

Stroke Width-Based Contrast Feature for Document Image Binarization

Le Thi Khue Van* and Guesang Lee*

Abstract—Automatic segmentation of foreground text from the background in degraded document images is very much essential for the smooth reading of the document content and recognition tasks by machine. In this paper, we present a novel approach to the binarization of degraded document images. The proposed method uses a new local contrast feature extracted based on the stroke width of text. First, a pre-processing method is carried out for noise removal. Text boundary detection is then performed on the image constructed from the contrast feature. Then local estimation follows to extract text from the background. Finally, a refinement procedure is applied to the binarized image as a post-processing step to improve the quality of the final results. Experiments and comparisons of extracting text from degraded handwriting and machine-printed document image against some well-known binarization algorithms demonstrate the effectiveness of the proposed method.

Keywords—Degraded Document Image, Binarization, Stroke Width, Contrast Feature, Text Boundary

1. INTRODUCTION

Document image analysis plays an important role in image processing and pattern recognition. In the entire process, binarization is an important initial step of almost all document image analysis systems, and its performance is crucial for success in the subsequent character recognition and segmentation.

Although many research studies have been carried out in the field of document image binarization, and various thresholding algorithms have been developed, the binarization of document images with poor and variable contrast, shadow, smudge, and variable foreground and background intensities are still challenging problems. The thresholding of degraded document images remains unsolved. This can be attributed to the difficulty in modeling different types of document degradation existing within many document images, such as uneven illumination, image contrast variation, bleeding-through, and smear.

The existing algorithms can be categorized into global method [1,2], local method [3,4,5], and hybrid algorithm [6,7]. Global binarization algorithms calculate a single threshold for a whole text image. They have good performance when there is good separation between foreground and background intensities; in cases of poor contrast and variable background and foreground intensities, however, these methods fail to achieve proper binarization. Local binarization

※ This study was financially supported by Chonnam National University, 2012.
Manuscript received February 26, 2013; first revision May 2; accepted July 21, 2013.

Corresponding Author: Guesang Lee (gslee@chonnam.ac.kr)

* Dept. of Electronics and Computer Engineering, Chonnam National University, Kwangju, 500-757, Korea
(mmlkvan@gmail.com, gslee@chonnam.ac.kr).

methods use local area information to calculate the threshold value for each pixel or sub-image. These methods can obtain satisfactory results in text extraction from degraded document images. Note, however, that local binarization methods generally have some predetermined parameters that decrease their efficiency in different conditions. On the other hand, hybrid methods are a combination of both global and local binarization algorithms, and they have also solved some cases of degraded document images. An evaluation of document binarization algorithms [8] has shown that there is no method that performs well in all types of degraded document image. Therefore, the binarization of degraded document images remains a challenge.

In this paper, we propose a method for the binarization of document images with some degradation such as poor contrast, badly illuminated, seeping, smear, etc., as shown in Fig. 1. The proposed method presents an extraction of a new local contrast feature based on the structural characteristics of the text. It performs an increase in the discrimination of the text and non-text areas. The text boundary is then detected in the image constructed from the extracted contrast feature. Afterward, local estimation is applied to get the binarization result. Finally, a post-processing step is performed to improve the quality of the final results.

The rest of this paper is organized as follows: Section 2 describes briefly some previous works on document image binarization methods; Section 3 is devoted to the proposed approach; Experimental results in comparison with some other methods are discussed in Section 4; Finally, Section 5 presents the conclusions.

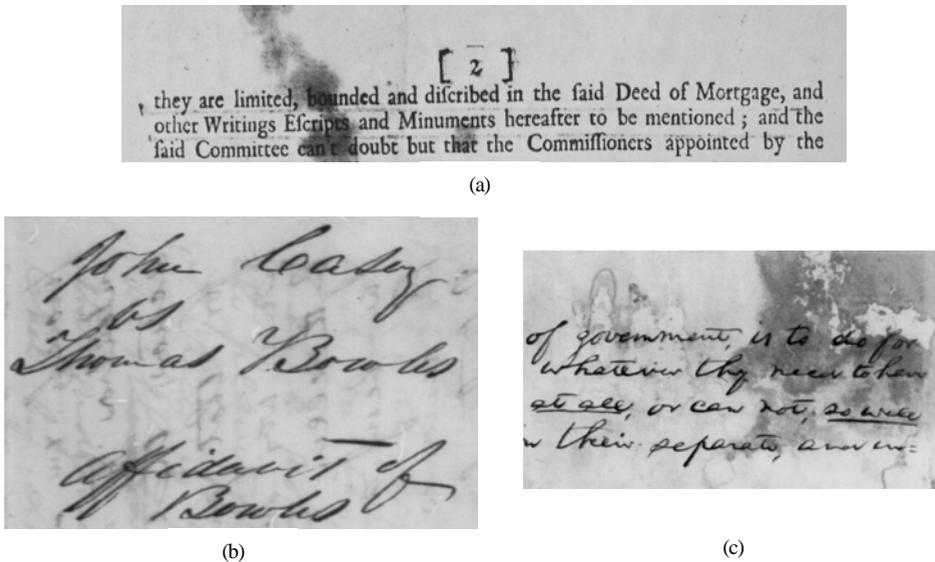


Fig. 1. Document image with different types of degradation: (a) smudge, (b) bleeding-through, and (c) image contrast variation and smear

2. PREVIOUS WORKS

Text information in images is essential and critical in many applications. Therefore, a large number of binarization methods have been proposed to extract text from background and reported in literature. For document images of good quality, global techniques such as Otsu's method [1] and Kitter's method [2] showed its efficiency in generating good text against a clean background. They are fast, but inappropriate for complex text images. To solve those cases, local binarization methods were developed. Some well-known local binarization algorithms are those of Bernsen [3], Niblack [4], and Sauvola [5], which calculate separate threshold for each pixel by shifting a sliding window across the image.

Bernsen's method [3] calculates the local threshold value based on the mean value of the minimum and maximum intensities of pixels within an $N \times N$ sliding window. If the window is centered at the pixel (x, y) , the threshold for $I(x, y)$ is defined by equation (1).

$$T(x, y) = \begin{cases} \frac{I_{\max} + I_{\min}}{2} & \text{if } I_{\max} - I_{\min} \geq L \\ \delta & \text{if } I_{\max} - I_{\min} < L \end{cases} \quad (1)$$

where I_{\max} and I_{\min} are the maximum and minimum intensities in the window centered in the pixel (x, y) , respectively, and δ indicates a global threshold value (can be calculated by applying Otsu's method to the entire image). Window size N and parameter L are two parameters of this method. It works well only when the contrast between foreground and background is huge.

In Niblack's method [4], the threshold for each pixel is obtained based on the local mean and variance inside a sliding window as in equation (2).

$$T(x, y) = m(x, y) + ks(x, y) \quad (2)$$

where $m(x, y)$ and $s(x, y)$ are the local mean and standard deviation of the gray level in the local window centered on the pixel (x, y) , respectively. k is a user-tuned parameter and is set to -0.2 as in [4].

Note, however, that this method only results in significant background noises due to a large variance that usually occurs in a poor illuminated document, making the threshold too sensitive to noise.

Sauvola's method [5] is an improvement that can reduce the noise and enhance the textual part well as defined in equation (3).

$$T(x, y) = m(x, y) \times \left(1 + k \left(1 - \frac{s(x, y)}{R} \right) \right) \quad (3)$$

where R is constant; for the image with 256 gray levels, it is equal to 128 in most cases. k is a positive value in the range [0.2, 0.5]. For the high contrast region in the image, $s(x, y) \approx R$, which results in $T(x, y) \approx m(x, y)$ as with Niblack's method. It gives a good result since the images have high contrast between foreground and background; in contrast, the text regions are eliminated.

Some different methods have also been reported in literature for document images

binarization such as hybrid algorithm. In [6], the difference between foreground and background is enhanced based on a water flow model, and Otsu's method is then applied to binarize such enhanced image. In [7], Gatos performed a method based on background subtraction, and document background surface is estimated based on the binary document image generated by Sauvola's method followed by a thresholding surface on the estimated background surface. These methods require the predetermined parameters, which serves as a limitation for application to a wide variety of degraded document images. Others, through texture analysis [9] or decomposition method [10], combine different types of image information and domain knowledge, tending to be complex and time-consuming.

3. PROPOSED APPROACH

In this section, we describe our proposed approach in detail. The overall diagram of our proposed binarization method is depicted in Fig. 2, a multi-stages method. In particular, we divide this section into subsections corresponding to each stage; these subsections deal with the pre-processing step at the first stage to remove noise and enhance image, feature extraction and text stroke edge detection in the next, and followed by local threshold estimation to get the binarization result. Finally, a refinement procedure as a post-processing step for the improvement of the quality of the final results is performed. A step-by-step description of the proposed method is drawn in the following subsections:

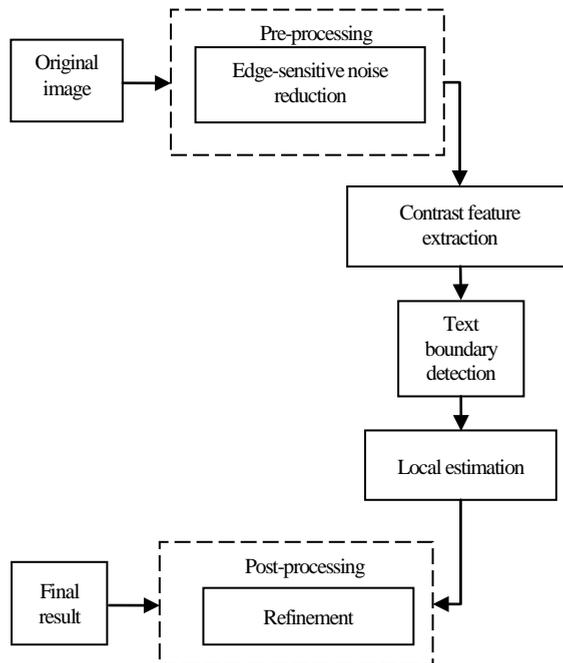


Fig. 2. Block diagram of the proposed method

3.1 Edge-Sensitive Noise Reduction

Due to degraded and poor-quality documents, a preprocessing stage is essential for the elimination of noisy areas and smoothing of background area. For this purpose, we perform an edge-sensitive noise reduction procedure [11].

Edge-sensitive noise reduction is as effective as the median filter in noise elimination but is also able to preserve even thin lines and sharp edges unlike the median filter. This procedure can be summarized in the following steps (the image is processed in subfields as a 5x5 window moves over all regions of the image):

Step 1: In the neighborhood window, evaluating whether the current pixel (central pixel) is lying at an edge of one of the following possible orientations:

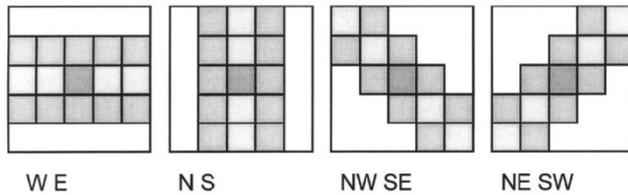


Fig. 3. Lanes around the central pixel in different orientations

Step 2: Computing the average of all pixels lying on the lanes accompanying the central row; this is performed along each lane according to the current orientation

Calculating the absolute Δ between the lanes

$$\Delta = \left| \bar{x}_1 - \bar{x}_2 \right| \tag{4}$$

The Δ value can be used as a measure of sharpness of the edge. Averaging along the lanes in advance makes this local criterion robust against noise-induced, single-pixel fluctuations.

Step 3: Searching the highest Δ among all orientations to find out which edge orientation fits best

Step 4: Replacing the central pixel by a directed mean over all pixels indicated as in Fig. 4 below, depending on the orientation of the edge detected

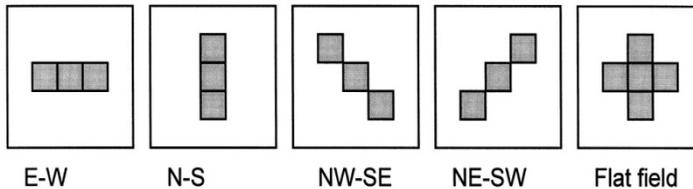


Fig. 4. Areas for the calculation of the directed mean in the corresponding orientation of the edge

When a central pixel is not an edge boundary, it is replaced by a cross-shaped neighborhood average (Flat field). The result is more smoothing and more effective noise reduction in these image areas.

The result of applying the edge-sensitive noise reduction procedure is shown in Fig. 5. As we can see, the result obtained after this step is the image with thin lines and sharp edges preserved in contrast to other noise removal method such as Median filter, which tends to blur the edges.

3.2 Contrast Feature Extraction and Text Boundary Detection

For document binarization, the most powerful features are those that take into account the structural characteristics of the characters. Stroke width is a major structural feature of a character stroke. In this paper, we present an extraction of the local contrast feature taking into account the structural characteristics and stroke width of the text. It performs an increase in the discrimination of the text and non-text areas. Based on the stroke width and gray level value, the contrast feature at pixel (x_0, y_0) is defined as follows:

$$SC(x_0, y_0) = \max_{k=0}^3 \{ \min(Avg_{SW}(p_k), Avg_{SW}(p_{k+1}), Avg_{SW}(p'_k), Avg_{SW}(p'_{k+1})) \} - I(x_0, y_0) \quad (5)$$

$$Avg_{SW}(p_k) = \left(\frac{\sum_{i=-SW}^{SW} \sum_{j=-SW}^{SW} I(p_{kx} - i, p_{ky} - j)}{(2 \times SW + 1)^2} \right) \quad (6)$$

where $I(x_0, y_0)$ represents the gray level of pixel (x_0, y_0) ; p_{kx}, p_{ky} are the coordinates of p_k in the x and y coordinates, and $p'_i = p_{(i+4) \bmod 8}$ and SW denote the stroke width of the characters.

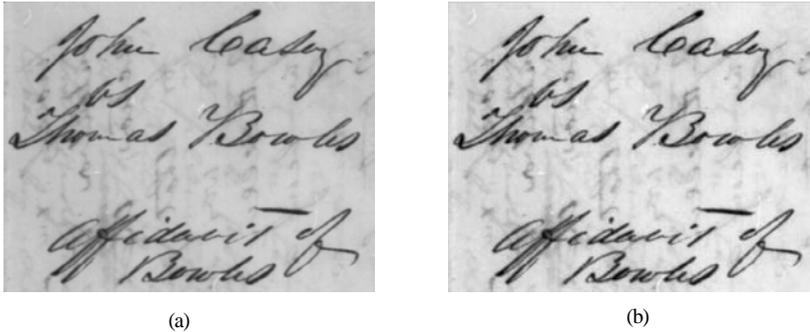


Fig. 5. Preprocessing result: (a) after applying Median filter and (b) after applying the edge-sensitive noise reduction procedure on the original image in Fig.1 (b)

Stroke width SW is known to be a useful structural feature of text. The quality of binarization result can be improved if we have an appropriate measurement of this feature. The run length

histogram [12] can be used to measure stroke width. Note, however, that this process is time-consuming. Here, we apply a method in [13] for the determination of stroke width. First, we use the Canny edge detector for the extraction of edge map. Applying this detector ensures avoiding the loss of weak edges, which can eliminate the characters. Stroke width is determined based on the distances between two successive edge pixels in the horizontal scan line. We compute the distance histogram between two successive edges in the horizontal direction, and it is denoted as one-dimensional array $His(d)$ where $d \in \{2, \dots, MD\}$, with MD as the maximum distance to be counted. $His(d)$ is the frequency of distance d . Stroke width SW is defined as the distance with highest frequency of distance d .

Based on the contrast feature extracted for every pixel in the image, we can construct the contrast feature image shown in Fig. 6(a). The purpose of this construction is to detect the high contrast image pixels lying around the text stroke boundary. This image has a bimodal histogram because the contrast value around the text stroke boundary varies within a small range but is obviously much larger compared with the image contrast within the background. Therefore, here, we apply Otsu's global thresholding method to detect the high-contrast pixels on the one hand and to detect the pixels around the text stroke boundary on the other. The obtained result shown in Fig. 6(b).

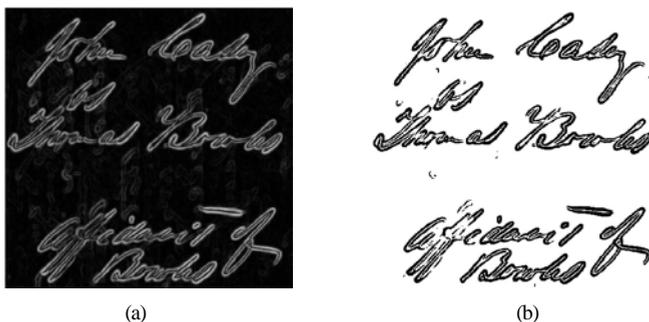


Fig. 6. (a) Contrast feature image and (b) Text stroke boundary detected by Otsu's method

3.3 Local Estimation

The foreground text will be extracted from the background based on the text stroke detected from the previous step. To determine whether a pixel belongs to text or background cluster, we make an estimation within a local neighborhood window. The neighborhood window must at least be larger than the stroke width to contain stroke edge pixels. Therefore, here, the size of the neighborhood window is set based on the stroke width of the document image determined as mentioned above. The extracted result is shown in Fig. 7 below.

$$R(x, y) = \begin{cases} 0 & I(x, y) \leq E_{mean} + \frac{E_{std}}{2} \\ 1 & otherwise \end{cases} \quad (7)$$

where E_{mean} refers to the mean and E_{std} denotes the standard deviation of intensity of the detected stroke edge pixels within the local neighborhood window. E_{mean} is computed using equation (8).

$$E_{mean} = \frac{\sum_{neighbor} I(x, y) * (1 - E)}{N_e} \quad (8)$$

where E refers to the determined stroke edge image and N_e denotes the number of detected stroke edge pixels within a local neighborhood window.

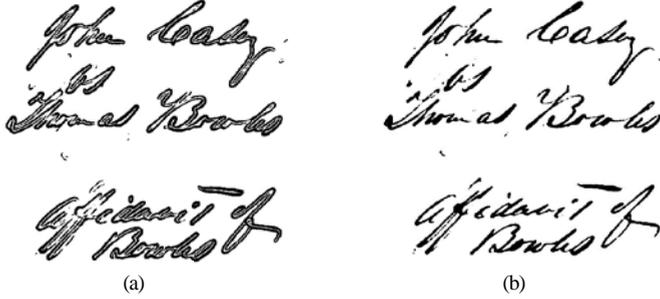


Fig. 7. Obtained binarization result (b) using local thresholding estimation on text stroke boundary (a)

3.4 Refinement

Performing a threshold on an image often introduces a certain amount of error that can be corrected through a post-processing step. This step is applied to the binary image resulting from the previous step to eliminate the noise and fill the breaks, gaps, or holes for the improvement of the quality of text regions. For this purpose, we perform a two-phase refinement method by applying shrink and swell filtering [14] successively. In the first phase, a shrink filter is used to remove the remaining noise from the background. An $N \times N$ sliding window is considered around each foreground pixel in the binary image. If the number of foreground pixels in the sliding window is smaller than a threshold, then this pixel is turned into background. Let $I_B(x, y)$ denote the binary image at pixel (x, y) obtained from the previous stage. According to shrink filtering, the following condition must be satisfied:

$$\begin{aligned} &\text{If } I_B(x, y) = 1 \text{ AND } \sum_{ky=y-w}^{y+w} \sum_{kx=x-w}^{x+w} I_B(kx, ky) < T_1 \\ &\text{then } I_B(x, y) = 0 \end{aligned} \quad (9)$$

where $N = 2w + 1$ is the window size and T_1 denotes the threshold for pixel density in the sliding window.

In the second phase, a swell filter is used to fill the possible breaks, gaps, or holes.

$$\begin{aligned} &\text{If } I_B(x, y) = 0 \text{ AND } \sum_{ky=y-w}^{y+w} \sum_{kx=x-w}^{x+w} I_B(kx, ky) > T_2 \\ &\text{then } I_B(x, y) = 1 \end{aligned} \quad (10)$$

In our experience, the value of parameters in turn is: $w = 2$, $T_1 = T_2 = 16$.

The final result after applying this stage is shown in Fig. 8. As we can see, the quality of the final result is improved. Some noise remaining from the previous steps is removed, and breaks or gaps are also filled for the enhancement of the binarization result.

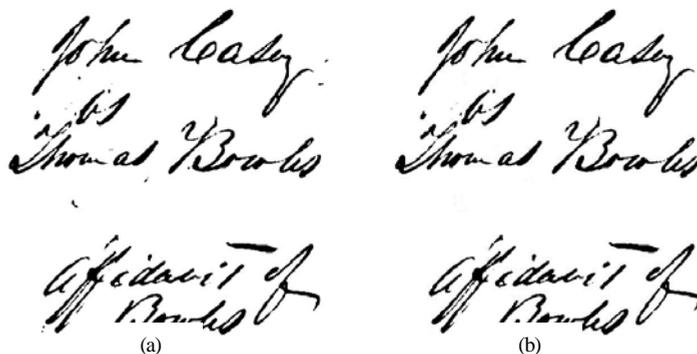


Fig. 8. Final result (b) after applying the refinement procedure for image (a)

4. EXPERIMENTAL RESULTS

In this section, we present some of our experimental results on DIBCO'09 and H-DIBCO'10 datasets. It contains a number of representative handwritten document images under different types of degradation. We also test on a set of typed document images. All images are of poor quality, with shadows, non-uniform illumination, smear, and strain. Fig. 1(a) presents an example of a typed document, whereas Figs. 1(b) and (c) show examples of a handwritten document. To make an evaluation of the obtained results, a comparison with other methods is carried out as in Fig. 9. We compare the performance of our method with that of Otsu[1], Bernsen[3], Niblack[4], and Sauvola[5]. The parameters of Bernsen's, Niblack's and Sauvola's methods including the window size, weight of the local standard variation, and weight of the local dynamic range of standard variation are all set according to the recommendations within the reported papers. For the evaluation of our performance vis-à-vis other state-of-the-art methods, we perform two comparisons corresponding to two criteria. First is holistic image visual judgment, and the other is quantitative comparison. As shown in all cases, in visual comparison, our algorithm outperforms all the rest of the algorithms in terms of the preservation of meaningful textual information. In the final obtained results, noise can be seen to have been removed from document images. To evaluate our algorithm better, quantitative evaluation is done by comparing with ground truth labels in the database as shown in Table 1.



Fig. 9. Binarization result of the proposed method for degraded document image in comparison with other methods

The quantitative evaluations of accuracy are displayed under the corresponding binary results in the form of F-measure, which combines the two complementary measures of precision and recall of pixels in a binary result relative to the ground truth. As an overall measure, F-measure is defined by equation (11),

$$F - measure = \frac{(1 + \beta^2)(Precision \times Recall)}{\beta^2(Precision + Recall)} \tag{11}$$

where $Recall = \frac{TP}{TP + FN}$

$$Precision = \frac{TP}{TP + FP}$$

with TP, FP, and FN denoting true-positive, false-positive, and false-negative, respectively.

In our work, β^2 is set to 1. This is also known as the F_1 measure, for balance where recall and precision are evenly weighted.

In Fig. 9, based on the visual comparison, we can see that our proposed method outperforms state-of-the-art methods. More experimental results – shown in Fig. 10, Fig. 11, and Fig. 12 – demonstrate the efficiency of our method.

Table 1. Quantitative evaluation

| Method | Precision | Recall | F-measure |
|----------------------|-----------|--------|-----------|
| Otsu's method [1] | 70.5 | 91.3 | 79.56 |
| Bemsen's method [3] | 63.41 | 85.62 | 72.86 |
| Niblack's method [4] | 45.11 | 96.2 | 61.42 |
| Sauvola's method [5] | 80.17 | 92.14 | 85.74 |
| Proposed method | 86.87 | 94.58 | 90.56 |

As we can see in the comparison, Otsu's method [1] is based on a bimodal histogram; therefore, it is efficient only in case of good separation between foreground and background. As shown in Fig. 10 (b), Fig. 11(b), and Fig. 12 (b), it fails to handle document images with poor contrast and background variation. Bemsen [3] also fails in the cases in Fig. 10 (c), Fig. 11 (c), and Fig. 12 (c). For adaptive thresholding method such as Niblack's method [4], the obtained results yield a large amount of background noise labeled as text as shown in Fig. 10(d). Another adaptive method in the comparison is Sauvola's method [5]. As a result, as shown in Fig.10 (e), Fig. 11(e), and Fig. 12(e), background regions can be removed properly, but the text regions are also eliminated or broken up for faded characters in cases of low contrast between foreground and background. In some cases, our method obtains results that are a little bit similar to Sauvola's result as shown in Fig. 10. In cases wherein the document image has smear or bleeding-through (Fig. 11, Fig. 12), however, the results of our method are better. Table 1 shows all comparative experimental results in terms of Recall, Precision, and F-measure for the evaluation. Based on these results, we can see that our method has the best overall performance compared with other methods.

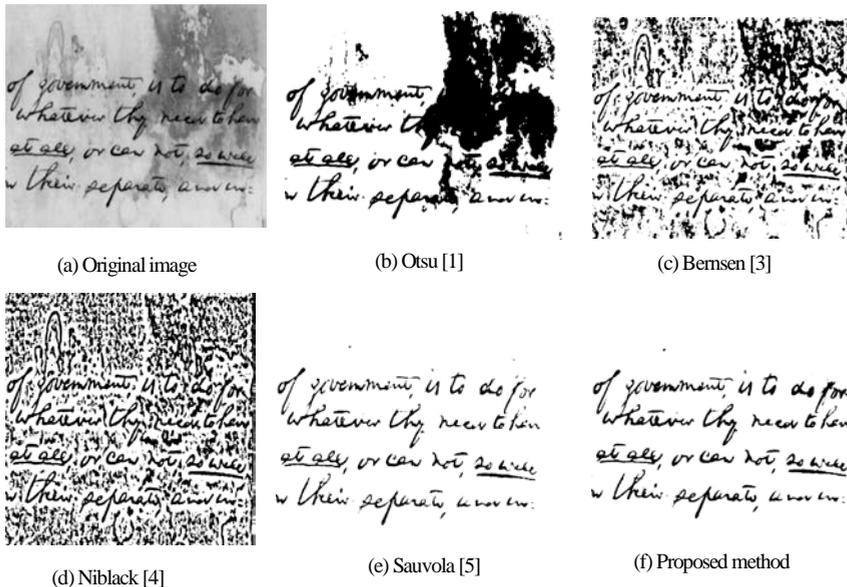


Fig. 10. Comparison of proposed method with other methods: (a) original images, (b) Otsu's results, (c) Bemsen's results, (d) Niblack's results, (e) Sauvola's results, and (f) proposed method's results

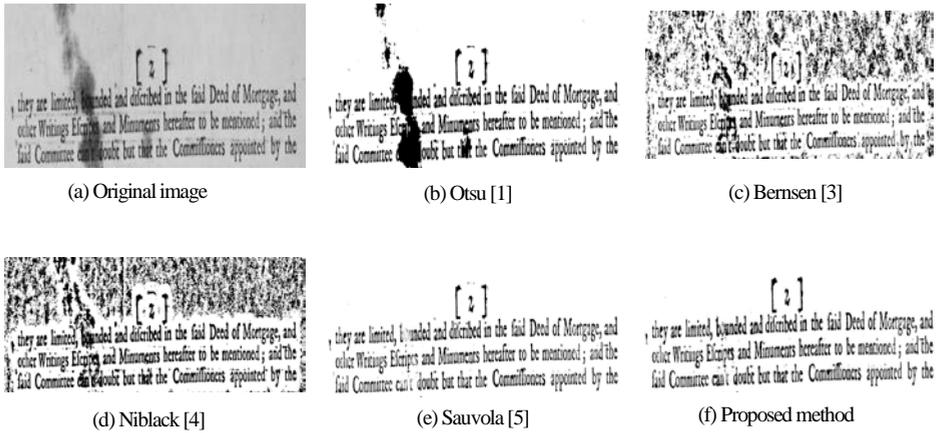


Fig. 11. Comparison of proposed method with other methods: (a) original images, (b) Otsu's results, (c) Bernsen's results, (d) Niblack's results, (e) Sauvola's results, and (f) proposed method's results

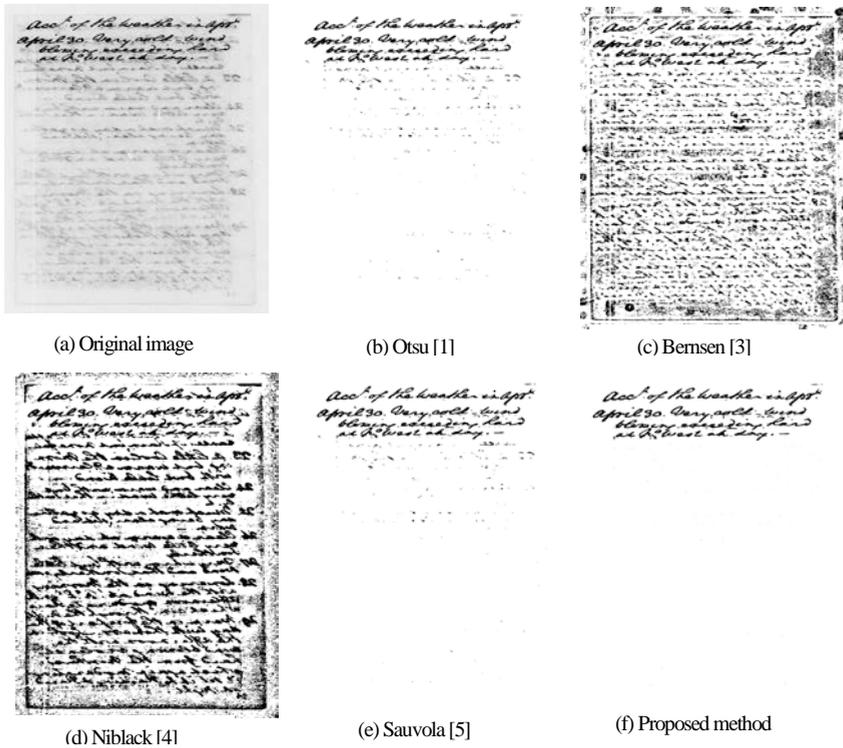


Fig. 12. Comparison of proposed method with other methods: (a) original images, (b) Otsu's results, (c) Bernsen's results, (d) Niblack's results, (e) Sauvola's results, and (f) proposed method's results

5. CONCLUSION

In this paper, we present a method for the binarization of degraded document images by exploiting the local contrast feature, which triggers an increase in the discrimination of the text and non-text areas. Then local thresholding estimation comes after the text boundary detection to generate appropriate results for the extraction of text pixels from the background area. To improve the obtained results, a refinement procedure is finally performed. Compared with other existing global as well as local thresholding methods, our proposed method performs better on low-quality document images, especially those with uneven illumination, shadow, and low contrast. Experimental results obtained from applying the proposed method to various challenging document images have been presented to demonstrate its effectiveness and superior performance. A more efficient post-processing method to preserve stroke connectivity and for the removal and filling of possible breaks, gaps, or holes can be considered the next development for this study for the significant improvement of the final results.

REFERENCE

- [1] N. Otsu, "A threshold selection method from grey level histogram," *IEEE Transactions on System, Man, and Cybernetics*, vol. 9, 1979, pp.62-66.
- [2] J. Kitter and J. Illingworth, "On threshold selection using clustering criteria," *IEEE Transactions on System, Man, and Cybernetics*, vol. SMC-15, 1985, pp.652-655.
- [3] J. Bernsen, "Dynamic thresholding of gray level images," Proceedings of the 8th International Conference on Pattern Recognition, vol. 2, 1986, pp. 1251-1255.
- [4] W. Niblack, *An Introduction to Digital Image Processing*, NJ, USA: Prentice Hall, Englewood Cliffs, 1986.
- [5] J. Sauvola, and M. Pietikainen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, 2000, pp. 225-236.
- [6] I. K. Kim, D. W. Jung, and R. H. Park, "Document image binarization based on topographic analysis using a water flow model," *Pattern Recognition*, vol. 35, 2002, pp.265-277.
- [7] B. Gatos, I. Pratikakis, and S. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, 2006, pp.317-327.
- [8] O. Trier and T. Taxt, "Evaluation of binarization methods for document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, pp.312-315.
- [9] Y. Liu and S. Srihari, "Document image binarization based on texture features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, pp.540-544.
- [10] Y. Chen and G. Leedham, "Decompose algorithm for thresholding degraded historical document images," *IEEE Proceedings on Vision, Image, and Signal Processing*, 2005, pp.702-714.
- [11] H. G. Adelman, "An edge-sensitive noise reduction algorithm for image processing," *Computers in Biology and Medicine*, vol. 29, 1999, pp.137-145.
- [12] Y. Yang, and H. Yan, "An adaptive logical method for the binarization of degraded document images," *Pattern Recognition*, vol. 33, 2000, pp.787-807.
- [13] M. Valizadeh, M. Komeili, N. Armanfard, and E. Kabir, "Degraded document image binarization based on combination of two complementary algorithms," *International Conference in Advances in Computer Tools for Engineering Applications*, 2009, pp.595-599.
- [14] R.J. Schilling, *Fundamentals of Robotics Analysis and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1990.



Le Thi Khue Van

She received the B.S degree in Mathematics and Computer Science from University of Science Ho Chi Minh City, Vietnam in 2009. She is currently a MS student in Electronics and Computer Science department of Chonnam National University, Korea. Her interesting researches are multimedia and image processing, document analysis, artificial intelligence and computer vision.



Guesang Lee

1980 He received the B.S degree in Electrical Engineering from Seoul National University.1982 He received the M.S degree in Computer Engineering from Seoul National University.1991 He received Ph.D. degree in Computer Science from Pennsylvania State University. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. Research Interests: Image processing, computer vision and video coding.