



A Survey on Document Image Binarization for Enhancing Degraded Document Image

Sangeetha.R¹, Rajkumar.T.C², Amali Asha.A³

Research scholar, Dept. of Computer Science, ST.Xavier's College, Tirunelveli, Tamilnadu, India¹

Associate Professor, Dept. of Computer Science, ST.Xavier's College, Tirunelveli, Tamilnadu, India²

Associate Professor, Dept. of Computer Science, Loyola College, Chennai, Tamilnadu, India³

ABSTRACT: Document image binarization is a prominent research area in computer science. Image binarization is a process of converting a gray scale image to binary image. Segmentation is a very challenging task to segregate a text from background images. Thresholding is the best way to use binarization of degraded document images. Thresholding is divide into two parts: local and global thresholding. The local thresholding divide the image into sub-images and then convert the binary image depending the threshold value. The global thresholding enumerate a single threshold for whole image. Thresholding is very fast calculating the accurate value to perform image segmentation. Thresholding is a pre-processing for the degraded image to enhance their quality. The main goal of this paper is to study and to evaluate the different image binarization techniques. In some method they propose the adaptive image contrast method it is the combination of local image contrast and local image gradient. Both the method are used to recover the image.

KEYWORDS: Document, Thresholding, Binarization

I. INTRODUCTION

Binarization is a popular research area in the field of image processing. Document image binarization is performed the preprocessing stage for segment the image. Binarization is a reform of gray image to binarized image. In many field they use the image such as in medical, document image, face recognition where it is a very challenging task to segment the text from background. This can be done in the process of binarization. In some place handwritten document are used, which can at times be faulty due to printing and manual error. Those type of document can recover through the binarization method. Papers use the preprocessing method and thresholding to segment the image. The goal of document image is to recover an image. A binary image processed better then the gray scale image. The basic approach of binarization method to degraded document images. Next they approach the thresholding to restoration of foreground from background of document image.

II. RELATED WORK

Method on adaptive image contrast

Bolan su (2013) [1] has proposed a document image binarization mainly focus on the issues by using adaptive image contrast. The contrast map is binarized and apply with canny's edge detection to identify the edge of the pixels. They use the local threshold to detect the text within the local tunnel. The main drawback of Beckley diary dataset is a bad quality document images through this technique, compared with other techniques.

Method on different document

Prachi K.More (2016) [2] has proposed the different methods to solve degraded document image. But none of the method display any good result. Every method has some limitation over recovering document image.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Method on canny edge map

Nikita mote (2016) [3] has proposed a degraded document input image to improve the quality of the output image. This method used adaptive image contrast is binarized and then apply the canny's edge map to identify the edge pixels. The document image is segmented by local threshold to detect the edge pixels. Filter is apply to reduce noise. The main drawback of this method is it works only few dataset.

Method on thresholding

Rekha chaudhari (2015) [4] has proposed the binarization technique using threshold only. They use the global and local threshold can segment the image. In this paper, they upgrade their old handwritten document and machine printed document. They proposed new binarization method using image segmentation through threshold segmentation. They overcome the drawback of canny edge map and also it generate good result. The disadvantage is the use of less parameter to recover the document image.

Method on OCR

D.D.Pukale (2016) [5] has proposed the OCR technique to retrieve the characters. They proposed otsu, local, manual thresholding for binarization. After binarization, if noise prevails the post processing method is used to remove the noise. The OCR technique implemented in some algorithm to reduce the human error. The algorithm are thinning, segmentation, scaling. These algorithm are time consumption that making the method is more effective.

Method on Different Document

J.Sauvola (2000) [6] has proposed different types of images which has been tested in binarization method. They use the benchmark algorithm shows the test result for each case qualitative and quantitatively. They use two types of threshold method for each pixel. A new method is text binarization is implemented for textual and line drawing areas. The disadvantage is it support only the textual component mostly.

Method on wiener filter

B.Gatos (2005) [7] has proposed the wiener filter to binarize the degraded image. They can process a binarization by a combination of wiener filter and sauvola method. To examine the pixel contrast to get the black and white image. The main disadvantage is does not improve the quality of image but the content of image.

Method on different techniques

Tarnjot Kaur Gill (2014) [8] has proposed the techniques to solve the degraded images. The method use various thresholding process to detect the degraded image but none of the method can solve the degraded images. They can survey many method to recover the problem. Each paper has its own benefits and limitations so no technique is best for every problem.

Method on improved binarization technique

Prachi K.more (2016) [9] has proposed the use of thresholding method that divide the text from degraded document and also they can apply the adaptive image contrast. It can implemented the canny edge map to get a clear document. They proposed the pre-processing method that involve the histogram equalization that helps to recover the quality and clarity of images. They use the local threshold to detect the degraded document image.

Method on morphology operator

Divya jyoti (2016) [10] has proposed a degraded document image to improve the quality of the output image through the binarization method and also use the filtering method to avoid the noise. A new method has been proposed a morphology operator for get a better result. They use two thresholding method to solve but it is not suitable for all images. The drawback of this method was unable to display the accurate output image.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Method on Thresholding

Ravinder Kumar(2015) [11] has proposed single thresholding algorithm. This method was not suitable for all images. So they use adaptive threshold algorithm. The main drawback of the method was foreground images suppress the background images.

Method on Niblack algorithm

Khuurran khurshid (2011) [12] has proposed the different algorithm to test the ancient document but the most of the algorithm cannot give the approximate solution. So they propose the Niblack algorithm in very detail. They apply all images and the better result are not shown. The disadvantage of the method is it can be low process.

I. SURVEY TABLE

S.No	Paper Name	Technique	Advantage	Disadvantage
1	Robust document image binarization technique for degraded document image.	Approach the adaptive image contrast technique.	To apply canny edge detection to identify the edge of the pixels.	Performance of some images still needed to be improved.
2	Binarization technique used for recovering degraded document images	A new method of combining the adaptive image contrast and canny edge map.	To improve the quality of the output image by using threshold and filtering.	It perform only few dataset.
3	Adaptive image binarization	A new approach of benchmarking algorithm to detect the images.	A text binarization is implemented to textual and drawing areas.	It can support only the textual component mostly.
4	Adaptive degraded document image binarization	The combination of wiener filter and sauvola method.	To get the exact content of the image.	This method cannot improve the quality of the image but it can change the content of the image

III. GAPS IN LITERATURE

Many techniques are proposed the document image binarization shows in literature survey. The existing research paper shown no techniques is work with perfect for every paper. Every paper has some drawback. The main limitations of the research work:

1. All the method use the same technique but all the method are displayed the same result no improvement to differentiate the result.
2. The size of image is a problem which can be resizing the image and work with it.

IV. PROPOSED METHODOLOGY

In proposed method, the RGB degraded images as an input image for noise removing. It converts the RGB image into gray scale. Gray scale image have many shades of gray with blur in the boundary. So avoid this problem to convert the gray image into binary image. Filtering method is used to modifying or enhance the degraded image and also used to reduce the unwanted elements. Morphology operator is applied to enhance the image. In this operator, a structuring element as input as compared as degraded image until the process to get a accurate value. Finally apply the



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 3, March 2017

thresholding value is applied to restrict the layer of the image to restoration of image is displayed. The various method to display the result are shown in the below table.

Method	F-measure	PSNR	DRD	MPM
Canny edge map	87.8	17.56	4.84	5.17
Beckley diary	78.54	13.15	12.92	16.71
Adaptive image contrast	90.05	-	-	-
Morphology operator	91.5	20.89	3.37	1.31

V. CONCLUSION

This survey work is comparing different document image binarization method to recover a degraded document image. The main goal of this paper is to classify the algorithms to degraded document image. In each technique there is some limitations and benefits are present. The above survey table can be explained the different methods ,advantages and disadvantages Document image binarization is the important application to all field. For example, medical, college, offices. In earlier days, they use the handwritten document or printed document. Now a days they use mobile phones to study and all purpose. If any image is damaged then detect the image through the above methodology. In future, propose a new algorithm based morphology operator to enhance the work easily and also enhance the pre-processing technique to improve the degraded document image.

REFERENCES

- [1] Su,Bolan,Shijian Lu, and chew Lim Tan. "Robust document image binarization technique for degraded document images."Image processing, IEEE Transactions on 22.4(2013): 1408-1417.
- [2] Prachi K.More ,D.D.Dighe. "A review on document image binarization technique for degraded document images" Image processing,IRJET(2016)
- [3] Nikita Mote, Shital avhad, Sonali jangale. "Binarization technique used for recovering degraded document images" IJIR Vol 2 Issue-2, (2016)
- [4] Rekha chaudhari, Dinesh patil. "Document image binarization using threshold segmentation" IJIRCCCE., vol 3., Issue 3 march 2015
- [5] D.D.Pukale ,aparna thorat,pooja bora,shruti barmecha,dipika suryawanshi. "Document image binarization for degraded document images along with OCR" IJIR Vol 2 Issue 6 2016
- [6] J.sauvola and M.pietikainen. "Adaptive image binarization," pattern recognition, vol 33,no2,00.255-236,2000
- [7] B.Gatos, i.pratikakis and s.perantonis, "Adaptive degraded document image binarization"pattern recognition, vol.39,no.3 2005
- [8] Tarnjot kaur gill, "Document image binarization techniques –A review" ,IJCA vol 98-no 12 (2014)
- [9] Prachi K.More,D.D.Dighe, "An improved hybrid binarization technique for degraded document digitization" IJESRT vol 5(8) 2016
- [10] Divya jyoti, Bodh Raj, arun Sharma, Kapil Kapoor, " Document image binarization technique for degraded document image by using morphology operators" IJARIT vol 2 issue 3 2016
- [11] Ravinderkumar, Mohinder Malhotra, "Document image Binarization" IJERSTE vol 4 issue 12 December 2015
- [12]Khurram Khurshid, ImranSiddiqi, Daudie Faure, Nicolevincent, "Comparision of niblack inspired binarization method for ancient document"2010.