Pattern Recognition 45 (2012) 4250–4262

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

PATTERN RECOGNITION

Parameter-free based two-stage method for binarizing degraded document images

Yung-Hsiang Chiu [a], Kuo-Liang Chung [a,*], Wei-Ning Yang [b], Yong-Huai Huang [c], Chi-Huang Liao [d]

[a] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 10607, Taiwan, ROC
[b] Department of Information Management, National Taiwan University of Science and Technology, No. 43, Section 4, Keelung Road, Taipei 10607, Taiwan, ROC
[c] Institute of Computer and Communication Engineering, Jinwen University of Science and Technology, No. 99, Anzhong Road, Xindian District, New Taipei City 23154, Taiwan, ROC
[d] System Online Co. Ltd. 8F, No. 247-1, Section 3, Jhong-Siao East Road, Taipei 10654, Taiwan, ROC

ARTICLE INFO

Article history:
Received 12 August 2011
Received in revised form
18 January 2012
Accepted 26 February 2012
Available online 5 March 2012

Keywords:
Adaptive binarization method
Degraded document image
Document image processing
Performance evaluation

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, an incremental 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 which incorporate the local mean gray and gradient values. Empirical results demonstrate that the proposed method is competitive when compared to the existing adaptive binarization methods and achieves better performance in precision, accuracy, and F-measure.

1. Introduction

In the past decades, many document image operations, such as document layout analyzing [1,2], data hiding [3], skew estimation [4,5], stroke extraction [6], document block classification [7], and optical character recognition [8], have been developed for storing, transmitting, and managing digital documents. Among the different types of document image operations, binarization is a preliminary process and the resultant binary images usually affect the performance of the succeeding processes, such as the document image segmentation and the optical character recognition. For binarization, each pixel in a document image is 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 [9] proposed an automatic thresholding scheme to determine a global threshold for the input image. Kapur et al. [10] determined a global threshold by maximizing the entropy of two partitioned subimages. For document images with satisfactory quality, Otsu's and Kapur et al.'s methods 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 degradation of document images, adaptive binarization schemes [11–16] which incorporate the information from local statistics of an image are proposed to improve the Otsu's and Kapur et al.'s methods. Niblack [11] 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 Pietikäinen [12] 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. [13] utilized geometric features and proposed an efficient window-based thresholding method. Kim et al. [17] proposed an adaptive method by first transforming an input document image into a three-dimensional terrain and then generating the binarized image by iteratively performing the water flowing process. Oh et al. [18] improved Kim et al.'s method by concentrating on the specific regions of importance and achieved better performance in terms of computational efficiency and image quality. Gatos et al. [14] binarized 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 Pietikäinen's method. Based on the edge map detected by the Canny edge-detector [19], Chen et al. [15] binarized the input document image using a pair of high and low thresholds. The resultant binary images achieve higher sensitivity (i.e. the proportion

[*] Corresponding author. Tel.: +886 2 27376771.
E-mail address: klchung01@gmail.com (K.-L. Chung).

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2012.02.023

of foreground pixels that are correctly binarized) but lower precision (i.e. the proportion of correctly binarized foreground pixels). Moghaddam and Cheriet [16] 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. Empirical results showed that the quality of the resultant binarized image was significantly affected by the binarized image generated by using the largest window size.

In this paper, we present a two-stage parameter-free window-based method to binarize the degraded document images. In the first stage, based on rough foreground pixels determined by Otsu's method, an incremental scheme is used to automatically determine a proper window size beyond which no substantial increase in the local variation of gray levels is observed. The reason for considering the foreground pixels only is because the variation of gray values within windows around foreground pixels increases substantially as the window size smaller than necessary increases until a proper window size is reached. In the second stage, based on the determined window size, a noise-suppressing scheme delivers the final binary image by contrasting two binarized images produced by two adaptive thresholding schemes which incorporate the local mean gray and gradient values. Empirical results demonstrate that the proposed method is competitive when compared to the six existing adaptive binarization methods—Niblack's method [11], Sauvola and Pietikäinen's method [12], Zhao et al.'s method [13], Gatos et al.'s method [14], Chen et al.'s method [15], and Moghaddam and Cheriet's method [16], and achieves better performance in precision, accuracy, and F-measure.

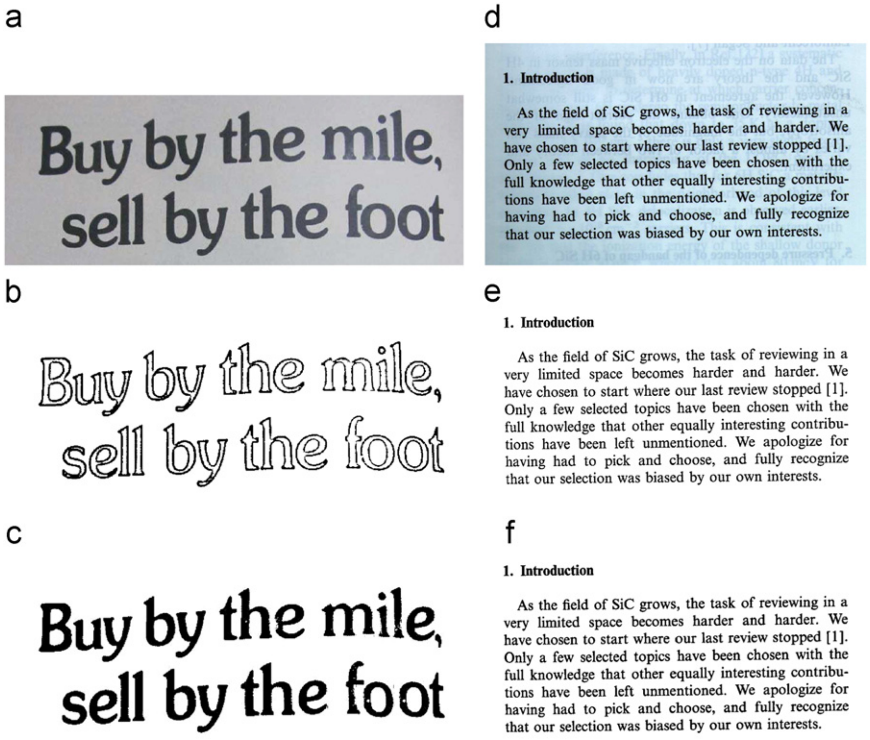
2. Challenges in adaptive binarization

The adaptive binarization scheme needs to deal with two challenges: (a) the determination of a proper window size used for extracting the local information and (b) the trade-off between detail preservation and noise suppression. Observations on the two challenges are given to 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 [11–14]. 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. For documents with large characters as shown in Fig. 1(a), using smaller than necessary window size in Sauvola and Pietikäinen's method may erroneously binarize foreground pixels to background pixels as shown in Fig. 1(b), where large characters are binarized as hollow characters. Fig. 1(c) illustrates a better binarized result of Fig. 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 Fig. 1(e) and (f), but incurs higher computational cost.

For proper window size, Gatos et al. [14] suggest that window size should cover at least one to two characters. However, detecting character size usually requires image segmentation and is difficult for degraded documents. Chen et al. [15] 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 [16] propose a scheme that 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, the proposed scheme suffers from the same problem as Gatos et al.'s method.

In addition to determining the proper window size, the trade-off between the preservation of detailed contents and noise



Buy by the mile,
sell by the foot

Buy by the mile,
sell by the foot

Buy by the mile,
sell by the foot

1. Introduction

As the field of SiC grows, the task of reviewing in a very limited space becomes harder and harder. We have chosen to start where our last review stopped [1]. Only a few selected topics have been chosen with the full knowledge that other equally interesting contributions have been left unmentioned. We apologize for having had to pick and choose, and fully recognize that our selection was biased by our own interests.

1. Introduction

As the field of SiC grows, the task of reviewing in a very limited space becomes harder and harder. We have chosen to start where our last review stopped [1]. Only a few selected topics have been chosen with the full knowledge that other equally interesting contributions have been left unmentioned. We apologize for having had to pick and choose, and fully recognize that our selection was biased by our own interests.

1. Introduction

As the field of SiC grows, the task of reviewing in a very limited space becomes harder and harder. We have chosen to start where our last review stopped [1]. Only a few selected topics have been chosen with the full knowledge that other equally interesting contributions have been left unmentioned. We apologize for having had to pick and choose, and fully recognize that our selection was biased by our own interests.

Fig. 1. Window size effect on large-scale and small-scale characters when using Sauvola and Pietikäinen's method. (a) Original image, (b) binarized image using 9 × 9 window, (c) binarized image using 33 × 33 window, (d) original image, (e) binarized image using 9 × 9 window and (f) binarized image using 33 × 33 window.

suppressing should be addressed in an adaptive binarization scheme. To discuss this issue, we briefly survey how the previous two adaptive methods, Niblack's method and Sauvola and Pietikäinen's method, determine the threshold of each pixel under a specific window size. Denote by $f(x,y)$, $0 \le f(x,y) \le 1$, the normalized intensity value of the pixel at position (x,y). 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

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

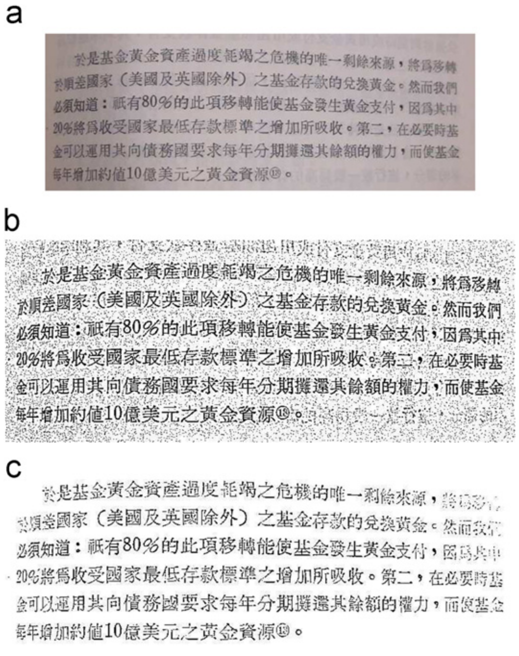
$$\sigma_w(f,x,y) = \sqrt{\frac{1}{w^2} \sum_{i=-r}^{r} \sum_{j=-r}^{r} (f(x+i,y+j) - \mu_w(f,x,y))^2}. \tag{3}$$

To improve Niblack's method, Sauvola and Pietikäinen [12] 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 [12].

Parameters k and k' used in Eqs. (1) and (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 Fig. 2(a)–(c) are binarized images obtained by Sauvola and Pietikäinen'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.



於是基金黃金資產過度耗竭之危機的唯一剩餘來源，將爲移轉
於順差國家（美國及英國除外）之基金存款的兌換黃金。然而我們
必須知道：祇有80%的此項移轉能使基金發生黃金支付，因爲其中
20%將爲收受國家最低存款標準之增加所吸收。第二，在必要時基
金可以運用其向債務國要求每年分期攤還其餘額的權力，而使基金
每年增加約值10億美元之黃金資源⑱。

Fig. 2. Effect of parameter k' in adaptive thresholding of Sauvola and Pietikäinen's method. (a) Original image, (b) binarized image with $k' = 0.01$ and (c) binarized image with $k' = 0.2$.

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 incorporate the local mean gray and gradient values.

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 from a window size which is smaller than necessary. 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 smaller than necessary window size may erroneously binarize foreground pixels to background pixels and using larger than necessary window size increases the computational cost without increasing the quality, we start with a small window size and keep increasing the window size until no substantial increase in the local variation of gray levels 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 over the rough foreground pixels 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

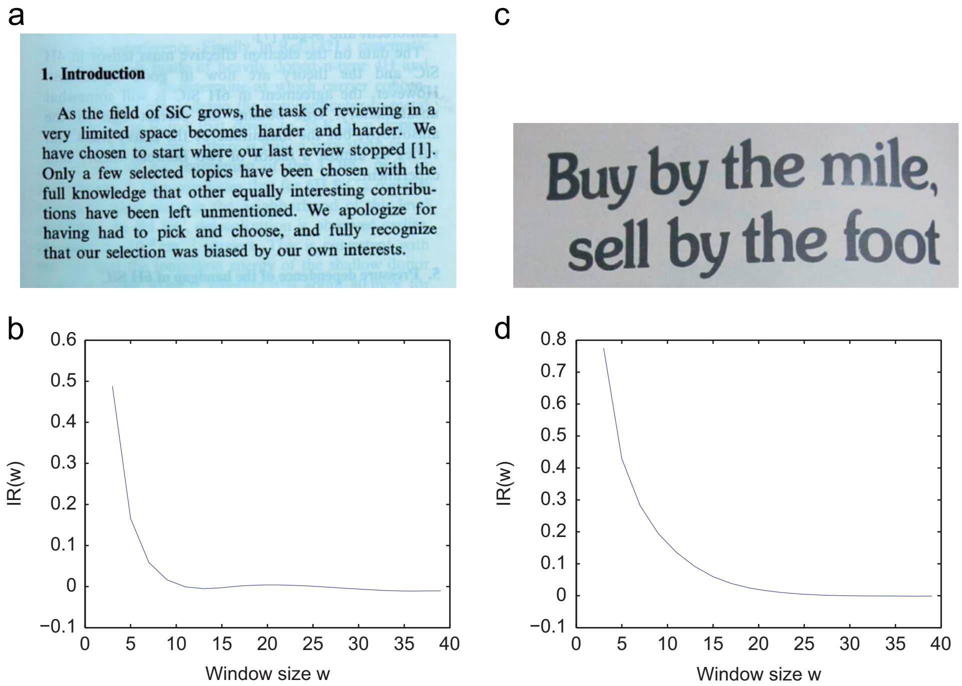
$$\overline{\sigma}_w = \frac{1}{|RFG|} \sum_{(x,y) \in RFG} \sigma_w(f,x,y), \tag{6}$$

where $|RFG|$ is the cardinality of the rough foreground set *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 Fig. 3(b) and (d). The proper window size w^* is the smallest window size such that $IR(w)$ is less than or equal to 0.01; that is, $w^* = \min\{w : IR(w) \le 0.01\}$. For saving computational efforts to determine the proper window size w^*, the fast calculation scheme proposed by Viola and Jones [20] is used to calculate the standard deviation for each window in constant time.

3.2. *Proposed noise-suppressing thresholding scheme*

When determining the binarization threshold for each pixel in degraded documents, information about the mean gray level $\mu_{w^*}(f,x,y)$ and the mean gradient $\mu_{w^*}(g,x,y)$ in the window of size $w^* \times w^*$ may be helpful, where the gradient corresponding to a pixel is determined using Sobel operator [21]. Figs. 4 and 5 show, respectively, the distributions of $\mu_{w^*}(f,x,y)$ and $\mu_{w^*}(g,x,y)$ for true foreground and background pixels of six test images.

When dividing $\mu_{w^*}(f,x,y)$ corresponding to each pixel into two groups according to the pixel status, the squared correlation



1. Introduction

As the field of SiC grows, the task of reviewing in a very limited space becomes harder and harder. We have chosen to start where our last review stopped [1]. Only a few selected topics have been chosen with the full knowledge that other equally interesting contributions have been left unmentioned. We apologize for having had to pick and choose, and fully recognize that our selection was biased by our own interests.

Buy by the mile,
sell by the foot

Fig. 3. *RFG* and associated *IR*(w) corresponding to the original images. (a) Original image, (b) *IR*(w) associated with (a), (c) original image and (d) *IR*(w) associated with (c).

coefficient between $\mu_{w*}(f,x,y)$ and the status is the ratio of the between-group variation to the total variation of $\mu_{w*}(f,x,y)$, which is the proportion of the variation in $\mu_{w*}(f,x,y)$ that can be explained by the grouping variable status. By symmetry, the squared correlation coefficient is the proportion of the variation in the pixel status that can be explained by $\mu_{w*}(f,x,y)$, indicating that $\mu_{w*}(f,x,y)$ may be useful in predicting the pixel status and is called the separability factor SP [22] associated with $\mu_{w*}(f,x,y)$ in separating the pixel status. Larger values of the separability factor SP indicate greater predicting capability. Table 1 lists the separability factor SP associated with $\mu_{w*}(f,x,y)$, $\mu_{w*}(g,x,y)$, $\sigma_{w*}(f,x,y)$, and $\sigma_{w*}(g,x,y)$, respectively, showing that the local gradients may even contain more information for predicting the pixel status than the mean gray values. Since significant amount of the variation in the pixel status can be accounted for by both $\mu_{w*}(f,x,y)$ and $\mu_{w*}(g,x,y)$ and a parsimonious model is often preferred, we present a new thresholding scheme for predicting the status of a pixel, by incorporating the information on the mean gray and gradient values in the neighborhood.

The magnitude of the gradient corresponding to the pixel located at (x,y) can be expressed as $g(x,y)=\sqrt{g_x^2(x,y)+g_y^2(x,y)}$, where

$$g_x(x,y)=\begin{bmatrix} f(x+1,y-1)-f(x-1,y-1) \\ f(x+1,y)-f(x-1,y) \\ f(x+1,y+1)-f(x-1,y+1) \end{bmatrix}\cdot\begin{bmatrix}1\\2\\1\end{bmatrix}, \tag{7}$$

$$g_y(x,y)=\begin{bmatrix} f(x-1,y+1)-f(x-1,y-1) \\ f(x,y+1)-f(x,y-1) \\ f(x+1,y+1)-f(x+1,y-1) \end{bmatrix}\cdot\begin{bmatrix}1\\2\\1\end{bmatrix} \tag{8}$$

with $\cdot$ denoting the dot product of vectors. Large values of $g(x,y)$ indicate that pixel $f(x,y)$ is around the boundary between foreground and background pixels. Based on the window size $w^*=2r^*+1$ determined in stage 1, compute the mean gradient

$$\mu_{w*}(g,x,y)=\frac{1}{w^{*2}}\sum_{i=-r^*}^{r^*}\sum_{j=-r^*}^{r^*}g(x+i,y+j). \tag{9}$$

When incorporating the information contained in the mean gray level $\mu_{w*}(f,x,y)$ and the mean gradient $\mu_{w*}(g,x,y)$ of the neighborhood around pixel (x,y), we propose a binarization threshold

$$T(x,y)=\mu_{w*}(f,x,y)(1-k''e^{-\mu_{w*}(g,x,y)/M}), \tag{10}$$

where $M=\max_{(x,y)\in D}\{\mu_{w*}(g,x,y)\}$ with D denoting the input document is used for normalizing. Eq. (10) indicates that pixels with lower values of $\mu_{w*}(f,x,y)$ or higher values of $\mu_{w*}(g,x,y)$ tend to be predicted as foreground pixels.

Denote by $|FG|(k'')$ the number of predicted foreground pixels using threshold T with parameter k'' in Eq. (10). Let

$$R(k'')=\frac{|FG|(k''-\Delta)-|FG|(k'')}{|FG|(k'')} \tag{11}$$

represent the relative increasing rate of $|FG|(k'')$ when decreasing k'' where decrement $\Delta=0.001$ is used in the experiments. Decreasing parameter k'' increases the threshold $T(x,y)$ and the number of pixels predicted as foreground pixels increases. Starting with $k''_0=0.3$, iteratively increase the threshold according to $k''_{i+1}=k''_i-0.001$. Fig. 6 plots the corresponding relative increasing rate $R(k'')$ with respect to k''. When decreasing parameter k'' from some large initial value, the relative increasing rate $R(k'')$ first decreases since the cumulative number of predicted foreground pixels increases slowly. The relative increasing rate $R(k'')$ starts increasing at some point where the slope of the intensity density of all pixels is large, leading to a surge on the number of the predicted foreground pixels.

The relative increasing rate $R(k'')$ starts decreasing at some point before the mode of the intensity of all pixels where the slope of the intensity density is low.

Parameters k''_1 and k''_2 are determined such that $R(k'')$ achieves the minimum and maximum, respectively. Threshold using

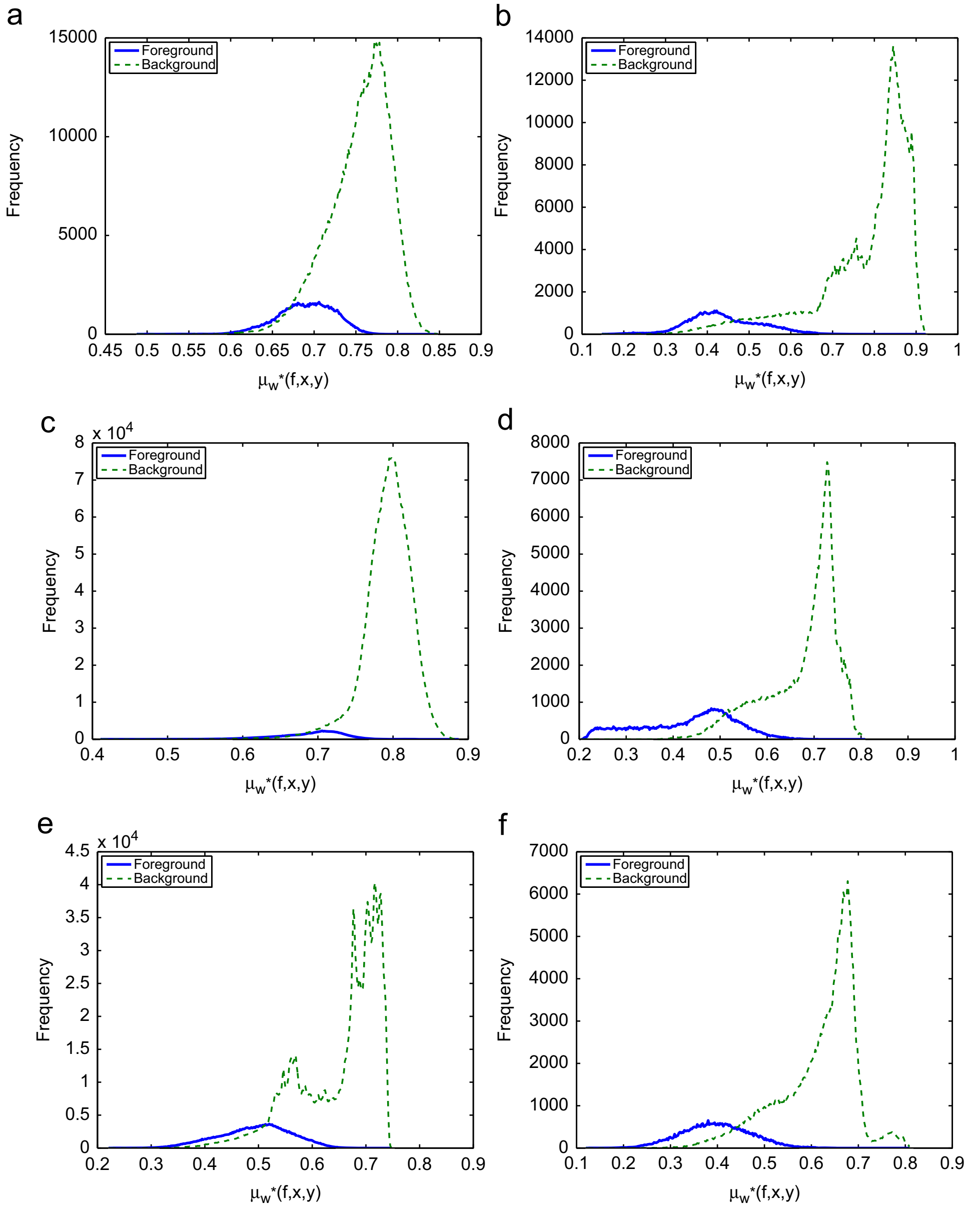


Fig. 4. $\mu_{w*}(f,x,y)$ distributions of true foreground and background pixels. (a) Test image 1, (b) test image 2, (c) test image 3, (d) test image 4, (e) test image 5 and (f) test image 6.

parameter k_1'' corresponds to the case where almost true foreground pixels are predicted as foreground pixels, whereas parameter k_2'' corresponds to the case where some of the true background pixels with lower intensities are predicted as foreground pixels. Two thresholds $T_1(x,y)$ and $T_2(x,y)$ for pixel $f(x,y)$ are determined as

$$\begin{cases} T_1(x,y) = \mu_{w*}(f,x,y)(1-k_1'' e^{-\mu_{w*}(g,x,y)/M}), \\ T_2(x,y) = \mu_{w*}(f,x,y)(1-k_2'' e^{-\mu_{w*}(g,x,y)/M}). \end{cases}$$

Let B_i denote the binarized image produced by the threshold $T_i(x,y)$, $i=1,2$. Thus, almost all the true foreground pixels are predicted as foreground pixels in B_1 and, since noisy pixels usually have lower intensities compared to those of background pixels, noisy pixels are predicted as foreground pixels in B_2. Two binarized images B_1 and B_2 are then contrasted to deliver the final binarized image B. Since $T_1(x,y) < T_2(x,y)$, if $f(x,y)$ is a background pixel in B_2, then $f(x,y)$ 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 predicted as a background pixel in B. Similarly, if $f(x,y)$ is a foreground pixel in B_1, then $f(x,y)$ 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 predicted as a foreground pixel in B. If $f(x,y)$ 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 process, based on the pixels which are predicted as foreground pixels in both B_1 and B_2, is proposed. For each pixel $f(x,y)$ predicted 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 predicted as a background in B_1 and a foreground in B_2, then it is predicted as a foreground pixel in the final binarized image B. Consequently, if a pixel which is predicted as a background in B_1 and a foreground in B_2 is not covered by any 3×3 window will be predicted as a

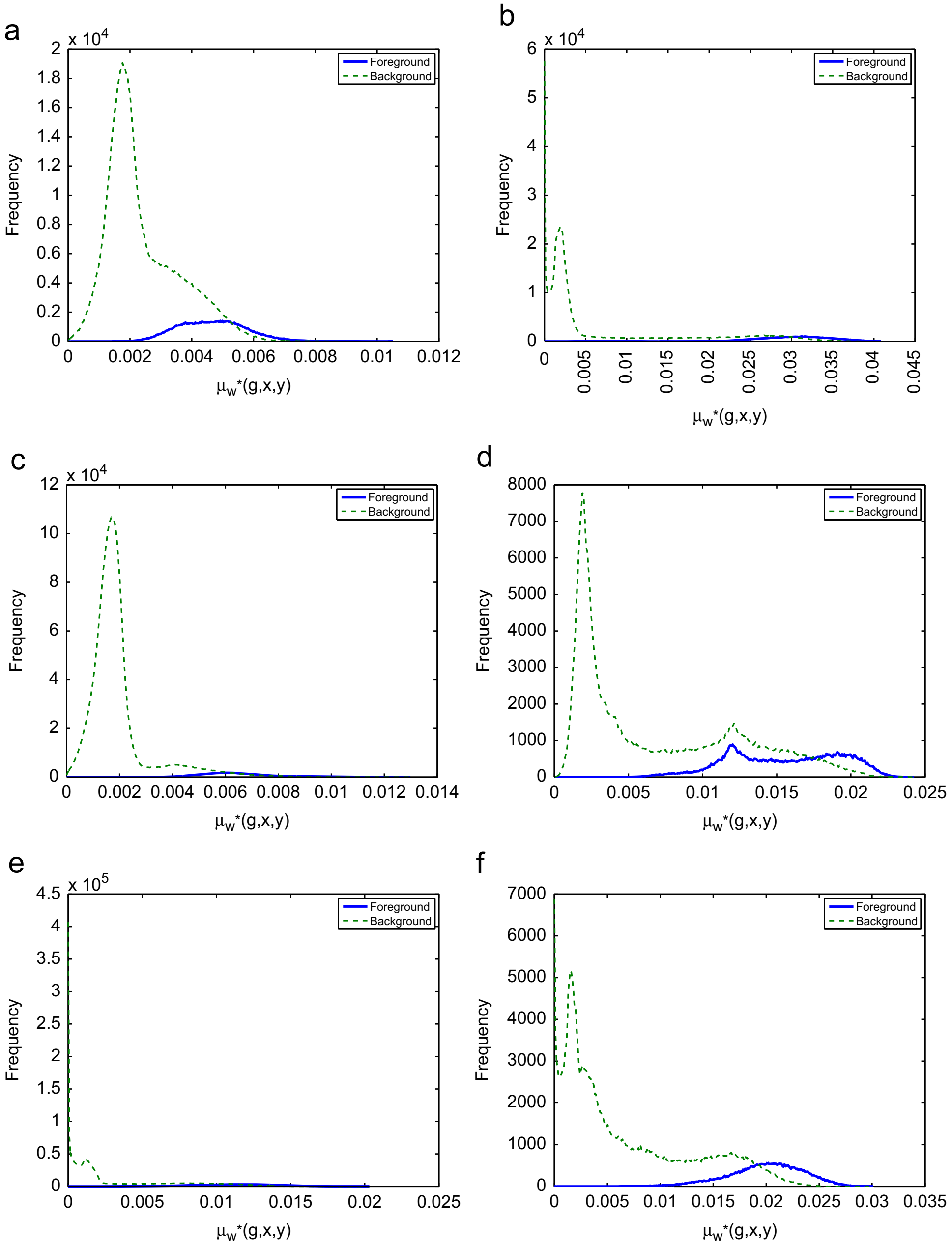


Fig. 5. $\mu_{w*}(g,x,y)$ distributions of true foreground and background pixels. (a) Test image 1, (b) test image 2, (c) test image 3, (d) test image 4, (e) test image 5 and (f) test image 6.

Table 1
Separability factor SP of test images.

Images	1	2	3	4	5	6
$\mu_{w*}(f,x,y)$	0.2396	0.2573	0.3400	0.4569	0.5887	0.4814
$\mu_{w*}(g,x,y)$	0.2737	0.3485	0.4103	0.4103	0.3145	0.4036
$\sigma_{w*}(f,x,y)$	0.1715	0.2564	0.3165	0.3406	0.2617	0.3312
$\sigma_{w*}(g,x,y)$	0.0919	0.1650	0.1649	0.1807	0.2104	0.1608

background in the final binarized image B. Furthermore, the region-growing process will be applied to the newly predicted foreground pixel by the region-growing process. This proposed region-growing process can recover some true foreground pixels and suppress the noise.

When dealing with the images with bleed-through degradation where the verso-side text or graphics appear on the recto side, the proposed region-growing process tends to predict the noisy pixels as the foreground pixels. To alleviate such problems, a refining process is used to further polish the final binarized image. We first determine if binarized image B needs further refinement by calculating the separability of the foreground pixels in B

$$s=\max_{0<i<1}\frac{\sigma_B^2(i)}{\sigma_T^2},$$

where $\sigma_B^2(i)$ denotes the variation between groups separated at intensity $i/255$ and σ_T^2 the total variation of the foreground intensities in B. If the separability s is greater than some threshold, say 0.7, the pixels in the binarized image B are further separated using Otsu's method. Before using Otsu's method, replace the intensities of the background pixels in B by their average intensity to enhance the refining effectiveness.

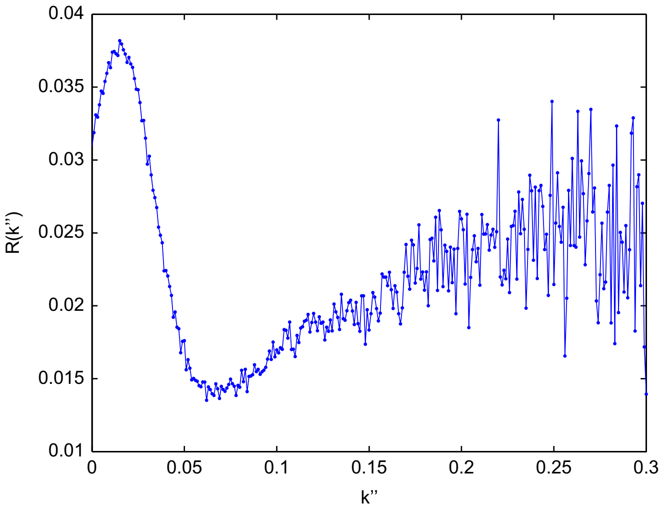


Fig. 6. $R(k'')$ for document in Fig. 8(a).

The proposed two-stage parameter-free window-based method for binarizing the input document image can be summarized as follows:

Step 1. Obtain from the smoothed image, by Otsu's method, the rough foreground pixel set *RFG*.

Step 2. Starting with $w=3$, compute $IR(w)$. If $IR(w) \leq 0.01$, let $w^* = w$ be the proper window size and go to Step 3; otherwise, let $w=w+2$ and recompute $IR(w)$ to check if condition $IR(w) \leq 0.01$ is satisfied.

Step 3. Determine $T_1(x,y)$ and $T_2(x,y)$ and produce binarized images B_1 and B_2.

Step 4. Based on B_2, the foreground pixels in B_1 will be expanded to the binarized image B by applying the region-growing process. Perform the refining process of B if the separability s of the foreground pixels in B is greater than 0.7.

4. Experimental results

In this section, we empirically compare the proposed method with the six existing methods—Niblack's method [11], Sauvola and Pietikäinen's method [12], Zhao et al.'s method [13], Gatos et al.'s method [14], Chen et al.'s method [15], and Moghaddam and Cheriet's method [16]. These six methods have been chosen for comparison because some are specifically designed for blueprint images and others are implemented based on adaptive thresholding scheme or edge detection.

Test images in Fig. 7 include blueprint images, scanned machine-printed images, and handwritten document images whose ground truth binary images are created by human eyes or provided by DIBCO'09 dataset [23]. Test images 1 and 2 in Fig. 7 are blueprint images of architectures with different proportion $\#FG/N$ of foreground pixels, where $\#FG$ is the number of true foreground pixels in the document with N pixels. Test image 3 is a document images with non-uniform background. Test images 4–6 are the bleed-through document images. Test image 7 is a textual document with ink stains and test images 8–10 are the handwritten document images.

The performance evaluations are based on five measures: (a) recall, (b) specificity, (c) precision, (d) accuracy, and (e) F-measure. Recall is the proportion of correctly binarized foreground pixels within the true foreground pixels. Specificity is the proportion of correctly binarized background pixels within the true background pixels. Precision is the proportion of true foreground pixels within the binarized foreground pixels. Accuracy is the weighted average of recall and specificity with weights proportional to the numbers of true foreground and background pixels. The F-measure is the harmonic mean of recall 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, the formulations of the measures can be expressed as

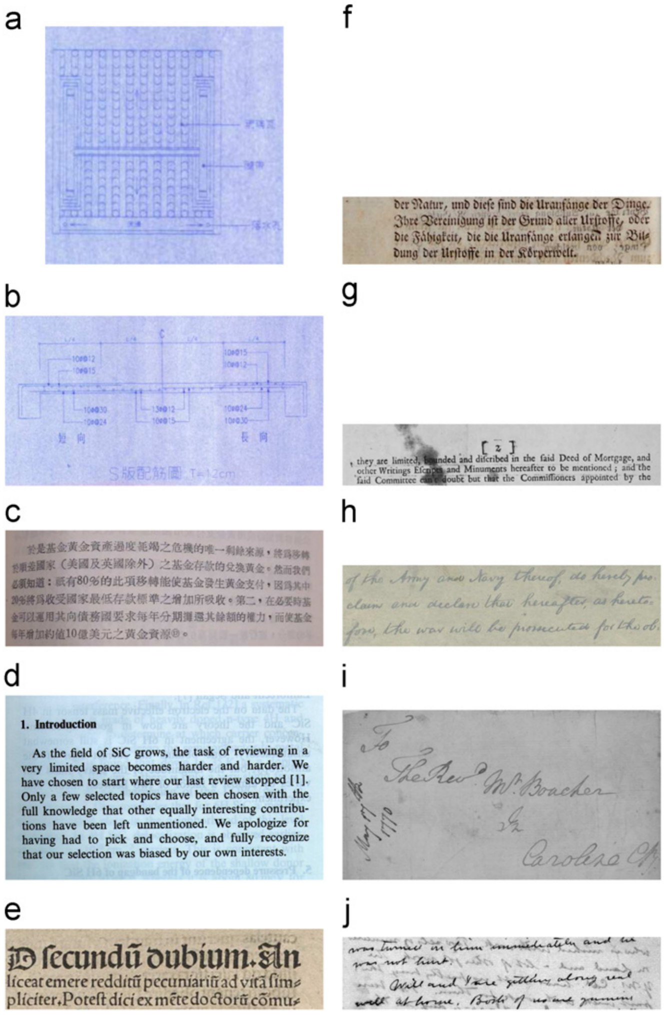

a
b
c
d
e
f
g
h
i
j

Fig. 7. Ten test images with different proportion ($\#FG/N$) of true foreground pixels. (a) Test image 1 with $\#FG/N=11.03\%$, (b) test image 2 with $\#FG/N=3.74\%$, (c) test image 3 with $\#FG/N=13.12\%$, (d) test image 4 with $\#FG/N=10.74\%$, (e) test image 5 with $\#FG/N=20.75\%$, (f) test image 6 with $\#FG/N=14.63\%$, (g) test image 7 with $\#FG/N=10.46\%$, (h) test image 8 with $\#FG/N=10.69\%$, (i) test image 9 with $\#FG/N=4.51\%$ and (j) test image 10 with $\#FG/N=5.78\%$.

$$\text{recall} = \frac{TP}{TP+FN},$$

$$\text{specificity} = \frac{TN}{TN+FP},$$

$$\text{precision} = \frac{TP}{TP+FP},$$

$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN},$$

$$\text{F-measure} = \frac{2}{\dfrac{1}{\text{recall}} + \dfrac{1}{\text{precision}}}.$$

For most images the proportion of foreground pixels is much smaller than that of the background pixels, meaning that high accuracy may result from high specificity and poor recall. Thus, the measure of accuracy is only used for reference. F-measure, the harmonic mean of recall and precision, is a more appropriate performance measure when binarizing images. The F-measure is high only when both recall and precision are high since the harmonic mean of two proportions tends to be low if one of the two proportions is low. When binarizing images, increasing threshold increases the recall but decreases the precision. Good thresholding scheme should obtain appropriate trade-off between the recall and the precision to achieve high values of the F-measure.

Empirical results are tabulated in Tables 2–11, accompanied by the visual comparisons, corresponding to 10 test images and Table 12 shows the average performance comparison over ten test images. For the average performance shown in Table 12, the proposed method has good performance in both recall and precision measures, leading to a higher F-measure. The binarized test images showed that the proposed method also has a better performance in terms of subjective evaluation. For blueprint images in Figs. 8 and 9, the proposed method is competitive, in

Table 2
The performance comparison of test image 1.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.5761	0.9372	0.6886	0.9098	0.6274
[12]	0.6969	0.9577	0.7841	0.9386	0.7379
[13]	0.7487	0.9865	0.5734	0.9290	0.6495
[14]	0.7567	0.9740	0.6518	0.9385	0.7003
[15]	0.5892	0.9245	0.8733	0.9189	0.7036
[16]	0.6267	0.9384	0.8348	0.9269	0.7160
Proposed	0.7402	0.9612	0.8918	0.9535	0.8090

Table 3
The performance comparison of test image 2.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.2628	0.9133	0.7955	0.9089	0.3951
[12]	0.6602	0.9832	0.8415	0.9779	0.7399
[13]	0.7424	0.9904	0.7088	0.9799	0.7252
[14]	0.5321	0.9766	0.6863	0.9657	0.5995
[15]	0.6866	0.9840	0.9018	0.9809	0.7709
[16]	0.5418	0.9709	0.8865	0.9677	0.6725
Proposed	0.7976	0.9914	0.8774	0.9871	0.8356

Table 4
The performance comparison of test image 3.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.8171	0.9765	0.6957	0.9396	0.7516
[12]	0.9650	0.9960	0.7224	0.9601	0.8263
[13]	0.8333	0.9807	0.6395	0.9359	0.7236
[14]	0.9494	0.9938	0.7720	0.9647	0.8516
[15]	0.8555	0.9778	0.8702	0.9637	0.8628
[16]	0.8427	0.9796	0.7222	0.9459	0.7778
Proposed	0.8958	0.9845	0.8848	0.9714	0.8903

Table 5
The performance comparison of test image 4.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.7003	0.9533	0.9068	0.9483	0.7903
[12]	0.9705	0.9965	0.9448	0.9910	0.9575
[13]	0.8882	0.9877	0.8141	0.9690	0.8495
[14]	0.8483	0.9787	0.9924	0.9801	0.9147
[15]	0.8261	0.9762	0.9423	0.9725	0.8804
[16]	0.8434	0.9781	0.9821	0.9785	0.9075
Proposed	0.9492	0.9939	0.9428	0.9884	0.9460

Table 6
The performance comparison of test image 5.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.7256	0.9114	0.8943	0.9079	0.8012
[12]	0.9529	0.9858	0.9432	0.9785	0.9480
[13]	0.7573	0.9450	0.6550	0.8848	0.7025
[14]	0.9476	0.9860	0.9647	0.9816	0.9561
[15]	0.9289	0.9804	0.9787	0.9800	0.9531
[16]	0.8592	0.9587	0.9610	0.9592	0.9072
Proposed	0.9754	0.9939	0.9274	0.9801	0.9508

Table 7
The performance comparison of test image 6.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.7926	0.9665	0.7468	0.9344	0.7690
[12]	0.8627	0.9758	0.8886	0.9630	0.8755
[13]	0.8435	0.9800	0.6296	0.9287	0.7210
[14]	0.8510	0.9719	0.9374	0.9668	0.8921
[15]	0.8875	0.9817	0.8401	0.9610	0.8632
[16]	0.7309	0.9410	0.9345	0.9401	0.8203
Proposed	0.9471	0.9918	0.8537	0.9716	0.8980

Table 8
The performance comparison of test image 7.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.7429	0.9667	0.8242	0.9518	0.7814
[12]	0.9026	0.9882	0.9385	0.9830	0.9202
[13]	0.7921	0.9761	0.7787	0.9555	0.7853
[14]	0.9604	0.9958	0.8683	0.9825	0.9120
[15]	0.7838	0.9686	0.9756	0.9693	0.8692
[16]	0.8228	0.9761	0.9511	0.9735	0.8823
Proposed	0.9303	0.9920	0.9094	0.9834	0.9197

Table 9
The performance comparison of test image 8.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.7518	0.9699	0.7620	0.9477	0.7569
[12]	0.9770	0.9980	0.7046	0.9667	0.8188
[13]	0.8717	0.9921	0.4460	0.9338	0.5900
[14]	0.9892	0.9990	0.7558	0.9730	0.8569
[15]	0.8795	0.9842	0.9635	0.9820	0.9196
[16]	0.9140	0.9916	0.7453	0.9653	0.8211
Proposed	0.9320	0.9922	0.8950	0.9818	0.9131

Table 10
The performance comparison of test image 9.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.4795	0.9662	0.6600	0.9524	0.5555
[12]	0.9802	0.9994	0.6038	0.9816	0.7473
[13]	0.5300	0.9707	0.6988	0.9585	0.6028
[14]	0.9644	0.9988	0.7020	0.9854	0.8125
[15]	0.6568	0.9755	0.9920	0.9763	0.7903
[16]	0.8464	0.9922	0.9139	0.9886	0.8789
Proposed	0.8979	0.9951	0.9135	0.9914	0.9056

terms of precision, with the method proposed by Zhao et al. which is specifically designed for blueprint images but has much higher recall, leading to much higher values of the F-measure.

Table 11
The performance comparison of test image 10.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.4948	0.9414	0.9360	0.9411	0.6474
[12]	0.8732	0.9914	0.9644	0.9899	0.9165
[13]	0.4394	0.9532	0.5981	0.9327	0.5066
[14]	0.8421	0.9889	0.9690	0.9877	0.9011
[15]	0.7434	0.9791	0.9889	0.9796	0.8487
[16]	0.6463	0.9669	0.9868	0.9680	0.7810
Proposed	0.9070	0.9941	0.9346	0.9907	0.9206

Table 12
The average performance comparison of test images 1–10.

Method	Precision	Specificity	Recall	Accuracy	F-measure
[11]	0.6343	0.9502	0.7910	0.9342	0.6876
[12]	0.8841	0.9872	0.8336	0.9730	0.8488
[13]	0.7447	0.9763	0.6542	0.9408	0.6856
[14]	0.8764	0.9888	0.8129	0.9722	0.8363
[15]	0.7837	0.9732	0.9326	0.9684	0.8462
[16]	0.7674	0.9693	0.8918	0.9614	0.8165
Proposed	0.8973	0.9830	0.9030	0.9799	0.8989

For test image 3 with non-uniform background in Fig. 10, the methods proposed by Sauvola and Pietikäinen and Gatos et al. have higher precision but lower recall and the proposed method outperforms both methods in terms of the F-measure.

For test images 4–6 with different degrees of bleed-through in Figs. 11–13, the proposed method is competitive, in terms of F-measure, with the methods proposed by Sauvola and Pietikäinen, Zhao et al., and Chen et al. and has good visual quality.

For test image 7 with ink stains in Fig. 14, the proposed method has competitive precision and good recall, leading to higher values of the F-measure when compared to Sauvola and Pietikäinen's and Gatos et al.'s methods.

For handwritten images 8–10 with different degrees of bleed-through in Figs. 15–17, the proposed method is competitive, in terms of the F-measure, with the methods proposed by Sauvola and Pietikäinen and Gatos et al. and has good visual quality.

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

1. The proposed method has significantly higher F-measure than the existing methods, indicating that the proposed method achieves better performance in both recall and precision simultaneously.
2. In terms of the specificity, the proposed method is better than most of the existing methods. The reason is because the thresholds $T_1(x,y)$ or $T_2(x,y)$ may be conservatively low, leading to the fact that most true background pixels are correctly identified.

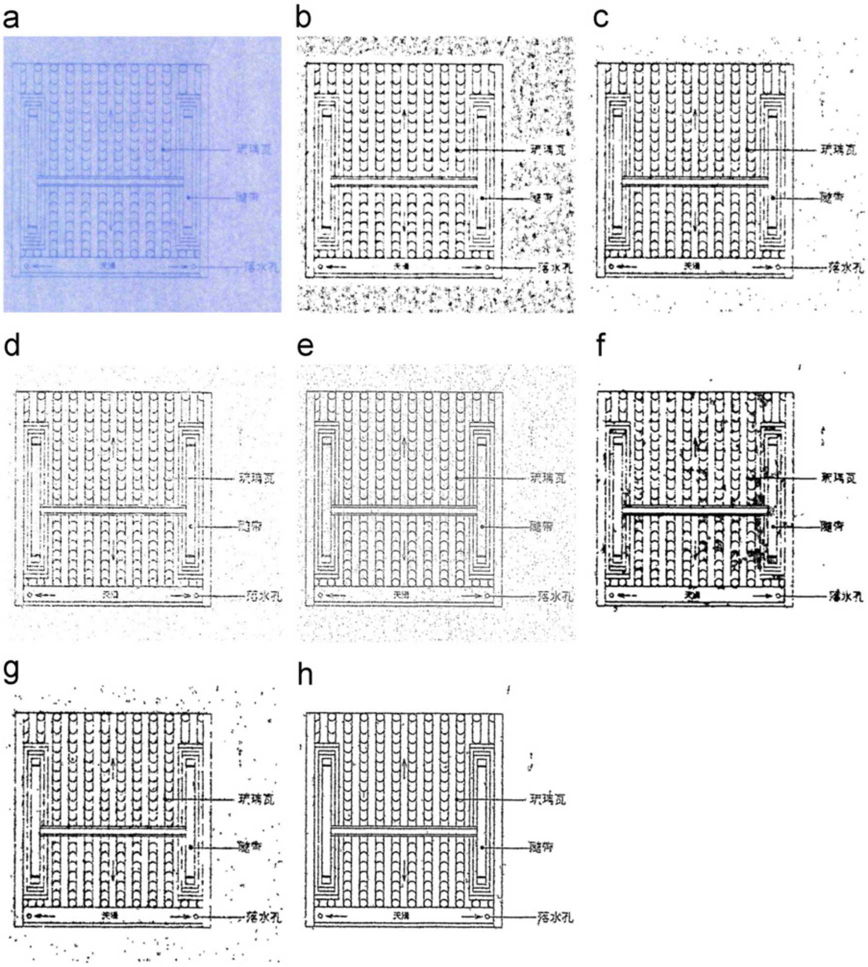


Fig. 8. Test image 1 and corresponding binarized images with different binarization methods. (a) Test image 2, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) Proposed method.

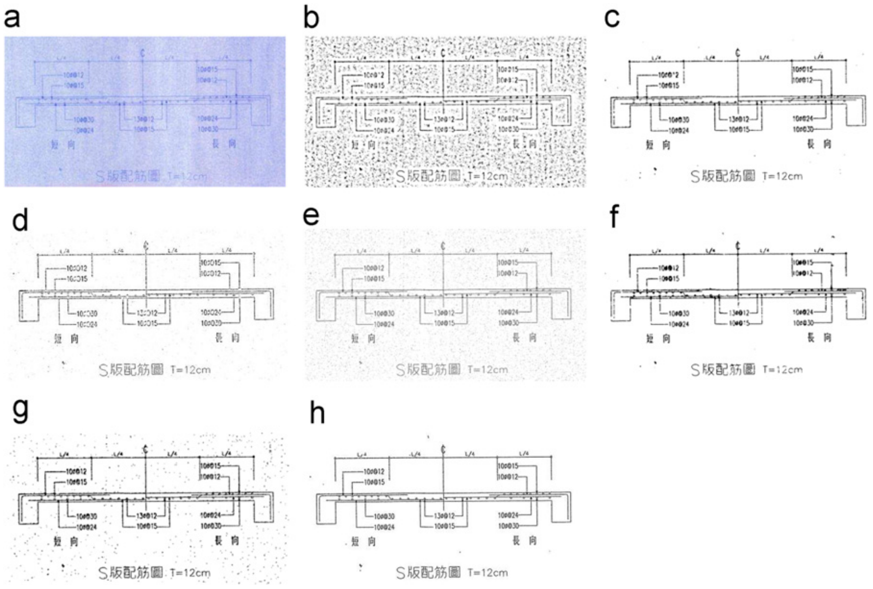


Fig. 9. Test image 2 and corresponding binarized images with different binarization methods. (a) Test image 2, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

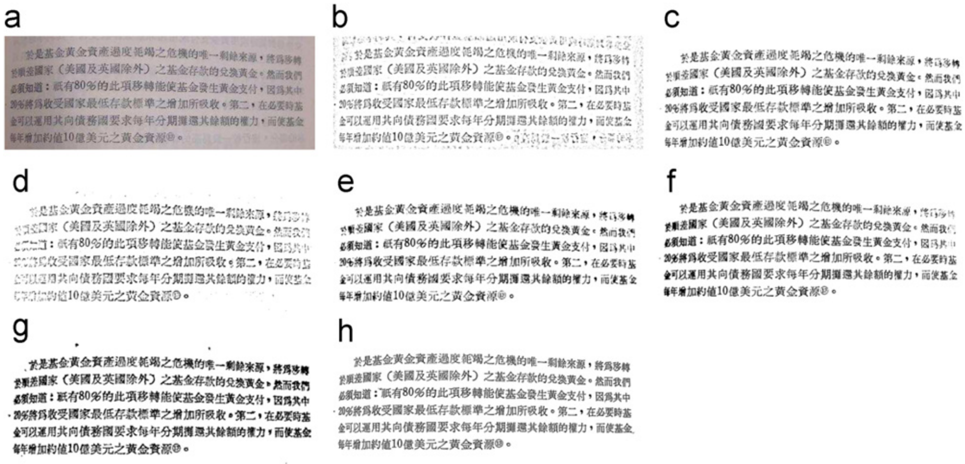


Fig. 10. Test image 3 and corresponding binarized images with different binarization methods. (a) Test image 3, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

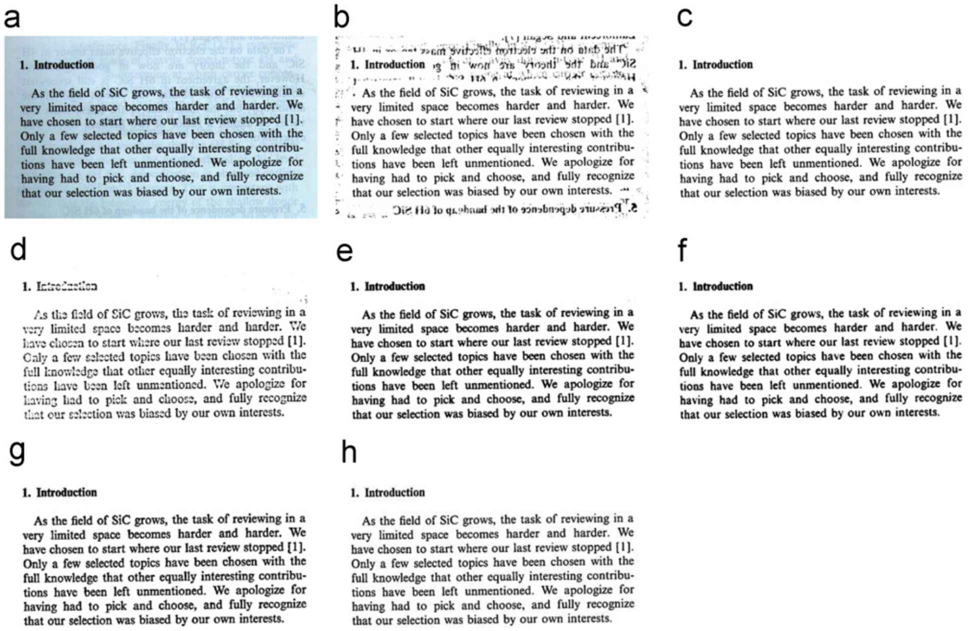


Fig. 11. Test image 4 and corresponding binarized images with different binarization methods. (a) Test image 4, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

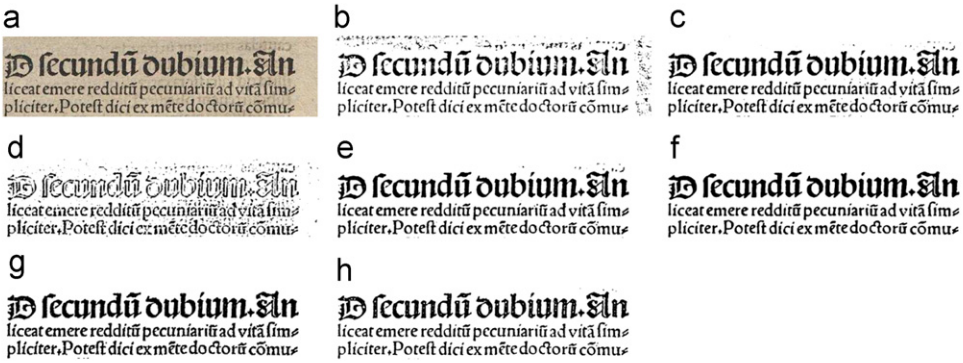


Fig. 12. Test image 5 and corresponding binarized images with different binarization methods. (a) Test image 5, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

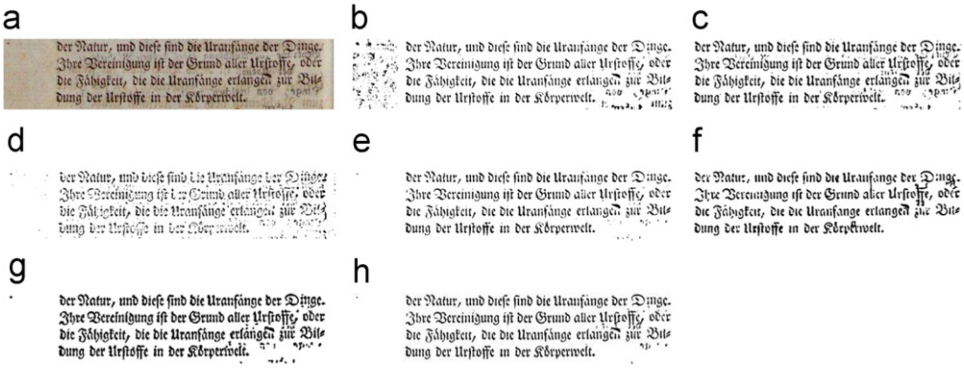


Fig. 13. Test image 6 and corresponding binarized images with different binarization methods. (a) Test image 6, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

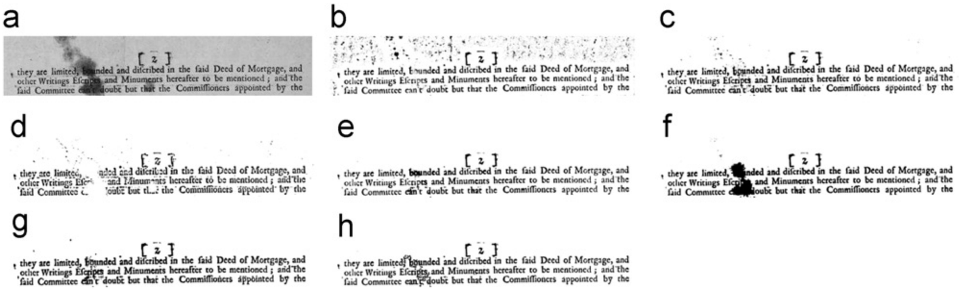


Fig. 14. Test image 7 and corresponding binarized images with different binarization methods. (a) Test image 7, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

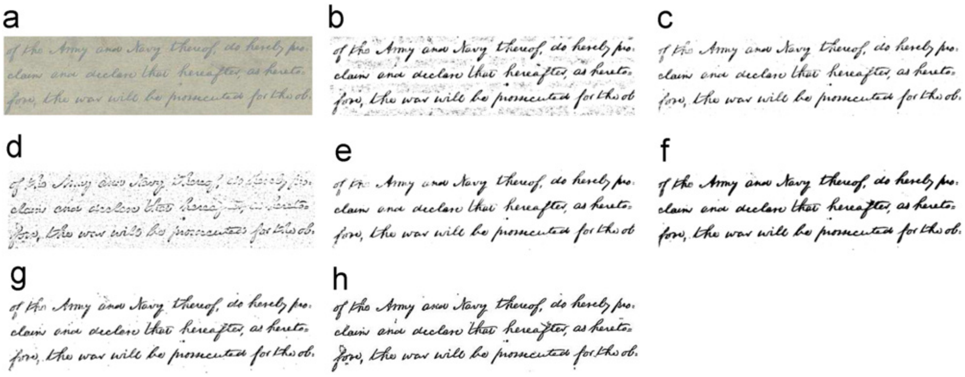


Fig. 15. Test image 8 and corresponding binarized images with different binarization methods. (a) Test image 8, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

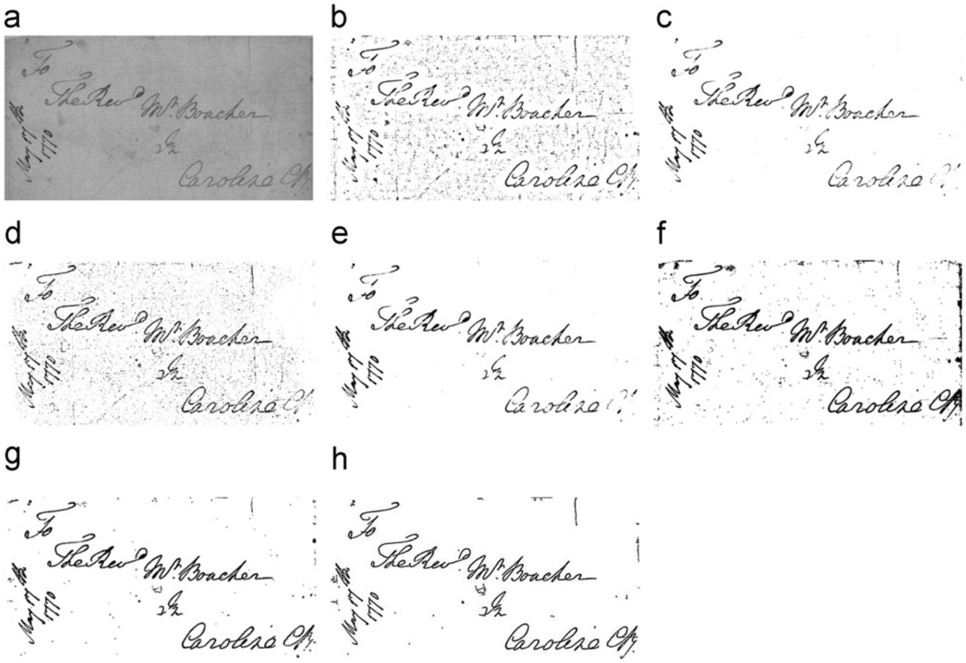


Fig. 16. Test image 9 and corresponding binarized images with different binarization methods. (a) Test image 9, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

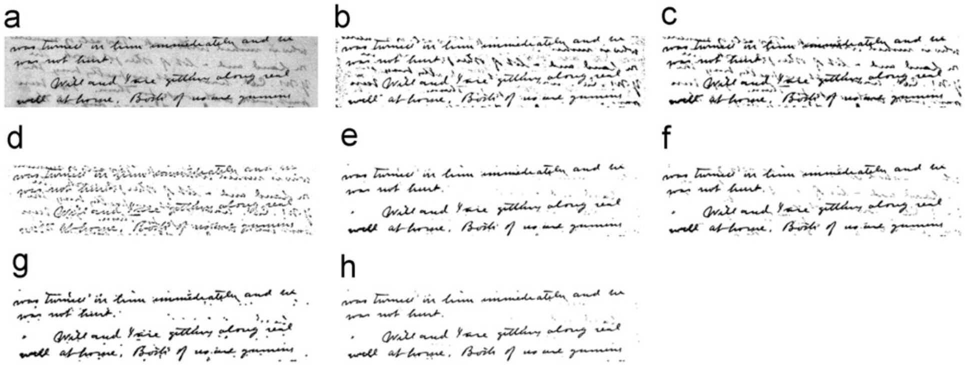


Fig. 17. Test image 10 and corresponding binarized images with different binarization methods. (a) Test image 10, (b) Niblack's method, (c) Sauvola and Pietikäinen's method, (d) Zhao et al.'s method, (e) Gatos et al.'s method, (f) Chen et al.'s method, (g) Moghaddam and Cheriet's method and (h) proposed method.

3. In terms of the accuracy, the proposed method is competitive with the existing methods. Since accuracy is the weighted average of recall and specificity, small recall of the proposed method is accompanied by large specificity and may lead to higher values of the F-measure.
4. For highly degraded documents such as blueprint images 1 and 2, results in Tables 1 and 2 show that the proposed method achieves higher precision with acceptable recall when compared to the existing methods.
5. In general, the proposed method has competitive recall and the good precision, specificity, accuracy, and F-measure among all concerned methods.

In addition to the numerical performance measures, the visual comparisons are also shown in Figs. 8–17. In the case of highly degraded documents, our method suppresses the noisy background pixels and preserves more texts and lines since the proposed thresholding scheme incorporates the information of the mean gray and gradient values which are highly correlated with the pixel status.

5. Conclusion

In this paper, we propose a two-stage parameter-free window-based binarization method. In the first stage, a proper window size is automatically determined based on the local variation of pixel intensities. Based on the determined window size, in the second stage, noise is suppressed by contrasting two binarized images produced using two thresholding schemes which incorporate information on local mean gray and gradient values. Furthermore, a refining process is proposed to improve the binarization quality, especially for bleed-through degraded images. In general, the proposed binarization scheme is competitive when compared to the existing methods. Specifically, the proposed method generates visually acceptable binarized images and has good performance in both recall and precision measures, resulting in high values of the F-measure.

Acknowledgments

K.-L. Chung was supported by the National Science Council of ROC under Contract NSC98-2923-E-011-001-MY3.

W.-N. Yang was supported by the National Science Council of ROC under Contract NSC100-2218-E-011-006.

Y.-H. Huang was supported by the National Science Council of ROC under Contract NSC99-2221-E-228-006.

References

[1] F. Chang, S.-Y. Chu, C.-Y. Chen, Chinese document layout analysis using adaptive regrouping strategy, Pattern Recognition 38 (2005) 261–271.
[2] C.-C. Wu, C.-H. Chou, F. Chang, A machine-learning approach for analyzing document layout structures with two reading orders, Pattern Recognition 41 (2008) 3200–3213.
[3] C.-L. Tsai, H.-F. Chiang, K.-C. Fan, C.-D. Chung, Reversible data hiding and lossless reconstruction of binary images using pair-wise logical computation mechanism, Pattern Recognition 38 (2005) 1993–2006.
[4] H.-F. Jiang, C.-C. Han, K.-C. Fan, A fast approach to the detection and correction of skew documents, Pattern Recognition 18 (1997) 675–686.
[5] C.-H. Chou, S.-Y. Chu, F. Chang, Estimation of skew angles for scanned documents based on piecewise covering by parallelograms, Pattern Recognition 40 (2007) 355–443.
[6] K.-C. Fan, W.-H. Wu, A run-length-coding-based approach to stroke extraction of chinese characters, Pattern Recognition 33 (2000) 1881–1895.
[7] K.-C. Fan, L.-S. Wang, Classification of document blocks using density features and connectivity histogram, Pattern Recognition 16 (1995) 955–962.
[8] A.-B. Wang, K.-C. Fan, Optical recognition of handwritten chinese characters by hierarchical radical matching method, Pattern Recognition 34 (2001) 15–35.
[9] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man and Cybernetics 9 (1979) 62–66.
[10] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision, Graphics, and Image Processing 29 (1985) 273–285.
[11] W. Niblack, An Introduction to Digital Image Processing, Prentice Hall, Englewood Cliffs, NJ, 1986.
[12] J. Sauvola, M. Pietikäinen, Adaptive document image binarization, Pattern Recognition 33 (2000) 225–236.
[13] M. Zhao, Y. Yang, H. Yan, An adaptive thresholding method for binarization of blueprint images, Pattern Recognition Letter 21 (2000) 927–943.
[14] B. Gatos, I. Pratikakis, S.J. Perantonis, Adaptive degraded document image binarization, Pattern Recognition 39 (2006) 317–327.
[15] Q. Chen, Q.S. Sun, P.A. Heng, D.S. Xia, A double-threshold image binarization method based on edge detector, Pattern Recognition 41 (2008) 1254–1267.
[16] R.F. Moghaddam, M. Cheriet, A multi-scale framework for adaptive binarization of degraded document images, Pattern Recognition 43 (2010) 2186–2198.
[17] I.K. Kim, D.W. Jung, R.H. Park, Document image binarization based on topographic analysis using a water flow model, Pattern Recognition 35 (2002) 265–277.
[18] H.H. Oh, K.T. Lim, S.I. Chien, An improved binarization algorithm based on a water flow model for document image with inhomogeneous backgrounds, Pattern Recognition 38 (2005) 2612–2625.
[19] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (1986) 679–698.
[20] P. Viola, M.J. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2004) 137–154.
[21] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison Wesley, 1992.
[22] H.H. Oh, K.T. Lim, S.I. Chien, An improved binarization algorithm based on a water flow model for document image with inhomogeneous backgrounds, Pattern Recognition 38 (2005) 2612–2625.
[23] B. Gatos, K. Ntirogiannis, I. Pratikakis, ICDAR2009 Document Image Binarization Contest (DIBCO2009), in: ICDAR09, 2009.

Yung-Hsiang Chiu received the B.S. degree in Computer Science and Information Engineering from Tamkang University, Danshui, Taipei, Taiwan, and he is now a Ph.D. student in Computer Science and Information Engineering of National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include document image processing and video coding.

Kuo-Liang Chung received his B.S., M.S., and Ph.D. degrees in Computer Science and Information Engineering from National Taiwan University in 1982, 1984, and 1990, respectively. In August 1990, he jointed the Department of Computer Science and Information Engineering of National Taiwan University of Science and Technology, Taiwan, as an Associate Professor. He was promoted to Professor and then Chair Professor in 1995 and 2009, respectively. During 2003–2006, he served as the head of the Department of Computer Science and Information Engineering of the National Taiwan University of Science and Technology. He has published numerous articles in well-known international journals. His current research interests include image processing, video coding, and data hiding. During 1996–1998, he served as the executive editor of Journal of the Chinese Institute of Engineers. In 2000, he served as co-chair of the 13th IPPR Conference on Computer Vision, Graphics, and Image Processing held at Taipei. He was a recipient of the Distinguished Engineering Professor Award from Chinese Institute of Engineers in 2001. In 2004, He received the Distinguished Research Award from the National Science Council and in 2007, received the best paper award from the Society of Computer Vision, Graphics, and Image Processing in Taiwan. He is a senior member of IEEE and a fellow of IET.

Wei-Ning Yang receives his Ph.D. degree in Industrial and Systems Engineering from The Ohio State University, Columbus Ohio, U.S.A., and is now an associate professor in the Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include statistical analysis, stochastic simulation, and image processing.

Yong-Huai Huang received the B.S. degree in Information Management from Aletheia University, Danshui, Taipei, Taiwan, and the M.S. and Ph.D. degrees in Computer Science and Information Engineering from the National Taiwan University of Science and Technology, Taipei, Taiwan. He is now an assistant professor in the Institute of Computer and Communication Engineering at Jinwen University of Science and Technology, Hsin-Tieny, Taipei, Taiwan. His research interests include image processing and compression, and algorithms.

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.