International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 4, April - 2014

A Novel Technique for Text Detection and Localization in Natural Scene Images

[1]Kumuda T, [2]L Basavaraj
[1] Adichunchanagiri Institute of Technology, Dept of E&C, Chikmagalur, India.
[2] A T M E, Dept of E&C, Mysore, India.

***Abstract*-Text that appears in images contains important and useful information. Text in images can be used for identification, indexing and retrieval etc. Text detection and localization in natural scene images is challenging due to the complex background,non-uniform illumination,andthe variations in font, size and orientation. In this paper we have proposed a novel technique for locating text regions in an image. The method is based on texture feature extraction using first and second order statistics. First texture features are extracted, then candidate text areas are obtained by applying a simple classification procedure using two discriminative functions. Finally the detected text blocks are merged and localized. Both ICDAR 2011 competition dataset and our own dataset are used for the experiments. Extensive experiments show that the proposed method can effectively detect texts of various sizes, colors,fonts and different orientation.**

Keywords-Text detection, Text localization, Texture features, Natural scene images.

1. INTRODUCTION

Natural scenes images have diverse objects and among them texts are important objects since they convey important meanings for image understanding. Text detection and segmentation from a natural scene images is very useful in many applications. In [23] automatic detection of traffic signs guarantees the normal operation of transportation systems. Scene text localization and translation of these texts will be of great help for foreign travelers and visually impaired people .In the massive data management respect, multimedia indexing and retrieval are

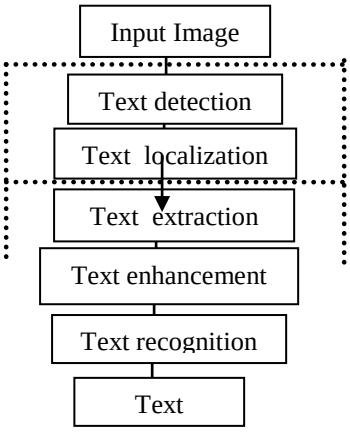

Fig. 1.Text Extraction System

some CC-based methods give encouraging results, there exist two difficulties degrading the system performance. also based on text localization technique.Shortly,text extraction has been a popular research topic in recent years.Common problems for text extraction from camera based images are the lack of prior knowledge of any kind of text features such as color, font size ,language and orientation, as well as the location of the probable text regions.

Jung et al [11] define an integrated image text information extraction system with five stages: text detection, text localization, text extraction, enhancement and recognition as shown in Fig. 1.Among these stages, text detection and localization are critical to the overall system performance. In the last decade, many methods as surveyed in [11], [15], [24], [13] have been proposed to address image text detection and localization problems, and some of them have achieved impressive results for specific application. However, fast and accurate text localization in natural scene images is still a challenging. In addition, text localization is generally time consuming, as is a challenging task to implement in a real system.

According to the feature utilized, text localization methods can be categorized into two types: Region based and Texture based methods. Region based methods are bottom up approaches where connected components (cc's) are found on the basis of their perceptive difference with the background.CC-based methods [26,16,8]are based on the fact that texts in images can be seen as sets of separating connected components, each of which has distinct intensity or color distributions and linked edge contours. Although First, CCs are hard to be accurately extracted due to image degrading and noises. Second, even if CCs can be extracted accurately, designing fast and highly credible CCs analysis algorithm is also difficult as there are too many text-like components in images.

The rest of the paper is organized as follows: Section II briefly reviews the related works. The proposed method is presented in section III to section V.Experimental results and discussions are presented in section VI and concluding remarks are provided in section VII.

II. RELATED WORK

Texture has long been an important research topic in image processing. Successful applications of texture analysis methods have been widely found in industrial, biomedical, and remote sensing areas.The methods of Zhong et al[25] and Gllavata et.al [7] employ image transformations, such as discrete cosine transform(DCT) and wavelet decomposition,toextract features. By thresholding the filter responses, non-text regions are removed and the remaining

text regions are grouped according to their spatial relationships.

Li et al[4,5] use mean, second and third order central moments in wavelet domain as the texture features and a neural network classifier is applied for text block detection. In their detection results, small isolated areas are filtered out and large text blocks are connected into text regions.Sin et.al[19] use frequency features such as the In recent years, texture-based methods [3,12,17,22] have been receiving more and more attentions with the development of image analysis, pattern recognition and machine learning techniques. These methods are based on the fact that text regions in images have distinct characteristics from non-text regions such as high density gradient distribution, distinctive texture and structure, which can be used to differentiate text regions from non-text regions effectively. Comparing with region-based methods, texture-based methods have similar localization accuracy but less sensitive to image noise.Texture-based methods are known to perform well even with noisy, degraded, textured or complex texts and backgrounds. In this paper we propose a texture-based method for text localization. Two features are computed for every block, using first and second order statistics. These features are used to classify the blocks as text and non-text.

Kwang ln Kim et.al[14] propose a texture based method using support vector machine. Textural properties of the texts are analyzed using SVM.Then CAMShift is applied to the results of the texture analysis.Although the SVM-based learning approach makes the algorithm fully automatic, it is still difficult to discriminate text with non-text using pure texture features in complex background since the feature is insufficient to discriminate text with general textures. Sunil Kumar et.all [21] proposed use of globally matched wavelet filters for extraction of textual areas of an image.Multiple,two-class Fisher classifiers are used to segment the document images into text, background and picture components. Segmentation results are finally refined using Markov random field.

Jain and Zhong[9] used a neural network (NN) to discriminate between text,graphics,and halftones in document images. A work similar to the proposed method is presented in [2].Angadi and Kodabagi presented texture based text detection segmentation from low resolution natural scene images each of size 240X320.Texture features homogeneity and contrast are computed from GLCM for every 50X50 blocks. The performance of the method has been tested for localizing kannada text in a indoor display board images.Harr discrete wavelet transform(DWT) based approach has been suggested by Liang and Chen[18].Kundu and Acharya[1] reported another scheme for segmentation of texts in the document images based on wavelet scale-space features. The method used M-band wavelet which decomposes an image into M X M based-pass channels so as to detect the text regions easily. The real text regions are then recognized based on the intensity of the text edges in an M-band image.Etemad et al [6] used a neural network to classify the output of wavelets into text and non-text regions.

number of edge pixels in horizontal and vertical directions and Fourier Spectrum to detect text regions in real scene images. Based on the assumption that many text regions are on a rectangular background, rectangle search is then performed by detecting edges, followed by the Hough transform. However it is not clear how these three stages are merged to generate the final result.

All these methods treats text as a type of texture. These methods usually divide a whole image into blocks. Then they use various approaches eg.Gabor filters,Fourier transform or Wavelet transform, to calculate the texture features of blocks. Methods based on the Fourier transform perform poorly in practice, due to its lack of spatial localization. Gabor filter provide means for better spatial localization, however, their usefulness is limited in practice because there is usually no single filter resolution at which one can localize a spatial structure in natural textures. Statistical texture analysis methods measure the spatial distribution of pixel values. They are well rooted in the computer vision world and have been extensively applied to various tasks. A large number of statistical texture features have been proposed, ranging from first order statistics to higher order statistics. Amongst many, histogram statistics, co-occurrence matrices, autocorrelation, and local binary patterns have been applied to texture analysis or classification. We have used both first-order and second-order statistics for texture feature computation.

III. SYSTEM

In this paper we propose text localization method, based on texture feature extraction, composed of two parts: text detection and text region localization. The flow chart of the proposed system is shown in Fig .2.Text detection consists of two stages, preprocessing and texture analysis. To increase the quality of texture classification a preprocessing stage is needed. After preprocessing texture features are calculated. Based on texture features ,blocks are classified as text and non-text blocks using newly defined discriminative functions.A merging procedure is applied in the final stage to merge the text block derived from the discriminative functions and to produce the final text localization result.

IV.TEXT DETECTION

A. Preprocessing:

At the preprocessing step, if the input image is color image then it is first transformed from RGB to gray-level space. The RGB components of the input color image are combined together to give an intensity image Y by using expression

$$Y=0.299R+0.587G+0.114B$$

Where R, G, B are the red, green and blue components of the input color image, respectively. The color components may differ in a text region, while having an almost constant intensity .So, the intensity image Y is processed in the next steps of the algorithm rather than the color components R, G and B.

The next step is to suppress the constant backgrounds such as building walls,doors, uneven lighting conditions, and

roughly estimate the text regions. A high pass filter in the DCT domain are used to remove constant background. The DCT co-efficient's globally map the periodicity of an image and can be a quite efficient solution for constant background suppression. The DCT co-efficient values are computed on every 8X8 block of the image. The DC component is the first entry in the DCT matrix. For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT block. While the high frequency components that are located at the bottom right corner contain zero values. Hence the constant background is removed by applying high pass filter. Then inverse DCT is

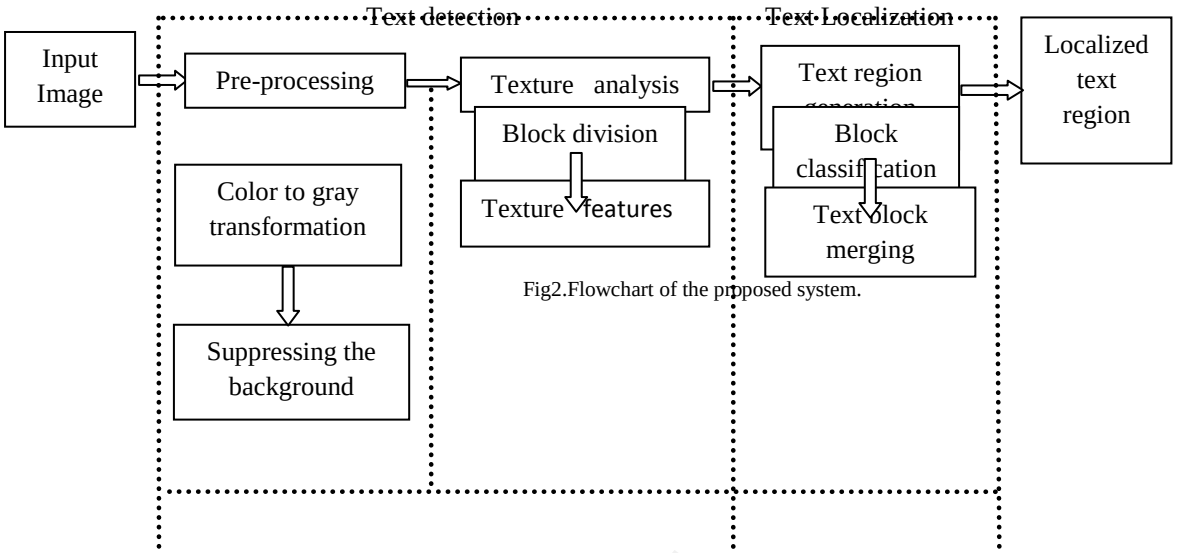

Fig2.Flowchart of the proposed system.

applied to obtain the background suppressed image. In the preprocessed image most of the unwanted details are removed,only gray level discontinuities belonging to the text and edges remain for further processing.An example of the preprocessing stage is shown in Fig.3 (b).

B.Texture analysis

In the proposed method block based texture segmentation is used. Image is divided into non-overlapping blocks of size 40X40, and then texture features are computed for every 40X40 blocks. The performance of texture based methods critically depends on how efficiently the block size is chosen. If the block size is too small, the classification errors will increase as unable to fully utilize the textural information, while a large block may have objects belonging to several different categories. Also a large block size will increase the storage requirements. After repeated analysis, we have used 40X40 block size which gives better result.

Textures are generally random, however posses consistent properties. Hence an obvious way to describe texture is their statistical properties. In this paper two basic feature groups are used for extracting textural properties of blocks. These features are calculated in the spatial domain, and the statistical nature of texture is taken into account. One of the simple ways to extract statistical features in an image is to use the first-order probability distribution of the amplitude of the quantized image. variance as given in equations (1) is calculated using first-order statistics. Where A is the input image matrix , r and c are number of rows and columns respectively. In order to approach higer classification accuracies it is necessary to consider texture information and neighborhood around each pixel. A GLCM stores the number of pixel neighborhoods in an image that have a particular gray scale combination, separated by a distance'd' in a direction 'ɵ'.Contrast as given in equation (2) is computed from p(i,j) ,which is the Gray level co-occurrence matrix[GLCM].

$$\text{Variance} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}(A(i,j)-\mu)^2}{(r*c)-1} \quad (1)$$

$$\text{Contrast} = \sum_{i,j}|i-j|^2 p(i,j) \quad (2)$$

Haralick et.all.[20],first introduced the use of co-occurrence probabilities using GLCM for extracting various texture features.GLCM is a second order texture measure. The dimension of GLCM is GXG where G is the number of gray levels used to construct the matrix. Number of gray levels is an important factor in GLCM computation. We have used GLCM of dimension30X30. The texture feature contrast is obtained from every preprocessed 40X40 block in four orientations(0^o,45^o,90^o,and 135^o) at d=1 . An average over all orientations is taken so that these matrices are rotation invariant. Block co-ordinates which corresponds to minimum and maximum row and column numbers of the block along with the extracted features are stored in a feature matrix .

V. TEXT LOCALIZATION

Texture features extracted from the blocks are used by the discriminate functions for classification decisions. After elaborate analysis based on experiment conducted, threshold values for each feature are selected. The threshold value for variance is selected as greater than 39 and for contrast it is greater than 1.31There are two

discriminative functions one for each feature. Discriminative functions classify every block into two classes text and non-text blocks based on threshold values. Discriminative functions checks whether the blocks feature vector satisfies the threshold values or not. Only those blocks which satisfy both discriminative functions are classified as text blocks.

The merging process combines the detected text blocks connected in rows and columns to obtain text regions. Four pairs of co-ordinates of the boundary blocks are determined by the maximum and minimum co-ordinates of the top, bottom,left and right points of the corresponding text blocks. Finally the detected text blocks are merged together spatially and text regions are enclosed within a bounding box.In order to avoid missing those character pixels which lie near or outside of the initial boundary, width and height of the boundary box are padded by a small amount. An example of the text region localization is shown in Fig.3.

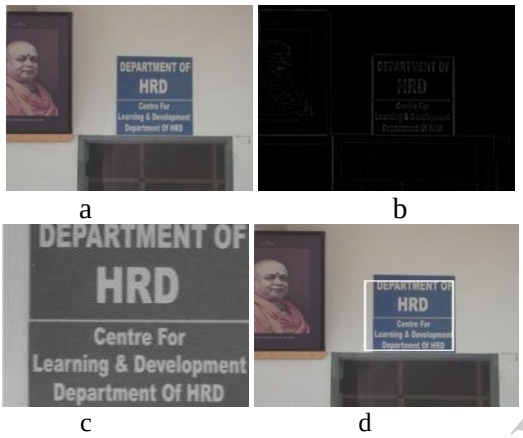

Fig. 3. (a) Input image (b) background suppressed image (c) text block merged image (d) text region localization result.

VI. RESULTS

We analyzed several types of images so as to demonstrate the performance of our algorithm. The system was tested on a public data set: ICDAR 2011 Robust reading competition data set and own image data set collected by us using digital camera and mobile phones in indoor and outdoor conditions. The test sets consists of a variety of cases, including text in different font-size, font-color, orientations and languages. Fig .4. Shows some examples of images from two datasets. The resolution of these images ranged from 960 X1280 to 2048X1536. All the images were in JPEG format. For each detected text region, the proposed system drew box encompassing that region in the input image.Intially the image is preprocessed to suppress the complex backgrounds, such as building walls, doors, and uneven lighting conditions using a high pass filter in the DCT domain.

After preprocessing, image is divided into blocks.If the block size is too small, the classification errors will increase because the classifier is unable to fully utilize the textural information. Conversely,

Fig.4.Examples of evaluation data sets (a)-(b) ICDAR 2011 competition dataset, (c)-(d) our dataset.

if the block is too large, this can result in unstable classifications at the texture boundary. Accordingly, selecting the appropriate input window size is a tradeoff between classification accuracy and processing time. As such, experiments were performed with different input block sizes of 50X50, 40X40, 30X30, 100X20, 20X20.It was observed that with the larger block sizes, the error rate is smaller, the best results were obtained with a 40X40 block. Our system was coded in Mat lab and operated on a 2.53GHz computer with windows-7 OS.And it was observed that the processing time lies in the range of 4 to 6 secs.

Fig.5. Fig6 and Fig7 show examples of text localization results from both data sets along with the computed text bounding box. We have not used any priori information regarding the font size or format of the text in the image. It can be seen from the results, that most of the text is well detected despite variations in font size and font style. Our algorithm is based on the difference between texture of text and non-text areas. Thus it works successfully in the cases where text is randomly oriented at different angles as shown in Fig.6. and different languages as shown in Fig.7.Precision rate and recall rate are used as a performance evaluation of the text detection results[21] ,which are given as in equation (3) and (4).Tables I shows the precision and recall rates for the test database images .

$$\text{Precisionrate}=\frac{Number\ \ of\ currently\ \ detected\ \ text\ blocks}{Total\ number\ \ of\ text\ blocks\ detected} \quad (3)$$

$$\text{RecallRate}=\frac{Number\ \ of\ currently\ \ detected\ \ text\ blocks}{Number\ \ of\ text\ blocks\ \ in\ image} \quad (4)$$

The method has detected larger region than the actual size of the text region, when images with more complex backgrounds containing trees, buildings are tested. It is shown that the proposed method can detect and localize most scene texts. Nevertheless, some false positives and negatives still exist.

TABLE I:TEXT REGION LOCALIZATION RESULTS

Image number	Precision rate (%)	Recall rate (%)
Fig.3a.	73.91	73.91
Fig..4a	100	75
Fig.4b.	81.8	90
Fig.4c	100	100
Fig.4d	95.65	95.65

VII.CONCLUSIONS

In this paper we have presented a novel technique for locating the text region based on texture features.Our method computationally less expensive and faster.The combination of first order and second order statistics make the method a truly robust one.We have applied our algorithm on several images with complex backgrounds and obtained encouraging results.Although the proposed system repoted encouraging performance,it still needs further improvements. Our approach fails in some cases where the background texture matches with the text texture. Currently, we only provide a text detection and localization method. Text should be clearly extracted from its background to obtain a good recognition result for the characters. Special technique should be investigated to segment the characters from their background before putting them into OCR software in the future work. In the future work, text extraction and recognition should be integrated with text localization to complete the need of text information extraction

Fig. 5. Text region localization examples on the ICDAR 2011 competition dataset (from left to right: original images, results of background suppression, merging of text blocks and text region localization)

Fig .6. Example on the oriented text region localization (from left to right: original images, results of background suppression, merging of text blocks and text region localization)

Fig .7. Example on the Kannada text region localization (from left to right: original images, results of background suppression, merging of text blocks and text region localization)

REFERENCES

[1] M. Acharya and M. K. Kundu, "Multiscale segmentation of document images using-band wavelets," in Proc. 9th Int. Conf. Compute. Analysis Images and Patterns, pp. 510–517, ISBN 3-540-42513-6.

[2] 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, Volume.3.Issue.5.2010

[3] D. T. Chen, J. M. Odobez, and H. Bourlard. Text detection and recognition in images and videos frames. PatternRecognition, 37(3):595–608, 2004.

[4] H. Li, D. Doermann, and O. Kia, "Automatic Text Detection and Tracking in Digital Video," IEEE Trans. Image Processing, vol. 9, no. 1, pp. 147-156, 2000.

[5] H. Li, D. Doermann, O. Kia, Automatic text detection and tracking in digital video, Maryland University LAMP Technical Report 028, 1998.

[6] K. Etemad and R. Chellappa, "Separability based tree structured local basis selection for texture classification," in Proc. Int. Conf. Image Processing, 1994, vol. 3, pp. 441–445.

[7] J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients, "in Proc. 17th Int. Conf. Pattern Recognition (ICPR'04), Cambridge,u.k 2004.pp.425-428

[8] T. Hiroki. "Region graph based text extraction from outdoor images". In Proc. 3rd Int'l Conf. Information Technology and Applications, volume 1, pages 680–685, 2005..

[9] A.K. Jain and Y. Zhong, "Page Segmentation Using Texture Analysis, "Pattern Recognition, vol. 29, no. 5, pp. 743-770, 1996.
[10] A. Jain and S. Bhattacharjee, "Text segmentation using gabor filters for automatic document processing," Machine Vis. Applicat., vol. 5, pp. 169–184, 1992.
[11] K. Jung, K. I. Kim, and A. K. Jain, "Text information extraction in images and video: A survey," Pattern Recogn., vol. 37, no. 5, pp. 977–997, 2004.
[12] K. I. 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 Analysis and Machine Intelligence, 25(12):1631–1639, 2003.
[13] Kumuda.T,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 VLSI,ICrtSIV2014,ISSN:22140344,DOI:03.AETS.2014.5.377,p412-416.
[14] Kwang In Kim, Keechul Jung, and Jin Hyung Kim," Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm." IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 12, DECEMBER 2003.
[15] J. Liang, D. Doermann, and H. P. Li, "Camera-based analysis of text and documents: A survey," Int. J. Document Anal. Recogn., vol. 7, no.2-3, pp. 84–104, 2005.
[16] Y. X. Liu, S. Goto, and T. Ikenaga. A contour-based robust algorithm for text detection in color images. IEICE Trans.Information and Systems , 89(3):1221–1230, 2006.
[17] H. P. Li and D. Doermann. A video text detection system based on automated training. In Proc. 15th Int'l Conf. Pattern Recognition, volume 2, pages 2223–2226, 2000
[18] C. W. Liang and P. Y. Chen, "DWT based text localization," Int. J.Appl. Sci. Eng., vol. 2, no. 1, pp. 105–116.
[19] B. Sin, S. Kim, B. Cho, Locating characters in scene images using frequency features, Proceedings of International Conference on Pattern Recognition, Vol. 3, Quebec, Canada,2002, pp. 489–492.
[20]Robert M.Haralick, K.Shanmugam, and Its'hak Dinstein."Textural Features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, vol.smc-3.no.6, November 1973, pp.610-621.
[21]Sunil kumar,Rajat Gupta,Nitin Khanna,Santanu Chaudhury,and Shiv Dutt Joshi."Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model", IEEE Transactions on Image Processing, Vol 16, No 8, August 2007.
[22] V. Wu, R. Manmatha, and E. M. Riseman. Finding text in images. In Proc. 2nd ACM Int'l Conf. Digital libraries, pages 3–12, 1997.
[23]Xiaoqian Liu,Ke lu,Weiqiang Wang."Effectively localize text in natural scene images"21st International conference on pattern recognition,November11-15,2012,Tsukuba,Japan.
[24] J. Zhang and R. Kasturi, "Extraction of text objects in video documents: Recent progress," in Proc. 8th IAPR Workshop on Document Analysis Systems (DAS'08), Nara, Japan, 2008, pp. 1–13.
[25] Y. Zhong, H. J. Zhang, and A. K. Jain, "Automatic caption localization in compressed video," IEEE Trans. Pattern Anal. Mach. Intel., vol. 22, no. 4, pp. 385–392, 2000.
[26] K. H. Zhu, F. H. Qi, R. J. Jiang, L. Xu, M. Kimachi, Y. Wu,and T. Aizawa. Using AdaBoost to detect and segment characters from natural scenes. In Proc. 1st Int'l Workshop on Camera Based Document Analysis and Recognition, pages 52–59, 2005.