A Two-stage and Parameter-free Binarization Method for Degraded Document Images

Yung-Hsiang Chiu[1], Kuo-Liang Chung[1], Yong-Huai Huang[2], Wei-Ning Yang[3], Chi-Huang Liao[4]
[1]Department of Computer Science and Information Engineering,
National Taiwan University of Science and Technology, Taipei, Taiwan, R. O. C.
[2]Institute of Computer and Communication Engineering,
Jinwen University of Science and Technology, Taipei, Taiwan, R. O. C.
[3]Department of Information Management,
National Taiwan University of Science and Technology, Taipei, Taiwan, R. O. C.
[4]System Online Co. Ltd., Taiwan, R. O. C.

Abstract

Binarization plays an important role in document image processing, especially in degraded documents. For degraded document images, adaptive binarization methods often incorporate local information to determine the binarization threshold for each individual pixel in the document image. We propose a two-stage parameter-free window-based method to binarize the degraded document images. In the first stage, a proposed scheme is used to determine a proper window size beyond which no substantial increase in the local variation of pixel intensities is observed. In the second stage, based on the determined window size, a noise-suppressing scheme delivers the final binarized image by contrasting two binarized images which are produced by two adaptive thresholding schemes depending on the change rate of the number of binarized foreground pixels. Empirical results demonstrate that the proposed method is competitive when compared to the existing adaptive binarization methods and achieves better performance in F-measure.

Keywords: *Adaptive binarization method, Degraded document image, Document image processing.*

1. INTRODUCTION

Document image processing is necessary for storing, transmitting, and managing digital documents. Among different types of document image processing, binarization is a preliminary process and the resultant binary images usually affect the performance of the succeeding processes, such as the document image segmentation, the optical character recognition, and so on. For binarization, each pixel in a document image can be classified as a foreground or a background pixel. Pixels inside characters, lines, and curves in a document image are foreground pixels and should be binarized as black pixels and the remaining background pixels should be binarized as white pixels.

For maximizing the between-class variance of foreground and background pixels, Otsu [1] proposed an automatic thresholding scheme to determine a global threshold for the input image. It usually yield good resultant binary images. However, the determined global threshold may not be applicable for degraded document images since intensities of foreground and background pixels are contaminated at different positions of the images. To alleviate the problem caused by the degraded document images, adaptive binarization schemes [2, 3, 4, 5, 7, 8] which incorporate the information from local statistics of an image are proposed to improve the global thresholding method. Niblack [2] presented a window-based method to determine the threshold for each pixel by incorporating the information of the mean and the standard deviation of gray levels within each window. Sauvola and Pietikainen [3] modified Niblack's method by proposing different weights on the mean and the standard deviation of gray levels within each window. For blueprint images, Zhao *et al.* [4] utilized geometric features and proposed an efficient window-based thresholding method. Gatos *et al.* [5] binarize the document image by contrasting the document image to the background surface which is constructed by interpolating the background pixels after removing the binarized foreground pixels via Sauvola and Pietikainen's method. Based on the edge map detected by the Canny edge-detector [6], Chen *et al.* [7] binarized the input document image using a pair of high and low thresholds. Moghaddam and Cheriet [8] proposed a multi-scale window-based thresholding scheme which first generates several binarized images based on different window sizes and then iteratively combines the binarized images to yield the final binarized image.

In this paper, we presented a two-stage parameter-free window-based method to binarize the degraded documents. In the first stage, a proposed scheme is used to determine a proper window size beyond which no substantial increase in the local variation of gray levels is observed. In the second stage, given the determined window size, a noise-suppressing scheme delivers the final binary image by contrasting two binarized images which are produced by two adaptive thresholding schemes depending on the change rate of the number of foreground pixels. Empirical results demonstrate that the proposed method is competitive when compared to the existing adaptive binarization methods and achieves better performance in F-measure.

2. CHALLENGES IN ADAPTIVE BINARIZATION

The adaptive binarization scheme needs to deal with two challenges: (a) the determination of a proper window size used to collect the local information and (b) the trade-off between detail preservation and noise suppression. These two challenges motivate the research of this paper and are addressed in this section.

The quality of the resultant binary document images produced by the existing adaptive binarization methods often are very sensitive to the window size used [2, 3, 4, 5]. Proper window size usually depends on the scale of objects in the document images. The document images with large objects require large window size in the adaptive binarization scheme.

Binarizing an image as shown in Figure 1 (a) with large objects using smaller than necessary window size may erroneously binarize foreground pixels to background pixels as shown in Figure 1 (b). Figure 1 (c) illustrates a better binarized result of Figure 1 (a) when a large windows size is used. However, adaptive binarization scheme with larger than necessary window size will not significantly increase the quality of the binarized images, as shown in Figure 1 (e) and (f), but incurs higher computational cost.

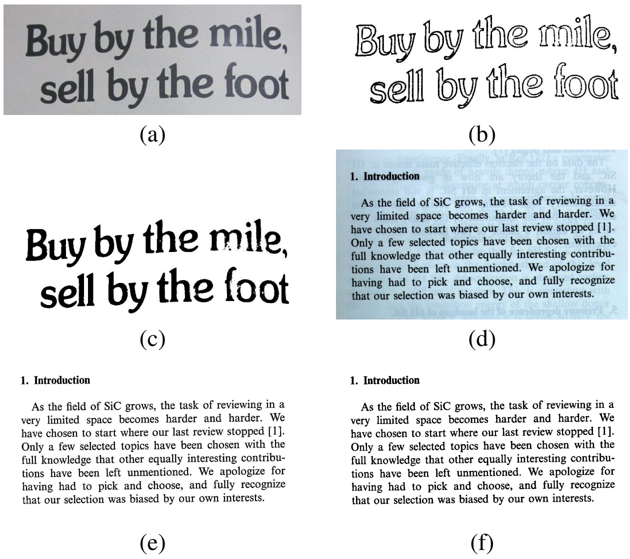


Figure 1: The effect of window size when using Sauvola and Pietikainen's method. (a) Document image with large-scale characters. (b) Binarized image of (a) using a 9×9 window. (c) Binarized image of (a) using a 33×33 window. (d) Document image with small-scale characters. (e) Binarized image of (d) using a 9×9 window. (f) Binarized image of (d) using a 33×33 window.

For proper window size, Gatos *et al.* [5] suggest that window size should cover at least 1 to 2 characters. However, detecting character size usually requires image segmentation and is difficult for degraded documents. Chen *et al.* [7] apply a 3×3 window and determine two thresholds based on the edge pixels detected by the Canny edge detector. The quality of the binarized image heavily depends on the correctness of the edge map which is usually poor for degraded documents. Moghaddam and Cheriet [8] propose a scheme which starts with a large window size determined by the average line height of the input document image and iteratively reduces to a proper window size. Since the average line height is usually determined by the image segmentation process and the proposed scheme suffers from the same problem as Gatos *et al.*'s method.

Free from other image pre-processings, we first apply Otsu's method to obtain a rough foreground image and then determine a proper window size based on the change rate of the variation of the foreground intensities within each window. In addition to determining the proper window size, the trade-off between the preservation of detailed contents and noise suppressing should be addressed in the adaptive binarization scheme.

Let f be the input document image and the intensity value of the pixel at position (x, y) is denoted by $f(x, y)$, $0 \leq f(x, y) \leq 1$. Given a specific window of size $w \times w$ with $w = 2r + 1$, the threshold used for binarization in Niblack's method is expressed as

$$T_{Nib,w}(x, y) = \mu_w(f, x, y) + k\sigma_w(f, x, y) \tag{1}$$

where k is a user-defined parameter and $\mu_w(f, x, y)$ and $\sigma_w(f, x, y)$ represent respectively the mean and standard deviation of intensities of the pixels within the window centered at (x, y) and can be expressed as

$$\mu_w(f, x, y) = \frac{1}{w^2} \sum_{i=-r}^{r} \sum_{j=-r}^{r} f(x + i, y + j), \tag{2}$$

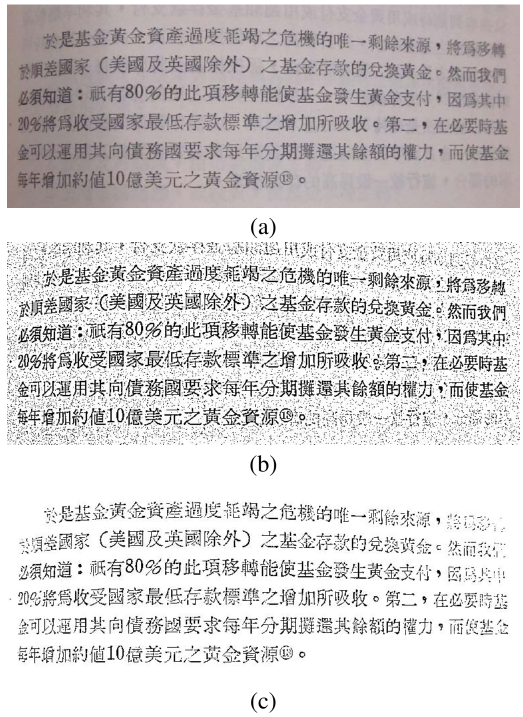


Figure 2: The effect of k' when using Sauvola and Pietikainen's method. (a) Degraded document image. (b) Binarized image using $k' = 0.01$. (c) Binarized image using $k' = 0.2$.

$$\sigma_w(f, x, y) = \sqrt{\frac{1}{w^2} \sum_{i=-r}^{r} \sum_{j=-r}^{r} (f(x + i, y + j) - \mu_w(f, x, y))^2}. \tag{3}$$

To improve Niblack's method, Sauvola and Pietikainen [3] proposed a modified threshold $T_{Sau,w}(x, y)$ which is expressed as

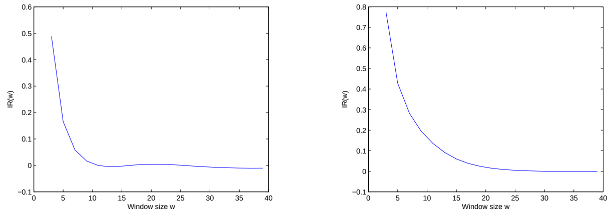
$$T_{Sau,w}(x, y) = \mu_w(f, x, y) \times \left(1 - k'\left(1 - \frac{\sigma_w(f, x, y)}{R}\right)\right) \tag{4}$$

where both R and k' are set to 0.5 in [3].

Parameters k and k' used in Eq. (1) and Eq. (4), respectively, are sensitive to the contents of the input document images and may not be applicable for degraded document images. For example, for a degraded document image in Figure 2 (a), Figure 2 (b) and (c) are binarized images obtained by Sauvola and Pietikainen's method with $k' = 0.01$ and $k' = 0.2$, respectively. The binarized image with smaller k' preserves more detailed contents but suffers from more noises. This observation motivates using two thresholding schemes to produce two binarized images from which the final binarized image is delivered.

3. THE PROPOSED TWO-STAGE AND PARAMETER-FREE BINARIZATION METHOD

In this section, we present a two-stage and parameter-free binarization scheme for degraded document images. The first stage determines a proper window size by considering the variation of foreground pixel intensities within windows. In the second stage, based on the window size determined in stage 1, a final binarized image is delivered by contrasting two binarized images produced by two adaptive thresholding schemes which consider the content preser-



(a) (b)

Figure 3: $IR(w)$ for documents in Figure 1(a) and (d)

vation and noise suppressing.

3.1 Determine the proper window size

To start the two-stage binarization scheme, we first apply the Gaussian low-pass filter to obtain the smoothed image and then the Otsu's method is used to determine the set of the rough foreground pixels, denoted by RFG. The variation of foreground pixel intensities within each window usually increases as the window size increases. Large window size usually delivers binarized images with better quality but suffers from larger computational cost, indicating that the window size larger than necessary for acceptable quality should not be adopted.

Since binarizing with small window size may erroneously binarize foreground pixels to background pixels and using large window size increases the computational cost without significantly increasing the quality, we start with a small window size and keep increasing the window size until no substantial increase in the variation of the pixel intensities within each window is observed.

Starting with a window of size 3×3, we compute the standard deviation of the foreground pixel intensities within each window and use the average of the standard deviations as the indicator to search for the proper window size.

Let $IR(w)$ denote the increasing rate of the average standard deviation when enlarging the window from $w \times w$ to $(w+2) \times (w+2)$ and is expressed as

$$IR(w) = \frac{\overline{\sigma}_{w+2} - \overline{\sigma}_w}{\overline{\sigma}_w} \tag{5}$$

with

$$\overline{\sigma}_w = \frac{1}{|RFG|} \sum_{(x,y) \in RFG} \sigma_w(f, x, y), \tag{6}$$

where $|RFG|$ is the cardinality of the set of rough foreground pixels RFG and $\sigma_w(f, x, y)$ is the standard deviation of pixel intensities within the $w \times w$ window centered at (x, y). The increasing rate $IR(w)$ decreases as the window size w increases as shown in Figure 3. The proper window size w^* is the smallest window size such that $IR(w)$ is less than or equal to 0.01; that is, $w^* = \min\{w : IR(w) \le 0.01\}$.

3.2 Proposed noise-suppressing thresholding scheme

For low contrasting documents, information contained in the neighborhood of a specific pixel can be helpful in determining the binarization threshold. Let $g(x, y)$ denote the gradient magnitude, proposed by Sobel operator [9], formed from the pixels in the neighborhood of pixel (x, y). Large values of $g(x, y)$ indicate that pixel (x, y) is around the boundary between foreground and background pixels. Based on the window size $w^* = 2r^* + 1$ determined in stage 1, compute

$$\mu_{w^*}(g, x, y) = \frac{1}{w^{*2}} \sum_{i=-r^*}^{r^*} \sum_{j=-r^*}^{r^*} g(x+i, y+j), \tag{7}$$

and

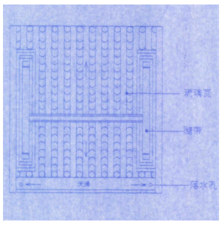
$$\sigma_{w^*}(g, x, y) = \frac{1}{w^*} \sqrt{\sum_{i=-r^*}^{r^*} \sum_{j=-r^*}^{r^*} (g(x+i, y+j) - \mu_{w^*}(g, x, y))^2} \tag{8}$$

When incorporating the information contained in the mean $\mu_{w^*}(g, x, y)$ and the standard deviation $\sigma_{w^*}(g, x, y)$ of local gradients around pixel (x, y), we propose a binarization threshold

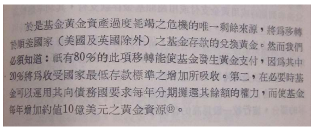
$$T(x, y) = \mu_{w^*}(f, x, y)\left(1 - k'' e^{-\left((\mu_{w^*}(g,x,y) + \sigma_{w^*}(g,x,y))/M\right)}\right), \tag{9}$$

where $M = \max_{(x,y) \in D}\{\mu_{w^*}(g, x, y) + \sigma_{w^*}(g, x, y)\}$ with D denoting the input document. Decreasing parameter k'' increases the threshold $T(x, y)$ and the number of pixels identified as foreground pixels increases. Thus when decreasing parameter k'' from some large initial value, the number of identified foreground pixels increases sharply and becomes saturated as most of the true foreground pixels are correctly identified as foreground pixels. If keep decreasing the parameter k'', the number of identified foreground pixels may increase sharply again since the present noises are erroneously identified as foreground pixels. Let $|FG|(k'')$ denote the number of identified foreground pixels using threshold $T(x, y)$ on pixel $f(x, y)$. Starting with $k''_0 = 0.2$, iteratively decrease the threshold by modifying the parameter k'' according to $k''_{i+1} = (0.9)k''_i$. Denote by k''_{1^*} and k''_{2^*}, with $k''_{1^*} > k''_{2^*}$, the reflection points of the function $|FG|(k'')$; that is, $\frac{d^2|FG|(k'')}{dk''^2}\big|_{k''=k''_{1^*}} = \frac{d^2|FG|(k'')}{dk''^2}\big|_{k''=k''_{2^*}} = 0$. Two thresholds $T_1(x, y)$ and $T_2(x, y)$ for pixel (x, y) are determined by Eq. 9.

Let B_i denote the binarized image produced by the threshold $T_i(x, y)$, $i = 1, 2$. Since $T_1(x, y) < T_2(x, y)$, in image B_1 noises are suppressed but some true foreground pixels are not correctly identified. On the other hand, in image B_2 almost all the true foreground pixels are identified but in the meantime the noises are about to be included. Two binarized images B_1 and B_2 are then contrasted to deliver the final binarized image. Since $T_1(x, y) < T_2(x, y)$, if (x, y) is a background pixel in B_2, then (x, y) must be a background pixel in B_1 and is very likely to be a true background pixel in the document. Thus the pixel appeared to be a background pixel in B_2 will be identified as a background pixel in the final binarized image. Similarly, if (x, y) is a foreground pixel in B_1, then (x, y) must be a foreground pixel in B_2 and is very likely to be a true foreground pixel in the document. Thus the pixel appeared to be a foreground pixel in B_1 will be identified as a foreground pixel in the final binarized image. If (x, y) is a background pixel in B_1 and a foreground pixel in B_2, then it can be a foreground or a noise in the document. To tackle such pixels, a region-growing scheme, based on the pixels which are identified as foreground pixels in both B_1 and B_2, is proposed. For each pixel (x, y) identified as a foreground pixel in both B_1 and B_2 images, a 3×3 window centered at (x, y) is considered. Within the window, for each of the eight pixels surrounding (x, y), if it is identified as a background in B_1 and a foreground in B_2, then it is identified as a foreground pixel in the final binarized image. Furthermore, the region-growing scheme will be applied to the newly identified foreground pixel by the region-growing scheme. This proposed region-growing scheme mends some true foreground pixels that are suppressed in B_1 when suppressing the noises.



(a) Test image 1
($\#FG/N = 11.03\%$)



(b) Test image 2
($\#FG/N = 13.12\%$)

Figure 4: The test images

Table 1: The performance comparison of test image 1

Method	Precision	Recall	Accuracy	F-measure
[3]	69.6861	78.4149	93.8572	73.7932
[4]	74.8738	57.3418	92.8991	64.9454
[5]	75.6668	65.1826	93.8480	70.0345
[7]	58.9203	87.3272	91.8872	70.3649
[8]	62.6726	83.4845	92.6945	71.5968
Proposed	83.0143	79.3807	95.9345	81.1568

4. EXPERIMENTAL RESULTS

In this section, we empirically compare the proposed method with six existing methods — Sauvola and Pietikainen's method [3], Zhao *et al.*'s method [4], Gatos *et al.*'s method [5], Chen *et al.*'s method [7], and Moghaddam and Cheriet's method [8]. All methods are implemented by Borland C++ Builder 6.0 and run on a standard PC with AMD Athlon 64X2 4800+ CPU(2.5 GHz) and 1.87 GB of RAM. The test images include scanned machine-printed image and blueprint image for which we create the true binary image by human eyes. Test images 1 in Figure 4 is the blueprint image of architectures with proportion $\#FG/N$ of foreground pixels, where $\#FG$ is the number of true foreground pixels in the document with N pixels. Test images 2 in Figure 4 is a textual image with non-uniform illumination. The performance evaluations are based on four accuracy measures: (a) recall, (b) precision, (c) accuracy, and (d) F-measure. Recall is the proportion of correctly binarized foreground pixels within the true foreground pixels. Precision is the proportion of true foreground pixels within the binarized foreground pixels. Accuracy is the weighted average of the proportions of correctly binarized foreground and background pixels within the true corresponding pixels with weights proportional to the numbers of true foreground and background pixels. The F-measure is the harmonic mean of recallz and precision. Let TP and TN denote respectively the number of pixels that are correctly binarized as foreground and background pixels. And denote respectively by FP and FN the number of pixels that are erroneously binarized as foreground and background pixels. Then we have recall $= TP/(TP + FN)$, precision $= TP/(TP+FP)$, accuracy $= (TP+TN)/(N)$, and F-measure $= 2 \times$ recall $\times$ precision/(recall $+$ precision). Empirical results are listed in Tables 1 and 2 respectively.

Based on the empirical results, the following general conclusions are obvious:

1. The proposed method has significantly higher F-measure than the existing methods, indicating that the proposed method achieves higher accuracy in both recall and precision.
2. In terms of accuracy, the proposed method is competitive when compared to the existing methods.
3. For highly degraded documents such as test blueprint image 1, results in Tables 1 show that the proposed method achieves higher precision with acceptable recall compared to the existing methods.

Table 2: The performance comparison of test image 3

Method	Precision	Recall	Accuracy	F-measure
[3]	96.5006	72.2408	96.0135	82.6268
[4]	85.1910	60.3617	93.4216	70.6586
[5]	90.8324	85.6041	96.9771	88.1408
[7]	43.8361	89.6292	83.5699	58.8767
[8]	74.9070	93.0555	94.9981	83.0008
Proposed	89.1313	88.9508	97.1267	89.0409

5. CONCLUSION

In this paper, a two-stage parameter-free window-based binarization method is proposed. In general, the proposed binarization scheme is competitive when compared with the existing methods. Specifically, the proposed method has good performance in both recall and precision measures, resulting a higher F-measure.

6. ACKNOWLEDGEMENTS

K.-L. Chung, Y.-H. Huang, and W.-N. Yang are supported by the National Science Council of R.O.C. under Contract NSC98-2923-E-011-001-MY3, NSC99-2221-E-228-006, and NSC100-2218-E-011-006 respectively.

7. REFERENCES

[1] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transcations on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[2] W. Niblack, "An introduction to digital image processing," Prentice Hall, Englewood Cliffs, NJ, pp. 115–116, 1986.

[3] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.

[4] M. Zhao, Y. Yang and H. Yan, "An adaptive thresholding method for binarization of blueprint images," *Pattern Recognition Letter*, vol. 21, no. 10, pp. 927–943, 2000.

[5] B. Gatos, I. Pratikakis and S. J. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, no. 3, pp. 317–327, 2006.

[6] J. Canny, "A computational approach to edge detection," *IEEE Transcations on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[7] Q. Chen, Q. S. Sun, P. A. Heng and D. S. Xia, "A double-threshold image binarization method based on edge detector," *Pattern Recognition*, vol. 41, no. 4, pp. 1254–1267, 2008.

[8] R. F. Moghaddam and M. Cheriet, "A multi-scale framework for adaptive binarization of degraded document images," *Pattern Recognition*, vol. 43, no. 6, pp. 2186–2198, 2010.

[9] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," Section 7:1.3: Edge Detection, Addison Wesley, 1992.

ABOUT THE AUTHORS

Yung-Hsiang Chiu is now a Ph.D. student in the Department of Computer Science and Information Engineering of National Taiwan University of Science and Technology, Taiwan. His research interests include document image processing and advance video coding. His contact email is D9915013@mail.ntust.edu.tw.

Kuo-Liang Chung is a Chair Professor in the Department of Computer Science and Information Engineering of the National Taiwan University of Science and Technology, Taiwan. His current research interests include image processing, video coding, and data hiding. His contact email is k.l.chung@mail.ntust.edu.tw.

Yong-Huai Huang is an assistant professor in the Institute of Computer and Communication Engineering at Jinwen University of Science and Technology, Taiwan. His research interests include image processing and compression, and algorithms. His contact email is yonghuai@ms28.hinet.net.

Wei-Ning Yang is now an associate professor in the Department of Information Management, National Taiwan University of Science and Technology, Taiwan. His research interests include statistical analysis, stochastic simulation, and image processing. His contact email is yang@cs.ntust.edu.tw.

Chi-Huang Liao is now the manager at System Online Co. Ltd., Taiwan. His research interests include geographic information system, system integration, and image processing. His contact email is chin.laiw@msa.hinet.net.