REVIEW ARTICLE

An Analysis of Image Denoising and Restoration of Handwritten Degraded Document Images

Sarika Jain[1], Pankaj Parihar[2]

[1]Department of computer science and Engineering, Institute of Technology and Management, Bhilwara, India

[2]Department of computer science and Engineering, Institute of Technology and Management, Bhilwara, India

[1] sarikajain03@gmail.com; [2] pankajsinghparihar2002@gmail.com

Abstract— The restoration of a blurry or noisy image is commonly performed with a MAP estimator, which maximizes a posterior probability to reconstruct a clean image from a degraded image. A MAP estimator, when used with a sparse gradient image prior, reconstructs piecewise smooth images and typically removes textures that are important for visual realism. The three public datasets that were used in the recent. Document Image Binarization Contest (DIBCO) 2009 & 2011 and Handwritten Document Image Binarization Contest (H-DIBCO) 2010 and achieves different accuracies. Experiments on the Bickley diary dataset that consists of several challenging bad quality document images also show the superior performance in image binarization technique which is compared with different techniques. The general objective is to identify current advantages in document image binarization using established evaluation performance measures.

Keywords— Image Processing, Pixel Classification, Degraded Document, Image Binarization, Adaptive Image Contrast

I. INTRODUCTION

Degradations in document images result from poor quality of paper, the printing process, ink blot and fading, document aging, extraneous marks, noise from scanning, etc. The goal of document restoration is to remove some of these artifacts and recover an image that is close to what one would obtain under ideal printing and imaging conditions. The ability to restore a degraded document image to its ideal condition would be highly useful in a variety of fields such as document recognition, search and retrieval, historic document analysis, law enforcement, etc. The emergence of large collections of scanned books in digital libraries [1, 10] has introduced an imminent need for such restorations that will aid their recognition or ability to search. Images with certain known noise models can be restored using traditional image restoration techniques such as Median filtering, Weiner filtering, etc. [9]. However, in practice, degradations arising from phenomena such as document aging or ink bleeding cannot be described using popular image noise models. Document processing algorithms improve upon the generic methods by incorporating document specific degradation models [20] and text specific content models [2, 16]. Document image binarization is an important step in the document image analysis. It's aims to segment the foreground text from the document background. A fast and accurate document image binarization technique is important for the ensuing document image processing tasks such as optical character recognition (OCR). As illustrated in Figure 1, the handwritten text within the degraded documents often shows a certain amount of variation in terms of the stroke width, stroke brightness, stroke connection, and document background. In addition, historical documents are often degraded by the bleed-through as illustrated in Figure 1(a) and (c) where the ink of the other side seeps through to the front. In addition, historical documents

are often degraded by different types of imaging artifacts as illustrated in Figure 1(e). These different types of document degradations tend to induce the document thresholding error and make degraded document image binarization a big challenge to most state-of-the-art techniques. The recent Document Image Binarization Contest (DIBCO) [1], [2] held under the framework of the International Conference on Document Analysis and Recognition (ICDAR) 2009 & 2011 and the Handwritten Document Image Binarization Contest(H-DIBCO)[3]held under the framework of the International Conference on Frontiers in Handwritten Recognition show recent efforts on this issue.
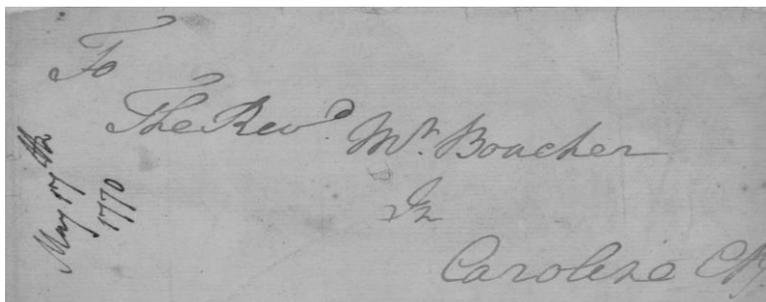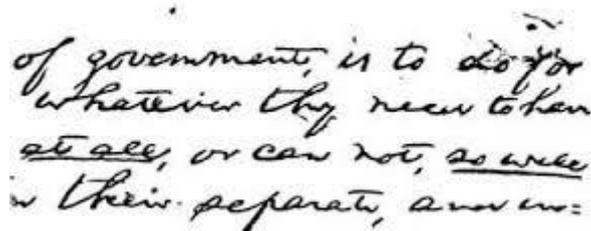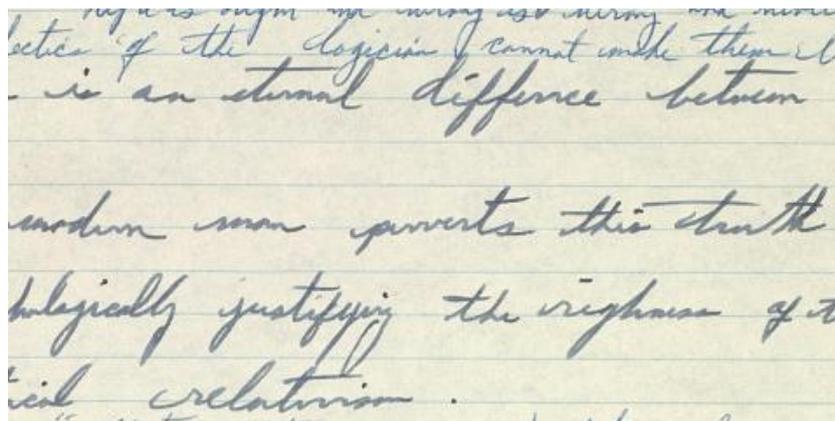

(a)
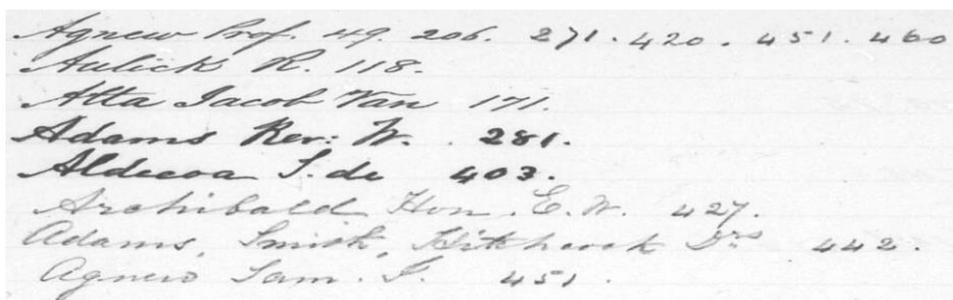
(b)
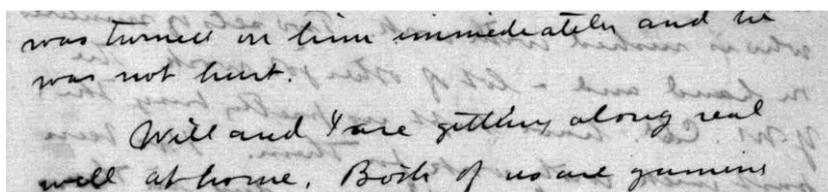
(c)

(d)

(e)

Fig.1 Five degraded document image examples taken from DIBCO, H-DIBCO and Bickley diary datasets.

II. RELATED WORK

A number of thresholding techniques [3], [6], [8], [11] have been reported for document image binarization. As many degraded documents do not have a clear bimodal pattern, global thresholding [10], [11], [12], [13] is usually not a suitable approach for the degraded document binarization. Adaptive thresholding [11],[12], which estimates a local threshold for each document image pixel, is often a better approach to deal with different variations within degraded document images. For example, the early window-based adaptive thresholding techniques [8], [9] estimate the local threshold by using the mean and the standard variation of image pixels within a local neighborhood window. The main drawback of these window-based thresholding techniques is that the thresholding performance depends heavily on the window size and hence the character stroke width. Other approaches have also been reported, including background subtraction [4] texture analysis, recursive method decomposition method and combination of binarization techniques .These methods combine different types of image information and domain knowledge and are often complex. The local image contrast and the local image gradient are very useful features for segmenting the text from the document background because the document text usually has certain image contrast to the neighboring document background. They are very effective and have been used in many document image binarization techniques [5], [9],[15], [16]. In Bernsen's paper [14], the local contrast is defined as follows:

$$C(i, j) = Imax(i, j) - Imin(i, j) \ldots\ldots\ldots(1)$$

where $C(i, j)$ denotes the contrast of an image pixel (i, j), $Imax(i, j)$ and $Imin(i, j)$ denote the maximum and minimum intensities within a local neighborhood windows of (i, j), respectively. If the local contrast $C(i, j)$ is smaller than a threshold, the pixel is set as background directly. Otherwise it will be classified into text or background by comparing with the mean of $Imax(i, j)$ and $Imin(i, j)$. Bernsen's method is simple, but cannot work properly on degraded document images with a complex document background. In a novel document image binarization method [5] by using the local image contrast that is evaluated as follows [14]

$$C(i, j)=[Imax(i, j)-Imin(i, j)]/[Imax(i, j)+Imin(i, j)+e] \ldots\ldots.(2)$$

where e is a positive but infinitely small number that is added in case the local maximum is equal to 0. Compared with Bernsen's contrast in Equation 1, the local image contrast in Equation 2 introduces a normalization factor (the denominator) to compensate the image variation within the document background.

III. METHODS

This section describes the document image binarization techniques. Given a degraded document image, an adaptive contrast map is first constructed and the text stroke edges are then detected through the combination of the binarized adaptive contrast map and the canny edge map. The text is then segmented based on the local threshold that is estimated from the detected text stroke edge pixels.

A. Image Contrast Construction

The image gradient has been widely used for edge detection [2] and it can be used to detect the text stroke edges of the document images effectively that have a uniform document background. On the other hand, it often detects many non-stroke edges from the background of degraded document that often contains certain image

592

variations due to noise, uneven lighting, bleed-through, etc. To extract only the stroke edges properly, the image gradient needs to be normalized to compensate the image variation within the document background. In our earlier method [5], The local contrast evaluated by the local image maximum and minimum is used to suppress the background variation as described in Equation 2.

B. *Pixel Edge Detection*

The purpose of the contrast image construction is to detect the stroke edge pixels of the document text properly. The constructed contrast image has a clear bi-modal pattern [5], where the adaptive image contrast computed at text stroke edges is obviously larger than that computed within the document background. We therefore detect the text stroke edge pixel candidate by using Otsu's global thresholding method. The binary map can be further improved through the combination with the edges by Canny's edge detector, because Canny's edge detector has a good localization property that it can mark the edges close to real edge locations in the detecting image.

C. *Threshold Estimation*

The text can then be extracted from the document background pixels once the high contrast stroke edge pixels are detected properly. Two characteristics can be observed from different kinds of document images [5]: First, the text pixels are close to the detected text stroke edge pixels. Second, there is a distinct intensity difference between the high contrast stroke edge pixels and the surrounding background pixels.

IV.   EVALUATION MEASURES

For the evaluation, the measures used comprise an ensemble of measures that have been widely used for evaluation purposes. These measures consist of:

(i) F-Measure
(ii) Negative Rate Metric and
(iii) Misclassification Penalty Metric

Definitions:

(i)   F-Measure:

$$FMeasure = \frac{(2 \times Recall \times Precision)}{(Recall + Precision)} \quad (1)$$

$$\text{Where} \quad Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

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

(ii)   Negative Rate Metric (NRM):

The negative rate metric NRM is based on the pixelwise mismatches between the GT and prediction. It combines the false negative rate NRFN and the false positive rate NRFP. It is denoted as follows:

$$NRM = \frac{NR(FN) + NR(FP)}{2}$$

$$\text{Where} \quad NR_{FN} = \frac{N(FN)}{N(FN) + N(TP)}$$

$$NR_{FP} = \frac{N(FP)}{N(FP) + N(TN)} \quad (3)$$

593

N_{TP} denotes the number of true positives, N_{FP} denotes the number of false positives, N_{TN} denotes the number of true negatives, N_{FN} denotes the number of false negatives. In contrast to F-Measure and PSNR, the binarization quality is better for lower NRM.

(iii) Misclassification penalty metric (MPM):

The Misclassification penalty metric MPM evaluates the prediction against the Ground Truth (GT) on object

-by-object basis. Misclassification pixels are penalized by their distance from the ground truth object's border.

$$MPM = \frac{MP(FN) + MP(FP)}{2} \quad (4)$$

$$\text{Where} \quad MP_{FN} = \frac{\sum_{i=1}^{N(FN)} d^{\wedge}i(FN)}{D}, \quad MP_{FP} = \frac{\sum_{j=1}^{N(FP)} d^{\wedge}j(FP)}{D}$$

d^i_{FN} and d^j_{FP} denote the distance of the ith false negative and the jth false positive pixel from the contour of the GT segmentation. The normalization factor D is the sum over all the pixel to-contour distances of the GT object. A low MPM score denotes that the algorithm is good at identifying an object's boundary.

V.   CONCLUSION AND FUTURE SCOPE

This paper concludes the different methods that have been tested on various datasets. The DIBCO 2009 Document Image Binarization Contest attracted 35 research groups that are currently active in document image analysis. The increased interest in this competition is a two-fold proof: first, it shows the importance of binarization as a step towards effective document image recognition and second, the need for pursuing a benchmark that will lead to a meaningful and objective evaluation.

REFERENCES

[1] Digital library of India. http://dli.iiit.ac.in/.

[2] E. Borenstein and S. Ullman. Combined top-down/bottomup segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(12):2109–2125, 2008.

[3] H. Cao and V. Govindaraju. Handwritten carbon form preprocessing based on markov random field. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007.

[4] K. Donaldson and G. K.Myers. Bayesian super-resolution of text in video with a text-specific bimodal prior. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[5] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1):41–54, 2006.

[6] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Examplebased super-resolution. *IEEE Comput. Graph. Appl.*, 22(2):56–65, 2002.

[7] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 146–165, 2004.

[8] O. D. Trier and A. K. Jain, "Goal-directed evaluation of binarization methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1191–1201, 1995.

[9] O. D. Trier and T. Taxt, "Evaluation of binarization methods for document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 312–315, 1995.

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

[11] J. Kittler and J. Illingworth, "On threshold selection using clustering criteria," *IEEE transactions on Systems, Man, and Cybernetics,*vol. 15, pp. 652–655, 1985.

[12] N. Otsu, "A threshold selection method from gray level histogram," *IEEE Transactions on System, Man, Cybernetics*, vol. 19, no. 1, pp. 62–66, January 1978.

[13] N. Papamarkos and B. Gatos, "A new approach for multithreshold selection," *Computer Vision Graphics and Image Processing*, vol. 56, no. 5, pp. 357–370, 1994.

[14] H. Lu, A. Kot, and Y. Shi. Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters*, 11(2):228–231, 2004.

[15] H. Q. Luong and W. Philips. Robust reconstruction of lowresolution document images by exploiting repetitive character behaviour. *International Journal of Document Analysis and Recognition*, 11(1):39–51, 2008.

[16] G. Myers and K. Donaldson. Bayesian super-resolution of text in video with a text-specific bimodal prior. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1188–1195, 2005.