Image Binarization Based On ICA Approach for Optical Character Recognition

M. Jyostna Grace[1], K. Subhashini[2]

PG Student [ECE], Dept. of ECE, K.L.M. College of Engineering for Women, Kadapa, Andhra Pradesh, India[1]

Assistant Professor, Dept. of ECE, K.L.M. College of Engineering for Women, Kadapa, Andhra Pradesh, India[2]

ABSTRACT:Image binarization plays a vital role in text segmentation which is used in OCR application. Binarization of text in degraded images is a challenging task due to the variations in colour, size, and font of the text and the results are often affected by complex backgrounds, different lighting conditions, shadows and reflections. A robust solution to this problem can significantly enhance the accuracy of scene text recognition algorithms leading to a variety of applications such as scene understanding, automatic localization, navigation, and image retrieval. In this paper, we propose a novel method to extract and binarize text from images that contains complex background. We use an Independent Component Analysis (ICA) based technique to map out the text region, which is inherently uniform in nature, while removing shadows, specularity and reflections, which are included in the background. This algorithm works better on images with different degradations. We implement our method on DIBCO dataset then we compare our robust algorithm with state-of-art criteria like binarization based on Otsu method and we can prove that our algorithm will give better results.

KEYWORDS: Image binarization,thresholding,adaptive local image contrast, connected components, Independent Component Analysis.

I.INTRODUCTION

In the recent years, content-based image analysis techniques have received more attention with the advent of various digital image capture devices. The images captured by these devices can vary dramatically depending on lighting conditions,reflections, shadows and specularities. These images contain numerous degradations such as uneven lighting, complex background, multiple colors, blur etc. We propose a method for removing reflections, shadows and specularities in natural scene text images and extracting out the text from a single image. There are many algorithms that aim at extracting foreground text from background in images but thresholdingremains one of the oldest form that is used in many image processing applications. Many sophisticated approaches often have thresholding as a pre-processing step. It is often used to segment images consisting of bright objects against dark backgrounds or vice versa [1], [3], [4]. It typically works well for images where the foreground and background are clearly defined. For color thresholding images, most algorithms convert the RGB image into grayscale but here we will make use of the RGB channel as three different sources.

Traditional thresholding based binarization can begrouped into two categories: the one which uses global threshold for the given images like Otsu [2], Kittler et al. [5] and the one with local thresholds like Sauvola [6], Niblack [9]. In global thresholding methods [2], [7], global thresholds are used for all pixels in image. These methods are fast and robust as they use a single threshold based on the global histogram of the gray-value pixels of the image. But they are not suitable for complex and degraded scene images. Also selecting the right threshold for the whole image is usually a challenge because it is difficult for the thresholding algorithmto differentiate foreground text from complex background. On the other hand, local or adaptive binarization [8] methods changes the threshold over the image according to local region properties. Adaptive thresholding addresses variations in local intensities throughout the image. In these methods, a pre-pixel threshold is computed based on a local window around each pixel. Thus, different threshold values are used for different parts of the image. These methods are proposed to overcome global binarization drawbacks but they can be sensitive to image artifacts found in natural scene text images like shadows, specularities

and reflections. Mishra et al [13] has recently formulated the problem of binarization as an MRF optimization problem. The method shows superior performance over traditional binarization methods on many images, and we use it as the basis for our comparisons. However, their method is sensitive to the initial auto seeding process. Zhou et al [14] also addresses the segmentation problem in text images which contains specular highlightsand focal blur. On the other hand, we propose a method that removes shadows, specularity and reflections and thus produces a clean binary images even for the images with complex background. The primary issue related to binarizing text from scene images is the presence of complex/textured background. When the background is uneven as a resultof poor or non-uniform lighting conditions, the image will not be segmented correctly by a fixed gray-level threshold. These complex backgrounds vary dramatically depending on lighting, specularities, reflections and shadows. The above methods applied directly to such images give poor results and cannot be used in OCR systems.
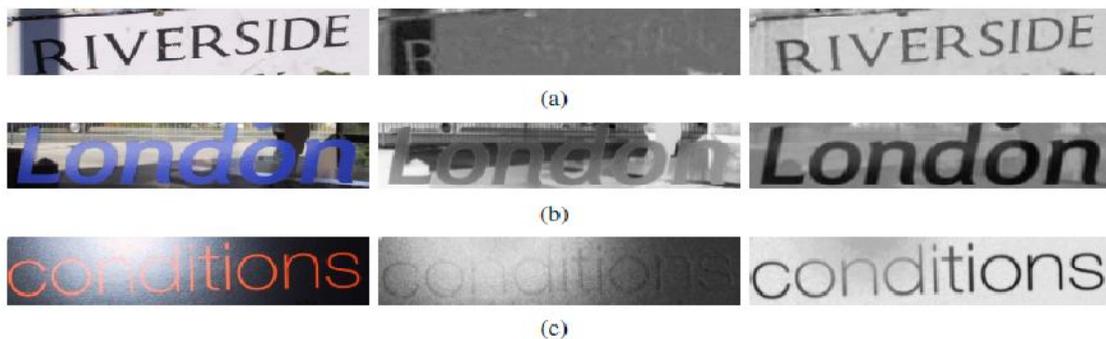

Figure 1. Foreground and Background Extracted: (a) Shadowed background and foreground text (b) Reflective background and foreground text (c) Specular background and foreground text.

In this paper, we do an ICA based decomposition which enables us to separate text from complex backgrounds containing, reflections, shadows and specularities. For binarization, we apply a global thresholding method on the independent components of the image and that with maximum textual properties is used for extracting the foreground text. Binarization results show significant improvement in the extraction of text over other methods. Some of the word Images that we used for experiments are shown in Fig 1. The remainder of the paper is organized as follows. We discuss the general ICA model in Section II followed by the detailed binarization process in section III. Our existing method is described in Section IV.We then show the results of the proposed method on a variety of images from the ICDAR dataset in section V, followed by the conclusions and potential directions for further improvement in section VI.

II. INDEPENDENT COMPONENT ANALYSIS (ICA) MODEL

Independent Component Analysis (ICA) has been an active research topic because of its potential applications in signal and image processing. The goal of ICA is to separateindependent source signals from the observed signals, which is assumed to be the linear mixtures of independent source components. The mathematical model of ICA is formulated by mixture processing and an explicit decomposition processing.

Assume there exists a set of 'n' unknown source signals $S = \{s_1, s_2, \ldots . s_n\}$. The assumptions of the components $\{s_i\}$ include mutually independent, stationary and zero mean. A set of observed signals $X = \{x_1, x_2, \ldots . x_n\}$, are regarded as the mixture of the source components. The most frequently considered mixing model is the linear instantaneous noise free model, which is described as:

$$x_i = \sum_{j=1}^{n} a_{ij} s_j \qquad (1)$$

Or in the matrix notation

$$X = A.S \qquad (2)$$

where A is an unknown full rank mixing matrix, which is also called mixture matrix. Eqn.1 assumes that there exists a linear relationship between the sources S and the observations X. In our case, 'n' is equal to 3.

III.BINARIZATION PROCESS

A wide variety of ICA algorithms are available in the literature [11], [12]. These algorithms differ from each other on the basis of the choice of objective function and selected optimization scheme. Here we use a fast fixed point ICAalgorithm to separate out the text from complex backgroundin images. A Blind Source Separation method based on Singular Value Decomposition [10] can also be used. Fig2 shows the complete framework for the proposed method.
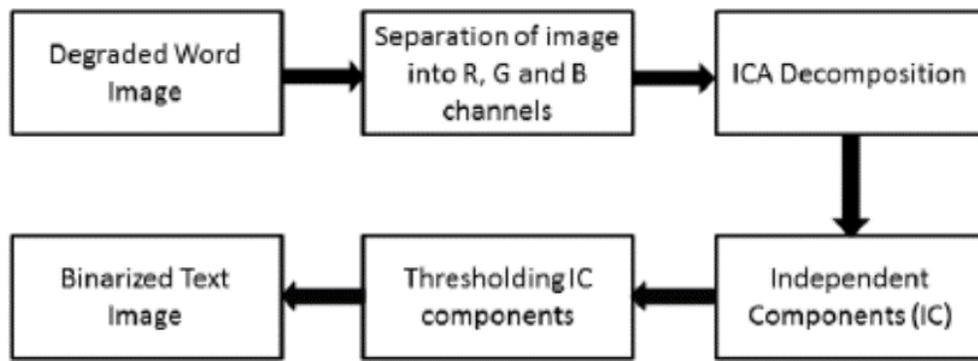

Figure 2. Framework for the proposed method

A. Separation Model

Consider the text image as a mixture of pixels from three different sources and assume it to be a noiseless instantaneous mixture. We use a single image i.e., its R, G and B channels as three observed signals. Therefore, we can define that the color intensity at each pixel from these three observed signals mix linearly to give the resultant color intensity at that pixel. Denoting these mixture images in row vector form as x_r, x_g and x_b, the linear mixing of the sources at a particular pixel k can be expressed in matrix form as follows:

$$\underbrace{\begin{bmatrix} x_r(k) \\ x_g(k) \\ x_b(k) \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} s_1(k) \\ s_2(k) \\ s_3(k) \end{bmatrix}}_{S} \qquad (3)$$

Where X is an instantaneous linear mixture of source images at pixel k, A is the instantaneous 3x3 square mixing matrix and S is the source images which add up to form the color intensity at pixel k. The mixed images in X contain a linear combination of the source images in S. We find the mixing matrix A and sources S using fixed point ICA algorithm. From this step, we get three independent sources or components. Fig. 3 shows the background and the foreground extracted.
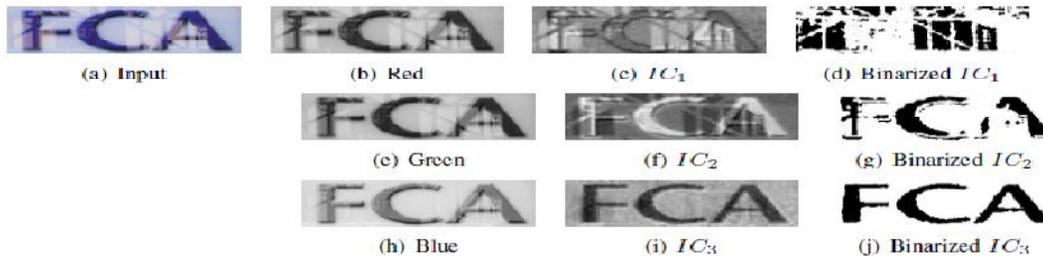
Figure 3. (a) Original word image; (b),(e),(h) R, G and B channel respectively; (c),(f), (i) Independent Components; (d),(g),(j) Binarized image

The resultant independent components for a particular word image can be seen in Fig. 3 which shows the independent component free from reflective background and containing maximum information of the foreground text.

B. Thresholding

Otsu thresholding [2] is a well-known algorithm that determines a global threshold for an image by minimizing the within-class variance for the resulting classes (foreground pixels and background pixels). This is done by equivalently maximizing the between-class variance σ_B^2 (T) for a given threshold T:

$$\sigma_B^2 = \alpha_1(T)\alpha_2(T)[\mu_1(T) \quad \mu_2(T)]^2 \qquad (4)$$

Whereα_i denotes the number of pixels in each class, μ_i denotes the mean of each class, and T is the value of the potential threshold. We apply this thresholding algorithm on all the three independent components to get the binarized image. We can also apply Kittler [5] algorithm which is also a global thresholding method. To find the IC that contains the foreground text, we examine the connected components (CC) in the binarization of each IC. For each binarized image, we extract the following features from the CCs: average aspect ratio, variance of CC size, and the deviation from linearity of their centroids. A simple linear classifier is designed to separate the text and non-text classes in the above feature space. After binarization, we identify the connected components and remove non-text portions based on size and aspect ratio. In practice, we note that a simpler global thresholding scheme works well in most cases.

IV. EXISTING METHOD

A. Contrast Image Construction

The image gradient has been widely used for edge detection and it can be used to detect the text stroke edges of the document images effectively that have a uniform document background. On the other hand, it often detects many nonstroke edges from the background of degraded document that often contains certain image variations due to noise, uneven lighting, bleed-through, etc. To extract only the stroke edges properly, the image gradient needs to be normalized to compensate the image variation within the document background.

The local contrast evaluated by the local image maximum and minimum is used to suppress the background variation. In particular, the difference between the local maximum and the local minimum, captures the local image difference that is similar to the traditional image gradient and a normalization factor suppresses the image variation within the document background. For image Pixels within bright regions, it will produce a large normalization factor to neutralize the numerator and accordingly result in a relatively low image contrast. For the image pixels within dark regions, it will produce a small denominator and accordingly result in a relatively high image contrast. However, this image contrast has one typical limitation that it may not handle document images with the bright text properly. This is because a weak

contrast will be calculated for stroke edges of the bright text. To overcome this over-normalization problem, we combine the local image contrast with the local image gradient and derive an adaptive local image contrast as follows:

$$C_a(i, j) = \alpha C(i, j) + (1 - \alpha)(I_{\max}(i, j) - I_{\min}(i, j))$$

B. Local Threshold Estimation

The text can then be extracted from the document background pixels once the high contrast stroke edge pixels are detected properly. Two characteristics can be observed from different kinds of document images [5]: First, the text pixels are close to the detected text stroke edge pixels. Second, there is a distinct intensity difference between the high contrast stroke edge pixels and the surrounding background pixels. The document image text can thus be extracted based on the detected text stroke edge pixels

$$R(x, y) = \begin{cases} 1 & I(x, y) \le E_{\text{mean}} + \frac{E_{\text{std}}}{2} \\ 0 & \text{otherwise} \end{cases}$$

Where E_{mean} and E_{std} are the mean and standard deviation of the intensity of the detected text stroke edge pixels within a neighborhood window W, respectively. The neighborhood window should be at least larger than the stroke width in order to contain stroke edge pixels. So the size of the neighborhood window W can be set based on the stroke width of the document image under study, EW, which can be estimated from the detected stroke edges.

V. EXPERIMENTAL RESULTS AND ANALYSIS

We used the ICDAR 2003 Robust Word Recognition Dataset [15] for our experiments. For qualitative evaluation, we selected the word images that had complex reflective, shadowed and specular background. We separate these word images into Red, Green and Blue channels assuming that these are the mixture images of the independent source images that contains the foreground (text) and background. These three images are used for extracting the foreground as described before. Text Localization and Recognition Results of proposed Binarization Method is shown in Fig. 4 below.

Figure4: Text Localization and Recognition Results of proposed Binarization Method

We compare the performance of our method with four well known thresholding algorithms i.e., Kittler [5], Otsu [2], Niblack [9] and Sauvola [6]. We also compare with the recent method by Mishra et al [13]. It although performs well for many images but severely fails in cases of shadows, high illumination variations in the image. This poor show is likely due to fact that performance of the algorithm heavily depends on initial seeds. We show both qualitative and quantitative results of the proposed method. The qualitative results are shown in Fig. 5. We took around 50 images from the dataset and generated its ground truth images for pixel level accuracy. We use well known measures like precision, recall and F-score to compare the proposed method with different binarization methods we also use OCR accuracy to show the effectiveness of our method. Note that we are only using the subset of images that are most degraded by shadowing illumination variations, noise and specular reflections. The results of thresholding schemes are too poor for the OCR.

Figure 5. Comparison of Binarization algorithms and the proposed method (From left to right Original, MRF, Kittler, Otsu, Niblack, Sauvola, Proposed); (a)-(d) Text containing reflective background; (e)-(f) Text containing shadowed background; (g)-(j) Text containing specular background; (k) Text containing reflective, shadowed and specular background.

The results show that the proposed method is an effective method and performs better than other methods in the case where images have complex background. Fig. 6 shows that our technique can also be applied to text image containing two different types of colored text.
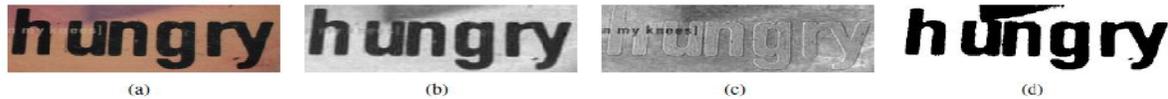
Figure 6. (a) Image containing Text over another Text; (b) Foreground Text; (c) Background Text; (d) Text extracted

We analyze that the above methods do not work in the case where there is a complex and textured background in the images. It is not that these methods do not work at all, no single algorithm works well for all types of images. Thus we can say that our method can extract out the text embedded in complex reflective, shadowed and specular background. Our method fails in cases where foreground text and the background are of the same color. Moreover, the approach works only with color images.

VI. CONCLUSION

We have proposed an effective method to binarize text from colored scene text images with reflective, shadowed and specular background. By using a blind source separation technique followed by global thresholding, we are able to clearly separate the text portion of the image from the background. ICA decomposition enables us to separate reflections, shadows and specularities from natural scene texts so that the global thresholding methods can be applied afterwards to binarize the text image. Experimental results on ICDAR dataset demonstrate the superiority of our method over other existing methods. Possible directions for improvement of the approach includes a patch-based SVM. Classification for thresholding as well as integration of the results with a spatially aware optimization such as MRF,working with text where the foreground and background have same color is also of great interest.

REFERENCES

1. R. M. Haralick and L. G. Shapiro, Image segmentation techniques, Computer Vision, Graphics and Image Processing, vol.
2. 29, pp. 100-132, 1985.
3. N. Otsu, A threshold selection method from gray-level histograms, IEEE Systems, Man, and Cybernetics Society, vol. 9, pp. 62-66, 1979.
4. P. K. Sahoo and S. Soltani and A. K. C. Wong and Y. C. Chen, A survey of thresholding techniques, Computer Vision,
5. Graphics and Image Processing, vol. 41, pp. 233-260, 1988.
6. N. R. Pal and S. K. Pal, A review on image segmentation techniques, Pattern Recognition, vol. 26, pp. 1227-1249, 1993.
7. J. Kittler and J. Illingworth and J. Foglein, Threshold selection based on a simple image statistic, Computer Vision, Graphicsand Image Processing, vol. 30, pp. 125-147, 1985.
8. J. J. Sauvola and M. Pietikainen, Adaptive document image binarization, Pattern Recognition, vol. 33, pp. 225-236, 2000.
9. P. Sahoo and G. Arora, A thresholding method based on twodimensional Renyis entropy, Pattern Recognition, vol. 37, pp.
10. 1149-1161, 2004.
11. J. Bernsen, Dynamic thresholding of gray level images, International Conference on Pattern Recognition, pp. 1251-1255,
12. 1986.
13. W. Niblack, An introduction to digital image processing, New York: Prentice Hall, 1986.
14. R. Szupiluk, A. Cichocki, Blind signal separation using second order statistics, Proc. of SPETO, pp. 485-488, 2001.
    A. Hyvarinen and J. Karhunen and E. Oja, Independent Component Analysis, John Wiley and Sons, New York, 2001.
15. Hyvarinen and E. Oja, Independent component analysis: Algorithms and applications, Neural Networks, vol. 13, pp.
16. 411-430, 2001.
    A. Mishra, K. Alahari, and C.V Jawahar, an MRF Model for Binarization of Natural Scene Text, ICDAR 2011
17. Y. Zhou, J. Feild, E Miller and R Wang, Scene Text Segmentation via Inverse Rendering, ICDAR 2013.
18. The ICDAR 2003 Robust Reading Datasets,http://algoval.essex.ac.uk/icdar/RobustWord.html