2004 IEEE International Conference on Multimedia and Expo (ICME)

ADAPTIVE BINARIZATION METHOD FOR DOCUMENT IMAGE ANALYSIS

Meng-Ling Feng and *Yap-Peng Tan*

School of Electrical and Electronic Engineering
Nanyang Technological University, Singapore

ABSTRACT

This paper proposed an adaptive binarization method, based on a criterion of maximizing local contrast, for document image analysis. The proposed method has overcome, to a large extent, the general problems of poor quality document images, such as non-uniform illumination, undesirable shadows and random noise. It was tested against a variety of challenging images, and the experimental results are presented to show the effectiveness and superiority of the proposed method.

1. INTRODUCTION

In these few years we have witnessed the migration of cheap and tiny imaging devices into portable and mobile computing applicants, such as PDAs and mobile phones [1]. Plenty of new applications and uses have become possible with this revolutionary combination. One notable example is the quick capture and analysis of document images for faxing, auto note taking, OCR, etc.

This combination not only stimulates new exciting applications, but it brings in new technical challenges to the field of document image understanding as well. This is because, due to the constraints of capturing devices' size and cost, the quality of document images captured by handheld portable devices is relatively low. The common problems in these poor quality text images include: (1) undesirable shadows because of the lack of appropriate lighting setup, (2) variable background intensity due to non-uniform illumination, (3) low contrast and (4) large amount of random noise due to limited sensitivity of low-cost camera. Therefore, to have an accurate analysis of the document image, a versatile binarization method, which can correctly remove noise and unnecessary background and reliably keep all useful information, becomes indispensable.

During the past decade, many binarization techniques have been proposed [2]. Few of them achieve satisfactory performance in dealing with poor quality images, say obtained from handheld cameras. This paper proposes an improved binarization method, based on a criterion of maximizing local contrast, to extract information from such degraded and bad quality document images. Our experiments show that the proposed method has outstanding robustness against non-uniform illumination, low image contrast, shadows and stochastic noise.

2. REVIEW OF RELATED WORK

Binarization, which scans gray-scale text images into two levels, is usually the first stage in document image understanding systems. This is because the use of bi-level information greatly reduces the computational load and the analysis algorithm complexity. Moreover, binarization is also the most critical stage, since any error in this phase will pass down to the following ones. One of the popular binarization approaches is gray value thresholding, and the corresponding techniques can be further classified into *global* and *local* thresholding.

2.1. Global Thresholding

Global thresholding uses only one threshold value, which is estimated based on statistics or heuristics on global image attributes, to classify image pixels into foreground or background. The major drawback of global thresholding techniques is that it cannot differentiate those pixels which share the same gray level but do not belong to the same group. Otsu's method [3] is one of the best global thresholding methods. It works well with clearly scanned images, but it performs unsatisfactorily for those poor quality images that have low contrast and non-uniform illumination.

2.2. Local Thresholding

In local thresholding, the threshold values are spatially varied and determined based on the local contents of the target image. In comparison with global techniques, local thresholding techniques have better performance against noise and error, especially when dealing with information near texts or objects. According to Trier's survey [4], Yanowitz-Bruckstein's method [5] and Niblack's method [6] are two of the best performing local thresholding methods. Yanowitz-Bruckstein's method is extraordinary complicated and thus requires very large computational power. This makes it infeasible and too expensive for real system implementations. On the other hand, Niblack's method is simple and effective. As a result, we decided to focus on Niblack's method. From the study of Niblack's algorithm, we were inspired and eventually derived a better local thresholding method.

The concept of Niblack's algorithm is to build a threshold surface, based on the local mean, m, and local standard deviation, s, computed in a small neighborhood of each pixel in the form of

$$T = m + k \cdot s$$

where k is a negative constant. This algorithm, however, produces a large amount of binarization noise in those areas that contain no text objects (see Fig. 2 (b)), resulting in costly post-processing.

Sauvola et al. [7] improved the algorithm by adding a hypothesis on the gray values of text and background pixels (text pixels have gray level near 0 and background pixels have gray level near 255), and formulate the local threshold value as

$$T = m \cdot (1 - k \cdot (1 - \frac{s}{R}))$$

where R is the dynamic range of standard deviation fixed to 128 and k is a constant set to 0.5. The improved method performs better for well-scanned document images, but it faces difficulties in dealing with images that do no correspond with the hypothesis, for instance, documents scanned with very dark lighting, especially those in which the gray levels of text and background are quite close to each other.

To overcome this problem, Christian et al. [8] further changed the formula as follow to normalize the contrast and the gray level mean of the image

$$T = (1-k)\cdot m + k\cdot M + k\cdot \frac{s}{R}\cdot (m-M)$$

where M is the minimum gray level of the whole image, R is set to the maximum of the standard deviations of the image and k is fixed at 0.5. This approach achieves the best binarization results among the three. Nevertheless, since the algorithm focuses only on the maximum standard deviation of the entire image, its performance degrades when the input image involves great changes in background luminance.

3. PROPOSED ALGORITHM

In order to extract useful information from document images, especially those poor quality ones with non-uniform illumination, low contrast, undesired shadows and random noise, we have devised a new and reliable local thresholding method by formulating the binarization decision in terms of contrast instead of gray values.

The proposed method first computes the local mean, m, minimum, M, and standard deviation, s, by shifting a primary local window (as shown in Fig. 1), whose size is large enough to cover 1-2 characters, across the input image. Then, in order to compensate the effect of illumination, the dynamic of standard deviation R_S is calculated for a larger local window, secondary local window, instead of the whole image as in Fig. 1. According to our finding, the appropriate size of this local window depends on the degree of illumination variation and the setup of image capturing equipments. However, through a few iterations of testing and retuning, the optimum window size can be easily obtained. Finally, the threshold value is calculated as

$$T = (1-\alpha_1)\cdot m + \alpha_2\cdot (\frac{s}{R_S})\cdot (m-M) + \alpha_3\cdot M,$$

$$\alpha_2 = k_1\cdot (\frac{s}{R_S})^\gamma \text{ and } \alpha_3 = k_2\cdot (\frac{s}{R_S})^\gamma,$$

where α_1, γ, k_1 and k_2 are positive constants.

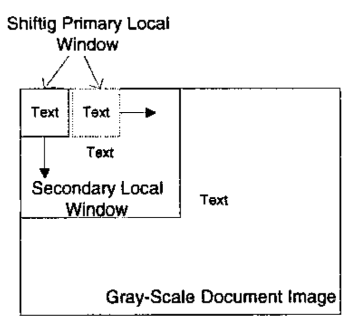

Fig. 1. Local windows definition

The proposed equation consists of three elements, and three corresponding coefficients, α_1, α_2 and α_3, are applied to the equation so that more flexible and versatile control on the weight of different equation elements is allowed. Coefficients α_2 and α_3 are adaptively formulated based on the normalized local standard deviation, s/R_S. This idea comes form the fact that windows that contain text have a larger standard deviation, and thus a larger s/R_S value, than those windows contain no texts. Therefore, including the s/R_S ratio in coefficients α_2 and α_3 enables the estimated threshold to separate texts from background more adaptively without much prior knowledge of the input image. Moreover, through experiments we found that, to a certain extent, the power of s/R_S ratio affects the binarization performance as well. As a result, an additional exponential parameter, γ, is introduced to the coefficient formulae.

In Fig. 2, the respective contributions of the three equation elements are further illustrated with an example scanline. The bar on top of the figure shows the enlarged contents of the scanline, where dark black portions represent pixels of text and gray portions correspond to background pixels, respectively. As can be seen, the first term in equation is to shift down the mean value by a fix amount controlled by α_1. Then, the second and third terms aim to adjust the threshold T according to the contrast of the image and the normalized standard deviation.

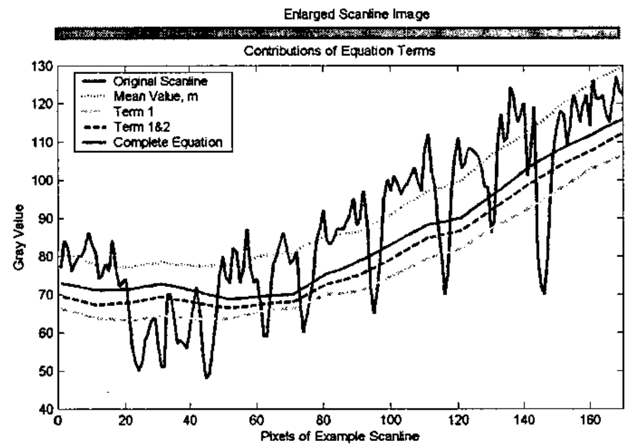

Fig. 2. Illustration of the function of each term in the binarization equation.

In our implementation of the proposed algorithm (see Fig. 3), the captured document picture was first transformed to a gray-scale image. A median filter was also applied to remove the random noise, as much as possible, of the input image. To minimize the computation load, threshold values were only calculated for the centers of every primary local window and the values for the rest of the pixels were obtained by bilinear interpolation. After the threshold surface was established, input image was binarized into black text and white background, and the binarization results were then fed to an OCR software for testing. Based on our empirical study, setting the values of α_1, k_1 and k_2 in the ranges of 0.1-0.2, 0.01-0.05 and 0.15-0.25 and γ to 2, respectively, can generally produce good binarization results.

4. EXPERIMENTAL RESULTS

We have applied the proposed method to a group of text images with various sizes and qualities, and promising results

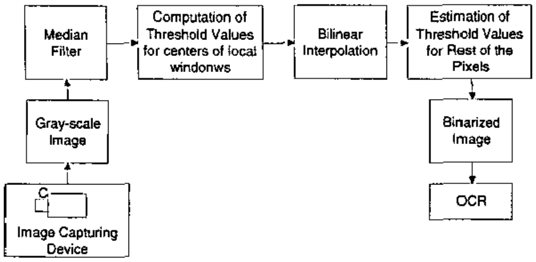

Fig. 3. Implementation flow chart.

were achieved. Fig. 4 shows the different thresholds computed along the same scanline in Fig. 2 based on the previous three and our proposed techniques. It is evident that compared with Niblack's and Christian's methods, the threshold obtained by the proposed algorithm is more closely adapted to the illumination variation of the input image. As presented in Fig. 4, our method successfully classified the right portion of the scanline (from pixel 140 onwards) as background, while the other two mistake part of it as texts. This testifies the effectiveness of the proposed method in filtering background noise. For Sauvola's method, due to the restriction from its additional hypothesis, its threshold line turned out to be very near the zero gray level line, which actually failed to binarize the example scanline.

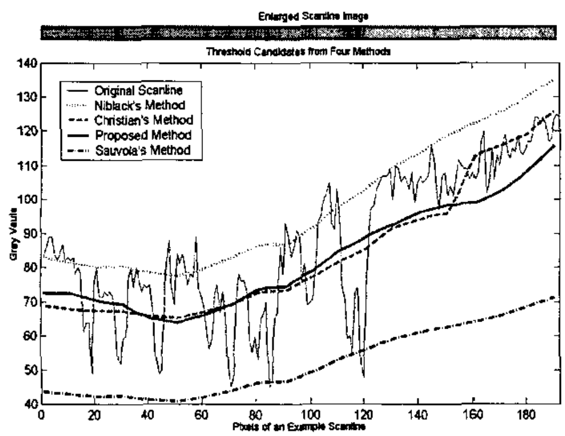

Fig. 4. Local binarization thresholds determined by different methods under comparison along a scanline of a test document image.

In Fig. 5, the performance of the proposed method is compared with the other methods. The two test images were captured by a low-resolution handheld camera. The figure shows that Niblack's method segments the characters well, but suffers from noise in regions without text. Sauvola's approach is found to be more suitable for good quality images, because badly scanned images generally do not correspond to the additional hypothesis. Both Christian's and the proposed algorithms produce relatively better results. Nevertheless, Christian's algorithm suffers from the effect of illumination, which causes background noise, broken lines and holes in its output image. Our solution is supreme in terms of the robustness to illumination variation, the noise suppression in regions without text and the clearance of the characters. Since the performance of Christian's method and the proposed method was found to be competitively close, more comparisons, as shown in Fig. 6, were carried out to justify the superiority of our method.

Fig. 7 presents an example of our OCR performance with Transym Optical Character Recognizer (TOCR) [9]. Based on a variety of tests, we achieve an average OCR character recognition rate of 90.8%, while Christian can only reach 85.4%. The OCR also confirms the better results of our binarization technique.

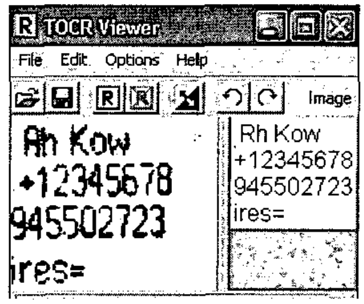

Fig. 7. An example of OCR result.

5. CONCLUSION

In this paper, we have presented a new local thresholding method, for binarizing document images based on a criterion of maximizing local contrast. We have shown that the improved method, which is formulated in terms of contrast instead of gray values, is robust against shadows, stochastic noise and illumination variations. By testing our binarization system against various challenging document images, we have also demonstrated the effectiveness of our proposed algorithm.

6. REFERENCES

[1] M. Pilu and S. Polland, "A Light-weight Text Image Processing Method for Handheld Embedded Cameras," *Hewlett-Packard Laboratories, Bristol, England.* March 2002

[2] H.-S. Don, "A Noise Attribute Thresholding Method for Document Image Binarization." *IEEE 0-8186-7128-9/95*, 1995.

[3] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," *IEEE Transactions on Systems,* Man and Cybernetics, 1979.

[4] O. D. Trier and T. Taxt, "Evaluation of Binarization Methods for Document Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* VOL. 17, NO. 3. March 1995.

[5] S. D. Yanowitz and A. M. Bruckstein, "A New Method for Image Segmentation. Computer Vision," *Graphics and Image Processing,* 1989.

[6] W. Niblack, "An Introduction to Digital Image Processing," pp 115-116. *Englewood Cliffs, N.J.: Prentice Hall,* 1986.

[7] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization," *The Journal of the Pattern Recognition Society,* PR 33 (2000) 225-236. 1999.

[8] C. Wolf and J.-M. Jolion, "Extraction and Recognition of Artifical Text in Multimedia Documents," *RFV RR-2002.01.* 2002.

[9] Transym Optical Character Recognizer (TOCR). [Available online] http://www.sorcery.demon.co.uk

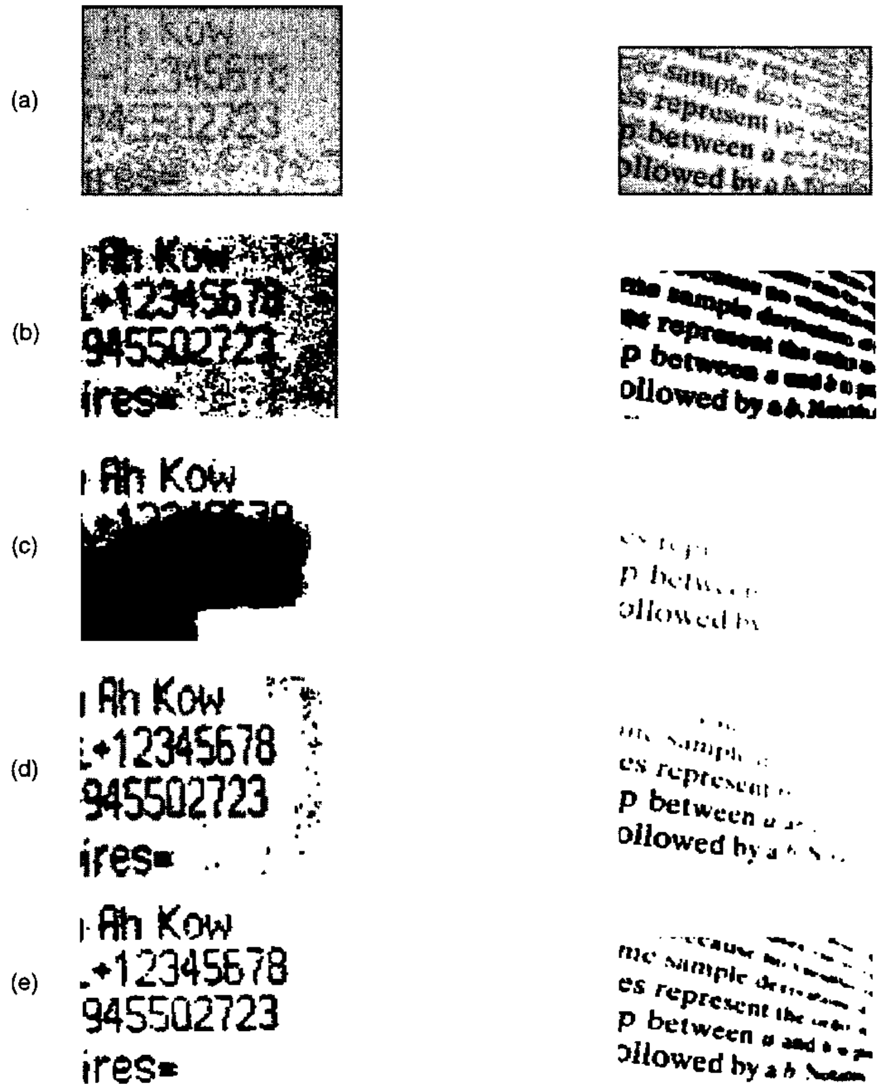

Fig. 5. Binarization results. (a) Original document image. (b) Niblack's method. (c) Sauvola et al,'s method. (d) Christian et al.'s method. (e) Our Proposed method.

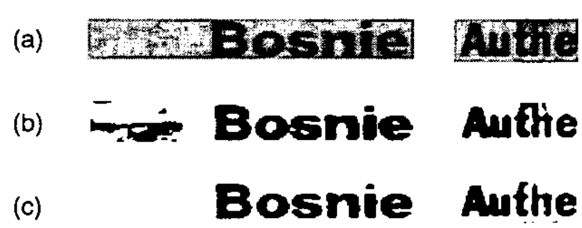

Fig. 6. More Binarization results. (a) Original image. (b) Christian's method. (c) Our proposed method.