A Review of Damaged Manuscripts Using Binarization Techniques.

Prajakta. S. Jadhav[1], Preeti. P. Dapkekar[2], Shubhangi. M. Jadhav[3], Kartik. K. Kondalkar [4],Shabnam S. Shaikh[5]

[1]Student (UG),Department Of Computer Engineering,

AISSMS COE Pune, Maharashtra, India.

praj11593@gmail.com

[2] Student (UG) ,Department Of Computer Engineering,

AISSMS COE Pune, Maharashtra, India.

dapkekarpreeti@gmai.com

[3]Student (UG) ,Department Of Computer Engineering,

AISSMS COE Pune, Maharashtra, India.

shubhangijadhav129@gmail.com

[4]Student (UG) ,Department Of Computer Engineering,

AISSMS COE Pune, Maharashtra, India.

kartikkon@gmail.com

[5]Assistant Professor, Department of Computer Engineering,

AISSMS COE, Pune, Maharashtra, India.

shabnamfsayyad@gmail.com

Abstract: *Image binarization is the procedure of separating pixel values into two parts, black as foreground and white as background. In binarization document image is converted into binary image using thresholding techniques. Many binarization algorithms have been proposed for different types of degraded document images. The main objective of image enhancement is to modify attributes of an image to make it more suitable for a given task . To remove the noise and improve the quality of the document binarization techniques are used. Thresholding is one of such binarzation technique which is used for this purpose. Thresholding is further divided into the global and local thresholding techniques. In the document with uniform contrast delivery of background and foreground, global thresholding has been found to be best technique. Local thresholding , is an approach for the situations in which single value thresholding does not yield proper result, to overcome this a hybrid approach is introduced which is a combination of local with Otsu's thresholding.*

Keywords: Binarization, Thresholding, Degraded Images, Pixels.

1. Introduction

Documents are important source of information but they suffer dreadful conditions such as degradation problems, especially in the case of historical and government documents. There are libraries in the world, which has collections of historical and ancient documents that are of great scientific and cultural importance. Degradation problems are quiet common in these documents. Few factors that check the legibility of the documents are strains, big variations, uneven illumination, presence of smear, seepage of ink, etc. Before libraries expose them to public view it is important to remove noise from historical document images and improve their quality. Noise is considered anything that is not valid with the textual information. The main objective of image enhancement is to modify attributes of an image to make it more suitable for a given task and observer. In this process, one or more attributes of the image are changed.

The selection of attributes and the way they are modified are specific to a given task. The enhancement system produces a document image that can be viewed in different ways using various methods for 3 different processes. A first method is converting the given RGB image to a grayscale image. Gray scale image is provided as input. Second process is of Thresholding. Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. Third process is Binarization. Objective of Binarization is to mark pixels that belong to true foreground regions with a single intensity and background regions with different intensities. Various types of thresholding methods are applied to binarize an image depending upon various binarization techniques.

The paper is organized as follows. Section I gives an introduction. Section II gives a description about the various binarization techniques used. This is followed by Section III which describes the literature study. Section IV concludes the paper.

2. Literature Survey

PreetiKale ,Dr.S.T.Gandhe, Prof.G.M.Phade and Prof.Pravin A.Dhulekar [1], "Enhancement of old images and documents by Digital Image Processing Techniques" ,2015 International Conference on Communication, Information and Computing Technology (ICCICT),978-1-4799-5522-0/15/2015 IEEE . Authors proposed that, Global and local thresholding techniques are combined and used to remove the variation of noise in the image   global thresholding is applied to the whole image firstly. If still the image has background noise the technique is reapplied to each area separately. In order to evaluate this approach, a degree is represented in which the legibility of the image has been improved.

Su. Bolan , Shijian Lu, and Chew Lim Tan [2], "Robust Document Image Binarization Technique for Degraded Document Images ", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 4, APRIL 2013.Authors proposed a local thresholding technique which include use of local image contrast that is based on local maximum and local minimum value.

Pratikakis Basilis Gatos and Konstantinos Ntirogiannis [3] ,"Document analysis and recognition (ICDAR 2013)" 12th international conference on IEEE .Authors intended that, in degraded documents, where extensive background noise or difference in contrast and brightness exists i.e. there exists many pixels that cannot be effortlessly categorized as foreground or background. In such cases, local thresholding has significant over available techniques. Objective is to evaluate the different image binarization techniques   to find the gaps in existing techniques.

Chinni Bhargavi S,B Priyanka,and Mamatha[5],"A Comparative Study of Binarization Techniques for Enhancement of Degraded Documents ",International Journal of Computer Applications (0975 8887) Volume 119 No.11, June 2015.Authors presented some methods suitable and effective for enhancement of degraded documents by applying effective techniques like Local thresholding, Global thresholding, Hybrid Binarization, Iterative Global Thresholding, which shows satisfactory results for Binarization after which remaining noise is removed by applying, filters to the image using methods like Smoothing and Sharpening.

Sehad, Abdenour.et al [6],"Ancient degraded document image binarization based on texture features",(ISPA 2013) 8th International Symposium on IEEE2013.Author intended that ,Few factors that impede the legibility of the documents are strains, big variations, uneven illumination, presence of smear, seepage of ink, etc. Before libraries expose them to public view it is important to remove noise from historical document images and improve their quality. Objective of Binarization is to mark pixels that belong to true foreground regions with a single intensity and background regions with different intensities.

Rabeux,Vincent. et al[7],"Quality evaluation of ancient digitized documents for binarization prediction",ICDAR 2013.Author intended that, objective of image enhancement is to modify attributes of an image to make it more suitable for a given task and a observer. In this process, one or more attributes of the image are changed. The selection of attributes and the way they are modified are specific to a given task. Algorithms belong to this category tries to combine the advantages of two categories of algorithms and removes their limitation to get accurate binarize image.

Jin-won Jangl, Sewon Leel ,Hee-Jung Hwangl and Kwang-Ryul Baekl [8], "Global thresholding algorithm based on boundary selection", 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013).To compensate the weak point of global thresholding methods, local thresholding algorithms such as Niblack's method have been introduced by authors. They also proposed that Otsu's method is an efficient solution for segmenting images into two classes when the shape of the image histogram is bi modal.

Sitti Rachmawati Yahya, S. N. H. Sheikh Abdullah , K. Omar, M. S. Zakaria and C. Y. Liong[9], "Review On Image Enhancement Methods Of Old Manuscript With The Damaged Background", 2009 International Conference on Electrical Engineering and Informatics 5-7 August 2009,2009 IEEE.Authors developed a procedure to separate the letters from the background in that manuscript. Their research conducted by has attempted streak elimination in both sides of document by document image by using edge detection method with double threshold value in foreground image. They introduced image enhancement methods, difficulties in Malay manuscript image enhancement, benefit and limitations in image enhancement.

3. Binarization Techniques.

A. **Processing of RGB image into Grayscale:**

Traverse through entire input image array. Read individual pixel color value (24-bit).Split the color value into individual R, G and B 8-bit values. Calculate the grayscale component (8-bit) for given R, G and B pixels .There are various ways to convert color values to grayscale. Any one can be used depending on the users needs. Calculate the grayscale component (8-bit) for given R, G and B pixels using a conversion formula.

$gs = (r +g +b) / 3;....(1)$

Here average of all three colors are calculated and saved in output image. Equation(1) can also be written as

$gs = r * 0.33 + g * 0.33 + b * 0.33;.....(2)$

 Compose a 24-bit pixel value from 8-bit grayscale value. In order to recompose a 24 bit value we need to update the values of r, g and b first and then store them back to form a 24-bit value. To do this we left shift the red component by 16 so that it comes to its exact position in 24-bits.We left shift green by 8 to do the same and leave blue as is since its original position is LSB. We then OR these together to for the actual 24-bit output pixel.. Store the new value at same location in output image.

B. **The global thresholding technique** computes an optimal threshold for the entire image; these techniques need few

computations and can work well in simple cases. But fails in complex backgrounds, such as non-uniform color and poor illuminated backgrounds. These methods are not suitable for degraded images, because foreground and background are not separated properly.

C. **The local binarization techniques** set different thresholds for different target pixels depending on their neighbourhood/local information. Generally, these techniques are sensitive to background noises due to large variance in case of a poor illuminated document or bleed-through degradation.

D. **Otsu's Method**: Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold, i.e. the pixels that either fall in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum

E. **Hybrid binarization approach** combines global and local thresholding. A first step consists in carrying out a global thresholding to classify a part of the background of the document image and keep only the part containing the foreground (graphics or text in our case). A second step aims to refine the image obtained by the previous step in order to obtain a sharper result by applying an adaptive thresholding technique.

4. Conclusion

The paper focuses on different types of thresholding algorithms in binarization techniques. After that, global and local thresholding techniques are introduced with analysis of pros and cons in terms of robustness. This paper has presented a hybrid thresholding technique which combines otsus and local thresholding methods which is a hybrid approach. A Grayscale image is taken as an input and local thresholding is carried upon using Otsus method at sub parts of an image using Threshold Binary for better results. From the existing algorithms, it is evident that this method can produce promising results compared to other methods.

References

[1] Preeti Kale, Dr. S. T. Gandhe , Prof. G. M. Phade and Prof. Pravin A. Dhulekar , Enhancement of old images and documents by Digital Image Processing Techniques, International Conference on Communication, Information and Computing Technology (ICCICT), 978-1-4799-5522-0,PP 1-5, 2015 IEEE.

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

[3] Pratikakis Basilis Gatos and KonstantinosNtirogiannis ,"Document analysis and recognition (ICDAR 2013)" 12th international conference on IEEE .

[4] Stathis Pavlos, Brgina Kavannieratou and Nikos Papamarkos ,"An evolution survey of binarization algorithms on historical documents",19th international conference on IEEE 2008.

[5] Chinni Bhargavi S, B Priyanka, Mamatha H R, A Comparative Study of Binarization Techniques for Enhancement of Degraded Documents, International Journal of Computer Applications,PP 39-41,2015.

[6] Sehad, Abdenour.et al,"Ancient degraded document image binarization based on texture features",(ISPA 2013) 8th International Symposium on IEEE 2013.

[7] Rabeux,Vincent. et al,"Quality evaluation of ancient digitized documents for binarization prediction",ICDAR 2013.

[8] Jin-won Jangl, Sewon Leel ,Hee-Jung Hwangl and Kwang-Ryul Baekl, Global thresholding algorithm based on boundary selection, 13th International Conference on Control, Automation and Systems,PP 704-706,2013.

[9] Sitti Rachmawati Yahya, S. N. H. Sheikh Abdullah, K. Omar, M. S. Zakaria, And C. Y. Liong, Review on Image Enhancement Methods of Old Manuscript with Damaged Background, International Journal on Electrical Engineering and Informatics,PP 1-14,2010.

[10] Aroop Mukherjee and Soumen Kanrar, Enhancement of Image Resolution by Binarization, International Journal of Computer Applications, 0975 888,PP
15-19,2010.

Author Profile

Prajakta S Jadhav is pursuing Bachelor's degree in Computer Engineering from Savitribai Phule Pune University from A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India.

Preeti P Dapkekar is pursuing Bachelor's degree in Computer Engineering from Savitribai Phule Pune University from A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India.

Shubhangi M Jadhav is pursuing Bachelor's degree in Computer Engineering from Savitribai Phule Pune University from A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India.

Kartik K Kondalkar is pursuing Bachelor's degree in Computer Engineering from Savitribai Phule Pune University from A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India.

Shabnam S. Shaikh is an Assistant Professor, Department of Computer Engineering, Savitribai Phule Pune University in AISSMS College Of Engineering, Pune, Maharashtra, India.