




A Variance Based Image Binarization Scheme and Its Application in Text Segmentation

Ranjit Ghoshal^(✉), Aditya Saha, and Sayan Das

St. Thomas' College of Engineering and Technology, Kolkata, India
ranjit.ghoshal.stcet@gmail.com , adi96saha@gmail.com, sayandas896@gmail.com

Abstract. This paper presents a novel variance based image binarization scheme for automatic segmentation of text from low resolution images. First, the variance based binarization scheme is separately carried out on the three color planes of the image. Then, we merge these planes to obtain final binarized image. This creates several connected components (CCs). Now, these CCs are studied in order to segment possible text CCs. Now, a number of features that classify between text and non-text components, are considered. Further, KNN and SVM classifiers are applied for the present two class classification problem. For the training of KNN and SVM, ground-truth information of text CCs and our laboratory made non-text CCs are considered. We conduct extensive experiments on publicly available ICDAR 2011 Born Digital Data set. Concerning comparison, we consider a number of previously reported methods. Our binarization scheme significantly outperforms the existing methods and segmentation results are also satisfactory.

1 Introduction

Text in scene images includes important information and is exploited in many content-based video and image applications [1]. Text segmentation is a challenging problem due to variations of font, color, size and orientation etc. Binarization is also a great challenge, especially in the process of text based scene images where binarization result can directly influence the OCR rate. Several methods exist for binarization in document images but they cannot be directly applied on low resolution images. Conventional binarization techniques are either *global* [2] or *local* ([3, 4]) thresholding. Existing techniques for scene text segmentation can generally be classified into two sets: sliding window [5] and CC [6] based schemes. Sliding window based schemes use a sliding window to find for possible texts in the scene image and then use machine learning methodologies to identify text. CC based methods separate out character candidates from scene images by CC analysis. Due to their relatively simple implementation, CC-based methods are widely used. Here, we take an interest into color images embedding text. In the following sections, our methods are presented.

2 Variance Based Image Binarization Scheme

Consider a color image which consists of red (R), green (G) and blue (B) planes. Now, it is required to retrieve the information from each plane. Further, the image contains highly varying gray pixel values which make binarization a difficult task. This can be overcome to a great extent using variance to perform binarization.

Each plane is separately passed through the binarization process. First, the variance matrix is calculated from the gray scale image, which marks the change in pixel intensities in the image. Binarizing the variance matrix, we separate the image into two regions, one having high variance values and other having low variance values. Now, using each region, two gray scale images (one from the white and other from the black region of the binarized image) are formed. Binarizing these two images separately will produce more even binarization as they don't contain any fluctuating pixel intensities. These gray scale images are binarized by a window based Otsu binarization method, as illustrated in the algorithm. Finally, binarized image from each plane is merged together to form the final binarized image. The details have been presented in Algorithm 1.

Algorithm 1. Algo. of Binarization

Input: Gray Scale Image **G**

Output: Binarized Image

- Step1: Find the variance matrix **M** of **G** by sliding a 5×5 window throughout **G** and replace the variance of the window at the centre pixel of the window in **M**.
 - Step2: Apply Otsu's method on variance matrix to obtain a binerized images (**MBW**).
 - Step3: Obtain two gray scale images from **MBW** by considering pixel intensities of **G**. Let these are **GW** and **GB** respectively.
 - Step4: Apply Canny's method On **GW** to obtain an edge image say **GWE**. Further, apply Otsu's method on **GW** and let it be **GWBW**. Let complement of **GWBW** is \sim **GWBW**.
 - Step5: Run two windows (**WBW** and **WBC**) of size 21×21 over the images **GWBW** and \sim **GWBW** respectively.
 - Step6: Now construct two matrices **M1** and **M2** and are defined as follows:
 $M1(i,j) = M1(i,j) + WBW(i,j)$, where $1 \leq i \leq m$ and $1 \leq j \leq n$
 $M2(i,j) = M2(i,j) + WBC(i,j)$, where $1 \leq i \leq m$ and $1 \leq j \leq n$
 - Step7: From **M1** and **M2** final binerized image of **GW** is calculated as follows:
 $GWBWF(i,j) = 1$, where $M1(i,j) > M2(i,j)$ and $1 \leq i \leq m$ and $1 \leq j \leq n$
 - Step8: Similiarly final binerized images of **GB** is calculates, say **GBBWF**.
 - Step9: Merge **GWBWF** and **GBBWF** to obtain the final resultant Binerized Image.
-

Consider the RGB image(Fig. 1(a)) as an input. The image has separated into three different planes(Figs. 1(b), (c) and (d)). Consider the R plane, which is passed into the proposed binarization algorithm. Figure 2(a) represents the binarized image of variance matrix, which is calculated by moving a 5×5 window throughout the image. For white pixels and black pixels, separate gray scale images are again formed and a window based Otsu algorithm is performed.



Fig. 1. (a) Input color image. (b) R plane. (c) G plane. (d) B plane. (Color figure online)

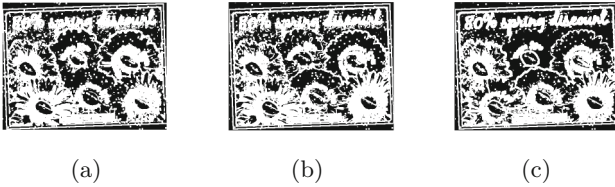


Fig. 2. Binarization for variance matrices on (a) R plane, (b) G plane and (c) B plane.

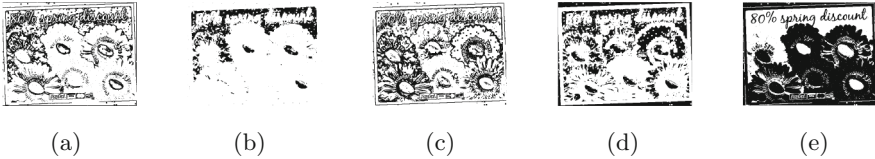


Fig. 3. Binarized image of gray scale corresponding to (a) White pixel in 2(a). (b) Black pixel in 2(a). (c) White pixel in 2(b). (d) Black pixel in 2(b). (e) Merged image of 3(a) and 3(b).

Results are presented in Figs. 3(a) and (b). These are merged to obtain binarized image for R plane(Fig. 3(e)). Similarly, binarized images for G and B planes are presented in Figs. 4(c) and (d). Binarized images from each planes are merged to obtain the final binarized image(Fig. 4(e)).

3 Shape Based Feature Extraction

Image binarization creates a number of CCs. In order to segment text, we have considered a number of features from each CC.

AL: Axial ratio (AL) of a CC is the ratio of the length of the two axes to each other - the longer axis divided by the shorter.

LO: Number of lobes in a CC [7].

A: Aspect ratio of a CC [7].

E: Elongation ratio of a CC [7].

O: Object to background pixels ratio of a CC [7].



Fig. 4. Binerized image of gray scale corresponding to (a) White pixel in 2(c) and (b) Black pixel in 2(c). Merged images (c) (3(c) and 3(d)), (d) (4(a) and 4(b)) and (e) (3(e), 4(c), and 4(d)).

AR: Area ratio of a CC. It is the ratio of (area of the CC and area of input image).

L: Length ratio (L) of a CC. It is the ratio of (max (height of CC, width of CC), max (height of the I, width of the I)), where I is the input image.

Now, we construct the feature vector $\mathbf{Y} = \{AL, LO, A, E, O, AR, L\}$ for a CC.

4 KNN and SVM Based Text Segmentation

To segment the text components, K-NN and SVM classifiers are applied. The feature vector \mathbf{Y} for text and non-text CCs are calculated. The dataset contains 420 train images and 102 test images. Ground truth information of train images are used to create the feature file for 21700 text components. Next, the input images are binarized with our binarization method. Then the components present in the ground truth images are eliminated. Thus we create 78800 non-text CCs. These are used to prepare the feature file for non-text components. Based on these feature files, K-NN and SVM classifiers are trained separately. To segment the text components from test images, an input image is binarized using our binarization method and the feature vector \mathbf{Y} is obtained. Now, each CC is fed to both the trained K-NN and SVM classifiers to decide whether the component is text or non-text. Thus, two output images from K-NN and SVM classifiers are obtained. Finally, these two images are merged using *logical OR* operation to get the final image consisting only text components.

5 Results and Discussion

The experimental results are obtained on ICDAR 2011 Born Digital Dataset [8]. These images are inherently low-resolution. So, automatic segmentation of text is therefore an important project. Our experiments are divided into two parts based on our aim of the paper.

5.1 Results of Binarization Scheme

Let us first pictorially observe some binarization results. A few example results are presented in Table 2. First column represents the sample input images and second column presents the corresponding binarized images. Evaluation of our binarization scheme is done in terms of the precision, recall and F-measure [7]. Also the performance of our binarization scheme has been compared with a few known methods in terms of recall, precision and FM on ICDAR 2011 Born Digital data set. It can be seen from the results (Table 1) that our binarization method has significantly outperformed.

Table 1. Recall, Precision and FM for different binarization technique.

	Proposed	Otsu	Niblack	Sauvola	Bhatt et al.	Kumar et al.
		[2]	[4]	[3]	[9]	[10]
Recall	91.15	88.98	87	91	91.14	85.56
Precision	71.36	65.36	36	14	47.85	47.09
FM	73.78	65.05	38.17	20.4	53.81	46.81

5.2 Text Identification Results

We present the text segmentation results obtained by our KNN and SVM classifiers. A few images and their corresponding segmented text using KNN, SVM and merged KNN and SVM classifier are presented in Table 2. Visually, it is clear that our approach good towards text segmentation. A robust comparison analysis has been performed by means of Recall, Precision and FM values of our different classification methods obtained on the basis of ICDAR 2011 Born Digital data set images are presented in the Table 3. Final evaluation of our scheme is

Table 2. Input images, binarized images and segmented text (using KNN, SVM and merged of KNN and SVM) are presented respectively 1st, 2nd, 3rd, 4th and 5th columns.












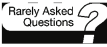



RGB	BW	KNN	SVM	Merged
				
				
				

Table 3. Recall, Precision and FM for different text segmentation methods.

	Merged KNN and SVM	SVM	KNN	TexStar	SASA
Recall	77.72	60.80	63.49	65.23	71.28
Precision	45.53	50.89	52.53	63.63	55.54
FM	53.63	51.07	52.85	64.64	62.52

presented by comparing with other known techniques. The ICDAR 2011 Robust Reading Competition presented evaluation results of a number of methods from different participants. In Table 3, a few of these techniques are compared with our scheme. Our scheme has achieved highest recall (77.72).

6 Summary and Future Scope

This paper provides a new variance based image binarization scheme and its application in text segmentation. A number of shape based features are defined towards segmentation of text. Then, SVM and KNN classifiers are trained for classification of text and non-text. Finally, the results obtain from SVM and KNN are merged to get the final segmented text. The proposed method is very effective for low resolution images. Future study may aim towards combining machine learning tools to improve the binarization scheme.

References

1. Yin, X.C., Hao, H.W., Sun, J., Naoi, S.: Robust vanishing point detection for mobile cam-based documents. In: Proceedings of ICDAR, pp. 136–140 (2011)
2. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 377–393 (1979)
3. Sauvola, J., Pietikinen, M.: Adaptive document image binarization. *Pattern Recogn.* **2**, 225–236 (2000)
4. Niblack, W.: *An Introduction to Digital Image Processing*. Prentice Hall, Englewood Cliffs (1986)
5. Lee, J.J., Lee, P.H., Lee, S.W., Yuille, A., Koch, C.: Adaboost for text detection in natural scene. In: ICDAR, pp. 429–434 (2011)
6. Yi, C., Tian, Y.: Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification. *IEEE Trans. Image Process.* **21**(9), 4256–4268 (2012)
7. Ghoshal, R., Roy, A., Parui, S.K.: A copula based statistical model for text extraction from scene images. In: Maji, P., Ghosh, A., Murty, M.N., Ghosh, K., Pal, S.K. (eds.) *PREMI 2013. LNCS*, vol. 8251, pp. 489–494. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-45062-4_67](https://doi.org/10.1007/978-3-642-45062-4_67)
8. Karatzas, D., Robles Mestre, S., Mas, J., Nourbakhsh, F., Roy, P.P.: Icdar 2011 robust reading competition-challenge 1: Reading text in born-digital images (web and email). In: ICDAR, pp. 1485–1490 (2011)

9. Bhattacharya, U., Parui, S.K., Mondal, S.: Devanagari and bangla text extraction from natural scene images. In: Proceedings of the International Conference on Document Analysis and Recognition (ICDAR), pp. 171–175 (2009)
10. Kumar, D., Ramakrishnan, A.G.: Octymist: otsu-canny minimal spanning tree for born-digital images. In: DAR, DAS 2012, pp. 389–393 (2012)