International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 15 (2017) pp. 5163-5168
© Research India Publications. http://www.ripublication.com

Scene Text Extraction by Combining Edge Based Stroke Segmentation and Morphological Filtering

Kumuda.T[1] and L.Basavaraj[2]

[1]Department of Electronics & Communication Engineering, AIT, Chikamagaluru-577102, Karnataka, India.

[2]Department of Electronics & Communication Engineering, ATME, Mysore-18, Karnataka, India

[1]0000-0003-4623-8965

Abstract

Extraction of text embedded in scene images has many applications, such as license plate recognition, content based retrieval, text based indexing etc. As the appearances of the text in images are unpredictable, separating text pixels from the background is a challenging one. A new frame work is introduced here which uses edge enhanced MSER along with connected components for text extraction. Complementary features of MSER and sobel edge methods are combined to obtain the better results. Feature vector obtained from the stroke width transform are used to recover the text connected components. Further morphological operations along with heuristic filters are used to separate the text pixels from the background. Validation of the method proposed is tested on ICDAR datasets and also on own dataset. The experimental outcome indicates that the suggested algorithm extracts the text characters efficiently even when there is a variation in font, size, color, under different conditions such as low resolution, complex background, varying illuminations etc.

Keywords: Scene text; edge detection; SWT; MSER; Morphological operation; Text extraction.

INTRODUCTION

In last one decade tremendous research work is going on in text analysis from images. This is due to the fact that scene text extracted from images can be used in many real time applications. Text Extraction from image is concerned with extracting the relevant text data from the images. Scene texts occurs naturally as a part of the scene image and contain important semantic information such as advertisements, names of streets, institutes, shops, road signs, traffic information, board signs, nameplates, street signs, bill boards, banners ,text on vehicle etc. Scene text extraction can be used in detecting text-based landmarks, vehicle license plate detection/recognition, keyword based image search, identification of parts in industrial automation, content based retrieval etc. However, text extraction from scene images is a challenging task due to the wide variety of text appearances, such as variations in font and style, geometric and photometric distortions, partial occlusions, and different lighting conditions, blurring effects of varying lighting, alignment & complexity of background.

Many methods have been developed for analysis of scene text from images. According to the features utilized, all these methods can be categorized into four types. They are connected component, edge, color and texture. CC-based methods [1-3] use a bottom-up approach by grouping small components into successively larger components until all regions are identified in the image. A geometrical analysis is needed to merge the text components using the spatial arrangement of the components so as to filter out non-text components and mark the boundaries of the text regions. Among the several textual properties in an image, edge-based methods [4-6] focus on the 'high contrast between the text and the background'. The edges of the text boundary are identified and merged, and then several heuristics are used to filter out the non-text regions. Color based methods [7-9] works under the assumption that, for easy detection all the characters in images will have the same color. This method groups the similar color pixels into single region, and then differentiates them as text and background pixel groups. Texture-based methods [10-12] use the observation that texts in images have distinct textural properties that distinguish them from the background.

Each method has its advantages and disadvantages. Some researchers have used hybrid approach of combining different methods, so that merit of each method can be used in overcoming the demerits of the other method. This type of approach has given very high fruitful results. In this algorithm edge and connected component methods are combined for extracting text from scene images. Even though edge based methods are faster and have high recall rate at the same time it gives high false alarms in images where other objects and background will also have edges same as text. In order to overcome this drawback each edge is detected and filtered in each stage, and then connected component method is used.

Rest of this paper is arranged as follows: Next section gives the detailed explanation of the proposed algorithm. Results and discussions are presented in the following section. Future work and conclusions are given in the last section.

PROPOSED METHOD

Proposed text extraction algorithm mainly consists of three stages: text detection, text region localization and text extraction. In the text detection stage combination of MSER and edge algorithms are used for finding the presence of text in the image. Next using SWT and geometric filtering text regions are localized. Finally in the text extraction stage text pixels are separated from the background using morphological and heuristic filtering. Flow of the proposed method is as shown in "Fig. 1,"

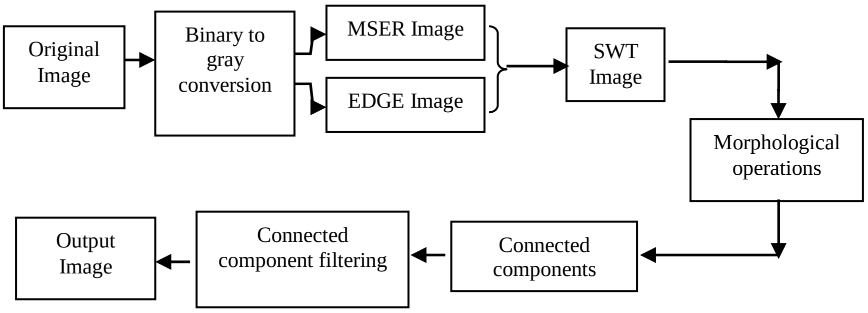


Figure 1: .Block diagram of the proposed method

A. Text Region Detection

MSER based text detection methods have reported promising results for many standard data base images [14-19].As MSERs are extracted from the gray scale image, in pre-processing stage RGB image is converted into gray scale image using "(1),".The algorithm proposed by matas et al [13] is used for extracting MSERs, which groups all the pixels having same intensity over a wide range of thresholds. MSER region contains pixels of same intensity either higher or lower than the outer boundary pixels. MSER regions are identified by first arranging all the pixels in the image according to their gray value, and then pixels are added to every region as threshold value changes. Next each region is analyzed and a region with less variation with respect to threshold is termed as maximally stable. If E_l is the extremely region then the variation of E_l is calculated using "(2)," E_l will be maximally stable extremely region if its variation is stable and more stable than E_{l-1} and E_{l+1} [13], where E_{l-1} is parent and E_{l+1} is child. One of the advantage of MSER is that it works satisfactorily even for the distinct homogeneous boundary regions.

$$y(i,j) = 0.2999R(i,j) + 0.587G(i,j) + 0.224 \quad (1)$$

$$V\,(E_l) = \frac{|E_{l+\Delta} - E_l|}{|E_l|} \quad (2)$$

Even though MSER has reported encouraging result, it is sensitive to image blur [19]. We propose a simple yet effective edge enhanced MSER algorithm which combines the complementary properties of MSER and sobel edge. In the proposed algorithm first MSER regions are extracted[13] then these regions are enhanced using sobel edges obtained from the original gray-scale image.Sobel operator uses the approximation to the derivative to find edges. It identifies those pixels in image which have maximum gradient as the edge pixels. The 3X3 sobel operators used are as shown in "Fig. 2,"

+1	+2	+1
0	0	0
-1	-2	-1

-1	0	1
-2	0	2
-1	0	-1

Figure 2: 3X3 Sobel operators

As shown in "Fig. 3," MSER pixels outside the boundary formed by sobel edges are removed. Further MSER filtering is done based on their geometric properties. Too lengthy and too short components are removed using aspect ratio. Complex contour MSERs are removed using the compactness of region as text won't have complex contour.

B. Text Localization

Text will have parallel edges of constant stroke width compare to other elements; hence from the MSER and edge enhanced MSER image, text regions are recovered using stroke width feature. Many recent works successfully employed stroke width for text detection [25-27] and achieved good results. We used SWT proposed by Epshtein[23], which results in stroke width image of the size same as input image. Epshtein[23] used bottom-up approach, where each pixel ‘p’ is assigned with ‘∞’ stroke width value, then gradient direction d_p perpendicular to the orientation of the stroke is considered. Another edge pixel q roughly opposite to d_p along the gradient exists, then each pixel along [p,q] is assigned with distance value of p and q ($\|p-q\|$).If the pixel is already assigned with a value, then lower among the two value is assigned to the pixel. If there is no pixel q in the opposite direction to d_p, then the ray is discarded. This algorithm faces problem in extracting stroke width value for corner pixels, to overcome this median stroke width transform is applied to each non-discarded ray again.

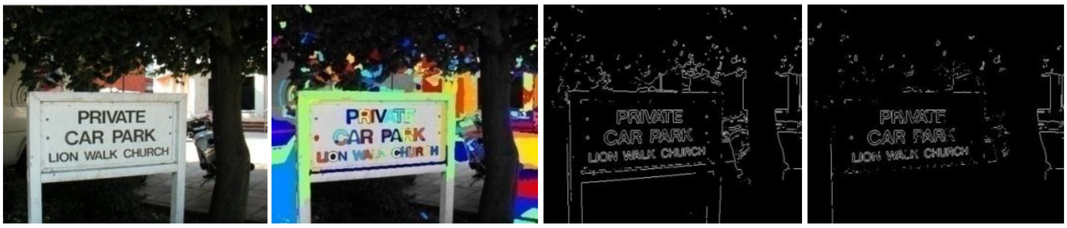


Figure 3: (a) Original (b) MSER (c) Edge Image (d) Edge Enhanced MSER.

Along with SWT ratio we have also included stroke width consistency proposed by [24], which makes use of standard deviation of stroke width. Stroke width maps generated in two opposite directions are named as SWT_1 and SWT_2. Stroke width consistency of pixel ‘p’ and ‘q’ are calculated using “(3), (4),” Finally the smaller among them is taken as the actual stroke width consistency between p and q as shown in equation “(5),”.

$$Cs_1^{p,q} = \sqrt{\frac{1}{len_{p,q}} \sum_{i=1}^{i=len_{p,q}} \left(\frac{SWT_1\left(p+\frac{q-p}{|q-p|}\cdot i\right)}{SWT_1^{p,q}} - 1 \right)^2} \quad (3)$$

$$Cs_2^{p,q} = \sqrt{\frac{1}{len_{p,q}} \sum_{i=1}^{i=len_{p,q}} \left(\frac{SWT_2\left(p+\frac{q-p}{|q-p|}\cdot i\right)}{SWT_2^{p,q}} - 1 \right)^2} \quad (4)$$

$$Cs^{p,q} = \min\left(Cs_1^{p,q} \, , \; Cs_2^{p,q}\right) \quad (5)$$

Next morphological operations are used for eliminating non-text, noisy objects and to retain text-like objects. In many images processing approach mathematical morphology is extensively used [28] for pattern recognition .Text edges are linked together using morphological dilation which merges the close region together, while leaving the isolated region. Size of the structuring element should be selected properly so that minimum number of non-text regions should be merged together. Erosion is used to remove the small non-text regions from the dilated image. By estimating the local region which probably contains text using edge feature, helps in segmenting candidate text components effectively in the next stage. Output of this stage is a localized image as shown in “Fig. 4,”.In the next stage accurate binary characters are extracted, which can be directly used as input to the OCR for recognition.

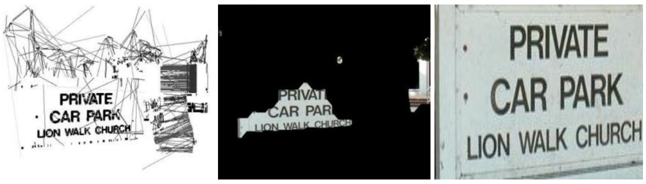


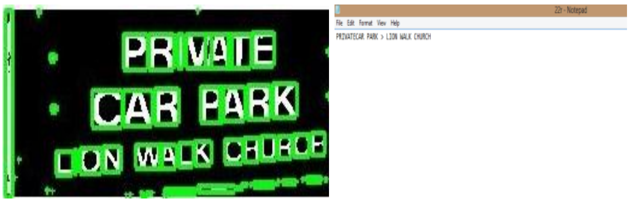
Figure 4: (a) SWT (b) geometric & morphological filtered (c) Localized.

C. Text Extraction

From the morphological filtered image eight-connected components and their bounding boxes are extracted. Text Pixels usually forms the CC with spatial and geometric relationships with the neighboring pixels. Connected component image contains some of the non-text components, which are needed to be removed before using as input to OCR.Text characters will have some proportionality between width and height, hence aspect ratio which is the ratio of height/width or width/height are used to remove too long with respect to width or too wide with respect to length connected components. As text components cannot have complex contour shape, hence complex contour shaped components are filtered using compactness feature. Edge density is used to filter out the CCs with only few edge pixels or too many edge pixels. The feature vectors of all CCs are used by the discriminative functions for filtering non-text components. The output of this stage is the binary image with only text pixels as shown in figure 5a.Which can be used as input to the standard available OCR for recognition. Extracted text and OCR output for the example image is as shown in "Fig. 5,"

RESULTS AND DISCUSSION

Results of the proposed methods are analyzed using standard data base images such as ICDAR 2003 Own database images collected from mobiles, camera, and internet are also used for evaluation. Figure (6-8) depicts some examples of the result ,which shows that even with considerable variations in contrast ,intensity and texture most of the text regions were correctly detected. The main advantage of using MSER is that it works well even for low resolution, noisy as well as low contrast images. Because of multilayer strategy of MSER both small and large characters are detected.



```
PRIVATECAR PARK > LION WALK CHURCH
```

(a) (b)

Figure 5: (a) Extracted text (b) OCR output

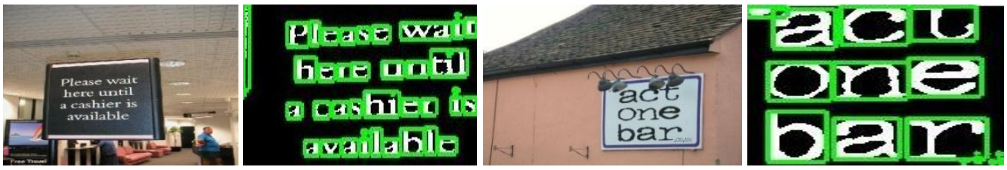


Figure 6: Example results from ICDAR 2003data base images

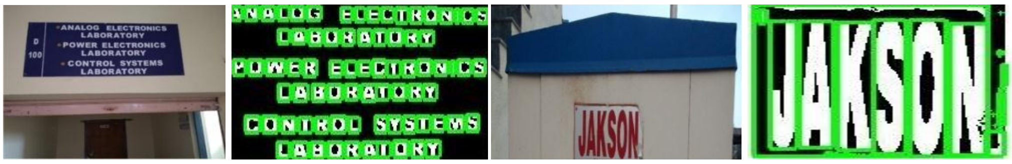

ANALOG ELECTRONICS LABORATORY
POWER ELECTRONICS LABORATORY
CONTROL SYSTEMS LABORATORY
JAKSON

Figure 7: Example results from OWN data base images.

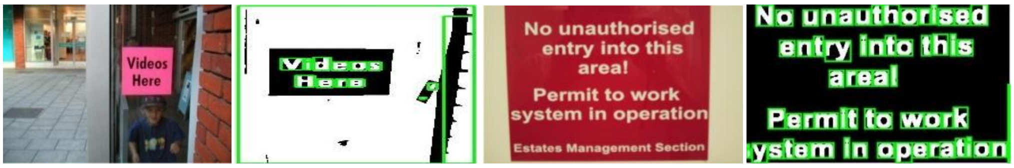

Videos Here
No unauthorised entry into this area!
Permit to work system in operation
Estates Management Section

Figure 8: Example results of the proposed method

The main drawback of MSER is that it detects large number of repeated components. Some researchers are addressed this problem [16, 17], even then also there is scope for getting better of it. MSERs pruning algorithm suggested [21, 22] does the pruning in two stages. First by border energy functions are maximized, results in reduction of linear components. In the second stage, filters in cascade are used with hierarchical filtering. By combining the complementary properties of sobel and MSER better results are obtained. Sobel masks are second degree gradient operator; hence they are much more aggressive in enhancing sharp changes. Stroke widths are effectively used for text detection [25], efficiency of the SWT depends upon the accuracy of the previous edge detection step. All the previous work uses SWT after canny edge but in suggested algorithm SWT is applied to the MSER and sobel edge intersection image. SWT algorithm is applied twice in the opposite directions for each pixel 'p' so that text on both dark as well as light backgrounds are detected .By using two methods advantages of each method is taken and at the same time drawback of these methods are suppressed. Here geometric and stroke width features are effectively used for filtering non-text regions.

Performance of the text extraction algorithms are evaluated using precession rate (PR), Recall rate (RR) [20, 25, 26] as given in equation (6) & (7). PR and RR are high for some example images ("Fig. 6," "Fig. 7,"). This is due to, the algorithm detects the structure such as window frames, bricks, leaves which as sharp edges as text. Performance evaluation comparison of the proposed algorithm with state of art methods are as shown in table I. As this algorithm integrates edge algorithms together in localizing text region, this works well compare to texture or color based technique. Some text were missed ("Fig. 8,") because of their small size or low contrast.

$$\text{precision rate} = \frac{\text{number of correctly detected characters}}{\text{number of characters detected by algorithm}} \quad (6)$$

$$Recall\ rate = \frac{number\ of\ correctly\ detected\ characters}{number\ of\ characters\ in\ image} \quad (7)$$

Table I: Comparisons with the state of art methods.

Method	ICDAR-2003		own	
	PR (%)	**RR (%)**	**PR (%)**	**RR (%)**
Color[8]	91..3	90	90	89.6
Texture([10]	85.8	87.3	91.8	92
Proposed	93.5	90.8	95.7	94.2

CONCLUSIONS

The method recommended in this work extract the scene text from the natural images using combination of CC, edge and morphological operations. The main aim of the proposed algorithm is to reduce the false positive so that the efficiency of the OCR is increased. From the results it can be observed that most of the characters have been identified, in spite of diversity in font, size, language with low resolution, varying illumination and complex background. Combining OCR with text extraction algorithm gives a useful system for the image analysis. The text recognized can be used in many applications such as license plate recognition, address location etc. Text can be converted to audio outputs for blind and visually impaired persons. The proposed method has many possible future extensions. One of the issues is detecting highly blurred text in low resolution images. Also detecting multi-oriented texts needs to be investigated. We intend to address all these issues in our future work.

REFERENCES

[1] Cunzhao S., Chunheng W., Baihua X., Yang Z., Song G., "Scene text detection using graph model built upon maximally stable extremal regions," Pattern recognition letters journal, Elsevier, vol. 34, no. 2, pp. 107-116, 2013

[2] Lukas N., Jiri M., "Real-time scene text localization and recognition," 25th IEEE conference on computer vision and pattern recognition, pp. 3538-3545,2012.

[3] Chitrakala G., and D. Manjula, "Statistical modelling for the detection, localization and extraction of text from heterogeneous textual images using combined feature scheme," Signal, Image and Video processing, vol. 5, pp. 165-183,2011.

[4] Lyu M. R., Jiqiang S., Min C., "A comprehensive method for multilingual video text detection, localization, and extraction," IEEE transactions on circuits and systems for video technology, vol. 15, issue 2, pp. 243-255,2005.

[5] Marios A., Basillis G., Loannis P., " A two-stage scheme for text detection in video images," IEEE International conference on Image and vision computing, pp. 1413-1426,2010.

[6] Cong Y., Xin Z., Xiang B., " Detecting texts of arbitrary orientations in natural images," In journal of computer vision and pattern recognition, pp. 1083-1090,2012.

[7] P. Shivakumara, H. T. Basavaraju, D. S. Guru, C. L. Tan, " Detection of curved text in video:Quad tree based method," 12th international conference on document analysis and recognition, pp. 594-598,2013.

[8] L. Sun, G. Liu, X. Qian, D. Guo, " A novel text detection and localization method based on corner response," IEEE international conference on multimedia and expo, pp. 390-393,2009.

[9] Chucai Y., " Text locating in scene images for reading and navigation aids for visually impaired persons," 12th International ACM SIGACCESS conference on computers and accessibility," pp. 325-326,2010.

[10] Shehzad Muhammad Hanif, Lionel Prevost."Texture Based Text Detection In Natural Scene Images: Ahelp to blind and visually impaired persons."Conference and workshop on Assistive for People with Vision & Hearing Impairments, CVHI 2007, M.A.Hersh (ed).

[11] Anhar R., P. Shivakumara, C. S. Chan, C. L.Tan, "A robust arbitrary text detection system for natural scene images," Expert systems with applications, Elsivier, pp. 8027-8048,2014.

[12] Mohammad K., A. Behrad., "Text localization, extraction and inpainting in color images," 20th Iranian conference on electrical engineering, May 15-17, Tehran, Iran,2012.

[13] Jiri M., Ondrej C., Martin U.,and Tomas P., "Robust wide-baseline stereo from maximally stable extremal

regions," Image and vision computing journal 22 (10) , Elsevier, pp. 761-767, 2004.

[14] Asif S., Faisal S., and Andreas D., "ICDAR robust reading competition challenge 2: Reading text in scene images," In proceedings of IEEE International conference on document analysis and recognition, DOI: 10.1109, pp. 1491-1496, 2011.

[15] Huizhong c., Sam S. T., "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," In proceedings of 18th IEEE international conference on image processing, DOI:10.1109/icip 20116116200, pp. 2609-2612, 2011.

[16] Lukas N., and Jiri M., "Real-time scene text localization and recognition," In proceedings of 25th IEEE conference on computer vision and pattern recognition , pp. 3538-3545,2012.

[17] Lukas N., and Jiri M., "Text localization in real-world images using efficiently pruned exhaustive search ," In proceedings of ICDAR, pp. 687-691,2011.

[18] Lukas N., and Jiri M., "A method for text localization and recognition in real-world images ," In proceedings of the 10th Asian conference on computer vision, Springer-verlag, vol. 3, pp. 770-783,2011

[19] Cunzhao S., Chunheng W., Baihua X., Yang Z., and Song G., " Scene text detection using graph model built upon maximally stable extremal regions ," Journal of pattern recognition, Elsevier science, vol. 34, Issue 2, pp. 107-116,2013.

[20] Xu-Cheng Y., X. Yin, K. Huang, and Hong-Wei H., "Robust text detection in natural scene images," IEEE transactions on pattern analysis and machine intelligence, vol. 36, Issue 5, pp. 970-983,2014.

[21] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," 18th IEEE International conference on image processing, Doi: 978-1-4577-1303-3, pp. 2609-2612,2011.

[22] T. Q. Phan, P. Shivakumara, S. Tian, and C. L.Tan, " Recognizing text with perspective distortion in natural scenes," IEEE International conference on computer vision , pp. 569-576,2013.

[23] Boris Epshtein, E. Ofek, Yonatan Wexler, " Detecting text in natural scenes with stoke width transform ," IEEE conference on Computer vision and pattern recognition, pp. 2963-2970,2010.

[24] L. Wu, P. Shivakumara, T. Lu, and C. L. Tan, "A new multi-oriented scene text detection and tracking ," IEEE Transactions on multimedia , vol. 17 , Issue 8 , pp. 1137-1152,2015.

[25] X. Zhang, Z. Lin, F. Sun, and Yi Ma, "Transform invariant text extraction ," The visual computer, vol. 30, issue 4, pp. 401-415,2014.

[26] C. Yi, and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification ," IEEE transactions on image processing , vol. 21, Issue 9, PP. 4256-4268,2012.

[27] A. Mosleh, N. Bouguila, and A. B. Hamza, " Automatic impainting scheme for video text detection and removal ," IEEE transactions on image processing , vol. 22,Issue 11, pp. 4460-4472.,2013.

[28] Y. M. Y. Hasan , and L. J. Karam , "Morphological text extraction from images ," IEEE transactions on image processing, vol. 9, Issue 11, pp. 1978-1983,2000.