Improved Document Image Binarization by Using a Combination of Multiple Binarization Techniques and Adapted Edge Information

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 document image binarization. The proposed method is mainly based on the combination of several state-of-the-art binarization methodologies as well as on the efficient incorporation of the edge information of the gray scale source image. An enhancement step based on mathematical morphology operations is also involved in order to produce a high quality result while preserving stroke information. The proposed method demonstrated superior performance against six (6) well-known techniques on numerous degraded handwritten and machine-printed documents. The performance evaluation is based on visual criteria as well as on an objective evaluation methodology.

1. Introduction

Document image 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 degradations (e.g. shadows and non-uniform illumination, ink seeping) the current trend is to use local information that guides the threshold value pixelwise in an adaptive manner ([2]-[6]).

Most of the adaptive local binarisation 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. Since no gold standard technique exists that performs best in all degradation cases, 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.

Motivated by the aforementioned use of edge information as well as the combination of several binarization approaches, we propose a novel methodology that is based on (i) efficient pre-processing; (ii) combination of the binarization results of several state-of-the-art methodologies; (iii) incorporation of the edge map of the grey scale source image and, finally, (iv) application of efficient image post-processing based on mathematical morphology for the enhancement of the final result. The remaining of the paper is structured as follows: Section 2 details the proposed methodology; experimental results are given in Section 3 and, finally, conclusions are drawn in Section 4.

2. Methodology

2.1 Pre-processing

A pre-processing stage of the grey scale source image is essential 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 to be efficient for the aforementioned goals. Our pre-processing module involves an adaptive Wiener method based on statistics estimated from a 5x5 neighborhood around each pixel.

Consider the gray scale source image:

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

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

where 0 corresponds to black and 255 to white.

The grey scale source image I_S is transformed to the filtered grey scale 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:

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

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 (see Fig. 1) is calculated as follows:

$$B_R(x,y) = \begin{cases} 1, \text{ if } \sum_{i=1}^{2m+1} R_i(x,y) > m \\ 0, \text{ otherwise} \end{cases} \qquad (3)$$

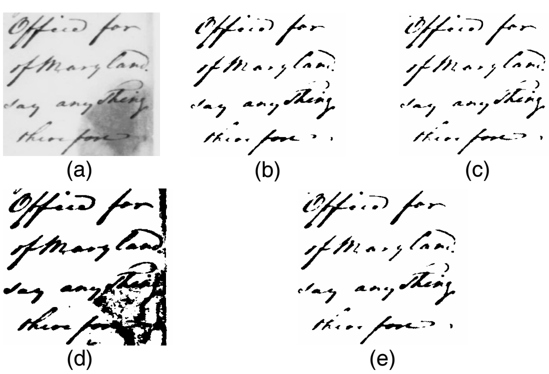


Figure 1. Example of combining several binarization results: (a) original image I_F; (b) - (d) binarization results R_1, R_2 and R_3, respectively and (e) combined result B_R.

2.3 Incorporation of edge information

At this step, we produce the edge map I_E of the filtered grey scale image I_F using the Canny edge detector [10]. According to this approach, edge magnitudes above an upper threshold th_e are preserved. Most of the times the selection of th_e is done in an ad hoc manner and the user is required to change its value based on the image at hand in order to achieve satisfactory edge detection results. To overcome this, we define threshold th_e as the minimum threshold that results to an edge map having more than 90% common pixels with binary image B_R (see Fig. 2).

Following the above procedure, the edge map I_E is defined as follows:

$$I_E(x,y) = \begin{cases} 1, \text{ if } (x,y) \in \partial I_F \\ 0, \text{ otherwise} \end{cases} \qquad (4)$$

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. 2(d).

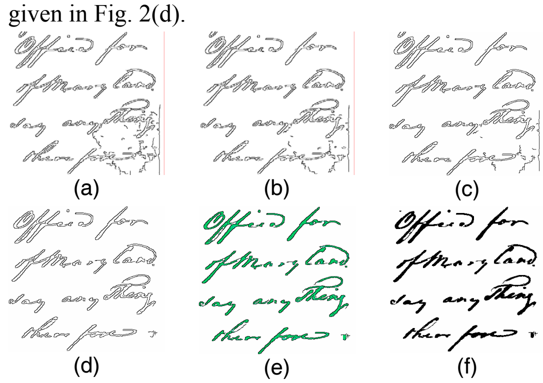


Figure 2. Incorporation of edge information: (a)-(c) Canny edge maps when threshold th_e is equal to 80, 90, and 100, respectively. The edge map I_E is achieved for th_e =100; (d) adapted edge map I'_E; (e) image B_E having the text areas filled; (f) the final b/w image I_B after the incorporation of edge information.

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 [11] 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. 3) is turned to a black run only if it is of short length and 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. 3). These conditions are expressed in the following:

$$x_2 - x_1 + 1 < th_l \quad \text{AND}$$

$$\frac{\sum_{\substack{i=x_1-2\\ j=y-1}}^{\substack{x_1\\ y+1}} I_F(i,j) + \sum_{\substack{i=x_1-2\\ j=y-1}}^{\substack{x_1\\ y+1}} I_F(i,j)}{18} - \frac{\sum_{\substack{i=x_1\\ j=y}}^{\substack{x_2\\ y}} I_F(i,j)}{x_2 - x_1 + 1} < 0 \qquad (5)$$

where th_l equals to half the average character height which is calculated as in [12].

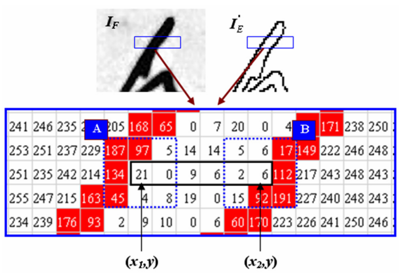
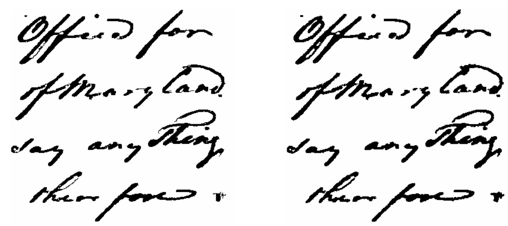


Figure 3. The extension of the Run Length Smoothing Algorithm that is guided by the grey scale image I_F. Example of a white run that is turned to a black run assuming $th = 10$.

In the 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 (see Fig. 2(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 (see Fig. 2(f)).

2.4 Enhancement

In order to enhance the quality of image I_B which results in the 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 gray value has not large difference (<5%) with the gray 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. Finally, we proceed to a successive application of shrink and swell filtering [13] 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.



(a) (b)

Figure 4. Enhancement stage: (a) b/w image I_B and (b) its corresponding enhanced image.

3. Experimental results

The proposed algorithm was tested using ten (10) representative historical and degraded, handwritten and machine-printed document images selected mainly from the Library of Congress on-line database 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 (3) 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 Adaptive degraded document method [6]. We compared the performance of our methodology with six (6) well-known binarization techniques: Otsu's global thresholding method [1], Bernsen's adaptive method [2], Niblack's adaptive thresholding 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 outperforms all algorithms that it is tested against with respect to image quality and preservation of meaningful textual information. Comparative binarization results are shown in Fig. 5.

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 after applying an adaptive binarization to the original image. Furthermore, in a user friendly environment, we provide a view of both the layers of the ground truth skeleton and the gray level image in order to aid the user to easily delineate the remaining character parts or remove spurious parts. Recall is defined as the percentage of the skeletonized ground truth image S that is detected in the resulting binary 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 both the edge image ∂S and the binary image which results in the estimated ground truth image *E*. Precision is defined as the percentage of the estimated ground truth image that is detected in the binary image. As an overall measure, we use the weighted harmonic mean of precision and recall also referred as the F-measure.

According to the comparison performed, the proposed methodology has the best overall performance with F-measure equal to 92.3% 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.

Figure 5. Binarization results: (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.

Table 1. Ranking list

Rank	Method	F-measure
1	Proposed method	**92.3 %**
2	Adapt.degr. 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 %

4. Conclusions

The main novelties of the proposed document image binarization approach consist of (i) combining the binarization results 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 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,* .9: 62-66, March 1979.

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

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

[4] J. Sauvola, T. Seppanen, S. Haapakoski, and M.Pietikainen, "Adaptive Document Binarization", *ICDAR '97* pp. 147-152, 1997.

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

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

[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*, 41:1254-1267, 2008.

[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, 20: 11-24, 2007.

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

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

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

[12] B. Gatos, T. Konidaris, K. Ntzios, I. Pratikakis and S. J. Perantonis, "A Segmentation-free Approach for Keyword Search in Historical Typewritten Documents", *ICDAR'05*, pp. 54-58, 2005.

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