Edge Based Segmentation Approach to Extract Text from Scene Images

Kumuda T
Research Scholar, Dept of E&C, ATME,
Mysore, Karnataka, India.

L Basavaraj
Principal, A T M E, Mysore,
Karnataka, India.

***Abstract*— Scene text in images is finding lots of application in real life. When text in images are extracted efficiently they can be used for guiding tourists, visually impaired people, license plate detection, providing location information etc. Extracting scene text faces many problems, as images are subjected to lots of deprivation such as noise, uneven lighting, blur etc. In our work, we propose an algorithm to extract and analyze the text in from complex scene images efficiently. In the earlier stage, edges are detected using DWT. Then connected component clustering and AdaBoost classifier are used for localizing the text regions. In the next stage morphological operations and heuristic rules are used for character extraction. Finally using OCR extracted texts are analyzed. The proposed algorithm evaluated on different database images has achieved good results in spite of variations in font type, style, orientations and complex background.**

Keywords-Complexsceneimages,Textsegmentation,DWT,Connected component,AdaBoost classifier,Morphological filtering.

I. INTRODUCTION

Numerous algorithms have been developed to extract text from the images for applications such as address identification, number plate recognition, and automatic robot navigation and in wearable or portable computers. Text extraction is a challenging problem, as images may contain many sources of variations such as text in images may have different font size, different font style, they may be aligned in different orientation, contrast of the image may vary, and text may appear in various colors. Lighting conditions and shadows will affect the images illumination. Outdoor images will usually have the complex backgrounds. These variations make it difficult to extract the text automatically from the images.

In last decade many methods have been suggested [4, 6, 7, 5] for analysis of text in scene images, and some have proved their efficiency for a particular application. Low-level features such as Color variance, gray-level variation are used by K. C. Kim et al [8] for text region extraction and extracted regions are verified using a high level text stroke feature. Text stroke in local image areas are verified by multi-resolution wavelet transforms and SVM. An experimental result of this method shows that the text with particular size and font can be extracted. This method finds it difficult to identify the text, if the illumination in the image varies a lot.

Chucai Yi and Yingli Tian [9] have proposed an algorithm which makes use of stroke components and Gabor filters for scene text extraction. Initially they train the data set that computes the Gabor filters set which helps to identify the text's principle stroke component. With the help of characteristic distributions that are generated by principle stroke component they separate text from nontext regions. This method is robust against the text images having complex backgrounds and varying text patterns, but it fails to extract text with different orientation.Wakahara and Kohei Kita [10] mainly concentrated on the problem of binarizing multi-color words that appear in the natural scene images which occurs due to heavy variation in the intensity of the image and hard backgrounds. In the preliminary step binarized images are generated in the HSI color space by using color clustering method. Binarized images obtained are the difference of twice of the total clusters. Each character like pixels are identified with the help of support vector machines. Edge-enhanced MSER is used to identify the basic letter regions by Huizhong Chen et al [11]. Non-text regions are removed with the help of stroke width transform and geometric filtering.

Chucai Yi and Yingli Tian [12] developed a new algorithm based on color-uniformity. Edge pixels are clustered, depending on the color dependency and spatial relations into boundary layers. Then, each character regions are extracted at each boundary layer with the help of stroke segmentation. They proposed an algorithm which combines the use of stroke and the color information of the text to do the structural analysis. Hyung il koo and duck hoon kim [15] have developed two machine learning classifiers in which one locates the text area whereas the other helps in removing the nontext regions.

SeongHun Lee [13] proposed a method where initially text pixels are grouped together using k-means clustering to form the text regions. Then candidate labels are verified using an MRF model, later co-linearity weight is joined to obtain the geometrical interconnection between text components. From the state of art we can see that there is no single algorithm which extracts text with all variations. We are proposing an algorithm which is capable of extracting text of varying size, font and orientation. The remaining paper is arranged as: Section one gives brief introduction to the existing algorithms. Explanation about our method is given in section two. Section three contains test results and discussions. Conclusions and future works are stated in section four.

II. PROPOSED METHOD

The projected method has of four steps: Pre-processing, Text localization, Text extraction and character recognition. Block diagram of the complete method is as shown in the "Fig. 1".

978-1-5090-1560-3/17 $31.00 © 2017 IEEE

DOI 10.1109/IACC.2017.138

A. Pre-processing

In this stage, the input image is converted from color to intensity image 'Y' by combining RGB components together using "(1)". Noises in the gray image are removed using median filters while retaining the sharpness of the image as it is. As the median of the neighborhood pixels are used by the median filters instead of mean, the sharpness of the image is not altered. The result of this stage is as given in "Fig. 2".

$$Y = 0.3r + 0.6g + 0.2b \tag{1}$$

B. Text Localization

Detection and localization of the text region is done by extracting edges in the image. Usually text and background will have high difference in contrast for the easy detection of the text. Hence edge features can be used in identifying the text pixels. We implemented Harr discrete wavelet transform for detecting the edges. Coefficients of Harr wavelets are either 1 or -1; hence it is fastest compare to other wavelets. Compared to traditional edge detection filters like canny, sobel etc, DWT is more efficient and less time consuming because in traditional method for detecting edges in different direction we have to use different masks which is not the case in DWT.

Input image is decomposed into four sub-bands by 2-D DWT. LL sub band gives the average component, detailed components are given by other three sub bands LH, HL and HH. Among LH, HL, HH, vertical edge information is obtained by HL-band, horizontal edge information is obtained by LH-band and diagonal edge information is obtained by HH-band. Connected components of the same objects are identified from the edge image. As each character has distinct geometric features, each character can be grouped into individual connected region. We used the cc algorithm proposed by Suzuki [2] as it is effective as well as it takes significantly minimum labeling time.

In the next step the geometric features of the cc's are used by the AdaBoost classifier to filter non text cc's. Cascade AdaBoost classifiers has been successfully used for text cc's filtering [1].To build up cascade AdaBoost classifiers we adopted algorithm suggested by Zhu et al [1]. Effective cascade AdaBoost classifiers are obtained by connecting number of strong AdaBoost classifiers sequentially. Every classifier is trained in such way that each will have a good text detection rate with small false positive rate for the training set. Only those connected components which are selected as text component from the previous classifiers will pass on to the next classifier. The result of this stage is given in "Fig. 3".

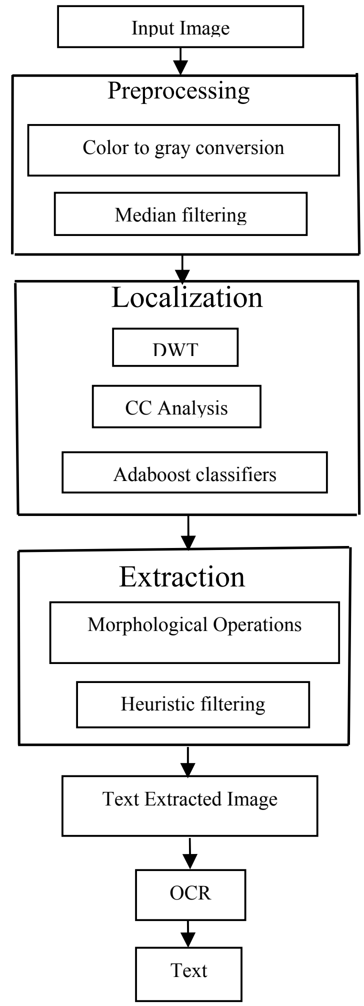


Fig .1.Block diagram of the proposed method

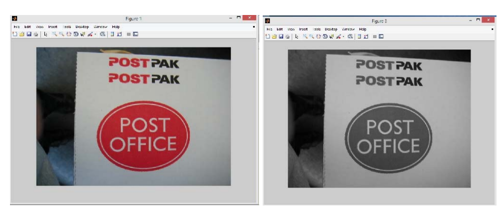


Fig .2. (a) Original image (b) median filtered image.

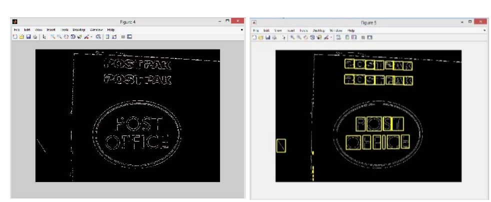


Fig .3. (a) Edge detected image (b) Valid CCs from Adaboost classifier

C. Text Extraction

In this step from the localized region, text pixels are separated from the background pixels. Efficiency of the OCR depends upon the accuracy of the text extraction system. Morphological filtering is widely used in image processing for extracting image components. Erosion and Dilation are the most commonly used morphological operations. Pixels are added to the boundaries of the object in dilation where as erosion removes the pixels from the object boundaries. Hence morphological operations are used for removing too small, too big objects and isolated regions in the image. Some assumptions are made for filtering non-character regions They are that character cannot be too small or too big as well as they appear in a group as a word. Threshold values are smaller for still camera and it depends upon the distance between the object and camera if the camera is moving. Rectangular bounding boxes are drawn around each letter by maximum and minimum coordinates of extracted character region. The output of this step is a binary image with only text pixels as given in "Fig. 4," and "Fig. 5(a) and Fig. 5(b)".

D. Character recognition

Recognition of text is done with the help of OCR. The output of the text extraction stage is fed to an OCR. OCR converts the text in the images into editable/computer readable data. OCR's are widely used for data entry from passport documents, invoices, bank statements. Earlier days training of data was required in order to identify the data from the images. But now with advanced technology one can easily make use of OCR to recognize the text from the images. OCR output in notepad is as shown in "Fig. 6".

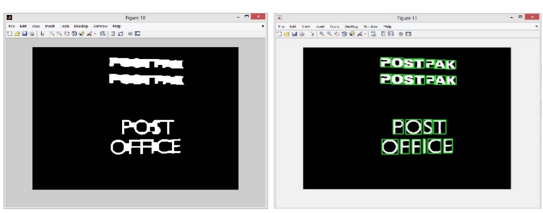


Fig .4. (a) Morphological Operation (b) Bounding Box

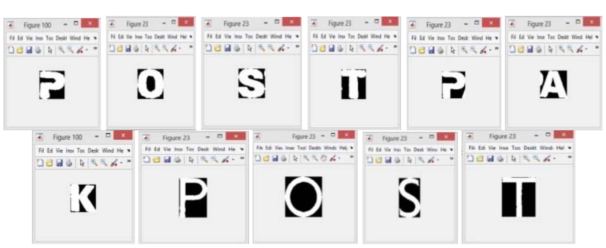


Fig .5(a).Extracted characters

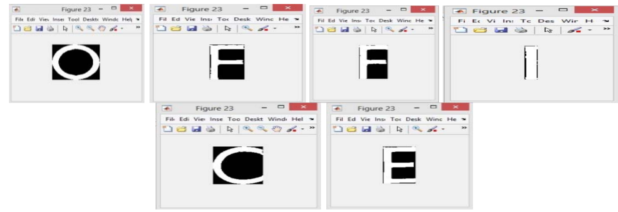


Fig .5(b).Extracted characters

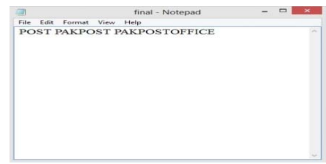


Fig .6.OCR output

III. RESULTS AND DISCUSSIONS

Images with text of different font, size, color and orientation from public dataset ICDAR 2011 and own data set collected by internet, camera, and mobile phones are analyzed to verify the quality of our method. From the outputs as shown in "Fig. 8," we can see that even without any prior knowledge, most of the texts are well detected by the proposed method. In the text extraction step the character pixels are segmented from the localized region using cc-based technique and morphological operations.CC-based method works accurately with prior information. As the text regions are already detected in the previous stage, cc method works effectively. This increases the accuracy of CC-based method.

Precision rate (PR) and recall rate (RR) are used as a tool to measure the performance of the proposed method. Precision rate and recall rate are calculated for set of images in database, and then average of each set is taken. Precision and recall rate are computed using equations (2) and (3) proposed by Chitrakala gopalan and Manjula D [3]. Number of text pixels identified by the algorithm which matches with the ground truth is considered as correctly detected pixels (CDP). False positives are considered in precision rate calculation. Non-Text pixels which are identified as text pixels by the algorithm are termed as false positives. False negatives are taken into consideration in the recall rate. Text pixels which are identified as non-text pixels by the algorithm are termed as false negatives._Hence accuracy of the algorithm in detecting text can be measured with help of precision and recall rate. Table I shows the PR and RR for the test data base images. We can see that the PR and RR of the proposed algorithm are remarkably high; however, some false positives may be detected due to periodical pattern or structural features that look like characters such as tree, branches, and leaves in complex images.

$$\text{Precision rate} = \text{CDP} \div (\text{CDP} + \text{FP}) \qquad (2)$$

$$\text{Recall rate} = \text{CDP} \div (\text{CDP} + \text{FN}) \qquad (3)$$

In order to give a an overall view about the status of the results obtained comparison with the other method [17,18] which uses second order statistics for texture feature and [16] which uses color feature for text extraction are carried out. Comparison outputs given in Table II shows that the proposed algorithm gives better results. Even though we got satisfactory results, but still proposed method fails in some cases such as classifying text from the structure such as leaves, window frames bricks etc which has edges similar to text. Precision rate and recall rate comparison results are shown in "Fig. 7".

IV. CONCLUSIONS

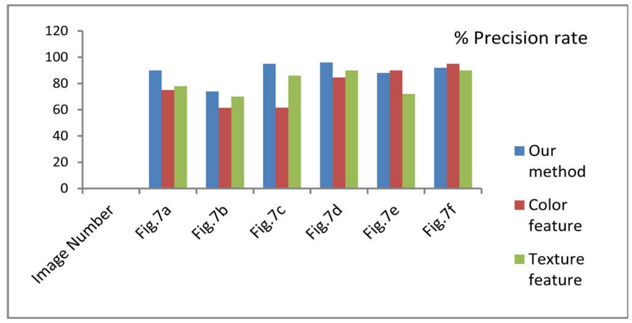
An algorithm suggested in this work uses Haar discrete wavelet transform along with sobel edge detector for text extraction from scene images. Adaboost classifiers with morphological operators are used for text extraction. After text extraction, they are given as input to OCR and then the results are displayed in the notepad. From the test results we can see that, the proposed algorithm extract text of different font, size, and orientation efficiently. Even though we got encouraging results, our method fails in detecting text from the structure such as leaves, window frames bricks etc which has edges same as that of text. This problem we need to overcome in the future works as well as we need to improve the speed of text extraction

TABLE I. Precision and Recall rate for different data set

Data-Set	Precision rate (%)	Recall rate (%)	Time (secs)
ICDAR	91	85	5-6
Own-Dataset	87	78	8-10

TABLE II. Comparisons of Precision and Recall rate with State of Art methods.

Image Number	Precision rate (%)			Recall rate (%)		
	Our Method	*State of Art Methods*		*Our Method*	*State of Art Methods*	
		Color feature	*Texture Features*		*Color feature*	*Texture Features*
Fig.7a	90	75	78	75	72	70
Fig.7b	73.9	61.4	70	52	64	60
Fig.7c	95	61.6	86	81.2	67	58
Fig.7d	96	84.6	90	43	70	63
Fig.7e	88	90	72	92	84	80
Fig.7f	92	95	90	89	93	96



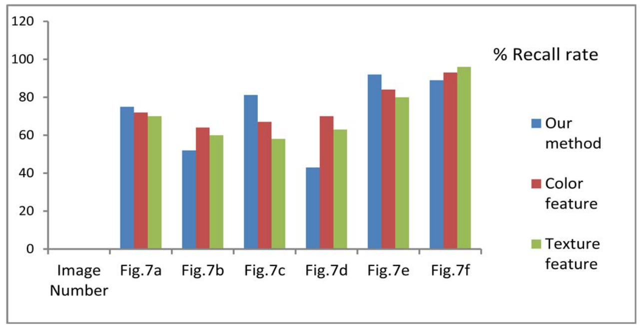


Fig .7. Comparisons of Precision and Recall rate with State of Art Methods.

REFERENCES

[1] Kaihua Zhu, Feihu Qi, Renjie Jiang, Masatoshi Kimachi Yue wu, Tomayoshi Aizawa, "Using Adaboost to detect and segment characters from natural scenes," In Proceedings of 1st International Workshop on Camera Based Document Analysis and Recognition, pp. 52-59, January 2005.

[2] Kenji Suzuki, Isao Honda, Noboru Sugie, "Linear-time connected-component labeling based on sequential local operations," Elsevier, Computer vision and image understanding, vol. 89, Issue 1, pp.,1-23, January 2003.

[3] Chitrakala G., and D. Manjula, "Statistical modelling for detection, localization and extraction of text from heterogeneous image using combined feature scheme," Springer, Signal, image, and video processing , vol. 5, issue 2,pp. 165-183, june 2011.

[4] Keechul Jung , Kwang In Kim , Anil K. Jain , "Text information extraction in images and video:A survey," Elsevier,Pattern Recognition, vol. 37, Issue 5, pp. 977-997,May 2004.

[5] Kumuda T., and L. Basavaraj, "Text extraction from natural scene images using region based methods-A survey," in Proceedings of ACEEE international conference on signal processing, image processing and computer vision, Issn. 22140344, DOI. 03, pp. 412-416, 2014.

[6] Jian Liang, David Doermann, Huiping Li, "Camera-based analysis of text and documents:A survey," Springer, International Journal of document analysis and recognition (IJDAR), vol. 7, Issue 2, pp. 84-104, July 2005.

[7] Jing Zhang, Rangachar Kasturi, "Extraction of text objects in video documents: Recent progress," in Proceedings of 8th IAPR international workshop on document analysis systems, Nara, Japan, pp. 5-17, Sept. 16-19, 2008.

[8] K. C. Kim, H. R. Byunl, Y. J. Song, Y. W. Choi, S. Y. Chi, K. K. Kim, and Y. K. Chung, "Scene text extraction in natural scene images using hierarchical feature combining and verification," Pattern recognition, Proceedings of the 17th international conference on pattern recognition, vol. 2, pp. 679-682,ISSN 1051-4651, ISBN:0-7695-2128-2,Aug 23-26, 2004.

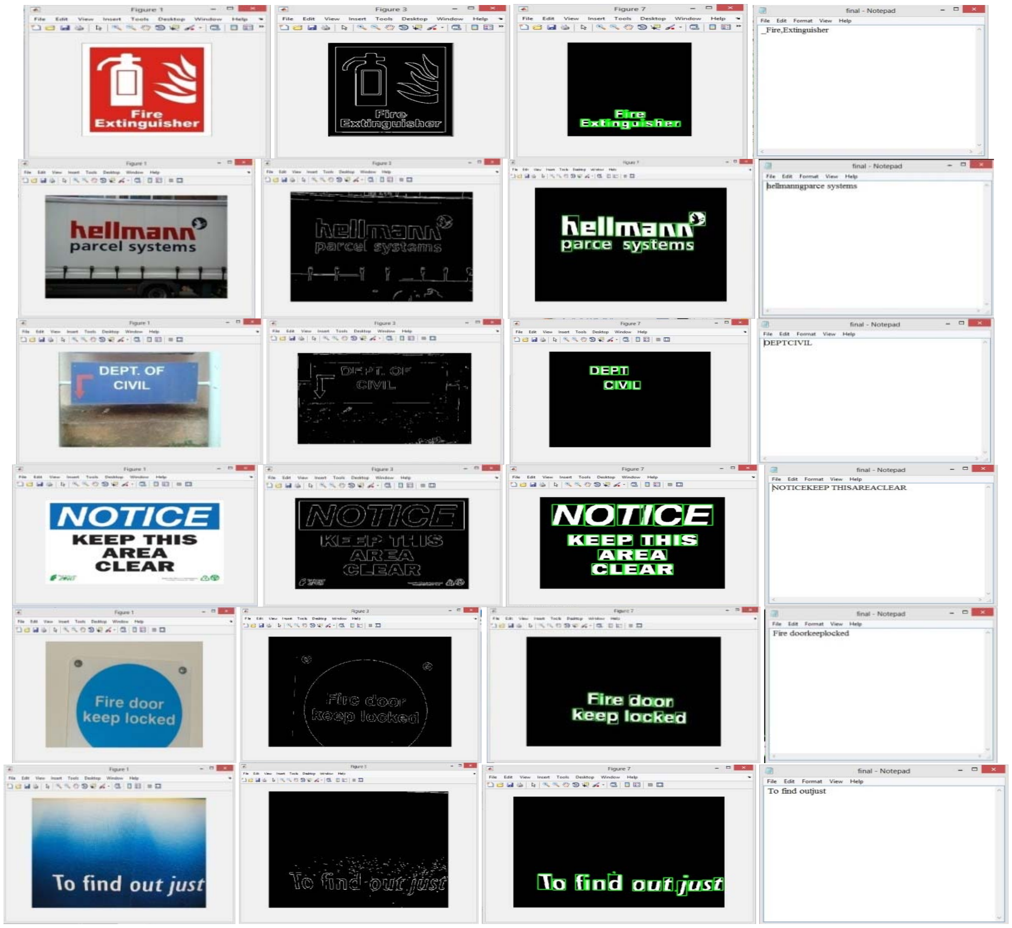


Fig .8. Text extraction examples (a, b, c, d) (from left to right: original images, edge image, text extracted image and OCR output

[9] Chucai Yi, and Yingli Tian, "Text detection in natural scene images by stroke gabor words," Proceedings of international conference on document analysis and recognition(ICDAR) , pp. 177-181,Issn 1520-5363,Beijing, Sept 18-21, 2011.

[10] Toru Wakahara, and Kohei Kita, "Binarazation of colour checter strings in scene images using k-means clustering and support vector machines," Proceedings of international conference on document analysis and recognition(ICDAR), pp. 274-278, Issn 1520-5363,Beijing, Sept 18-21, 2011.

[11] Huizhong . Chen, Sam S. Tsai, Georg Schroth, David M. Chen., R. Grzeszezuk, and B. Girod, "Robust text detection in natural images with edge enhanced maximally stable extremal regions," 18th IEEE international coference on image processing, pp. 2609-2612, ISSN 1522-4880, Brussels, Belgium, September 11-14, 2011.

[12] Chucai Yi, and Yingli 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, Sept 2012.

[13] SeongHun Lee, Min Su Cho, Kyomin Jungz, and Jin Hyung Kim, "Scene text extraction with edge constraint and text colinearity," Pattern recognition (ICPR), 20th international conference on pattern recognition, pp. 3983-3986, Istanbul, Aug 23-26, 2010.

[14] Ranjit G., Anandarup R., and Swapan K. P., "Text extraction from scene images using statistical distribution," 3rd International conference on emerging application technology, 2012.

[15] Hyung il koo, and Duck Hoon Kim, "scene text detection via connected component clustering and nontext filtering," IEEE transactions on image processing, vol. 22, issue. 6, pp. 2296-2305, June, 2013.

[16] Celine Mancas-Thillou, and Bernard Gosselin, "Color text extraction with selective metric based clustering," Elesevier,Computer vision and image understanding, vol. 107, issue 1-2, pp. 97-107, August 2007.

[17] S. A. Angadi, and M. M. Kodabagi, "A texture based methodology for text region extraction from low resolution natural scene images," International journal of image processing, vol. 3,Issue. 5, 2010.

[18] Muhammad Shehzad Hanif, Lionel Prevost, "Texture based text detection in natural scene images: a help to blind people with vision and hearing impairments," Proceedings of the conference and workshop on Assistive technology for people with vision and hearing impairments.Granada,Spain,August 28-31, 2007.