A COMPREHENSIVE REVIEW ON ANALYSIS OF IMAGE BINARIZATION FOR DEGRADED DOCUMENTS

Geetanjali Thakur

[1]*Computer Science and Engineering, Indo Global Colleges of Engineering*

Punjab Technical University, Jalandhar, (India)

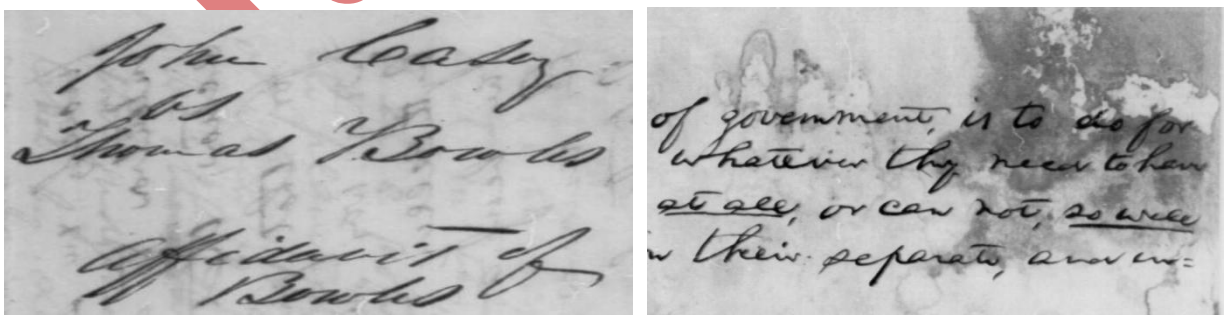
ABSTRACT

Document binarization is an active research area for many years. In the degraded document images, text segmentation is very challenging task because of variation in background and foreground text of different document images. 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. The general objective is to identify current advances in document image binarization using established evaluation performance measures.

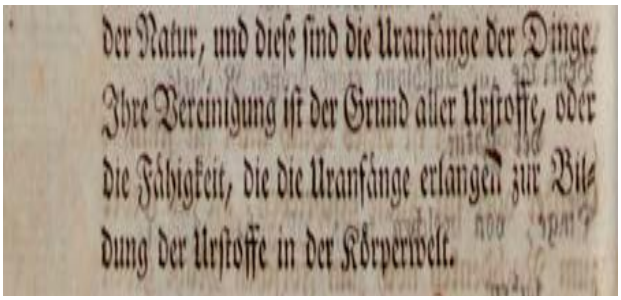
Keywords: Document Analysis, Document Image Processing, MATLAB.

I.INTRODUCTION

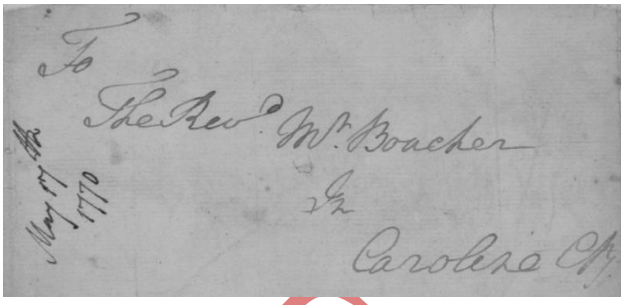
Document image binarization is an important step in the document image analysis. It's aims to segment the foreground text from the document background. As illustrated in Figure 1, the handwritten text within the degraded documents often shows a certain amount of variation in terms of the stroke width, stroke brightness, stroke connection, and document background. Document image binarization is usually performed in the document preprocessing stage.



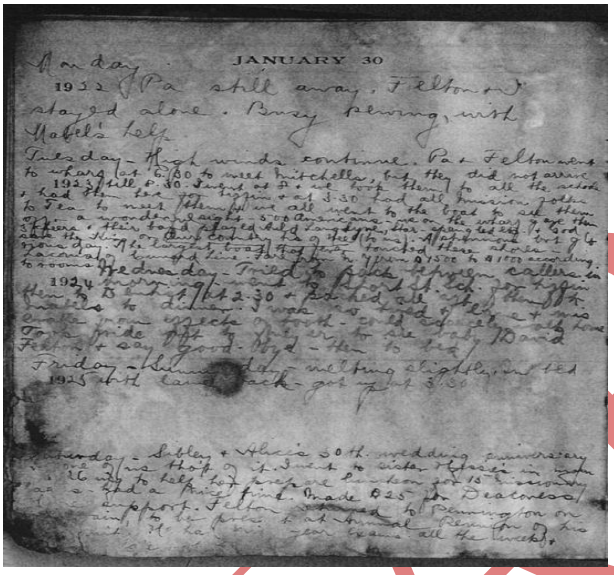
(a) (b)



(c)



(d)



(e)

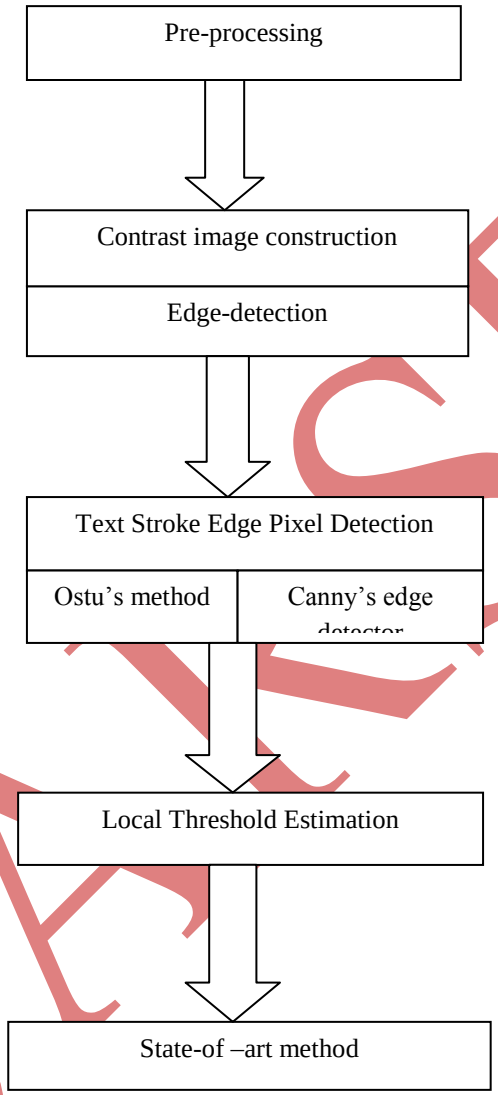
Fig.1.a,b,c,d,e [6].Five degraded document image examples taken from DIBCO, H-DIBCO and Bickley diary datasets.

II.METHODOLOGY

2.1 Document Image Binarization Technique

In document image binarization technique, an adaptive contrast map is first constructed and the text stroke edges are then detected through the combination of the binarization adaptive contrast map and the canny edge map. Text segmentation is based on the local thresholding that is estimated from the detected text stroke edge pixel.

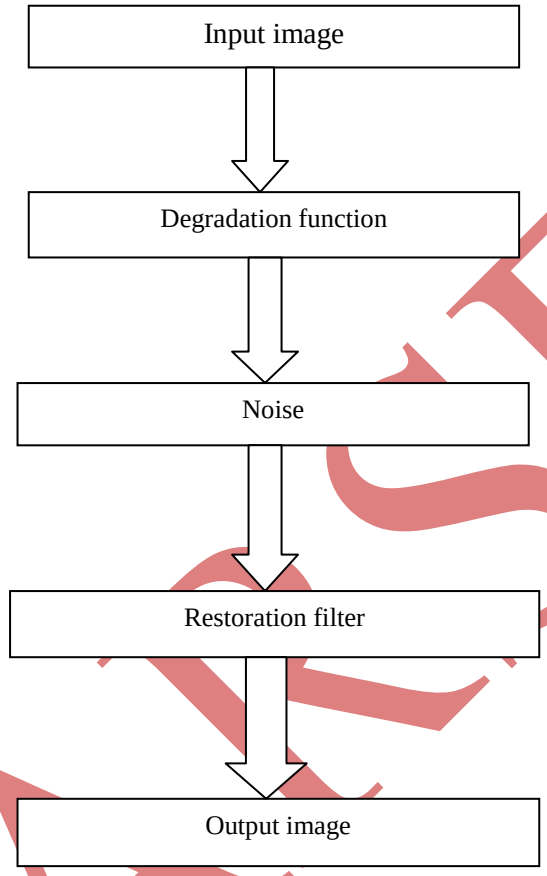
BLOCK DIAGRAM



2.2 Image Restoration

Restoration attempts to reconstruct or recover an image that has been degraded by using a prior knowledge of the degraded phenomenon **.**Restoration technique are oriented towards modeling the degraded and applying the inverse process in order to recover the original image.

BLOCK DIAGRAM



III CONCLUSIONS 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

Journal Papers:

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

[2] 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.

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

[4] J. Bernsen, "Dynamic thresholding of gray-level images," International Conference on Pattern Recognition, pp. 1251–1255, October 1986.

[5] L. Eikvil, T. Taxt, and K. Moen, "A fast adaptive method for binarization of document images,"International Conference on Document Analysis and Recognition, pp. 435–443, September 1991.

[6] Bolan Su and Chew Lin,Tan,"A Robust Document image binarization technique for degraded document images" IEEE transactions on Image processing,vol.22,pp.841-1263,April-2013.

Books:

[7] R.Gonzalez , R. Woods & Steven L, 'Digital image processing using MATLAB.