International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 6

ISSN: 2321-8169
4104 - 4106

Recover Degraded Document images Using Binarization Technique

Nimbalkar Amruta A

ME Computer Engineering,
DGOI,FOE swami chincholi, pune,
Maharashtra
India.

Prof.Amrit Priyadarshi

Assistant professor
ME computer Engineering
DGOI,FOE swami chincholi , pune, Maharashtra
India.

Abstract: In now a days,whole world is connected through the internet. The different types of data ,we can save,copy and backup in the digital form. But old data which is in the form of traditional paper. This old data plays important role in a major task.Many of the paper data is being degraded due to lack of reason. The front and rear data are mix up together so segmention of text from badly degraded document is very challenging task.To solve this problem by using binarization technique. In this paper ,we propose four binarization technique for recovering degraded document images.we firstly apply contrast inversion mechanism on degraded document images. The contrast map is then converted to grayscale image so as to clearly identify the text stroke from background and foreground pixels.Detected text is further segmented using local threshold method that is estimated based on intensities of detected text stroke edge pixel.Finally applying post processing to improve the quality of degraded document images.This binarization technique is simple,robust and efficient for recovering degraded document images.

Keywords-*Binarization Technique,Contrast Inversion,Degraded Document Image,Threshold Estimation*

****

I. INTODUCTION

Document image binarization technique is used for segment the foreground text from the document background.A fast and accurate document image binarization technique is important for the ensuring document image processing tasks such as optical character recognition.In now days,there are different types Degraded documents are available ,which is in unreadable format due to lack of attentions,such as foreground text mix up with background text. In historical documents,other sides of ink seeps through to the front.In Handwritten text within degraded documents often shows different amount of variation like stroke width,stroke brightness,stroke connection etc.

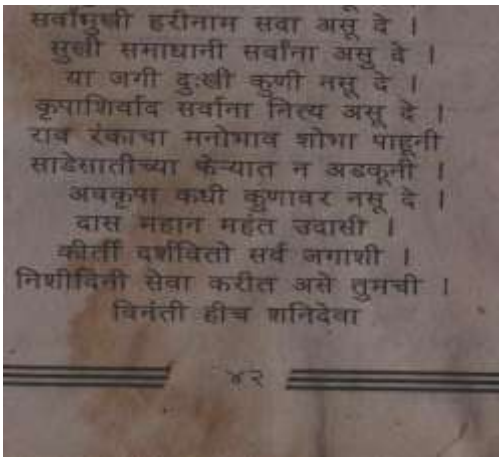

Fig 1- Example of Degraded document image

As time passes, Text data on document is in unredable format. 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 binarization technique for recovery of degraded documents.

II. LITERATURE SURVEY

In this paper ,they proposed the binarization technique which is useful for degraded document image.firstly they construct the adaptive contrast map which is a combination of local image contrast and local image gradient.constrast map is combine through canny edge map.canny edge detector is used for detection of text edge. The drawback of canny edge detector is they does not detect inside text edge.[1] Many thresholding techniques [2]have been reported for document image binarization. As many degraded documents
do not have a clear bimodal pattern, global thresholding [3] is usually not a suitable approach for the degraded document binarization.

III. PROPOSED METHODOLOGY

In this section we describes four binarization technique .which is useful for recovering degraded document images.

ISSN: 2321-8169
4104 - 4106

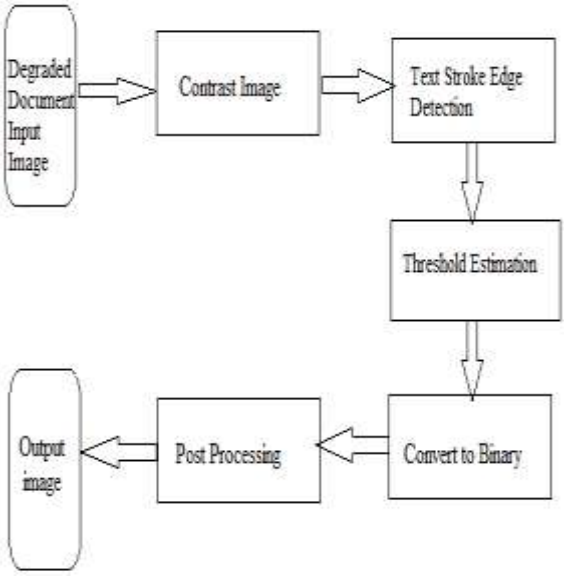

Fig 2- System Architecture

1. Contrast Image Construction

To adjust the level of contrast in the image is very necessary ,so we detect the exact text stoke edges. In this method we are keeping the image contrast at minimum or maximum level.Applying the contrast inversion on image that is we are reversing the color of image.Contrast inversion of above input image as shown in fig 3

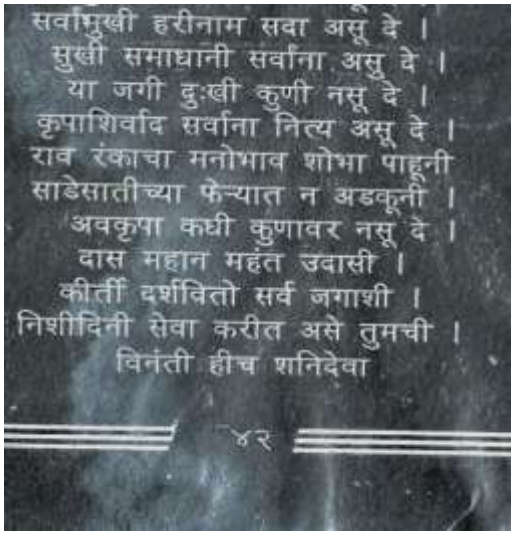

Fig 3-Contrast Image

2. Text Stroke Edge Detection

Todetect text stroke edges through gray scale conversion of contrast image.Before applying Otusu's global thresholding algorithm,we first convert the image into grayscale.the grayscale version of contrast image has properly veriation between the background and foreground pixel.The gray scale conversion helps to extract the text stroke edge pixel accurately.

Fig 4- Text Stroke Edge Detected image

3. Local Threshold Estimation

After detection of text stroke edge pixel 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. The edge detected image is then converted into binary format of 0's and 1's. 0 indicates that the image pixels are non-connected pixels and 1 indicate that image pixels are connected pixels and the represents the text strokes. The 0's are removed from the image because they are part of background image.

4. Post Processing

In proposed system, post processing technique plays important role,because it eliminate the non-stroke images from binary image.It also eliminate background pixel which is not related with resultant binarised image.It gives output as a clear image , which is in readable format.we are use following algorithm for post processing.

Input : The Input Document Image I , Initial Binary Result B and Corresponding Binary Text Stroke Edge Image Edg
Output**:** The Final Binary Result B*f*
1: Find out all the connect components of the stroke edge pixels in Edg.
2: Remove those pixels that do not connect with other pixels.
3: for Each remaining edge pixels (x,y): do
4: Get its neighbourhood pairs: (x − 1, y) and (x + 1, y); (x, y − 1) and (x, y + 1)
5: if the pixels in the same pairs belong to the same class (both text or background) then
6: Assign the pixel with lower intensity to foreground

 ISSN: 2321-8169
4104 - 4106

class (text), and the other to background class.
7: end if
8: end for
9: Remove single-pixel artifacts along the text stroke boundaries after the document thresholding.
10: Store the new binary result to B*f* .

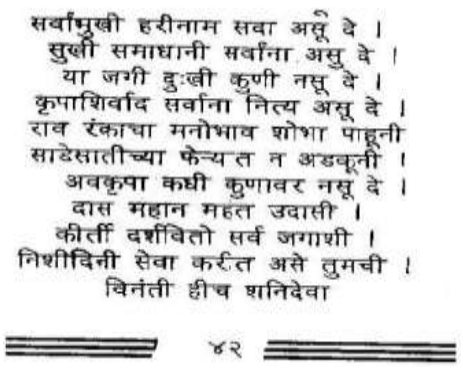

Fig 5 - Recovered Image by Proposed System

After post processing technique ,we calculate the evaluation parameter Mean Square Error(MSE),Signal to Noise Ratio (SNR) and Peak Signal to Noise Ratio(PSNR) of document image.

Mean Squared Error (MSE)	Signal to Noise Ration (SNR)	Peak Signal to Noise ration (PSNR)
1.946791430769731E13	14.120172978435516	18.476239065065172

IV. CONCLUSION

This paper present the different binarization Technique,Which helps to give clear output image. The proposed technique is simple ,robust and efficient.After contrast inversion ,we have used gray scale conversion which helps to detect the text stroke edges of degraded document images.The proposed method is more stable and easy to use for document images with different kinds of degradation. In proposed method, we calculate the evaluation parameter such as MSR,SNR and PSNR. . In future we are going to add character set into our system.This will allow system to recognize the character of each language. It will help to find the actual text strokes.

V. ACKNOWLEDGEMENT

We sincerely thanks to my Department Head, Guide and n all other staff members to give me the guidelines for this paper.

REFERENCES

[1] Bolan Su, Shijian Lu, and Chew Lim Tan, Senior Member, IEEE, "Robust Document Image Binarization Technique for Degraded Document Image,"IEEE Transactions On Image Processing, Vol. 22, No. 4, April 2013.
[2] G. Leedham, C. Yan, K. Takru, J. Hadi, N. Tan, and L. Mian, "Comparison of some thresholding algorithms for text/background segmentation in difficult document images," in Proc. Int. Conf. Document Anal. Recognit., vol. 13. 2003, pp. 859864.
[3] O. D. Trier and T. Taxt, "Evaluation of binarization methods for document images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 17, no. 3, pp. 312315, Mar. 1995.
[4] B. Gatos, I. Pratikakis, and S. Perantonis, "Adaptive degraded document image binarization," Pattern Recognit., vol. 39, no. 3, pp. 317327, 2006.
[5] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," Pattern Recognit., vol. 33, no. 2, pp. 225236, 2000.
[6] B. Su, S. Lu, and C. L. Tan, "Binarization of historical handwritten document images using local maximum and minimum filter," in Proc. Int. Workshop Document Anal. Syst., Jun. 2010, pp. 159166.
[7] J. Bernsen, "Dynamic thresholding of gray- level images" in Proc. Int. Conf. Pattern Recognit., Oct. 1986, pp. 12511255.
[8] M. van Herk, "A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels," Pattern Recognit. Lett., vol. 13, no. 7, pp. 517521, Jul. 1992.
[9] S. Kumar, R. Gupta, N. Khanna, S. Chaudhury, and S. D. Joshi, "Iterative multimodel subimage binarization for handwritten character segmentation," IEEE Trans. Image Process., vol , no. 9, pp. 12231230, Sep. 2004.
[10] Y. Chen and G. Leedham, "Decompose algorithm for thresholding degraded historical document images," IEEE Proc. Vis., Image Signal Process., vol. 152, no. 6, pp. 702714, Dec. 2005.