Document Image Binarization using Structural Symmetry of Strokes

Fuxi Jia, Cunzhao Shi, Kun He, Chunheng Wang and Baihua Xiao
The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences (CASIA)
Beijing, China
Email: {jiafuxi2014, cunzhao.shi, hekun2014, chunheng.wang, baihua.xiao}@ia.ac.cn

***Abstract*—In this paper, a novel local threshold binarization method using structural symmetry of strokes is proposed. Different from most existing local threshold methods which use the whole region to compute the threshold, we estimate the local threshold by only using the structural symmetric pixels (SSP) of the region so as to suppress the non-text pixels and maintain the text ones as well. The SSP is defined as those pixels around strokes whose gradient magnitudes are big enough and directions are opposite. As the gradient map is our basis for computing the SSP, we further propose to estimate background surface first and extract potential SSP in the compensated image so as to deal with degradations of document images such as uneven illumination, low contrast and stain. To prove the effectiveness of our method, tests on two public document image datasets are preformed and the experimental results show that our method outperforms other local threshold binarization approaches on both F-measure and PSNR.**

***Keywords*-document image binarization; local threshold; stroke structure; background estimation**

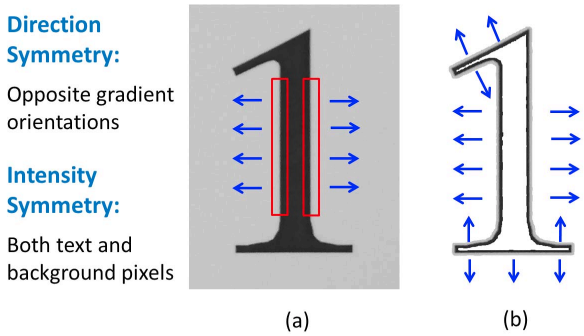

Figure 1. The illustration of structural symmetric pixels (SSP). We estimate the local threshold by only using SSP in neighborhood so as to suppress the possible random noise and background disturbance. (a) The motivation of SSP : pixels around strokes contain both text and background candidates. (b) Completed SSP (white pixels represent Non-SSP, black and gray pixels denote text and background candidates respectively). The blue arrows denote the gradient orientations of stroke edges.

I. Introduction

Optical character recognition (OCR) has been one of the earliest and most active topics in pattern recognition due to its scientific significance and practical applications. Binarization is an indispensable part of OCR since it is the basis of entire system. Although a lot of binarization methods have been applied in different situations successfully, there are still many unsolved problems in document images which have different types of degradation such as uneven illumination, bleeding-through and stain.

By definition, text binarization means labeling each pixel in the image as text or background. In current techniques, thresholding is one of the most useful methods for document images [1]. It is further divided into the global and local thresholding techniques. The global threshold approaches select a static threshold for the whole image, such as Otsu's [2], kittler's [3] and Brink's [4] methods. These techniques can work well on even background images but fail on complex background images. The local threshold approaches compute a local threshold based on the information in the neighborhood of each pixel. In Bernsen's [5] method, the threshold is a function of the lowest and highest gray values. In Niblack's [6] and Sauvola's [7] method, it is a function of the mean and standard deviation. Eikvil et al. [8] divide an input image into blocks and choose different binarization methods for each block. Some other literatures [9] [10] are published based on the similar idea. Although these local techniques can work better than global ones in some images, they are sensitive to background noises due to large variance such as uneven illumination or bleeding-through degradation. The reason might lie in the fact that the local threshold is estimated by all pixels in neighborhood including the possible random noise and background disturbance. To suppress the noises, Lu et al. [11] compute the local threshold based on the detected text stroke edges and the performance is improved to some extent. Additionally, similar to Gatos's [12], Su's [13] and Moghaddam's [14] methods, Lu et al. [11] estimate the background first and then binarize the compensated image instead of the original one so as to solve the problem of complex background. Other non-threshold approaches have been reported, including Markov Random Field [15] [16], self-learning [17] gabor filters [18] [19] and Laplacian energy [20] [21], these methods combine different types of image information and domain knowledge and are often complex.

In this paper, the structural properties of strokes have been utilized and we estimate the local threshold by only

2167-6445/16 $31.00 © 2016 IEEE
DOI 10.1109/ICFHR.2016.78

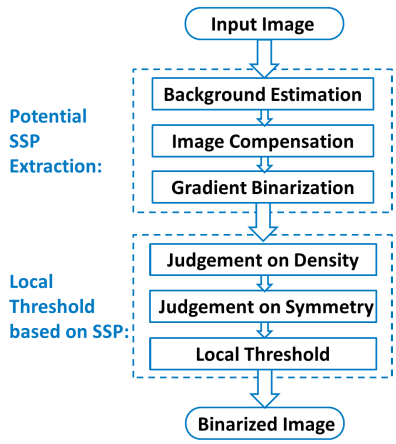

Figure 2. The flowchart of the proposed binarization method.

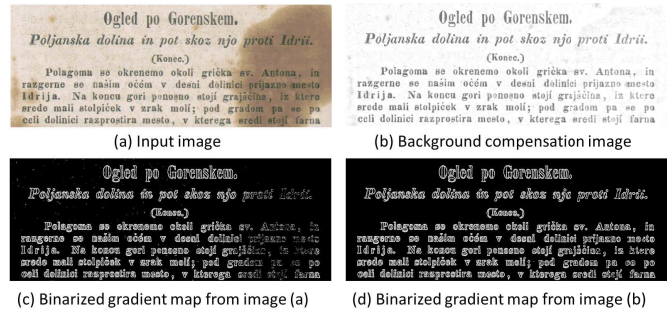

Figure 3. The process of potential SSP extraction: From (a) to (b), ended at (d). As we can see, the undetected stroke edges caused by the uneven background of original image are restored through the background compensation process.

using the structural symmetric pixels (SSP) of the region so as to suppress the non-text pixels and maintain the text ones as well. As shown in Fig. 1, the SSP is defined as the pixels around strokes which contain both text and background candidates and these pixels have big gradient magnitudes and opposite gradient directions. In order to deal with degradations of document images, we further propose to estimate background surface first and then extract gradient map in the compensated image. Experiments on two public document image datasets are preformed and the results show that our method outperforms other local threshold binarization approaches on both F-measure and PSNR.

The remainder of this paper is organized as follows: In the section II, we introduce the proposed binarization method. The concrete experiments are presented in the section III, and conclusions are drawn in section IV.

II. The Proposed Method

Fig. 2 shows the flowchart of the proposed pipeline. Given a document image, firstly, we extract the potential SSP by the following three substeps: estimating the background surface, compensating the variation of background and computing the binarized gradient map. Then we compute the local threshold based on SSP for each pixel after the judgment on density and symmetry of potential SSP. The concrete procedures will be introduced in subsection A and B, respectively.

A. Potential SSP Extraction

In order to suppress the non-text pixels, we estimate the local threshold by only using stroke-related pixels instead of the total ones in neighborhood. A fact can be observed that the pixels around strokes contain both text and background candidates and they are suitable for threshold estimation. This is the motivation of the SSP definition. Fig. 1 is a illustration of SSP. Pixel set **L** and **R** denote the two vertical edges in Fig. 1(a). In **L**, the gradient orientations are towards left while they are towards right in **R**. In addition, both text and background candidates exist in each edge and the quantities of them are roughly equal. Then, pixels in the union of set **L** and **R** are called local SSP since they have opposite direction symmetry and dual intensity symmetry in neighborhood. As shown in Fig. 1(b), the completed SSP denote edge pixels which satisfy the direction and intensity symmetry constraints. Then edge pixels are called potential SSP because they may contain some asymmetric noises.

Since the gradient map is responsive to both sides of the stroke edges, we extract potential SSP by binarizing the gradient magnitude image. For some document images suffering from different types of degradations such as uneven illumination or document smear, the gradient map may be binarized improperly. In particular, Fig. 3(c) shows an example that some SSP is not detected because of the low intensity contrast in the right side of image. For this reason, we estimate the background surface and compensate the variation of degradation to obtain an appropriate gradient map. Three procedures will be implemented as follows.

1) Estimating background image: Firstly, the input image of which the size is $M \times N$ is divided into some non-overlapping blocks of which the size is $P \times Q$. Then Sauvola's [7] thresholding method is used for each block b as follows:

$$T(b) = m(b)[1 + k(\frac{\sigma(b)}{R} - 1)] \quad b = 1, 2, ..., P \times Q. \quad (1)$$

where $m(b)$ and $\sigma(b)$ are the mean and standard variation of the intensities on block b. The parameter R is set to be 128 in the 8-bit input image and k is set to be 0.2 experimentally. For each block b, the background intensity value is computed by the mean of pixel intensities which are bigger than $T(b)$. Finally, resize the $P \times Q$ image into $M \times N$ background surface which refers to as B;

2) Compensating the variation of background: The document contrast compensation is performed by using the estimated background surface B as follows:

$$I_{norm} = \frac{I}{B} \times c \quad (2)$$

where c is a coefficient of intensity range of I_{norm}. In experiment, the minimum and maximum values of I_{norm} are normalized linearly as 0 and 255, respectively. Fig. 3(b) shows the normalized image of the input image in Fig. 3(a).

3) Computing binarized gradient map: In order to ensure the intensity symmetry of potential SSP, we need to detect double-edges in the images. The sobel operator is used to obtain the gradient of each pixel in the normalized image I_{norm}. It is believed that the variation of the background is compensated by the process of image normalization. Therefore the gradient amplitude can be binarized by a global threshold value and the threshold is computed based on otsu's [2] method. Then some techniques are used to remove small noises in the binarized gradient image.

After the above processing, the potential SSP is extracted: the binarized gradient image indicates its positions and the original image I indicates the intensities. Fig. 3(d) shows the binarized gradient map. Compared with Fig. 3(c), we can see that the undetected stroke edges caused by the uneven background of original image are restored through the background compensation process.

B. Local Threshold Binarization

In this paper, we use the mean intensity value of SSP in neighborhood as the local threshold due to the intensity symmetry of SSP. But the threshold can not be computed by potential SSP directly because there may contain some asymmetric noises in it. Therefore, some preprocess steps need to be done. For each pixel p in the image, its label is defined as text first. Then the label is changed to background directly if its local SSP can not satisfy the density or symmetry constraint. Otherwise, it will be determined by the local threshold computed based on local SSP. Fig. 4 is a illustration of the total processes.

1) Judging density of potential SSP: Firstly, computing the number of pixels of the local SSP according to pixel p. If the number is smaller than a threshold, p will be defined as background. The process can be specified as follows:

$$S(p) = \begin{cases} 1 & p \in potential\ SSP \\ 0 & else \end{cases} \tag{3}$$

$$N_{total}(p) = \sum_{q \in N_p} S(q) \tag{4}$$

$$L(p) = \begin{cases} 0 & N_{total}(p) < \alpha * W_{stroke} \\ L(p) & else \end{cases} \tag{5}$$

where L denotes the label map of the image ('1' denotes text and '0' denotes background), SSP mask map $S(p)$ denotes whether the pixel p belongs to potential SSP. N_p is the neighborhood window of the pixel p. $N_{total}(p)$ denotes the total number of local SSP corresponding to pixel p. W_{stroke} denotes the stroke width in the image and is a significant input parameter of binarization method. α is a coefficient of threshold. Fig. 4(f) shows three local windows: there are only a few potential SSP in W_1 and W_2 and no SSP in W_3 at all. The binarized image only processed by this step is shown in Fig. 4(b).

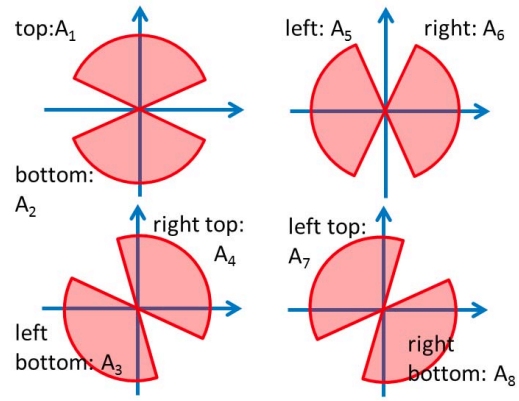

Figure 5. The eight angle ranges of gradient orientations for direction symmetry judgement: A_1, A_2 ,...,A_8.

2) Judging symmetry of potential SSP: The local SSP corresponding to text pixel must satisfy the direction and intensity symmetry constraints. Otherwise, the pixel should belong to background class. Therefore, we label pixel p as background if its local SSP is not symmetric.

Since the direction and intensity symmetries come in pairs and the direction one is defined more easily than the other, we only define and check the symmetry of direction. We divide angle plane into eight overlapping intervals, as shown in Fig. 5. They are evenly distributed throughout the plane of 360 degrees and each one of them is a range of 135 degrees. A_i denotes the i th group of the eight angle ranges, i = 1,2,...,8. If the gradient orientations focus on only one group, the local SSP is determined to be asymmetric. The process can be specified as follows:

$$N_{orint}(p,i) = \sum_{q \in N_p, Orint(q) \in A_i} S(q) \tag{6}$$

$$N_o(p) = \max_i N_{orint}(p,i) \tag{7}$$

$$L(p) = \begin{cases} 0 & N_o(p) > \beta * N_{total}(p) \\ L(p) & else \end{cases} \tag{8}$$

where $Orint(q)$ denotes the gradient orintation of the pixel q. In the local SSP corresponding to pixel p, $N_o(p)$ denotes the max number of pixels among eight ranges. β is a coefficient of threshold, it is set 0.75 in this paper. Three asymmetric local windows are shown in Fig. 4(g): W_4, W_5 and W_6 focuses on A_5, A_3 and A_2. The binarized image processed by first two steps is shown in Fig. 4(c).

3) Computing local threshold: The mean intensity of local SSP corresponding to the pixel p is defined in Eq. 9. Considering the purpose to make the binarized strokes thicker or thinner, the offset term δ is added in the Eq.

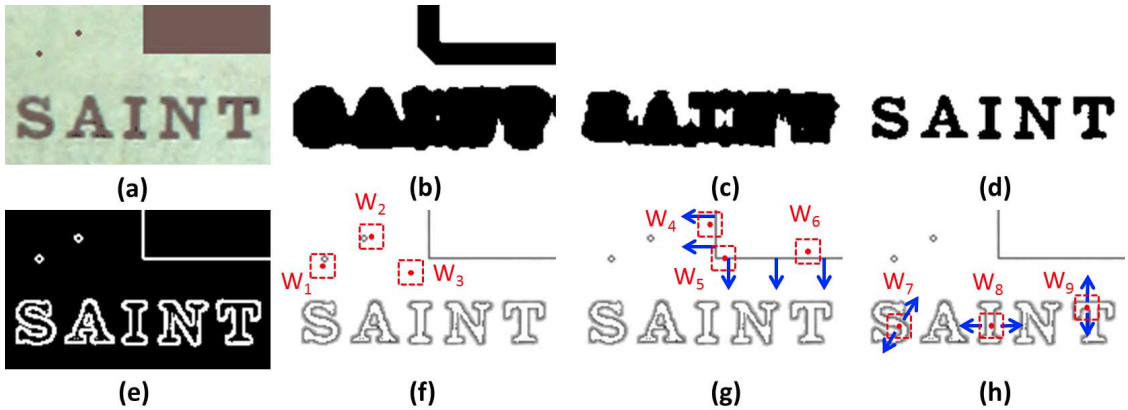

Figure 4. The illustration of binarization process. (a) Input image. (b)-(d) Binarization images achieved by density judgement, symmetry judgement and local threshold sequentially. (e) Binarized potential SSP. (f)-(h) Different local windows processed in density judgement, symmetry judgement and local threshold steps respectively.

10. Generally, the bigger δ is assigned the thicker stroke is binarized. In this step, pixel p will be labeled as text if its intensity is smaller than local threshold, otherwise it will be labeled as background. Fig 4(h) shows some symmetric windows. The final binarization image is shown in Fig. 4(d).

$$T(p) = \frac{\sum_{q \in N_p} I(q) \times S(q)}{N_{total}(p)} \tag{9}$$

$$L(p) = \begin{cases} 0 & I(p) > T(p) + \delta \\ L(p) & else \end{cases} \tag{10}$$

Fig. 4 gives an intuitive example which demonstrates the significance of each binarization step. In the first step, the two little noises in the upper left corner are labeled as background because of their low SSP density. Then pixels around the big block noise which lies in the upper right corner will be removed since their gradient orientations can not satisfy the direction symmetry constraint. Finally, the clear binarization image shown in Fig. 4(d) demonstrates the effectiveness of our local threshold computation method.

III. Experimental Results

In this section, the performance of the proposed method is evaluated on the DIBCO 2009 [22] and DIBCO 2013 [23] datasets, which consist of 10 and 16 different degraded, handwritten or printed images with their binarized ground truth images, respectively. To demonstrate the effectiveness of our binarization method, the proposed technique is evaluated between internal processes first. Then it is compared with other binarization algorithms including Otsu's [2] global thresholding method, Niblack's [6], Sauvola's [7], Howe's [20] and Lu's [11] local thresholding methods. In our experiment, parameters for the frontal four methods were chosen by selecting the best among several runs with different parameters. For Lu's method, the results are from their paper [11].

Table I
COMPARISON OF THE PROPOSED METHOD WITH DENSITY AND SYMMETRY JUDGEMENT OR NOT

	Method	F	PSNR
DIBCO 2009	ours without density judgment	89.37	17.63
	ours without symmetry judgment	90.93	18.36
	ours	**91.37**	**18.49**
DIBCO 2013	ours without density judgment	87.45	18.24
	ours without symmetry judgment	89.21	19.14
	ours	**89.50**	**19.30**

The evaluation measures from the DIBCO report [22][23] are including F-measure and peak signal-to-noise ratio (P-SNR). In particular, the two metrics is defined as follows:

$$F = \frac{2 \times P \times R}{P + R} \tag{11}$$

where P and R denote the binarization precision and recall, respectively.

$$PSNR = 10log(\frac{C^2}{MSE}) \tag{12}$$

where $MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(L(x,y)-L'(x,y))^2}{MN}$, C denotes the difference between text and background and equals to 1 in the label map level. Both of the two metrics measure how close the binarization result image is to the ground truth image.

A. Comparisons with density and symmetry judgement

To prove the necessity of the density and symmetry judgement processes quantitatively, we did a series of experiments on the two datasets. We skip the density and symmetry judgement respectively in the first two experiments and run the total steps in the last one. In order to avoid confusion,

Table II
EVALUATION ON DIBCO 2009 DATASET

Method	P	R	F	PSNR
Otsu's [2]	73.66	94.25	78.60	15.31
Niblack's [6]	43.74	71.13	51.17	9.25
Sauvola's [7]	86.77	86.02	85.12	16.33
Howe's [20]	87.95	93.20	90.37	18.05
Lu's [11]	-	-	91.24	**18.66**
ours	91.32	91.68	**91.37**	18.49

Table III
EVALUATION ON DIBCO 2013 DATASET

Method	P	R	F	PSNR
Otsu's [2]	83.16	84.73	80.04	16.63
Niblack's [6]	42.25	72.31	50.53	10.01
Sauvola's [7]	87.22	83.33	82.72	17.02
Howe's [20]	88.76	89.72	87.66	19.02
ours	92.35	88.03	**89.50**	**19.30**

parameters of the three kinds of experiments are set to be the same. As shown in table I, both the F-measure and PSNR of results without density judgement are quite lower than the completed ones due to the great quantity of small noises in history documents. The performance produced by skipping the symmetry judgement is slightly poor and the reason lies in the fact that the big block noises are fewer than the small ones. Therefore, the three steps of our method are indispensable.

B. Comparisons with other binarizaton methods

Experimental results of different binarization methods on the two datasets are shown in Table II and Table III. As can be seen in tables, our proposed method based on SSP outperforms most other methods in DIBCO 2009 database and achieves the highest score in both F-measure and PSNR in DIBCO 2013 database. Compared with Otsu's global thresholding technique, the performance of our method is higher due to the background compensation process which can balance the uneven background. The performance of the Niblack and Sauvola's local thresholding method is inferior to ours. This is because the SSP in local window can suppress the random noise and background disturbance. Compared with Lu's method which achieves the best performance in the competition DIBCO 2009, our method use a fixed value of stroke width for the whole datasets and the background estimation method is simple. We achieve higher F-measure thanks to the structural symmetry of SSP. But the PSNR score is slightly lower, the reason might lie in that we do not take any post-processing operations in experiment.

Fig. 6 shows the experimental results of different binarization methods. Otsu's method performs poorly on the image with uneven background but achieve acceptable result in the image without background interference, as shown in Fig. 6(e)(f). The binarization results of Sauvola's method shown in Fig. 6(g)(h) are applicable for uneven background image but too sensitive to bleeding-through noises. Our method

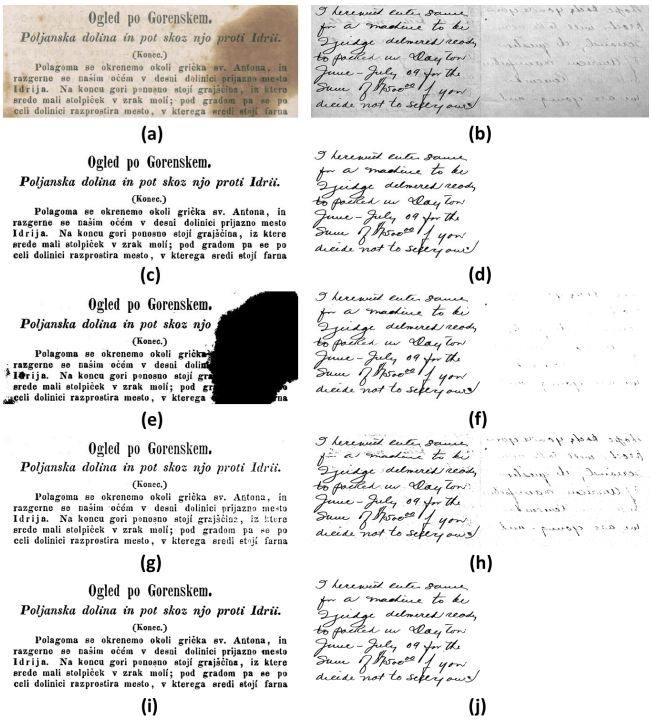

Figure 6. binarization results of two document images. (a)-(b) input images. (c)-(d) binarization ground truth. (e)-(f) results of otsu's [2] method. (g)-(h) results of Sauvola's [7] method. (i)-(j) results of ours method

produces better results thanks to the structural properties of SSP. On one hand, the background estimation can suppress the global uneven background. On the other hand, the structural symmetry of strokes can suppress the non-text pixels and maintain the text ones as well in local window.

IV. CONCLUSION

This paper presents a novel document binarization method using structural symmetry of strokes. We only use the SSP to estimate the local threshold in neighborhood so as to suppress the non-text pixels. In order to extract SSP properly and weaken the influence of document degradations, we estimate the background surface first and then use the compensated image to detect stroke edges. Experimental results on two public datasets show that our method can achieve superior performance for degraded document image binarization.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No. 61271429 and No. 61531019.

REFERENCES

[1] J. Kaur and R. Mahajan, "A review of degraded document image binarization techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 5, pp. 6581–6586, 2014.

[2] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[3] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern recognition*, vol. 19, no. 1, pp. 41–47, 1986.

[4] A. Brink, "Thresholding of digital images using two-dimensional entropies," *Pattern recognition*, vol. 25, no. 8, pp. 803–808, 1992.

[5] J. Bernsen, "Dynamic thresholding of grey-level images," in *International conference on pattern recognition*, 1986, pp. 1251–1255.

[6] W. Niblack, *An introduction to digital image processing*. Strandberg Publishing Company, 1985.

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

[8] L. Eikvil, T. Taxt, and K. Moen, "A fast adaptive method for binarization of documents images," in *Proc. 1st Int. Conf. Document Analysis and Recognition*, 1991.

[9] C.-H. Chou, W.-H. Lin, and F. Chang, "A binarization method with learning-built rules for document images produced by cameras," *Pattern Recognition*, vol. 43, no. 4, pp. 1518–1530, 2010.

[10] J. H. Park, I. H. Jang, and N. C. Kim, "Skew correction of business card images in pda," in *Communications, Computers and signal Processing, 2003. PACRIM. 2003 IEEE Pacific Rim Conference on*, vol. 2. IEEE, 2003, pp. 724–727.

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

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

[13] B. Su, S. Lu, and C. L. Tan, "Binarization of historical document images using the local maximum and minimum," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. ACM, 2010, pp. 159–166.

[14] R. F. Moghaddam and M. Cheriet, "Rsldi: Restoration of single-sided low-quality document images," *Pattern Recognition*, vol. 42, no. 12, pp. 3355–3364, 2009.

[15] T. Lelore and F. Bouchara, "Document image binarisation using markov field model," in *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*. IEEE, 2009, pp. 551–555.

[16] C. Shi, B. Xiao, C. Wang, and Y. Zhang, "Adaptive graph cut based binarization of video text images," in *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*. IEEE, 2012, pp. 58–62.

[17] B. Su, S. Lu, and C. L. Tan, "A self-training learning document binarization framework," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3187–3190.

[18] A. K. Jain and S. Bhattacharjee, "Text segmentation using gabor filters for automatic document processing," *Machine Vision and Applications*, vol. 5, no. 3, pp. 169–184, 1992.

[19] A. Sehad, Y. Chibani, and M. Cheriet, "Gabor filters for degraded document image binarization," in *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*. IEEE, 2014, pp. 702–707.

[20] N. R. Howe, "A laplacian energy for document binarization," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 2011, pp. 6–10.

[21] K. R. Ayyalasomayajula and A. Brun, "Document binarization using topological clustering guided laplacian energy segmentation," in *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*. IEEE, 2014, pp. 523–528.

[22] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "Icdar 2009 document image binarization contest (dibco 2009)," in *2009 10th International conference on document analysis and recognition*. IEEE, 2009, pp. 1375–1382.

[23] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "Icdar 2013 document image binarization contest (dibco 2013)," in *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE, 2013, pp. 1471–1476.