Efficient Binarization of Historical and Degraded Document Images

B. Gatos, I. Pratikakis and S.J. Perantonis

Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Research Center "Demokritos",
153 10 Athens, Greece
{bgat, ipratika, sper}@iit.demokritos.gr

Abstract

This paper presents a new adaptive approach for the binarization and enhancement of historical and degraded documents. The proposed method is based on (i) efficient pre-processing; (ii) the combination of the results of several state-of-the-art binarization methodologies; (iii) the incorporation of edge information and (iv) the application of efficient image post-processing based on mathematical morphology for the enhancement of the final result. The proposed method demonstrated superior performance against six well-known techniques on numerous historical handwritten and machine-printed documents mainly from the Library of Congress of the United States archive. The performance evaluation was based on a consistent and concrete methodology.

1. Introduction

Document image binarization is the initial step of most document image analysis and understanding systems that converts a grey scale image into a binary image aiming to distinguish text areas from background areas. Binarization plays a key role in document processing since its performance affects quite critically the degree of success in subsequent character segmentation and recognition.

In general, approaches that deal with document image binarization are categorised in two main classes: (i) global and (ii) local. In a global approach, threshold selection results in a single threshold value for the entire image. Global thresholding [1] has a good performance in the case that there is a good separation between the foreground and the background. However, in the case of historical documents, there exist degradations that diminish robustness of this class of binarization. Examples of degradations include shadows and non-uniform illumination, ink seeping, smear and strains. To deal with degradations, the current trend is to use local information that guides the threshold value pixelwise in an adaptive manner. Examples of adaptive binarization techniques contain the works in [2]-[6].

Most of the adaptive local binarization methods ignore the edge property and lead to erroneous results due to the creation of fake shadows. For this, there exist approaches that also incorporate edge information as in [7] wherein they find seeds near the image edges and present an edge connection method to close the image edges. Then, they use closed image edges to partition the binarized image that is generated using a high threshold, and obtain a primary binarization result by filling the partitioned high-threshold binary image with the seeds. In this way, they have achieved an effective solution in the case of low contrast, noise and non-uniform illumination.

Although the use of adaptive local binarization methods obtain satisfactory results for degraded documents, no gold standard technique exists that performs best in all degradation cases. Due to this, it has been recently proposed [8] to combine and take into account the results of a set of binarization techniques in order to use the complementarity in success of each technique. In particular, they take advantage of the benefits of a set of selected binarization algorithms by combining their results using a Kohonen self-organizing map neural network. Specifically, they follow a two stage approach, wherein in the first stage the best parameter values for each independent binarization technique are estimated while in the second stage, the neural network is fed by the binarization results obtained by those techniques using their estimated best parameter values.

978-0-7695-3337-7/08 $25.00 © 2008 IEEE
DOI 10.1109/DAS.2008.66

Motivated by the aforementioned use of edge information as well as the use of combined binarization approaches, we propose a novel methodology that is based on (i) efficient pre-processing; (ii) combination of the binarization result of several state-of-the-art methodologies; (iii) incorporation of the edge map in the grey scale image and (iv) application of efficient image post-processing based on mathematical morphology for the enhancement of the final result.

The remaining sections of the paper are structured as follows: Section 2 details all stages of the proposed methodology and experimental results are given in Section 3. Finally, conclusions are drawn in Section 4.

2. Methodology

The proposed methodology for efficient binarization of historical and degraded document images is illustrated in Fig. 1. It consists of four distinct steps. At the first step, an appropriate pre-processing based on Wiener filtering is applied. At the next step, several binarization results are combined in order to produce a binary (b/w) image taking into account the agreement in the majority of the binarization methodologies. At the next step, the edge information of the grey level image is combined with the binary result of the previous step. From all edge pixels, only those are selected that probably belong to text areas according to a criterion. An appropriate smoothing algorithm is then applied in order to fill text areas in the edge map. Finally, an enhancement step based on mathematical morphology operations is used in order to produce a high quality result while preserving stroke information. A detailed description of all steps is given in the following subsections.

2.1 Pre-processing

A pre-processing stage of the grey scale source image is essential for historical and degraded documents for the elimination of noisy areas, smoothing of background texture as well as contrast enhancement between background and text areas. The use of Wiener filtering [9] has been proved efficient for the aforementioned goals. The Wiener filter is commonly used for image restoration. Our pre-processing module involves an adaptive Wiener method based on statistics estimated from a local neighborhood around each pixel.

Consider the grey scale source image:

$$I_S(x,y) = \{0,1,\ ...,255\},\ 1 \le x \le I_x,\ 1 \le y \le I_y \tag{1}$$

where 0 corresponds to black and 255 to white.

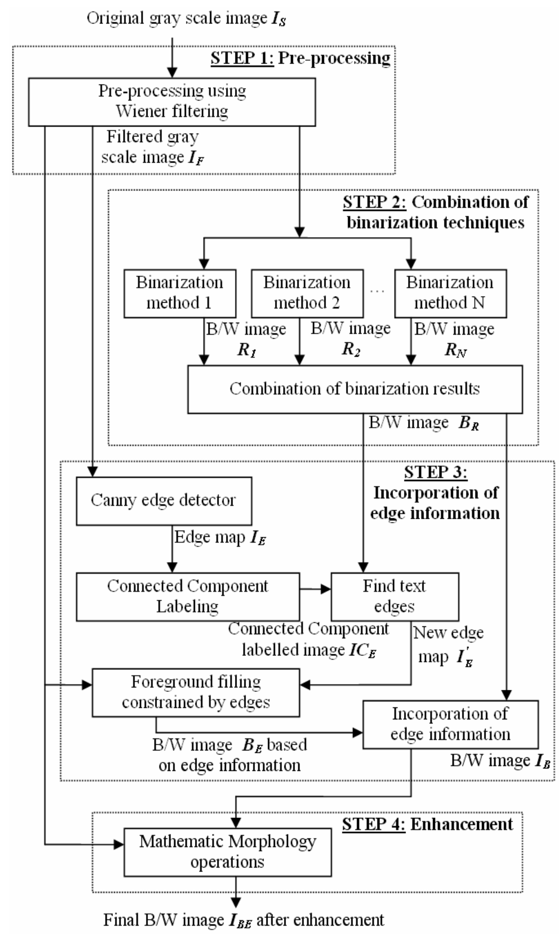
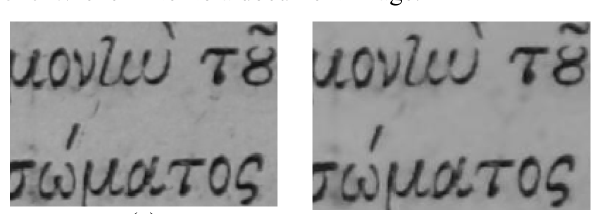


Figure 1. Block diagram of the proposed methodology

The grey scale source image I_S is transformed to the filtered grey scale image I_F according to the following formula:

$$I_F(x,y) = \mu + \frac{(\sigma^2 - \nu^2)(I_S(x,y) - \mu)}{\sigma^2} \tag{2}$$

where μ is the local mean, σ^2 the variance in a *5x5* neighborhood around each pixel and ν^2 is the average of all estimated variances for each pixel in the neighborhood. Fig. 2 shows the results of applying a 5x5 Wiener filter to a document image.



(a) (b)

Figure 2. Pre-processing example: (a) original grey scale image I_S and (b) resulting image I_F after applying a 5x5 Wiener filter.

2.2 Combination of binarization techniques

Let $R_1(x,y),\ R_2(x,y),...,R_N(x,y)$ represent the results of N different binarization methods, which have been applied to image $I_F(x,y)$. N is selected as an odd number ($N = 2m + 1$). Images R_i are defined as follows:

$$R_i(x,y) = \begin{cases} 1, & \text{foreground} \\ 0, & \text{background} \end{cases},\ \text{where } 1 \le i \le 2m+1 \qquad (3)$$

At this step, we calculate a binary image B_R which combines the N binarization results. Since we aim to mark as foreground pixels only those pixels that the majority of the binarization methodologies classify as foreground, B_R is calculated as follows:

$$B_R(x,y) = \begin{cases} 1, & \text{if } \sum_{i=1}^{2m+1} R_i(x,y) > m \\ 0, & \text{otherwise} \end{cases} \qquad (4)$$

Figure 3. Example of combining several binarization results: (a) original image I_F; (b) - (d) binarization results R_1, R_2 and R_3 respectively and (e) combined result B_R.

2.3 Incorporation of edge information

At this step of our methodology, we produce the edge map of the filtered grey scale image I_F. Several methodologies are used in the literature for computing the edge map of an image [10]. In our approach, we use the Canny edge detector [11]. Canny uses Sobel masks in order to find the edge magnitude of the grey scale image and then uses non-Maxima suppression and hysteresis thresholding. The produced edge map I_E is defined as follows:

$$I_E(x,y) = \begin{cases} 1, & \text{if } (x,y) \in \partial I_F \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

Fig. 4(a) shows the edge map I_E of a filtered grey scale image I_F.

Starting from the edge map I_E we calculate an adapted edge map $I_E^{'}$ containing only those connected components (CCs) of I_E that have a significant overlap with the binary image B_R. In this way, we exclude edges that do not belong to text areas. We assume that (i) an edge pixel of I_E overlaps with a foreground pixel from image B_R only if a B_R foreground pixel exists in the 3x3 neighbour of the edge pixel and (ii) a CC of I_E has a significant overlap with binary image B_R if more than 10% of the pixels of the CC overlaps with a foreground pixel from image B_R. An example of an edge image $I_E^{'}$ is given in Fig. 4(b).

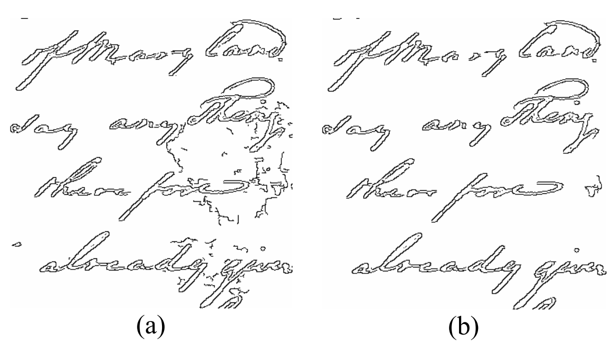


Figure 4. Example of edge images: (a) edge map I_E of the original grey scale image I_F of Fig. 3(a) after applying the Canny edge detector and (b) adapted edge map $I_E^{'}$ containing only those connected components of I_E that have a significant overlap with the binary image B_R of Fig. 3(e).

In order to incorporate the edge information at the final binarization result, we have to proceed to a foreground filling constrained by the edge image $I_E^{'}$. For this purpose, we use an extension of the Run Length Smoothing Algorithm [13] that is guided by the grey scale image I_F in order to turn white runs (successive 0's) into black runs (successive 1's) only if the white runs correspond to text areas. More specifically, a horizontal white run of the edge image $I_E^{'}$ that starts from background pixel (x_1,y) and ends at the background pixel (x_2,y) (see Fig. 5) is turned to a black run only if (i) is of short length and (ii) the corresponding average grey level value of the white run pixels is darker (of smaller value) than the corresponding average grey level value at the limits of the white run (A,B dotted rectangular areas of Fig. 5). These conditions are expressed in the following:

$$x_2 - x_1 + 1 < th \quad \text{AND}$$
$$\frac{\sum_{\substack{i=x_1-2 \\ j=y-1}}^{\substack{x_1 \\ y+1}} I_F(i,j) + \sum_{\substack{i=x_1-2 \\ j=y-1}}^{\substack{x_1 \\ y+1}} I_F(i,j)}{18} - \frac{\sum_{\substack{i=x_1 \\ j=y}}^{\substack{x_2 \\ y}} I_F(i,j)}{x_2 - x_1 + 1} < 0 \qquad (6)$$

where th equals to half the average character height which is calculated as in [14].

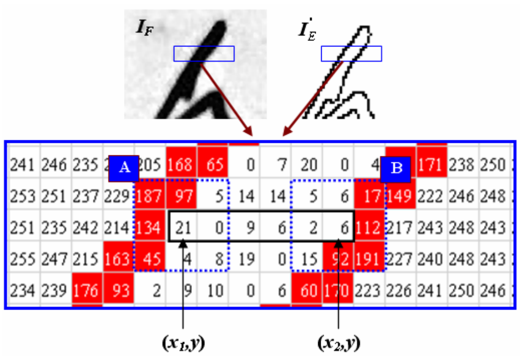
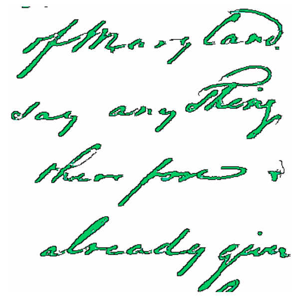


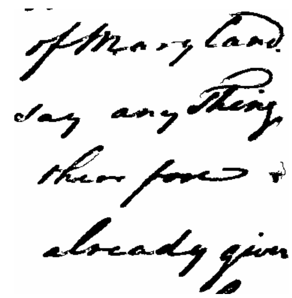
Figure 5. The extension of the Run Length Smoothing Algorithm that is guided by the grey scale image I_F. Example of a white run that is turned to a black run assuming *th* = 10.

In the example of Fig. 5, the white run from (x_1,y) to (x_2,y) is turned to a black run since (i) it is of short length (x_2-x_1+1=6 is less than th=10) and (ii) the corresponding average grey level value of the white run pixels which equals to (21+0+9+6+2+6)/6 = 7.3 is less than the corresponding average grey level value at the limits of the white run which is the sum of all values in the dotted areas A and B divided by 18 (52.6).

In the same way, we proceed to a vertical run length smoothing guided by the grey scale image I_F in order to completely fill all text areas and produce b/w image B_E. The final b/w image I_B after the incorporation of edge information is produced by applying an OR operation between images B_E and B_R. An example of the resulting images B_E and I_B is demonstrated in Fig. 6.



(a)



(b)

Figure 6. Example of incorporating the edge information in the b/w image B_R of Fig. 3(e) based on the adapted edge image $I_E^{'}$ of Fig. 4(b): (a) b/w image B_E having the text areas filled and (b) the final b/w image I_B after the incorporation of edge information.

2.4 Enhancement

In order to enhance the quality of image I_B which results from previous step, we first proceed to a conditional dilation with a 3x3 4-connected structuring element. The condition we apply is that a foreground pixel is added only if its corresponding grey value has not large difference with the grey value of the already existing foreground pixel in the 3x3 neighbourhood. This condition ensures that character thickness remains the same while several gaps in the character body will diminish. According to our methodology, $I_B(x,y)$ is turned to 1 only if the following condition is true:

$$\begin{aligned}
& I_B(x,y) = 0 \text{ AND} \\
& (I_B(x-1,y) = 1 \text{ AND} \\
& |I_F(x-1,y) - I_F(x,y)| < 0.05 I_F(x,y)) \text{ OR} \\
& (I_B(x+1,y) = 1 \text{ AND} \\
& |I_F(x+1,y) - I_F(x,y)| < 0.05 I_F(x,y)) \text{ OR} \\
& (I_B(x,y-1) = 1 \text{ AND} \\
& |I_F(x,y-1) - I_F(x,y)| < 0.05 I_F(x,y)) \text{ OR} \\
& (I_B(x,y+1) = 1 \text{ AND} \\
& |I_F(x,y+1) - I_F(x,y)| < 0.05 I_F(x,y))
\end{aligned} \qquad (7)$$

Finally, we proceed to a successive application of shrink and swell filtering [15] in order to further improve the quality of text regions and preserve stroke connectivity by isolated pixel removal and filling of possible breaks, gaps or holes. According to shrink filtering, $I_B(x,y)$ is turned to 0 if the following condition is true:

$$I_B(x,y)=1 \text{ AND } \sum_{\substack{ix=x-d\\iy=y-d}}^{\substack{x+d\\y+d}} I_B(ix,iy)<t_1 \quad (8)$$

where d denotes the window size and t_1 denotes the threshold for the pixel density in the window.

Similarly, according to swell filtering, $I_B(x,y)$ is turned to 1 if the following condition is true:

$$I_B(x,y)=0 \text{ AND } \sum_{\substack{ix=x-d\\iy=y-d}}^{\substack{x+d\\y+d}} I_B(ix,iy)>t_2 \quad (9)$$

After experimentation, we have used the following parameter values: d=2 and t_1=t_2=16.

The final b/w image I_{BE} that results after enhancing the quality of image I_B is demonstrated in Fig. 7.

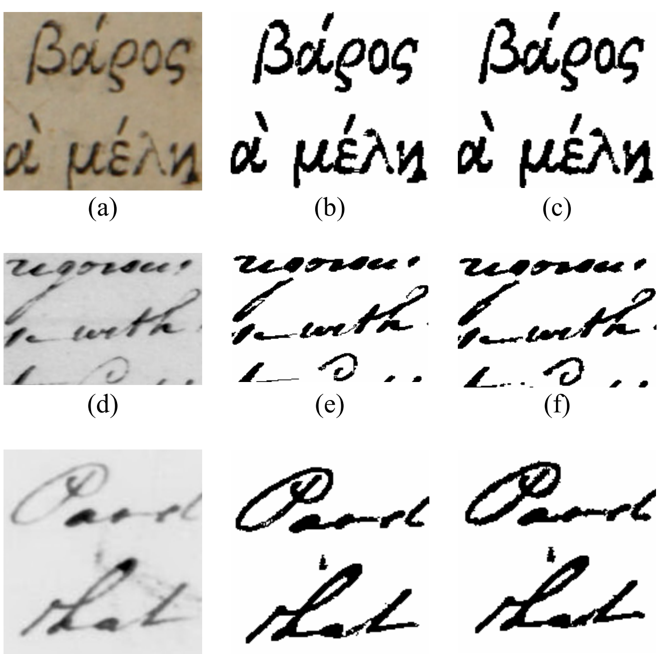


Figure 7. Enhancement examples: (a),(d),(i) original image I_F; (b),(e),(j) corresponding b/w image I_B; (c),(f),(k) enhanced b/w images I_{BE}.

3. Experimental results

The proposed algorithm was tested using ten representative historical and degraded, handwritten and machine-printed document images selected mainly from the Library of Congress on-line database [16] as well as from private collections. All document images are of poor quality and have shadows, non-uniform illumination, ink seeping, smear and strain.

For the implementation of the proposed methodology we used three state-of-the-art binarization techniques in order to combine the corresponding results. These are: Otsu's global thresholding method [1], Adaptive Logical method [5] and the Adapive degraded document method [6]. We compared the performance of our methodology with six well-known binarization techniques: Otsu's global thresholding method [1], Bernsen's adaptive method [2], Niblack's adaptive threholding method [3], Sauvola *et al.* adaptive method [4], Adaptive Logical method [5] and the Adaptive degraded document method [6].

Based on visual criteria, we can state that the proposed algorithm out-performs all algorithms that it is tested against with respect to image quality and preservation of meaningful textual information. Handwritten and machine-printed document example results are shown in Fig. 8-9.

For the quantification of the proposed binarization method's efficiency, additional experimentation was performed. We manually created the skeletonized ground truth image for all 10 images. To guide the user in this procedure, we computed the skeleton [17] after applying an adaptive binarization to the original image. In this stage, the Kamel and Zhao method [18], was used for binarization. To avoid to be biased towards a methodology, the Kamel and Zhao method is not included in our comparisons. Furthermore, in a user friendly environment, we provide a view of both the layers of the ground truth skeleton and the grey level image (Fig. 10(d)) in order to aid the user to easily delineate the remaining character parts or remove spurious parts. After a second skeletonization pass in order to ensure one pixel wide text, the skeletonized ground truth image S is created (Fig. 10(e)).

Recall is defined as the percentage of the skeletonized ground truth image S that is detected in the resulting binary image I_{BE}:

$$\text{Recall} = \frac{\sum_{x=1,y=1}^{x=Ix,y=Iy} S(x,y)\cdot I_{BE}(x,y)}{\sum_{x=1,y=1}^{x=Ix,y=Iy} S(x,y)}\ 100\ \% \quad (10)$$

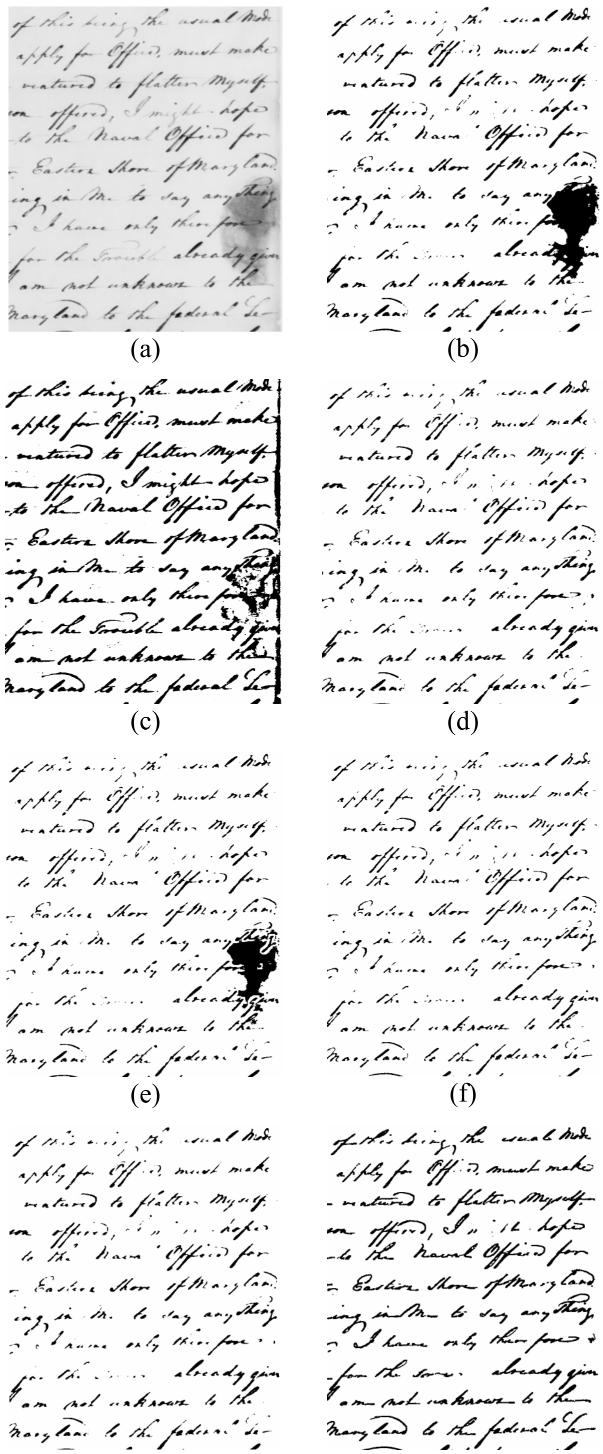


Figure 8. Binarization results for a handwritten document: (a) Original image; (b) Otsu's approach [1]; (c) Niblack's approach [3]; (d) Sauvola's approach [4]; (e) Bernsen's approach [2]; (f) Adaptive Logical approach [5]; (i) Adaptive degraded document approach [6] and (j) the proposed approach.

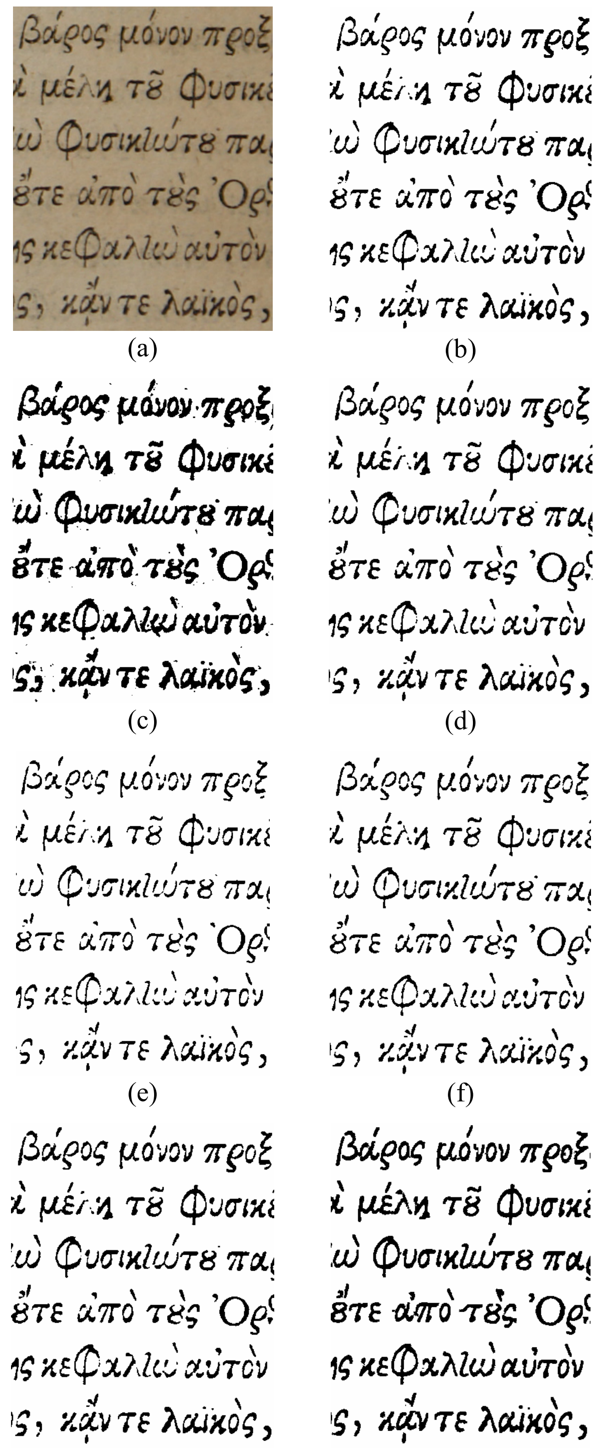


Figure 9. Binarization results for a machine-printed document: (a) Original image; (b) Otsu's approach [1]; (c) Niblack's approach [3]; (d) Sauvola's approach [4]; (e) Bernsen's approach [2]; (f) Adaptive Logical approach [5]; (i) Adapive degraded document approach [6] and (j) the proposed approach.

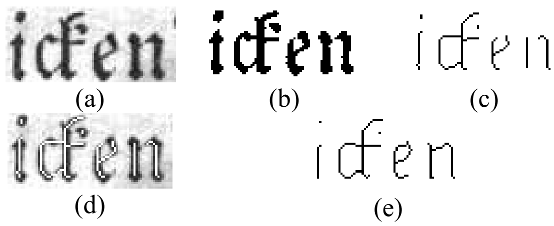


Figure 10. Skeletonized ground truth creation example: (a) original image; (b) the corresponding binary image after adaptive binarization; (c) binary image after skeletonization; (d) a view of both the layers of the ground truth skeleton and the grey level image in order to help the user edit the ground truth skeleton and (e) final skeletonized ground truth image.

Precision requires considering ground truth characters as much close as the original ones. For this purpose, we automatically estimate the ground truth taking into account that a skeletonized ground truth image has been achieved. Given the skeletonized ground truth image S, we apply a dilation constrained by the edge image ∂S and the binary image I_{BE} and produce the estimated ground truth image E (see Fig. 11). This procedure is summarized as follows: For every connected component (cc_i) of the binary image I_{BE}, the corresponding skeletonized ground truth component or components start to dilate. When half of the edges that belong to cc_i are covered by the dilated skeletonized ground truth components (Figures 12(c)-(d)), the dilation process will stop. The dilated components represent the estimated ground truth component. Dilated components cannot exceed cc_i, denoting that the dilated component cannot be larger than the original one.

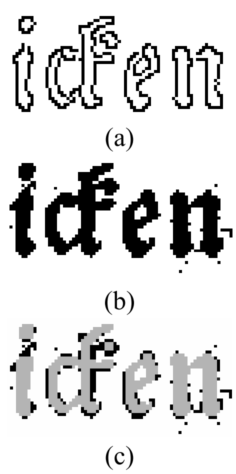


Figure 11. Construction of the estimated ground truth image E for the grey scale image of Fig. 10(a): (a) edges of original image; (b) binary image under evaluation and (c) estimated ground truth image (in grey).

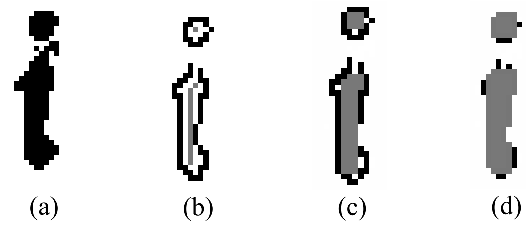


Figure 12. Example of the dilation procedure: (a) part of binary image I_{BE}; (b) the corresponding edges (in black) along with the skeletonized ground truth S (in grey); (c) the dilated S components after one dilation (in grey), 9.59% of the edges covered; (d) the dilated S components after two dilations (in grey), 78.08% of the edges covered indicated that dilation has to stop.

Precision is defined as the percentage of the estimated ground truth image that is detected in the binary image:

$$\text{Precision} = \frac{\sum_{x=1,y=1}^{x=Ix,y=Iy} E(x,y)\cdot I_{BE}(x,y)}{\sum_{x=1,y=1}^{x=Ix,y=Iy} I_{BE}(x,y)} 100\% \tag{11}$$

As an overall measure we use the weighted harmonic mean of precision and recall also referred as the F-measure:

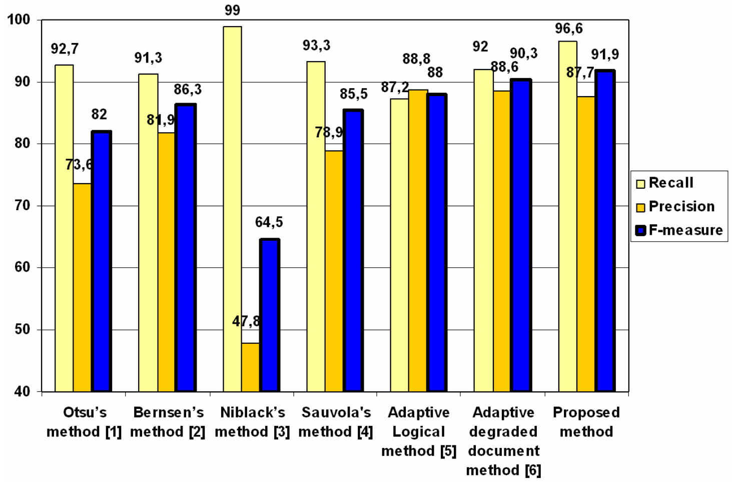
$$\text{F-measure} = \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

In Table 2, all comparative experimental results are presented in terms of average Recall, Precision and F-measure. According to these results, the proposed methodology has the best overall performance with F-measure equal to 91.9% while the Adaptive degraded document method [6] is second with F-measure equal to 90.3%. The complete ranking list is shown in the Table 1.

Table 1. Ranking list

Rank	Method	F-measure
1	Proposed method	**91.9 %**
2	Adapt.degraded docum. method [6]	90.3 %
3	Adaptive Logical method [5]	88.0 %
4	Bernsen's adaptive method [2]	86.3 %
5	Sauvola's adaptive method [4]	85.5 %
6	Otsu's global method [1]	82.0 %
7	Niblack's adaptive threholding [3]	64.5 %

Table 2. Comparative experimental results



4. Conclusions

In this paper we present a new adaptive approach for the binarization and enhancement of historical and degraded documents. The main novelties introduced of the proposed approach consist of (i) combining the binarization result of several state-of-the-art methodologies; (ii) incorporating the edge map of the grey scale image; and (iii) applying efficient image post-processing based on mathematical morphology for the enhancement of the final result. Experimental results in several historical and degraded documents prove the effectiveness of the proposed technique compared to other state-of-the-art methodologies.

Our future work will focus on the optimization of the step in which we combine the binarization result of several state-of-the-art methodologies.

References

[1] N. Otsu, "A thresholding selection method from gray-level histogram", *IEEE Trans. on Systems, Man and Cybernetics*. vol.9, pp. 62-66, March 1979.

[2] J. Bernsen, "Dynamic thresholding of grey-level images", *Proc. of the Eighth International Conference on Pattern Recognition,* Paris, France, 1986, pp. 1251-1255.

[3] W. Niblack, "An Introduction to Digital Image Processing", Prentice-Hall, Englewood Cliffs, NJ, 1986 pp. 115–116.

[4] J. Sauvola, T. Seppanen, S. Haapakoski, and M.Pietikainen, "Adaptive Document Binarization", *Fourth Int. Conference on Document Analysis and Recognition (ICDAR '97)*, Ulm, Germany, 1997, pp. 147-152.

[5] Y. Yang and H. Yan, "An adaptive logical method for binarization of degraded document images", Pattern Recognition, vol. 33, pp. 787-807, 2000.

[6] B. Gatos, I. Pratikakis and S.J. Perantonis, "Adaptive degraded document image binarization", *Pattern Recognition*, Elsevier, 39(3), 2006, pp. 317-327.

[7] Q. Chen, Q-S Sun, P.A. Heng and D-S. Xia, "A double-threshold image binarization method based on edge detector", *Patt. Recognition*, vol. 41, 2008, pp. 1254-1267.

[8] E. Badekas and N. Papamarkos, "Optimal combination of document binarization techniques using a self-organizing map neural network", *Engineering Applications of Artificial Intelligence*, Elsevier, vol. 20, 2007, pp. 11-24.

[9] A. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.

[10] R. Gonzalez and R. Woods, *Digital Image Processing,* Addison Wesley, pp 414 – 428, 1992.

[11] J. Canny, "A computational approach to edge detection", *IEEE Trans. PAMI*, 8, 1986, pp. 679-698.

[12] F. Chang, C.-J. Chen, C.-J. Lu, "A Linear-Time Component-Labeling Algorithm Using Contour Tracing Technique", *Computer Vision and Image Understanding, Vol. 93, No.2,* February 2004, pp. 206-220.

[13] F.M. Wahl, K.Y. Wong, R.G. Casey, "Block Segmentation and Text Extraction in Mixed Text/Image Documents", *Comp. Grap. and Im. Proc.*, 1982, pp. 375-390.

[14] B. Gatos, T. Konidaris, K. Ntzios, I. Pratikakis and S. J. Perantonis, "A Segmentation-free Approach for Keyword Search in Historical Typewritten Documents", *8th International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005, pp. 54-58.

[15] R.J. Schilling, *Fundamentals of Robotics Analysis and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[16] The Library of Congress (http://memory.loc.gov/)

[17] H. J. Lee, and B. Chen, "Recognition of Handwritten Chinese Characters via Short Line Segments", Pattern Recognition, 25(5), pp. 543-552, May 1992.

[18] M. Kamel, and A. Zhao, "Extraction of Binary Character/Graphics Images from Grayscale Document Images, *CVGIP*, vol. 55, pp. 203-217, May 1993.