Binarization of Historical Document Images Using the Local Maximum and Minimum

Bolan Su	Shijian Lu	Chew Lim Tan
Department of Computer Science	Department of Computer Vision and Image Understanding	Department of Computer Science
School of Computing	Institute for Infocomm Research	School of Computing
National University of Singapore	1 Fusionopolis Way	National University of Singapore
Computing 1, 13 Computing Drive	#21-01 Connexis	Computing 1, 13 Computing Drive
Singapore 117417	Singapore 138632	Singapore 117417
subolan@comp.nus.edu.sg	slu@i2r.a-star.edu.sg	tancl@comp.nus.edu.sg

ABSTRACT

This paper presents a new document image binarization technique that segments the text from badly degraded historical document images. The proposed technique makes use of the image contrast that is defined by the local image maximum and minimum. Compared with the image gradient, the image contrast evaluated by the local maximum and minimum has a nice property that it is more tolerant to the uneven illumination and other types of document degradation such as smear. Given a historical document image, the proposed technique first constructs a contrast image and then detects the high contrast image pixels which usually lie around the text stroke boundary. The document text is then segmented by using local thresholds that are estimated from the detected high contrast pixels within a local neighborhood window. The proposed technique has been tested over the dataset that is used in the recent Document Image Binarization Contest (DIBCO) 2009. Experiments show its superior performance.

Categories and Subject Descriptors

I.7.5 [**DOCUMENT AND TEXT PROCESSING**]: Document Capture—*Document analysis*; I.4.6 [**IMAGE PROCESSING AND COMPUTER VISION**]: Segmentation—*Pixel classification*

General Terms

ALGORITHMS, EXPERIMENTATION

Keywords

Document Image Analysis, Document Image Binarization, Image Contrast, Image Pixel Classification

1. INTRODUCTION

Document image binarization aims to divide a document image into two classes, namely, the foreground text and the document
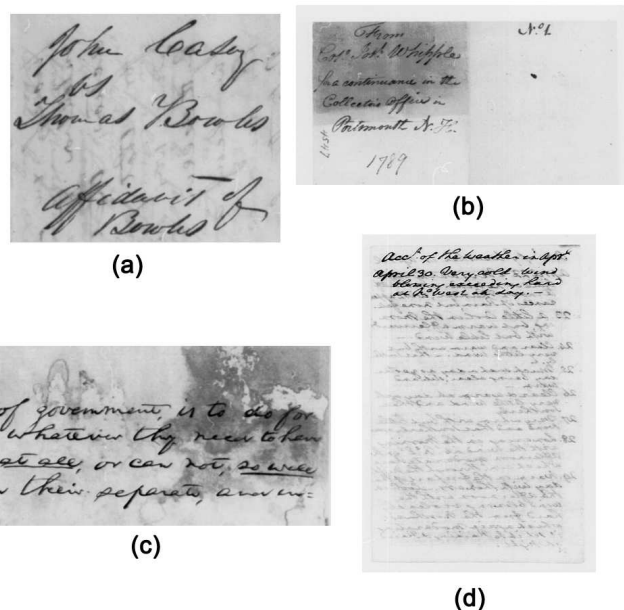
Figure 1: Four example historical document images that suffer from different types of document degradation

background. It is usually performed in the document preprocessing stage and is very important for ensuing document image processing tasks such as optical character recognition (OCR). As more and more paper documents are digitized, fast and accurate document image binarization is becoming increasingly important.

Though document image binarization has been studied for many years, the thresholding of historical document images is still an unsolved problem due to the high variation within both the document background and foreground as illustrated in Figure 1. The latest Document Image Binarization Contest (DIBCO) [7] held under the framework of the International Conference on Document Analysis and Recognition (ICDAR) 2009 also shows recent efforts on this issue. Compared with most machine-printed text documents, the historical documents often have a more variable background as illustrated in Figure 1(b) and 1(c) that tends to introduce the thresh-
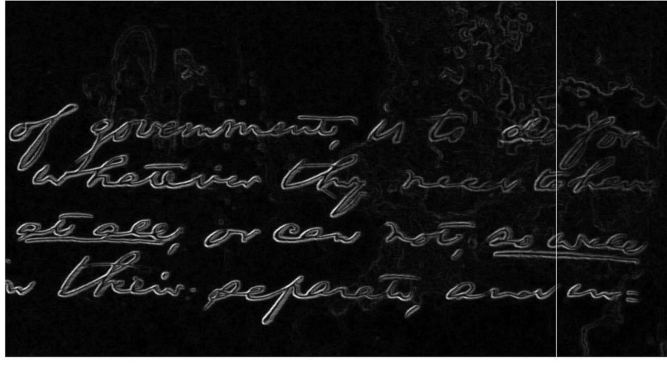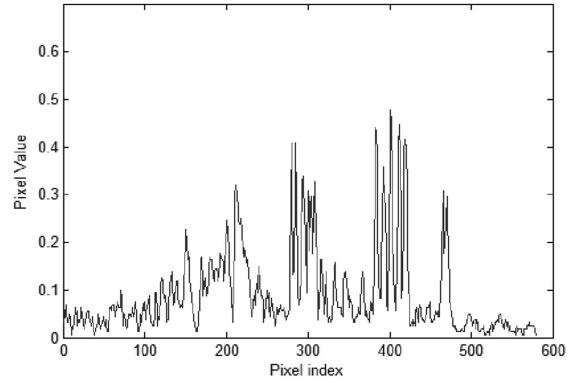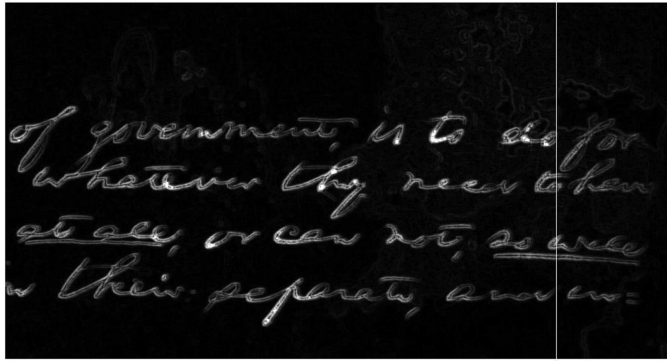
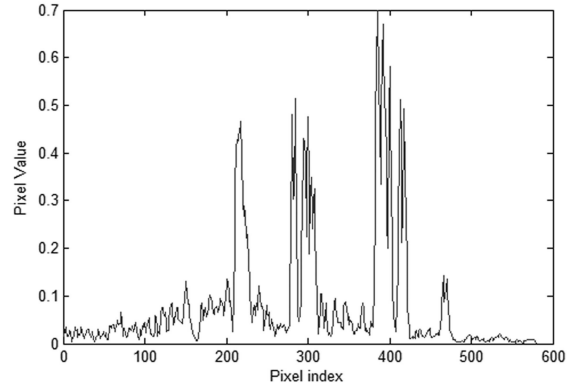**Figure 2: The gradient and contrast of the historical document in Figure 1(c): (a) The traditional image gradient that is obtained using Canny's edge detector [3]; (b) The image contrast that is obtained by using the local maximum and minimum [19];(c) One column of the image gradient in Figure 2(a) (shown as a vertical white line);(d) The same column of the contrast image in Figure 2(b).**

olding error. Another type of variation that often comes with the historical document background is the bleeding-through as shown in Figure 1(d) where the ink of one paper side seeps through to the other. In addition, the foreground handwritten text within the historical documents often shows a certain amount of variation in terms of the stroke width, stroke brightness, and stroke connection as illustrated in Figure 1(a).

Many works [11, 18] have been reported to deal with the high variation within historical document images. As many historical documents do not have a clear bimodal pattern, global thresholding [2, 10, 15] is usually not a suitable approach for the historical document binarization. Adaptive thresholding [1, 6, 9, 16, 17, 14, 20], which estimates a local threshold for each document image pixel, is usually a better approach to handle the high variation associated with historical document images. For example, the early window-based adaptive thresholding techniques [17, 14] estimate the local threshold by using the mean and the standard variation of image pixels within a local neighborhood window. The main drawback of this window-based thresholding approach is that the thresholding performance depends heavily on the window size and hence the character stroke width. Other approaches have also been reported to binarize historical document images through background sub-

traction [13, 8], texture analysis [12], decomposition method [4], and cross section sequence graph analysis [5], and so on. These approaches combine different types of image information and domain knowledge and are often complex and time consuming.

This paper presents a document thresholding technique that is able to binarize degraded historical document images efficiently. One distinctive characteristic of the proposed technique is that it makes use of the image contrast that is evaluated by using the local maximum and minimum. Compared with the image gradient, such image contrast is more capable of detecting the high contrast image pixels (lying around the text stroke boundary) from historical documents that often suffer from different types of document degradation. And compared with the Lu&Tan's method which was modified from [13] and used in DIBCO contest [7], the proposed method is better while handling document images with complex background variation. Given a historical document image, the proposed technique first determines a contrast image based on the local maximum and minimum. The high contrast image pixels around the text stroke boundary are then detected through the global thresholding of the determined contrast image. Lastly, the historical document image is binarized based on the local thresholds that are estimated from the detected high contrast image pixels. Compare with

previous method based on image contrast [1], the proposed method uses the image contrast to identify the text stroke boundary, which can be used to produce more accurate binarization results. Experiments over the dataset used in the recent Document Image Binarization Contest (DIBCO) 2009 show the superior performance of the proposed technique.

2. PROPOSED METHOD

The section describes the proposed historical document image thresholding technique. In particular, we divide this section into three subsections, which deal with the contrast image construction, the high contrast pixel detection, and the local threshold estimation, respectively.

2.1 Contrast Image Construction

The image gradient has been widely used in the literature for edge detection [3]. However, the image gradient is often obtained by the absolute image difference within a local neighborhood window, which does not incorporate the image intensity itself and is so sensitive to the image contrast/brightness variation. Take an unevenly illuminated historical document image as an example, the gradient of an image pixel (around the text stroke boundary) within bright document regions may be much higher than that within dark document regions. To detect the high contrast image pixels around the text stroke boundary properly, the image gradient needs to be normalized to compensate for the effect of the image contrast/brightness variation. At the same time, the normalization suppresses the variation within the document background as well.

In the proposed technique, we suppress the background variation by using an image contrast that is calculated based on the local image maximum and minimum [19] as follows:

$$D(x,y) = \frac{f_{max}(x,y) - f_{min}(x,y)}{f_{max}(x,y) + f_{min}(x,y) + \epsilon} \qquad (1)$$

where $f_{max}(x,y)$ and $f_{min}(x,y)$ refer to the maximum and the minimum image intensities within a local neighborhood window. In the implemented system, the local neighborhood window is a 3×3 square window. The term ϵ is a positive but infinitely small number, which is added in case the local maximum is equal to 0.

The image contrast in Equation 1 lowers the image background and brightness variation properly. In particular, the numerator (i.e. the difference between the local maximum and the local minimum) captures the local image difference that is similar to the traditional image gradient [3]. The denominator acts as a normalization factor that lowers the effect of the image contrast and brightness variation. For image pixels within bright regions around the text stroke boundary, the denominator is large, which neutralizes the large numerator and accordingly results in a relatively low image contrast. But for image pixels within dark regions around the text stroke boundary, the denominator is small, which compensates the small numerator and accordingly results in a relatively high image contrast. As a result, the contrasts of image pixels (lying around the text stroke boundary) within both bright and dark document regions converge close to each other and this facilitates the detection of high contrast image pixels lying around the text stroke boundary (to be described in the next subsection).

At the same time, the image contrast in Equation 1 suppresses the variation within the document background properly. For document background pixels, the local minimum is usually much brighter

than that of the image pixels lying around the text stroke boundary. As a result, the contrast of the document background pixels will be suppressed due to the high denominator. With the same reason, the image pixels with similar image gradient lying around the text stroke boundary in dark regions will have a much higher image contrast. This enhances the discrimination between the image pixels around the text stroke boundary and those within the document background region with high variation because of the document degradation shown in Figure 1.
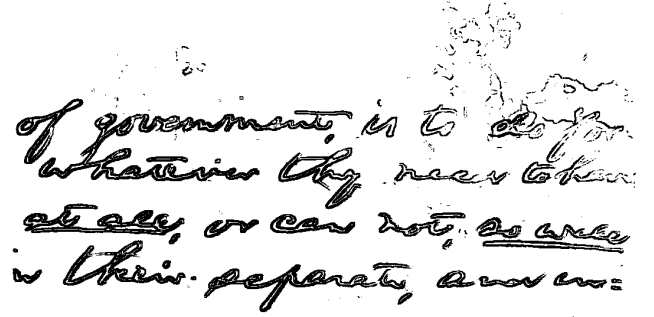

(a)
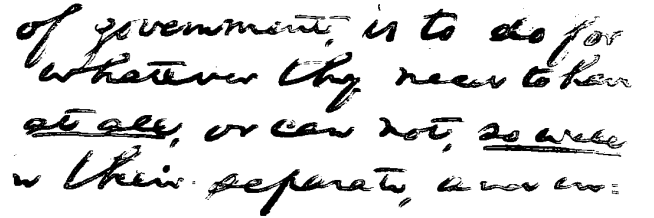
(b)

Figure 3: High contrast pixel detection: (a) Global thresholding of the gradient image in Figure 2(a) by using Otsu's method; (b) Global thresholding of the contrast image in Figure 2(b) by using Otsu's method.

Figure 2 illustrates the difference between the image gradient and the image contrast defined in Equation 1. In particular, Figure 2(a) and 2(b) show the gradient image and the contrast image that are constructed based on the document image in Figure 1(c), respectively. As Figure 2(a) shows that the image gradient around text stroke boundary varies visibly from the bright document regions to the dark document regions. However,, as shown in Figure 2(b), the image contrast around text stroke boundary varies little from the bright document regions to the dark document regions. At the same time, the discrimination between the image contrast around the text stroke boundary and that around the document background is much stronger compared with the discrimination of the the image gradient around the text stroke boundary and that around the document background. These two points can be further illustrated in Figure 2(c) and 2(d) where the same column of the gradient image in Figure 2(a) and the contrast image in Figure 2(b) is plotted, respectively.
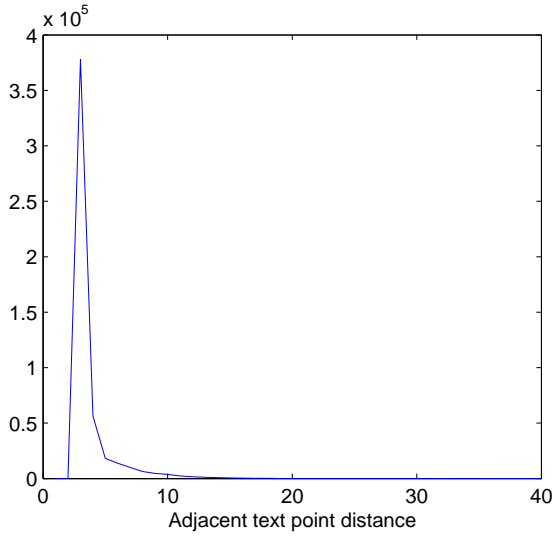
Figure 4: **The histogram of the high contrast image in Figure 2(b) that records the frequency of the distance between adjacent pixels with the peak contrast value.**

## 2.2 High Contrast Pixel Detection

The purpose of the contrast image construction is to detect the desired high contrast image pixels lying around the text stroke boundary. As described in the last subsection, the constructed contrast image has a clear bimodal pattern where the image contrast around the text stroke boundary varies within a small range but is obviously much larger compared with the image contrast within the document background. We therefore detect the desired high contrast image pixels (lying around the text stroke boundary) by using Otsu's global thresholding method.

Figure 3(a) and (b) show the binarization results of the gradient image in Figure 2(a) and the contrast image in Figure 2(b), respectively, by using Otsu's global thresholding method. As Figure 3(b) shows, most of the high contrast image pixels detected through the binarization of the contrast image correspond exactly to the desired image pixels around the text stroke boundary. On the other hand, the binarization of the gradient image in Figure 3(a) introduces a certain amount of undesired pixels that usually lie within the document background.

2.3 Historical Document Thresholding

The text pixels can be classified from the document background pixels once the high contrast image pixels around the text stroke boundary are detected properly. The document thresholding from the detected high contrast image pixels is based on two observations. First, the text pixels should be close to the detected high contrast image pixels because most detected high contrast image pixels lie around the text stroke boundary. Second, the intensity of most text pixels should be close or lower than the average intensity of the detected high contrast image pixels within a local neighborhood window. This can be similarly explained by the fact that most detected high contrast image pixels lie around the text stroke boundary.

For each document image pixel, the number of the detected high
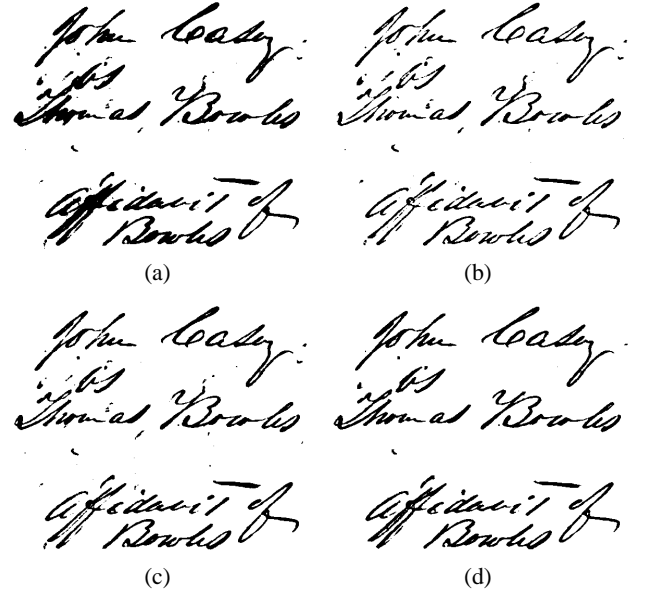
Figure 5: **Binarization results of the degraded document image in Figure 1(a) by using Otsu's global thresholding method in (a), Niblack's adaptive thresholding method in (b), Sauvola's adaptive thresholding method in (c), and our proposed method in (d)**

contrast image pixels is first determined within a local neighborhood window. The document image pixel will be considered a text pixel candidate if the number of high contrast image pixels within the neighborhood window is larger than a threshold. The document image pixel can thus be classified based on its intensity relative to that of its neighboring high contrast image pixels as follows:

$$R(x,y) = \begin{cases} 1 & \begin{aligned} N_e \geq N_{min} & \quad \&\& \\ I(x,y) \leq E_{mean} & + E_{std}/2 \end{aligned} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where E_{mean} and E_{std} are the mean and the standard deviation of the image intensity of the detected high contrast image pixels (within the original document image) within the neighborhood window that can be evaluated as follows:

$$E_{mean} = \frac{\sum\limits_{neighbor} I(x,y) * (1 - E(x,y))}{N_e}$$

$$E_{std} = \sqrt{\frac{\sum\limits_{neighbor} ((I(x,y) - E_{mean}) * (1 - E(x,y)))^2}{2}}$$

where I refers to the input document image and (x,y) denotes the position of the document image pixel under study. E refers to the binary high contrast pixel image where $E(x,y)$ is equal to 0 if the document image pixel is detected as a high contrast pixel. N_e refers to the number of high contrast image pixels that lie within the local neighborhood window. So if N_e is larger than N_{min} and $I(i,j)$ is smaller than $E_{mean} + E_{std}/2$, $R(i,j)$ is set at 1. Otherwise, $R(i,j)$ is set at 0.

(a) (b)

(c) (d)
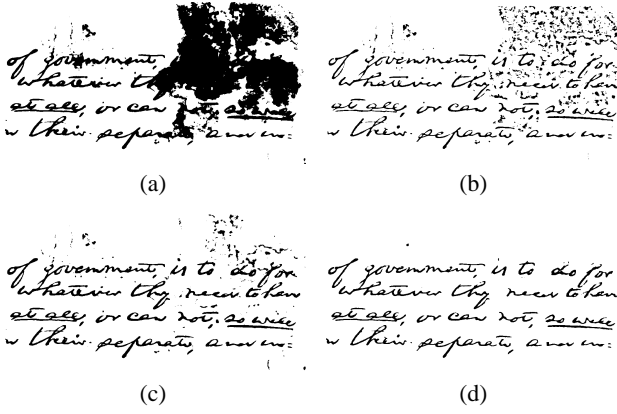
Figure 6: Binarization results of the degraded document image in Figure 1(b) by using Otsu's global thresholding method in (a), Niblack's adaptive thresholding method in (b), Sauvola's adaptive thresholding method in (c), and our proposed method in (d)

(a) (b)

(c) (d)

Figure 7: Binarization results of the degraded document image in Figure 1(c) by using Otsu's global thresholding method in (a), Niblack's adaptive thresholding method in (b), Sauvola's adaptive thresholding method in (c), and our proposed method in (d)

There are two parameters that need to be set properly, namely, the size of the neighborhood window and the minimum number of the high contrast image pixels N_{min} within the neighborhood window. These two parameters are both correlated to the width of the text strokes within the document image under study. In particular, the size of the neighborhood window should not be smaller than the text stroke width. Otherwise the text pixels within the interior of the text strokes may not be classified correctly because the local neighborhood window may not enclose enough high contrast image pixels. At the same time, the minimum number of the high contrast image pixels (within the neighborhood window) should be around the size of the local neighborhood window based on the double-edge structure of the character strokes.

We estimate the text stroke width from the constructed contrast image shown in Figure 2(b). In particular, we first scan the contrast image horizontally row by row. The image pixels with the peak contrast value is then located within a one-dimensional window, most of which usually correspond to the text stroke edge pixels within the text regions. Considering the fact that the text strokes should be larger than one pixel width, we set the size of the one-dimensional window at 3. Once the image pixels with the peak contrast value are located, a histogram is constructed that records the frequency of the distance between two adjacent peak pixels. For text documents with a certain amount of text, the most frequent distance between the adjacent peak pixels usually gives a rough estimation of the text stroke width. Figure 4 shows the distance histogram that is determined from the contrast image in Figure 2(b). As Figure 4 shows, the most frequent distance can be obviously located from the distance histogram.

3. EXPERIMENT AND DISCUSSIONS

The proposed method has been tested over the handwritten images of the dataset that is used in the recent Document Image Binarization Contest (DIBCO), 2009 [1]. The dataset is composed of a number of representative document images that suffer from different types of document degradation. We compare our method with three well known binarization methods including Otsu's global thresh-

[1]http://users.iit.demokritos.gr/~bgat/DIBCO2009/benchmark
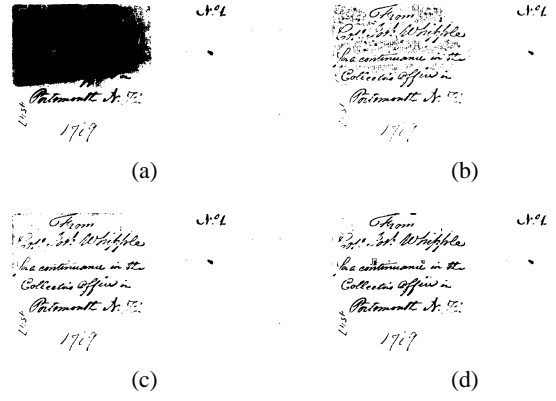
olding method [15] and Niblack's and Sauvola 's adaptive thresholding methods [14, 17]. In particular, the parameters of Niblack's and Sauvola's methods including the window size, the weight of the local standard variation, and the weight of the local dynamic range of standard variation are all set according to the recommendations within the reported papers [14, 17]. In addition, we also compare the proposed method with our earlier method that participated in the DIBCO contest and achieves the top performance among 43 algorithms that are submitted from 35 international research groups.

The evaluation measures are adapted from DIBCO's report [7] including F-Measure, Peak Signal to Noise Ratio (PSNR), Negative Rate Metric (NRM), and Misclassification Penalty Metric (MPM). In particular, the F-Measure is defined as follows:

$$FM = \frac{2 * RC * PR}{RC + PR} \qquad (3)$$

$$RC = \frac{CTP}{CTP + CFN}$$
$$PR = \frac{CTP}{CTP + CFP}$$

where RC and PR refer the recall and the precision of the method in Equation 3. CTP, CFP, and CFN denote the numbers of true positive pixels, false positive pixels, and false negative pixels, respectively. This measure evaluates how well an algorithm can retrieve the desire pixels.

The measure PSNR is defined as follows:

$$PSNR = 10\log(\frac{C^2}{MSE}) \qquad (4)$$

$$MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y) - I'(x,y))^2}{MN}$$

where C is a constant that denotes the difference between foreground and background. This constant can be set to 1. The $PSNR$ measures how close the resultant image to the ground truth image.

Table 1: Evaluation results of Lu and Tan, Otsu, Niblack, Sauvola's methods and proposed method over the handwritten images of the DIBCO dataset

Method	F-Measure(%)	PSNR	NRM($\times 10^{-2}$)	MPM($\times 10^{-3}$)
Lu and Tan	88.53	19.42	5.11	0.32
Otsu	66.13	13.98	7.52	23.5
Niblack	77.34	16.46	11.8	5.9
Sauvola	80.44	16.98	7.84	3.4
Proposed Method	89.93	19.94	6.69	0.3

The measure NRM is defined as follows:

$$NRM = \frac{NR_{FN} + NR_{FP}}{2} \tag{5}$$

$$NR_{FN} = \frac{N_{FN}}{N_{FN} + N_{TP}}$$

$$NR_{FP} = \frac{N_{F}P}{N_{FP} + N_{TN}}$$

where N_{TP}, N_{FP}, N_{TN}, and N_{FN} denote the number of true positives, false positives, true negatives, and false negatives respectively. This metric measures pixel mismatch rate between the ground truth image and resultant image.

The measure MPM is defined as follows:

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \tag{6}$$

$$MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^i}{D}$$

$$MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^j}{D}$$

where d_{FN}^i and d_{FP}^j denote the distance of the ith false negative and the jth false positive pixel from the contour of the ground truth segmentation. The normalization factor D is the sum over all the pixel-to-contour distances of the ground truth object. This metric measures how well the resultant image represents the contour of ground truth image.

Experiment results are shown in Table 1. Compared with the other four methods, our proposed method performs better in term of the F-Measure, $PSNR$, and MPM. This means that the proposed method produces a higher precision and preserves the text stroke contour better. At the same time, the performance of the proposed method is close to that of our earlier method that achieved the top performance in the DIBCO 2009. Figure 5, 6, 7, and 8 further show four document binarization examples. As shown in the four figures, our proposed method extracts the text properly from document images that suffer from different types of document degradation. On the other hand, the other three methods often produce a certain amount of noise due to the variation within the document background.

Our proposed method can suppress the noise efficiently because it suppress the contrast of the document background through the normalization as described in Section 2.1. As a result, few high contrast image pixels are detected from the document background area and most document background pixels therefore will be excluded from the further classification (based on the intensity) as defined in Equation 2. As a comparison, Otsu's global thresholding method simply classifies dark background pixels as the text pixels

improperly. Niblack's and Sauvola's method cannot classify some dark background pixels properly either because the image variation there is often as large as that around the text stroke boundary.

(a)   (b)

(c)   (d)
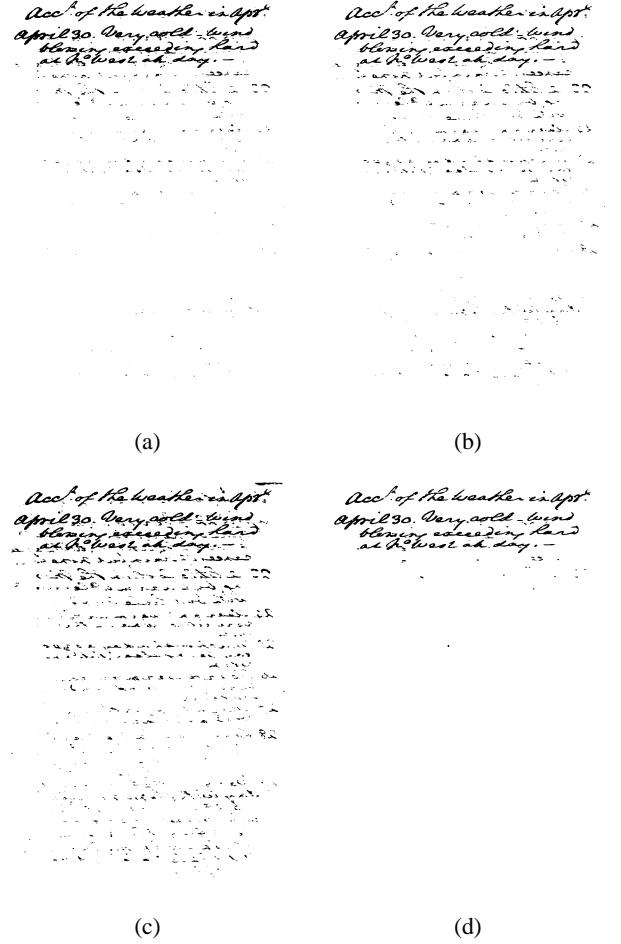
Figure 8: Binarization results of the degraded document image in Figure 1(d) by using Otsu's global thresholding method in (a), Niblack's adaptive thresholding method in (b), Sauvola's adaptive thresholding method in (c), and our proposed method in (d)

The proposed document binarization method has a few limitations. First, the proposed method can deal with the ink-bleeding as illustrated in Figure 2(d) when the back-side text strokes are much weaker compared with the front-side text. But when the back-side text strokes are as dark as or even darker than the front-side text

strokes, the proposed method cannot classify the two types of character strokes correctly. In addition, the proposed method depends heavily on the high contrast image pixels. As a result, it may introduce error if the background of the degraded document images contains a certain amount of pixels that are dense and at the same time have a fairly high image contrast. We will study these two issues in our future works.

4. CONCLUSION

This paper presents a simple but efficient historical document image binarization technique that is tolerant to different types of document degradation such as uneven illumination and document smear. The proposed technique makes use of the image contrast that is evaluated based on the local maximum and minimum. Given a document image, it first constructs a contrast image and then extracts the high contrast image pixels by using Otsu's global thresholding method. After that, the text pixels are classified based on the local threshold that is estimated from the detected high contrast image pixels. The proposed method has been tested on the dataset that is used in the recent DIBCO contest. Experiments show that the proposed method outperforms most reported document binarization methods in term of the F-measure, PSNR, NRM, and MPM.

5. REFERENCES

[1] J. Bernsen. Dynamic thresholding of gray-level images. *International Conference on Pattern Recognition*, pages 1251–1255, October 1986.

[2] A. Brink. Thresholding of digital images using two-dimensional entropies. *Pattern Recognition*, 25(8):803–808, 1992.

[3] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, January 1986.

[4] Y. Chen and G. Leedham. Decompose algorithm for thresholding degraded historical document images. *IEEE Proceedings on Vision, Image and Signal Processing*, 152(6):702–714, December 2005.

[5] A. Dawoud. Iterative cross section sequence graph for handwritten character segmentation. *IEEE Transactions on Image Processing*, 16(8):2150–2154, August 2007.

[6] L. Eikvil, T. Taxt, and K. Moen. A fast adaptive method for binarization of document images. *International Conference on Document Analysis and Recognition*, pages 435–443, September 1991.

[7] B. Gatos, K. Ntirogiannis, and I. Pratikakis. ICDAR 2009 document image binarization contest (DIBCO 2009). In *International Conference on Document Analysis and Recognition*, pages 1375–1382, July 2009.

[8] B. Gatos, I. Pratikakis, and S. Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39(3):317–327, 2006.

[9] I.-K. Kim, D.-W. Jung, and R.-H. Park. Document image binarization based on topographic analysis using a water flow model. *Pattern Recognition*, 35(1):265–277, 2002.

[10] J. Kittler and J. Illingworth. On threshold selection using clustering criteria. *IEEE transactions on systems, man, and cybernetics*, 15:652–655, 1985.

[11] G. Leedham, C. Yan, K. Takru, J. H. N. Tan, and L. Mian. Comparison of some thresholding algorithms for text/background segmentation in difficult document images. *International Conference on Document Analysis and Recognition*, pages 859–864, 2003.

[12] Y. Liu and S. Srihari. Document image binarization based on texture features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):540–544, May 1997.

[13] S. J. Lu and C. L. Tan. Binarization of badly illuminated document images through shading estimation and compensation. *International Conference on Document Analysis and Recognition*, pages 312–316, 2007.

[14] W. Niblack. *An Introduction to Digital Image Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[15] N. Otsu. A threshold selection method from gray level histogram. *IEEE Transactions on System, Man, Cybernetics*, 19(1):62–66, January 1978.

[16] J. Parker, C. Jennings, and A. Salkauskas. Thresholding using an illumination model. *International Conference on Document Analysis and Recognition*, pages 270–273, October 1993.

[17] J. Sauvola and M. Pietikainen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.

[18] O. Trier and T. Taxt. Evaluation of binarization methods for document images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):312–315, March 1995.

[19] M. van Herk. A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels. *Pattern Recognition Letters*, 13(7):517–521, 1992.

[20] J.-D. Yang, Y.-S. Chen, and W.-H. Hsu. Adaptive thresholding algorithm and its hardware implementation. *Pattern Recognition Letters*, 15(2):141–150, 1994.