

Adaptive binarization method for degraded document images based on surface contrast variation

Bilal Bataineh¹ · Siti Norul Huda Sheikh Abdullah¹ · Khairuddin Omar¹

Received: 29 January 2015 / Accepted: 17 October 2015 / Published online: 30 October 2015
© Springer-Verlag London 2015

Abstract Document binarization is an important technique in document image analysis and recognition. Generally, binarization methods are ineffective for degraded images. Several binarization methods have been proposed; however, none of them are effective for historical and degraded document images. In this paper, a new binarization method is proposed for degraded document images. The proposed method based on the variance between pixel contrast, it consists of four stages: pre-processing, geometrical feature extraction, feature selection, and post-processing. The proposed method was evaluated based on several visual and statistical experiments. The experiments were conducted using five International Document Image Binarization Contest benchmark datasets specialized for binarization testing. The results compared with five adaptive binarization methods: Niblack, Sauvola thresholding, Sauvola compound algorithm, NICK, and Bataineh. The results show that the proposed method performs better than other methods in all binarization cases.

Keywords Document image binarization · Laplacian filter · OFF center-surround cells · OCR · Degraded images · Historical images

✉ Bilal Bataineh
b.btnh@yahoo.com

Siti Norul Huda Sheikh Abdullah
mimi@ftsm.ukm.my

Khairuddin Omar
ko@ftsm.ukm.my

¹ Pattern Recognition Research Group, Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

1 Introduction

Document image binarization is an essential preparation step in document image analysis and recognition (DIAR) techniques. The aim of document image binarization is to remove unwanted information while retaining useful information by classifying the pixels into only two levels, usually black text on a white background [1–3]. Normally, the document consists of dark textual parts, represented by low intensity pixels that are placed on a light background, represented by high intensity pixels. However, some degradation effects (e.g., aging), physical effects (e.g., heat damage, spilled liquid, leaked ink from other surfaces, artistic features, and non-uniform illumination), and color effects occur randomly, causing pixel intensity values to overlap. Because most of the previous binarization methods are based mainly on pixel value [4–10], there are varying degrees of deficiencies in the outcome of the binarization, which can appear as binarization noise on the background and in missing parts of the text. The degrees of deficiencies vary according to the amount of degradation and the ability of the binarization methods to address the deficiencies [1, 3].

Many binarization methods have been proposed to address binarization challenges. Usually, document image binarization methods are categorized as global or local approaches [1, 3]. The global approach finds a single thresholding value where the pixels are split into foreground and background, while the local thresholding method finds different thresholding values for each sub-region of the image, depending on the properties of its pixels. Otsu [4] presented a global thresholding method. Due to spatial variance of pixel intensities in degraded images, a fixed single threshold is unsuitable for all regions in the image. Therefore, the local thresholding methods are

preferable because they are more adaptable and accurate [1, 3]. A well-known local thresholding method, namely the Niblack method [5], can detect the text perfectly, but it also produces a large amount of noise within its surroundings. The Sauvola method [6] is another local thresholding competitor which overcomes the problem of the surrounding noise. However, drawbacks can occur when the contrast between the foreground and background is low. The Khurshid et al. [2] method also aims to overcome both the problem of noise in the Niblack and the low contrast problem in the Sauvola thresholding methods. Regardless, it still fails to recognize the images with low contrast or text in the form of thin pen strokes. Additionally, Nick [2], Niblack [5] and Sauvola [6] also encounter difficulties with manually determining the certain parameter values and the binarization window size. Consequently, the local method, proposed by Bataineh et al. [7], has partially overcome these issues, but critical problems remain unsolved, such as failure of the binarization windows in the text body.

Several of the binarization methods apply serial integrated steps such as compound methods [8–11]. These compound methods are able to solve more complex problems. They merge a set of processes to improve the performance of binarization. Some compound methods rely on document features and recognition stages [8], while other methods are based on additional preparatory pre-processing and post-processing steps such as filters, segmentation, text strokes extraction, or an existing simple thresholding method [9–11].

Chou et al. [8] proposed a method for non-uniform illumination of document images based on decision rules that determine the suitable binarization method for each binarization window separately. This method yields better results than many of the simple thresholding methods [1–3]. However, this method does not address different types of binarization challenges. Thus, Gatos et al. [9] proposed a binarization algorithm that depended on preparatory stages. At the pre-processing stage, the image is prepared using a Wiener filter to enhance the image contrast. At the next stage, the text regions estimate and segment the foreground from the background by applying the Sauvola binarization method. In the following stage, the background surface is further estimated and the final thresholding process is performed; in this step, the resulting binary text and background are joined. Lastly, the quality of the binary result is improved in the post-processing stage. This method yields better results than several of the simple thresholding methods [1, 3]. However, this method has shown a negative response to faint characters and critical problems such as text parts removing.

Additionally, Sauvola and Pietikainen [10] proposed a binarization algorithm to address several binarization challenges in the image. This algorithm uses image surface

analysis for the selection of regions according to the document image properties. Then, the algorithm determines the local threshold value for each pixel to be computed, based on the current region. The degraded parts are filtered within each algorithm by removing the degradation problems. The results are then combined into the binarization result image from Sauvola [6] prior to obtaining the final results. Based on several evaluation studies, this method is one of the best binarization methods. However, it has not yet achieved consistent performance across all binarization problems. Howe [11] proposed a method based on energy-based segmentation methods and pixel intensity. This method is based on several steps, such as a Laplacian filter to find the local likelihood of foreground and background labels, a Canny filter to determine the likely discontinuities, and the graph cut implementation to find the minimum energy solution. This method has shown high performance in selected images of DIBCO'09, H-DIBCO'10 and DIBCO'11 benchmark datasets. However, this method is slow, introduces binarization noise problems and misses the text parts.

In general, no binarization technique is completely accurate for all, or even most, document challenges. The current state-of-the-art methods are effective for several cases, but not others. Moreover, in some binarization methods such as Niblack, Sauvola, and NICK [2, 5, 6], certain parameter values are determined manually. There is a tradeoff between each method's ability to reduce noise and maintain the text because most binarization methods depend mainly on pixel intensity values and ignore the structural properties of a document. For document images, it is notable that the variance of pixel intensity between the text and its neighboring background pixels usually changes abruptly, while between background and degraded pixels the variance of pixel intensity changes gradually. We can observe that the textual parts have sharp and smooth edges compared to their neighboring pixels for normal and degraded cases (e.g., spots, leaked ink and illumination and color changes, as shown in Fig. 1). At the same time, the level of pixel values for the text (foreground) may look similar to or higher than background regions, thus resulting in binarization noise in the background.

Based on the above arguments and various types of text and non-text pixel changing properties, an adaptive binarization method for degraded document images is proposed. This adaptive method aims to overcome the noise problems effectively while retaining the text body in all degraded images suffering from aging, physical effects, leaked ink, artistic features, and non-uniform illumination degradations.

First, during the pre-processing stage, the image is smoothed and areas of low contrast is prepared. Then, during the region extraction stage, each homogeneous region is classified as “textual or non-textual” based on pixel contrast changes and is extracted using a Laplacian

Fig. 1 Samples of degraded textual and background pixels in documents image



convolution filter. Additionally, the positions of the textual regions are determined by the OFF center-surround cells method. Next, during the selection stage, the textual regions are retained while the non-textual regions are removed using the connected-component labeling method with the outputs of the previous stage. Finally, during the post-processing stage, the quality of the final result is improved using a thresholding process to restore the missing, small textual parts and the low-constricted text and to fill in the wide text. To evaluate the proposed method, it is compared with set of the most recent and well-known methods. The evaluation experiments were conducted using the series of the International Document Image Binarization Contest (DIBCO) benchmark datasets for binarization methods [12–16]; these five datasets contain the most critical challenges for binarization.

We organize the research findings into four sections in the present paper. Section 2 introduces and explains our proposed method in detail. Section 3 presents the analyses and experimental results. Finally, we present the conclusions of our findings in Sect. 4.

2 Proposed method

In this section, the proposed binarization method is presented. It can be effectively implemented for even the worst cases of image degradation, such as the noise caused by leaked ink, spilled liquid, an inhomogeneous background, and low contrast between the foreground and background. This method consists of four stages. First, the document images are prepared in the pre-processing stage. In the second stage, the geometrical features representing the regions' edges of the rapid changes of the pixel contrast are extracted and the positions of the true text are determined. In the third stage, the regions of the true text are retained and the regions of the noise are removed. In the final stage, the text body is completely reconstructed in the

post-processing stage. A schematic of this process is shown in the flowchart in Fig. 2.

2.1 Pre-processing

The main aim of the pre-processing stage is to prepare the contrast between the pixels in the document images. In this stage, two processes are implemented: a 5×5 median filter and modified gray-level image histogram equalization.

(a) Median filter

This step provides image texture smoothing to eliminate the contrast due to such noises as page edges, ink leaks, and dirt or any other physical contamination [27]. The smoothing process is required because the following stages depend strongly on the variance of the pixel contrast in the document image texture. A 5×5 kernel median filter is used to smooth the contrast [27, 28]. The kernel size of the median filter should be equal the kernel size of the surround masks in the OFF center-surround cells image (Sect. 2.2). That to keep enough variance of the pixels contrasts between the background and foreground pixels for the geometric feature extraction stage.

(b) Adjusting the pixel contrast

The histogram equalization method adjusts the pixel contrast based on pixel histogram of the image [17]. This method represents the distribution of the histogram by increasing the contrast of the pixel levels, with the signal being strong if the contrast between the foreground and background is small. This procedure enhances the image to support the binarization performance. However, the noise appearance and its effect on the binarization increase strongly over a wide range of document cases in this method (Fig. 3b). To overcome these problems, a modified histogram

Fig. 2 Flowchart of the proposed method: **a** pre-processing stage, **b** feature extraction stage, **c** labeling and selection stage, and **d** post-processing stage

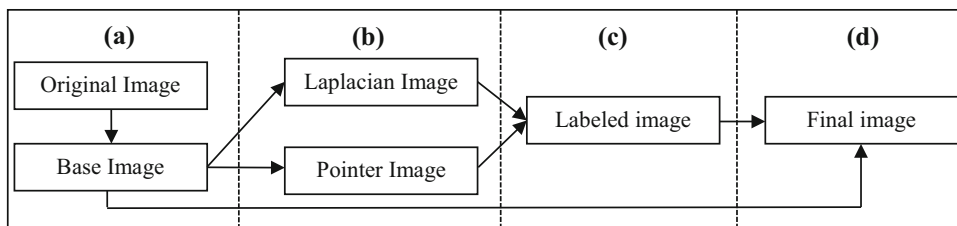
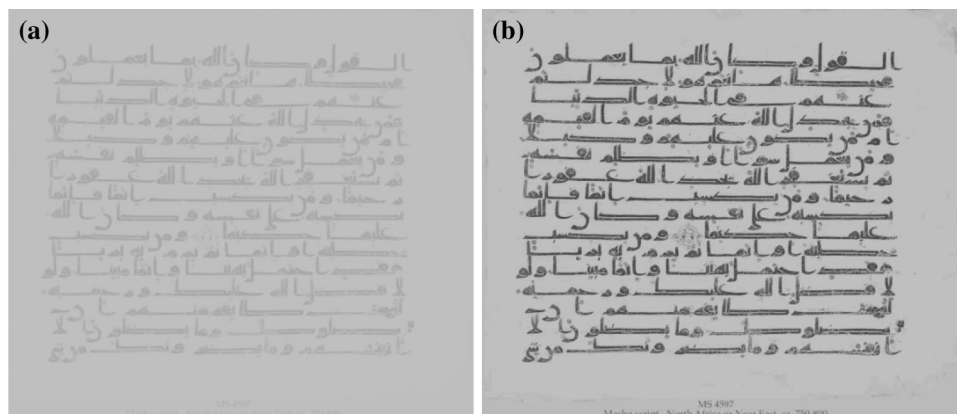


Fig. 3 An example of **a** a low contrast image and **b** the result of the modified histogram equalization method (image_{base})



method based on the level orders instead of level accumulations is applied. First, the existing pixel values are indexed in sequential order by excluding the non-existent pixels. This order index values called the order distribution value (odv), and started from 1 to the number of all existing values. The current odv of each occurrence is equal to one plus the odv of the preceding occurrence (Eq. (1)). The first existing pixel value of odv is initialized as 1 or odv(1) = 1.

$$\text{odv}(i) = \text{odv}(i - 1) + 1, \tag{1}$$

where odv(i) is the order distribution value of the current occurrence i. The odv values are then normalized into the gray-level range (0–255) by Eq. (2). Finally, the new values replaced the original values in the resultant base image (image_{base}).

$$\text{odv}_{\text{new}}(i) = \frac{\text{odv}(i) - \text{odv}(i_{\min})}{\text{odv}(i_{\max}) - \text{odv}(i_{\min})} \times G, \tag{2}$$

where odv_{new}(i) is the new value of occurrence i, odv(i_{min}) is the minimum odv value, odv(i_{max}) is the maximum odv value, and G is the maximum gray-level value (in this work, G = 255). Finally, the new values of odv_{new}(i) are replaced by the equivalent pixel values in the original image to produce the base image (image_{base}). An example of an original image and its base image (image_{base}) is shown in Fig. 3.

2.2 Geometric feature extraction

When the pre-processing stage is complete, the geometric features in the image texture are extracted using a modified Laplacian filter to produce the ground image. OFF center-surround cells and global thresholding methods are used to produce the pointer image.

- (a) **Generating the Laplacian image**
A Laplacian filter is one of the simplest and most used second-derivative image enhancement methods. It has the advantage of detecting edges in all directions of the rapid pixel contrast changes [18]. The main disadvantage of this method is that it extracts the edges for all of the pixel contrast, including unwanted intensity regions [18]. Therefore, before applying the Laplacian filter, the image is usually smoothed to reduce its sensitivity to noise. The Laplacian kernel is constructed in several forms, but the most commonly used form is the 3 × 3 kernel, as shown in Fig. 4a–d [18]. The input is a kernel center pixel value, the output gray-level value is comprised of the sum of the kernel multiplied by the pixel values. In this work, a 7 × 7 kernel Laplacian filter is used, as shown in Fig. 4e. The 7 × 7 size is used to increase the effect of the Laplacian on the image surface, which increases the strength of the contrast edge extraction [19]. That improves the ability of OFF center-surround cells

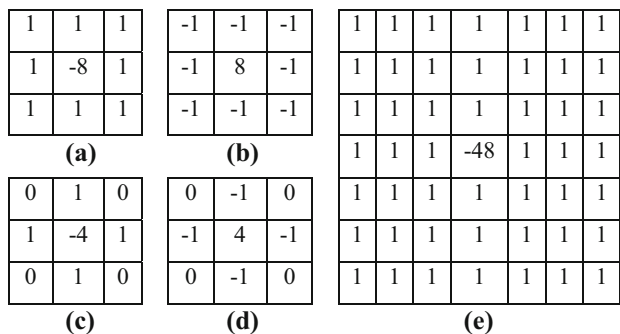


Fig. 4 a–d Some of the most common forms of the Laplacian filter, including (e) the 7×7 Laplacian filter used in this study

method to determine the position of the text body in the next stage. Equation (3) is used to compute the Laplacian value $L(x, y)$ of each pixel $P_{(0,0)}$ in the ($\text{image}_{\text{base}}$).

$$L(x, y) = \left(\left[\sum_{x,y=-k}^k P_{(x,y)} \right] - P_{(0,0)} \right) + (C \times P_{(0,0)}),$$

$$C = -(2k + 1)^2 - 1 \tag{3}$$

where $k = 3$ in this work, $2k + 1$ is the kernel value, and C is the value of the center kernel. In this study, $C = -48$ and each cell value is 1. The zero-crossing values show the abrupt changes between the background and foreground, while the static values present the static surface. To generate the final Laplacian image ($\text{Laplacian}_{\text{image}}(x, y)$), which has white background, black regions’ edges and minimum amount of noise based on gray-level/2 thresholding, the following equation is applied:

$$\text{Laplacian}_{\text{image}}(x, y) = \begin{cases} 255, & L(x, y) \leq 127 \\ 0, & 127 > L(x, y) \end{cases}, \tag{4}$$

Finally, the regions that are related by one pixel only are separated by removing the black pixel trapped between two white pixels in opposite directions. The resultant Laplacian image of this stage is shown in Fig. 6b.

(b) Generating the pointer image

In the previous section, all details in the images were extracted in the Laplacian image, including both textual and non-textual parts. In this stage, the textual parts will be denoted using OFF center-surround cells of the Human Visual System (HVS). The OFF center-surround cells of the HVS function based on the interaction to the brightness and darkness by two different cells in the retina, called ON/OFF centers ganglion cells [20, 21]. Vonikakis

et al. [22] claimed that HVS can recognize text under complex illumination conditions better than any artificial system. Moreover, the HVS can extract the text in images with background regions that are darker than the text, unlike other global thresholding methods. For these reasons, the OFF center-surround cells method is used to generate a pointer document image to denote the position of the text body in the degraded document image.

First, the OFF center-surround cells image of the base image ($\text{image}_{\text{base}}$) is extracted. The size of the center and surround masks is optimized to fit the proposed method, whereas the surrounding mask kernel is 5×5 and the center mask kernel is 1×1 . Those kernels sizes were selected because [22, 23] claimed that in HVS, the kernel of the surround mask must be larger than center mask by a factor of 4–8, and the 5×5 surround mask responds better with small to medium fonts and high-frequency contrast changes, leading to the extraction of small fonts and edges of large fonts and the exclusion of noise or low-frequency constricted regions.

In this work, the pointer image $P_{\text{image}}(x, y)$ is generated based on following. For every pixel $P_{(x,y)}$ in the base image ($\text{image}_{\text{base}}$) the surrounding mask $S_{x,y}(k)$ and center mask $C_{x,y}(l)$ average values are extracted. The OFF function value $f_{x,y}(S, C)$ of each pixel represents its surrounding average value minus its center average value, as shown in Eq. (5).

$$f_{x,y}(S, C) = S_{x,y}(k) - C_{x,y}(l), \tag{5}$$

$$S_{x,y}(k) = \frac{1}{(2k + 1)^2} \sum_{x,y=-k}^k P_{(x,y)},$$

$$C_{x,y}(l) = \frac{1}{(2l + 1)^2} \sum_{x,y=-l}^l P_{(x,y)},$$

where the surround kernel mask = 5 and therefore ($k = 2$); the center kernel mast = 1 and therefore ($l = 0$); and the pointer image P_{image} is generated based on Eq. (6). Each pixel value in the pointer P_{image} is extracted based on the fine correlation between the $C_{x,y}(l)$ and $f_{x,y}(S, C)$ values under the conditions that the surrounding average is greater than the center average. Otherwise, the $P_{\text{image}}(x, y)$ pixel is equal 0. The $C_{x,y}(l)$ value is set to the intensity value to produce the text level in the result image. The output is scaled to the interval $[0, 255]$.

$$P_{\text{image}}(x, y) = \begin{cases} \frac{f_{x,y}(S, C) \times (C_{x,y}(l) + 255)}{C_{x,y}(l) + f_{x,y}(S, C)}, & f_{x,y}(S, C) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

After this step, a thresholding method is applied to the resulting image to determine the final form of the pointer image. Globally, the method of Bataineh [7] is applied to the $P_{\text{image}}(x, y)$ image. This thresholding method is used because of its ability to extract the critical low contrast pixel levels and overcome the problem of noise. The Eq. (7) presents the form modified into a global thresholding method:

$$T = m - \frac{m^2 \times \sigma}{(m + \sigma)(127 + \sigma)} \tag{7}$$

where T is the thresholding value, m is the mean value of the global image pixels, σ is the standard deviation of the image pixels, and 127 is the median of the gray-level value (255). Finally, the shrink filter is used to remove pepper noise. In this work, the binary image is scanned to a 7×7 window and each black pixel in this window is counted. Next, these pixels are exchanged for white pixels if the total number of the black pixels is less than 15. The result of the final pointer image (P_{image}) is shown in Fig. 6c.

2.3 Region selection stage

Upon obtaining the Laplacian and pointer images, the regions that present solely text patterns are extracted from the Laplacian image using the pointer image using the following two steps:

- (a) Region labeling

The connected-component labeling algorithm is applied to the Laplacian image. This algorithm scans the pixels in the binary image and label the pixels based on pixel connectivity. The connected pixels in each separated region will have unique labeling value. To achieve this, a two-pass algorithm is adopted [24].

- (b) Region selection

After the regions in the Laplacian image are labeled, all pixels in each separated region in the Laplacian image have the same unique label value. Next, the labeled regions that belong to the textual pattern are selected. The textual region is recognized based on the existence of black pixels in both the Laplacian image ($\text{Laplacian}_{\text{image}}$) and the pointer image (P_{image}), where all pixels in the pointer image are matched with their equivalent pixels positions in Laplacian image. If any black pixel in the pointer image is matched with a labeled pixel in Laplacian image, all pixels with the same label value are kept. The label values that are not matched with any black pixels in the pointer image are removed. For example, in Fig. 5a the black pixels in the pointer image (P_{image}) of positions (3, 2), (3, 3), (3, 4),

and (4, 2) are matched with black pixels of region that is labeled as 1 in the Laplacian image (Fig. 5b). Therefore, all pixels with label “1” are kept in the region selection image (Fig. 5c). On the other hand, all equivalent pixels’ positions of label “2” (2, 9), (2, 10), (2, 11), and (3, 9) do not match with any black pixels in the pointer image. Therefore, this region is also removed. Figure 6d shows the resulting region selection image ($\text{image}_{\text{base}}$) at this stage.

2.4 Post-processing Stage

Finally, a post-processing stage is applied to improve the text region quality by filling the text body and to restore the small textual parts that are missing during the filtering processes. Based on the description above, the result selection image ($\text{image}_{\text{regions}}$) of the region selection stage is text and the edges of Laplacian image extracted in Sect. 3.2(A). These edges represent the text body itself in the thin-stroke text and the edge of the wide-stroke text. Therefore, the holes and empty text body are filled to obtain the final text. The Laplacian image is divided into several binarization windows. It is important that the height of each binarization window should be very small and consists of small number of pixels only, while its preferable width should be approximately double of the average text width in prior to minimize the impact of the unrelated pixel. Based on our experimental study, the optimal binarization window is set as 70×5 pixels. In addition, two values are extracted from the equivalent pixels in the base image: the mean value of the window (Mean_w) and the mean value of the pixels that are black in the region selection image ($\text{Mean}_{\text{black}}$). Next, a thresholding value (Eq. (8)) is applied to the base image (Eq. (9)). If the pixel’s value is less than the threshold (T_{Lap}) value, the equivalent

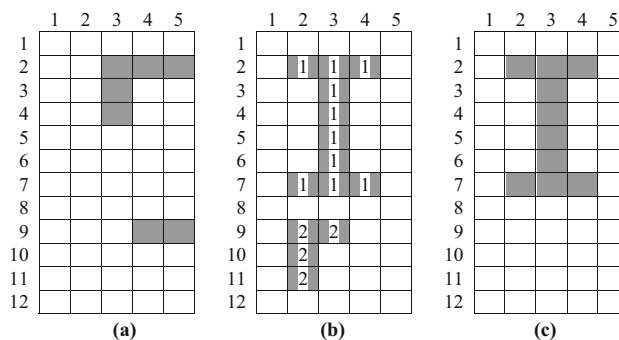


Fig. 5 **a** The pointer image P_{image} , **b** the labeled Laplacian result image ($\text{Laplacian}_{\text{image}}$), and **c** the region selection image ($\text{image}_{\text{regions}}$)

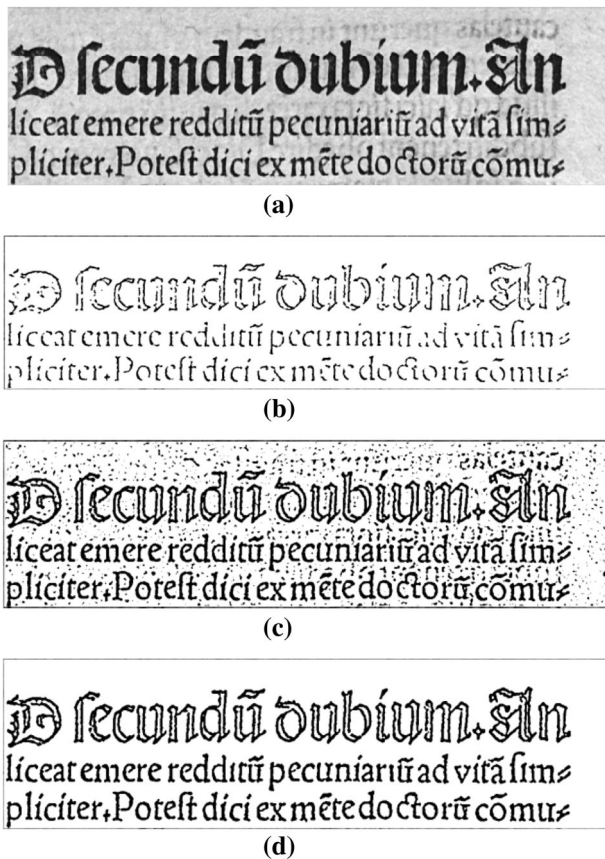


Fig. 6 **a** The base image ($image_{base}$), **b** its Laplacian result image ($Laplacian_{image}$), **c** its pointer image (P_{image}), and **d** the region selection image ($image_{regions}$)

pixel in the region selection image ($image_{region}$) will be converted from white to black. The binarization result is shown in Fig. 7. The output of this stage presents the final binary image ($image_{final}$) of the proposed method.

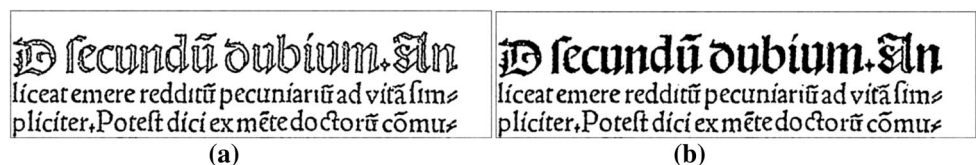
$$T_{Lap} = (Mean_{black} + Mean_W) \times 0.5 \tag{8}$$

$$image_{final}(x,y) = \begin{cases} image_{regions}(x,y) = 255, & (image_{base}(x,y)) \geq T_{Lap} \\ image_{regions}(x,y) = 0, & (image_{base}(x,y)) < T_{Lap} \end{cases} \tag{9}$$

3 Experiment and results

In this section, the proposed method is compared with five binarization methods: the Niblack [5], Sauvola [6], Sauvola and Pietikainen [10], NICK [2], and Bataineh [7] methods.

Fig. 7 **a** Region selection ($image_{regions}$) and **b** its result after the post-processing stage ($image_{final}$)



Based on the literature, these methods are chosen because they are some of the better adaptive binarization methods. In other side, most of the previous methods are based on a self-selected dataset [10, 25]; which yields biased information about the method performances. Thus, this approach yields an unfair evaluation for the binarization methods because it overlooks the method performances (and potential weaknesses) for other binarization challenges.

To meaningfully evaluate this work, the experiments utilize five binarization benchmark datasets (DIBCO). These datasets present 50 document images in total, containing printed, handwritten with color, and gray-scale images. These datasets were carefully collected by experts from several recourses to include all of the critical challenges facing document binarization processes.

In this work, two types of experiments are conducted: visual and statistical experiments.

3.1 Visual experiments

The visual experiments clearly illustrate the binarization problems and how each binarization method addresses them. These experiments are conducted on all the aforementioned datasets. Because of the large number of results, only one case from each dataset is show in this section. The selected images present compound binarization problems. The first visual experiment conducted on the image “P03”, selected from the DIBCO2009 dataset, is shown in Fig. 8a. This printed text image includes various text sizes and colors. Moreover, point noise is present in the non-smooth background. The second experiment is conducted on the image “PR2” from the H_DIBCO2010_test_images dataset. This image exhibits low contrast between the text and background, dirty spots, degraded handwritten text, various text slopes, and dark background, as shown in Fig. 9a. The next experiment used the “H02” printed text image from the DIBCO11-machine_printed dataset. As shown in Fig. 10a, this image has leaked ink from other surfaces, low contrast, and an inhomogeneous background. The fourth experiment was conducted on the “HW4” image from the DIBCO11-handwritten dataset (Fig. 11a). This image includes noise from spilled liquid, low contrast, and inhomogeneous text and background. The final visual experiment used “H06” from the H-DIBCO2012 dataset, as shown in Fig. 12a. This handwritten image has strong text leaking and an inhomogeneous background.

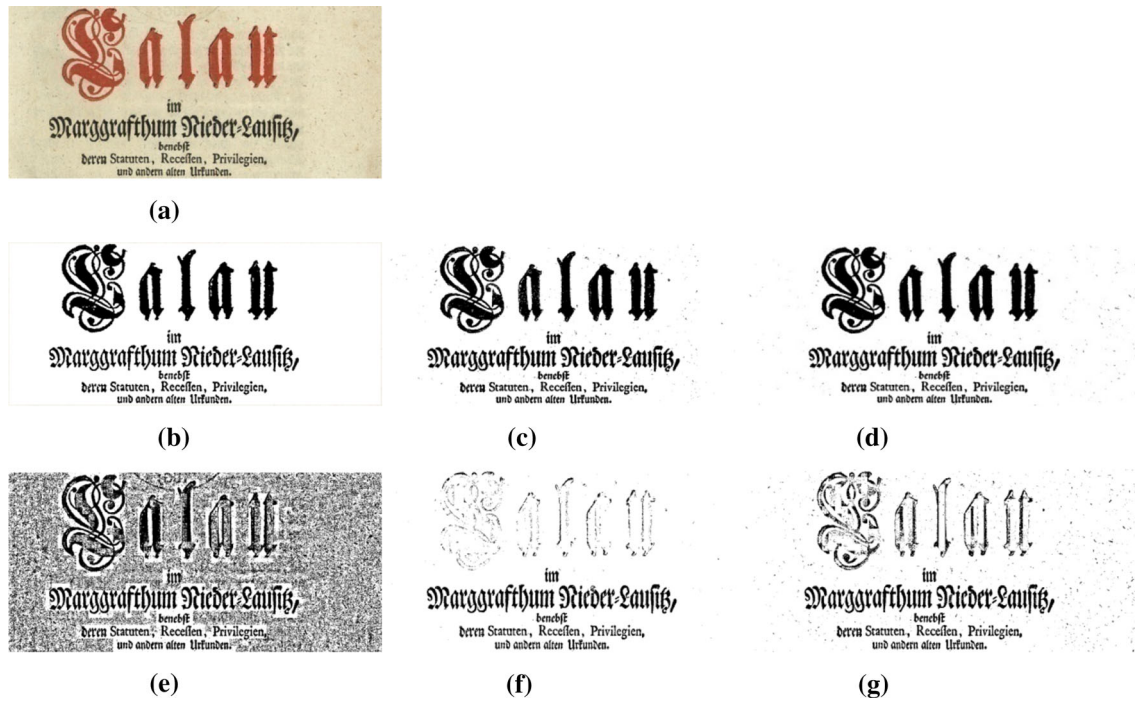


Fig. 8 a Original image “P03” from the DIBCO2009 dataset and its binarization results using **b** the proposed method and the methods of **c** Bataineh [13], **d** Sauvola compound, **e** Niblack, **f** Sauvola, and **g** NICK

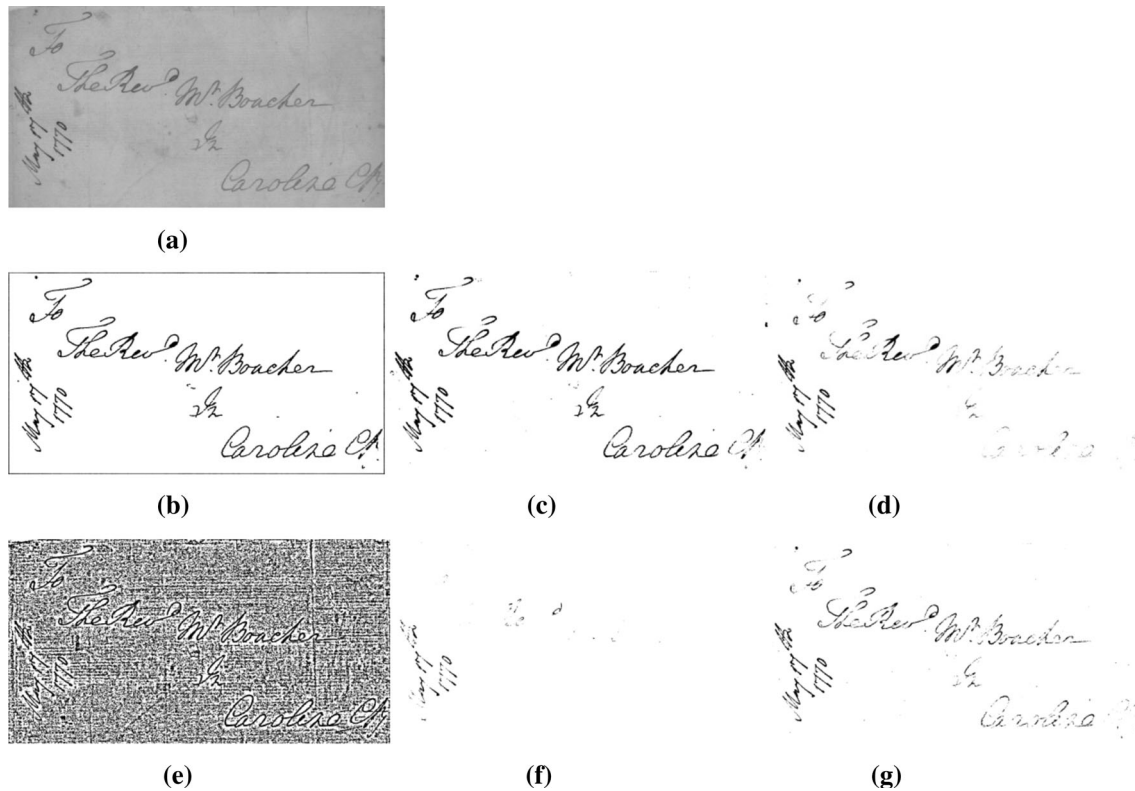


Fig. 9 a Original image “PR2” from the H_DIBCO2010 dataset and its binarization results using **b** the proposed method and the methods of **c** Bataineh [13], **d** Sauvola compound, **e** Niblack, **f** Sauvola, and **g** NICK

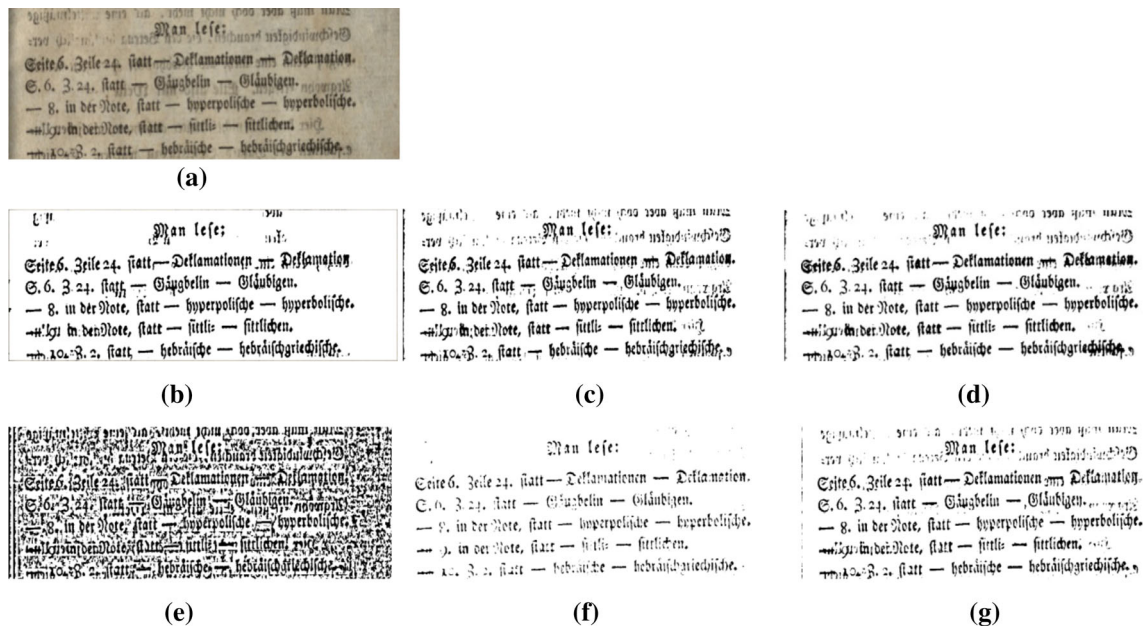


Fig. 10 a The original image “H02” from the DIBCO11-machine_printed dataset and its binarization results using the **b** proposed method and the methods of **c** Bataneh [13], **d** Sauvola compound, **e** Niblack, **f** Sauvola, and **g** NICK

Based on the visual results, the proposed method achieved the best performance among the methods studied, as shown in Figs. 8, 9, 10, 11, 12b. In the first experiment, the proposed method gives the best results (Fig. 8b), producing less noise and extracting the text of all sizes and colors better than other methods (Fig. 8c–g).

The results of the second experiment show that the proposed method provides the best text quality, without any disconnection, destruction, or noise (Fig. 9). In the third experiment, the proposed method yields the best balance between the text and noise. The Sauvola method presents the less leaked ink noise than the other methods, but it distorts the text. The other methods perform very poorly regarding the leaked ink, whereas the proposed method resulted in only a small amount of noise with perfect text presentation (Fig. 10).

This fact is also shown in the fourth experiment (Fig. 11b), wherein the proposed method overcame the noise of the spilled liquid while retaining an acceptable text quality better than other methods (Fig. 11c–g). In the final experiment (Fig. 12b), the proposed method shows the best binary representation. The leaked text is absent, and the text of interest is preserved in great detail. The other methods fail to exhibit both of these traits. As indicated by these results, the proposed method performed better than the other methods for almost all images in the datasets. These findings are presented in the numerical experiments.

3.2 Statistical experiments

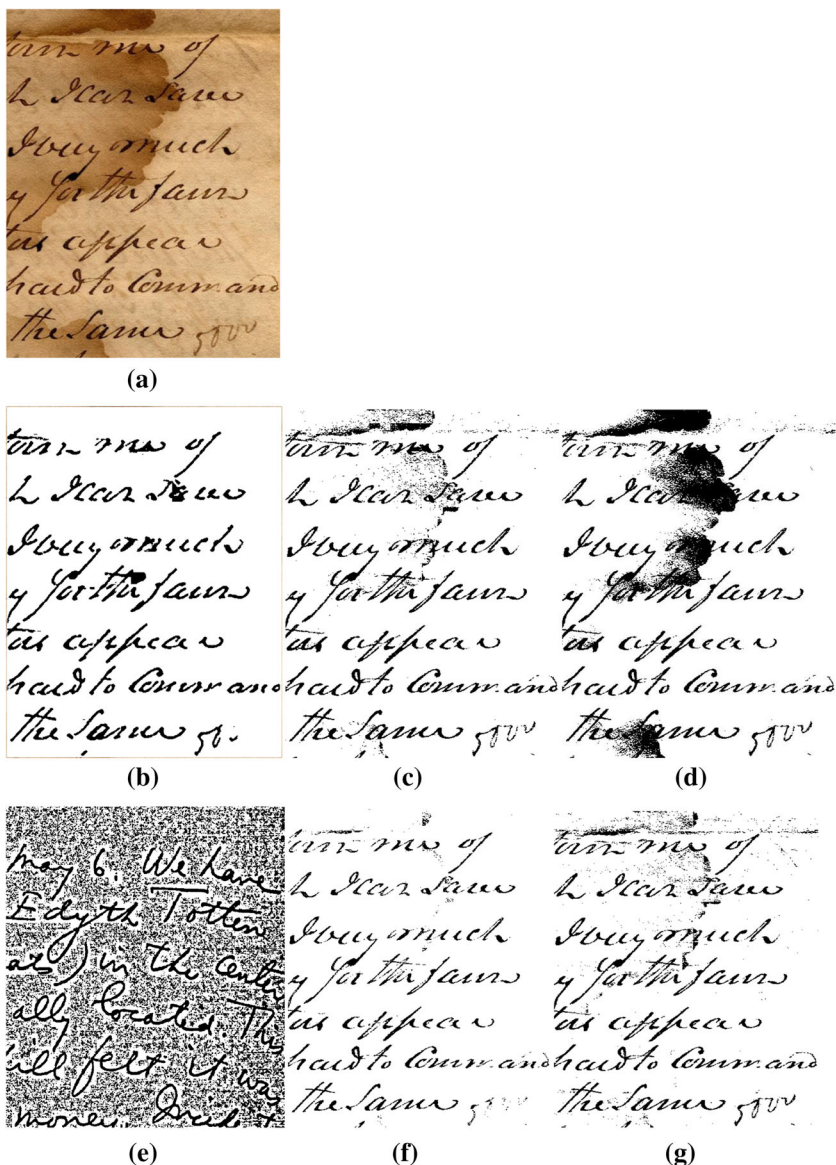
The visual experiments provided an accessible representation of the binarization problems and methods performances. However, a limited number of cases can be presented. Therefore, statistical experiments and analysis are conducted to better understand the overall trends. This work adopted the ICDAR benchmark evaluation measurement and datasets. This was proposed in the International Document Image Binarization Contest (DIBCO 2009) organized in the context of the ICDAR 2009 Conference. It has been adopted as the main benchmark evaluation technique for the binarization methods from 2009 to date. The evaluation defines the similarity percentage between the result image and ground image by F-measure, PSNR, negative rate metric (NRM) measurements, and F-Error [12–16, 26], wherein the F-measure represents the overall binarization accuracy, the precision represents the binarization noise, and the recall represents the text body accuracy.

$$F - \text{measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \times 100 \%, \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (12)$$

Fig. 11 **a** The original image “HW4” from the DIBCO11-handwritten dataset and its binarization results using **b** the proposed method and the methods of **c** Bataineh [13], **d** Sauvola compound, **e** Niblack, **f** Sauvola, and **g** NICK



The PSNR represents the similarity between the ground image and the binarization result image.

$$PSNR = 10 \log \left(\frac{C^2}{MSE} \right), \tag{13}$$

where $MSE = \frac{\sum_{x,y}^{m,n} (\text{original image}(x,y) - \text{binary image}(x,y))^2}{MN}$, and c is the maximum gray-scale level value. The F-Error represents the amount of the black noises to the text in the binarization result image

$$F - Error = \frac{FP + FN}{TP + FN} \tag{14}$$

The negative rate metric (NRM) represents the mismatch between the ground image and the binarization result image.

$$NRM = \frac{FN_* + FP_*}{2}, \tag{15}$$

where $FN_* = \frac{FN}{FN + TP}$,

$$FP_* = \frac{FP}{FP + TN},$$

and TP is the number of true positives, TN is the number of false negatives, FP is the number of false positives, and FN is the number of false negatives.

The experiments are conducted on each dataset, and the results are shown in Tables 1, 2, 3, 4, 5. The first analytical experiment was conducted on DIBCO2009 dataset. As shown in Table 1, the proposed method performs better than the other methods on this dataset. The average F-measure is 89.34 % for the proposed method, compared

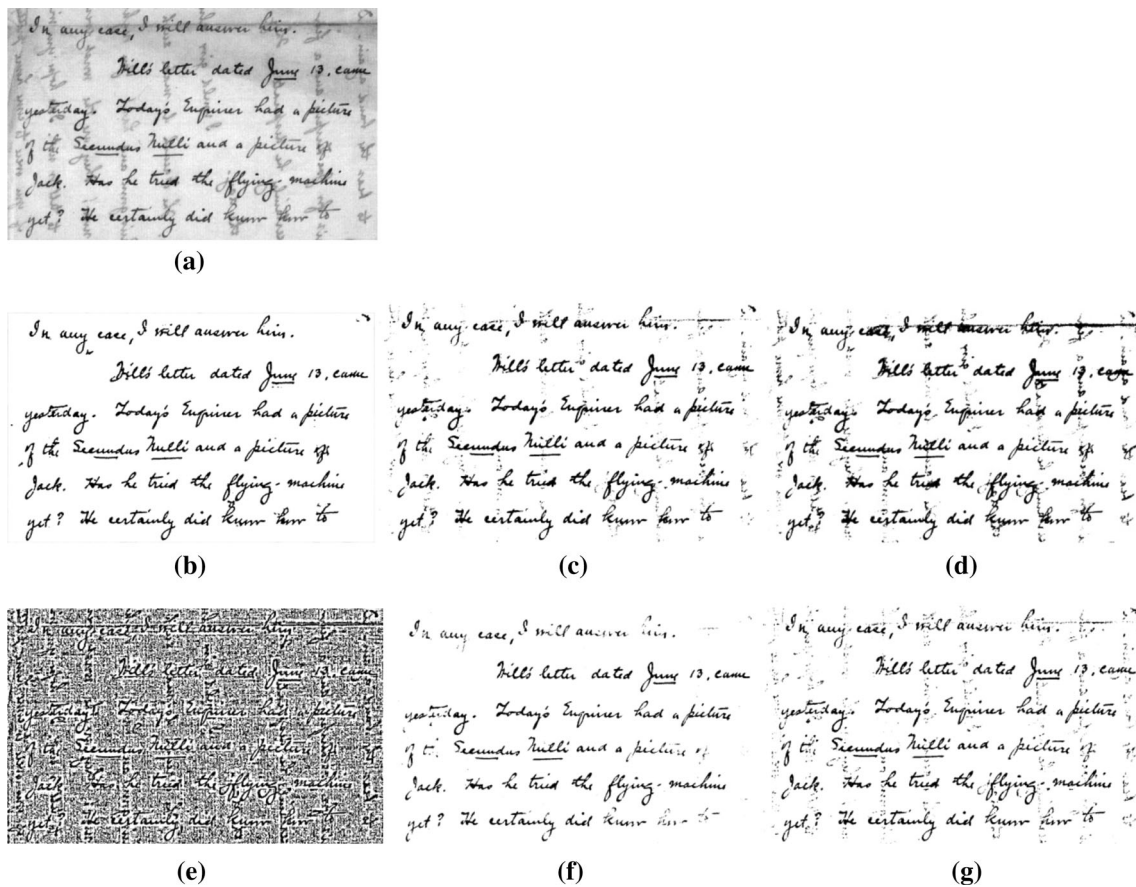


Fig. 12 a The original image “H06” from the H-DIBCO2012 dataset and its binarization results using b the proposed method and the methods of c Bataineh [13], d Sauvola compound, e Niblack, f Sauvola, and g NICK

to 41.78, 68.87, 78.06, 80.96, and 88 % for the Niblack, Sauvola thresholding, Sauvola compound algorithm, NICK and Bataineh binarization methods, respectively. The proposed method also achieved the best performance in terms of PSNR, NRM and F-Error.

As shown in Table 2, the proposed method outperformed the other methods for the H_DIBCO2010_test_images dataset. The average F-measure for the proposed method is 84.75 %, compared to 28.51, 42.46, 61.72, 65.89, 82.82, and 84.75 % for the Niblack, Sauvola thresholding, Sauvola compound algorithm, NICK, and Bataineh binarization methods, respectively. The proposed method also achieved the best performance in terms of PSNR and NRM. Moreover, the proposed method performed best in terms of PSNR, NRM and F-Error.

As shown in Table 3, the proposed method achieved an F-measure of 85.91 % on the DIBCO11-handwritten dataset, which is higher than that of the Niblack, Sauvola thresholding, Sauvola algorithm, NICK and Bataineh methods (30.52, 67.28, 73.81, 79.56, and 83.81 %, respectively). In addition, the proposed method achieved an F-measure of 83.42 % for the DIBCO11-machine_printed

dataset, as shown in Table 4. This value is higher than the F-measure average of the Niblack, Sauvola thresholding, Sauvola compound, NICK and Bataineh methods, which were 42.76, 65.26, 84.89, 78.55, and 83.42 %, respectively. Additionally, the proposed method achieved the best PSNR, NRM and F-Error for both datasets.

Based on Table 5, the proposed method performs better than the other methods. The average F-measure is 87.98 % for the proposed method, compared to 28.53, 52.71, 69.62, 74.56, and 86.88 % for the Niblack, Sauvola thresholding, Sauvola compound algorithm, NICK, and Bataineh binarization methods, respectively. Additionally, the proposed method achieved better PSNR and NRM than the other methods. As shown in Fig. 13, the proposed method outperformed the other methods for all datasets.

Based on the above-mentioned experiments, some conclusions have been drawn:

- The proposed method is based on extracting the edges where a rapid change in the pixel contrast occurs. This approach ignores most of the noise, such as leaked ink, spilled liquid and inhomogeneous background, which

Table 1 Performance of the DIBCO2009 dataset

	F-measure (%)	Recall (%)	Precision (%)	PSNR	NRM
Niblack	39.64	91.2	27.1	5.8	18.94
Sauvola	63.25	51	98	13.95	24.6
Sauvola_2	78.06	88.3	76.4	15.3	7.4
Nick	80.07	74.5	89.6	15.27	13.05
Bataineh	88	85.85	89.75	17.03	7.48
Proposed	89.34	90.86	88.28	17.5	5.14

Table 2 Performance of the H_DIBCO2010_test_images dataset

	F-measure (%)	Recall (%)	Precision (%)	PSNR	NRM
Niblack	28.9	89.5	17.5	5.26	21.22
Sauvola	41.75	32	88.9	13.97	34.05
Sauvola_2	63.04	55.4	93.8	15.4	22.55
Nick	65.91	55.1	95.2	15.08	22.6
Bataineh	82.9	75.7	93.3	16.95	12.36
Proposed	84.77	82.89	87.6	17.22	9

Table 3 Performance of the DIBCO11-handwritten dataset

	F-measure (%)	Recall (%)	Precision (%)	PSNR	NRM
Niblack	30.86	90.06	19	5.23	20.91
Sauvola	67.31	57	96	15.43	21.73
Sauvola_2	74.01	76.63	78	14.393	13.62
Nick	79.7	75.75	86.63	15.94	12.69
Bataineh	83.97	83.37	85.38	16.84	9.02
Proposed	86.5	86.7	86.6	17.6	7.19

Table 4 Performance of the DIBCO11-machine_printed dataset

	F-measure (%)	Recall (%)	Precision (%)	PSNR	NRM
Niblack	40.68	86.37	28.25	5.61	21.97
Sauvola	65.26	51.88	99.5	13.52	24.05
Sauvola_2	84.99	82.3	90.63	16.03	9.48
Nick	78.62	73.25	88.13	14.25	14
Bataineh	83.57	86.5	83.25	15.18	7.52
Proposed	87.98	84.65	91.88	16.54	8.17

Table 5 Performance for the H-DIBCO2012 dataset

	F-measure (%)	Recall (%)	Precision (%)	PSNR	NRM
Niblack	28.82	89.85	17.43	5.22	21.14
Sauvola	52.72	43.65	98.29	14.99	28.27
Sauvola_2	69.63	66.43	88.29	15.23	17.31
Nick	74.54	65	94	16.05	17.71
Bataineh	86.89	84	91.21	17.89	8.28
Proposed	88.11	86.35	90.96	18.39	7.16

appear as noise in other methods that focus mainly on the pixel values themselves.

- The principle of the proposed method is capable of extracting a fine text body because a constricted process

on the text is applied. Additionally, there is less noise than other methods because the principle of the text constriction does not correspond with the nature of the noise structure.

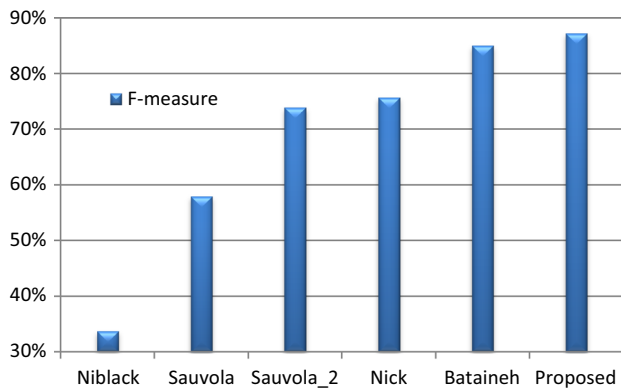


Fig. 13 Average F-measure of the Niblack, Sauvola thresholding, Sauvola compound algorithm, NICK, and Bataineh methods and the proposed binarization method for all five datasets

- The proposed method is independent of the binarization window size and is not a thresholding method, both of which are primary principles of most binarization methods. Those components have been implemented perfectly for some cases, but failed with others, especially in the case of compound binarization challenges.
- The proposed method may face problems in some cases if the contrast between the text and background is very low or the contrast between the noise and background is very high. However, these problems affect other binarization methods more strongly than the proposed method.
- In general, the proposed method can be implemented well with any binarization case. In contrast, the other methods perform well for some binarization cases and fail for others.

4 Conclusion

In this paper, an adaptive binarization method for document images was proposed based on the variance between pixel contrasts. The proposed method was evaluated based visual and statistical experiments. The experiments were conducted using DIBCO2009, H-DIBCO2010_test_images, DIBCO11-handwritten, DIBCO11-machine_printed, and H-DIBCO2012-dataset benchmark datasets. The proposed work has also been compared with: Niblack, Sauvola thresholding, Sauvola compound algorithm, NICK, and Bataineh binarization methods. The results show that the proposed method achieved better performance than the other methods in all binarization cases and benchmark datasets. It removes noise and remains the text body. The proposed work effectively overcomes almost all of the noise problems, such as leaked ink, spilled liquid and

inhomogeneous background. Moreover, this proposed method is considered as a global binarization approach which skips the problems of local methods such as window size, thresholding methods and empty window problems. It can also be implemented with any document image under any language and binarization challenges.

References

1. Kefali A, Sari T, Sellami M (2010) Evaluation of several binarization techniques for old Arabic documents images. In: 1st International symposium on modeling and implementing complex systems “MISC’2010”, pp 88–99
2. Khurshid K, Siddiqi I, Faure C, Vincent N (2010) Comparison of Niblack Inspired Binarization Methods for Ancient Documents. In: 16th International conference on Document Recognition and Retrieval, pp 1–10
3. Stathis P, Kavallieratou E, Papamarkos N (2008) An evaluation technique for binarization algorithms. *J Univers Comput Sci* 14:3011–3030
4. Otsu N (1979) A thresholding selection method from gray-scale histogram. *IEEE Trans Syst Man Cybern* 9:62–66
5. Niblack W (1985) An introduction to digital image processing. Prentice Hall, pp 115–116
6. Sauvola J, Seppanen T, Haapakoski S, Pietikainen M (1997) Adaptive document binarization. In: Fourth international conference document analysis and recognition (ICDAR), pp 147–152
7. Bataineh B, Abdullah SNHS, Omer K (2011) An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows. *Pattern Recogn Lett* 32:1805–1813
8. Chou C, Lin W, Chang F (2010) A binarization method with learning-built rules for document images produced by cameras. *Pattern Recognit* 43:1518–1530
9. Gatos B, Pratikakis I, Perantonis S (2006) Adaptive degraded document image binarization. *Pattern Recognit* 39:317–327
10. Sauvola J, Pietikainen M (2000) Adaptive document image binarization. *Pattern Recognit* 33:225–236
11. Howe R (2011) A laplacian energy for document binarization, “ICDAR 2011”. In: International conference on document analysis and recognition, pp 6–10
12. Gatos B, Ntirogiannis K, Pratikakis I (2009) ICDAR 2009 document image binarization contest. In: proceedings 10th international conference on document analysis and recognition, pp 1375–1382
13. Gatos B, Ntirogiannis K, Pratikakis I, DIBCO 2009 (2009) Document image binarization contest. *Int J Doc Anal Recognit* 14(2011):35–44
14. Pratikakis I, Gatos B, Ntirogiannis K (2010) H-DIBCO 2010—handwritten document image binarization competition. In: 12th international conference on frontiers in handwriting recognition, pp 727–732
15. Pratikakis I, Gatos B, Ntirogiannis K (2011) ICDAR 2011 document image binarization contest (DIBCO 2011). In: International conference on document analysis and recognition “ICDAR2011”, pp 1506–1510
16. Pratikakis I, Gatos B, Ntirogiannis K (2012) ICFHR 2012 competition on handwritten document image binarization (H-DIBCO 2012). In: 2012 international conference on frontiers in handwriting recognition, pp 813–818
17. Bassiou N, Kotropoulos C (2007) Color image histogram equalization by absolute discounting back-off. *Comput Vis Image Underst* 107:108–122

18. Gonzalez RC, Woods RE (2007) Digital image processing, 3rd ed. Prentice Hall, Upper Saddle River, NJ
19. Bar-Yosef I, Beckman I, Kedem K, Dinstein I (2007) Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. *Int J Doc Anal Recognit* 9:89–99
20. Chichilnisky E, Kalmar JRS (2002) Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neurosci* 22:2737–2747
21. Fiorentini A (2004) Brightness and lightness. In: *The visual neurosciences*, vol 2. MIT Press, Cambridge, pp 881–891
22. Vonikakis V, Andreadis I, Papamarkos N (2011) Robust document binarization with OFF center-surround cells. *Pattern Anal Appl* 14:219–234
23. Konstantinidis K, Vonikakis V, Panitsidis G, Andreadis I (2011) A center-surround histogram for content based image retrieval. *Pattern Anal Appl* 14:251–260
24. Shapiro LG, Stockman G (2001) *Computer vision*, 1st edn. Prentice Hall PTR, Upper Saddle River, NJ
25. Chiu Y, Chung K, Yang W, Huang Y, Liao C (2012) Parameter-free based two-stage method for binarizing degraded document images. *Pattern Recogn* 45:4250–4262
26. Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45:427–437
27. Narendra, Patrenahalli M (1981) A separable median filter for image noise smoothing. In: *IEEE transactions on pattern analysis and machine intelligence*, pp 20–29
28. Wang J, Lin L (1997) Improved median filter using minmax algorithm for image processing. *Electron Lett* 33:1362–1363