detection of text regions and extraction of connected components as character regions by local binarization.The non-characters are eliminated with a Conditional Random Fields model [8], and characters are finally be grouped into text. The MSER algorithm detects many repeating components and most of these repeating components, apart from the character components, need to be removed before further processing.

In this paper, we propose a potent and precise MSER based detection of text in scene images method. First, by going through the hierarchical structure of MSERs and considering simple features, we design an accurate and fast MSERs cut short algorithm. The number of character regions to be processed is significantly reduced with a high accuracy.

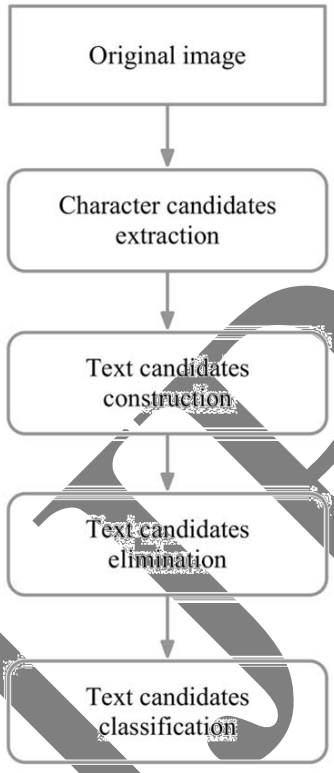
II. ROBUST SCENE DETECTION OF TEXT:

In this approach, a potent and very precise method based on MSER detection of text in scene images has been discussed. Initially, analysis of the hierarchical structure of MSERs has been made and acquiring some simple features, fast and precise MSERs cut short algorithm has been developed in this procedure. The character regionsare very significantly decreased to detected text characters of image with high accuracy. A self-training learning algorithm having an ability to learn distance weights is implemented. Also, it clustersthe threshold automatically when an image is given as an input for recognition. Candidates characters are clustered or grouped into text regions through a

clustering algorithm with the help of the already learned parameters as computer program. At next step, character categorizer is used to estimate the subsequent probabilities of input image. Text regions corresponding to the existing non-text regions in the image are described in this process. The algorithm removes the text regionswith high non-text probable regions within the image. These kind of elimination helps for training a more accurate, efficient and reliable text categorizer for identifying and recognizing text in the image. Lastly, an precise and potentdetection of text in scene images system has been built.

By applying various key improvements over traditional MSER-based methods, a novel MSER-based detection of textin scene images has been developed. The proposed detection of text in scene images method includes the following stages:

Fig. 1. Flowchart of the system used and results after each



step of the sample.

A. EXTRACTION OF CHARACTER REGIONS FROM THE INPUT IMAGE:

From the image, Character candidates are extracted by applying the MSERs algorithm. The components which are repeated are decreased by MSERs algorithm using the minimizing regularized variations.Repeating components is the major pitfall when applying the MSER technique as a character segmentation algorithm. The stratified structure of MSERs is useful for designing a cut short algorithm. Based on the lower variation between the parent or child, MSERs which having less variation with sharp borders are mostly characters.

B. TEXT CANDIDATE CONSTRUCTION:

Text regions are constructed from received results. The metric learning algorithm is used to learn the distance weights and grouped threshold of the detected text. Character regions are grouped into text regions through a single-link clustering algorithm. Such grouping produces groups that are elongated and is particularly better for the text region construction. Single-link clustering belongs to the ordered family of clustering i.e. in hierarchical clustering, each input data is taken as a single group and they are successively merged till all points have been merged into a single group. In single-linkclustering, the two groups having smallest distance between the two closest groups are merged together. A threshold is set such that the grouping process is terminated when the threshold is exceeded by the distance between nearest clusters.

C. TEXT REGIONS ELIMINATION:

Most of the non-text regions are needed to be removed before training the categorizer. Hence, it is not easy to train an effective text categorizer using such an unbalanced database. We propose the use of a character categorizer to judge the subsequent probabilities of text regions corresponding to non-text and remove text regions with more non-text probabilities. Some features used to train the character categorizer are: height of text region, width, smoothness, aspect ratio and stroke width features. Characters with lesser aspect ratios such as “i”, “j” and “l” are labeled as negative samples.

D. TEXT REGIONS’S CLASSIFICATION:

Text regions which correspond to real texts of image are identified by using the text categorizer. A categorizer is trained to finalize whether in an image the text regionis corresponding to the true text or not.

III. STEPS INVOLVED:

A. DETECTION OF TEXT REGIONS USING MSER:

Fig. 2. MSER Region in multi-colour

The MSER feature detector works well for finding text regions. The constant color and high contrast of text from

the background leading to stable intensity profiles makes it work well for the text.

B. USE OF BASIC GEOMETRIC PROPERTIES FOR REMOVAL OF NON-TEXT REGIONS:

Although the MSER technique picks most of the text, it detects many stable regions in the image too that are not text. A rule-based approach is used to separate these non-text regions. For example, to filter out non-text regions, geometric properties of text are used using simple thresholds. Alternatively, to train a text against non-text a machine learning approach categorizer can be used. A combination of the two approaches can produce better results. The several geometric properties that can be used for discriminating between text and non-text regions can be aspect ratio, eccentricity, Euler number, extent and solidity.

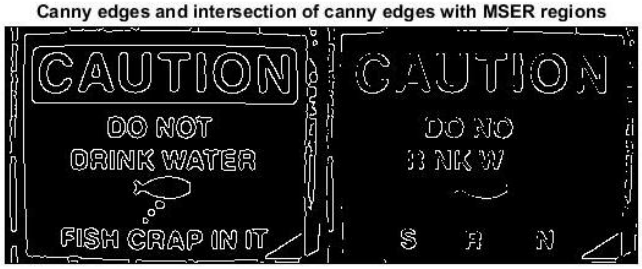


Fig.3. Canny edges and intersection of canny edges with MSER region.

Use of stroke width variation for removal of non-text regions Stroke width can be another parameter to distinguish between the text and the non-text regions. Stroke width is a measure of the width of the curves and lines that make up a character. Non-text regions are likely to have larger stroke width variations while text regions have little. Then we estimate the stroke width of one of the detected MSER regions. This can be done by using a distance transform and binary thinning operation.

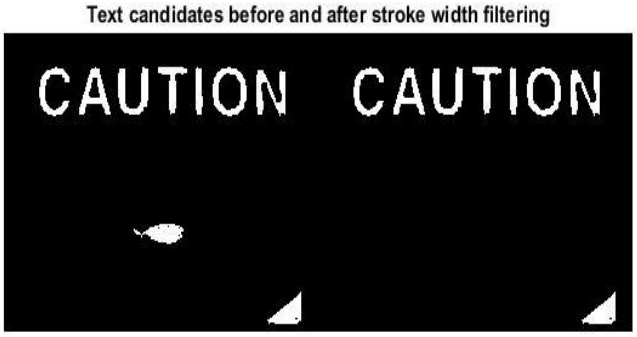


Fig. 4. Text candidate before and after stroke width filtering

C. MERGE OF TEXT REGIONS FOR FINAL DETECTION:

Till now, we have results composed of individual text characters. Now, these individual letters must be merged to recognized as a single word and then to line. This helps in recognition of the actual words in an image than just the individual characters. For example, recognizing the string 'EXIT' vs. the set of individual characters {'X','E','T','I'}, where the meaning of the word is lost without the correct ordering.

D. TEXT RECOGNITION:

Finally, the text is recognized using template based technique.

Fig. 4. Final image after text region detection and masking with original image.

CONCLUSION:

This paper presents the MSER method for detection of text in scene images with several techniques. A swift and precise MSERs cut short algorithm allows detecting most characters even in low quality images. Also, novel distance metric technique where distance between the text regions is used as a parameter to distinguish between the text and non- text regions, is used. Now, the individual text regions are grouped together to form the meaningful words. The text categorizer eliminated the the non-text regions compared with the text regions.Asystem with potent detection of text in scene images exhibits superior performance over state-of-the-art methods on a variety of public databases.

For the further research in this method we put forth several limitations. It is still difficult to detect highly blurred texts in images from scenes which are of lower resolutions. Second, for different language text parameter such as geometric parameters change like in English and in Chinese. Third, the highly-skewed distortion of the text with multiple orientations needs further research. Best results have been observed with bright font colors (E.g. white) distinguished from the background.

ACKNOWLEDGMENT:

I would like to express my deep sense of respect and gratitude towards my advisor Dr.R.R. Sawant and guide Prof. G.B. Ubale Head of Electronics & Telecommunication Department of our institute, who has been the guiding force behind this work. I am greatly indebted to him for his constant encouragement, invaluable advice and for propelling me further in every aspect of my academic life.

REFERENCES:

1) J. Matas, O. Chum, M. Urban, and T. Pajdla. *Potent wide-baseline stereo from maximally stable extremal regions.* Image and Vision Computing, 2004.
2) B. Epshtein, E. Ofek, and Y. Wexler. *Detecting text in natural scenes with stroke width transform*. In CVPR, 2010
3) D. Doermann, J. Liang, and H. Li. *Progress in camera-based document image analysis*. In ICDAR. IEEE, 2003.
4) V. Wu, R. Manmatha, and E. M. Riseman. *Textfinder: An automatic system to detect and recognize text in images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999.
5) L. Neumann and J. Matas. *Text localization in real-world images using efficiently pruned exhaustive search.* In ICDAR, 2011.
6) Y.-F. Pan, X. Hou, and C.-L. Liu, *"A hybrid approach to detect and localize texts in natural scene images*," IEEE Trans. Image Process., vol. 20, no. 3, pp. 800–813, Mar. 2011.
7) A. Shahab, F. Shafait, and A. Dengel, *"ICDAR potent reading competition challenge 2: Reading text in scene images,"* in Proc. ICDAR, 2011, pp. 1491–1496.
8) J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "*Conditional random fields: Probabilistic models for segmenting and labeling sequence data*," in Proc. Int. Conf. Mach. Learn., San Francisco, CA,USA, 2001, pp. 282–289.