# An Efficient Binarization Technique for Recovering Degraded Document Images

**Archana Thange[1], Amina.N[2], Amruta Nimbalkar[3]**

[1] ME (IT), DKGOI'S COE, Swamichincholi, Maharashtra, India
[2] ME (COMPUTER), DKGOI'S COE, Swamichincholi, Maharashtra, India
[3] ME (COMPUTER), DKGOI'S COE, Swamichincholi, Maharashtra, India

***Abstract:*** *Degraded document images is very difficult to segment in text format due the high inter /intra variation between the document background and foreground text from different document images. In this paper we propose a binarization technique for recovering degraded document images. In the proposed technique firstly adaptive image contrast map is constructed by giving input degraded document images and detect the text stroke edge pixel. Text of the document is segmented by using local threshold estimation then further applying the post processing to improve document binarization quality. The proposed method is simple, effective and involves minimum parameter tuning.*

***Key Words*****:** *Degraded document images, adaptive image contrast, binarization technique.*

## 1. INTRODUCTION

The document image binarization has been studied for many years; the thresholding of document image is still unsolved problem. The handwritten text within the degraded document images shows some variation such as brightness, stroke connection, stroke width and some historical degraded document shows. Some variation in terms of ink of the other side seeps through to the front. So we can use the different binarization technique to improve the degraded document images.

As time passes, the documents degrade making the data unreadable. We need to recover the data from this degraded document. Various other techniques were proposed to recover data, but were less efficient. To provide maximum accuracy and exact recovery of documents, we propose a robust technique for recovery of degraded documents.

## 2. OBJECTIVES

Main objective is to make use of the adaptive image contrast that combines the local image contrast and the local image gradient adaptively and therefore is tolerant to the text and background variation caused by different types of document degradations. In particular, the proposed technique addresses the over-normalization problem of the local maximum minimum algorithm. Proposed system presents a document binarization technique that extends our previous local maximum-minimum method.

### 2.1 SCOPE

Document Image Binarization system is image processing based system.
1. To improve the quality of novel document images.
2. Used for different kinds of degraded document images.

## 3. ARCHITECTURE OF PROPOSED SYSTEM

In this section we propose four methods, which are used in binarization technique. Such as

1. Contrast Image Construction
2. Text Stroke Edge Pixel Detection
3. Local Threshold Estimation
4. Post-Processing

Description

1. Contrast Image Construction

Adaptive image contrast is a combination of local image gradient and local image contrast.

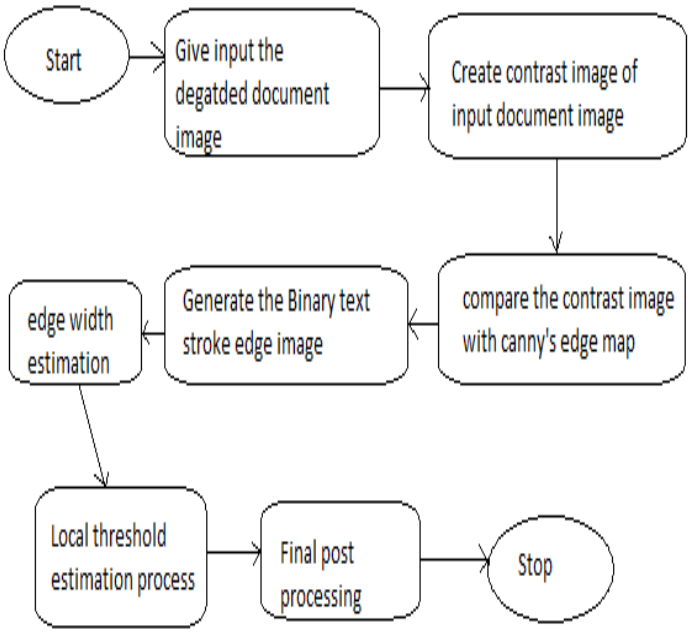


Figure 1:- Architecture of system

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 [3] [4].

Ideally, the image contrast will be assigned with a high weight (i.e. large $\alpha$) when the document image has significant intensity variation. So that the proposed binarization technique depends more on the local image contrast that can capture the intensity variation well and hence produce good results. $C\alpha(i,j) = \alpha C(I,j) + (1-\alpha)(Imax(I,j) - Imin(I,j))$.

The adaptive combination of the local image contrast and the image gradient in above equation can produce proper contrast maps for document images with different types of degradation.

## 2. Text Stroke Edge Pixel Detection

The purpose of the contrast image construction is to detect the stroke edge pixels of the document text properly. The pixel at both sides of the text stroke will be selected as the high contrast pixel.

## 3. Local Threshold Estimation

After detecting the text stroke edge pixel, we calculate the most frequent distance between two adjacent edge pixels in horizontal direction and use it as the estimated stroke width.

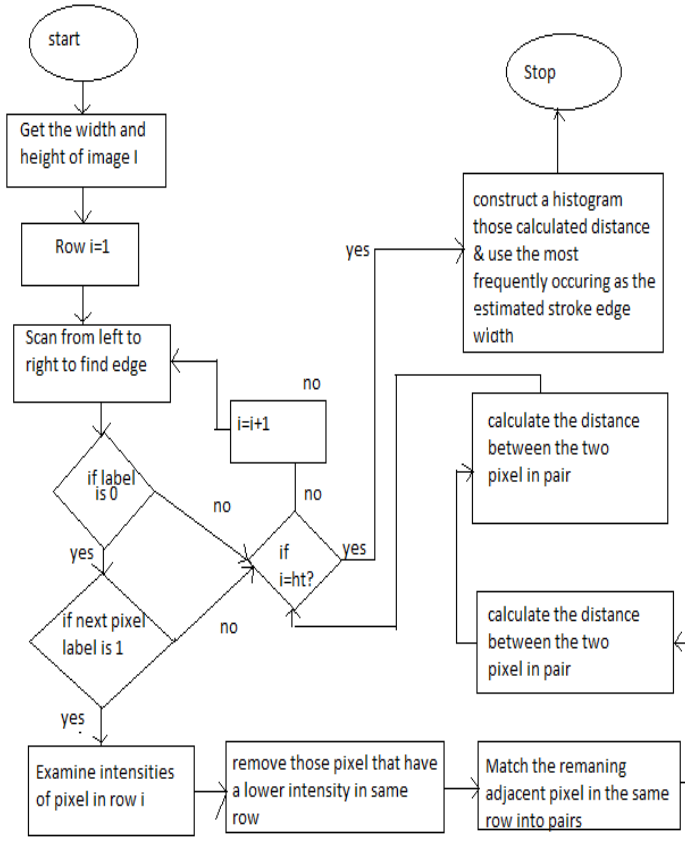


Figure 2:- Flowchart of local threshold estimation

## 4. Post-Processing

Binarization result can be improved by using Post-Processing method. By using the algorithm of post processing we remove single pixel artifacts along the text stroke boundaries after the document thresholding.

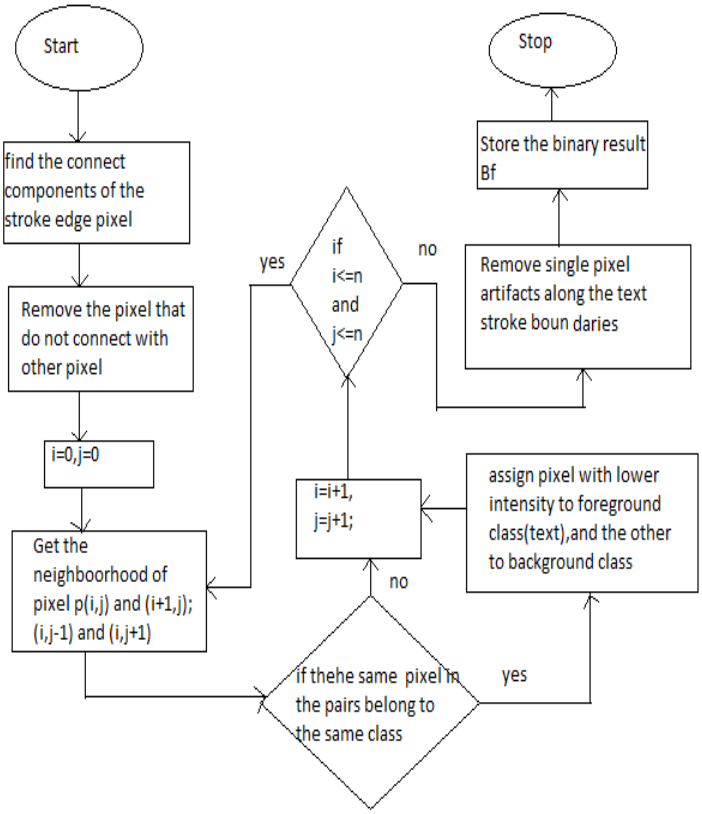


Figure 3:- Flowchart of Post-Processing

## CONCLUSION

The proposed system presents an adaptive image contrast based document image binarization technique that is tolerant to different types of document degradation such as uneven illumination and document smear. This proposed system has its work in all areas that are concerned with managing documents for preserving historical data. The Proposed system emphasizes on recovering the data that tends to be damaged due to damaging of the documents.

## REFERENCES

[1] Bolan Su, Shijian Lu, and Chew Lim Tan, Senior Member, IEEE, *"Robust Document Image Binarization Technique for Degraded Document Images"*,IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 4, APRIL 2013.

[2] Gatos, K. Ntirogiannis, and I. Pratikakis, "*ICDAR 2009 document image Binarization contest (DIBCO 2009),*" in *Proc. Int. Conf. DocumentAnal. Recognition.*, Jul. 2009, pp. 1375–1382.

[3] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "*ICDAR 2011 document image binarization contest (DIBCO*

*2011)*," in *Proc. Int. Conf. DocumentAnal. Recognit.*, Sep. 2011, pp.1506–1510.

[4] Pratikakis, B. Gatos, and K. Ntirogiannis, "*H-DIBCO 2010 handwritten document imagebinarization competition*," in *Proc. Int. Conf.Frontiers Handwrit. Recognit.*, Nov. 2010, pp. 727–732.

[5] S. Lu, B. Su, and C. L. Tan, "*Document image binarization using background estimation and stroke edges*," *Int. J. Document Anal. Recognit.*, vol. 13, no. 4, pp. 303–314, Dec. 2010.