© 2015 IJEDR | Volume 3, Issue 4 | ISSN: 2321-9939

Segmentation Framework for Multi-Oriented Text Detection and Recognition

Shashi Kant, Sini Shibu
Department of Computer Science and Engineering, NRI-IIST, Bhopal

***Abstract* - Here in this paper a new and efficient technique for the text detection from natural scenes is implemented. The proposed methodology is based on the concept of Otsu's segmentation method which segments the higher intensity texts from the natural scenes. Although there are various text detection techniques implemented, but the proposed methodology implemented here for text detection provides high precision and recall rate as compared to the existing technique. The methodology also provides less computational time to detect texts from scenes.**

I. INTRODUCTION

Text detection and extraction method play an important role in many applications over segmentation. It is challenging task due to rapidly increase the digitization of all the materials. It is complex because we need to find out where the text located in the scene image. Text extraction process involves text detection, segmentation and recognition of text. Scene images contains the text, such as the advertising boards, banners, vehicle number plates, street sign boards and traffic sign board etc, which is captured by cameras. This text is difficult to detect and recognized due to their various styles, font, color, contrast and orientation.

Although text detection has been studied extensively in the past [6.7], the problem remain unsolved. The difficulties mainly come from two aspects: one the diversity of text and other complexity of background. Most of existing methods for text detection [1], [8], [12] and recognition [4], [5], [13] can only handle horizontal texts. To rectify this issue we proposed an algorithm which is segmentation based recognition that can detect and recognized texts with different scales, colors, fonts and orientation in various real world scenarios.

To evaluate the effectiveness of the proposed system, we have conducted extensive experiments on segmentation based text detection and end-to-end text recognition using, the ICDAR dataset [14] and MSAR dataset [15]. MSAR dataset includes both English and Chinese characters. So we generate a large collection of natural scene images with only English letters. The proposed algorithm achieves highly performance on ICDAR and MSAR datasets.

II. LITERATURE REVIEW

There have been a large number of methods dealing with text detection in natural images, scene text detection and recognition have been important research for decades [4], [5] and have recently seen a surge in research effort in computer vision community [2], [3] [11]-[13], [16], [17], [18]. This literature review will involve in text recognition system.

The approaches of Chen et al. [1] and Epshtein et al. [8] are recent attempts towards end-to-end text recognition systems. These approaches mainly focus on the problem of text detection and achieve further recognition using Off-the-shelf OCR engines. Wu et al. [18] introduced automatic text detection and recognition based Text Finder system. This system treated text as a distinctive texture and procedures called texture segmentation and chip generation were applied succession to extract candidate regions. So the detectd text regions were binarized and go to OCR software for character recognition.

Andrej Ikica et al. [9] gave an ivmproved edge profile based method for text detection in image of natural scenes that can first process image and detect edges of text for segmentation. Jing Zhang et al. [10] introduced text detection using edge gradient and graph spectrum method for segmentation and detection of edges of the text from scene image In [16], text recognition system, Wang introduced an end-to-end scene text recognition system, which aim to tackle a special case of scene text understanding where in addition to the natural image it is also given a lists of words (lexicon) to be detected and read. However, the usability of this algorithm in general text is limited.

Azadboni et al. [11] have provided a text detection and character extraction in color image system for text character segmentation. He used two methods namely FFT domain filtering and SVM classifier for text extraction from natural images. In [17], Neumann and Matas presented a Maximally Stable External Region (MSER) based method for text localization and recognition result on challenging real-world images. The authors achieved scene text recognition using trained OCR software. Moreover to achieve higher performance this algorithm replaced in the tradition feed-forward pipeline with verification framework.

Most related to our work, Cong Yao and Xiang Bai et al. [19], suggested to tightly couples of aspects of text detection and recognition and share features for these two tasks. In this paper, we achieve high text detection and recognition accuracy concurrently with exactly same features and classification schemes. For this we realize a unified framework for scene text segmentation, detection and recognition. The algorithm in our previous work is able to detect scene text of different orientation and features are effective as well for recognition on different (English, Arabic numbers) characters. Therefore, we base current work to build an end-to-end segmentation based scene text recognition system, which is simple yet effective and provides high performance. Through widely used in the community, the ICDAR dataset [14] only contain horizontal English texts. Here we collect new dataset [15] with 500 images of indoor and outdoor scene. This is suitable to handle texts of arbitrary orientations.

III. PROPOSED METHODOLOGY

The proposed methodology implemented here for text detection using segmentation based algorithm is applied here for the detection of text from the natural scenes. The proposed methodology implemented here consists of following phases.

1. Take an input Training dataset of MSAR and ICDAR Dataset.
2. Filter each of the input training images using Gaussian filter.
3. Perform segmentation on each of the trained images and stores the features.
4. Take an input Testing dataset of MSAR and ICDAR Dataset.
5. Filter each of the input testing images using Gaussian filter.
6. Perform the segmentation on each of the test image and match the segmented portion with the stored features and classify the text from images.

A. Segmentation by Thresholding

Segmentation is the simple segmentation method in which the pixels are partitioned depending o n their intensity value. Global thresholding using an appropriate threshold T:

g(x, y) = { 1, *if f(x, y) > T*, 0 *if f(x, y) <= T*

Variable thresholding if T can change over the image.

B. Global Thresholding

Global thresholdind consists of the following steps:

1. Initial estimate of T.
2. Segmentation using T:
   a. G1, pixels brighter than T.
   b. G2, pixels darker than T.
3. Computation of the average intensities m1 and m2 of G1 and G2.
4. New threshold values:

$$\mathrm{T_{new}} = \frac{\mathrm{m1} + \mathrm{m2}}{2}$$

5. If $|T - T_{new}| > \Delta T$, back to the step 2, otherwise stop.

C. Proposed Algorithm for Segmentation

The proposed methodology based segmentation can be used for the detection of the texts in natural scenes. This methodology is aimed in finding the optimal value for global threshold. The proposed methodology is based on the interclass variance maximization where well threshold classes have well discriminative intensity values. It is a very simple form of segmentation.

A threshold is defined than every pixel in the image compared with this threshold. If the pixel lies above the threshold it will be marked as foreground. The threshold will most often be intensity or color value. Other forms of the thresholding exist where the threshold is allowed to vary across the image, but thresholding is a primitive segmentation task.

Thresholding is non-linear operation that converts a gray-scale image to binary image where the two level are assigned to pixels that are below or above the specified threshold value. In this method the section of initial threshold value is depends upon the histogram of an image and the gray the gray scale of an image.

1. The methodology starts with the creation of M*N image histogram:
2. L is intensity level, [0.....L-1].
3. Ni is the number of pixels of intensity i:
   a. $MN = \sum_{i=0}^{L-1} n_i$
4. Normalized the histogram:

$$p_i = \frac{n_i}{MN}$$

$$\sum_{i=0}^{L-1} p_i = 1\,, p_i \geq 0$$

5. Similarly, m2 mean intensity of the pixel C2

$$m_2 = \frac{1}{P_2} \sum_{i=k+1}^{L-1} i.p_i$$

6. While the mean intensity up to the k level m:

$$m = \sum_{i=0}^{k} i.p_i$$

7. Hence:

$$P_1 m_1 + P_2 m_2 = m_G$$
$$P1 + P2 = 1$$

D. Performance Parameters

These are the basic functions that we have used to find the efficient accuracy for our proposed system in term of precision, recall and f-measure.

Precision: Precision (p) is known as the ratio of CDTP to the sum of CDTP and FP pixels.

$$p = \frac{CDTP}{CDTP + FP}$$

Recall: Recall (r) is defined as the ratio of CDTP to the sum of FN pixels.

$$r = \frac{CDTP}{CDTP + FN}$$

F-Measure:

$$f = \frac{2 * p * r}{p + r}$$

Where: CDTP is Correctly Detected Text Pixels.
FP is False Positive or Pictorial Pixels.
FN is False Negative or Texture Pixels

IV. RESULT AND DISCUSSION

A. Datasets

ICDAR

The ICDAR dataset includes 450 images in total. Some of them are from ICDAR 2003 and 2005, except few images added by the organizers of ICDAR Robust Reading Competition Challenge [14]. ICDAR dataset contains 229 training images and 221 testing images in the dataset.

Figure 4.1 Examples of Train and Test scene text of ICDAR Dataset.

MSAR

The MSAR dataset is benchmark for multi-oriented text detection in natural scenes for evaluating algorithms of detection. This dataset includes 500 images in total of horizontal as well as skewed text in complex natural scenes. MSAR consist 300 text images of natural scenes for training and 200 text images of natural scenes for testing

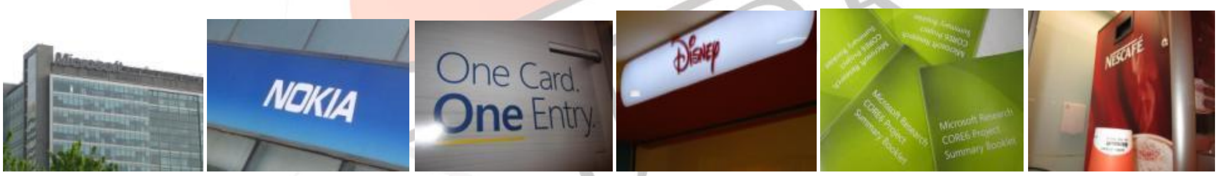

Figure 4.2 Examples Train and Test Scene text of MSAR dataset

B. Experiments Results

The table shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on ICDAR Dataset which consists of training and testing images. The Existing and proposed methodology when applied on ICDAR Dataset provides precision and Recall and F-measure. The proposed methodology implemented here for the text detection provides more Precision & Recall as compared to the existing Text Detection Methodology.

Table 4.1 Analysis & Comparison on ICDAR Dataset

	ICDAR Dataset		
Algorithm	**Precision**	**Recall**	**F-Measure**
Proposed	0.5089	0.036	0.0675919
Base	0.4911	0.036	0.0674296

The table shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on ICDAR Dataset which consists of training and testing images. The Existing and proposed methodology when applied on ICDAR Dataset provides Computational time in seconds. The proposed methodology implemented here for the text detection provides less Computational time as compared to the existing Text Detection Methodology.

Table 4.2 Analysis & Comparison of Time on ICDAR Dataset Images

No. of ICDAR Images	Time in sec	
	Existing Work	**Proposed Work**
10	8.4865	2.5272
20	17.6749	3.2916
30	20.5609	5.2884
40	26.7854	6.1776
50	28.0022	7.41
60	32.5886	8.3149
70	40.5915	8.9077
80	44.4603	7.7844
90	52.3071	11.2165
100	54.2103	14.7421

The table shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on MSRA Dataset which consists of training and testing images. The Existing and proposed methodology when applied on MSRA Dataset provides Computational time in seconds. The proposed methodology implemented here for the text detection provides less Computational time as compared to the existing Text Detection Methodology.

Table 4.3 Analysis & Comparison of Time on MSRA Dataset Images

No. of MSRA Images	Time in sec	
	Existing Work	**Proposed Work**
10	13.5565	4.5708
20	32.417	8.5801
30	40.1079	11.1853
40	66.4252	11.4349
50	66.5812	16.2397
60	82.1033	21.1693
70	104.4427	26.1614
80	94.0062	31.5122
90	128.264	32.8694
100	135.0345	32.027

The table shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on MSRA Dataset which consists of training and testing images. The Existing and proposed methodology when applied on MSRA Dataset provides precision and Recall and F-measure. The proposed methodology implemented here for the text detection provides more Precision & Recall as compared to the existing Text Detection Methodology.

Table 4.4 Analysis & Comparison on MSRA Dataset

Algorithm	MSRA Dataset		
	Precision	**Recall**	**F-Measure**
Proposed	0.6	0.5	0.5455
Base	0.4	0.5	0.44

The table shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on MSRA & ICDAR Dataset which consists of training and testing images. The Existing and proposed methodology when applied on ICDAR & MSRA Dataset provides Computational time in seconds. The proposed methodology implemented here for the text detection provides less Computational time as compared to the existing Text Detection Methodology.

Table 4.5 Comparison of Computational Time on various Datasets

Algorithm	ICDAR Dataset	MSAR Dataset
Proposed	22.0429	39.8895
Base	110.8855	277.3074

The figure shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on ICDAR Dataset which consists of training and testing images. The Existing and proposed methodology when applied on ICDAR Dataset provides precision and Recall and F-measure. The proposed methodology implemented here for the text detection provides more Precision & Recall as compared to the existing Text Detection Methodology.

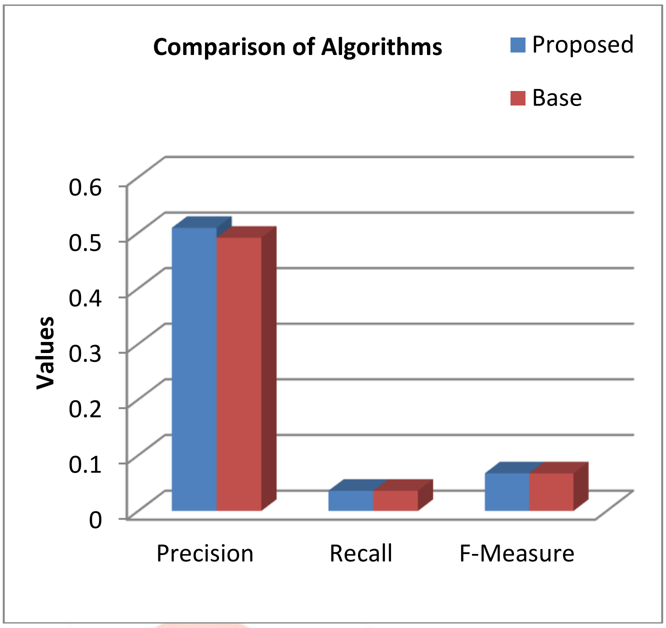

Figure 4.3 Analysis & Comparison on ICDAR Dataset

The figure shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on ICDAR Dataset which consists of training and testing images. The Existing and proposed methodology when applied on ICDAR Dataset provides Computational time in seconds. The proposed methodology implemented here for the text detection provides less Computational time as compared to the existing Text Detection Methodology.

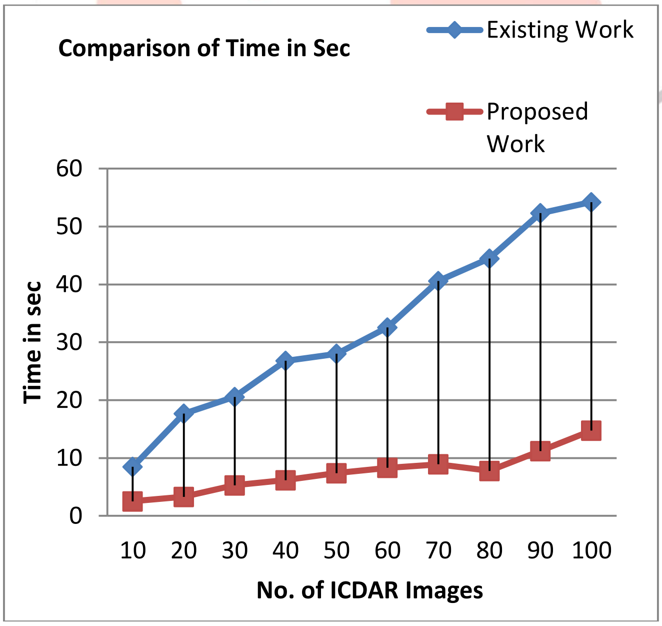

Figure 4.4 Analysis & Comparison of Time on ICDAR Dataset Images

The figure shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on MSRA Dataset which consists of training and testing images. The Existing and proposed methodology when applied on MSRA Dataset provides Computational time in seconds. The proposed methodology implemented here for the text detection provides less Computational time as compared to the existing Text Detection Methodology.

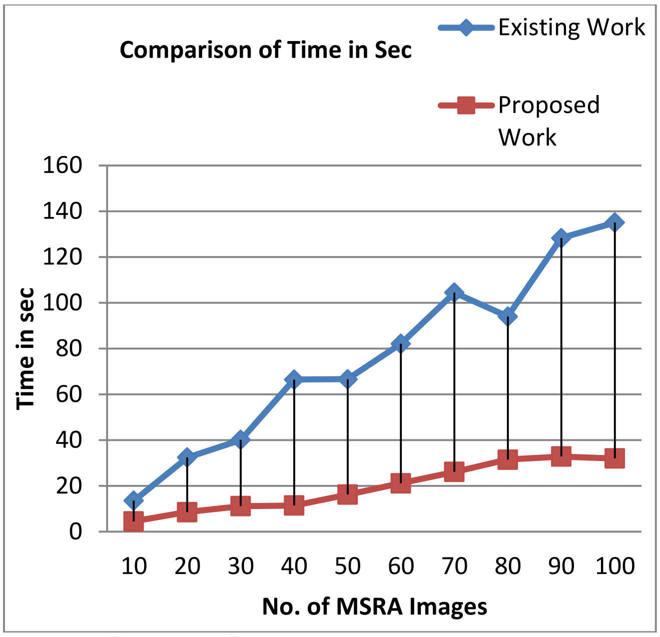

Figure 4.5 Analysis & Comparison of Time on MSRA Dataset Images

The figure 4.6 shown below is the analysis and comparisons of existing text detection and the proposed text detection methodologies. The Experimental results are performed on MSRA Dataset which consists of training and testing images. The Existing and proposed methodology when applied on MSRA Dataset provides precision and Recall and F-measure. The proposed methodology implemented here for the text detection provides more Precision & Recall as compared to the existing Text Detection Methodology.

The figure 4.7 shown below is the analysis and comparison of existing text detection and the proposed text detection methodologies. The Experimental results are performed on MSRA & ICDAR Dataset which consists of training and testing images. The Existing and proposed methodology when applied on ICDAR & MSRA Dataset provides Computational time in seconds. The proposed methodology implemented here for the text detection provides less Computational time as compared to the existing Text Detection Methodology.

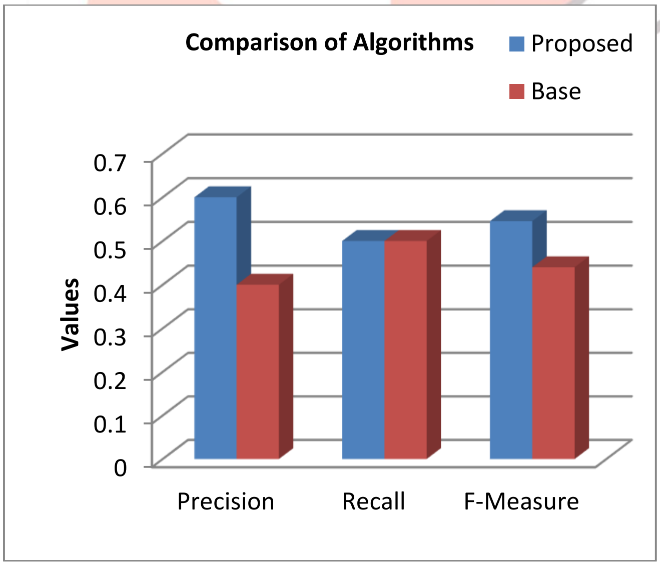

Figure 4.6 Analysis & Comparison on MSRA Dataset

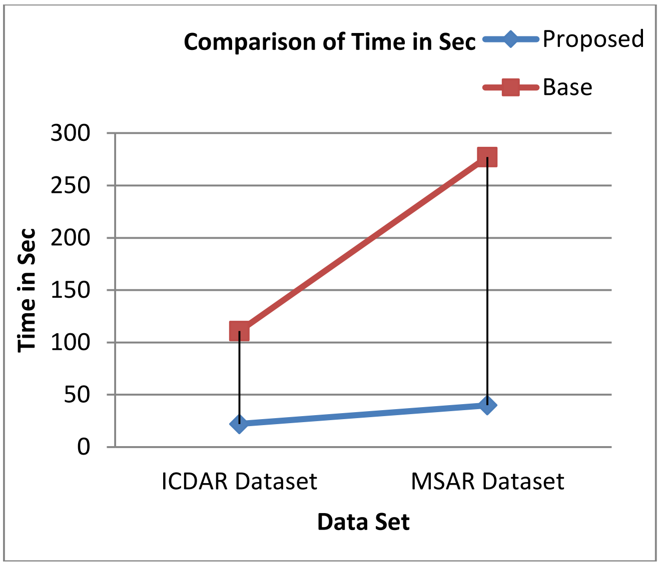

Figure 4.7 Comparison of Computational Time on various Datasets

V. CONCLUSION

The text detection is a technique of detecting text from the natural scenes that the detected text can be used for various applications. The existing methodology implemented here for text detection using multi-oriented framework based text detection provides less accuracy of detecting text. Hence an efficient methodology is implemented here for the text detection using segmentation based text detection. The experimental results are performed on two datasets namely ICDAR and MSAR dataset. The datasets contains number of images including natural scenes. The various experimental results show that the proposed methodology provides higher precision and recall as well as accuracy as compared to the existing methodology implemented for the text detection as shown above in the tables 4.1 and 4.4. The proposed methodology implemented also provides less computational time for text detection as shown in tables 4.2, 4.3 and 4.5. The various analyses are done for whole dataset images and images with combination of 10, 20, 30, 40.......100 images from each dataset. Here also the proposed methodology efficient results as compared to the existing methodology.

REFFERENCES

[1] X. Chen and A. Yuille, “Detecting and reading text in natural scenes”, Proc. of CVPR, 2004.
[2] A. ZramdiniS, R. Ingold, “Optical font recognition using typographical features,” In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pp. 877-882, 1998.
[3] C. Yi andY. Tian , “Text string detection from natural image by structured-based pertition and grouping,” IEEE Transction in Image Processing, Vol. 20, no. 9, pp. 2594-2605, 2011.
[4] D. Chen, J. M. Odongibez, and H. Bourlard, “Text detection and recognition in images and video frames” Pattern Recognition, Vol. 37, no. 3, pp, 595-608, 2004.
[5] M.R. Lyu, J. Song and M. Chi, “ A comprehensive method for muvltilingual video text detection, localization and extraction,” IEEE Trans. CSVT, Vol. 15, no. 2, pp. 243-255, 2005.
[6] K. Jung, K. Kim, and A. Jain. Text information extraction in images and video: a survey. PR, 2004.
[7] J. Liang, D. Doermann, and H. Li. Camera-based analysis of text and documents: a survey, IJDAR, 2005.
[8] B. Epshtein, E. Ofek, and Y. Wexler, “Detecting text ivn natural scenes with stroke width transform,” in Proc. of CVPR, 2010.
[9] Andrej Ikica, Peter Peer, “An improved edge profile based method for text detection in images of natural scenes”, EUROCON-International Conference on Computer as a tool (EUROCON) 2011 IEEE: 1-4.
10] Jing Zhang and Rangachar Kasturi, “Text Detection Using Edge Gradient and Graph Spectrum”, 2010. International Conference on Pattern Recognition, 2010, IEEE, pp 3979-3982.
[11] Azadboni, M.K.; Behrad, A. “Text detection and character extraction in color images using FFT domain filtering and SVM classification”, 2012 Sixth International Symposium on Telecommunications (IST), 2012, IEEE, pp794-799. DOI: 10.1109/ISTEL.2012.6483094
[12] K. Wang and S. Belongie, “Word spotting in wild,” in Proc. of ECCV, 2010.
[13] T. Novitova, O. Barinova, P. Kohli and V. Lempitsky, “Large-lexican attribute-consostent text recognition in natural images,” in Proc. of ECCV, 2012.
[14] A. Shahab, F. Shafait, and A. Dengel, “ICDAR 2011 robust reading competition challenge 2: Reading text in scene images,” in Proc. of ICDAR, 2011.
[15] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu, ‘’Detecting texts of arbitrary orientation in natural images,” in Proc. of CVPR, 2012.
[16] K. Wang, B. Babenko and S. Belongie, “End-to-end scene text recognition,” in Proc. of ICCV, 2011.
[17] L. Neumann and J. Matas, “A method for text localization and recognition in real-world images,” in Proc. of ACCV, 2010.
[18] V. Wu, R. Manmatha and E. M. Riseman, “ Text Finder: an automatic system to detect and recognize text in images,” IEEE Trans. PAMI, Vol. 21, no. 11, pp. 1224-1229, 1999.

[19] C. Yao, X. Bai and W. Liu, "A Unified Framework for Multi-Oriented Text Detection and Recognition," IEEE Transaction on Image Processing, Vol. 23, no. 11, pp. 4737-4749, 2014