

## Research Article

# Text Extraction from Historical Document Images by the Combination of Several Thresholding Techniques

Toufik Sari,<sup>1</sup> Abderrahmane Kefali,<sup>1,2</sup> and Halima Bahi<sup>1</sup>

<sup>1</sup> LabGED Laboratory, Badji Mokhtar, BP 12, 23000 Annaba, Algeria

<sup>2</sup> 08 May 1945 University, BP 401, 24000 Guelma, Algeria

Correspondence should be addressed to Abderrahmane Kefali; kefali@labged.net

Received 13 June 2014; Revised 4 September 2014; Accepted 10 September 2014; Published 29 September 2014

Academic Editor: Chengcui Zhang

Copyright © 2014 Toufik Sari et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a new technique for the binarization of historical document images characterized by deteriorations and damages making their automatic processing difficult at several levels. The proposed method is based on hybrid thresholding combining the advantages of global and local methods and on the mixture of several binarization techniques. Two stages have been included. In the first stage, global thresholding is applied on the entire image and two different thresholds are determined from which the most of image pixels are classified into *foreground* or *background*. In the second stage, the remaining pixels are assigned to *foreground* or *background* classes based on local analysis. In this stage, several local thresholding methods are combined and the final binary value of each remaining pixel is chosen as the most probable one. The proposed technique has been tested on a large collection of standard and synthetic documents and compared with well-known methods using standard measures and was shown to be more powerful.

## 1. Introduction

Binarization is an important step in the process of document analysis and recognition. It has as goal to segment the image into two classes (foreground and background in the case of document images). The resulting image is a binary image in black and white where the black represents the foreground and the white represents the background. In fact, document image binarization is critical in the sense that bad separation will cause the loss of pertinent information and/or add useless information (noise), generating wrong results. This difficulty increases for old documents which have various types of damages and degradations from the digitization process itself, aging effects, humidity, marks, fungus, dirt, and so forth, making the automatic processing of these materials difficult at several levels.

A great number of techniques have been proposed in the literature for the binarization of gray-scale or colored documents images, but no one between them is generic and efficient for all types of documents.

The binarization techniques of grayscale images may be classified into two categories: global thresholding and local

thresholding [1, 2]. Another category of hybrid methods can be added [3]. The global thresholding methods are widely used in many document image analysis applications for their simplicity and efficiency. However, these methods are powerful only when the original documents are of good quality and well contrasted and have a clear bimodal pattern that separates foreground text and background. For historical document images which are generally noisy and of poor quality, global thresholding methods become not suitable because no single threshold is able to completely separate the foreground from the background of the image since there is no sufficient distinction between the gray range of background and foreground pixels. This kind of document requires a more detailed analysis, which may be guaranteed by local methods. Local methods calculate a different threshold for each pixel based on the information of its neighborhoods. These methods are more robust against uneven illumination, low contrast, and varying colors than global ones, but they are very time consuming since a separate threshold is computed for each pixel of the image considering its neighborhoods. This calculation becomes slower when increasing the size of neighborhood considered. Hybrid

methods, in contrast, combine global and local information for segmenting the image.

In this paper, we propose a new hybrid thresholding technique for binarizing images of historical documents. The proposed approach uses a mixture of thresholding methods and it combines the advantages of the two families of techniques: the computation speed and the efficiency. The remainder of this paper is organized as follows. In Section 2, we present some existing binarization methods. Then in Section 3, we describe the proposed approach. The experiments performed and the results will be shown in Section 4, before concluding.

## 2. State of the Art

According to [4], existing methodologies for image binarization may be divided under two main strategies: *grouping based* and *thresholding based*. Thresholding based methods use global or local threshold(s) to separate the text from the background. In grouping based methods we distinguish two categories: *region based grouping* and *clustering based grouping* methods. *Region based grouping* methods are mainly based on spatial-domain region growing or on splitting and merging. Also, *clustering based grouping* methods are based on classification of intensity or color values as a function of a homogeneity criterion. However, several techniques have been employed to reach this classification: *K-means* algorithm, artificial neural networks, and so forth.

Sezgin and Sankur [5] established a classification of binarization methods according to the information that they exploit in 6 categories.

- (i) Histogram-based methods: the methods of this class perform a thresholding based on the form of the histogram.
- (ii) Clustering-based methods: these methods assign the image pixels to one of the two clusters: object and background.
- (iii) Entropy-based methods: these algorithms use the information theory to obtain the threshold.
- (iv) Object attribute-based methods: they find a threshold value based on some similarity measurements between original and binary images.
- (v) Spatial binarization methods: they find the optimal threshold value taking into account spatial measures.
- (vi) Locally adaptive methods: these methods are designed to give a new threshold for every pixel. Several kinds of adaptive methods exist. We find methods based on local gray range, local variation, and so forth.

We present in this section some binarization methods, the most frequently cited in the literature, and we consider only thresholding based methods.

**2.1. Global Methods.** Note  $I$  the grayscale image of which the intensities vary from 0 (black) to 1 (white) and  $H$  its

histogram of intensities. The number of pixels having a gray level  $i$  is denoted as  $H(i)$ .

**2.1.1. Otsu's Method.** Otsu's method [6] tries to find the threshold  $T$  which separates the gray-level histogram in an optimal way into two segments (which maximize the intersegments variance or which minimize the intrasegments variance). The calculation of the interclasses or intraclasses variances is based on the normalized histogram  $H_n = [H_n(0) \cdots H_n(255)]$  of the image, where  $\sum H_n(i) = 1$ .

The interclasses variance for each gray level  $t$  is given by

$$V_{\text{inter}} = q_1(t) \times q_2(t) \times [\mu_1(t) - \mu_2(t)]^2 \quad (1)$$

such that

$$\begin{aligned} \mu_1(t) &= \frac{1}{q_1(t)} \sum_{i=0}^{t-1} H_n(i) \times i, \\ \mu_2(t) &= \frac{1}{q_2(t)} \sum_{i=t}^{255} H_n(i) \times i, \\ q_1(t) &= \sum_{i=0}^{t-1} H_n(i), \quad q_2(t) = \sum_{i=t}^{255} H_n(i). \end{aligned} \quad (2)$$

**2.1.2. ISODATA Method.** Thresholding using ISODATA [7] consists in finding a threshold by separating iteratively the gray-level histogram into two classes, with the a priori knowledge of the values associated with each class. This method starts by dividing the interval of nonnull values of the histogram into two equidistant parts, and next we take  $m_1$  and  $m_2$  as the arithmetic average of each class. Repeat until convergence the calculation of the optimal threshold  $T$  as the closest integer to  $(m_1 + m_2)/2$  and update the two averages  $m_1$  and  $m_2$ .

**2.1.3. Kapur et al.'s Method.** Kapur et al.'s method [8] is an entropy based method which takes into account the foreground likelihood distribution  $P_f$  and the background likelihood distribution ( $P_b = 1 - P_f$ ) in the determination of the division entropy. The binarization threshold  $T$  is chosen for which the value  $E = E_f + E_b$  is maximal, such that

$$\begin{aligned} E_f &= - \sum_{i=0}^t \frac{P_i}{P_f} \times \log \left( \frac{P_i}{P_f} \right), \\ E_b &= - \sum_{i=t+1}^{255} \frac{P_i}{1 - P_f} \times \log \left( \frac{P_i}{1 - P_f} \right), \end{aligned} \quad (3)$$

where  $P_i$  is the occurrence probability of the gray level  $i$  in the image and  $P_f = \sum_{i=0}^t P_i$ .

**2.1.4. Iterative Global Thresholding (IGT).** The proposed method selects a global threshold to the entire image based on an iterative procedure [9]. At each iteration  $i$ , the following steps are performed:

- (a) calculating the average gray level ( $T_i$ ) of the image;
- (b) subtracting  $T_i$  from all pixels of the image;

- (c) histogram equalization to extend the pixels over the whole gray levels interval.

The algorithm stops when  $|T_i - T_{i-1}| < 0.001$ .

**2.2. Local Methods.** Local methods compute a local threshold for each pixel by sliding a square or rectangular window over the entire image.

**2.2.1. Bernsen's Method.** It is an adaptive local method [10]. Thus for each pixel of coordinates  $(x, y)$ , the threshold is given by

$$T(x, y) = \frac{Z_{\text{low}} + Z_{\text{high}}}{2} \quad (4)$$

such that  $Z_{\text{low}}$  and  $Z_{\text{high}}$  are the lowest and the highest gray levels, respectively, in a squared window  $w \times w$  centered over the pixel  $(x, y)$ .

However, if the local contrast  $C(x, y) = (Z_{\text{high}} - Z_{\text{low}})$  is below a threshold  $l$  ( $l = 15$ ), then the neighborhood consists of a single class: *foreground* or *background*.

**2.2.2. Niblack's Method.** The local threshold  $T(x, y)$  is calculated using the mean  $m$  and standard deviation  $\sigma$  of all pixels in the window (neighborhood of the pixel in question) [11]. Thus, the threshold  $T(x, y)$  is given by

$$T(x, y) = m + k \times \sigma \quad (5)$$

such that  $k$  is a parameter used for determining the number of edge pixels considered as object pixels and takes a negative values ( $k$  is fixed  $-0.2$  by authors).

**2.2.3. Sauvola and Pietikäinen's Method.** Sauvola and Pietikäinen's algorithm [3] is a modification of that of Niblack in order to give more performance in the documents with a background containing a light texture or too variation and uneven illumination. In the modification of Sauvola, the local binarization threshold is given by

$$T(x, y) = m \times \left( 1 - k \times \left( 1 - \frac{\sigma}{R} \right) \right), \quad (6)$$

where  $R$  is the dynamic range of the standard deviation  $\sigma$  and the parameter  $k$  takes positives values in the interval  $[0.2, 0.5]$ .

**2.2.4. Nick Method.** This method improves considerably the binarization of lighted images and low contrasted images, by downwards moving, the threshold of binarization [2]. The threshold calculation is done as follows:

$$T(x, y) = m + k \sqrt{\frac{(\sum p_i^2 - m^2)}{\text{NP}}} \quad (7)$$

such that  $k$  is the Niblack factor and varies between  $-0.1$  and  $-0.2$  according to the application need,  $m$  is the average gray level,  $p_i$  is the gray level of pixel  $i$ , and NP is the total number of pixels. In their tests, the authors used a window of size  $19 \times 19$ .

**2.2.5. Sari et al's Method.** This method uses an artificial neuron network of multilayer perceptron (MLP) type to classify the image pixels into two classes: *foreground* and *background* [12]. The MLP has one hidden layer, 25 inputs, and one single output. To assign a new value (black or white) to a pixel, the MLP takes as input a vector of 25 values corresponding to the intensities of the pixel in a  $5 \times 5$  window centered on the processed pixel. The MLP parameters (structure, the input statistics, etc.) have been chosen after several experiments.

### 2.3. Hybrid Methods

**2.3.1. Improved IGT Method.** This method [13] is an improvement of IGT technique from [9] and it consists of two passes. In the first pass, a global thresholding is applied to the entire image and in the second pass a local thresholding processes areas still containing noise. To do this, the binary image resulting from global thresholding is divided into several segments of size  $n \times n$  and for each segment the frequency  $f$  of black pixels is calculated. The segments satisfying the following criteria are kept:  $f > \mu + k \times \sigma$  such that  $\mu$  and  $\sigma$  denote the mean and the standard deviation of the black pixels frequency in the segment, and  $k$  is a constant (equal to 2 according to the authors). For each detected area, the IGT method is applied to the corresponding area in the original image. Areas of size  $50 \times 50$  give good results according to the authors.

**2.3.2. Gangamma and Srikanta's Method.** Gangamma and Srikanta [14] proposed a method based on a simple and effective combination of spatial filters with gray scale morphological operations to remove the background and improve the quality of historical document image of palm scripts. The first step of this technique is to apply adaptive histogram equalization (AHE) to overcome the problem of uneven illumination in the document image. On the resulting image, a morphological opening operation is applied and the opened image is added later with the histogram equalized image. After that, the morphological closing operation is applied to the image for smoothing. The histogram equalized image is subtracted from the smoothed image and the result is subtracted again from the previous addition image. A Gaussian filter is subsequently applied in order to remove the noise. A final improvement is obtained by adding the last image with the histogram equalized image. Finally, a global thresholding (Otsu's algorithm) is required to separate the text from the background.

**2.3.3. Thresholding by Background Subtraction.** This technique has been proposed in [15] and it consists of three steps. The background is modeled by removing the handwriting by applying a closing of the original image with a small disk as a structuring element. After that, the background is subtracted from the original image which only leaves the foreground. Finally, the resulting image is segmented using Otsu's algorithm multiplied by an empirical constant.

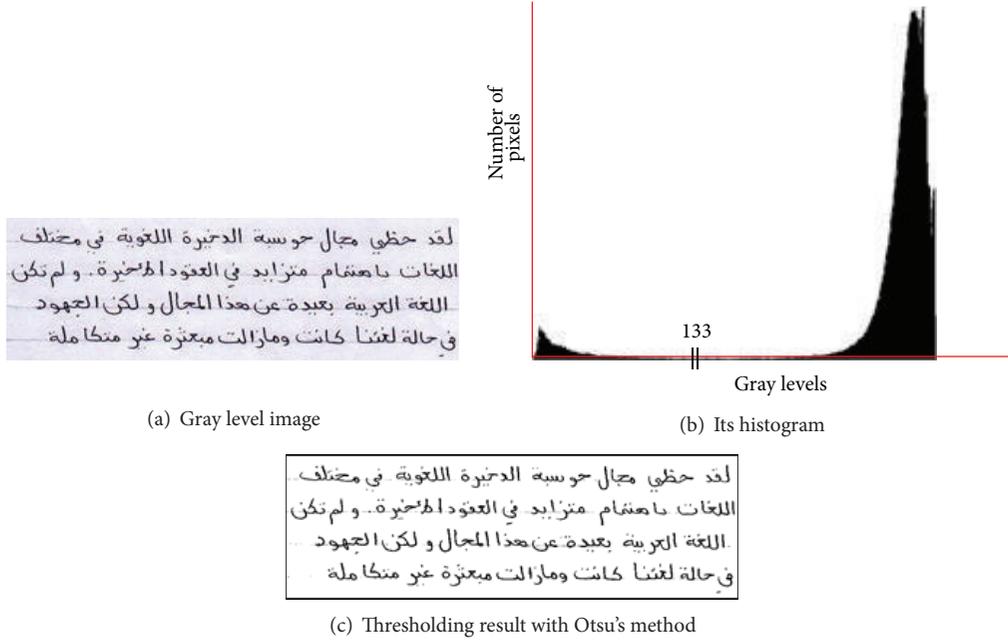


FIGURE 1: Global thresholding of a bimodal image.

**2.3.4. Tabatabaie and Bohlool's Method.** It is a nonparametric method proposed for the binarization of bad illuminated document images [16]. In this method, the morphological closing operation is used to solve the background uneven illumination problem. Indeed, closing may produce a reasonable estimate of the background if we use the appropriate structuring element. Experiments show that a structuring element of size equal to twice the stroke size gives the best results. The appropriate structuring element size is estimated as follows. A global threshold is first applied to the original image. Then we look for the size of the largest black square for each pixel and we save these values in a matrix  $M$ . The biggest value of  $M$  in each connected set of pixels is calculated and assigned to the other elements of the set. After that, the  $S$ -*histogram* ( $f$ ) is made starting from the matrix  $M$ . The value of  $f$  at the point  $i$  is equal to the number of elements of  $M$  having the value  $i$ . Finally, we determine  $x_{\max}$ , the greatest value of  $x$ , with  $x$  satisfying

$$\sum_{i=1}^x f(i) > 0.02 \times \sum_{i=1}^{\infty} f(i), \quad \sum_{i=x}^{2x} f(i) = 0. \quad (8)$$

The structuring element size will be  $2x_{\max}$ .

### 3. Proposed Technique

As we said earlier, the global thresholding techniques are generally simple and fast which tend to calculate a single threshold in order to eliminate all background pixels and preserve all foreground pixels. Unfortunately, these techniques are only applicable when the original documents are of good quality, well contrasted, and with a bimodal histogram. Figure 1 shows an example.

When the documents are of poor quality, containing different types of damages (stains, transparency effects, etc.), with a textured background and uneven illumination or when the gray levels of the foreground pixels and the gray levels of the background pixels are close, it is not possible to find a threshold that completely separates the foreground from the background of the image (Figure 2).

In this case, a more detailed analysis is needed, and we have recourse to local methods. Local methods are more accurate and may be applied to variable backgrounds, quite dark or with low contrast, but they are very slow since the threshold calculation, based on the local neighborhood information, is done for each pixel of the image. This computation becomes slower with larger sliding windows.

To solve this problem, we propose a hybrid thresholding approach that will be fast and at the same time effective as well as local methods and that is achieved by combining the advantages of both families of binarization methods. The proposed technique uses two thresholds  $T_1$  and  $T_2$  and it runs in two passes. In the first pass, a global thresholding is performed in order to class the most of pixels of the image. All pixels having a gray-level higher than  $T_2$  are removed (becomes white) because they represent the background pixels. All pixels having a gray level lower than  $T_1$  are considered as foreground pixels and therefore they are kept and colored in black. The remaining pixels are left to the second pass in which they are locally binarized by combining the results of several local thresholding methods to select the most probable value.

We detail in the following the processing steps.

**3.1. Estimation of Two Thresholds  $T_1$  and  $T_2$ .** The first step in the binarization process is the calculation of the two thresholds  $T_1$  and  $T_2$ . Since the thresholding purpose is to divide the

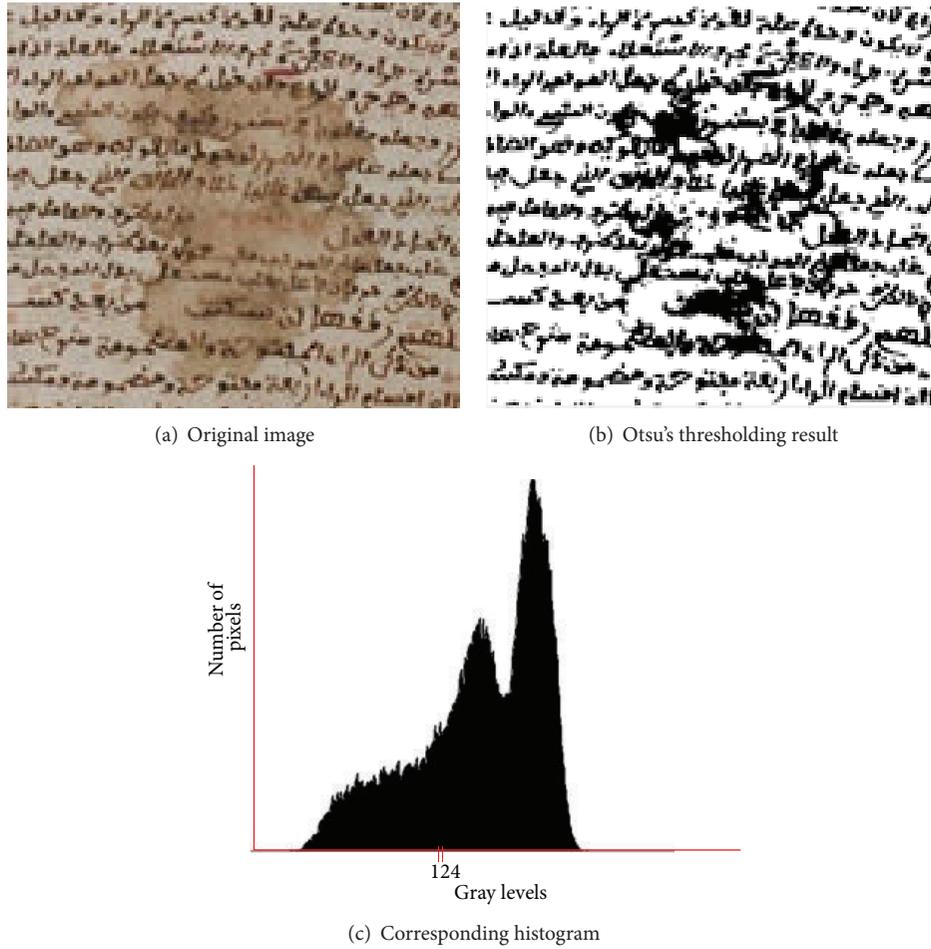


FIGURE 2: Global thresholding of degraded image.

image into two classes: foreground and background and since a single threshold is not able to accomplish this task, the use of more separation thresholds seems to be a perfect solution.

These two thresholds are estimated from the gray levels histogram of the original image and represent the average intensity of the foreground and background, respectively.

To obtain these two thresholds, we first compute a global threshold  $T$  using a global thresholding algorithm which can be Otsu's algorithm, Kapur, or any other global algorithm. In our approach, we opted for the Otsu algorithm because this technique has shown its efficiency and overcame other global methods in several comparative studies [17, 18].  $T$  separates the gray levels histogram of the image into two classes: foreground and background.

$T_1$  and  $T_2$  are estimated from  $T$ . Noting  $d_{\min}$  the minimum distance between the average intensity of the foreground is represented by the first half of the histogram and the average intensity of the background is represented by the second half:

$$\begin{aligned}
 T_1 &= T - \frac{d_{\min}}{2}, \\
 T_2 &= T + \frac{d_{\min}}{2}.
 \end{aligned}
 \tag{9}$$

**3.2. Global Image Thresholding Using  $T_1$  and  $T_2$ .** After the estimation of the two thresholds  $T_1$  and  $T_2$ , all pixels having a gray level higher than  $T_2$  are transformed into white which eliminates most of the image background, and those whose gray level is less than  $T_1$  are colored in black. Note  $I$  the resulting image. These pixels are certainly foreground pixels. The resulting image still contains some noise, but all the foreground information is preserved.

**3.3. Local Thresholding of the Remaining Pixels.** The pixels unprocessed in the previous step (those with a gray level between  $T_1$  and  $T_2$ ) may be from the foreground and thus must be preserved; likewise they may be background's or noise's pixels and should be removed. The decision to assign the remaining pixels to one of the two classes: foreground or background is performed using a local process by examining the neighborhood of these pixels. To guarantee a more correct classification, we propose to apply several local thresholding methods. In our experiments, we chose the following methods: Niblack, Sauvola, and Nick, since these methods were ranked in first places in several previous comparative studies [2, 19, 20].

For each pixel  $(x, y)$  of  $I$  not yet classified, we calculate locally its new binary values (0 for black and 1 for white)

obtained by applying Niblack's, Sauvola's, and Nick's methods and we obtain thus three temporary images  $I_1$ ,  $I_2$ , and  $I_3$ , respectively. Each one of the three local methods computes the binary value of each remaining pixel  $(x, y)$  by

$$I_i(x, y) = \begin{cases} 0, & \text{if } I(x, y) < LT_i(x, y), \\ 255, & \text{otherwise,} \end{cases} \quad (10)$$

$$1 \leq i \leq 3$$

with  $LT_1$ ,  $LT_2$ ,  $LT_3$  being the local threshold computed using Niblack's, Sauvola's, and Nick's methods, respectively.

The final binary value  $I_b(x, y)$  of each remaining pixel is that resulted of at least two of the three methods:

$$I_b(x, y) = \begin{cases} 0, & \text{if } \sum_{i=1}^3 I_i(x, y) < 2, \\ 1, & \text{otherwise.} \end{cases} \quad (11)$$

## 4. Experiments and Results

Experiments have been performed in order to estimate the performance of our approach. We applied the proposed technique over a large test set and compared the obtained results with well-known methods, including global, local, and hybrid methods. The comparison will estimate both the binarization quality and the execution time.

Firstly, for parameterized methods a series of experiments have been performed in order to set their optimal parameter values. Note  $PS(m)$  the parameters set of a specific thresholding method  $m$ . For example,  $PS(\text{Niblack}) = \{k, w\}$ ,  $PS(\text{Sauvola}) = \{k, w, R\}$ , and so forth. We try to find the optimal values of  $PS(m)$  giving the binarization results which are closest to the ground truth images. A specific range of values  $[a, b]$  is first defined for each parameter. To improve the accuracy of the process, we used a wide initial range for every parameter. For Niblack's method per example, the range of the parameter  $w$  is defined as  $[a, b] = [3, 299]$ . After that, we apply the binarization method  $m$  with the different values of  $PS(m)$  from the predefined ranges on the test set described in Section 4.1. We compare the binarization results with the ground truth images using the evaluation measures detailed in Section 4.2. A ranking of the obtained results is then performed according to each measure separately. By calculating the sum of all ranks, we can infer the optimal set of parameter values as leading to the top ranks. The optimal parameter values of the parameterized methods are summarized in Table 1.

**4.1. Test Bases.** Two test sets have been used for the evaluation of the proposed method. The first is a public set composed of document images from the four collections proposed within the context of the competitions DIBCO 2009 (<http://users.iit.demokritos.gr/~bgat/DIBCO2009/benchmark/>), H-DIBCO 2010 (<http://www.iit.demokritos.gr/~bgat/H-DIBCO2010/benchmark>), DIBCO 2011 (<http://utopia.duth.gr/~ipratika/DIBCO2011/benchmark>), and H-DIBCO 2012 ([TABLE 1: Optimal parameters values of the parameterized methods.](http://utopia.duth.gr/~ipratika/HDIBCO2012/bench-</a></p>
</div>
<div data-bbox=)

| Binarization technique      | Optimal parameter values           |
|-----------------------------|------------------------------------|
| Niblack [11]                | $k = -0.2$ and $w = 35$            |
| Sauvola and Pietikäinen [3] | $k = 0.2$ , $R = 128$ and $w = 27$ |
| Nick                        | $k = -0.1$ and $w = 19$            |
| Improved IGT                | $n = 50$ and $k = 3$               |

mark). These four collections contain a total of 50 real documents images (37 handwritten and 13 printed) coming from the collections of several libraries, with the associated ground truth images. All the images contain representative degradations which appear frequently (e.g., variable background intensity, shadows, smear, smudge, low contrast, and bleed-through). Figure 3 shows some images from these collections.

The second set of images is a synthetic collection composed of 150 synthetic images of documents constructed by the fusion of 15 different backgrounds and 10 binary images (Figure 4). The fusion is done by applying the image mosaicing by superimposing technique for blending [21]. The idea is as follows: we start with some images of documents in black and white, which represent the ground truth, and with some backgrounds extracted from old documents and we apply a fusion procedure to get as many different images of old documents. However, Stathis et al., in [22], proposed two different techniques for the blending: maximum intensity and image averaging. We adopt using the image averaging technique in order to have a more natural result.

**4.2. Evaluation Measures.** In addition to the execution time, the qualitative assessment is performed in terms of five standard evaluation measures used in DIBCO 2009, H-DIBCO 2010, DIBCO 2011, and H-DIBCO 2012 which are  $F$ -measure, PSNR, NRM, MPM, and DRD.

Note TP, TN, FP, and FN the true positive, true negative, false positive, and false negative values, respectively.

**4.2.1.  $F$ -measure.**  $F$ -measure was introduced first by Chinchor in [23]:

$$F\text{-Measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}, \quad (12)$$

where

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (13)$$

**4.2.2. PSNR.** PSNR is a similarity measure between two images. However, the higher the value of PSNR, the higher the similarity of the two images [24, 25]:

$$\text{PSNR} = 10 \cdot \log \left( \frac{C^2}{\text{MSE}} \right), \quad (14)$$



FIGURE 3: Images taken from the collections of (a) DIBCO 2009, (b) H-DIBCO 2010, (c) DIBCO 2011, and (d) H-DIBCO 2012.



FIGURE 4: Image of document obtained by fusing binary image and background. (a) Background from old document, (b) ground truth binary image, and (c) the resulting synthetic image.

where

$$\text{MSE} = \frac{\sum_{x=1}^M \sum_{y=1}^N (I(x, y) - I_2(x, y))^2}{MN}. \quad (15)$$

$I$  and  $I_2$  represent the two images matched.  $M$  and  $N$  are their height and width, respectively.  $C$  is the difference between foreground and background.

4.2.3. *NRM (Negative Metric Rate)*. NRM is based on the pixelwise mismatches between the ground truth and the

binarized image [26]. It combines the false negative rate  $\text{NR}_{\text{FN}}$  and the false positive rate  $\text{NR}_{\text{FP}}$ . It is denoted as follows:

$$\text{NRM} = \frac{\text{NR}_{\text{FN}} + \text{NR}_{\text{FP}}}{2} \quad (16)$$

with

$$\text{NR}_{\text{FN}} = \frac{\text{FN}}{\text{FN} + \text{TP}}, \quad \text{NR}_{\text{FP}} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (17)$$

Contrary to  $F$ -measure and PSNR, The better binarization quality is obtained for lower NRM value.

TABLE 2: Average results obtained on the test dataset.

|        |                                     | Execution time (ms) | F-measure (%) | PSNR         | NRM ( $\times 10^{-2}$ ) | MPM ( $\times 10^{-3}$ ) | DRD          |
|--------|-------------------------------------|---------------------|---------------|--------------|--------------------------|--------------------------|--------------|
| Global | Otsu [6]                            | 458,75              | 80,565        | 15,716       | 8.83                     | 2.961                    | 26.208       |
|        | Kapur et al. [8]                    | 520,828             | 81,921        | 15,692       | 8.109                    | 2.059                    | 23.104       |
|        | IGT                                 | 309,543             | 76,556        | 14,307       | 10.11                    | 2.778                    | 28.73        |
| Local  | Niblack [11]                        | 78504.1             | 66.88         | 7.78         | 11.47                    | 18.239                   | 43.42        |
|        | Sauvola and Pietikäinen [3]         | 73914.2             | 85.68         | 18.62        | 7.95                     | 1.254                    | 6.13         |
|        | Nick                                | 72480.9             | 82.22         | 19.047       | 9.31                     | 2.201                    | 5.27         |
| Hybrid | Improved IGT                        | 487,943             | 78,039        | 14,728       | 12.05                    | 3.72                     | 22.78        |
|        | Gangamma and Srikanta [14]          | 73999,157           | 69,187        | 13,469       | 21.71                    | 15.43                    | 38.81        |
|        | Background Subtraction thresholding | 1122,057            | 68,907        | 13,432       | 22.34                    | 17.75                    | 40.97        |
|        | Tabatabaei and Bohlool [16]         | 75875               | 85.82         | 26.69        | 8.2                      | 1.183                    | 3.439        |
|        | Proposed Approche                   | <b>1417.86</b>      | <b>87.44</b>  | <b>27.97</b> | <b>6.74</b>              | <b>1.025</b>             | <b>3.161</b> |

4.2.4. *MPM (Misclassification Penalty Metric)*. The misclassification penalty metric MPM evaluates the binarization result against the ground truth on an object-by-object basis [26]:

$$\text{MPM} = \frac{\text{MP}_{\text{FN}} + \text{MP}_{\text{FP}}}{2}, \quad (18)$$

where

$$\text{MP}_{\text{FN}} = \frac{\sum_{i=1}^{\text{FN}} d_{\text{FN}}^i}{D}, \quad \text{MP}_{\text{FP}} = \frac{\sum_{j=1}^{\text{FP}} d_{\text{FP}}^j}{D}. \quad (19)$$

$d_{\text{FN}}^i$  and  $d_{\text{FP}}^j$  denote the distance of the  $i$ th false negative and the  $j$ th false positive pixel from the contour of the ground truth segmentation. The normalization factor  $D$  is the sum over all the pixel-to-contour distances of the ground truth object. A low MPM score denotes that the algorithm is good at identifying an object's boundary.

4.2.5. *DRD (Distance Reciprocal Distortion Metric)*. DRD is an objective distortion measure for binary document images, and it was proposed by Lu et al. in [25]. This measure properly correlates with the human visual perception and it measures the distortion for all the  $S$  flipped pixels as follows:

$$\text{DRD} = \frac{\sum_{k=1}^S \text{DRD}_k}{\text{NUBN}}. \quad (20)$$

NUBN is the number of the nonuniform  $8 \times 8$  blocks in the GT image.

$\text{DRD}_k$  is the distortion of the  $k$ th flipped pixel of coordinate  $(x, y)$  and it is calculated using a  $5 \times 5$  normalized weight matrix  $W_{Nm}$ . This last is defined in [25] as follows:

$$W_{Nm}(i, j) = \frac{W_m(i, j)}{\sum_{i=1}^m \sum_{j=1}^m W_m(i, j)} \quad (21)$$

such that

$$W_m(i, j) = \begin{cases} 0, & \text{for } i = i_C, j = j_C, \\ \frac{1}{\sqrt{(i - i_C)^2 + (j - j_C)^2}}, & \text{otherwise.} \end{cases} \quad (22)$$

With  $m = 5$  and  $i_C = j_C = (1 + m)/2$ .

$\text{DRD}_k$  is given as follows:

$$\text{DRD}_k = \sum_{i=-2}^2 \sum_{j=-2}^2 |I_{\text{GT}}(x + i, y + j) - I_B(x, y)| \times W_{Nm}(i + 2, j + 2). \quad (23)$$

4.3. *Results and Discussion*. The average results obtained over the test images are summarized in Table 2. The final ranking of the compared methods is shown in Table 3, which also summarizes the partial ranks of each method according to each evaluation measure and the sum of ranks.

From Tables 2 and 3, it is clear that our proposed method is totally ranked first and has the best performances according to all measures of binarization quality. It exceeded local methods such as the famous Sauvola and Pietikainen method, which ranked 3rd in our experiments, and even other hybrid techniques. Indeed, the combination of three local thresholding techniques enabled a more robust determination of the binary value of each pixel by taking the most likely value.

Regarding the execution time, our method is very fast compared to local methods (about 52 times more than Sauvola and Pietikainen's method), which enabled us to earn about than 98% of the execution time. This is logical because only a portion of the pixels (having a gray level between the two thresholds  $T_1$  and  $T_2$ ) is locally analyzed.

## 5. Conclusion

In this paper we tackled the problem of foreground/background separation from the images of historical documents.

TABLE 3: Final ranking of the compared methods on the test set.

| Rank | Method                                 | Execution time (ms) | F-measure (%) | PSNR | NRM ( $\times 10^{-2}$ ) | MPM ( $\times 10^{-3}$ ) | DRD | Sum of ranks |
|------|--|---------------------|---------------|------|--------------------------|--------------------------|-----|--------------|
| 1    | Proposed Approche                      | 6                   | 1             | 1    | 1                        | 1                        | 1   | 11           |
| 2    | Tabatabaei and Bohlool [16]            | 10                  | 2             | 2    | 4                        | 2                        | 2   | 22           |
| 3    | Sauvola and Pietikäinen [3]            | 8                   | 3             | 4    | 2                        | 3                        | 4   | 24           |
| 4    | Kapur et al. [8]                       | 4                   | 5             | 6    | 3                        | 4                        | 6   | 28           |
| 5    | Nick                                   | 7                   | 4             | 3    | 6                        | 5                        | 3   | 28           |
| 6    | Otsu [6]                               | 2                   | 6             | 5    | 5                        | 7                        | 7   | 32           |
| 7    | IGT                                    | 1                   | 8             | 8    | 7                        | 6                        | 8   | 38           |
| 8    | Improved IGT                           | 3                   | 7             | 7    | 9                        | 8                        | 5   | 39           |
| 9    | Gangamma and Srikanta [14]             | 9                   | 9             | 9    | 10                       | 9                        | 9   | 55           |
| 10   | Thresholding by Background Subtraction | 5                   | 10            | 10   | 11                       | 10                       | 10  | 56           |
| 11   | Niblack [11]                           | 11                  | 11            | 11   | 8                        | 11                       | 11  | 63           |

We proposed a hybrid approach of degraded document images binarization. The proposed approach runs on two passes. Firstly, a global thresholding using Otsu's algorithm is applied on the entire image, and two different thresholds are determined. All pixels below the first threshold are preserved and all pixels higher than the second threshold are eliminated as they represent surely background pixels. The remaining pixels are then processed locally based on their neighborhood information. In this step, three local thresholding methods are combined in order to obtain a more accurate decision. Since the number of pixels processed locally is very small compared to the total number of pixels, the time required for the binarization is reduced considerably without decreasing the performances. To validate our approach, we compared it with the state-of-the-art methods from the literature and the obtained results on standard and synthetic collections are encouraging and confirm our approach.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] N. Arica and F. T. Yarman-Vural, "An overview of character recognition focused on off-line handwriting," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 31, no. 2, pp. 216–233, 2001.
- [2] K. Khurshid, I. Siddiqi, C. Faure, and N. Vincent, "Comparison of Niblack inspired binarization methods for ancient documents," in *16th International conference on Document Recognition and Retrieval*, Proceedings of SPIE, San Jose, Calif, USA, January 2009.
- [3] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.
- [4] B. Fernando and S. Karaoglu, "Extreme value theory based text binarization in documents and natural scenes," in *Proceedings of the 3rd IEEE International Conference on Machine Vision (ICMV '10)*, pp. 144–151, Hong Kong, December 2010.
- [5] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–168, 2004.
- [6] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [7] F. R. D. Velasco, "Thresholding using the ISODATA clustering algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 10, no. 11, pp. 771–774, 1980.
- [8] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 273–285, 1985.
- [9] E. Kavallieratou, "A binarization algorithm specialized on document images and photos," in *Proceedings of the 8th International Conference on Document Analysis and Recognition*, pp. 463–467, September 2005.
- [10] J. Bernsen, "Dynamic thresholding of grey-level images," in *Proceedings of the 8th International Conference on Pattern Recognition*, pp. 1251–1255, Paris, France, 1986.
- [11] W. Niblack, *An Introduction to Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1986.
- [12] T. Sari, A. Kefali, and H. Bahi, "An MLP for binarizing images of old manuscripts," in *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition (ICFHR '12)*, pp. 247–251, Bari, Italy, September 2012.
- [13] E. Kavallieratou and S. Stathis, "Adaptive binarization of historical document images," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 3, pp. 742–745, Hong Kong, August 2006.
- [14] B. Gangamma and M. K. Srikanta, "Enhancement of degraded historical kannada documents," *International Journal of Computer Applications*, vol. 29, no. 11, pp. 1–6, 2011.
- [15] G. Leedham, C. Yan, K. Takru, J. H. N. Tan, and L. Mian, "Comparison of Some thresholding algorithms for text/background segmentation in difficult document images," in *Proceedings of the 7th International Conference on Document Analysis and Recognition*, pp. 859–864, 2003.

- [16] S. A. Tabatabaei and M. Bohlool, "A novel method for binarization of badly illuminated document images," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 3573–3576, Hong Kong, China, September 2010.
- [17] P. K. Sahoo, S. Soltani, and A. K. C. Wong, "A survey of thresholding techniques," *Computer Vision, Graphics and Image Processing*, vol. 41, no. 2, pp. 233–260, 1988.
- [18] M. A. Ramírez-Ortegón and R. Rojas, "Unsupervised evaluation methods based on local gray-intensity variances for binarization of historical documents," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 2029–2032, IEEE Computer Society, Istanbul, Turkey, August 2010.
- [19] A. J. O. Trier, "Goal-directed evaluation of binarization methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1191–1201, 1995.
- [20] A. Kefali, T. Sari, and M. Sellami, "Evaluation of several binarization techniques for old Arabic documents images," in *Proceedings of International Symposium on Modelling and Implementation of Complex Systems*, Constantine, Algeria, 2010.
- [21] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, 1992.
- [22] P. Stathis, E. Kavallieratou, and N. Papamarkos, "An evaluation survey of binarization algorithms on historical documents," in *Proceedings of the 19th International Conference on Pattern Recognition*, pp. 742–745, December 2008.
- [23] N. Chinchor, "MUC-4 evaluation metrics," in *Proceedings of the 4th Message Understanding Conference*, pp. 22–29, 1992.
- [24] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "DIBCO 2009: document image binarization contest," *International Journal on Document Analysis and Recognition*, vol. 14, no. 1, pp. 35–44, 2011.
- [25] H. Lu, A. C. Kot, and Y. Q. Shi, "Distance-reciprocal distortion measure for binary document images," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 228–231, 2004.
- [26] J. Aguilera, H. Wildenauer, M. Kampel, M. Borg, D. Thirde, and J. Ferryman, "Evaluation of motion segmentation quality for aircraft activity surveillance," in *Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS '05)*, pp. 293–300, October 2005.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

