Tech Science Press

Novel Adaptive Binarization Method for Degraded Document Images

Siti Norul Huda Sheikh Abdullah[1], Saad M. Ismail[1,2], Mohammad Kamrul Hasan[1,*] and Palaiahnakote Shivakumara[3]

[1]Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selangor, 43600, Malaysia
[2]The Faculty of Computer and Information Sciences, Aljouf University, Saudi Arabia
[3]Faculty of Computer Systems and Information Technology, University of Malaya, Malaysia
[*]Corresponding Author: Mohammad Kamrul Hasan. Email: mkhasan@ukm.edu.my

Abstract: Achieving a good recognition rate for degraded document images is difficult as degraded document images suffer from low contrast, bleed-through, and nonuniform illumination effects. Unlike the existing baseline thresholding techniques that use fixed thresholds and windows, the proposed method introduces a concept for obtaining dynamic windows according to the image content to achieve better binarization. To enhance a low-contrast image, we proposed a new mean histogram stretching method for suppressing noisy pixels in the background and, simultaneously, increasing pixel contrast at edges or near edges, which results in an enhanced image. For the enhanced image, we propose a new method for deriving adaptive local thresholds for dynamic windows. The dynamic window is derived by exploiting the advantage of Otsu thresholding. To assess the performance of the proposed method, we have used standard databases, namely, document image binarization contest (DIBCO), for experimentation. The comparative study on well-known existing methods indicates that the proposed method outperforms the existing methods in terms of quality and recognition rate.

1 Introduction

The digitization of historical documents by authors has recently become an active research area in document analysis. This is because digitization helps preserve old records for a long time and understand their history, which could lead to new inventions in their respective fields [1]. The literature on document analysis revealed that optical character recognition (OCR) is a successful system in the area of image processing and pattern recognition owing to its capability to achieve more than 90% recognition for scanned plain background images [2,3]. However, when applied to degraded historical document images, the OCR system cannot achieve a high recognition rate

for plain background images. The current OCR has its inherent limitations, such as homogeneous background, binary images, character sizes, and character fonts [4–6].

Unlike plain background images Fig. 1a, degraded historical documents suffer from bleed-through and display severe noise due to the folding and nonuniform illumination effect. The blur effect due to aging and degradation has caused the developed OCR to have severe issues in performing well for degraded documents. Hence, achieving a high recognition rate for degraded document images is still an elusive goal for researches on document analysis [7–9]. Several existing methods have been developed to address the issue of degraded document images [10–14]. Most of these methods are based on thresholding. Almost all of the methods were developed by targeting plain background images but not degraded images. As a result, the existing thresholding techniques fix constant thresholds and window sizes when binarizing document images [15–18]. Due to the unpredictable characteristics of degraded document images, the fixed thresholds and window sizes may not work well. Fig. 1a presents the plain; 1b, the degraded historical document images; 1c, the results of Otsu thresholding [16]; and 1d, the results of the Bataineh adaptive thresholding technique [12]. Therefore, there is immense scope for studying automatic thresholding and window sizes to address to the issue of background complexity. Thus, a new binarization approach is important to solve the current issue.
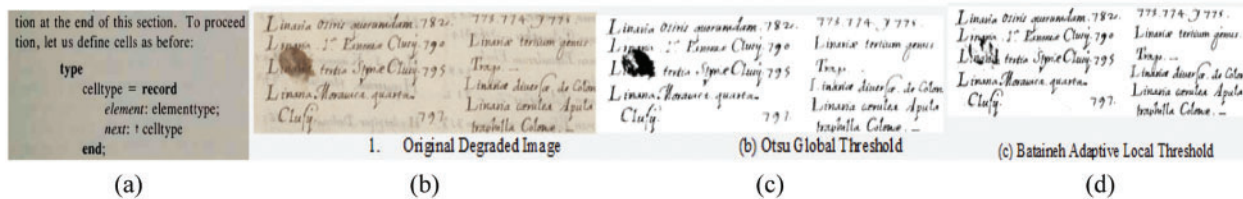

Figure 1: (a) Plain background, (b) degraded images, (c) Otsu global thresholding images [16], and (d) Bataineh adaptive local thresholding images [12]

This paper aimed to address the existing document background challenges, such as low contrast, bleed-through, and nonuniform illumination effects, using a new technique. The proposed method is capable of solving both the problem of a plain background and degraded document images. The main contributions of this paper are summarized as follows:

(1) It integrates the simple and hybrid methods called the adaptive compound threshold technique;
(2) It formulates a new mean histogram stretching (MHS) method for suppressing noisy pixels in the background and increasing pixel contrast at edges or near edges to enhance the image;
(3) It derives adaptive local thresholds for dynamic window size determination using Otsu thresholding.

The remainder of this paper is organized as follows. Section 2 presents related works. Section 3 describes the proposed binarization method. Section 4 discusses the measurements of the binarization method for evaluation metrics; the experimental results present the comparison among the methods using several benchmark datasets with varying types of degradation. Finally, Section 5 concludes the paper.

2 Related Works

According to the literature [19–22], two methods can be applied to recognize degraded document images. The first one is to develop the features and classifiers for recognizing text, and the second one is to binarize the given image and use the available OCR application for recognition rather than to develop a separate OCR for the degraded document images. This work employs the second method because the first one has issues with regard to feasibility and advisability. After all, the existing methods function only in specific cases, and their performances depend on the trained classifier.

The second method is a generalized one. It provides binarized output with a clear character shape. Therefore, in this work, we review the literature on thresholding techniques and binarization algorithms used in the available OCR application [11–39]. The existing methods can be categorized into two: Simple methods and hybrid methods. The simple methods can be classified further into global and local thresholding methods, whereas the hybrid methods use both the global and local thresholding methods for binarization, as presented in Fig. 2.
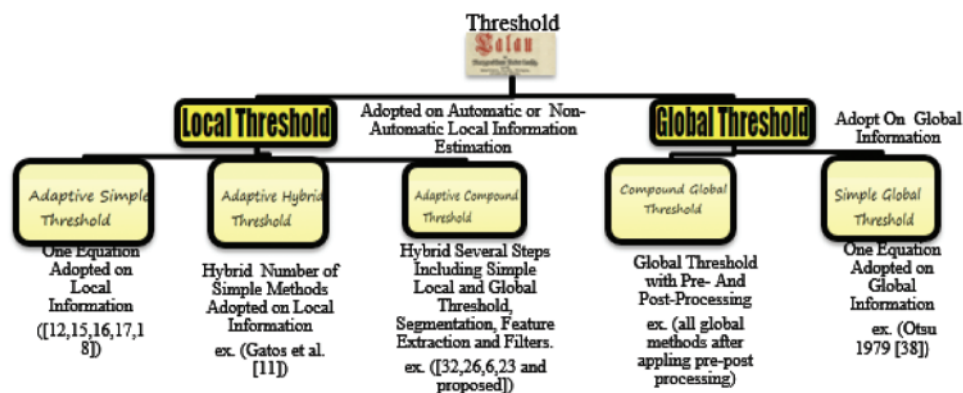

Figure 2: Categorization of binarization methods based on the type, characteristics, and information content

The enhancement of degraded images using a simple thresholding or a compound approach concerning local approach adaptation has become a cutting-edge research topic [11,12, 14,16,17,23–29]. Initially, Sehad et al. [30] proposed a simple local binarization method based on the mean and variance of the computed pixels of the original image and the local binary pattern (LBP) image. These features are adopted in a Sauvola threshold-based method. The LBP operator is a texture analysis method used in this method to solve the problem of poor contrast exhibited by the document images. Wagdy et al. [31] presented a simple global binarization consisting of two steps: Degradation enhancement and global binarization. First, the retinex theory is employed to enhance the degraded document image by decomposing the image into two components, namely, illumination and reflectance. Then, the Otsu global threshold is employed for binarization. A simple local threshold method, namely, the Niblack binarization algorithm [16], calculates the local threshold for each pixel by gliding a rectangular window over the whole image. The threshold is calculated using the mean and standard deviation of all the pixels in the window. Additionally, Sauvola and Pietikäinen [17] proposed a modified version of the Niblack method to provide enhanced performance in the documents with a background containing a low contrast texture and uneven illumination. Khurshid et al. [15] proposed a threshold method called NICK, which

is derived from the Niblack binarization method. This method has several advantages, such as the ability to significantly improve the binarization of "white" pages and lighted images, as well as low-contrast images. Bataineh et al. [12] proposed a simple threshold method with dynamic window generation. This method resolves the problems of low-contrast images and thin pen strokes. Howe [32] also proposed an adaptive method that uses the Laplacian operator to evaluate the local likelihood of foreground and background labels and involves Canny edge detection to identify potential discontinuities and graph cut implementation to find the minimum energy.

Until now, none of the abovementioned compound methods were able to solve the challenges of a single approach, such as low contrast, thin pen strokes, black spots, and nonuniform background, nor yield optimal results. Gradually, numerous thresholding developers have improved their previous methods [12]. For example, in 2015, Bataineh et al. [14] proposed an adaptive compound method based on pixel contrast variance. This method consists of four stages: Pre-processing, geometrical feature extraction, feature selection, and post-processing. They claimed that their proposed [12] method could effectively solve the problems of thin pin stroke and low contrast. Mandal et al. [33] proposed the hybrid binarization method based on several steps, starting with morphological operations. Then, they employed background estimation to increase the text region contrast and the image contrast histogram to achieve the initial threshold text region segmentation. Finally, the local threshold method was applied for the final binarization.

Similarly, a compound binarization technique proposed by Adak et al. [34] adopted an LBP and contour analysis to filter out the noise pixels before separating the foreground pixels via background estimation. After enhancing the filtered image using the spatial smoothing technique, it used a local NICK [15] thresholding technique for image binarization. Contrarily, Singh et al. [24] proposed a compound binarization method by fusing Gatos et al. [11] and Otsu [16], plus dilation and logical AND operations to uncover the disconnected characters or the detection of false-negative text from severely degraded document images. Moreover, Ntirogiannis et al. [35] presented a newly compound binarization method that employs image normalization based on background compensation. They fused the global and local adaptive binarization methods. Additionally, Moghaddam et al. [36] proposed an adaptive threshold with parameterless behavior that combined the grid-based modeling with their threshold method to estimate the background map.

Conversely, several studies have adopted a contrast map or filter-based approach [11,13,37] to provide a better binarization process. Su et al. [13] introduced an adaptive binarization method by constructing a contrast map. The contrast map was then binarized and combined with a Canny edge map to identify the text stroke edge pixels. The document text was further segmented local threshold that is estimated within a local window. In a separate effort, Gatos et al. [11] presented an adaptive binarization approach consisting of the following steps: Pre-processing procedure using a low-pass Wiener filter, a rough estimation of the foreground regions, background surface calculation by interpolating neighboring background intensities, thresholding by combining the calculated background surface with the original image while incorporating image upsampling, and post-processing procedure.

Additionally, Chiu et al. [38] presented a method based on two stages to binarize the degraded document images. An incremental scheme determined a proper window size to prevent any significant increase in local pixel variation. Then, the final binarized image adopted the noise-suppressing scheme. The noise-suppressing scheme was accomplished by contrasting binarized images via adaptive thresholding, incorporating the local mean gray and gradient values. Unlike others, Lu et al. [37], Varish et al. [40] proposed an adaptive binarization algorithm that includes

the following steps: Background extraction, stroke edge detection, local thresholding, and post-processing. Recently, machine learning has become a promising approach to document image analysis [25,26]. In 2015, Pastor-Pellicer et al. [25] proposed a binarization method based on a convolutional neural network; this method uses a sliding window and its neighboring pixels to classify the central pixel foreground or background.

Similarly, Kaur et al. [26] employed another popular supervised learning method called support vector machine (SVM) in a binarization method. This approach was derived from the method proposed by Sauvola et al. [17]. It employs several parameter selections and uses the SVM to reconstruct the binary image. However, the methods mentioned previously can be categorized into compound and hybrid binarization methods. Moreover, using several approaches in the category of simple and compound have been proposed. Raj et al. [23] proposed another form of multilevel thresholding method that employed optimization, namely, Tsallis fuzzy entropy and differential evolution. This approach relies on differential evolution to determine the optimal multilevel thresholds for image segmentation. Unlike the handcrafted approaches, their approach requires less human intervention, rule of thumb or heuristic formulation in advance.

Simultaneously, simple methods work well with plain background images and binary information but not for gray images with a complex background. Conversely, hybrid methods work well with gray images with a complex background as they employ adaptive thresholds that utilize local information. However, when an image, such as a degraded historical document image, is affected by severe illumination, nonuniform color, and low contrast, the hybrid methods cannot perform well as the derived adaptive threshold method that uses local information may not be able to handle such situations. This finding indicates the shortage of new methods for solving the problem of degraded historical document images. Therefore, in this paper, we proposed a new method for binarizing degraded historical document images, integrating the advantages of hybrid and straightforward methods to derive dynamic thresholds and windows.

3 Proposed Method

For enhancing low-contrast text information of degraded historical document images, an MHS method was proposed. Motivated by the work in [41] on image contrast enhancement, in which background subtraction is proposed to enhance low-contrast pixels, we explored the same idea of background subtraction in a new way with a new operation called stretching through the normalization of pixels. To enhance the degraded document image, the global information or the local information alone is insufficient to yield good results. Therefore, the proposed method integrates global and local information to determine the dynamic threshold using an adaptive compound threshold. This step involves the determination of both adaptive compound threshold and dynamic window. This approach outputs the binarized image of the input degraded document image. Sometimes, noise can appear during the binarization process. Therefore, we performed median-filter and area-opening operations to remove such noise in the image. This step is referred to as post-processing in this work. Finally, the steps were validated through different experiments using standard databases. The pipeline of the proposed method is presented in Fig. 3.

3.1 Mean Histogram Stretching for Image Enhancement

For the input image presented in Fig. 4a, the proposed method subtracts each pixel from the global mean and then performs a normalization process after the subtraction output. The mean value of the whole image is computed and is considered as the global mean value. To stretch the pixels obtained from the subtraction operation, we proposed normalization, as expressed by

Eq. (1). This process results in an enhanced image, and the whole process is called MHS; here, the stretching is defined as a re-distribution of the pixel values of the input image over a wider or a narrower range of values into the output image. In general, stretching is used to enhance the contrast. This effect of this operation is presented in Fig. 4, which results in the loss of information in the binarization result yielded by the proposed method for the input image in Fig. 4b without MHS. Similarly, Fig. 4c demonstrates the enhancement that was accomplished by MHS, whereas Fig. 4d demonstrates the binarization results of the proposed method for the image in Fig. 5c; it appears that the lost information in Fig. 4b is restored. This difference is the advantage of MHS, which helps $=$ restore low-contrast text pixels.

$$I_{Enhance}(x,y) = 1 - \frac{1-(I(x,y)-m_g)}{1-min(I(x,y)-m_g)} \tag{1}$$

where $I_{Enhance}$ denotes the final enhanced image; $I(x,y)-m_g$, the result of the input image after performing subtraction from the global mean (m_g); and min, the minimum pixel value in the subtraction image. Note that to visually demonstrate the effect of this process, we employed the proposed binarization method in the input image and the enhanced image obtained from this step, as presented in Fig. 3 with and without the enhancement.
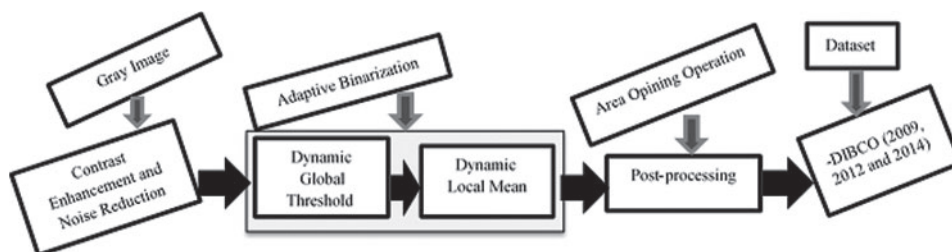

Figure 3: The procedure of the proposed adaptive method based on dynamic global threshold and dynamic local mean

(a)   (b)   (c)   (d)
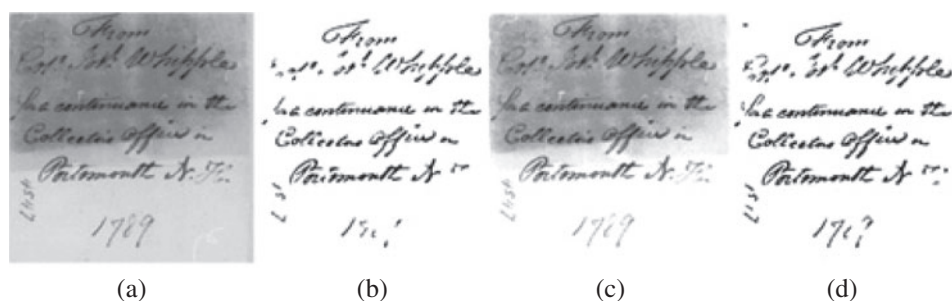
Figure 4: Part of the image presents the effects of the proposed enhancement: (a) Original gray, (b) binarization of gray image, (c) enhanced gray, and (d) binarization of enhanced image

3.2 Global Threshold Value Determination

For the enhanced gray image obtained from the previous step, we determined the global threshold value that was integrated with the local threshold value in subsequent sections. Inspired by the Otsu thresholding [39], which automatically classifies the pixels in a gray image into two,

we proposed the same operation to obtain two classes for the input enhanced image, namely, C_0 and C_1. For each class, the proposed method calculates the probability of the pixel distribution, as expressed by Eqs. (3) and (4). Next, the proposed method calculates the means for the classes of probability pixels, as expressed by Eqs. (5) and (6). Finally, the proposed method calculates the global threshold using Eqs. (3) to (6), as defined in Eq. (7), where T_g denotes the global threshold value, and k denotes a constant value. Due to the difficulty of handling degraded document images, using T_g is insufficient. Therefore, we introduced a constant to derive the optimal threshold value using T_g. Fig. 5a presents the sample results of two classes for the enhanced image, as well as the effect of the k parameter. As the k value changes, the image result is altered. Thus, it is important to automatically determine the correct k value. Let the pixels of a given image be represented as L gray levels $[1, 2 \ldots L]$. N_i denotes the number of pixels at level i, and the total number of pixels is expressed as $N = n_1 + n_2 + \cdots + n_l$. To simplify, the gray-level histogram is normalized as a probability distribution (P), as expressed by Eq. (2):

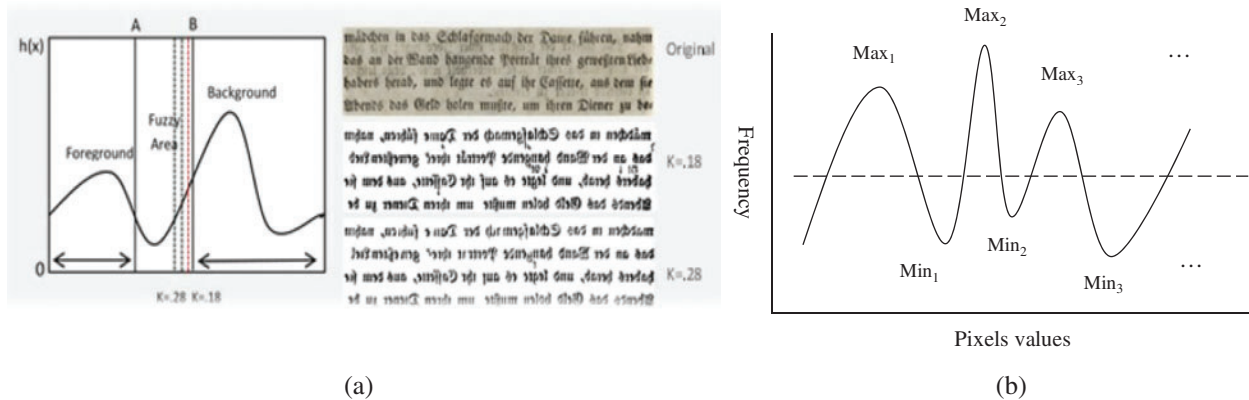$$P(i) = \frac{N_i}{N}, \quad P_i \geq 0, \quad \sum_{i=1}^{L} P(i) = 1 \tag{2}$$

(a)   (b)

Figure 5: The effectiveness of parameter (a) k when it takes two different values, and (b) an example of the maximum and minimum peak values in the histogram

Next, suppose that we dichotomize the pixels into two classes, namely, C_0 and C_1 (background and object, or vice versa), by a threshold at level k; C_0 denotes the pixels at levels $[1 \ldots T_g]$, whereas C_1 denotes the pixels at levels $[T_g + 1 \ldots L]$. Then, the probabilities of class occurrence (u) and the class mean levels (w), respectively, are given as follows:

$$u_0 = P(c_0) = \sum_{i=1}^{T_g} P(i) \tag{3}$$

$$u_1 = P(c_1) = \sum_{i=T_g+1}^{L} P(i) \tag{4}$$

$$w_0 = P(c_0) = \frac{\sum_{i=1}^{T_g} iP(i)}{u_0}, \text{ the means for class } c_0 \tag{5}$$

$$w_1 = P(c_1) = \frac{\sum_{i=T_g+1}^{L} iP(i)}{u_1}, \text{ the means for class } c_1 \tag{6}$$

The final global threshold is expressed by Eq. (7):

$$T_g = max((u_0 \cdot u_1)(w_1 - w_0)^2) \times k \tag{7}$$

3.3 Dynamic Window Size Determination Through K

Typically, a document image contains acluster of text pixels, with dense pixels in the area of text cluster, and usually the text pixels share the same color. These cues indicate that the text pixels in the document image are statistically and spatially correlated. Based on this information, we proposed a method for deriving a dynamic window size determination and for determining the correct k value required by Eq. (7). Therefore, the proposed method calculates the global mean, global standard deviation, and average of the maximum and minimum values in the enhanced image. The rules using global values are derived from above-mentioned parameters, as presented in Eqs. (8) until (10) which utilize the representations R_1, R_2, and R_3. Three relationships are estimated from these features, which are used to establish the rules as follows,

$$R_1 = \frac{\mu_g}{\sigma} \tag{8}$$

$$R_2 = \mu_g - \frac{\frac{\sum_{i=1}^{n} max}{n} + \frac{\sum_{i=1}^{n} min}{n}}{2}, \tag{9}$$

$$R_3 = \sigma, \tag{10}$$

where (μ_g) denotes the global mean; (σ) denotes the global standard deviation; and (n) denotes the total number of maximum (*max*) and minimum (*min*) peak values in the gray image. Fig. 5a presents an example of the maximum and minimum peak values in the histogram.

The window sizes are generated using a hierarchical rule-based approach by generating the window size at two levels, with each level presenting a type of degradation. The established rules were based on the properties of the image which were estimated to determine the window size and the values of several parameters, namely, the mean, standard deviation, maximum, and minimum, for the grayscale image. For each termination level, each rule presents a degraded case suitable for a specific window size generated. Given H_s and W_s are the width and height of the source image, the window size, W_{size} can be formulated as below:

$$W_{size} = \left\| \frac{1}{5} \left[\frac{H_s}{a} \times \frac{W_s}{b} \right] \right\|, \quad \forall a \in (1,5), \ \forall b \in (1,11), \tag{11}$$

We define the window size in two levels. The large window size is used to obtain the text with high contrast, large text, and text with no bleed-through on the first level. On the second level, all of the other cases have a smaller window size. The small window size is used especially for degraded document images or those with severe noise to avoid thin pen stroke and medium to low contrast problems. Additionally, an efficient, dynamic fixed parameter is utilized to control this problem by multiplying the global threshold by a dynamic fixed value determined by similar W_{size} features. Using these features, the proposed method identifies four different text cluster areas

that are common in the case of degraded document images, namely, high-contrast variation, low-contrast variation, large text with high density, and text with bleed-through. The proposed method determines the W_{size} and k values based on the image complexity, as presented in Algorithm 1.

Algorithm 1: Dynamic window size, W_{size} generation and parameter k value determination

---

Input: Pre-processed image
Output: Window size, W_{size} and K value
Start
Step 1: Get H_s and W_s from height and width of source image.
Step 2: Set a $= 5$, b $= 11$ and K $= 0.1$. //*Rule 1: High Contrast Variation where* $\|R_3 - R_1\| < R_3$
Step 3: if $\|R_3 - R_1\| > R_3$ **then** //*Rule 2: Very Low Contrast with thin bin stroke*
Step 4:   **if** $R_2 < R_3$ **or** $(R_1)^2 > R_2$ **then** //*Large Text With High Red Pixels Density*
Step 5:   {**if** $(R_1)^2 < R_2$ **then** K $= 0.16$ //*Rule A1*
Step 6:   **else** a $= 1$, b $= 1$} //*Rule A2*
Step 7:   **if** $R_2 < R_3$ **or** $(R_1)^2 > R_2$ **then** //*Text Reflection, Spot Ink And Ink Bleeding*
Step 8:   {**if** $R_2 < R_3$ **then** K $= 0.16$ //*Rule B1*
Step 9:   **else** K $= 0.38$} //*Rule B2*
Step 10: Set window size, W_{size} to $\left\| \dfrac{1}{5} \left[\dfrac{H_s}{a} \times \dfrac{W_s}{b} \right] \right\|$.
End.

This hierarchical rules determine the suitable window size, W_{size} and the parameter value (K), where $\forall k, k \in [0-1]$. For each type or level of noise, we used different basic statistical feature relationships. The important thing is the small window size; remove noise, and the small parameter values to retrieve text pixels removed by the smallest window size. In a large window size, the pixels are greater than the object itself, and thus, a large parameter value removes unwanted pixels surrounding the object border without affecting the object pixels. If the image satisfies rule 2 in level 1, then the image has low contrast and thin pen stroke, and the window size and parameter value in this level are suitable for solving this problem. If rule 2 is not satisfied in level 1, then the image has more challenges, and the window size and parameter value in this level are not suitable; thus, level 2 is preferable. In level 2, there are four rules (A1, A2, B1, B2). Each rule has a window size and parameter a, b, k values suitable for the degradation type, and we determine the appropriate rule for each image based on its characteristics. Rules that are dependent on these characteristics have been established, and Fig. 6 presents the effect of each rule on the given example of degraded image.

3.4 Adaptive Compound Threshold for Binarization

In this section, we propose compound thresholding using global and local values as expressed in Eq. (12). To determine the adaptive compound threshold using local and global information, we defined an operation as in Eq. (13), where the local mean calculated using local information convolves with the complement of the global threshold value. As this operation involves a dynamic window size, global information, and local information, we named it adaptive compound threshold. This method is new in the field of document analysis.

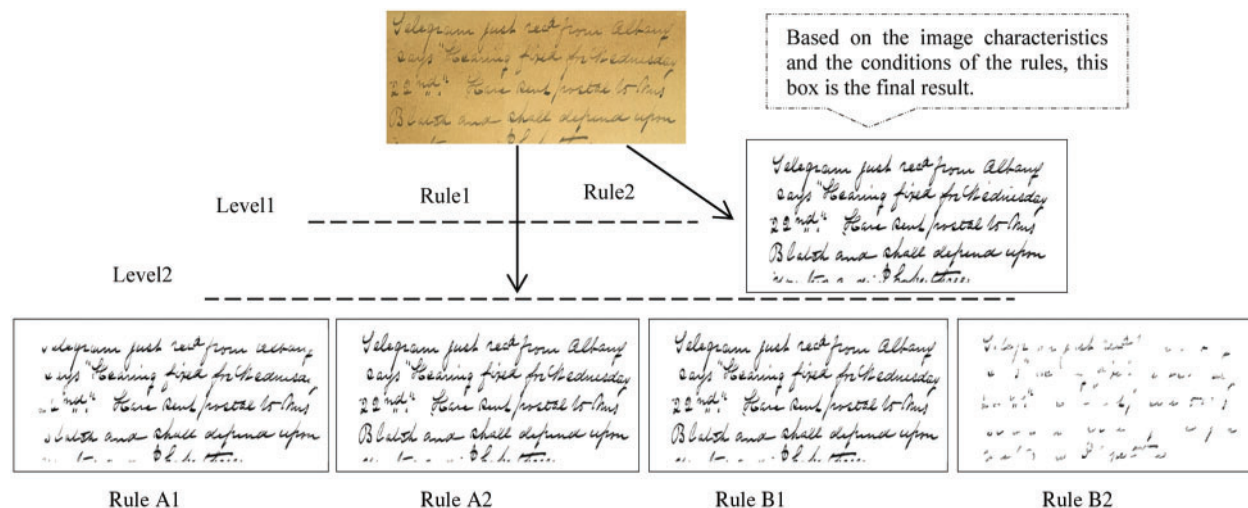$$M_L = \frac{I(x,y)}{W_{size}}. \tag{12}$$

Figure 6: Degraded image that shows the process and the effect of each rule for level 1 and 2 on the image

The local information M_L is computed using Eq. (8), where $I(x, y)$ denotes the enhanced image. The window size was provided in the previous section. With this step, only a small amount of local information can be obtained, and the same information is used for calculating the mean (M_L):

$$TH = M_L * (1 - T_g). \tag{13}$$

The adaptive compound threshold is used for the binarization of degraded document images, as expressed in Eq. (14), which yields a binarized image.

$$I_{binary}(x, y) = \{black, I(x, y) < TH; \; white, otherwise \tag{14}$$

The effect of the adaptive compound thresholding is presented in Fig. 7e where good results are achieved for the input degraded image with the chosen window size and k value. Fig. 7a is the source image, whereas Figs. 7b–7e are presenting the effect of binarization images with different window sizes and k values. The results in Fig. 7e are more improved than those in Figs. 7a–7d. From this, we can see that the proposed adaptive compound threshold works well in degraded historical document images, as presented in Figs. 8a–8d. Due to the difficulty in handling a degraded document image, some noise may be introduced during binarization. Therefore, we proposed the median filter to reduce the noise and preserve the edges, as well as the morphological filter to remove small areas that are not considered as part of the text region for the binarized image; this step is referred to as post-processing.
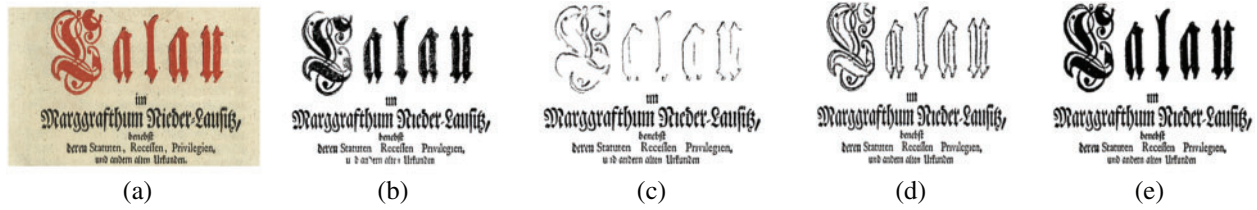
(a)   (b)   (c)   (d)   (e)

Figure 7: Examples of showing the effect of binarization images when its height $H_s = 493$, and width, $W_s = 1153$ based on different parameter values (K) substituting into dynamic window size formula, $W_{size} = \left\| \frac{1}{5} \left[\frac{H_s}{a} \times \frac{W_s}{b} \right] \right\|$, (a) source image (b) K = 0.3, a = 1, b = 1, $W_{size} = [99 \times 231]$ (c) K = 0.3, a = 5, b = 11, $W_{size} = [45 \times 105]$ (d) K = 0.16, a = 1, b = 1, $W_{size} = [99 \times 231]$ and (e) K = 0.16, a = 5, b = 11, $W_{size} = [45 \times 105]$
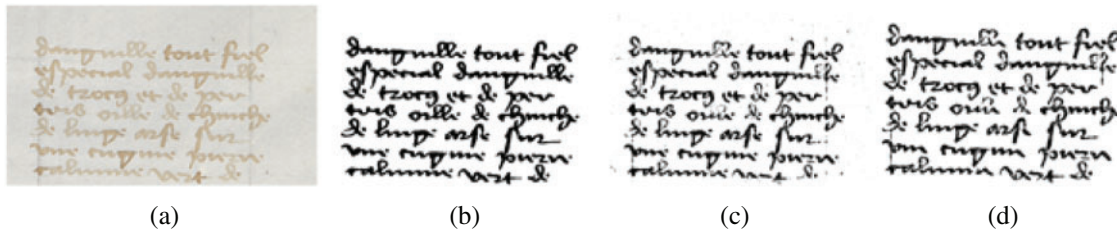

(a)   (b)   (c)   (d)

Figure 8: Binarization result: (a) Original image, (b) groundtruth image, (c) binary image without post-processing, (d) binary post-processing: Median filter and opening area

4 Experimental Results

This section presents the datasets and performance measurement for evaluating the proposed binarization method and the qualitative and quantitative results of the proposed binarization method compared to existing methods.

4.1 Datasets and Evaluation Measures

To evaluate the proposed binarization method, we used standard benchmark databases available in the literature, namely, DIBCO 2009, DIBCO 2012, and DIBCO 2014 [42–44]. These datasets comprise degraded document images suffering from degradations, such as smear, smudge, bleed-through, and low contrast. The DIBCO 2009 dataset [41] contains 10 testing images consisting of 5 degraded handwritten documents and 5 degraded printed documents; the DIBCO 2012 dataset [43], 14 degraded handwritten documents; and the DIBCO 2014 dataset [44], 10 degraded handwritten document images. To measure the performance of the proposed binarization method, we used standard measures similar to those used in the evaluation of degraded document images in the literature [39–44], namely, the F-measure, peak signal-to-noise ratio (PSNR), negative rate metric (NRM), and the Friedman test. The definitions and details of the measures are as follows:

F-measure: F-measure is the harmonic mean of precision and recall. Let Ntp, Nfn, and Nfp denote the true positive, false negative, and false positive values, respectively. This measure combines precision and recall as follows:

$$FM = \frac{2 * Recall * Precision}{Recall * Precision} \text{ where } Recall = \frac{Ntp}{Ntp + Nfn}, \ Precision = \frac{Ntp}{Ntp + Nfp} \qquad (15)$$

PSNR: PSNR measures the similarity of an image to another. Therefore, the higher the PSNR value, the higher the similarity of the two $M \times N$ images. We consider that the difference between the foreground and background pixels equals to C.

$$PSNR = \left(\frac{C^2}{MSE}\right) \text{ where } MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y) - Ibinary(x,y))^2}{M*N} \tag{16}$$

NRM: NRM represents the relationship (defined in Eq. (18)) between the ground truth pixels and the binarized image pixels (Ntp, Nfn, Nfp, and Ntn indicate the number of true positives, false positives, false negatives, and true negatives, respectively). Therefore, a low NRM value indicates a higher similarity between the two images.

$$NRM = \frac{\frac{Nfn}{Nfn+Ntp} + \frac{Nfp}{Nfp+Ntn}}{2} \tag{17}$$

The Friedman test: The Friedman test is a non-parametric test for evaluating the difference between several related algorithms. This measure is calculated as follows:

$$FQ = \frac{12}{Ni*Na(Na+1)} \sum_{j=1}^{Na} S_j^2 - 3Ni(Na+1) \tag{18}$$

where Na denotes the number of algorithms (treatments); Ni, the number of dataset images; and S_j, the sum of the ranks for the jth algorithms. The null hypothesis is rejected when $Q > X^2(Na-1)$. Where the null hypothesis is present, there is no difference between the algorithms. If the null hypothesis is rejected, then a significant difference exists between the results of the binarization algorithm.

To determine the effect of the proposed binarization method, we compared it with state-of-the-art methods, namely, the Niblack, Sauvola, NICK, Bataineh, and Lazzara MS_k [12,15–18] methods. These methods are classified as local and simple methods that heavily rely on window size. Moreover, they are considered as the most commonly used methods for binarization. Thus, these methods have been chosen for a comparative process.

4.2 Experiments for Validating the Effect of the Dynamic Window Size Determination

The proposed method employs the experimental approach to study the window size effect, unlike existing methods that typically use a fixed window size for binarization. The experimental results of the window size used in [12,15–18] and adoption of our proposed dynamic window size are reported in Tabs. 1a to 1c for DIBCO-2009, DIBCO-2012, and DIBCO-2014. The experimental results also indicate that the proposed dynamic window size determination improves the outcomes in terms of the F-measure, PSNR, and NRM compared with the default window used by the existing methods. Therefore, it can be concluded that the proposed dynamic window helps improve the current methods for the binarization of document images.

4.3 Evaluation of the Proposed Binarization Method

Sample qualitative results of the proposed and existing methods for the input images in Fig. 9 are presented in Figs. 10a–10f, which indicates that the proposed method yields better results than the existing Niblack, Sauvola, NICK, Bataineh, and Lazzara MS_k methods [12,15–18]. The main

reason why the Niblack method failed is the low variation in the background regions and unseen nearby text. For the Sauvola method, the text stroke width had low variations, which leads to considering some of the text region pixels to be part of the background, especially in the images that have very low contrast or bleed-through. The NICK method yielded good results; however, the disadvantage of this method is that it is inefficient in some cases, such as in images that have very low contrast, variations in text size, or thin pen stroke with low contrast. The Lazzara MS_k method does not exceed the performance of the Sauvola method in exceptional cases; however, it has demonstrated improvements in some of the low-contrast images owing to its dynamic window size. The Bataineh method yielded better results than the previous ones [15–18] as it can solve most of the problems in binarization; its only disadvantage is its inefficiency in images with very low contrast. The quantitative results of the proposed and existing methods are reported in Tab. 3 using DIBCO (2009, 2010, and 2014) databases.

Table 1: The F-measure, PSNR, and NRM performances of the proposed by the author, fixed and our proposed window size generation, Niblack, Sauvola, NICK, Bataineh, and Lazzara_ MS_k methods for the (a) DIBCO 2009, (b) DIBCO12, and (c) DIBCO 2014 datasets in three experiments

| Window method	Binarization method	Proposed by authors			Fixed window (20 × 20)			Our proposed window generation		
Database	Methods	FM	PSNR	NRM	FM	PSNR	NRM	FM	PSNR	NRM
(a) DIBCO2009	Proposed	91.03	18.47	6.05	91.03	18.47	6.05	91.03	18.47	6.05
	Niblack	41.67	6.09	16.71	39.59	7.05	18.52	47.69	7.23	12.59
	Sauvola	81.79	17.81	13.95	84.31	16.61	10.77	87.06	16.49	6.51
	NICK	79.89	15.4	14.2	80.4	15.48	13.75	85.22	16.66	9.2
	Bataineh	88	17.22	6.32	84.97	11.15	8.84	87.81	17.09	6.25
	Lazzara_MS_k	78.09	15.24	7.34	78.08	15.24	7.45	78.32	15.31	7.35
(b) DIBCO2012	Proposed	91.16	19.43	5.13	91.19	19.48	5.4	91.16	19.43	5.13
	Niblack	32.21	5.71	17.23	30.56	5.45	18.86	37.66	6.7	14.2
	Sauvola	71.98	16.34	19.62	76.64	16.76	16.77	81.17	17.14	13.15
	NICK	70.79	16.04	20.31	71.64	16.05	19.8	76.5	16.49	16.52
	Bataineh	86.88	17.88	8.28	83.87	16.55	4.4	88.95	18.48	6.54
	Lazzara_MS_k	69.56	15.12	17.29	69.74	15.2	17.41	70.41	15.28	17.05
(c) DIBCO2014	Proposed	91.61	18.5	6.85	90.58	18. 5	6.85	91.61	18.5	6.85
	Niblack	45.67	6.75	15.12	41.88	6.07	17.37	43.27	6.75	15.12
	Sauvola	82.73	17.1	12.38	60.68	13.67	25.79	84.73	17.1	12.38
	NICK	76.1	15.39	17.3	78.55	15.79	15.54	76.1	15.39	17.3
	Bataineh	87.07	18.15	8.72	85.86	17.23	10.9	87.07	18.15	8.72
	Lazzara _MS_k	80.08	16.96	13.14	80.02	16.96	13.18	80.29	17.03	13.01
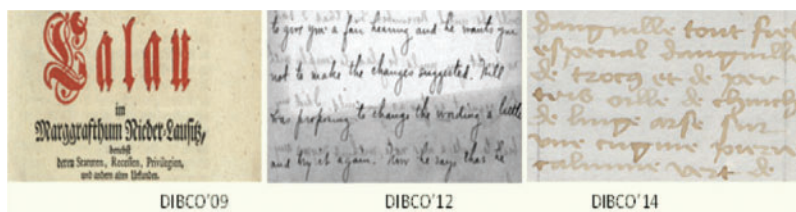
Figure 9: Three example images obtained from the three datasets used in this experiment: DIBCO (2009, 2012, and 2014) [15,16,21]

(a)   (b)

(c)   (d)

(e)   (f)
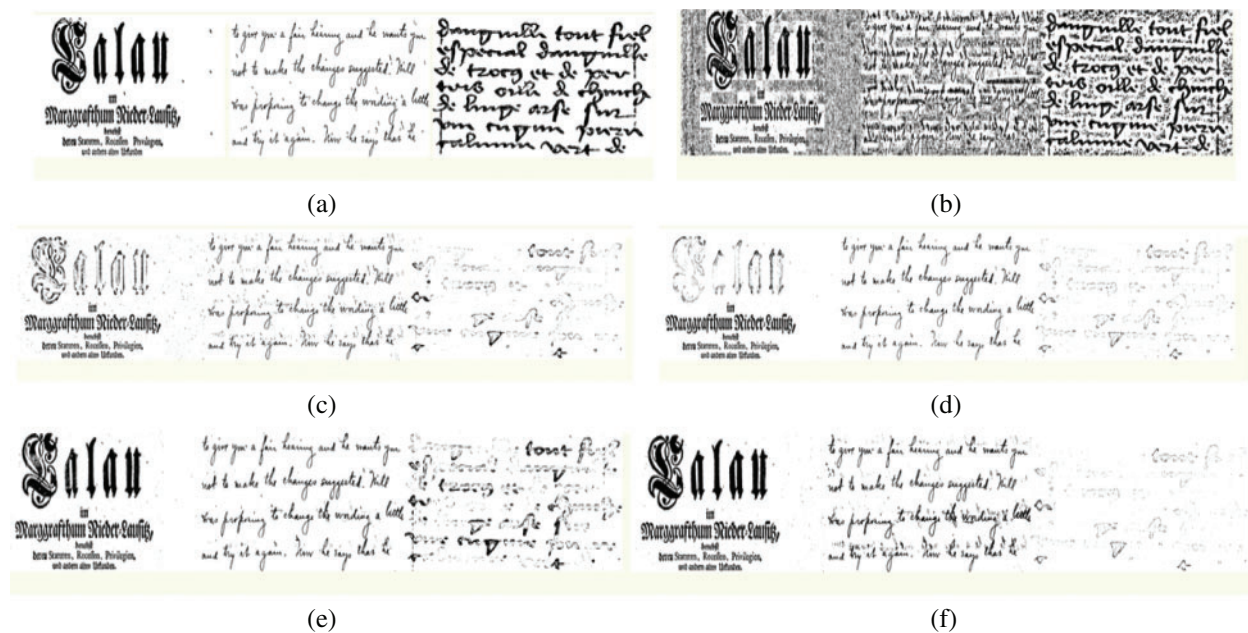
Figure 10: Example of the binarization results after their application to the images in Fig. 9: (a) Proposed, (b) Niblack [16], (c) Sauvola [17], (d) NICK [15], (e) Bataineh [12], and (f) Lazzara MS_k [18]

Tab. 2 demonstrates that the proposed method outperforms the existing methods in terms of the F-measure and PSNR for all three databases. This is mainly because the proposed method takes advantage of the adaptive compound threshold and dynamic window size determination, whereas the existing methods usually have a constant threshold and fixed window size.

Table 2: Experimental result for the different state-of-the-art methods consisting of simple local and adaptive compound methods

	DIBCO (2009)		DIBCO (2012)		DIBCO (2014)	
Methods	FM	PSNR	FM	PSNR	FM	PSNR
Proposed	91.03	18.47	92.20	19.43	91.61	18.50
Lazzara_MS_k [25]	78.08	15.24	69.56	15.12	80.08	16.96
Bataineh [12]	88.01	17.22	86.88	17.88	87.07	18.15
NICK [15]	79.48	15.14	73.85	16.10	76.10	15.39
Niblack [16]	41.67	06.09	32.21	05.71	45.67	06.75
Sauvola [17]	63.25	13.88	50.50	14.71	82.73	17.10
AdOtsu [36]	91.57	18.90	89.29	18.59	93.39	19.51
Howe [32]	94.70	22.60	90.16	18.67	–	–
Su [42]	91.07	18.50	88.90	19.60	–	–

In addition to the comparison presented above, our proposed method is also compared with the previous methods using the DIBCO 2009 and DIBCO 2012 datasets via the Friedman test. The Friedman test is a non-parametric test for evaluating the difference between several related binarization methods in terms of the data's F-measure (Proposed, Lazzara_MS_k, Bataineh, NICK, Niblack, and Sauvola). The analysis provides a test of the hypothesis that each of the five algorithms used insignificantly. It is concerning the F-measure generated by the alternative hypothesis that has the five methods used differ significantly. The following tables present a summary of the Friedman test results, i.e., Tab. 3a for DIBCO 2009, 3b for DIBCO 2012, and 3c for DIBCO 2014.

Table 3: The Friedman test to determine the F-measure for (a) DIBCO 2009, (b) DIBCO 2012, and (c) DIBCO 2014

| Methods | (a) DIBCO 2009 | | | | (b) DIBCO 2012 | | | | (c) DIBCO 2014 | | | |
	Count images	M-rank	Mean	Var.	Count images	M-rank	Mean	Var.	Count images	M-rank	Mean	Var.
Proposed	10	55	91.027	10.0	14	83	91.2	4.3	10	49	91.61	6.6
Lazzara_MS_k	10	42	78.083	342.2	14	44	69.6	335.5	10	35	80.08	670.6
Bataineh	10	45	88.003	26.3	14	67	86.9	14.3	10	48	87.07	244.7
NICK	10	33	79.482	64.4	14	48	3.8	186.2	10	21	76.10	520.2
Niblack	10	12	41.667	339.0	14	21	32.2	51.7	10	14	45.67	213.0
Sauvola	10	23	63.253	554.8	14	31	50.5	1163.2	10	43	82.73	490.4
Friedman test	Q (Observed value)	Q (Critical value)	DF	Alpha	Q (Observed value)	Q (Critical value)	DF	Alpha	Q (Observed value)	Q (Critical value)	DF	Alpha
	35.03	11.07	5	.05	53.35	11.07	5	.05	30.46	11.07	5	.05

As can be seen from Tabs. 3a–3c, Q (Observed value) is greater than Q (Critical value), which means that the first hypothesis does not differ significantly and is rejected, and the alternative hypothesis that the six methods used differ significantly for the F-measure is accepted. Thus, the proposed method is considered to yield the best results about the F-measure.

5 Conclusions

In this study, we proposed a novel method for the binarization of degraded document images. We also introduced a new concept for deriving an adaptive compound threshold, which integrates the advantage of global and local information for binarization. Additionally, we proposed a new method for deriving a dynamic window size according to the complexity of the image content. Experiments were conducted to verify the effectiveness of the proposed method for dynamic window size determination compared with the existing methods. The experimental results of the existing methods using the three standard databases indicate that the proposed method outperforms the existing ones in terms of suitable measures. We have planned to extend the same methodology for testing using images of a natural scene in the near future.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Marinai, "A survey of document image retrieval in digital libraries," HAL-00111996, pp. 193–198, 2006. https://hal.archives-ouvertes.fr/hal-00111996/document.

[2] T. A. Abu-Ain, S. N. H. S. Abdullah, K. Omar and S. Z. A. Rahman, "Advanced stroke labelling technique based on directions features for arabic character segmentation," *Asia-Pacific Journal of Information Technology and Multimedia (Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik)*, vol. 8, no. 1, pp. 97–127, 2019.

[3] A. Baldominos, Y. Saez and P. Isasi, "A survey of handwritten character recognition with mnist and emnist," *Applied Sciences*, vol. 9, no. 15, pp. 31–69, 2019.

[4] K. L. Bouman, G. Abdollahian, M. Boutin and E. J. Delp, "A low complexity method for detection of text area in natural images," in *2010 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Dallas, TX, pp. 1050–1053, 2010. https://doi.org/10.1109/ICASSP.2010.5495331.

[5] A. Mishra, K. Alahari and C. Jawahar, "An MRF model for binarization of natural scene text," in *Proc. Int. Conf. on Document Analysis and Recognition*, Beijing, pp. 11–16, 2011.

[6] V. Khare, P. Shivakumara, P. Raveendran and M. Blumenstein, "A blind deconvolution model for scene text detection and recognition in video," *Pattern Recognition*, vol. 54, pp. 128–148, 2016.

[7] S. Kavitha, P. Shivakumara, G. Hemantha Kumar and C. Tan, "A robust script identification system for historical indian document images," *Malaysian Journal of Computer Science*, vol. 28, no. 4, pp. 283–300, 2015.

[8] B. Su, S. Tian, S. Lu, T. A. Dinh and C. L. Tan, "Self-learning classification for degraded document images by sparse representation," in *Proc. Int. Conf. on Document Analysis and Recognition*, Washington, DC, pp. 155–159, 2013.

[9] F. Jia, C. Shi, K. He, C. Wang and B. Xiao, "Degraded document image binarization using structural symmetry of strokes," *Pattern Recognition*, vol. 74, no. 2, pp. 225–240, 2018.

[10] R. Hedjam, R. F. Moghaddam and M. Cheriet, "A spatially adaptive statistical method for the binarization of historical manuscripts and degraded document images," *Pattern Recognition*, vol. 44, no. 9, pp. 2184–2196, 2011.

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

[12] B. Bataineh, S. N. H. S. Abdullah and K. Omar, "An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows," *Pattern Recognition Letters*, vol. 32, no. 14, pp. 1805–1813, 2011.

[13] B. Su, S. Lu and C. L. Tan, "Robust document image binarization technique for degraded document images," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1408–1417, 2013.

[14] B. Bataineh, S. N. H. S. Abdullah and K. Omar, "Adaptive binarization method for degraded document images based on surface contrast variation," *Pattern Analysis and Applications*, vol. 20, no. 3, pp. 639–652, 2017.

[15] K. Khurshid, I. Siddiqi, C. Faure and N. Vincent, "Comparison of niblack inspired binarization methods for ancient documents," *Proc. Document Recognition and Retrieval XVI*, vol. 7247, pp. 72470–72479, 2009.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23–27, 1975.

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

[18] G. Lazzara and T. Géraud, "Efficient multiscale Sauvola's binarization," *Proc. Int. Conf. on Document Analysis and Recognition*, vol. 17, no. 2, pp. 105–123, 2017.

[19] S. Bhowmik, R. Sarkar, B. Das and D. Doermann, "Gib: A game theory inspired binarization technique for degraded document images," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1443–1455, 2018.

[20] M. Diem and R. Sablatnig, "Recognizing characters of ancient manuscripts," *Computer Vision and Image Analysis of Art*, vol. 7531, pp. 753106–753112, 2010.

[21] D. Kumar and A. Ramakrishnan, "Power-law transformation for enhanced recognition of born-digital word images," in *Proc. Int. Conf. on Signal Processing and Communications*, Bangalore, pp. 1–5, 2012.

[22] L. Neumann and J. Matas, "Real-time lexicon-free scene text localization and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1872–1885, 2015.

[23] A. Raj, G. Gautam, S. N. H. S. Abdullah, A. S. Zaini and S. Mukhopadhyay, "Multi-level thresholding based on differential evolution and tsallis fuzzy entropy," *Image and Vision Computing*, vol. 91, no. 14, pp. 103792, 2019.

[24] B. M. Singh, "Efficient binarization technique for severely degraded document images," *CSI Transactions on ICT*, vol. 2, no. 3, pp. 153–161, 2014.

[25] J. Pastor-Pellicer, S. España-Boquera, F. Zamora-Martínez, M. Z. Afzal and M. J. Castro-Bleda, "Insights on the use of convolutional neural networks for document image binarization," in *Advances in Computational Intelligence*. Cham: Springer, pp. 115–126, 2015.

[26] A. Kaur, U. Rani and G. S. Josan, "Modified sauvola binarization for degraded document images," *Engineering Applications of Artificial Intelligence*, vol. 92, no. 1, pp. 103672, 2020.

[27] H. Michalak and K. Okarma, "Optimization of degraded document image binarization method based on background estimation," in *Proc. Int. Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision*, Pilsen, pp. 89–98, 2020.

[28] U. Rani, K. Amandeep and J. Gurpreet, "A new contrast based degraded document image binarization," in *Cognitive Computing in Human Cognition*. Cham: Springer, pp. 83–90, 2020.

[29] U. Rani, K. Amandeep and J. Gurpreet, "A new binarization method for degraded document images," *International Journal of Information Technology*, vol. 9, no. 1, pp. 1–19, 2019.

[30] A. Sehad, Y. Chibani, R. Hedjam and M. Cheriet, "LBP-based degraded document image binarization," in *Proc. Int. Conf. on Image Processing Theory, Tools and Applications*, Orleans, pp. 213–217, 2015.

[31] M. Wagdy, I. Faye and D. Rohaya, "Degradation enhancement for the captured document image using retinex theory," in *Proc. of the 6th Int. Conf. on Information Technology and Multimedia*, Putrajaya, pp. 363–367, 2014. https://doi.org/10.1109/ICIMU.2014.7066660.

[32] N. R. Howe, "A laplacian energy for document binarization," in *Proc. Int. Conf. on Document Analysis and Recognition*, Beijing, pp. 6–10, 2011.

[33] S. Mandal, S. Das, A. Agarwal and B. Chanda, "Binarization of degraded handwritten documents based on morphological contrast intensification," in *Proc. Third Int. Conf. on Image Information Processing*, Waknaghat, pp. 73–78, 2015.

[34] C. Adak, P. Maitra, B. B. Chaudhuri and M. Blumenstein, "Binarization of old halftone text documents," in *Proc. TENCON 2015—2015 IEEE Region 10 Conf.*, Macao, pp. 1–5, 2015.

[35] K. Ntirogiannis, B. Gatos and I. Pratikakis, "A combined approach for the binarization of handwritten document images," *Pattern Recognition Letters*, vol. 35, no. 6, pp. 3–15, 2014.

[36] R. F. Moghaddam and M. Cheriet, "Adotsu: An adaptive and parameterless generalization of Otsu's method for document image binarization," *Pattern Recognition*, vol. 45, no. 6, pp. 2419–2431, 2012.

[37] S. Lu, B. Su and C. L. Tan, "Document image binarization using background estimation and stroke edges," *Proc. Int. Journal on Document Analysis and Recognition*, vol. 13, no. 4, pp. 303–314, 2010.

[38] Y. H. Chiu, K. L. Chung, W. N. Yang, Y. H. Huang and C. H. Liao, "Parameter-free based two-stage method for binarizing degraded document images," *Pattern Recognition*, vol. 45, no. 12, pp. 4250–4262, 2012.

[39] S. N. Kalyani and A. Lalitha, "Simple and efficient method of contrast enhancement of degraded document image through binarization," *Int. Journal of Research Studies in Science, Engineering and Technology*, vol. 1, no. 8, pp. 64–70, 2014.

[40] N. Varish, A. K. Pal, R. Hassan, M. K. Hasan, A. Khan *et al.,* "Image retrieval scheme using quantized bins of color image components and adaptive tetrolet transform," *IEEE Access*, vol. 22, no. 8, pp. 117639–117665, 2020.

[41] B. Gatos, K. Ntirogiannis and I. Pratikakis, "Icdar 2009 document image binarization contest (DIBCO 2009)," in *Proc. Int. Conf. on Document Analysis and Recognition*, Barcelona, pp. 1375–1382, 2009.

[42] B. Su, S. Lu and C. L. Tan, "Binarization of historical document images using the local maximum and minimum," in *Proc. The Ninth IAPR Int. Workshop on Document Analysis Systems, DAS 2010*, June 9–11, Boston, Massachusetts, USA, pp. 159–166, 2010. https://doi.org/10.1145/1815330.1815351.

[43] B. Gatos, K. Ntirogiannis and I. Pratikakis, "DIBCO 2009: Document image binarization contest," *IJDAR*, vol. 14, pp. 35–44, 2011. https://doi.org/10.1007/s10032-010-0115-7.

[44] I. Pratikakis, B. Gatos and K. Ntirogiannis, "ICFHR 2012 Competition on Handwritten Document Image Binarization (H-DIBCO 2012)," in *2012 Int. Conf. on Frontiers in Handwriting Recognition*, Bari, pp. 817–822, 2012. https://doi.org/10.1109/ICFHR.2012.216.