RESEARCH ARTICLE OPEN ACCESS

Remove Noise from Scanned Handwritten De-Graded Document Images Using Various Approaches

Kuljeet Singh [1], Gurinder Singh [2]
LCET, Katani kalan
Ludhiana -Punjab Technical University
India

ABSTRACT

Image is defined as sequential collection of pixels. Some images may be corrupted by degradation such as linear frequency distortion, noise and blocking artifacts. These sources of degradation may arise during image capture or processing and have a direct bearing on visual quality. Most offline handwriting recognition approaches proceed by segmenting characters into smaller pieces which are recognized separately. The recognition result of a word is then the composition of the individually recognized parts. In past few years a number of algorithms were developed by researchers for de-noising and restoration of the degraded documents. The restoration of handwritten degraded historical documents plays a major role in degradation process. The algorithms used in past few years were Ni-blacks algorithm, canny edge map algorithm, OTSU method, Back ground estimation, dynamic thresholding and image binarization etc. The results obtained by the previously implemented algorithms were not sufficient. So, in this thesis work three filters are proposed to improve the quality of degraded documents. Wiener filter algorithm, Sobel operator and Prewitt filter are used in this research work to check the output quality by implementing these three different approaches. In this research work the performance of all filters is measured by calculating size of Input and Output images, Mean square error and Peak signal to noise ratio. This shows that wiener filter is more suitable filter for de-noising and restoration of the handwritten degraded document images.

Keywords:- Degraded documents, de-noising, Signal to Noise ratio, Mean Square Error, Wiener filter algorithm.

I. FILTERS

There are a number of filters used in image processing for adding and removing noise from images like photographs, hand-written images, scanned images etc. Filters used in image processing are Prewitt, Sobel, Roberts, canny and wiener filter. We choose wiener filter to clear the de-graded documents scanned images, Wiener filter is itself an algorithm for clearing scanned documents, and rest all filters are used for edge detection.

II. WIENER FILTER

Wiener filters are a class of optimum linear filters which involve linear estimation of a desired signal sequence from another related sequence. In the statistical approach to the solution of the linear filtering problem, we assume the availability of certain statistical parameters (e.g. mean and correlation functions) of the useful signal and unwanted additive noise. The problem is to design a linear filter with the noisy data as input and the requirement of minimizing the effect of the noise at the filter output according to some statistical criterion. A useful approach to this filter-optimization problem is to minimize the mean-square value of the error signal that is defined as the difference between some desired response and the actual filter output. For stationary inputs, the resulting solution is commonly known as the Weiner filter. Its main purpose is to reduce the amount of noise present in a signal by comparison with an estimation of the desired noiseless signal.

III. DEGRADED IMAGES

Degradation in scanned document images result from poor quality of paper, the printing process, ink blot and fading, document aging, extraneous marks, noise from scanning, etc. The goal of document restoration is to remove some of these artifacts and recover an image that is close to what one would obtain under ideal printing and imaging conditions. The ability to

restore a degraded document image to its ideal condition would be highly useful in a variety of fields such as document recognition, search and retrieval, historic document analysis, law enforcement, etc. The emergence of large collections of scanned books in digital libraries has introduced an imminent need for such restorations that will aid their recognition or ability to search. Images with certain known noise models can be restored using traditional image restoration techniques such as Median filtering, Weiner filtering, etc.

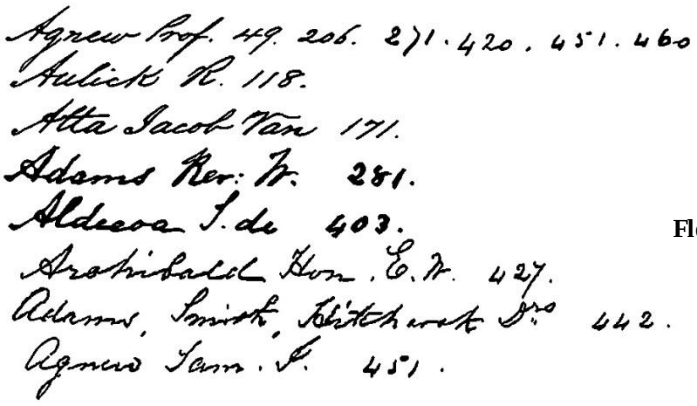


Figure 1: Degraded document (1)

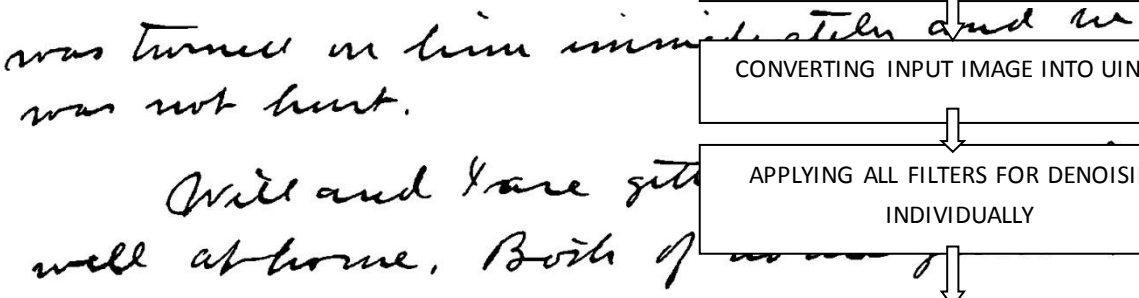
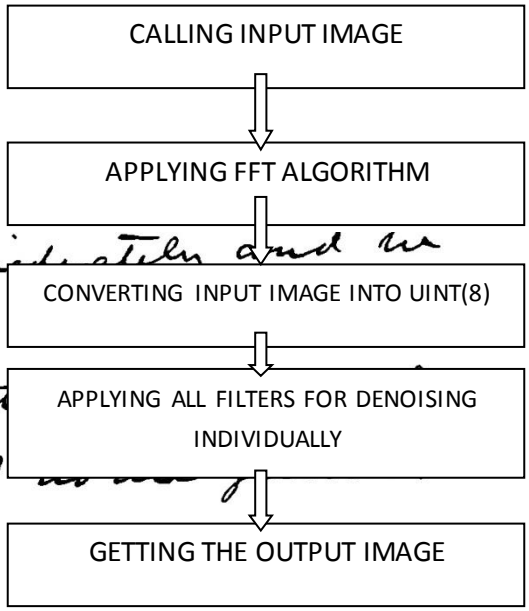


Figure 2: Degraded document (2)

However, in prac-tice, degradations arising from phenomena such as document aging or ink bleeding cannot be described using popular image noise models. Document processing algorithms improve upon the generic methods by incorporating document specific degradationmodels and text specific content models. Approaches that deal with highly degraded documents take a more focused approach by modeling specific types of degradations. For instance, ink-bleeding or backside reflection is one of the main reasons for degradation of historic handwritten documents. In this paper, we approach document restoration in a different way, and useful setting. We consider the problem of restoration of a degraded 'collection of documents' such as those from a single book. Such a collection of documents,arising from the same source, is often highly homogeneous in the script, font and other typesetting parameters. The availability of such a uniform collection of documents for learning allows us to:

- To reduce the noise from the scanned de-graded document images area by using wiener filter algorithm degraded images.
- To reduce Mean Square Error and calculate Peak Signal to Noise Ratio.
- To calculate execution time for our final implemented code.

Flow chart



At the next step, the edge information of the grey level image is combined with the binary result of the previous step. From all edge pixels, only those are selected that probably belong to text areas according to a criterion, number of pixels in output image and input image is calculated. Smoothing algorithm is then applied in order to fill text areas in the edge map. Finally, different parameters are calculated using different formulas.

IV. EVALUATION MEASURES

I MSE is **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_imageno_pixels_in_input_image).^2)./((Size_Of_Image).^2)

II 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: , where MSE defined in 2 refers to mean square error.

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.

III 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.

Results and Discussion

In proposed algorithm, are used to provide more clearity than in previous work. In this, results of all the intermediate steps of the proposed methods are highlighted. Implementation is done on MATLAB Experimental results of intermediate steps show the efficiency of the proposed approach. Results includes following steps:

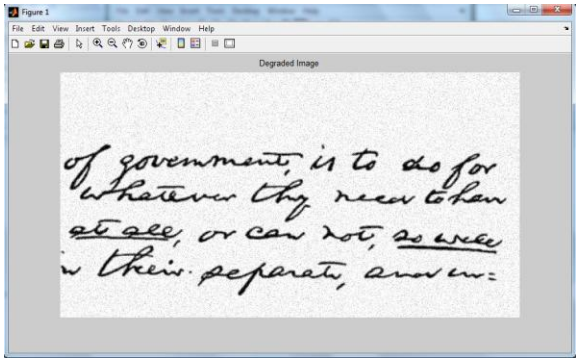


Figure 3: Degraded document

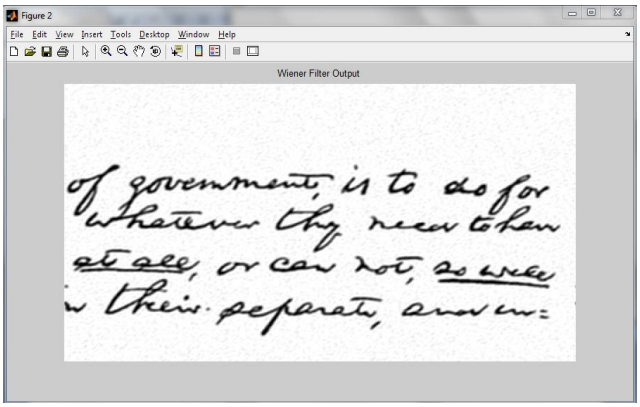


Figure 4: Restored image by Wiener Filter

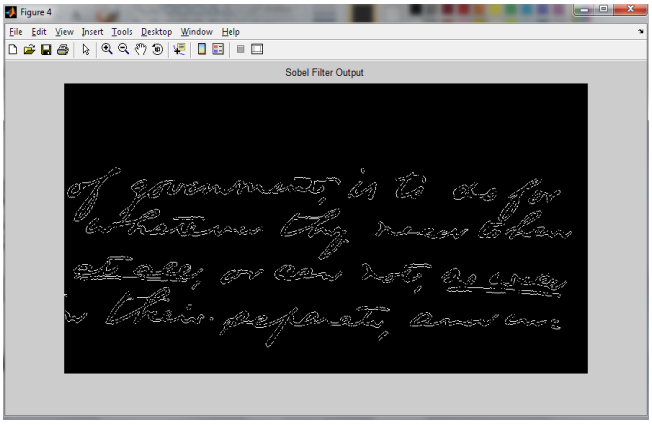


Figure 5: Restored image by Sobel Filter

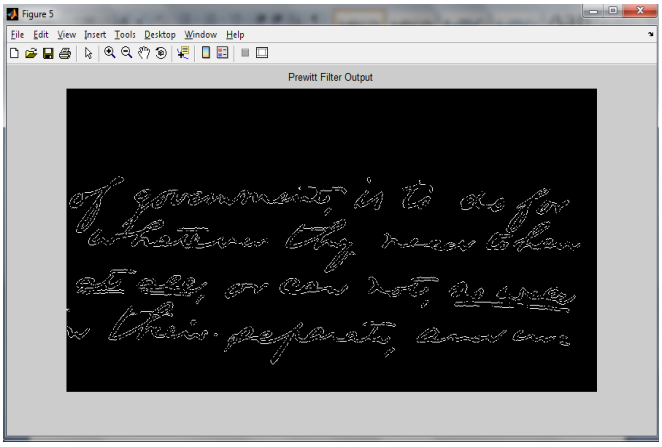


Figure 6: Restored image by Prewitt Filter

Table 1:- Table for PSNR

PSNR is Peak Signal to Noise Ratio, it is calculated after calculating Mean Square Error it should be maximum for the perfect output. The table above shows that the PSNR calculated by the Wiener Filter is perfect among all the filters. Graph for the PSNR calculated by different filters is given below.

Sr. No.	IMAGE TYPE	PSNR OF WIENER FILTER	PSNR OF SOBEL FILTER	PSNR OF PREWITT FILTER
1.	HT-01.jpeg	34.7065	24.6602	24.6624
2.	HT-02.jpeg	41.1011	24.2462	24.2465
3.	HT-03.jpeg	32.5148	24.9249	24.9210
4.	HT-04.jpeg	34.1580	24.6959	24.6935
5.	HT-05.jpeg	38.0003	24.3906	24.3908

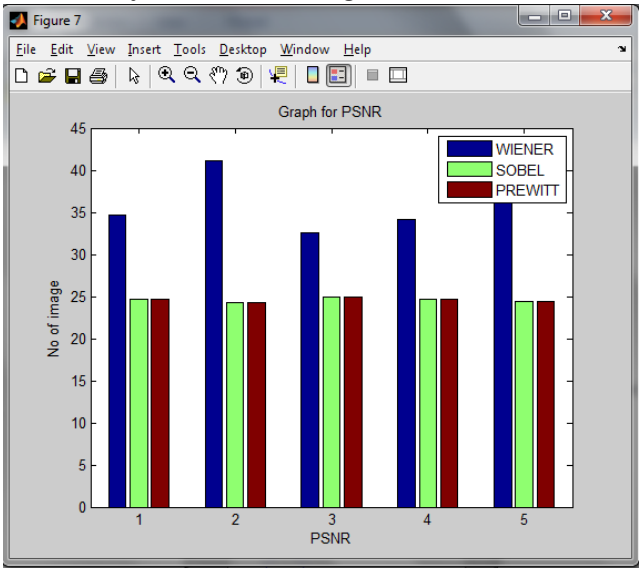


Figure 7: Plot for PSNR

Sr. No.	IMAGE TYPE	MSE OF WIENER FILTER	MSE OF SOBEL FILTER	MSE OF PREWITT FILTER
1.	HT-01.jpeg	0.0016	0.8720	0.8726
2.	HT-02.jpeg	0.0001	0.9592	0.9592
3.	HT-03.jpeg	0.0034	0.8204	0.8212
4.	HT-04.jpeg	0.0019	0.8649	0.8653
5.	HT-05.jpeg	0.0005	0.9279	0.9278

Table 2:- Table for MSE

MSE is Mean Square Error; it should be minimum for image processing images output. As shown in the table it is clear that the MSE calculated by the Wiener filter is minimum among all other filters. The graph for MSE is given below.

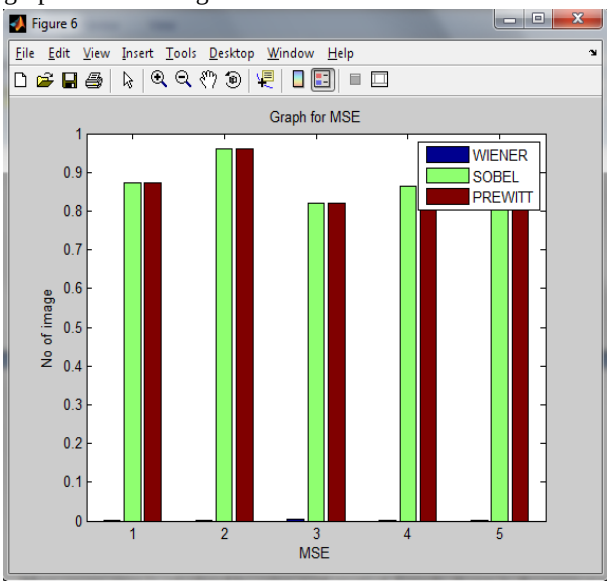


Figure 8: Plot for MSE

V. CONCLUSION

This research work is based on removing noise from degraded images (handwritten documents). The implemented algorithm is Wiener Filter Algorithm. Coding is done on MATLAB tool, the code is written and tested on a number of images from different DIBCO datasets. This method includes de-blurring or de noising of degraded documents. This research work develops a system which is used to clear the degraded documents. Parameters like Peak Signal to Noise Ratio, Image size, Mean Square Error etc. are calculated to show the improvement for our work. Comparison with other filters is also shown in the chapter above. The comparison shows that the Wiener filter gives better PSNR and MSE as compared to Sobel and Prewitt filters.

VI. FUTURE SCOPE

To develop an image technique that will become efficient for de-noising degraded images, blur effects and other noisy images. In this research work I took number of images to calculate various parameters like MSE, PSNR and Time to implement our design. The implementation is done on MATLAB tool with three different filter algorithms. In future someone

can use some other technique to implement same design with reduced time and more improved PSNR and can also calculate some other parameters to check the quality of improvement.

REFERENCES

[1] Sonia Saini and Ms. Richa Dogra (October,2015), "A Novel In To Approach for Image Binarization for Degraded Document Images" in Global Journal of Computers & Technology, Volume No 3 and Issue No 2, pp 176-186.

[2] Megha Paul (July-August 2015), "Effective Process to Remove Broken Lines Effect from Degraded Document Images Using MATLAB Algorithm" in International Journal of Computer Science Trends and Technology (IJCST) – Volume No 3, Issue No 4 and pp119-124.

[3] Er Neha Kundal and Er Anantdeep (july 2015), "Performance Evaluation of Novel Historical Documents Restoration Algorithm" in IJCSET- Volume No 5, Issue No 7 and pp 278-282.

[4] Sanjay Chandra Arya, Rajesh Shyam Singh and Hardwari Lal Mandoria (july 2015), "Image Denoising in Hand Written Document for Degraded Documents using Wiener Filter Algorithm"in INTERNATIONAL JOURNAL FOR RESEARCH IN EMERGING SCIENCE AND TECHNOLOGY-VOLUME No2, ISSUE No 7 and pp 50-56.

[5] K. SRUJANA and D. C. VINOD R KUMAR (june 2015), "Evaluation of Ancient Documents and Images by using Phase Based Binarization" in IJIT- Volume No 3, Issue No 2 and pp 0077-0082.

[6] Er Neha Kundal and Er. Anantdeep (May 2015), "FUZZY FILTER BASED HISTORICAL DOCUMENT WITH GEODESIC MORPHOLOGY-A REVIEW" in International Journal For Technological Research In Engineering- Volume No 2, Issue No 9 and pp 2024-2026.

[7] Rakchambigai.S.K, Vasmiya.G, Nivetha.M, Santhiya.J and G.Dheepak (March 2015),"Denoising and Phase–based Binarization of Ancient Document using Adaptive Contrast Map" in International Journal of Innovative Research in Computer and Communication Engineering- Volume No 3, Issue No 2 and pp 28-32.

[8] Sarika Jain and Pankaj Parihar (Jan 2015), "Remove Noise and Reduce Blurry Effect From Degraded Document Images Using MATLAB Algorithm" in International Journal of Engineering Research and General Science- Volume No 3, Issue No 1 and pp 1333-1339.

[9] Bency Jacob and Prof. S. B. Waykar (October 2014), "A Survey On Binarization Of Historical Degraded Documents" in IJIFR- Volume No 2, Issue No 2 and pp 400-405.

[10] Bolan Su, Shijian Lu and Chew Lim Tan (APRIL 2013), "Robust Document Image Binarization Technique for Degraded Document Images" in IEEE TRANSACTIONS ON IMAGE PROCESSING, Volume No 22 and Issue No 4.

[11] J. Bharathi and Dr. P. Chandrasekar Reddy (November 2012), "Variational Background Modeling Using Grid Point Sampling for Document Image Binarization" in ISSN 2250-2459, Volume No 2 and Issue No 11, pp 460-468.

[12] Arathy Reghukumar, Divya Haridas, Poornima Rajan, Sachin Kumar S and K. P. Soman, "Active Contour based Document Image Segmentation and Restoration using Split-Bregman and Edge Enhancement Diffusion" in International Journal of Computer Applications- Volume No 54, Issue No 13 and pp 14-20.

[13] Oke Alice, Omidiora Elijah, Fakolujo Olaosebikan, Falohun Adeleye, Olabiyisi (AUGUST 2012)," Effect of Modified Wiener Algorithm on Noise Models" in International Journal of Engineering and Technology Volume No 2 and Issue No 8, pp 1439-1448.

[14] Taeg Sang Cho, C. Lawrence Zitnick,Neel Joshi, Sing Bing Kang, Richard Szeliski (APRIL 2012), "Image Restoration by Matching Gradient Distributions" in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Volume No 34 and Issue No 4, pp 683-694.

[15] Thirimachos Bourlai, Arun Ross and Anil K. Jain (June 2011), "Restoring Degraded Face Images : A Case Study in Matching Faxed, Printed and Scanned Photos" in IEEE

TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY- Volume No 6, Issue No 2 and pp 371-384.

[16] Xiang Li, Xiuqin Su, and Lei Ji (August 2010), "Image Denoising via Doubly Wiener Filtering with Adaptive Directional Windows and Mean Shift Algorithm in Wavelet Domain" in Proceedings of the 2010 IEEE International Conference on Mechatronics and Automation (ICMA), Chinese Acad. of Sci., Xi"an, China, pp 114-118.

[17] RezaFarrahi Moghaddam and Mohamed Cheriet (August 2010), "A Variational Approach to Degraded Document Enhancement" in IEEE Volume No 32, Issue No 8 and pp 1347-1361.

[18] Oliver Whyte, Josef Sivic, Andrew Zisserman and Jean Ponce (2010), " Non-uniform Deblurring for Shaken Images " in Computer Vision and Pattern Recognition (CVPR) IEEE Conference, San Francisco, CA, pp 491-498.

[19] Jacob Benesty , Jingdong Chen , and Yiteng (Arden) Huang (2010), " STUDY O F THE WIDELY LINEAR WIENER FILTER FOR NOISE REDUCTION " in Acoustics Speech and Signal Processing (ICASSP) IEEE conference, Univ. of Quebec, Montreal, QC, Canada, pp 205-208.

[20] Gangamma, Srikanta Murthy K (2010),"A Combined Approach for Degraded Historical Documents Denoising Using Curvelet and Mathematical Morphology " in Computational Intelligence and Computing Research (ICCIC) IEEE Conference Dept. of ISE, PES Inst. of Technol., Bangalore, India. pp 1-6.