RESEARCH ARTICLE OPEN ACCESS

De-Noising of Historical Document Images Using Ni-Black Thresholding

Geetika Gupta [1], Rupinder Kaur [2]
M.Tech Student [1], Assistant Professor [2]
Doaba Institute of Engineering and Technology, Kharar
Arun Bansal [3]
MD [3] AB Tech Labs
Punjab – India

ABSTRACT

The historical documents are of great importance. They present the Nation's Heritage and tradition. But with time, these documents start ageing. They get encountered with various noises. This leads to difficulty in reading those documents. Such a phenomenon is called as Degradation of document images. It is very much necessary to remove those noises, so that the historical documents can be preserved in a better way and condition. Various algorithms and techniques have been proposed for removing the noise from the degraded documents such as Canny's edge detector, Otsu's Global thresholding, MAP estimator, Markov Random Field, Adaptive Binarization, Weiner Filtering, Adaptive Bilateral Filtering, etc.. The proposed algorithm, Ni-Black thresholding proves better than the other algorithms. It is a Local Thresholding technique and removes the noise from the degraded document image far more than the algorithms used by various researchers. To further improve the output of the Ni-Black algorithm, filter is used thus giving a noise free as well as background eliminated image.

Keywords:- Historical document, Degraded Document Image, Binarization, Local thresholding, Ni-Black Thresholding.

I. INTRODUCTION

A. Degraded Document Images

The historical documents dated hundreds of years back suffer from ageing due to which they cannot be read properly [1]. The ageing is caused by the addition of noise through various sources. Some documents are degraded due to ink-bleed, whereas other suffer from backside reflection. Some documents get faded over time and some documents have intensity variation in foreground and background. In some documents the text get blurred. [7]

Some of the degraded document images taken from DIBCO dataset have been displayed below.

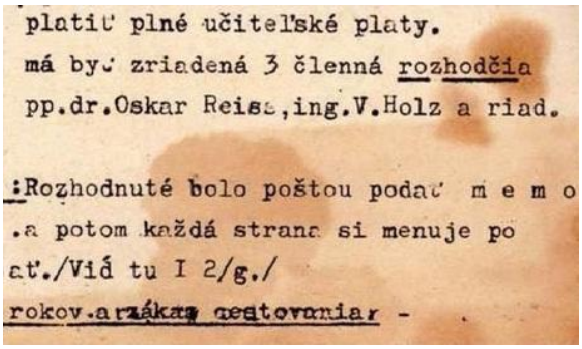


Fig (1): Image 1

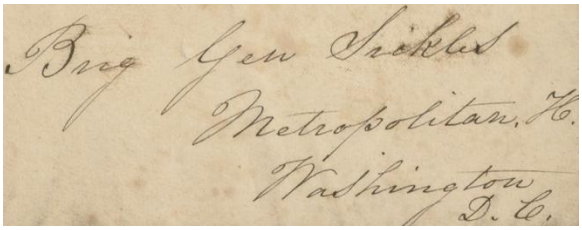


Fig (2): Image 2

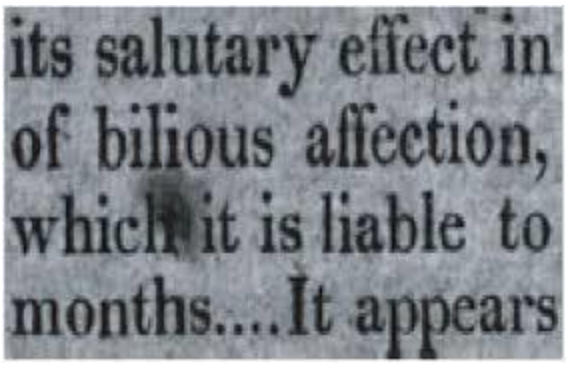


Fig (3): Image 3

B. Degradation Model

In a simplest image degradation model the degradation function is modeled as a low pass filter which resulted in a blurry effect. Figure 4 shows the block diagram of image degradation and restoration process. Fundamentally the image restoration process involves in reversing the distortion effects.[2]

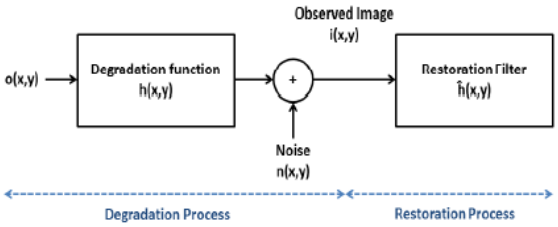


Fig 4

II. NI-BLACK ALGORITHM

The aim of the work is to recover the noise free image from the degraded document image. To achieve the aim, a local thresholding technique, Ni-Black algorithm have been proposed [3].

Thresholding is a type of image segmentation. It converts a gray-level image into binary image by replacing the pixels in the image having intensity less than a threshold value to zero (black) and the pixels having intensity greater than the threshold value to one(white) [4].

Thresholding is of two types: Global Thresholding and Local Thresholding. Global thresholding selects one threshold value for the entire image. [5] Whereas, Local thresholding selects different threshold values for different parts of the image and thus is more advantageous than the Global thresholding. [8].

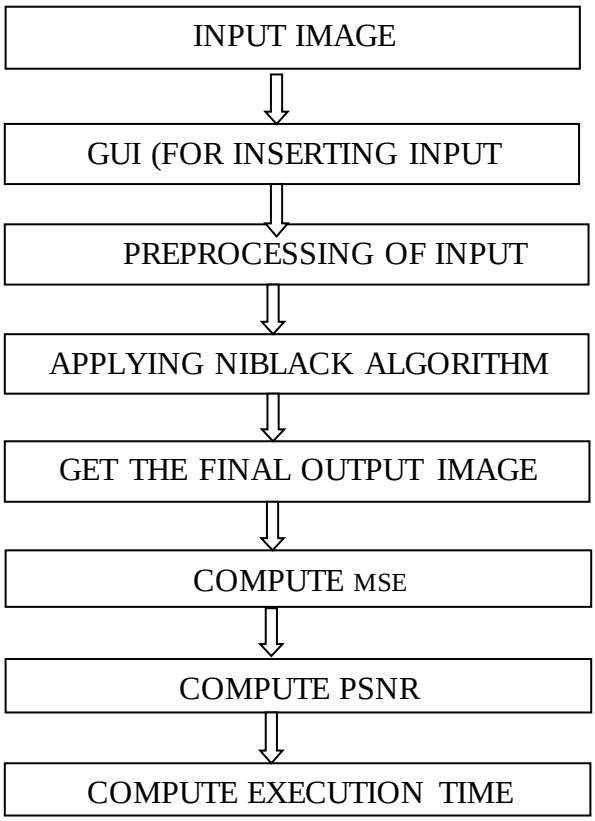
Ni-black thresholding works on the mean and standard deviation of the degraded input image. [12]

$T(i, j) = m(i, j) + k \cdot s(i, j)$

where 'm' is the mean of the number of pixels in that window is any constant that can be different for different type of documents and 's' is the standard deviation.[6]

The output image from Ni-Black thresholding has very less noise as compared to the original input image. But Ni-Black cannot remove the background noise. This is further achieved by applying a filter. The final output is noise free and eliminated background.

III. FLOW CHART



IV. EVALUATION MEASURES

The parameters being evaluated by the proposed algorithm are Mean Square Error, PSNR and Execution time of the code. [9]

1. Calculate Mean Square Error- f (i,j) is pixel value of output image, F(i,j) is pixel value of input image. Given by Formula:

MSE= ((no_pixels_in_output_image - no_pixels_in_input_image).^2)./((Size_Of_Image).^2)

2. PSNR (Peak Signal to Noise Ratio)- is used to measure the quality of restored image compared to the original image. Larger is the value, better will be the quality of image. It is calculated using equation as follow:

PSNR = 20 log10(255 / MSE)

The quality of the image is higher if the PSNR value of the image is high. Since PSNR is inversely proportional to MSE value of the image, the higher the PSNR value is, the lower the MSE value will be. Therefore the better the image quality is the lower the MSE value will be.

3. Time calculation- To use MATLAB command CLOCK to calculate time for our code to be executed, CLOCK is inbuilt command to show the real time, we use this command twice to calculate time consuming parameter.

V. RESULTS

Seven Degraded Document images have been taken from DIBCO Dataset. Ground Tooth images have been taken for experiments. The intermediate steps of the algorithm output are highlighted. The experimental results show that the proposed algorithm is more efficient than other de-noising algorithms.

Image 1:

Fig 5 shows the original input degraded image.

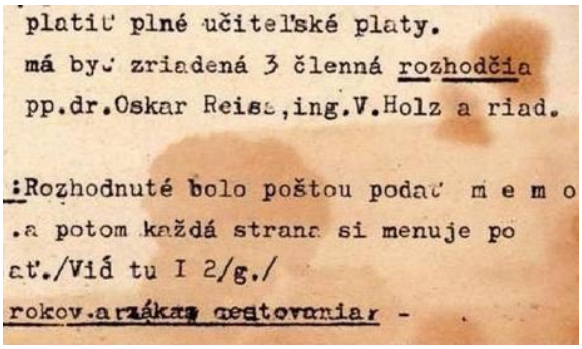


Fig 5: Original Image 1

Fig 6 shows the output of the Ni-Black algorithm after applying it on the original image.

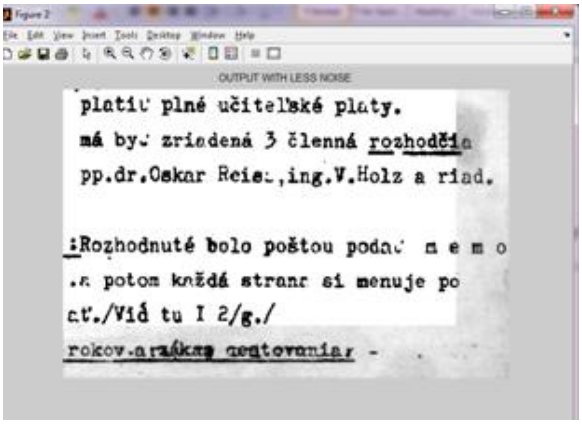


Fig 6: Restored Image of fig 5 Using Ni-Black Algorithm

Fig 7 shows the improved output of the Ni-Black algorithm after applying the filter on Ni-Black's output.

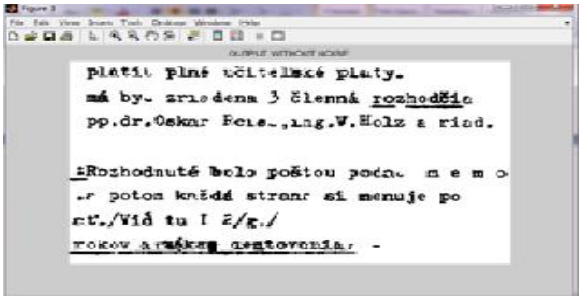


Fig 7: Output of Improved Ni-Black Algorithm of fig 6

Image 2:

Fig 8 shows the original input degraded image.

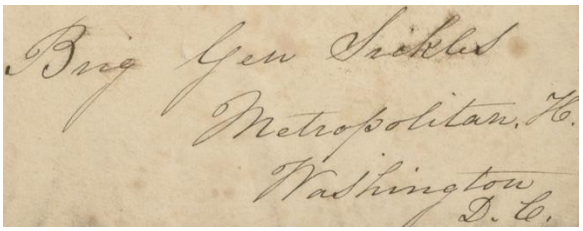


Fig 8: Original Image 2

Fig 9 shows the output of the Ni-Black algorithm after applying it on the original image.

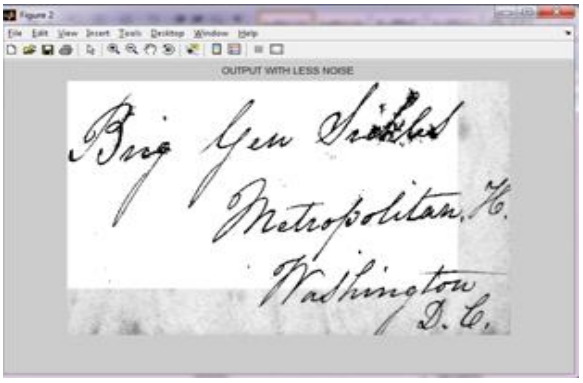


Fig 9: Restored Image of fig 8 Using Ni-Black Algorithm

Fig 10 shows the improved output of the Ni-Black algorithm after applying the filter on Ni-Black's output.

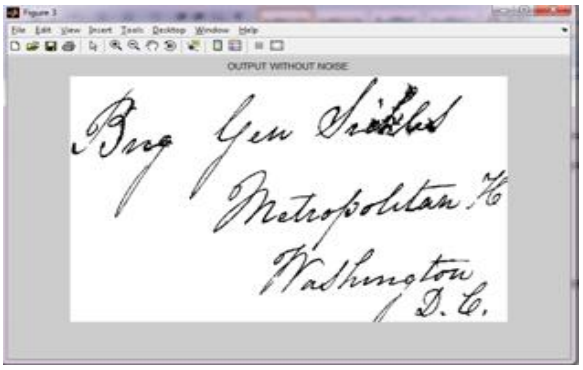


Fig 10: Output of Improved Ni-Black Algorithm of fig 9

Image 3:

Fig 11 shows the original input degraded image.

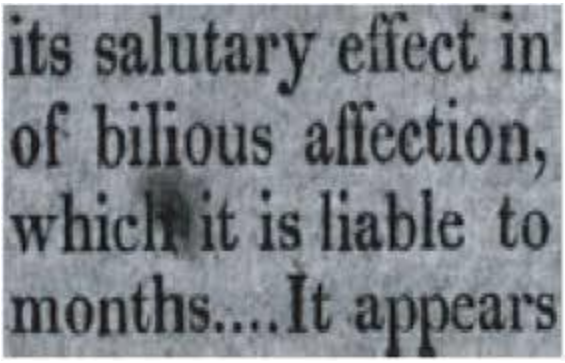


Fig 11: Original Image 3

Fig 12 shows the output of the Ni-Black algorithm after applying it on the original image.

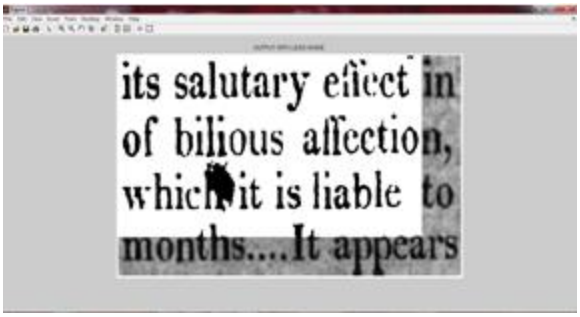


Fig 12: Restored Image of fig 11 Using Ni-Black Algorithm

Fig 13 shows the improved output of the Ni-Black algorithm after applying the filter on Ni-Black's output.

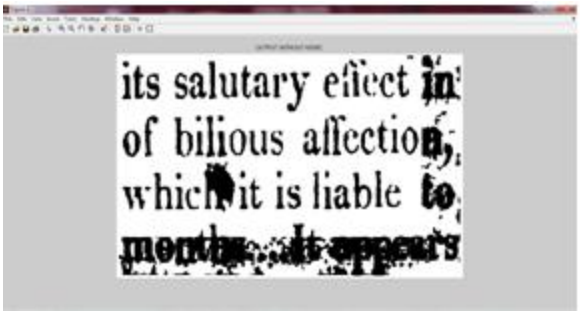


Fig 13: Output of Improved Ni-Black Algorithm of fig 12

Thus from the above images it is clear that the proposed algorithm gives a noise free image at the output.

Sr. No.	IMAGE TYPE	MSE	PSNR
1.	HT-01.png	0.8339	33.1393
2.	HT-02.png	0.8655	32.9234
3.	HT-03.png	0.7489	33.7614
4.	HT-04.png	0.8510	33.0216
5.	HT-05.png	0.8362	33.1233
6.	HT-06.png	0.8339	33.1393
7.	HT-07.png	0.8408	33.0910

The plot for the calculated parameters are being displayed here below:

Fig 14 gives the plot for MSE of the seven images.

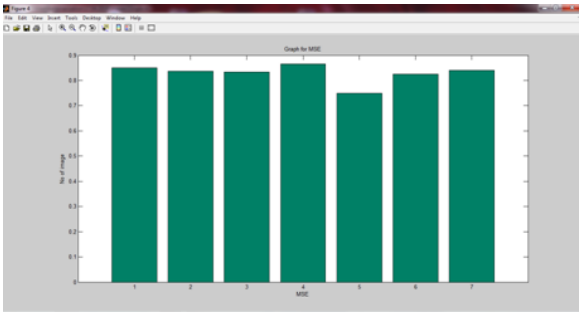


Fig 17: Plot for MSE

Fig 18 gives the plot for PSNR of the seven images.

` Fig 18: Plot for PSNR

Fig 19 gives the plot for Execution time of the seven images.

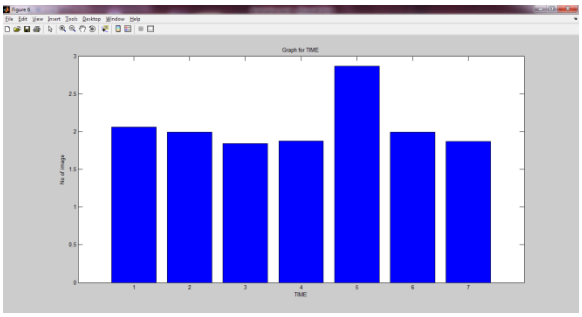


Fig 19: Plot for Execution Time

Given below are the tables of the parameters calculated using the proposed algorithm.

Table 1: Table for MSE and PSNR by proposed algorithm

Sr. No.	IMAGE TYPE	EXECUTION TIME (in seconds)
1.	HT-01.png	1.8397
2.	HT-02.png	1.8756
3.	HT-03.png	2.8705
4.	HT-04.png	2.0614
5.	HT-05.png	1.9958
6.	HT-06.png	1.8397
7.	HT-07.png	1.868262

Table 2: Table for Execution Time of proposed algorithm

The proposed method uses a local thresholding technique named Ni-Black thresholding, which is very efficient in removing noise from the degraded historical document images. The proposed Ni-Black algorithm with further improvement using filtering have greatly improved the degraded image as well as its PSNR. The average PSNR achieved by previous algorithm is 30.69 and the average PSNR obtained by the proposed algorithm is 33.19. Thus the proposed algorithm proves to be more efficient than other algorithms.

Algorithm	Avg. MSE	Avg. PSNR	Avg. Execution Time
Hybrid Binarization Technique	55.59	30.69	-
Proposed Method	0.83	33.19	2.07

Table 3: Comparison of Hybrid Binarization technique and proposed algorithm

REFERENCES

[1] Kavallieratou, E. and Stathis, S., "Adaptive Binarization of Historical Document Images", Proceedings of the 18th International Conference on Pattern Recognition (ICPR), 2006, pp 742-745.

[2] Ali, M. B. H., "Background Noise detection and Cleaning in Document Images", Proceedings of ICPR, 1996, pp 758-762.

[3] Niblack, W., "An Introduction to Digital Image Processing", Prentice Hall, 1986, pp 115 – 116.

[4] Morse, B. S., "Thresholding", Lecture 4.

[5] Anasuya Devi, H.K., "Thresholding: A Pixel level image processing methodology processing technique for an OCR system for Brahmi script", Ancient Asia, vol 1, 2006, pp 161- 165.

[6] Khurshid, K., Siddiqi, I., Faure, C. and Vincent, N., "Comparison of Ni-black inspired binarization methods for ancient documents", IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2009, pp. 72 470U-72 470U.

[7] Trier, D. and Taxt, T., "Evaluation of binarization methods for document images", Proceedings of the 1994 IEEE International Conference on Image Processing, 1994, pp 31-36.

[8] Kaur, J. and Mahajan, R., "A Review of Degraded Document Image Binarization Techniques", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol 3, 5, 2014, pp 6581-6586.

[9] Arya, S. C., Singh, R.S. and Mandoria, H.L., "Image Denoising in Hand Written Document for Degraded Documents using Wiener Filter Algorithm" International Journal for Research in Emerging Science and Technology, vol 2, 7, 2015, pp 50-56.

[10] Su, B., Lu, S. and Tan, C. L., "Robust Document Image Binarization Technique for Degraded Document Images", IEEE Transanctions on Image Processing, vol. 22, 4, 2013.

[11] Ranganatha D. and Ganga Holi, "Hybrid Binarization Technique for Degraded Documents", IEEE, 2015, pp 893-898

[12] Farid, S. and Ahmed, F., "Application of Ni-black's Method on Images", International Conference on Emerging Technologies, 2009, pp 280-286.