

Hybrid Binarization for Restoration of Degraded Historical Document

Rohini Umbare¹, M.D Mali², Sunita Sagat³

P.G. Student, Department of E&TC Engineering, N.B. Navale Sinhgad College of Engineering, Solapur,
Maharashtra, India¹

Assistant Professor, Department of E&TC Engineering, N.B. Navale Sinhgad College of Engineering, Solapur,
Maharashtra, India²

Assistant Professor, Department of E&TC Engineering, N.B. Navale Sinhgad College of Engineering, Solapur,
Maharashtra, India²

ABSTRACT: Historical Documents are the valuable source of information but they suffer degradation problems, such as ink seepage, uneven illumination, big variations, strains etc. for that binarization technique play very important role to remove the noise and improve the quality of documents. This paper focuses on degraded historical documents, which are in the form of machine printed and handwritten. The proposed hybrid binarization which is combination of local and global thresholding method consist of five stages such as noise elimination, foreground layer extraction, degradation of background layer estimation, thresholding, and vicinity analysis. Initially, a technique named global thresholding is applied to the whole image and local thresholding is applied to sub image. Therefore, a better adaptability is achieved for the algorithm where various kinds of noise exist in different areas of same image. Advantage of applying global thresholding, is that it avoids the computational time and cost of applying a local thresholding in the entire image. Hence it is indicated that this proposed method is effective in removing background noise and improving the quality of degraded images.

KEYWORDS: Image Processing, Image Enhancement, Binarization Method, Thresholding Method.

I. INTRODUCTION

Degraded historical documents are preserved in academic libraries, institutions, historical museums for their extensive usage but The historical documents suffers degradation problem which is caused by a combination of factors such as temperature level, environmental conditions and low quality paper. The Electronic scanning is the approach for handling such documents to preserve the culture of heritage. But the resulting images have low contrast and corrupted by various artefacts so they are often difficult to read. The restoration and enhancement of degraded historical document images are considered a transformation process which concentrated to restore its original representation of document images As well as restoration and enhancement, which are desired to improve the quality of subsequent segmentation.

Degraded historical document image are considered a combination of multilayer information that are foreground layer, background layer and degraded layer. The image processing technique is applied to restore and enhance the quality of degraded document images. In general, there are three steps for restoration of historical document images: pre-processing, binarization and post-processing [17].in this pre-processing step eliminate the noise on the image, binarizaion[10] step to transform gray –level image into binary image and post-processing step to enhance the quality of binary image. According to image processing techniques, the binarization methods play an important role in the document image processing. The binarization method classified as global and local thresholding. The binarization methods can be classified as global and local thresholding. A global thresholding, such as Otsu's, Kittler's and Kapur's methods and it provide a single threshold to classify an image into foreground and background, a

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

local thresholding calculate an adaptive threshold value in local block. The block size must be small to indicate local details and large to remove noise. Bernsen's, Niblack's and Sauvola's methods [7] are well-known local thresholding methods. The Niblack's method calculates local threshold by using the mean and standard deviation value of gray-level image in a local block.

Let's $g(x, y)$ is a gray level image, and $\mu(g(x, y))$ and $\sigma(g(x, y))$ is the average and standard deviation of gray level values of $g(x, y)$. A variable 'k' is used to adjust a ratio of foreground pixels particularly for edge of character. The threshold of Niblack's method defined as $TNibg(x, y)$ is calculated by the following formula:

$$TNibg(x, y) = \mu(g(x, y)) + (k \cdot \sigma(g(x, y))) \text{ ----- (1)}$$

The Niblack's method generates poor quality result. Because of this, the Sauvola's method as an improved Niblack's method is proposed to solve this problem. A variable "r" is added to Niblack's formula to change the behaviour from static to dynamic range standard deviation. The threshold of Sauvola's method defined as $TSaug(x, y)$ is calculated by using the following formula:

$$TSaug(x, y) = \mu(g(x, y)) \cdot \left((1 - k) \cdot \left(1 - \frac{\sigma(g(x, y))}{r} \right) \right) \text{ ----- (2)}$$

Niblack's method is applied to estimate foreground regions and then background regions are estimated sequentially. The background regions estimation is guided by the value of the initial binary image. After the binarization process by using local thresholding, post-processing is performed to reduce noise and enhance the quality of text regions.

II. LETERATURE SURVEY

Zhixin Shi and Srirangaraj Setlur and Venu Govindaraju [9] proposed a Transform based strategy for improving computerized pictures of palm leaf structures. The technique utilizes an alertly chose pivoting background color in a linear transform to improve the legibility of the forefront content. At that point a blend of two other image processing algorithms. Histogram normalization and background normalization are used to the transformed image.

Nikolaos NTOGAS and Dimitrios VENTZAS [10] propose a straightforward and strong binarization procedure. This work concentrates on text image enhancement and restoration, denoising and binarization using Mat lab. Binarization is obtained by global and local thresholding and at last refinement is applied to further clarify text and foreground compared to background. Ketki R. Ingole and V K. Shandilya [11] propose a technique for Historical Manuscripts document enhancement that improves the quality of historical Arabic manuscripts which shows uneven background and low contrast due to manufacturing and the effect of getting old and degradation then a background normalization algorithm applied to smoothen out the background and produces more legible images to the eye. B. Gangamma [15] proposed a Simple and efficient method for degraded historical document enhancement. This method enhances images using adaptive histogram equalization for setting contrast followed by gray scale morphological operations to eliminate noise, eliminate background from images with pure foreground contents. Blurred and skewed images are not in the scope of this technique. The binarization process separates foreground text from background. If the histogram of the image is bimodal or sparsely distributed, then binarization process separates text from background. Dimitrios Ventzas propose a work on denoising and binarization. That was introduce an innovative sequential procedure for digital image acquisition of historical documents including image preparation, image type classification according to their condition and their spatial structure, global and local features or both, including document image data mining. Image processing pixel alterations, allow one-pass iterations only by near neighborhood of alteration reprocessing algorithms. Gangamma and Srikanta Murthy [13] proposed a combination of spatial domain methods along with set theory operations to enhance the historical manuscript images. Bilateral filter is an efficient in eliminating the noise without smoothing the edges. Mathematical morphology which is based on set theory approach uses simple operations which are computationally less complex. This eliminates noise, rough background and improves the contrast of the script image. The restored images will have understandable consistent background and foreground with enhanced character

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

emergence. The enhanced document image can be used further to segment the document into lines, words and character for recognition purpose. And result of this planned technique is compared with Mean and Gaussian filter, and proved to be better than these techniques. D.N. Satange and Swati S. present [16] a recursive technique which includes iterated steps that makes it more stretchy regarding the needs of the user for enhancing and cleaning of historical manuscript documents. In this paper five filtering algorithms were applied on Salt & Pepper noise which was developed in handwritten Devnagari documents during image capturing and transmission. Rupinder K and Jaspreet K proposed a technique which works for both sided documents at the same time and give best results than existing techniques to remove the show-through from historical manuscript documents. In this by text segmentation through binarization the foreground and background should be separated. Using binarization a digital image converts into 0 (White) and 1 (Black) that is background as white and text as black, so by this text became more clear and read able, also it require less memory for storage.

III. PROPOSED METHODOLOGY AND DISCUSSION

This section presents the description of the proposed method based on a hybrid binarization method for restoration of degraded historical document images. The overview of the proposed method is illustrated in Fig. 1 the proposed method consists of five stages.

1. The noise elimination stage aims to eliminate noise areas by using a Wiener filter. The Wiener filter is a proper method and is proved an efficient technique for degraded document image filtering. The original graylevel image will be separated into 5×5 local blocks around corresponding pixel (x, y) . Let μ and σ be mean and variance in a local block, $\text{Avg}(\sigma)$ be an average variance of the original image. The gray-level value of the original and filtered image of pixel (x, y) are defined as $G_o(x, y)$ and $G_w(x, y)$ respectively. The $G_o(x, y)$ is transformed to $G_w(x, y)$
2. Binarization (local) stage aims to extracts the foreground pixels from the binary images of three well known binarization methods. Such as Niblack's method, Sauvola's [7] and laplacian gauss's method. The filtered image $G_w(x, y)$ is transformed to three binary images $B_1(x, y)$, $B_2(x, y)$ and $B_3(x, y)$. The extracted foreground layer in form of binary image defined as $B_f(x, y)$.
3. The degradation of the background layer estimation stage is based on the cluster analysis method, estimates the degradation of the background layer by replacing the foreground area with the estimated background which is the average value of cluster pixel. The gray value of pixel (x, y) of the degradation of the background layer defined as $G_{db}(x, y)$.
4. The thresholding stage transforms gray-level image into a binary image by calculating the threshold value in accordance with the gray value of the estimated degradation of the background layer, which is derived from otsu's method is combined with logistic sigmoid function. Based on the threshold $T_{ada}(x, y)$, the binary image defined as $B_{ada}(x, y)$.
5. The vicinity analysis stage enhances the quality of the binary image by analysing and categorizing the pixels of binary image into the correct group.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

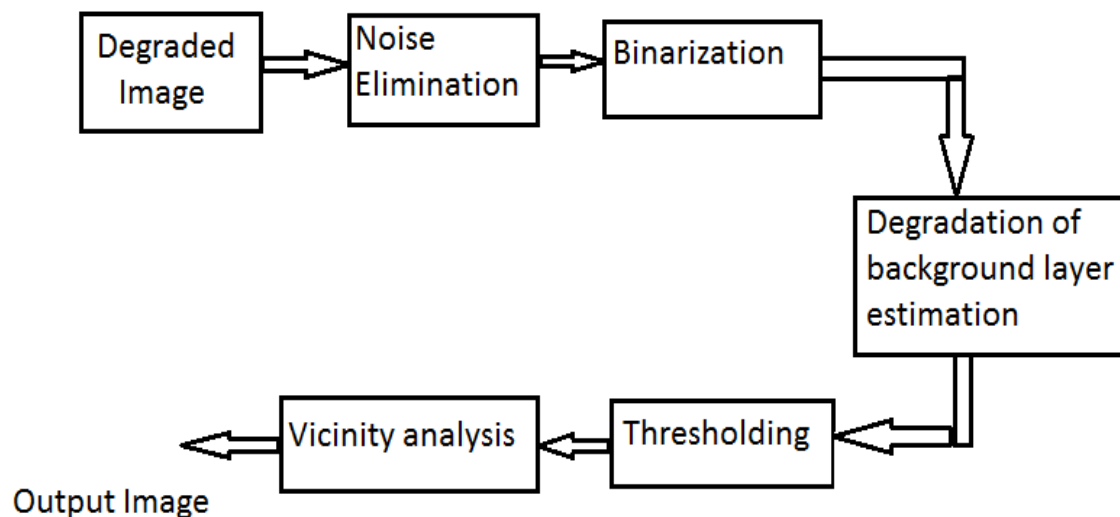


Fig 1- The proposed method.

IV. EXPERIMENTAL RESULT

In order to investigate and demonstrate the advantage of the proposed method, the experiments are on 110 degraded historical document images. The experimental results are evaluated by using following parameters such as Mean Square Error (MSE), Power to Signal Noise Ratio (PSNR), Average Difference (AD), and Structural Content (SC). The output of proposed method is compared with the Niblack's and sauvola's method.

S.No	Parameters	Description
1.	Mean Square Error (MSE)	$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (x(i,j) - y(i,j))^2$
2.	Power to Signal Noise Ratio(PSNR)	$PSNR = 10 \log_{10} \frac{(2^n-1)}{\sqrt{MSE}}$
3.	Average Difference(AD)	$AD = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (x(i,j) - y(i,j))$
4.	Structural Content(SC)	$SC = \frac{\sum_{i=1}^M \sum_{j=1}^N (y(i,j))^2}{\sum_{i=1}^M \sum_{j=1}^N (x(i,j))^2}$

Table1: Existing Measure of Quality Metrics

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

HW	Methods	MSE	PSNR	AD	SC
IMAGE1	Niblack	461.21	31.49	211.26	1.38
	Sauvola	463.82	31.46	211.84	1.35
	Proposed	25.07	44.13	-16.91	1.24
IMAGE2	Niblack	430.27	31.79	202.17	1.52
	Sauvola	432.5	31.77	202.66	3.85
	Proposed	31.4	43.16	-13.03	1.25
IMAGE3	Niblack	148.39	36.41	115.89	4.12
	Sauvola	149.84	36.37	116.48	2.5
	Proposed	142.02	36.6	-97.26	3.6
IMAGE4	Niblack	283.63	33.6	165.09	2.09
	Sauvola	285.57	33.57	165.64	1.77
	Proposed	71.57	39.58	-59.82	1.98

Table2: Three Method Performance Comparisons.

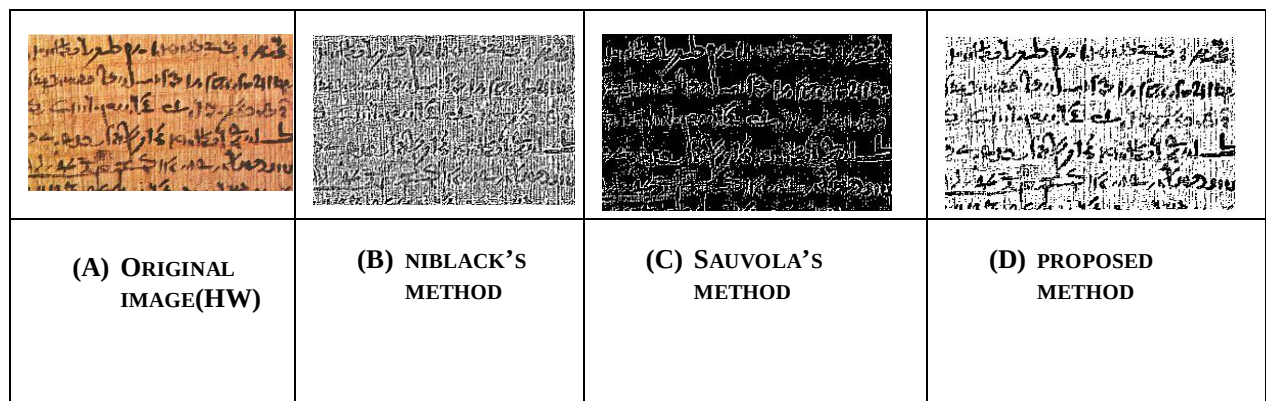


FIGURE 1: EXPERIMENTAL RESULT OF THREE METHODS.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

V. CONCLUSION

In this research, the new binarization method based on an adaptive multilayer-information for the restoration of degraded historical document images is proposed. The experiments are implemented by using MATLAB. The experimental results of the proposed method with 110 document images perform and analysed by evaluating PSNR, MSE, AD and SC. Moreover, the proposed method demonstrates superior performance against two well-known adaptive binarization methods on various degraded historical handwritten and machine printed document images. Furthermore, the proposed method can be applied with any degraded document images which have the same characteristics as the test set. But the parameters and techniques used in this method must be adjusted to be suitable for those document images.

REFERENCES

- 1) N.Maheshwari,P.singh ,A.Maloo "A Review of Digital Image Enhancement method of degraded Indian Ancient Manuscripts" International Journal for scientific research & development vol.3,Issue 03,2015.
- 2) R.Hedjam, M.cheriet "Novel data representation for text extraction from multispectral historical document Images" International conferences on Document analysis and Recognition 2011.
- 3) P.Kale , S.T Gandhe " Enhancement of old images by hybrid binarization Method" International Journal of advanced Research in computer Engineering & Technology vol 3 Issue 9, September 2014.
- 4) NTOGAS, NIKOLASO-"A Binarization algorithm for historical manuscripts" 12th WSEAS International conference on communication Heraklion, Greece, July 2008.
- 5) P.Jadhav, S.Shaikh "A Review Of Damaged Manuscripts Using Binarization Technique" International Journal of Engineering & computer Science.vol 5, Issue 2016.
- 6) A.Mukherjee, K. Soumen "Enhancement of Image Resolution By Binarization" International Journal of computer Application.vol-10, Issue 2010.
- 7) T.Romen Singh, O.Imoch Singh "A New Local Adaptive Thresholding Technique in Binarization"IJCSI vol-8, Issue 2011.
- 8) Niti Kamboj, V.Kumar "Degraded Document Image Enhancement Using Global Thresholding" International Journal for Technological Research in Engineering vol.1, Issue 2014.
- 9) Zhixin Shi, Srirangaraj Setlur and Venu Govindaraju, "Digital Enhancement of Palm Leaf Manuscript Images using Normalization Techniques", Dec. 2006.
- 10) Nikolaos, Ventzas, Dimitrios "a binarization algorithm for historical manuscripts"12th WSEAS International Conference on communication, Heraklion, Greece, July 23-25, 2008 Feb.
- 11) Ms. Ketki R. Ingole, Prof. V K. Shandilya, "Image Restoration of Historical Manuscripts," Aug. 2010.
- 12) K. Sitti Rachmawati Yahya, S. N. H. Sheikh Abdullah, K. Omar, M. S. Zakaria, "Review on Image Enhancement Methods of Old Manuscript with Damaged Background," April 2010.
- 13) B Gangamma, Srikanta Murthy K "Enhancement of Degraded Historical Kannada Documents", Sept. 2011.
- 14) Dimitrios Ventzas, Nikolaos Ntogas and Maria-Malamo Ventza "Digital Restoration by Denoising and Binarization of Historical Manuscripts Images", July 2012.
- 15) B. Gangamma, Srikanta Murthy K, Arun Vikas Singh "Restoration of Degraded Historical Document Image", May 2012.
- 16) Prof. D.N. Satange1, Ms. Swati S. Bobde2, Ms. Snehal D. Chikate "Historical Document Preservation using Image Processing Technique", IJCSMC, Vol. 2, Issue. 4, April 2013, pg.247 – 255.
- 17) Hugvainn Zarien, Senior Member, Jinendra Pallavi, Research Fellow Restoration of Degraded Historical Document Image: An Adaptive Multilayer-Information Binarization Technique", IEEE JOURNAL ON SELECTED AREAS IN IMAGE PROCESSING, VOL. NO. 67, JANUARY 2015.