Binarization Techniques for Degraded Document Images - A Review

Jyotsna[1], Shivani Chauhan[2], Ekta Sharma[3], Amit Doegar[4]

[1,2,3]Deptt. of Computer Science & Engineering, National Institute of Technical Teachers' Training and Research, Sector-26, Chandigarh 160019, India

[4]Assistant Professor, Deptt. of Computer Science & Engineering, National Institute of Technical Teachers' Training and Research, Sector-26, Chandigarh 160019, India

[1] jyotsnabhumbla@gmail.com, [2] shivanichauhan.1953@gmail.com

[3] ektasharma0013@gmail.com, [4]amit@nitttrchd.ac.in

Abstract: **Document Image binarization is the segmentation of the document into foreground text and background. It is done to obtain the clear images from which text can be retrieved easily. Thresholding is used for the segmentation of the document images. This paper, presents a review on various document image binarization techniques. Evaluation performance metrics used for the evaluation of the binarization techniques are also explained. Comparison of the performance of the binarization techniques based on the performance metrics like PSNR, F-Measure, NRM and MPM is shown. Performance of the techniques is evaluated on the dataset of DIBCO-2009 and DIBCO-2010.**

Keywords: **segmentation, document image binarization, degraded document images, thresholding**

I. INTRODUCTION

Document Image Binarization is the pre-processing step for document image analysis and processing. It enhances the performance of document processing techniques like OCR and layout analysis. Image Binarization is the conversion of document image into bi-level document image. Image pixels are separated into dual collection of pixels, i.e. black and white. The main goal of image binarization is the segmentation of document into foreground text and background.

The simplest approach to binarization is thresholding. In thresholding an optimal threshold value is chosen and the pixels are classified as foreground or background by comparison with this threshold value. But in real life, some documents suffer from degradation like uneven illumination, background noise, bleed through and variation in illumination and contrast. To choose an optimal threshold value for such documents is a challenging task [1]. The wrong estimation of threshold value results in misclassification of the pixels as foreground or background [2]. This affects the results of binarization and the accuracy of pattern recognition applications.

Generally, binarization is categorized as Global and Local [3]. Global binarization is the technique in which a single threshold value is applied to binarize the entire image. It is fast process but fails for the documents with complex background. In Local binarization technique instead of single threshold for the whole image, a different value of threshold is chosen for every pixel. The threshold is chosen depending upon the neighbourhood pixels. Local thresholding do not give good results for the documents suffering from background noise.

Binarization techniques can also be categorized based upon the criteria to choose the threshold value. Some of the methods to choose threshold are based upon histogram, clustering, entropy, local adaptive methods. Other methods uses decision tree [4], combination of various binarization technique [5] and iterative method [6].

II. REVIEW OF BINARIZATION TECHNIQUES

A. Otsu's Method

Otsu method [7] is a global thresholding method that converts gray scale image into bi-level image. This technique divides the pixels into two classes one is foreground and the other is background. It chooses an optimal threshold that separates the image into two different classes. The threshold value is chosen such that the within-class variance is minimized and the between-class variance is maximized [3]. The weighted within class variance of two classes is given as shown in Eq. 1

$$\sigma^2 p(t) = p_1(t)\sigma_1^2 + p_2(t)\sigma_2^2(t) \qquad (1)$$

Otsu method gives its best performance for only those images that have clear bi-modal pattern. But, degraded documents normally don't have such clear-cut pattern. Besides this, it does not perform well for images with uneven illumination and shadow.

B. Niblack's Method

Niblack [8] is a local thresholding method. In local thresholding methods, a different threshold value is calculated for each and every pixel. It uses local statistics of the image, such as variance, range to calculate the threshold. In Niblack method a rectangular window is slided over the gray scale image to estimate threshold of the pixels. It uses the local statistics mean and standard deviation of the window to estimate the threshold. Threshold T(i,j) is estimated as shown in Eq. 2

$$\mathrm{T(i,j)} = \mu + \mathrm{k} \times \sigma \tag{2}$$

In Eq.2 μ represents the mean of the window and σ represents the standard deviation of the window. The value of k is a constant and it defines the size and quality of binarization.

As this method is dependent upon the local features of the image, it gets affected by blank areas in the image and is also not efficient for the images with background noise.

C. Sauvola Method

Sauvola method [9] is the improvement of the Niblack method. It is local variance method that uses standard deviation. Threshold is calculated as shown in Eq. 3

$$\mathrm{T(i,j)} = \mu \times [1 + \mathrm{k}\frac{\sigma}{\mathrm{R}} - 1] \tag{3}$$

In Eq. 3 μ is the mean and σ is the standard deviation of the window. Values suggested for k and R are 0.5 and 128. The window size and value of k will affect the quality of image but R will have very little affect. This method is used for documents having uneven illumination, light texture and stained documents. But, Sauvola method thins the text after its application.

D. Bernsen

Bernsen method [10] uses the contrast of the image. The threshold is estimated as the average of highest and lowest intensity values in the window. The Eq. 4 calculates the local contrast of the window

$$C(i.j) = I_{max} - I_{\min} \tag{4}$$

The pixels are categorized as foreground or background by comparing the local contrast with threshold value. The pixel will be classified as background, if the local contrast is found to be less than the threshold and vice-versa. Bernsen method doesn't perform well for the images having more complex background.

E. Local Maxima and Minima

This method [11] uses contrast which depends upon the local minimum and maximum. A normalization factor is introduced which will compensate the effect of variation in image background. Image contrast is calculated as shown in Eq. 5

$$C(i,j) = \frac{I_{max} - I_{min}}{I_{\max} - I_{min} + \mathcal{E}} \tag{5}$$

High contrast image pixels are found from the contrast image. Then local thresholding is performed with threshold value calculated from the found high contrast image pixels. It is not suitable for the bright text with proper bright background.

F. Adaptive Contrast

Adaptive contrast [12] is the consolidation of image contrast and image gradient. Adaptive contrast method removes the over-normalization problem of Local maxima minima method. Contrast map is constructed and binarized. Canny edge map is also constructed and combined with the resultant of the first step. This detects the actual text stroke edges. The text is extracted using local thresholding. It is estimated from the mean and standard deviation of the detected text stroke pixels within a window. The limitation of this particular method is that the canny edge detector extracts the fake edges also.

G. Global-to-Local Approach

Global thresholding followed by local thresholding is performed to binarize the image. In pre-processing Gaussian filter is used. Edge map is constructed using canny edge detector. Histogram of the pixels other than the edge pixels is obtained. The lower of the two clear peaks of the histogram is chosen as global threshold and the pixel values exceeding this threshold are turned background. Remaining pixels are classified as foreground or background depending upon the local threshold. Local threshold is estimated using the average of the highest and lowest gray value of the window. Shrink and Swell filters are used to amend the final results of binarization in post-processing [13].

H. Morphological Contrast Intensification

Gray scale morphological tools are used for background estimation and contrast intensification. It is a hybrid approach of global and local thresholding. Initially size of structuring element is chosen. The histogram of the distance between the consecutive edges is used to get the size of the structuring element. Background of the image is estimated using morphology. Global threshold is estimated by morphological contrast intensification. Text regions are separated with the help of histogram of contrast image. Local thresholding is performed to binarize the document image. Post processing is done for removing the noise from the binarized image [14].

III. PERFORMANCE METRICES

The metrics that are suitable for the comparison of the performance of different binarization algorithms are PSNR, F-Measure, NRM and MPM.

A. PSNR

PSNR is used to check the similarity between two images. It is used for images having noise. PSNR is given as shown in Eq. 6

$$\mathrm{PSNR} = 10\log_{10}\frac{\mathrm{C} \times \mathrm{C}}{\mathrm{MSE}} \tag{6}$$

Where C is a constant and MSE (mean square error) depicts the difference between the distorted image and the original image. The value of PSNR should be more for better results [15].

B. SNR

SNR is the ratio between signal and noise. Higher the value of SNR lower is the noise.

C. PRECISION

Precision is true pixels extracted divided by total pixels extracted. Precision is given as shown in Eq. 7

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

TP denotes true positive i.e. the pixels that are foreground in both ground truth and binarized image. FP denotes false positive i.e. pixels identified as foreground in the binarized image but are actually background in the ground truth image.

D. RECALL

Recall is the true pixels extracted divided by the total number of true pixels. Recall is given as shown in Eq. 8

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

TP denotes true positive i.e. the pixels that are foreground in both ground truth and binarized image. FN denotes false negative i.e. the pixels identified as background in binarized image but are actually foreground in ground truth image.

F-MEASURE

F-Measure is given as shown in Eq. 9

$$\text{F} - \text{Measure} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{9}$$

F-Measure is the harmonic mean of Precision and Recall. Its value should be high for better results[13].

F. NRM

NRM is calculated using the unmatched pixels between the binarized image and ground truth image.

Negative Rate Metric is given as shown in Eq. 10

$$\text{NRM} = \frac{\text{R}_{\text{FN}} + \text{R}_{\text{FP}}}{2} \tag{10}$$

Where $\text{R}_{\text{FN}} = \frac{\text{FN}}{\text{FN+TP}}$

$$\text{R}_{\text{Fp}} = \frac{\text{FP}}{\text{FP+TN}}$$

Where R_{FN} is the rate of false negatives and R_{FP} is the rate of false positives. Lesser the value of NRM better is the result of binarization.

G. Misclassification Penalty Metric(MPM)

This metric is used to evaluate how the binarized image constitutes the contour of ground truth image. MPM is given as shown in Eq. 11

$$\text{MPM} = \frac{\text{MP}_{\text{FN}} + \text{MP}_{\text{FP}}}{2} \tag{11}$$

Where $\text{MP}_{\text{FN}} = \sum_{i=1}^{\text{NFN}} \frac{d_{\text{FN}}^{i}}{\text{D}}$

$$\text{MP}_{\text{FP}} = \frac{d_{\text{FP}}^{i}}{\text{D}}$$

d_{FN}^{i} represents i^{th} false negative and d_{FP}^{i} represents j^{th} false positive pixel. Normalization factor is given by D, the sum of overall pixel-to-contour distances of the GT object. Smaller the value of MPM better is the quality of the algorithm.

IV. COMPARISON OF TECHNIQUES ON DIBCO 2009 AND H-DIBCO 2010 DATASET

The results of different algorithms are compared in this section. The dataset of DIBCO-2009 and H-DIBCO 2010 is provided by the Document Image Binarization Contest (in the framework of ICDAR 2009) and Handwritten Document Image Binarization Contest (in conjunction with ICFHR 2010) respectively online. DIBCO 2009 contains 10 document images (5 printed and 5 handwritten) and H-DIBCO 2010 contains 10 handwritten document images. The results are cited from the literature.

TABLE I: RESULTS ON DIBCO 2009[10][11]

Method	F-Measure	PSNR	NRM	MPM
Otsu	78.72	15.34	5.77	13.3
Niblack	55.82	9.89	16.4	61.5
Sauvola	85.41	16.39	6.94	3.2
Bernsen	52.48	8.89	14.29	113.8
LMM	91.06	18.5	7	0.3
AC	93.5	19.65	3.74	0.43
Global-Local	83.66	15.60	4.80	32.9

TABLE II: RESULTS ON H-DIBCO 2010[10][12]

Method	F-Measure	PSNR	NRM	MPM
Otsu	85.27	17.51	9.77	1.35
Niblack	74.1	15.73	19.06	1.06
Sauvola	75.3	15.96	16.31	1.96
Bernsen	41.3	8.57	21.18	115.98
LMM	85.49	17.83	11.46	0.37
AC	92.03	20.12	6.14	0.25
MCI	90.67	19.15	4.7	0.6

V. CONCLUSION

This paper is a survey of various document image binarization techniques. Evaluation of results is shown on DIBCO 2009 and H-DIBCO 2010 dataset. Different techniques are compared on the basis of PSNR, F-Measure, NRM and MPM. Different techniques work for different type of degradation. No single technique can be generalized for all types of degradations. From the techniques discussed above Adaptive Contrast outperforms all the other techniques for both the datasets.

REFERENCES

[1] L. Jiang, K. Chen, S. Yan, Y. Zhou, and H. Guan, “Adaptive Binarization for Degraded Document Images”, IEEE, pp. 5–8, 2009.

[2] N. Chaki, S. H. Shaikh, and K. Saeed, “Exploring Image Binarization Techniques”, Studies in Computational Intelligence, vol. 560, pp. 5–16, 2014.

[3] T. K. Gill, “Document Image Binarization Techniques- A Review”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 5, May 2014

[4] C. Chou, W. Lin, and F. Chang, “Learning to Binarize Document Images Using A Decision Cascade”, IEEE., no. iii, 2009.

[5] B. Gatos, I. Pratikakis, and S. J. Perantonis, “Improved Document Image Binarization by Using a Combination of Multiple Binarization Techniques and Adapted Edge Information”, IEEE, pp. 1–4, 2008.

[6] S. H. Shaikh, A. Maiti, and N. Chaki, “Image binarization using iterative partitioning: A global thresholding approach,” International Conference on Recent Trends in Information Systems Image, ReTIS 2011 – Proceedings, pp. 281–286, 2011.

[7] P. Smith, D. B. Reid, L. Palo, P. Alto, and P. L. Smith, “Smith et al. - 1979 - A Tlreshold Selection Method from Gray-Level Histograms,” IEEE Transactions On Systrems, Man, And Cybernetics, Vol. Smc-9, No. 1, January 1979.

[8] Niblack, W.: An introduction to digital image processing, pp. 115–116. Prentice Hall, Eaglewood Cliffs (1986)

[9] J. Sauvola and M. Pietikäinen, “Adaptive document image binarization,” *Pattern Recognition Society.*, vol. 33, no. 2, pp. 225–236, 2000.

[10] Bernsen, J.: Dynamic thresholding of gray level images.”, Proceedings of International Conference on Pattern Recognition (ICPR), pp. 1251–1255 (1986)

[11] S. Lu, F. Way, and C. L. Tan, “Binarization of Historical Document Images Using the Local Maximum and Minimum,” *Image (Rochester, N.Y.)*, 2010.

[12] B. Su, S. Lu, C. L. Tan, and S. Member, “Robust Document Image Binarization Technique for Degraded Document Images,” IEEE Transactions On Image Processing, Vol. 22, No. 4, pp. 1408-1417, April 2013

[13] B. Biswas, U. Bhattacharya, and B. B. Chaudhuri, “A Global-to-Local Approach to Binarization of Degraded Document Images”, 22nd International Conference on Pattern Recognition, pp. 3008–3013, 2014.

[14] S. Mandail, A. Agarwae, and B. Chanda, “Binarization of Degraded Handwritten Documents Based On Morphological Contrast Intensification”, Third International Conference on Image Infornation Processing, pp. 73–78, 2015.

[15] G. Chutani, T. Patnaik and V. Diwedi, “An Improved Approach for automatic denoising and Binarization of Degraded Document Images Based on Region Localization”, IEEE, pp. 2272–2278, 2015.