A Morphology based Approach for Binarization of Handwritten Documents

Vassilis Papavassiliou[1], Fotini Simistira[1, 2], Vassilis Katsouros[1], George Carayannis[2]
[1]*Institute for Language and Speech Processing, Athena - Research and Innovation Center in Information, Communication and Knowledge Technologies, Athens, Greece*
[2]*School of Electrical and Computer Engineers*
National Technical University of Athens
Athens, Greece
fotini@ mail.ntua.gr
{vpapa, vsk}@ ilsp.athena-innovation.gr
carayan@softlab.ece.ntua.gr

Abstract

Document image binarization is an initial though critical stage towards the recognition of the text components of a document. This paper describes an efficient method based on mathematical morphology for extracting text regions from degraded handwritten document images. The basic stages of our approach are: a) top-hat-by-reconstruction to produce a filtered image with reasonable even background, b) region growing starting from a set of seed points and attaching to each seed similar intensity neighboring pixels and c) conditional extension of the initially detected text regions based on the values of the second derivative of the filtered image. The method was evaluated on the benchmarking dataset of the International Document Image Binarization Contest (DIBCO 2011) and show promising results.

1. Introduction

Binarization of document images is an image segmentation problem that attempts to extract text regions from the document. Therefore, binarization is an important stage in layout analysis upon which further tasks such as line or word segmentation and off-line character recognition may be developed. Although several algorithms have been proposed for document binarization, it remains an open issue since various types of degradations may appear. Non-uniform background illumination, smearing and shadows are the most common challenges one has to face for converting degraded grayscale document images into binary. For instance, the ink of one paper side often seeps through to the other. Another distinctive feature of the handwritten text is the variation in terms of stroke brightness, and stroke connection.

The organization of the rest of the paper is as follows: A short review of the related work is presented in Section 2. The proposed method is explained in detail in Section 3. Experimental results and conclusions are discussed in Section 4 and 5, respectively.

2. Related work

A detailed survey on binarization methods is presented in [1]. In general, binarization techniques may be categorized into global and local. The former approaches attempt to select a single value of intensity that distinguishes text regions from background areas. By reviewing the relevant bibliography, we conclude that global thresholding is of limited use in processing document images nowadays since it cannot handle common degradations effectively. However, well-known global thresholding techniques such as Otsu's [2] are still used in intermediate calculations or in combination with image enhancement methods.

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.158

Local thresholding uses a locally varying threshold function that depends either on a selected property of a pixel considering in its neighborhood [3] or on local statistics [4-7]. Other methods examine the histogram of the image to make conclusions relevant to the levels of intensity that are dominant in the image and/or exploit edge detection algorithms to localize the boundaries of text regions [8]. A comparative evaluation report on adaptive binarization methods is presented in [9].

Besides the methods proposed for threshold estimation, several filtering techniques such as Wiener and morphological filters have been used in document image binarization with the purpose of reducing noise and enhancing contrast [10]. Other approaches employ texture analysis [15] to locate foreground text or cross section sequence graph analysis [16] to separate text pixels from background.

3. The proposed method

The proposed methodology consists of three main steps: a) pre-processing, b) initial estimation of foreground areas and c) final localization of text regions.

3.1. Pre-processing

The goal of this step is to produce a filtered image with a roughly even background and homogeneous text areas. To accomplish that, we first apply median order filtering $\mathbf{Y} = med_{\mathbf{W}}(\mathbf{X})$, where $\mathbf{Y}$ denotes the resulting image, $\mathbf{W}$ is the flat structural element with dimension 3×3 and $\mathbf{X}$ denotes the original image. As a result, probable specks in the intensities of the original image are suppressed. Moreover, this non linear spatial filter does not blur the boundaries of text regions as a default averaging filter does.

Then, we apply the top-hat-by-reconstruction procedure with the purpose of compensating for non-uniform background illumination [11] and provide similar intensities to the pixels of each text region. Specifically, we make an initial estimation of the background by eroding $\mathbf{Y^C}$, i.e. the complement of $\mathbf{Y}$, by a large structuring element (SE), e.g. a disk with radius 25 pixels,. Due to the special nature of handwritten document images, it is very likely that no text element fits this structuring element[1]. Thus, the erosion will discard text areas and produce an image $\mathbf{Z}$ that could be used as a marker, i.e. none of its values exceeds the corresponding value in the mask. $\mathbf{Y^C}$.

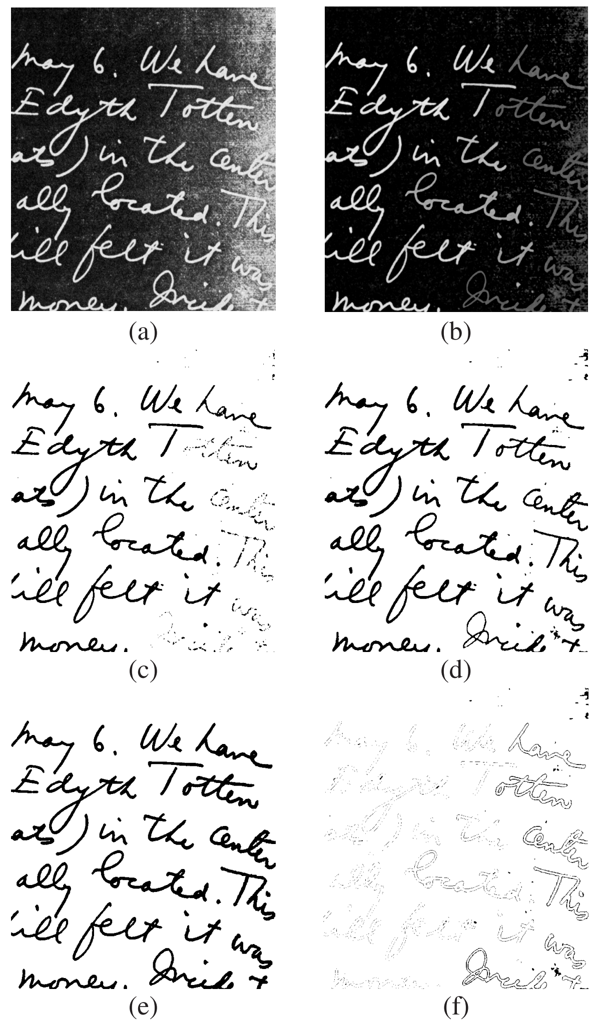

Figure 1. The initial steps of the method: (a) Original image $\mathbf{X}$ (HW1.png of DIBCO2011 dataset); (b) Filtered image $\mathbf{F}$; (c) The seeds of text regions $\mathbf{S}$; (d) The initial estimation of text regions $\mathbf{I}$, (e) The ground truth image $\mathbf{GT}$, and (f) $\mathbf{GT} \setminus \mathbf{I}$. Note that in binary images text is in white for visibility reasons.

By reconstructing $\mathbf{Y^C}$ from $\mathbf{Z}$, the intensities of the marker increase until they are forced downward by the local extrema of the mask. This procedure aims to extract the background of the document image. Consequently, the subtraction of the produced image from $\mathbf{Y^C}$ results into a filtered image $\mathbf{F}$ in which the

[1] Obviously, the selected size of the SE depends on the resolution of the scanned document image. The reported value (25 pixels) has been selected with respect to the resolution (96 dpi) of the images in DIBCO 2011 dataset.

smaller peaks (i.e. the background areas) are eliminated while the higher peaks (i.e. the text areas) are clipped (see fig. 1b).

3.2. Initial estimation of foreground areas

In order to detect areas for which we are very confident that belong to text we use $0.9*T$ as a reasonable global threshold, where T is the value calculated by the Otsu's method. Therefore, we generate a binary image $\mathbf{S}$ which contains the seeds of the text regions of the document. As expected, some text regions are almost fully covered while a few foreground pixels have been detected for others (fig. 1c). However, the localization of just one pixel of a foreground segment is enough to construct the entire component in the next step.

To accomplish this, we apply the reconstruction transformation [12] starting from the seed image and growing the regions as long as there are neighboring pixels (8-connectivity) that have intensity lower than a predefined value (e.g. $1.1*T$). The resulting image $\mathbf{I}$ includes the initial estimation of text regions and is illustrated in fig. 1d. The differences between this intermediate result and the ground truth image (see fig. 1e) are illustrated in fig 1f.

3.3. Final localization of text regions

Even though the initially detected foreground segments cover the main parts of the "real" text regions, it is likely that some critical parts are still missing. For instance, the bright part of the first symbol in the third text line of fig. 2c has not been assigned as text.

To deal with this shortcoming, we estimate the second derivative of the filtered image (fig. 2d). Considering that text elements are surrounded by negative values, we produce a binary image $\mathbf{BW_1}$ with 1s at the locations where the Laplacian has positive values (fig 2e). Obviously, $\mathbf{BW_1} \cup \mathbf{I}$ includes the missing text parts. Therefore, by using $\mathbf{BW_1} \cap \mathbf{I}$ as a marker and $\mathbf{BW_1} \cup \mathbf{I}$ as a mask, the binary reconstruction transformation results to a binary image $\mathbf{R}$ which contains the initial text parts, the missing parts and a few noisy areas originated from the second derivative (fig. 2f). Hence, we keep only the regions (connected components of $\mathbf{BW_1}$) of the second derivative that have been restored due to their 8-n connectivity with the connected components of $\mathbf{I}$ (i.e. the initially located text parts).

These areas can easily highlighted by estimating the local smoothness of $\mathbf{R}$ by the following equation:

$$M = 1 - 1/(1+\sigma^2),$$

where σ is the local standard deviation. The result of this process is presented in fig. 2g. In order to classify the areas of $\mathbf{R}$ as coarse or smooth, we calculate the normalized histogram (i.e. the probability density function) of M and estimate a global threshold in the most right local minimum of the histogram.

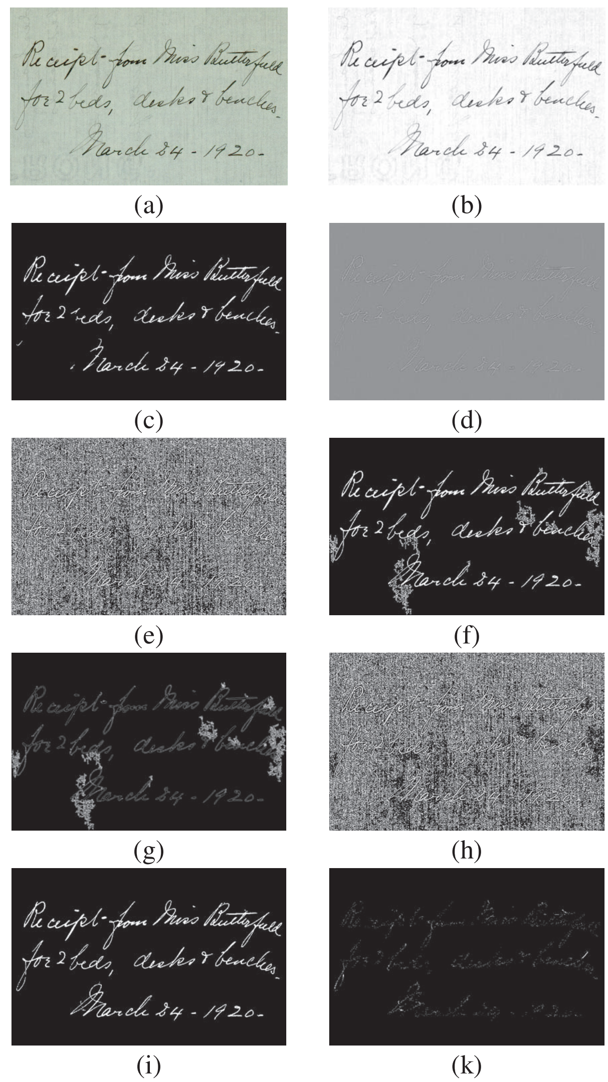

Figure 2. The steps of the process: (a) Original image (HW2.png of DICBO2011 dataset); (b) $\mathbf{F}$; (c) $\mathbf{I}$; (d) The Laplacian of $\mathbf{F}$; (e) $\mathbf{BW_1}$; (f) $\mathbf{R}$; (g) Visualization of local smoothness; (h) $\mathbf{BW_2}$; (i) The result of binarization; (k) The differences between the result and the ground truth.

Then, we produce another binary image $\mathbf{BW_2}$ by setting the values of pixels in $\mathbf{BW_1}$ that correspond to the coarse areas of $\mathbf{R}$ equal to 0. The elimination of these areas is evident by comparing the "dark" areas of $\mathbf{BW_2}$ and $\mathbf{BW_1}$.

The final step concerns the refinement of the detected text regions. We iteratively convert each OFF pixel of $\mathbf{I}$ into ON when it is a neighbor of an ON pixel in $\mathbf{I}$ and an ON pixel of $\mathbf{BW_2}$. The differences between the result and the ground truth image are presented in fig. 2k.

4. Experimental results

In order to evaluate the proposed method we used the eight handwritten documents included into the dataset of the testing dataset of DIBCO 2011. Even though the size of the testing dataset is small, the organizers reported that the selected images include representative types of degradations. The four measures that were adopted to evaluate the participated techniques were: a) F-measure, b) Peak Signal-to-Noise Ratio (PSNR), c) Distance Reciprocal Distortion Metric (DRD), and d) Misclassification Penalty Metric (MPM). More details about the selected measures, the ranking method as well as short descriptions of the participated algorithms are reported in the [13].

The scores of the proposed technique and the top 3 algorithms for each image are presented in Table 1.
Method 8 [17] employs a hybrid approach for edge detection by combining local statistics with Canny's algorithm. Adaptive local thresholding is applied on the produced edge map and proper heuristics are used as post-processing to provide better results. Similarly, method 10, adopts an edge detection method to roughly estimate text locations. Then, a clustering algorithm is applied to classify the pixels around the edges as "text" or "background". Method 11 considers binarization a graph cut problem that attempts to minimize a global energy function controlled by the Laplacian of the original image and the result of Canny's edge detector.
By comparing the evaluation results, we conclude that the binarization of degraded handwritten documents remains an open issue since no method performs almost excellent results on all images and outperforms the others. Regarding our method, we observed that the misses mainly concern the lack of exact match of the strokes' boundaries (see fig. 2k). However, the shapes of the text elements are not damaged. Another issue arises from the existence of small smudges in the binary image. To this end, simple heuristics (e.g. opening-by-reconstruction with special structuring elements such as a square of size 5 with 1s in the main anti-diagonal) could be introduced to remove these connected components without influencing the text regions.

Table 1. Evaluations Results

	Method	F-M	PSNR	DRD	MPM
HW1	10	88.2	15.1	6.6	14.0
	8	80.2	12.3	13.8	41.0
	11	79.1	11.8	15.3	48.0
	Ours	92.9	17.5	2.4	0.4
HW2	10	95.1	23.4	1.4	0.1
	8	93.7	22.6	1.7	0.1
	11	94.4	22.9	1.7	0.8
	Ours	94.4	23.0	1.4	0.1
HW3	10	92.8	19.8	1.8	0.2
	8	92.1	19.5	2.0	0.1
	11	93.2	20.0	1.8	0.6
	Ours	91.9	19.2	2.8	0.9
HW4	10	89.5	17.3	2.5	0.7
	8	87.9	16.8	3.0	0.7
	11	89.1	17.1	2.8	3.1
	Ours	87.6	16.4	3.6	3.0
HW5	10	95.2	19.7	1.6	1.1
	8	95.1	19.6	1.8	1.0
	11	90.6	16.4	4.6	12.0
	Ours	90.6	16.6	3.3	1.0
HW6	10	92.2	19.5	2.0	0.1
	8	76.4	15.3	6.3	0.7
	11	87.3	17.4	3.9	2.3
	Ours	75.6	14.8	9.3	3.9
HW7	10	92.0	22.0	1.7	0.1
	8	91.1	21.6	2.0	0.0
	11	88.5	20.2	3.4	2.0
	Ours	87.8	19.7	4.1	0.9
HW8	10	94.0	22.6	1.3	0.0
	8	93.4	22.3	1.5	0.1
	11	94.6	23.0	1.3	0.1
	Ours	92.5	21.6	2.1	0.1

5. Conclusions

We have presented an effective morphology-based technique for converting degraded handwritten documents into binary images. Median filtering and the top-hat-by-reconstruction transform are adopted in order to smooth the background. Then, we locate the seeds of text regions by using a global threshold lower than the one calculated by the Otsu's method. The region growing procedure follows to expand the seeds up to the boundaries of the regions corresponding to foreground pixels. However, it is likely that parts of text may be missed mainly due to low local contrast (i.e. faint text in bright background). We address this problem by estimating the Laplacian of the filtered image in order to enhance these image parts. Even though the second derivative is a noisy image, we remove areas with coarse texture and keep only the smooth areas which are adjacent to areas already classified as text regions. Therefore, the progressive expansion of the initial text areas is steered by the values (negative or positive) of the second derivative.

By observing the results, we conclude that the adoption of another technique for the initial detection of foreground areas would either produce a slightly more accurate system (e.g. by discarding small smudges) or speed up the whole process (i.e. by locating larger initial parts of text). On the other hand, the selection of the size of the sliding window, as in the case of adaptive local thresholding, would be an issue. In addition, the selection of a proper size and shape for the structuring element to detect redundant connected components would be another issue.

We also notice that even in the cases with not so high scores, text elements are not destructed. Therefore, we believe that the proposed method, as well as the other approaches submitted, could be considered either effective enough to be incorporated in an entire document analysis system, or promising enough to work towards its enhancement.

References

[1] M. Sezgin and B. Sankur. *Survey over image thresholding techniques and quantitative performance evaluation*. J. Electron. Imaging 13 (1): 146-165, 2004.

[2] N. Otsu. *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems Man Cybernet., 9(1):62–66, 1979.

[3] I.-K. kim, D.-W. Jung and R.-H. Park. *Document image binarization based on topographic analysis using a water flow model*. Pattern Recognition 35: 265-277, 2002

[4] W. Niblack. *An Introduction to Digital Image Processing*. Prentice-Hall, 115–116, 1986.

[5] J. Sauvola and M. Pietikainen. *Adaptive document image binarization*. Pattern Recognition 33:225–236, 2000.

[6] R. F.Moghaddam and M. Cheriet. *A multi-scale framework for adaptive binarization of degraded document images*. Pattern Recognition, 42:2186-2198, 2010.

[7] T. Obafemi-Ajayi and G. Agam. *Statistical multiresolution schemes for historical document binarization.* In Document Recognition and Retrieval XVIII. Proc. SPIE, 2011.

[8] M. A. Ramírez-Ortegón, E. Tapia, R. Rojas, and E. Cuevas. *Transition Pixel: A Concept for Binarization Based on Edge Detection and Gray-Intensity Histograms.* Pattern Recognition 43(4):1233-1243, 2010.

[9] O. D. Trier and A. K. Jain. *Goal-directed evaluation of binarization methods*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(12):1191–1201, 1995.

[10] B. Gatos, I. Pratikakis, S. J. Perantonis. *Adaptive degraded document image binarization*. Pattern Recognition, 39:317–327, 2006.

[11] R.C. Gonzalez and R.E. Woods. *Morphological Image Processing*. Prentice-Hall, 519-566, 2002.

[12] L. Vincent. *Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms*. IEEE Transactions on Image Processing, 2(2):176-201, 1993.

[13] E. Alpaydin, Introduction to Machine Learning, The MIT Press, Cambridge USA, 2004.

[14] I. Pratikakis, B. Gatos and K. Ntirogiannis. *H-DIBCO 2010 – Handwritten Document Image Binarization Competition*. ICFHR, 727-732, 2010.

[15] Y. Liu and S. Srihari. *Document image binarization based on texture features*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5):540–544, 1997.

[16] A. Dawoud. *Iterative cross section sequence graph for handwritten character segmentation*. IEEE Transactions on Image Processing, 16(8):2150–2154, 2007.

[17] B. Su, S. Lu and C. L. Tan. Binarization of historical document images using the local maximum and minimum. In Document Analysis Systems 159-166, 2010