An Improved Scene Text and Document Image Binarization Scheme

Ranjit Ghoshal
St. Thomas' College of Engg. & Tech.
Kolkata, India
Email: ranjit.ghoshal.stcet@gmail.com

Ayan Banerjee
Lexmark Research & Development Corporation
Kolkata, India
Email: ayanbanerjee.stcet@gmail.com

***Abstract*— Identification of text portions have a crucial impact on intelligent transport systems, document image processing, robotics and content based image retrieval systems. So, an accurate text identification method is necessary for text based scene image processing tasks such as OCR. Scene text image binarization plays an important role in any text identification algorithm and hence in the OCR performance. In this work a novel approach to natural scene text image binarization by tracking the text boundary based on edge and gray level variance information. Further, broken boundaries are linked to construct the complete boundary map. Here, an adaptive threshold is determined based on boundary edge information to binarize the image effectively. Compared to other well known binarization methods, our method has been proved more effective in cases where the natural scene images have low contrast, low resolution, non-uniform illumination and noise. Our experiments are conducted on the datasets of ICDAR 2003 Robust Reading Competition, ICDAR 2011 Born Digital Dataset, Street View Text (SVT) Dataset, DIBCO dataset and our laboratory made Bangla Dataset. The experimental results are satisfactory.**

Keywords: Text Identification; Scene Image; Binarization; Connected components; Feature Extraction; SVM classifier

I. Introduction

Mobile phones, digital cameras, various hand held devices and state of the art intelligent robotic systems are becoming increasingly popular. They are equipped with various technologies that are useful in all spheres of life. Manufacturers are now enabling these devices with technologies like extraction and recognition of text from scene images, text-to-speech converter etc. These text based softwares work by extracting the text from the image. Text binarization is an important part of text extraction. Numerous methods exist for text binarization in document images but they cannot be directly applied on natural scene images as the size of the text can vary drastically from a few pixels to a large part of the image. Their orientation also changes from image to image. Scene images are much more complex than document images as the background is in most cases never simple and uniform as in the case of document images. Moreover, illumination variation caused by light reflection, shadows and other noise sources add significant complexity to the problem. Thus, to solve this problem, a dynamic approach is required. Conventional binarization techniques are classified into two categories (i) global thresholding and (ii) local adaptive thresholding. Global thresholoding is quite simple and effective for simple images with bimodal histogram. But in scene images the histogram is much more complex to give any fruitful result of Otsu's method [1]. Binarization of such an image using a single threshold value often leads to loss of textual information against the background. On the other hand local thresholding methods, Niblack [2], Sauvola [3] and Lu [4], apply window-based processing where the size of the window is crucial for text detection. In scene images the size of the text can vary drastically, thus to choose a window size is an impossible task. Among recent works, Gatos et al. [5] proposed a document image binarization model by combining multiple global and local adaptive methodologies and reinforcing it with edge information.

The paper is organized as follows: The proposed binarization scheme is presented in Sec.II and the experimental results and discussions are described in Sec.III. Finally, the conclusion is presented in Sec.IV.

II. Proposed Binarization Methodology

This section describes the details of our proposed binarization scheme. At first, a color scene text image, is converted into a grayscale image. This grayscale image is the input of our proposed binarization scheme. Our proposed method for scene text binarization is divided into three major parts: (i) Variance Computation (ii) Broken Edges (Boundary) Linking and (iii) Adaptive Thresholding.

Text connected components have a few basic properties based on which they can be separated from the image - (a) they have a distinct boundary that separates it from non-text regions (b) the whole grayscale value of the text components region is almost the same and is dissimilar from the background non-text components. These basic features of text regions are exploited to classify the image into text and non-text clusters. Generally, edge detection techniques are applied to find the boundary of a text. There are numerous methods to extract text based on its edge properties. But the basic problem with this method is that it is highly susceptible to noise. Edge detection works by computing the gradient of the image and then finding the high gradient line segments. Noise causes high gradient values and thus contribute to faulty edges. Here we use another method to detect the boundary. A common observation about a text boundary is that, if we traverse along a text boundary, the boundary points will produce a large variance of the gray

978-1-5386-3039-6/18/$31.00©2018 IEEE

values of the neighbouring pixels. If we slide a 5×5 window over the entire image it will produce high variance in the text boundary regions and comparatively low value in non-boundary (or uniform) portions of the image. Text boundary regions have comparatively higher gradient magnitudes as they consist of two different portions - text region and background region. But the value of this variance will be much smaller in the smooth portions of the image. Even noise points do not contribute to a significant increase in variance magnitude. Thus, this method is immune to noise and also effectively detects boundary regions with high accuracy. So, first we convert the input gray image into variance image. A higher value of the variance around a pixel makes the pixel darker while a pixel around which this variance is low, will tend to be more white. This is similar to the gradient map in edge detection techniques. Canny [6] edge detector uses the gradient map to find the edges. Here, we use the principle of Canny edge detector on the *variance image* (instead of gradient map) to find the edge map that we call the boundary lines.

The *boundary lines* (EG_1) come from the *variance image* and partially gives the boundaries of the text regions. But there are discontinuities in the boundaries owing to low variance regions. Thus to complete the boundary description we take the help of canny edge detector. We find the edge (EG_2) of the input gray image using canny edge detector with a very small threshold so that even the small edges become visible.

Out of all the edges detected by the Canny edge detector we keep only those edges which are connected with the *boundary*. To do this we perform *logical OR* operation on EG_1 and EG_2 and keep the output in L. Now we form a new matrix *(F1)* where we store all the connected components (CC's) of L which are connected with the boundary pixels of EG_1. Now, after performing this step, there might be small discontinuities in the boundary. We connect them using morphological bridge operation and store this in a map named *boundary*. Thus we get the *complete boundary map* of image which is almost noise free and can be used to binarize the image.

We obtain the horizontal run in a row of the *boundary* image from one boundary pixel to the next boundary pixel. We move from left to right for each row over the entire *boundary* image. Then we compute the mean value over the two 3×3 neighourhoods around these two boundary pixels (starting and ending). This is the place where there are major changes in the gray scale values. So we compute the mean of these eighteen pixels and then compare all the pixels inside these two neighourhoods with this mean. If it has a higher value than the mean, we assign 1 to the corresponding pixel; otherwise, we assign 0 to it. Then the same procedure is applied in vertical direction. Using this we create two matrices, namely, *horizontal run matrix* and *vertical run matrix*. We then find the final matrix by *logical AND* on both *horizontal run matrix* and *vertical run matrix*. This gives us the final binarized image *(F1)*. The algorithm is summarized as follows:

Proposed Binarization Algorithm

Step 1: Read the Image.

Step 2: Slide a $m \times m$ window (where m is user choice) over the entire image and compute the variance matrix. Here, we consider m=5.

Step 3: Apply Canny algorithm on variance matrix to find the edges (EG_1).

Step 4: Link the broken edges (boundaries). This can be done as follows:

1) First find the edges (EG_2) from the input gray image using canny edge detector with a very small threshold (here, 0.001) to obtain all possible edges. Out of all the edges (EG_2), keep only those edges which are connected with the EG_1.
2) Perform *logical OR* operation on EG_1 and EG_2 and consider the output in L. Now we form a new matrix ($F1$) where we store all the connected components (CC's) of L which are attached with the boundary pixels of EG_1.
3) Now, after performing this, there might be small discontinuities in the boundary. Connect them using morphological bridging operation.

Output of this step is *complete boundary map* (after linking the broken edges).

Step 4: Apply the following Steps for adaptive thresholding technique to obtain the final binarized image. First, obtain the horizontal run in a row of the *complete boundary map* image from one boundary pixel to the next boundary pixel. Move from left to right for each row over the entire *complete boundary map* image.

Step 5: Compute the mean value over the two 3×3 neighourhoods around these two boundary pixels (starting and ending). This is the place where there are major changes in the gray scale values. So, the mean of these eighteen pixels are computed and all the pixels inside these two neighourhoods with this mean are compared.

Step 6: If a pixel has a higher value than the mean, we assign 1 to the corresponding pixel; otherwise, we assign 0 to it. Let the outcome image is BW_1. The same procedure is applied in vertical direction.

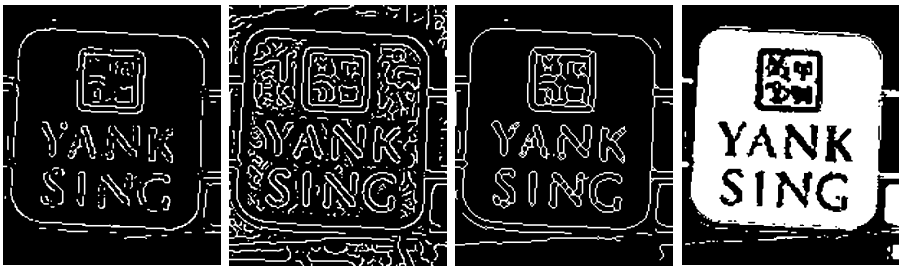
Step 7: Apply the same procedure for vertical direction and let the outcome is BW_2.

Step 8: Merge BW_1 and BW_2 to obtain the final binary image (F_1). This is done by applying *logical AND* operation.

978-1-5386-3039-6/18/$31.00©2018 IEEE

(a) (b) (c)

Fig. 1. (a) Input image. (b) Variance map. (c) Boundary lines (EG_1).



(a) (b) (c) (d)

Fig. 2. (a) Enlarged view of text portion of EG_1. (b) Canny edge map (EG_2) of input image. (c) Complete boundary map *(boundary)*. (d) Binarized image *(F1)*.

Let us consider an input image (Fig.1(a)). The variance image is presented in Fig.1(b). After applying Canny edge detection method, the boundary lines are presented in Fig.1(c). Enlarge view of the text portionis shown in Fig. 2(a). Further, Fig. 2(b) represents after applying Canny edge detector on the gray scale image with a low threshold value. Next, Fig. 2(c) presents, after linking the broken boundaries of Fig.1(c) with the help of Fig. 2(b). After applying the adaptive thresholding technique, we present the final binarized outcome in Fig.2(d). Full view of the final binarized image is presented in Fig.3

Fig. 3. Full view of final binarized image.

III. Results and Discussion

We have tested the proposed approach on four publicly avaibale datasets i.e. ICDAR 2003 Robust Reading Competition [7], ICDAR 2011 Born Digital Dataset, Street View Text (SVT) dataset [8] and DIBCO dataset ([9], [10]). Further we apply the proposed binarization method on our laboratory made Bangla Dataset. We consider ICDAR 2011 Born Digital Dataset for evaluation purpose since ground truths are available in this dataset. Some difficult images and their corresponding results obtained by applying Otsu, Niblack, Sauvola and our proposed methods are shown in Table I. These images are selected from ICDAR 2011 Born Digital Dataset and ICDAR 2003 Robust Reading Competition dataset. Further, we have applied the proposed binarization scheme on our laboratory made Bangla dataset. The results are presented respectivelly in Fig.4 and Fig.5. Finally, the proposed method is applied on DIBCO dataset. Fig.6, Fig.7 and Fig.8 represent the corresponding results. It is clearly visible that our proposed method outperforms the other methods like Otsu, Niblack and Savuola. The aim of our methodology is to binarize the text from scene images and not to extract text from the image. Thus to compare our binarization technique we consider only the text region of the binarized image with the ground truth to compute the precision, recall, F-measure. The performance assessment is based on a well established scheme Dance et al. [11] that calculates true positive (TP), false positive (FP) and false negative (FN) pixels in order to compute recall and precision metrics.

- A pixel is consider as TP if it is ON in both Ground Truth (GT) and binarization outcome images.
- A pixel is consider as FP if it is ON only in the binarization result image.
- A pixel is consider as FN if it is ON only in the GT image.

The recall metric describes the ratio of the number of pixels, which our method truly categorized as foreground, to the number of all pixels categorized as foreground from the ground truth image. Precision metric is the ratio of the number of pixels, which our method truly categorizes as foreground, to the number of all pixels which categorized as foreground. Setting C_{TP} as the number of TP pixels, C_{FP} as the number of FP pixels and C_{FN} as the number of FN pixels, recall (RC) and precision (PR) metrics are given as follows:

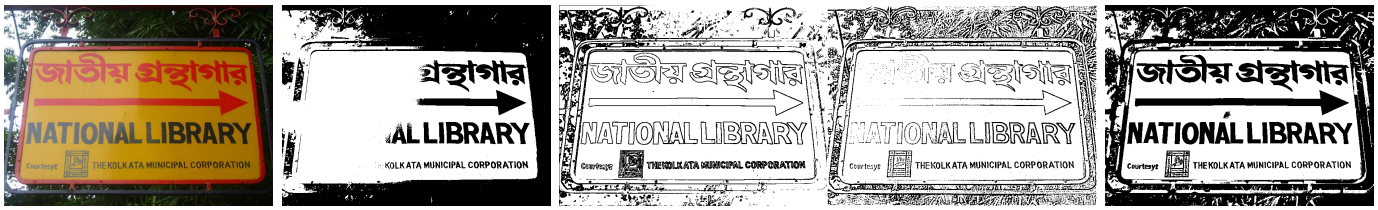
978-1-5386-3039-6/18/$31.00©2018 IEEE

TABLE I

RESULT OF VARIOUS BINARIZATION TECHNIQUES BASED ON ICDAR BORN DIGITAL AND ROBUST READING COMPETITION DATASET.

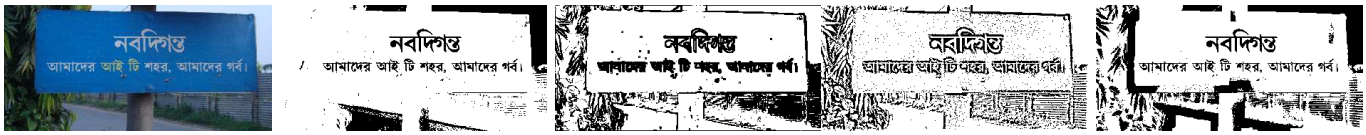
Input Image	Otsu	Niblack	Savola	Proposed Method

978-1-5386-3039-6/18/$31.00©2018 IEEE



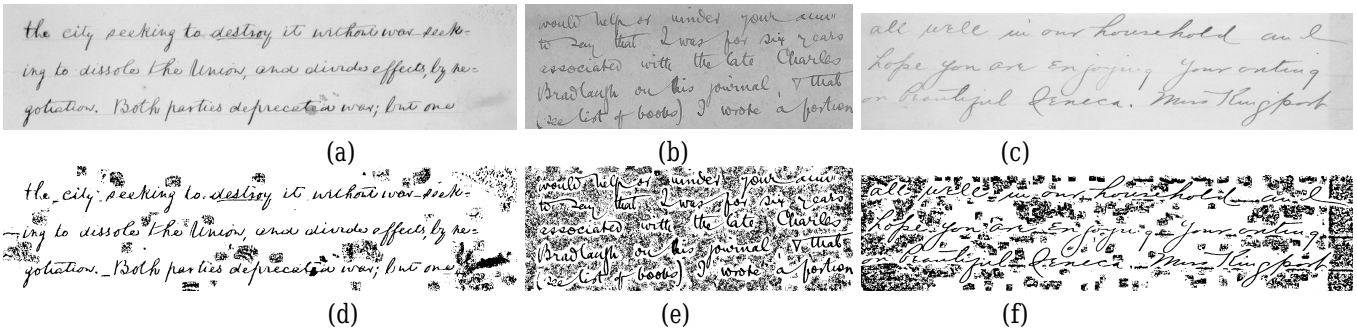
(a) (b) (c) (d) (e)

Fig. 4. Sample image from our laboratory made Bangla Dataset: (a) Input image. (b) Otsu. (c) Niblack. (d) Savola. and (e) Proposed TIBEV.



(a) (b) (c) (d) (e)

Fig. 5. Sample image from our laboratory made Bangla Dataset: (a) Input image. (b) Otsu. (c) Niblack. (d) Savola. and (e) Proposed TIBEV.



(a) (b) (c)

(d) (e) (f)

Fig. 6. Sample Images From DIBCO-2012 Dataset: (a), (b) and (c) are Input images. (d), (e) and (f) are corresponding Binarized Images.

(a) (b) (c)

(d) (e) (f)

Fig. 7. Sample Images From DIBCO-2010 Dataset: (a), (b) and (c) are Input images. (d), (e) and (f) are corresponding Binarized Images.

978-1-5386-3039-6/18/$31.00©2018 IEEE

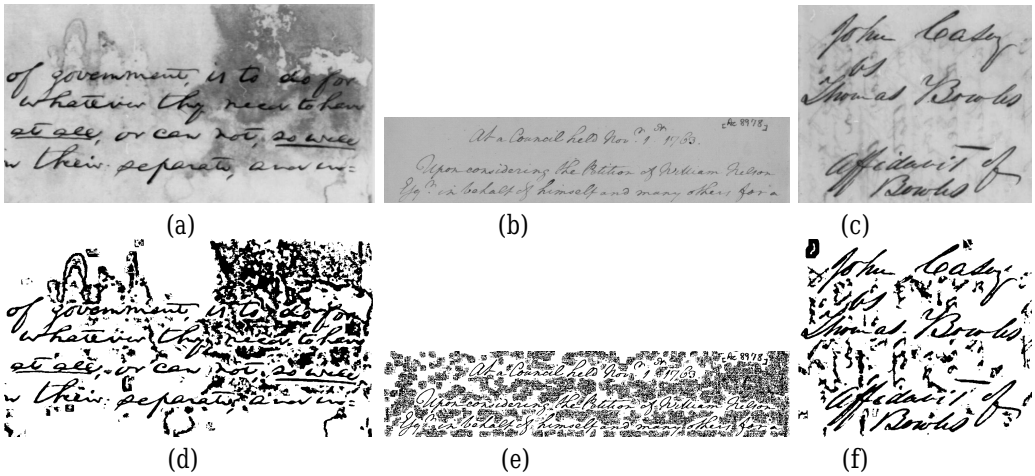


Fig. 8. Sample Images From DIBCO-2009 Dataset: (a), (b) and (c) are Input images. (d), (e) and (f) are corresponding Binarized Images.

$$RC = \frac{C_{TP}}{(C_{FN} + C_{TP})}$$

$$PR = \frac{C_{TP}}{(C_{FP} + C_{TP})}$$

Recall and Precision metric have values between zero and one. As these metrics approach one, the results get better.

The overall metric that is used for evaluation is the F-Measure (FM) which is calculated as follows:

$$FM = (2 \times RC \times PR/(RC + PR)) \times 100\%$$

TABLE II
AVERAGE VALUE OF F-MEASURE, RECALL AND PRECISION FOR EACH BINARIZATION TECHNIQUE.

	Otsu	Niblack	Sauvola	Proposed
Recall	**0.92**	0.87	0.91	0.88
Precision	0.67	0.36	0.14	**0.92**
FM	73.13	38.17	20.4	**89.50**

We have evaluated the proposed technique with a few well known image binarization schemes like Otsu, Niblack and Sauvola in terms of recall, precision and F-measure on the dataset of ICDAR 2011 Born Digital Dataset. As can be seen from the results shown in Table. II our proposed method significantly outperforms existing methods in terms of F-measure and precision.

IV. Conclusion and Future Scope

This work provides an improved scene text and document image binarization methodology. It uses both the edge and variance information of the input image. The proposed scheme is not very sensitive to image color, text font, skew and perspective variation. The proposed method is effective in terms of low contrast, non-uniform illumination and noisy text based scene images. The proposed method has been tested on four well known publically available datasets. The proposed method is also tested on our laboratory made Bangla dataset. Our experiments show that the proposed method outperforms most commonly used document binarization methods in terms of precision and F-measure. As far as recall is concerned, our method's performance is very close to the other three methods. Our future plan is to study the use of different types of machine learning tools to further improve the quality of the binarization scheme.

References

[1] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. System Man and Cybernetics*, vol. 9, no. 1, pp. 377–393, 1979.

[2] W. Niblack, *An Introduction to digital image processing.* Prentice Hall, Englewood Cliffs, 1986.

[3] J. Sauvola and M. Pietikinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 2, pp. 225–236, 2000.

[4] S. Lu, B. Su, and C. L. Tan, "Document image binarization using background estimation and stroke edge," *Int. Journal of Document Analysis and Recognition (IJDAR)*, vol. 13, no. 4, pp. 303–314, 2010.

[5] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Document image binarization by using a combination of multiple binarization techniques and adapted edge information," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2008.

[6] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence.*, vol. 8, no. 6, pp. 679–698, 1986.

[7] A. Shahab, F. Shafait, and A. Dengel, "Icdar robust reading competition challenge 2: Reading text in scene images," in *Proceedings of the International Conference of Document Analysis and Recognition*, 2011, pp. 1491–1496.

[8] K. Wang and B. Babenko, "End-to-end scene text recognition," in *ICCV*, 2011.

[9] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "Icdar 2009 document image binarization contest (dibco 2009)," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, 2009, pp. 1375–1382.

[10] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-dibco 2010 handwritten document image binarization competition," in *Proceedings of the International Conference on ICFHR*, 2010, pp. 727–732.

[11] C. R. Dance and M. Seegar, "On the evaluation of document analysis components by recall, precision, and accuracy," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition (ICDAR)*, 1999, pp. 713–716.

978-1-5386-3039-6/18/$31.00©2018 IEEE