Pattern Recognition Letters 32 (2011) 1805–1813

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows

Bilal Bataineh *, Siti Norul Huda Sheikh Abdullah, Khairuddin Omar

Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 3 January 2011
Available online 22 August 2011
Communicated by G. Borgefors

Keywords:
Document image binarization
Local binarization
OCR
Text
Thresholding methods

ABSTRACT

Binary image representation is essential format for document analysis. In general, different available binarization techniques are implemented for different types of binarization problems. The majority of binarization techniques are complex and are compounded from filters and existing operations. However, the few simple thresholding methods available cannot be applied to many binarization problems. In this paper, we propose a local binarization method based on a simple, novel thresholding method with dynamic and flexible windows. The proposed method is tested on selected samples called the DIBCO 2009 benchmark dataset using specialized evaluation techniques for binarization processes. To evaluate the performance of our proposed method, we compared it with the Niblack, Sauvola and NICK methods. The results of the experiments show that the proposed method adapts well to all types of binarization challenges, can deal with higher numbers of binarization problems and boosts the overall performance of the binarization.

1. Introduction

Binary image representation is the preferred format for image analysis in most textual image techniques (Stathis et al., 2008), and the performance of the subsequent steps in document analysis applications depends on the accuracy of the binarization process. The purpose of the binarization of document images is to extract the text from the images, remove the noise and reduce the size of the images in memory. This is done by removing non-useful information to increase the visibility of the useful information in the image. Kefali et al. (2010) have claimed that the binarization process aims to decrease the presence of unwanted data and preserve only the desired data in the document images. This can be done by converting all levels in the image into only two levels, black text and white background.

According to reviews presented in previous evaluation studies, there are many different methods and techniques used for document image binarization. In some evaluation studies, such as Sezgin and Sankur (2004), as many as 40 methods have been used. We will review the best methods found in the evaluation studies.

Stathis et al. (2008) proposed a technique to evaluate binarization methods that is better than previously used techniques such as segmentation or OCR testers. The experiments were conducted on 30 binarization algorithms, representing local, global and hybrid approaches. In this study, two datasets were modified to generate an average-intensity document image dataset and a max-intensity document image dataset. The results show that Sauvola's method achieved the highest ranks with the max-intensity document image dataset, whereas the Johansen algorithm achieved the highest ranks with the average intensity document image dataset. This study offers a clear view of the different categories of binarization methods, the most well-known methods in each category, the performance of these methods and the techniques used to evaluate their performance.

Kefali et al. (2010) evaluated 12 well known binarization methods to find the best one for historical Arabic manuscript images. This study focused on simple methods to determine the value of the thresholding methods and ignored more complex algorithms. The selected global methods are the fixed global thresholding, Otsu's method, the ISODATA method, Kapur's method, the Cheng-Chen method and the Li-Lee method, whereas the selected local methods are Bernsen's method, Niblack's method, Sauvola's method, Wolf's method, fuzzy hierarchical segmentation and the NICK method. In (Kefali et al., 2010) study, an experiment was conducted on 120 historical document images containing different types of binarization challenges. The results showed that the best performance was achieved by the NICK method. The advantage of Kefali et al. (2010) study is that it provides a summary and explanation of the most well-known simple methods that will be used in our study.

In continuation, Bataineh et al. (2011a) compared four binarization methods on degraded Arabic Calligraphy document images

* Corresponding author.
E-mail addresses: bilal.bataineh82@gmail.com (B. Bataineh), mimi@ftsm.ukm.my (S.N.H.S. Abdullah), ko@ftsm.ukm.my (K. Omar).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.08.001

namely Niblack's method, Sauvola's method, NICK method and a proposed method based on adaptive thresholding and fixed window size. Those methods were used in preprocessing stage of Arabic Calligraphy font recognition system and the proposed binarization method outperformed comparing to other methods. Later, Bataineh et al. (2011b) extended and tested their work on benchmark DIBCO 2009 datasets to solve the problem of the low contrast images and thin pen stroke problems. Again, their proposed method obtained better performance followed by NICK's, Niblack's and Sauvola's methods subsequently. Unfortunately, the drawback of the proposed method as well NICK's, Niblack's and Sauvola's methods is that they requires prior window size tuning.

Sezgin and Sankur (2004) surveyed and compared 40 binarization algorithms based on a set of objective segmentation quality metrics. They used a dataset divided into two subsets, one containing neutral texture images such as X-ray, ultrasonic images, and thermal images, and the other containing document images. They compared the performance of the algorithms based on the average misclassification error, edge mismatch, relative foreground area error, modified Hausdorff distance, and region non-uniformity. The results show that the local characteristic techniques are given the highest rank on the document dataset. This study provided an alternative classification of binarization techniques based on related information.

Based on the above literature review, binarization process for documents images can be applied based on two approaches: global and local approaches (Kefali et al., 2010; Stathis et al., 2008). The global approach determines one thresholding value for the entire document image. The disadvantage of this approach is that it is not effective on documents with varying levels of noise and complex structure, errors in scanning, poor quality and multiple illumination levels (Gatos et al., 2006). To solve these problems, the local approach determines the thresholding values of each of the windows in the image independently. Both the global and local approaches may consist of simple or complex methods. The simple methods (Khurshid et al., 2010; Sauvola et al., 1997) are used to determine the thresholding value where the pixels will be separated into black or white. They are easy to implement and provide high performance in most cases, but they fail in some cases and require the values of some factors to be determined manually. The compound methods consist of several steps, some based on document features and recognition stages (Chou et al., 2010; Lelore and Bouchara, 2009), others based on adding preparatory steps to an existing simple thresholding method (Tong et al., 2009; Zhang and Yang, 2010), and still others on filters, segmentation, text extraction, pre-processing and post-processing and several of the methods mentioned above (Armanfard et al., 2009; Cecotti and Belad, 2008). The complex methods are able to solve some of the more complex problems; however, they depend on the performance of other techniques, and they are often costly and complex to design.

Logically, simple methods will be better when using pre-processing and post-processing, as in compound methods. In this study, we focus on simple methods that depend only on the values of pixels in the image to determine the thresholding values. Based on the above analysis, the most widely used methods are Niblack's, Sauvola's and the NICK method.

Niblack's method was proposed by Niblack (1985) The thresholding values of each window over the image are calculated separately by the following formula:

$$T = m + k\sigma, \tag{1}$$

where the window size may be determined by the user; for example, the window size is defined as 15×15, m is the mean value and σ is the standard deviation value of the pixels inside the window (Sezgin and Sankur, 2004). The value of k is -0.2 by Niblack (1985). Niblack's method is one of the most commonly used methods, and it is the basic principle behind several recent methods. It can detect the text body very well no matter how low the image contrast. However, its disadvantage is that it produces a large amount of black noise in the empty windows. It also requires the values of the factors k and the window size to be determined manually.

Sauvola's method was developed from Niblack's method by Sauvola et al. (1997). It aims to solve the problem of black noise depending on the impact on the standard deviation value by using a range of gray-level values in the images. The thresholding formula is

$$T = m\left(1 - k\left(1 - \frac{\sigma}{R}\right)\right), \tag{2}$$

where k is a factor in the range [0.2, 0.5] and R is a gray-level range value; $k = 0.2$ and $R = 125$ (Sauvola et al., 1997). This method solves the problem of black noise. However, it fails if the contrast between the foreground and background is small or if the text is in thin pen stroke text. It also requires the values of the factors k and the window size to be determined manually.

The NICK method is an improvement on Niblack's method developed by Khurshid et al. (2010). It aims to solve the problem of black noise in Niblack's method and the low contrast problem in Sauvola's method by shifting the thresholding value downward. The thresholding formula is as below,

$$T = m + k\sqrt{\frac{\sum p_i^2 - m^2}{N}}, \tag{3}$$

where k is a factor in the range $[-0.2, -0.1]$, P_i is the pixel value of the gray-scale levels in the image, and N is the total number of pixels in the image. In (Khurshid et al., 2010) study the window size is 19×19, and $k = -0.1$. This method solves the black-noise and low-contrast problems (Khurshid et al., 2010). However, it still fails when the contrast is too small or the text is in thin pen stroke text. It also does not solve the problem of determining the factors manually.

In general, the accuracy of the binarization depends on two factors. The first factor is the general condition of the document image. The second factor is the efficiency of the method used to deal with this condition. The condition of the document image is a visual description, and some of its properties may pose a challenge to the binarization process. Ntogas and Veintzas (2008) listed some other challenges, including dirty and dark spots, poor quality in aging manuscripts, low contrast between the text and background, multi-color and multi-size scripts, ink seeping from other documents, and thin pen strokes. Old and historical documents are good examples of images that present these challenges. Usually, some methods give good performance on particular types of document images but fail with other types. Most related studies claimed that many binarization methods have been developed; however, there are still some challenges in the field of document binarization, especially in special cases such as historical manuscripts (Stathis et al., 2008; Kefali et al., 2010; Ntogas and Veintzas, 2008).

The objective of this study is to propose an enhanced local binarization method for document images. Based on a novel simple thresholding method with a novel dynamic and flexible window partition method, we have developed a binarization method for document images. This method consists of two steps: first, partitioning the document images into suitable-sized windows based on their properties and then applying a new, improved thresholding method that can overcome the weaknesses of previous methods and can address more challenges than the previous methods. This method can also solve the problem of determining the values of the factors, which had to be determined manually in the previous methods. To evaluate the proposed method, we compare it with some of the most recent, effective and simple local thresholding methods. Based on the literature review, the chosen methods are Niblack's, Sauvola's and the NICK methods. The datasets used consist of selected document images for visual results and a benchmark dataset with its evaluation techniques for analytical results. This paper is organized as follows. Section 2 explains the proposed method, and Section 3 presents and analyzes the experimental results. Finally, our conclusions are presented in Section 4.

2. The proposed method

In the proposed method, we want to overcome the problems and weaknesses of the previous simple methods. Therefore, we develop a method that is generally capable of dealing with all binarization challenges with high performance and is particularly effective in solving the problems of low contrast between the foreground and background and thin pen stroke text. The proposed method adopted the local approach and it is used two elements: a dynamic method to divide the image into the windows based on the characteristics of the image and a new method that can determine the appropriate value of the thresholding of each window.

2.1. The proposed thresholding method for binarization process

The aim of the proposed method is to address all binarization challenges and to overcome the weaknesses of previous methods. In line with Niblack's, Sauvola's and NICK's inspirations, we propose a thresholding equation used the mean value and the standard deviation as well as effective factors that help to address the previous challenges of binarization. These factors are gray-level values presented by mean value, and standard deviation for each current window and global image respectively.

(a) *Window mean value:* Each image is divided into many sub-images represented by the binarization window areas. Each sub-image has special values and properties. The gray-level value of the window mean value points to approximate thresholding value of each binarization window independently.
(b) *Global mean value:* The document image is divided into several windows whereby some of these windows contain irregular values. These may cause inconsistency when performing window binarization process. Therefore, in order to tolerate these inconsistency, we calculate and use a gray-level value which presented by global mean value of the whole image as one the important factors.
(c) *Window standard deviation value:* A standard deviation indicates a measure of dispersion in a distribution data. In relation to image case, standard deviation corresponds to the contrast of the pixel values. Based on experiment, as we reduce the standard deviation value, the resulted image pixel values become converging or vise versa. For that reason we also use window standard deviation gray-level value as a control mechanism of the mean value like other research (Khurshid et al., 2010; Sauvola et al., 1997; Niblack, 1985).
(d) *Adaptive standard deviation value:* The standard deviation of the binarization windows (sub-images) are varies from one window to other window in the same image, also it is small and insignificant in bright or low-contrast images. Therefore, we adapt standard deviation values for each sub-image that lies to gray-level value in the range from 0 to maximum gray level value. This leads to an equal effect regardless of the image and thus overcomes the problem of bright or low-contrast images. The standard deviation gray-level values of all windows are normalized to the interval from 0 to maximum gray level value as in Eq. (4).

$$\sigma_{Adaptive} = \frac{\sigma_W - \sigma_{min}}{\sigma_{max} - \sigma_{max}} max_{level}, \tag{4}$$

where $\sigma_{Adaptive}$ is the gray-level value of the adaptive standard deviation for the target window, σ_W is the standard deviation of the target window, σ_{min} is the minimum standard deviation value of all windows, σ_{max} is the maximum standard deviation value of all windows and max_{level} is the maximum gray level value.

Pertaining to above best effect factors, we propose a thresholding method to determine optimal the gray-level threshold value for classifying pixel into either black or white level. We apply this proposed thresholding method on each binarization window independently. Similar to Niblack's, Sauvola's and NICK's methods, our proposed thresholding method is inspired by shifting the *Window mean value* with a fitting value. Those state of the art methods calculate this fitting value using factors initiated by image nature such as standard deviation of gray-level distribution, or predefined value (k). Unlike them, we determine that fitting value automatically based on gray-level values presented by *Window mean value*, *Global mean value*, *Window standard deviation value* and *Adaptive standard deviation value*.

Based on experiments and analysis, the fitting value is inversely proportion to standard deviation of sub-image. As the standard deviation becomes lower, that observed a higher fitting value is required to remove the noise or to overcome the possibility of converting the non-white pixels into white pixels from the background, non-textual or dirty image. In relation to that, firstly an adaptive standard deviation is proposed, $\sigma_{Adaptive}$ to relatively change this fitting value as the following:

$$\frac{1}{\sigma_{Adaptive}}. \tag{5}$$

Secondly, we also monitor that the fitting value is linearly proportion to the ratio of mean value of the sub-image over the mean value of the global image. The higher the ratio denotes higher non-white pixels transforming into white pixels or non-textual sub-image. Regardless, we maximize the appearance of textual sub-image by doubling-up the window mean value and present it as below,

$$\frac{m_W^2}{m_g}, \tag{6}$$

where m_W is the mean value of the pixels in the binarization window and m_g is the mean value of the global image pixels.

In spite of them, another factor in relation to the fitting value is the standard deviation of the sub-image or window standard deviation value, σ_W. After multiplying Eqs. (5) and (6), we observe that the optimal fitting value responds positively as σ_W is subtracted from the numerator and σ_W is added to all factors individually in the denominator as in Eq. (7)

$$\frac{m_W^2 - \sigma_W}{(m_g + \sigma_W)(\sigma_{Adaptive} + \sigma_W)}. \tag{7}$$

Finally, we propose the thresholding value for the sub-image or window namely T_W by combining all above factors (Eq. (7)) to determine the fitting value and shift it to the *Window mean value* as given as in Eq. (8)

$$T_W = m_W - \frac{m_W^2 - \sigma_W}{(m_g + \sigma_W)(\sigma_{Adaptive} + \sigma_W)}, \tag{8}$$

where T_W is the gray-level value of the thresholding value of the binarization window, m_W is the gray-level of the mean value of the pixels in the binarization window, m_g is the gray-level of the mean value of the global image pixels; $\sigma_{Adaptive}$ is the gray-level of the adaptive standard deviation of the binarization window as defined in Eq. (4), σ_W is the gray-level of the standard deviation of the window. In some rare cases the value of $\sigma_{Adaptive}$ and σ_W together are equal 0, that leads to infinity value for the fitting value. In this case, we make the thresholding value is equal the max gray-level value in the image. Based on this T_W values, the binarization process is defined in Eq. (9)

$$I_{binary}(x,y) = \begin{cases} black, & i(x,y) < T_W, \\ white, & i(x,y) \geq T_W, \end{cases} \tag{9}$$

where I_{binary} is the binary image and $i(x,y)$ is the pixel value of the input image.

2.2. Generating the binarization dynamic windows

In the previous local binarization methods, the window size was defined by users. This is not practical with all images, including those with different characteristics and sizes. The window size is a critical factor in the local binarization methods. Small windows are effective at removing noise, whereas large windows are more effective at preserving the text. However, small windows may destroy large text, whereas large windows do not remove all of the noise. This motivated us to find suitable windows for each image independently or even different windows for different regions within a single image, depending on their characteristics. For that reason, we convert the image into presentable format that shows the regions where the binarization process may face a binarization problem. Then based on the image properties and percentage of the problem regions in the image, a suitable dynamic windows are examined. These windows are defined by the following steps.

At first, the thresholding Eq. (8) is applied on global image to extract the value of the confusion threshold T_{con}. That requires some modification in order to compromise global approach as the following equation,

$$T_{con} = m_g - \frac{m_g^2 - \sigma_g}{(m_g + \sigma_g)((0.5max_{level}) + \sigma_g)}, \tag{10}$$

where T_{con} is the confusion threshold of the global image, m_g is the mean value of the image pixels, max_{level} is the max level value of the global image either σ_g is the standard deviation of the image pixels. The T_{con} value classifies each pixel into three levels, black for the foreground, white for the background or red[1] for confusion values. We define those three values as below,

$$I = \begin{cases} black & i(x,y) \leq T_{con} - \left(\frac{\sigma_g}{2}\right), \\ red, \quad T_{con} - \left(\frac{sigma_g}{2}\right) < i(x,y) < T_{con} + \left(\frac{\sigma_g}{2}\right), \\ white, & i(x,y) \geq T_{con} + \left(\frac{\sigma_g}{2}\right). \end{cases} \tag{11}$$

where *I(x, y)* is the image, $i\,(x,y)$ is the pixel value, m_g is the mean value of all pixel values and σ_g is the standard deviation of all pixel values in the image.

Then, the probability p of black pixels over red pixels is computed $\left(p = \frac{number\ of\ black\ pixels}{number\ of\ red\ pixels}\right)$. Based on the probability result, large windows relative size are determined for low contact ($\sigma_g < 0.1 * $ maximum level value) or high probability of confusing ($p > 2.5$) images. These properties refers to the multi-levels, large text size, low contact images. Whereas, a medium windows relative size are determined for the fine and normal images ($1 < p < 2.5$) or small image size (high + width are less than 400 pixel). While the degraded images get a small windows relative size, when ($p < 1$). Based on the previous presentation, the primary window size is defined as in Eq. (12)

$$PW_{size} \begin{cases} \left(\frac{I_h}{4}, \frac{I_W}{6}\right), & \geq 2.5 \text{ or } (\sigma_g < 0.1 * max_{level}), \\ \left(\frac{I_h}{20}, \frac{I_W}{30}\right), & 1 < p < 2.5 \text{ or } (I_h + I_w < 400), \\ \left(\frac{I_h}{30}, \frac{I_W}{40}\right), & p \leq 1, \end{cases} \tag{12}$$

where PW_{size} is the primary window size, I_h is the image height, I_w is the image width, p is the ratio of black pixels to red pixels, σ_g is the standard deviation of all pixel values in the image and max_{level} is the maximum gray scale value of the image. Fig. 1(b) shows the results of generating the binarization windows.

Lastly, we check each primary window for its appropriateness as presented below:

$$SW_{size} = \begin{cases} \left(\frac{PW_h}{2}, \frac{PW_w}{2}\right) & red\_pixels > black\_pixels, \\ WP_{size} & \text{otherwise}, \end{cases} \tag{13}$$

where SW_{size} is the secondary window size, PW_h, PW_w are the primary window height and width correspondingly, *red_pixels* is the total number of red pixels in the primary window, *black_pixels* is the total number of black pixels in the primary window. In line to that, if the total number of red pixels higher than the total of black pixels in the primary window, therefore the sub-image requires a smaller binarization window. Consequently, the primary windows is divided into four new quadrant or secondary windows. After this process, the final results of the generated binarization windows are produced. An illustrative example is shown in Fig. 1(c).

Algorithm 1. The proposed adaptive binarization method

Step 1. Start
Step 2. Apply thresholding confusion equation onto global image
Step 3. Classify the global image into three levels black, red or white pixels
Step 4. Determine primary window size based on probability of red over black pixels
Step 5. Check appropriateness of each primary window size. If population of red pixels higher than black pixels then create secondary window size
Step 6. Apply thresholding window equation for each adaptive window size or local image
Step 7. End

We can simplify our proposed adaptive binarization method as shown above in Algorithm 1.

3. The experiments and results

In this work, two types of experiments were conducted. The first visual experiment provides a clear idea of the images and the results of the binarization. This experiment was conducted on selected images presenting various types of binarization challenges. The second type is an analytical test that provides a statistical measurement based on a benchmark dataset and evaluation measurement. This is the dataset adopted by the Document Image Binarization Contest "DIBCO 2009" organized by the International Conference on Document Analysis and Recognition "ICDAR 2009" (Gatos et al., 2009, 2011). This dataset is available in

[1] For interpretation of colour in Figs. 1–3, the reader is referred to the web version of this article.

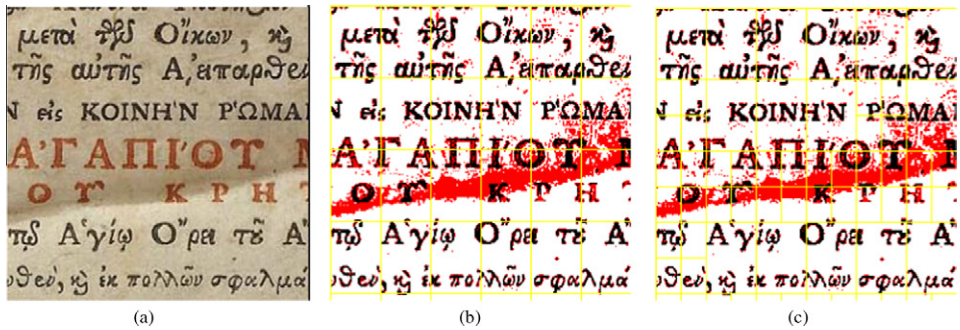


Fig. 1. (a) The original image imported from Ntogas and Veintzas (2008); (b) the confusion binarization with primary windows and (c) the confusion binarization with secondary windows.

(http://www.iit.demokritos.gr/~bgat/DIBCO2009/benchmark); it contains 10 document images presenting several types of challenges to the process of binarization. These include color and gray-scale images that are divided into five printed and five handwritten document images.

3.1. The visual experiment

This experiment was conducted to show the visual performance of our method in several challenges such as disparity in the size of text; large, non-uniform illumination, low-quality images; thin pen stroke lines; and low-contrast between the text and background. Six selected samples were used in this experiment. The selected samples were collected from several sources. Fig. 2(a), (b) and (f) are various sizes of printed text, low-quality handwritten and low-quality non-uniform illumination handwritten samples successively collected from DIBCO 2009, 2010, Fig. 2(c) and (e) are low-contrast text with thin pen strokes and non-uniform illumination samples collected imported from Ntogas and Veintzas (2008), and Fig. 2(d) is a historical Arabic manuscript that we modified to contain many challenges in a single text, such as very multi-low contrast, different text color and colored marks between the text. To evaluate the performance of the proposed method, we compared its results with those of Niblack's, Sauvola's and the NICK methods. The factor values used by the methods were identified based on the values of the authors: $k = -0.2$ and 25×25 window size like (Kefali et al., 2010) for Niblack's method, $k = 0.2$, $R = 128$ and 15×15 window size for Sauvola's method, and $k = -0.2$ and 19×19 window size for the NICK method. In the proposed method, the window size is determined automatically based on the generated binarization windows. The experimental results are shown in Fig. 3.

Based on the results in Fig. 3, the proposed method provided the best results for all of the challenges selected. As we note in the first image, Fig. 2(a), the proposed method, gave the best results with large text, as shown in Fig. 3(s), whereas we note a negative impact of large text on the other methods in Fig. 3(a), (g), and (m). In the case of the spotted and low quality image, Sauvola's method outperforms other methods where it able to eliminate dirt and noise perfectly (Fig. 3(h)). Nevertheless, our proposed method achieve better visual text results in comparison to other methods (Fig. 3(t)). Unfortunately, this advantage of Sauvola's method in this case equates to a disadvantage in thin pen stroke or low-contrast images. As shown in the case of Fig. 3(i) and (j), Sauvola's method and the other previous methods (Fig. 3(c), (d), (o), and (p)) fail to deal with these cases, whereas the proposed method gives the better performance on these images as in Fig. 3(u) and (v). Referring to Fig. 3(w) and (x), the proposed method outperforms other methods in dealing with non-uniform illumination and low-quality challenges. As a result, our proposed method can overcome almost all challenges in degraded document images.

3.2. The analytical experiment

This experiment was conducted to show the analytical performance of the proposed method on different types of binarization challenges. Therefore, the benchmark dataset (DIBCO 2009, 2010) devoted to the binarization method was used in these experiments.

Fig. 2. Example of (a) a multi-color image with different size fonts; (b) a spotted, low-quality image; (c) a thin pen stroke, low-contrast and small size; (d) a poor quality and very low-contrast image; (e) a non-uniform illumination and (f) a non-uniform illumination with thin pen stroke and low-quality image.

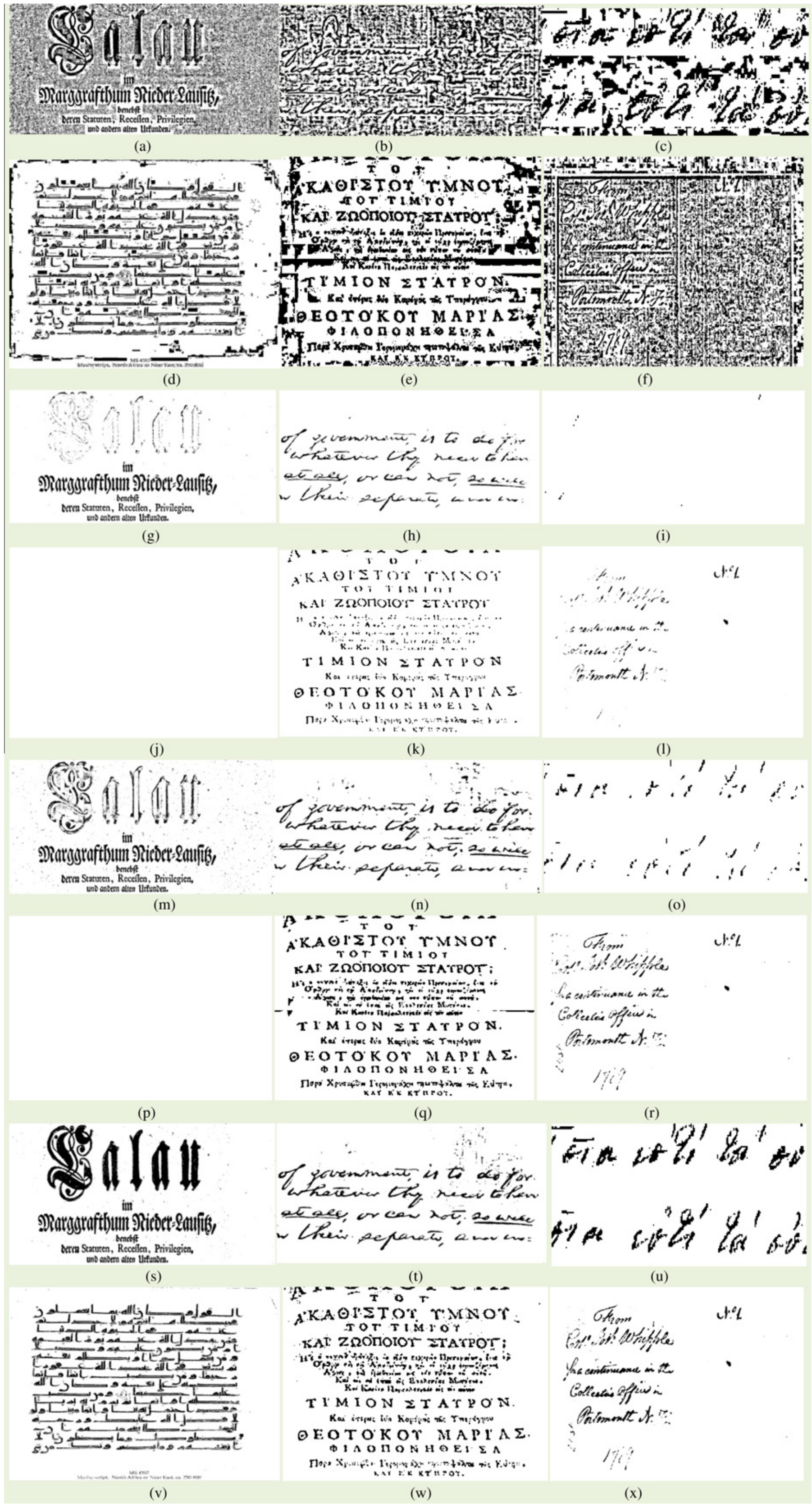
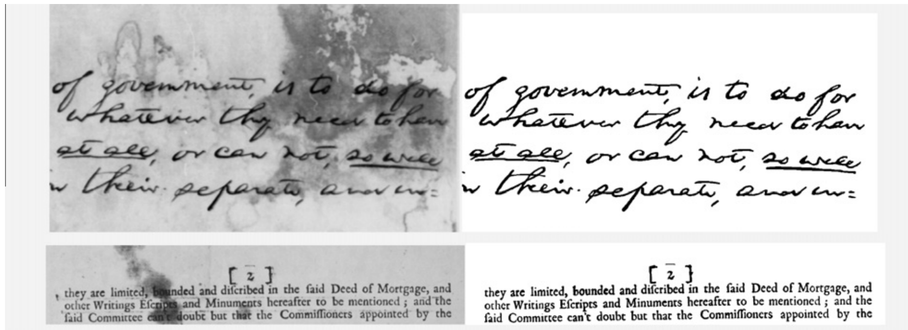


Fig. 3. Example of binarization results after Appling (a–f) Niblack, (g–l) Sauvola, (m–r) NICK, and (s–x) the proposed method correspondingly.

Many evaluation techniques have been used in previous works. Some works depend on the OCR accuracy testers (Chou et al., 2010; Khurshid et al., 2010; Sauvola et al, 1997; Cecotti and Belad, 2008), some depend on segmentation testers (Sezgin and Sankur, 2004), others are based on vision test (Ntogas and Veintzas, 2008), and still others have used spatial evolution measurements



(Left) (Right)

Fig. 4. Two examples of (DIBCO 2009, 2010) datasets (Left) and its binary ground image (Right).

Table 1
The *F*-mean, PSNR and NRM performance of the proposed, Niblack's, Sauvola's and NICK methods.

	Hand-written images			Printed text images			
	F-mean (%)	PSNR	NRM	*F*-mean (%)	PSNR	NRM	Average *F*-mean (%)
Proposed	85.1	11.79	6.65	90.93	10.5	6.16	88.002
Niblack	25.57	6.5	19.86	52.36	7.52	18.24	38.97
Sauvola	54.77	11.61	27.84	71.74	9.97	21.35	63.25
NICK	76.776	11.76	13.53	83.34	10.31	12.57	80.06

for binarization performance (Stathis et al., 2008). The OCR or segmentation testers are related to the accuracy of the techniques employed; this method cannot be applied to languages that do not include OCR or segmentation systems. The visual evolution is also a weak way to evaluate the binarization performance. However, there are existing techniques to test binarization performance. Some effective evaluation methods are proposed to evaluate the performance of binarization methods (Gatos et al., 2009, 2011; Sokolova and Lapalme, 2009). These evaluation measurements were adopted as a benchmark by DIBCO 2009 (2010). The evaluation measurements consist of the *F*-mean, PSNR and Negative rate metric (NRM). They measured the binarization performance based on the closeting to the ground binary image. The DIBCO 2009 (2010) benchmark dataset consists the binary ground image of each document image. Some examples are shown in Fig. 4.

The *F*-mean measurement denotes the percentage accuracy of the binary image

$$F\text{-mean} = \frac{2 \times recall \times \text{precision}}{recall + \text{precision}}, \tag{14}$$

where $recall = \frac{TP}{TP+FN}$, $\text{precision} = \frac{TP}{TP+FN}$, TP is the true-positive value, FP is the false-positive value, and FN is the false-negative value.

The PSNR measurement denotes how much a given image is similar to another image. Therefore, a higher value of PSNR indicates a higher similarity between the two images

$$\text{PSNR} = 10lg\left(\frac{C^2}{\text{MSE}}\right), \tag{15}$$

where

$$\text{MSE} = \frac{\sum_{x,y}^{M,N}(I(x,y) - I(x,Y))^2}{Mn} \tag{16}$$

and *c* is the max gray-scale level value of the input image, and the negative rate metric (NRM) denotes the mismatches between the resulting binary image and the original image. Therefore, a low value of NRM indicates a higher similarity between the two images

$$\text{NRM} = \frac{FN_* + FP_*}{2}, \tag{17}$$

where $FN_* = \frac{FN_n}{FN_n+TN_n}$, $FP_* = \frac{FP_n}{FP_n+TN_n}$, FN_n is the number of true positives, FP_n is the number of false positives, and TN_n is the number of false negatives.

The first analytical experiment was conducted based on the values of the factors as identified by each author, as above in the visual experiments. These experiments were conducted on the (DIBCO 2009, 2010) dataset and evaluated by the (DIBCO 2009, 2010) evaluation measurements. The proposed method was compared with the Niblack, Sauvola and NICK methods. As shown in Table 1, the results of the proposed method for both hand-written and printed text are better than those of other methods. The average *F*-mean for both types is 88.002% for the proposed method, compared to 38.97%, 63.25% and 80.06%, respectively, for Niblack's method, Sauvola's method, and the NICK method. The proposed method also achieved the best performance in terms of PSNR and NRM, followed by the NICK method. Fig. 5 shows the *F*-mean performance of the proposed, Niblack's, Sauvola's and NICK methods.

The second experiment used windows of the same size in all the methods to test the performance of each method independently of the effect of the size of the window. This provided a measure of the efficiency of the method compared to the other methods. In this experiment, 20 × 20 window size was adopted because it is the

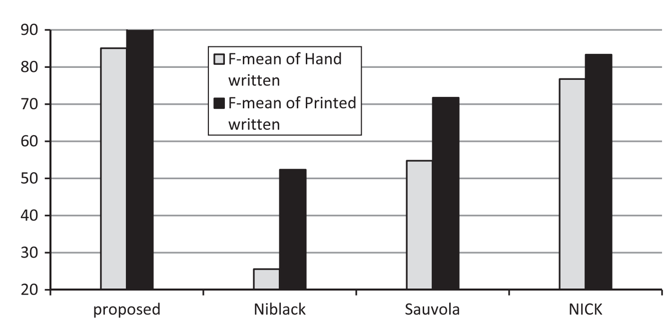


Fig. 5. The *F*-mean of the proposed, Niblack, Sauvola and NICK methods.

Table 2
The *F*-mean, PSNR and NRM performance of the proposed, Niblack's, Sauvola's and NICK methods with 20 × 20 window size.

	Hand-written images			Printed text images			
	F-mean (%)	PSNR	NRM	*F*-mean (%)	PSNR	NRM	Average *F*-mean (%)
Proposed	82.82	11.86	8.76	87.12	10.44	8.92	84.97
Niblack	26.02	6.54	19.35	53.15	7.56	17.68	39.59
Sauvola	59.17	11.67	25.5	74.63	10.1	19.4	66.91
NICK	76.77	11.76	13.53	83.34	10.31	12.57	80.1

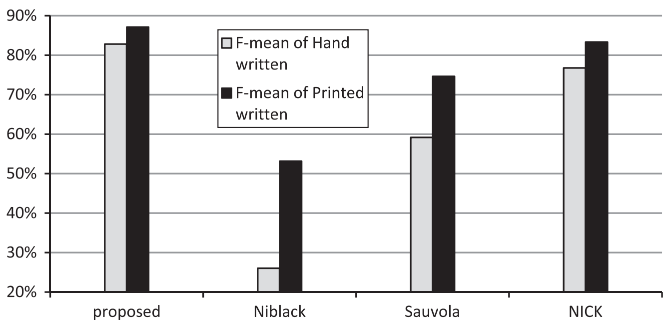


Fig. 6. The *F*-mean of the proposed, Niblack's, Sauvola's and NICK methods with 20 × 20 window size.

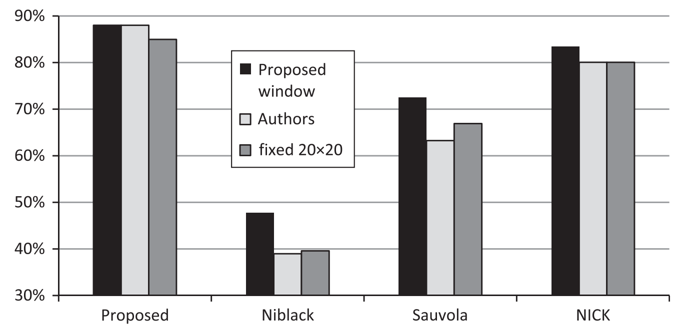


Fig. 7. The average *F*-means of the proposed, Niblack's, Sauvola's and NICK methods based on proposed adaptive window generation, authors (Khurshid et al., 2010; Sauvola et al., 1997; Niblack, 1985) and fixed 20 × 20 windows sizes for both handwritten and printed text images.

closest to the size chosen by the authors for each method. As shown in Table 2, the values of the proposed method for both hand-written and printed text images are higher than those of the other methods. The *F-mean* values for both types are 84.97% for the proposed method, compared to 39.59%, 66.91% and 80.1%, respectively, for Niblack's, Sauvola's, and the NICK methods. The proposed method also achieved the best performance in terms of PSNR and NRM, followed by the NICK method. Fig. 6 shows the *F-mean* performance of the proposed, Niblack's, Sauvola's and NICK methods.

The third experiment was conducted based on automatic generated binarization windows of the proposed methods. This experiment used dynamic windows of the same size in all the methods to test the effect of the windows of the proposed generating windows. As shown in Table 3, the results of the proposed thresholding method are higher than other methods. The *F-mean* values for both types are 88.002% for the proposed method, compared to 47.78%, 72.48% and 83.4%, for Niblack's, Sauvola's, and the NICK methods subsequently. The proposed method also achieved the best performance in terms of PSNR and NRM, followed by the NICK method. Moreover, Fig. 7 shows that the Niblack's, Sauvola's, and the NICK methods got the best performance based on the dynamic generated windows of the proposed method.

3.3. Discussion

In this study, we have compared the results obtained under two sets of conditions, based on the windows chosen and their impact on the results of methods and based on the performance of the method with a given window size. The window size is an influential factor in the local binarization approach, as can be seen in the two case studies. First, if the size of the windows is smaller than the size of the text, some parts of the text may be removed in the binarization process, as shown in Fig. 3(a), (g) and (m). The other case is non-textual windows, in which binarization noise may appear, and the volume of that noise depends on the size of the empty window and the method used. Based on our experimental results, the noise appears prominently in all empty windows in Niblack's method Fig. 3(a)–(f) and less prominently in the results of the other methods, as shown in Fig. 3(g)–(x).

The ability of the binarization methods to overcome the widest range of binarization challenges with the best results is the essence of all previous studies. Based on our literature review, the Niblack, Sauvola and NICK methods are the best of the simple thresholding methods for document image binarization. Based on our experiments, the results can be extended to include our case as follows:

- Niblack's method was able to identify the body of the text of all the images, regardless of the situation of the image or how small the values of contrast were. However, it generates a great deal of binarization noise in the empty windows. This method is strongly influenced by the size of the windows; the results are better with large windows, which have a greater likelihood of containing a portion of the text.
- Sauvola's method produced results with the least binarization noise, regardless of the size of the window or the situation of the image. However, it was only able to identify high-quality

Table 3
The *F*-mean, PSNR and NRM performance of the proposed, Niblack's, Sauvola's and NICK methods when applying proposed dynamic window size.

	Hand-written images			Printed text images			
	F-mean (%)	PSNR	NRM	*F*-mean (%)	PSNR	NRM	Average *F*-mean (%)
Proposed	85.1	11.8	6.7	90.93	10.5	6.16	88.002
Niblack	32.33	7.15	16	63.23	8.64	15.57	47.78
Sauvola	58.28	11.65	25.81	86.67	9.66	19.31	72.48
NICK	79.11	11.74	11.42	87.68	10.42	9.22	83.4

text. It failed to identify illuminated texts compared to dark texts, texts in images with low contrast or thin pen stroke text.

- The NICK method produced better results than the previous two methods in that it solved much of the problem of binarization noise that occurred with Niblack's method and was also able to identify the text in some cases where Sauvola's method failed. In general, this method has provided good results and is a worthwhile approach; however, the disadvantage of this method is that it did not exceed the performance of Niblack's method in special cases, such as images with very little contrast, variation in the thickness of the text size or thin pen stroke text with low contrast.
- The proposed method aimed to resolve the problem of binarization noise resulting from empty windows, depending on the global mean value and to solve the problem of low contrast and thin pen stroke text depending on the value of the fixed standard deviation. Depending on the previous results, the proposed method was able to determine the text in all images, like Niblack's method. At the same time, it was able to solve the problem of binarization noise almost as well as Sauvola's method. The results of the experiments showed that our method was able to solve the previous problems and special cases better than the NICK method.

The above analysis shows that the proposed method is more effective than previous methods. Whereas each of the previous methods was able to deal very well with a few challenges, the proposed method was able to deal very well with all types of challenges. The proposed method was able to solve the problems of binarization, including a thresholding method that can deal with all binarization challenges, regardless of the status of the image, and that produces a binary image containing the text with the lowest possible binarization noise. The final outcome shows that Niblack's method was the best at extracting the text, but it produced the most binarization noise. Sauvola's method was the best at avoiding binarization noise but the worst at extracting the text from the noise. The NICK method gave good results for both of these cases, but the results in some special cases were less impressive. The proposed method gave the best results for text extraction and noise minimization on all image challenges.

4. Conclusion

The objective of this work was to propose an adaptive local binarization method for document images. The proposed method is based on two steps: a dynamic method to partition the document images into binarization windows followed by the application of a novel simple thresholding method on those windows. This method aims to solve the binarization problems that were not well addressed by the previous simple thresholding methods, such as thin pen stroke lines and images with low-contrast between the text and background. It also avoids the weakness of the previous methods that require the user to determine some of the factors or the size of the windows manually. To evaluate the proposed method, we compared it with the most well-known and effective simple thresholding methods. These methods are Niblack's method, Sauvola's method, and the NICK method. The experiments were applied based on visual experiments on the selected dataset and analytical experiments based on the (DIBCO 2009, 2010) dataset. To evaluate the proposed window-generating method, we conducted experiments on each method by using the window sizes proposed by the authors. Another experiment was conducted using the same size window to neutralize the effect of the window size on the methods.

Based on these experiments, the proposed method gave the best performance compared to Niblack's method, Sauvola's method, and the NICK method. The proposed method solved the problem of identifying the factors and size windows. It is highly adapted to deal with any binarization problems, it can solve some special challenges such as thin pen stroke and low-contrast images, and it can solve more different binarization problems than any of the previous methods.

Acknowledgments

Special thanks are due to Dr. Khurram Khurshid, University Paris Descartes, France, for his assistance and cooperation. This project research was funded by the UKM-TT-03-FRGS0129-2010 grant entitled "Determining adaptive threshold for image segmentation", UKM-TT-03-FRGS0130 entitled "Automatic Background Subtraction for Image Enhancement" and UKM-GGPM-ICT-119-2010 entitled "Logo and Text Detection for moving object using vision guided".

References

Armanfard, N., Valizadeh, M., Komeili, M., Kabir, E., 2009. Document image binarization by using texture-edge descriptor. In: Proc. 14th Internat. CSI Computer Conference 2009. CSICC 2009, Tehran, Iran, pp. 134–139.
Bataineh, B., Abdullah, S.N.H.S., Omer, K., 2011. Arabic calligraphy recognition based on binarization methods and degraded images. In: Proc. Internat. Conf. Pattern Anal. Intelligent Robotics (ICPAIR 2011), Malaysia, pp. 65–71.
Bataineh, B., Abdullah, S.N.H.S., Omer, K., Faidzul, M., 2011. Adaptive thresholding methods for documents image binarization. In: Proc. 3rd Mexican Conf. on Pattern Recognition (MCPR2011), Mexico, pp. 230–239.
Cecotti, H., Belad, A., 2008. Dynamic filters selection for textual document image binarization. In: Proc. 19th Internat. Conf. on Pattern Recognition, 2008, ICPR 2008, Tampa, FL, pp. 1–4.
Chou, C., Lin, A.W., Chang, A.F., 2010. A Binarization Method with Learning-Built Rules for Document Images Produced By Cameras. Pattern Recognition. Elsevier, pp. 1518–1530.
DIBCO 2009, 2010. <http://www.iit.demokritos.gr/~bgat/DIBCO2009/benchmark>.
Gatos, B., Pratikakis, I., Perantonis, S.J., 2006. Adaptive Degraded Document Image Binarization. Pattern Recognition. Elsevier, pp. 317–327.
Gatos, B., Ntirogiannis, K., Pratikakis, I., 2009. ICDAR 2009 document image binarization contest. In: Proc. 10th Internat. Conf. on Document Analysis and Recognition, Barcelona, pp. 1375–1382.
Gatos, B., Ntirogiannis, K., Pratikakis, I., 2011. DIBCO 2009: Document image binarization contest. Int. J. Doc. Anal. Recognit. 14, 35–44.
Kefali, A., Sari, T., Sellami, M., 2010. Evaluation of several binarization techniques for old Arabic documents images. In: The First Internat. Symp. on Modeling and Implementing Complex Systems (MISC'2010), Algeria, pp. 88–99.
Khurshid, K., Siddiqi, I., Faure, C., Vincent N., 2010. Comparison of niblack inspired binarization methods for ancient documents. In: Proc. 16th International Conference on Document Recognition and Retrieval, USA.
Lelore, T., Bouchara, F., 2009. Document image binarisation using markov field model. In: Proc. 10th Internat. Conf. on Document Analysis and Recognition, 2009. ICDAR '09. Beijing, China, pp. 551–555.
Niblack, W., 1985. An introduction to digital image processing.
Ntogas, N., Veintzas, D., 2008. A binarization algorithm for historical manuscripts. In: Presented at the Proc. 12th WSEAS Internat. Conf. on Communications, Heraklion, Greece, pp. 41–51.
Sauvola, J., Seppanen, T., Haapakoski, S., Pietikainen, M., 1997. Adaptive document binarization. In: Proc. Fourth Internat. Conf. Document Analysis and Recognition (ICDAR), Ulm, Germany, pp. 147–152.
Sezgin, M., Sankur, B., 2004. Survey over image thresholding techniques and quantitative performance evaluation. J. Electron. Imaging, 146–168.
Sokolova, M., Lapalme, G., 2009. A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing & Management. Elsevier, pp. 427–437.
Stathis, P., Kavallieratou, E., Papamarkos, N., 2008. An evaluation technique for binarization algorithms. J. Univ. Comput. Sci., 3011–3030.
Tong, L., Chen, K., Zhang, Y., Fu, X., Duan, J., 2009. Document image binarization based on NFCM. In: Proc. 2nd Internat. Congress in Image and Signal Processing, 2009. CISP '09, Tianjin, China, pp. 1–5.
Zhang, C., Yang, J., 2010. Binarization of document images with complex background. In: Proc. 6th Internat. Conf. in Wireless Communications Networking and Mobile Computing (WiCOM), pp. 1–4.