MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

A Binarization Method for Degraded Document Images with Morphological Operations

Akihiro OKAMOTO, Hiromi YOSHIDA and Naoki TANAKA
Graduate School of Maritime Sciences, Kobe University, Japan
Graduate School of Engineering Science, Osaka University, Japan
128w103w@stu.kobe-u.ac.jp, yoshida@sys.es.osaka-u.ac.jp, ntanaka@maritime.kobe-u.ac.jp

Abstract

In this paper, we propose an effective binarization method for de-graded document images in this paper. This method employs morphological operations throughout its algorithm to suppress uneven illumination in the background region, to detect the character location and to reconstruct text regions. Moreover, a technique for estimating stroke width of characters is introduced to remove noises in a robust manner and preserve text regions. In order to confirm its validity, several of the experiments are conducted on the datasets including a wide variety of degraded document images which are provided by DIBCO 2011 (Document Image Binarization Contest). It is shown that our method achieves a good performance, compared with other methods submitted in the contest.

1 Introduction

A binarization step plays an important role in the document image recognition and analysis since its performance exceedingly affect subsequent processes. Although many binarization methods have been proposed[1], there is much room for discussing the degraded document image binarization. Generally speaking, binarization methods are divided into two classes, global method, in which method a single threshold value is applied to the whole of an image, and local method, in which method threshold values are computed locally, say, for each pixel. In the case of the degraded document image binarization, global methods such as Otsu's method[2] and Kittler's method[3] fail to separate the text regions from the background region because of degrading components such as stains bleed-through and uneven illumination, while local methods such as Niblack's method[4] and Sauvola's method[5] require sensitive parameter settings and may cause annoying noises from the background region.

With the intention of evaluating the performance of binarizatiom methods for document images objectively, Document Image Binarization Contest (DIBCO)[6] series are held every year. Judging from many state-of-the-art techniques being submitted to the contest, the document image binarization is still unsolved and a challenging field.

In this paper, we present a parameter free binarization method for degraded document images with morphological operations.

In the following, morphological operations are presented in Section 2 and then the proposed method is described in detail in Section 3. Experimental measures and results are shown in Section 4, and finally conclusions are drawn in Section 5.

2 Morphological Operation

Several of the morphological techniques[9] are introduced to our document image binarization method. Outlines of these techniques are mentioned in this section.

2.1 Basic Operations

The dilation of a grayscale image f by structuring element b is denoted by $f \oplus b$, and the erosion by $f \ominus b$. Therefore, the opening and closing operations of f by b are denoted as follows respectively:

$$f \circ b = (f \ominus b) \oplus b \qquad (1)$$

$$f \bullet b = (f \oplus b) \ominus b \qquad (2)$$

The effect of opening is to suppress regions which is smaller than the structuring element b and that of closing is to fill holes or gaps in the contour by b.

2.2 Black Tophat Transformation

The black tophat transformation preserves the foreground region which a structuring element fits and removes the background region which it does not, and the intensity of each region is inverted. The transformation of a grayscale image f is defined as follows:

$$T_{black}(f) = (f \bullet b) - f \qquad (3)$$

2.3 Morphological Gradient

The morphological gradient, which denoted by g, of grayscale image f is subtraction between the dilated f and the eroded f, where

$$g = (f \oplus b) - (f \ominus b) \qquad (4)$$

The contour regions on foreground objects are emphasized by the operation.

2.4 Conditional Dilation

The conditional dilation [7] is one of the regional reconstruction techniques which makes the marker image grow under constraint of the mask image. Let X denote the mask image and Y the marker image, and the conditional dilation concerning the mask X denoted by $\delta_X^{(n)}$ is defined as

$$\delta_X^{(n)} = (\delta_X^{(n-1)} \oplus b) \cap X \qquad (5)$$

where $\delta_X^{(0)} = Y$, and n denotes the number of iteration. When the operation reaches a steady state, $\delta_X^{(i)} = \delta_X^{(i-1)}$, the reconstruction is complete due to i times repetitive dilations.

3 Proposed Method

The processing steps of the proposed method are shown in Figure 1. Each stage of process is described in detail respectively following subsections.

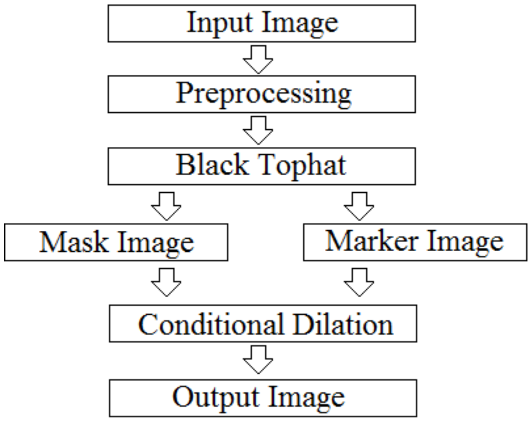

Figure 1. Processing flow

3.1 Preprocessing

First, a color input image is converted into a grayscale image, which employs Y channel of YUV color space. Then the text stroke width is roughly estimated for the following the black tophat operation. After binarizing the grayscale image with Otsu's method, prospective text regions are thinned with Hilditch's algorithm[8], and both the largest connected line and small components which are composed of less than 50 pixels are regarded as components causing noises and removed. The remainder of the lines are dilated iteratively with a cross shaped 3x3 structuring element until it becomes less than 0.2 that the ratio of the number of added pixels which is on the foreground region of previous binarized image to all added pixels. The value of *estimated stroke width* is obtained by $[\textit{times of iteration} \times 2 + 1]$. Finally, to remove speckle noises a 3x3 median filter is applied to the previous grayscale image.

3.2 Generating Mask and Marker

In order to suppress the uneven background intensity, the black tophat operation is applied to the previous image. Then the diameter of a round shaped structuring element is determined by $[\textit{estimated stroke width} \times 1.5]$ to completely fit the text regions (Figure 2). The obtained image of this process is applied to following both generating each the mask and marker image step.

Mask image

The mask image is obtained by binarizing the image with a threshold value calculated by Otsu's method which is shifted from the original value to the lower one by $[\textit{difference between both class mean values} \times 0.2]$ (Figure 3(a)).

Marker image

The marker image on which binary contour lines of text regions are located is obtained by applying Canny edge detector[10] to the image processed with morphological gradient (Figure 3(c)(d)). Then threshold values, high and low ones, of the edge detecting algorithm are determined by the Sobel gradient intensity distribution generated from the morphological gradient image. Based on the computation of the intensity distribution with Otsu's algorithm, the upper class mean value and Otsu's threshold value are set on the high and low threshold respectively.

3.3 Text Region Reconstruction

Through the conditional dilation, regions only on the mask image disappear and regions on the marker image grow within the mask image. The process of this operation is shown in Figure 4. Isolated regions which are only on Mask image (a) are not grown and removed.

(a) Input image (b) Gray image

(c) Closing (d) Black tophat

Figure 2. Black tophat transformation

4 Experiments

4.1 Dataset for Evaluation

The testing dataset provided at DIBCO 2011[6] is exploited to evaluate the proposed method. It consists of 16 degraded document images including 8 handwritten and 8 machine-printed images with each of the Ground Truth (GT). Every image contains distinctive features of degradation such as stains, bleed-through, uneven illumination and so forth, which could make accurate thresholding process difficult.

4.2 Experimental Measures

Four experimental measures are employed to assess the performance of the proposed method objectively in accordance with the framework of DIBCO 2011. Each measure is described simply in numerical expressions

(a) Mask image (b) Eroded image

(c) Gradient image (d) Marker image

Figure 3. Mask and marker image

(a) Mask image (b) Marker image

(c) Conditional dilation (d) Output

Figure 4. Conditional dilation

in this section. To know more about experimental measures please refer to DIBCO 2011 paper[6].

F-Measure

$$F-Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{6}$$

where $Recall = \frac{TP}{TP+FN}$, $Precision = \frac{TP}{TP+FP}$

TP, FP, FN denote the True positive, False positive and False Negative values, respectively.

PSNR

$$PSNR = 10 \log \left(\frac{C^2}{MSE} \right) \tag{7}$$

where $MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y) - I'(x,y))^2}{MN}$

The difference between foreground and background equals to C.

Distance Reciprocal Distortion Metric (DRD)

DRD is used to measure the visual distortion in binary document images[11].

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{8}$$

$$DRD_k = \sum_{i=-2}^{2} \sum_{j=-2}^{2} |GT_k(i,j) - B_k(x,y)| \times W_{Nm}(i,j) \tag{9}$$

where DRD_k is calculated using a weight matrix W_{Nm} defined in[11]. NUBN is the number of the non-uniform blocks in the GT image.

Misclassification penalty metric (MPM)

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \tag{10}$$

where $MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^i}{D}$, $MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^j}{D}$

d_{FN}^i and d_{FP}^j denote the distance of the i^{th} false negative and the j^{th} false positive pixel from the contour of the GT. D is the sum over all the pixel-to-countour distances of the GT object.

4.3 Experimental Results

The experimental results are shown in Table 1 and Table 2. The scores of the methods ranked as 1–3 in DIBCO 2011 plus that of Otsu's method are also listed in addition to the proposed one.

Our method outperforms Otsu's method in every test image and achieves a good performance compared with other methods.

5 Conclusions

We present an effective binarization method for degraded document images which employs morphological operations to reduce noises and reconstruct text regions. The proposed method shows good performance in experiments using the dataset provided by DIBCO 2011.

However, there being lack of information to some extent—the loss of connectivity within a character and the undetected characters from very low contrast regions, our future work is to cope with such problems.

Acknowledgements

This work was supported by a kakenhi Grant-in-Aid for Scientific Research (C)22500154 from Japan Society for the Promotion of Science.

Table 1. Results of hand-written images

	Method	FM	PSNR	DRD	MPM
HW1	rank1	88.2	15.1	6.6	14.0
	rank2	80.2	12.3	13.8	41.1
	rank3	79.1	11.8	15.3	48.0
	Otsu	67.6	9.3	27.5	80.7
	Proposed	93.6	17.8	2.9	3.8
HW2	rank1	95.1	23.4	1.4	0.1
	rank2	93.7	22.6	1.7	0.1
	rank3	94.4	22.9	1.7	0.8
	Otsu	89.0	20.3	2.8	0.1
	Proposed	93.1	22.0	1.7	0.1
HW3	rank1	92.8	19.8	1.8	0.2
	rank2	92.1	19.5	2.0	0.1
	rank3	93.2	20.0	1.8	0.6
	Otsu	86.7	17.3	3.4	1.4
	Proposed	91.0	18.8	2.2	0.2
HW4	rank1	89.5	17.3	2.5	0.7
	rank2	87.9	16.8	3.0	0.7
	rank3	89.1	17.1	2.8	3.1
	Otsu	49.3	7.7	35.7	81.1
	Proposed	89.0	17.1	2.5	0.4
HW5	rank1	95.2	19.7	1.6	1.1
	rank2	95.1	19.6	1.8	1.0
	rank3	90.6	16.4	4.6	12.0
	Otsu	90.2	16.5	3.9	6.7
	Proposed	93.9	18.5	2.2	2.2
HW6	rank1	92.2	19.5	2.0	0.1
	rank2	76.4	15.3	6.3	0.7
	rank3	87.3	17.4	3.9	2.3
	Otsu	65.2	12.2	15.8	17.1
	Proposed	89.2	18.2	2.8	0.3
HW7	rank1	92.0	22.0	1.7	0.1
	rank2	91.1	21.6	2.0	0.0
	rank3	88.5	20.2	3.4	2.0
	Otsu	82.1	18.4	5.3	3.1
	Proposed	90.7	21.3	2.2	0.4
HW8	rank1	94.0	22.6	1.3	0.0
	rank2	93.4	22.3	1.5	0.1
	rank3	94.6	23.0	1.3	0.1
	Otsu	88.9	20.2	2.4	0.1
	Proposed	91.6	21.1	1.9	0.1

Table 2. Results of machine-printed images

	Method	FM	PSNR	DRD	MPM
PR1	rank1	94.9	17.8	2.5	1.2
	rank2	92.9	16.4	3.5	2.4
	rank3	94.2	17.2	3.0	3.4
	Otsu	94.0	17.0	3.0	4.3
	Proposed	96.2	18.9	1.9	1.8
PR2	rank1	77.2	11.9	12.8	34.9
	rank2	82.0	13.2	9.0	26.0
	rank3	70.3	10.2	19.6	53.1
	Otsu	76.6	11.7	13.0	35.9
	Proposed	78.6	12.0	11.7	25.9
PR3	rank1	94.8	17.3	1.8	0.4
	rank2	93.8	16.5	2.3	0.9
	rank3	96.5	18.9	1.3	0.5
	Otsu	91.9	15.4	2.9	3.1
	Proposed	94.1	16.7	2.3	0.9
PR4	rank1	95.0	19.6	2.0	0.1
	rank2	92.0	17.7	3.5	0.1
	rank3	94.8	19.5	2.0	0.1
	Otsu	93.5	18.5	2.7	0.6
	Proposed	96.1	20.5	1.6	0.1
PR5	rank1	92.3	16.7	2.4	1.0
	rank2	92.7	17.1	2.0	0.2
	rank3	94.8	18.5	1.5	0.3
	Otsu	80.0	11.8	9.6	19.8
	Proposed	90.7	15.7	2.9	0.6
PR6	rank1	9.9	0.6	575.0	478.5
	rank2	92.6	21.4	3.1	0.1
	rank3	84.9	17.9	9.3	5.9
	Otsu	90.2	20.0	4.7	1.8
	Proposed	90.3	19.9	4.6	0.6
PR7	rank1	4.6	0.2	1052.7	498.0
	rank2	21.1	7.6	191.3	71.1
	rank3	79.1	19.2	11.0	4.1
	Otsu	86.4	21.5	6.0	1.3
	Proposed	90.0	22.8	3.4	0.2
PR8	rank1	86.1	14.6	3.6	0.4
	rank2	86.2	14.6	3.8	0.5
	rank3	88.5	15.3	3.2	2.6
	Otsu	82.3	13.7	4.5	1.3
	Proposed	85.8	14.4	3.8	0.8

References

[1] M. Sezgin and Bülent Sankur, : "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, p. 220, 2004.

[2] N. Otsu, : "A threshold selection method from gray-level histograms," *Automatica*, vol. 20, no. 1, pp. 62–66, 1975.

[3] J. Kittler and J. Illingworth, : "Minimum error thresholding," *Pattern Recogn*, 19,41–47, 1986.

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

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

[6] I. Pratikakis, B. Gatos, and K. Ntirogiannis, : "ICDAR 2011 Document Image Binarization Contest (DIBCO 2011)," *2011 International Conference on Document Analysis and Recognition*, no. Dibco, pp. 1506–1510, Sep. 2011.

[7] L. Vincent, : "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *Image Processing, IEEE Transactions on*, vol. 2, no. 2, 1993.

[8] S. Hilditch, : "Linear Skeleton from Square Cupboards," *J. Appl. Phys.*, vol.74, pp.403–419, Aug. 1993.

[9] Rafael C. Gonzalez, Richard E. Woods, : Digital Image Processing, Prentice Hall, 2007

[10] J. Canny, : "A computational approach to edge detection.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 8, no. 6, pp. 679?98, Jun. 1986.

[11] H. Lu, A. Kot, and Y. Shi, : "Distance-reciprocal distortion measure for binary document images," *Signal Processing Letters*, IEEE, vol. 11, no. 2, pp. 228–231, 2004.