RESEARCH ARTICLE OPEN ACCESS

Effective Process to Remove Broken Lines Effect from Degraded Document Images Using MATLAB Algorithm

Megha Paul

Gurukul Vidyapeeth College Of Engineering And Technology (Ptu) - Banur

Punjab – India

ABSTRACT

Most offline handwriting recognition approaches proceed by segmenting characters into smaller pieces which are recognized differently. The result of recognition is composition of the individually recognized parts. Inspired by results in cognitive researchers have begun to focus on holistic word recognition approaches. Here we present an effective approach for degraded documents, which is motivated by the fact that for severely degraded documents a segmentation of words into characters will produce very poor results. The poor quality of original documents does not allow us to recognize them with high accuracy - our goal here is to produce transcriptions that will allow successful retrieval of images, which has been shown to be feasible even in such noisy environments. We believe that this is the first systematic approach to recognizing words in historical manuscripts with extensive experiments. Our experiment is to clear the degraded documents using filtering approach. We will use wiener filter for removing the broken lines from different images using wiener filter algorithm. We will also implement this design using GUI for selecting different images from self-created data-base.

Keywords:- Degraded documents, broken lines, GUI, Wiener filter algorithm.

I. INTRODUCTION

There are 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 have chosen wiener filter algorithm to remove the broken lines effect from the degraded documents. 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 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.

II. 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. In this paper, we approach document restoration in a different 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 typeset parameter. The most important technique for removal of blur in images due to linear motion or unfocussed optics is the Wiener filter. From a signal processing standpoint, blurring due to linear motion in a photograph is the result of poor sampling. Each pixel in a digital representation of the photograph should represent the intensity of a single stationary point in front of the camera. Unfortunately, if the shutter speed is too slow and the camera is in motion, a given pixel will be an amalgam of intensities from points along the line of the camera's motion.

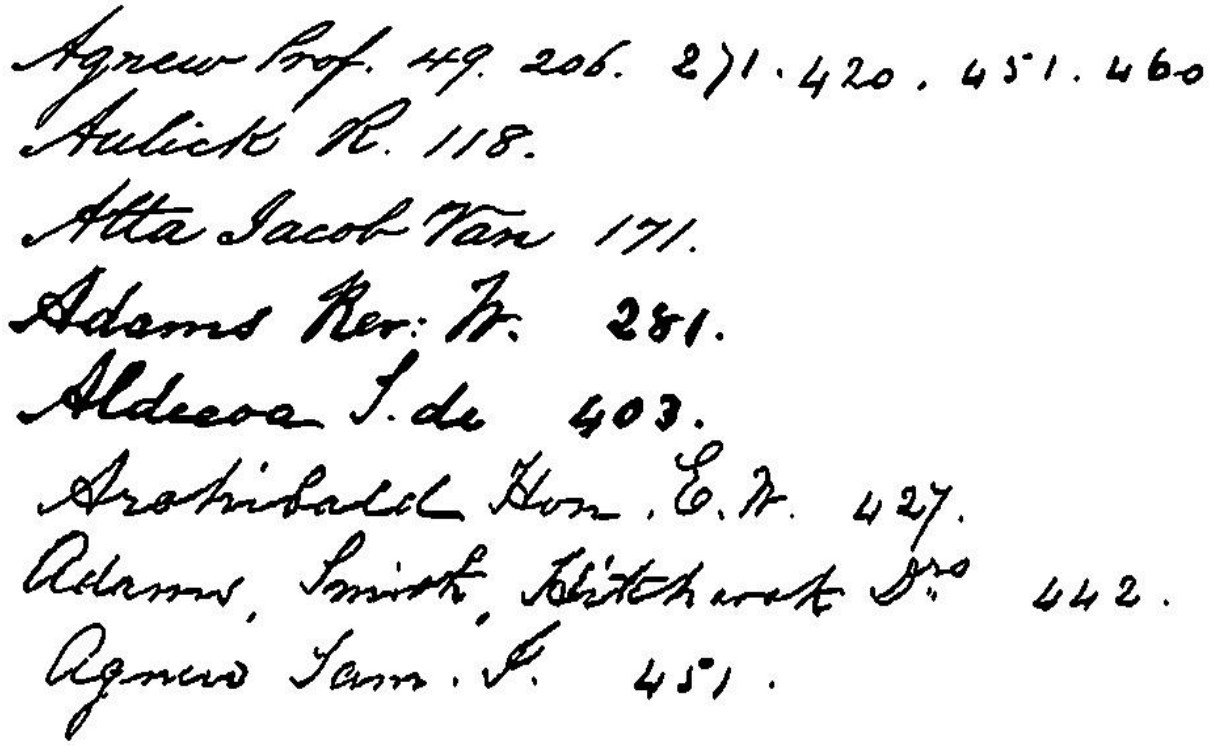


Figure 1: Degraded document (1)

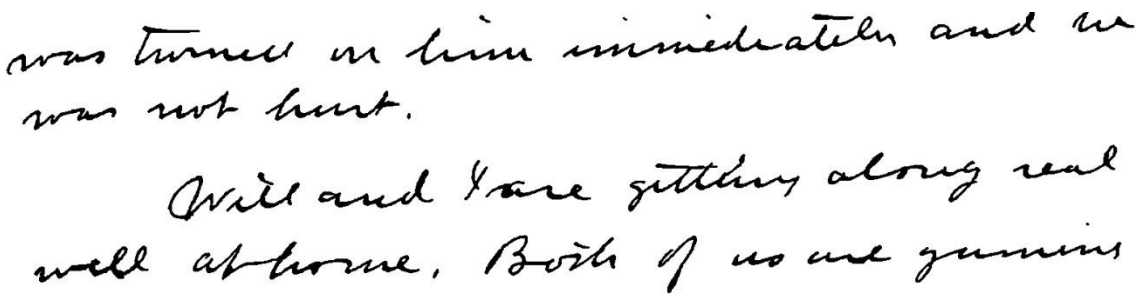


Figure 2: Degraded document (2)

However, in practice, 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 degradation models 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

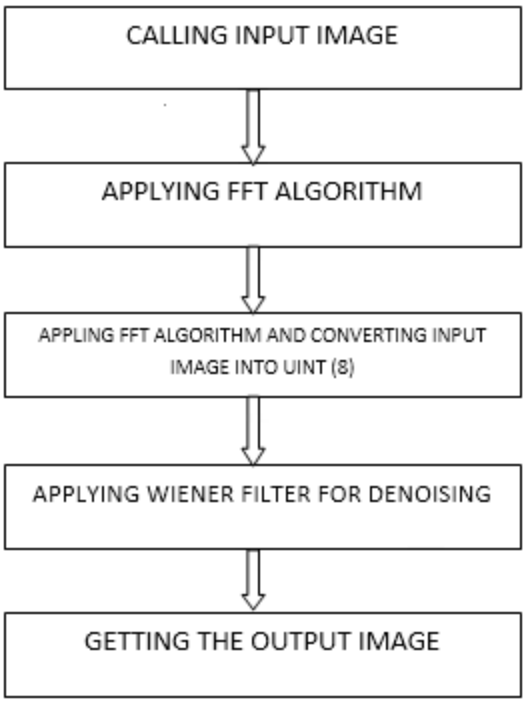
degraded 'collection of documents' such as those from a single book. Such a collection of documents, arising from the same source, is often highly homogeneous in the script, font and other typesetting parameters. The availability of such a uniform collection of documents for learning allows us to:

• To improve the document quality by filling the broken lines.
• To implement wiener filter algorithm for filling broken lines in degraded images.
• Evaluating various parameters for studying percentage of improvement.

III. RELATED WORK

The work related to our work done so far is as, Xiang Li, Xiuqin Su [5] proposed model a for degraded image as an original image that has been subject to linear frequency distortion and additive noise. he develop a distortion measure of the effect of frequency distortion, and a noise quality measure of the effect of additive noise. Mohammed M. Siddeq Dr. Sadar Pirkhider Yaba et. al. stated an algorithm for image de-nosing based on two level discrete wavelet transform and Wiener filter. At first The DWT transform noisy image into sub-bands, consist of low-frequency and high-frequencies and then estimate noise power for each of the sub-band. The noise power is computed through two important computations, compute square of variance for each sub-band then compute the mean of the variance. Jacob Benesty [8] et. al. stated that the images of outdoor scenes captured in bad weather often suffer from poor contrast. Under bad weather conditions, the light reaching a camera is severely scattered by the atmosphere and the resulting decay in contrast varies across the scene and is exponential in the depths of scene points. Gangamma, Srikanta Murthy K [9] designed speech recognition system using cross-correlation and FIR Wiener Filter. The algorithm is designed to ask users to record the words three times. The first and second recorded words are different words which will be used as the input signals. The third recorded word is the same word as one of the first two recorded words. The recorded signals corresponding to these words are then used by the program based on cross-correlation and FIR Wiener Filter to perform speech recognition. Bolan Su, Shijian Lu [1] et. al. concluded that Segmentation of text from badly degraded document images in a very challenging task due to the high inter/intravariation between the document background and the foreground text of different document images. He propose a novel document image binarization technique that addresses these issues by using adaptive image contrast. The adaptive image contrast is a combination of the local image contrast and the local image gradient that is tolerant to text and background variation which are caused by different types of document degradations.

IV. FLOW CHART



The edge information of the grey level image is combined with the whole binary result of the previous step. From the 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.

V. EVALUATION MEASURES

(i) Mean Square Error is calculated using the mathematical formula given below

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

(ii) 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

PSNR = 20LOG$_{10}$ (255 / MSE)

The quality of the image is higher if the PSNR value of the image is high. So we used to improve the PSNR, where PSNR is inversely proportional to MSE value of the image, the higher the PSNR value is, the lower the MSE value will be obtained.

(iii) Time calculation:-To calculate the execution time for the code, we use inbuilt TIC & TOC commands from MATLAB library.

VI. RESULTS AND DISCUSSION

In proposed algorithm, are used to provide more clarity 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:

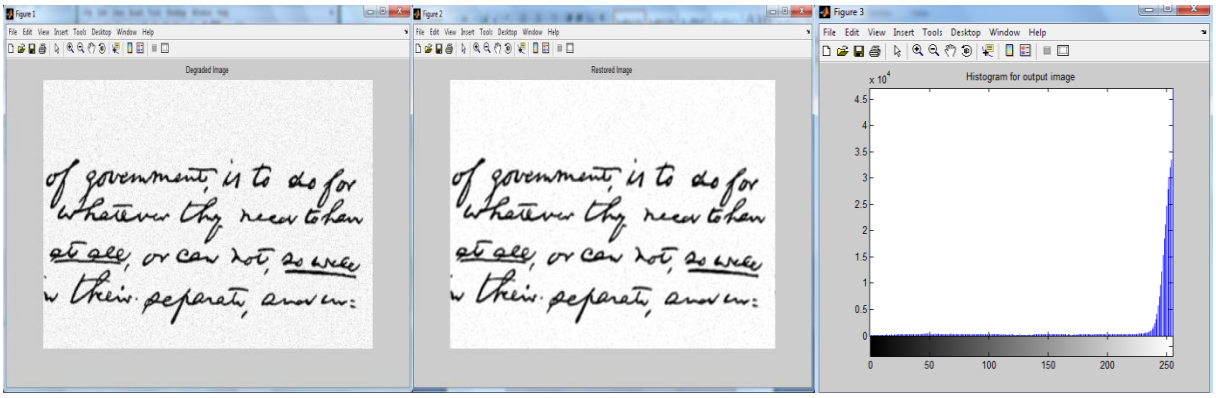
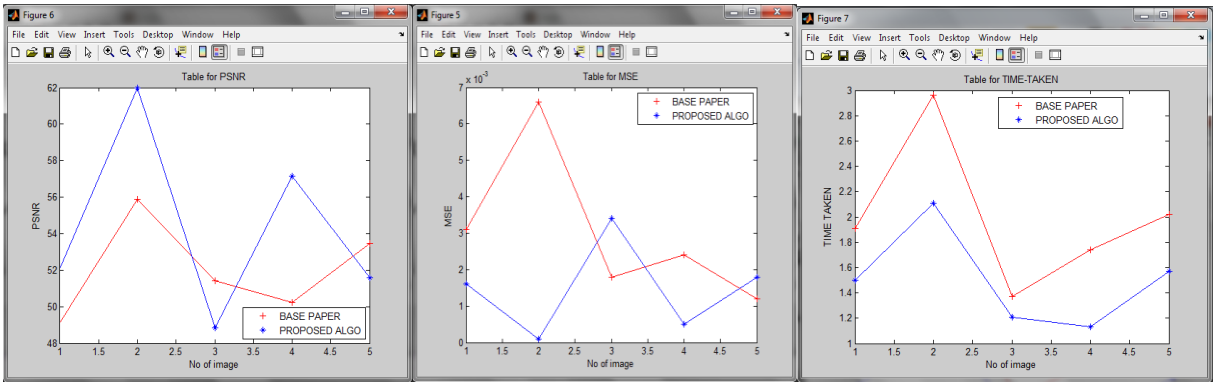


Figure1 a) input image b) restored image c) Histogram



d) PSNR calculation e) MSE calculation f) Execution Time calculation

Sr. No.	Image type	PSNR	MSE	Time-Taken (in Seconds)
1.	DB-01.jpeg	52.14	0.0016	1.50
2.	DB-02.jpeg	61.96	0.0001	2.11

3.	DB-03.jpeg	48.85	0.0034	1.21
4.	DB-04.jpeg	57.16	0.0005	1.13
5.	DB-05.jpeg	51.60	0.0018	1.57

Table 1: Evaluation parameters for DIBCO 2010 dataset

VII. CONCLUSION

This research work is based on removing noise from degraded images (handwritten documents). Our implemented algorithm is Wiener Filter Algorithm. This method includes debluring or de noising of degraded documents. This project develops a system which is used to clear the degraded documents. Parameters like PSNR, Image size, MSE etc are also calculated to show the improvement for our work. We have improved the PSNR using MATLAB algorithms as mentioned in the base paper. We have also used the same formula for calculating PSNR. The paper shows the average PSNR calculated on DIBCO data set 2010. The average PSNR calculated is 54.342.

VIII. 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 we took number of images for our thesis work, we calculate various parameters like MSE, PSNR and Time to implement our design. One could use some other technique to implement same design with reduced time and could also calculate some other parameters and some improved GUI design.

REFERENCES

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

[2] J. Bharathi1, Dr. P. Chandrasekar Reddy in November 2012, " Variational Background Modeling Using Grid Point Sampling for Document Image Binarization " in ISSN 2250-2459, Volume 2, Issue 11.

[3] Oke Alice, Omidiora Elijah, Fakolujo Olaosebikan, Falohun Adeleye, Olabiyisi ,in AUGUST 2012," Effect of Modified Wiener Algorithm on Noise Models" in International Journal of Engineering and Technology Volume 2 No. 8.

[4] Taeg Sang Cho, C. Lawrence Zitnick,Neel Joshi, Sing Bing Kang, Richard Szeliski in APRIL 2012, "Image Restoration by Matching Gradient Distributions" in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 34, NO. 4, APRIL 2012.

[5] Xiang Li, Xiuqin Su, and Lei Ji in 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 August 4-7, 2010, Xi'an, China.

[6] RezaFarrahi Moghaddam and Mohamed Cheriet in August 2010, " A Variational Approach to Degraded Document Enhancement " in IEEE VOL. 32, NO. 8, AUGUST 2010.

[7] Oliver Whyte ,Josef Sivic ,Andrew Zisserman , Jean Ponce in 2010, " Non-uniform Deblurring for Shaken Images " in IEEE 978-1-4244-6985-7.

[8] Jacob Benesty , Jingdong Chen , and Yiteng (Arden) Huang in 2010, " STUDY O F THE WIDELY LINEAR WIENER FILTER FOR NOISE REDUCTION " in IEEE 978-1-4244-4296-6.

[9] Gangamma, Srikanta Murthy K in 2010 ," A Combined Approach for Degraded Historical Documents Denoising Using Curvelet and Mathematical Morphology " in IEEE 978-1-4244-5967-4.

[10] S. Suhaila and T. Shimamura in 2010, " Power Spectrum Estimation Method for Image Denoising by Frequency Domain Wiener Filter " in IEEE 978-1-4244-5586-7.

[11] Mohammed M. Siddeq and Dr. Sadar Pirkhider Yaba in 2009, "Using Discrete Wavelet Transform and Wiener filter for Image De-nosing" in Wasiit Journall for Sciience & Mediiciine 2009 2(2): (18-30).

[12] B. Gatos, K. Ntirogiannis and I. Pratikakis in 2009, " ICDAR 2009 Document Image Binarization Contest (DIBCO 2009) " in 2009 10th International Conference on Document Analysis and Recognition 978-0-7695-3725-2.

[13] Huaigu Cao ,and Venu Govindaraju in 2009 ," Preprocessing of Low - Quality Hand written Documents Using Markov Random Fields " in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 31, NO. 7, JULY 2009.

[14] Jung-San Lee, Member, IEEE, Chin-Hao Chen, and Chin-Chen Chang, Fellow, IEEE in July 2009, " A Novel Illumination-Balance Technique for Improving the Quality of Degraded Text–Photo Images" in IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 19, NO. 6, JUNE 2009.

[15] B. Gatos, I. Pratikakis and S.J. Perantonis in 2008 , " Efficient Binarization of Historical and Degraded Document Images" in The Eighth IAPR Workshop on Document Analysis Systems.

[16] Russell Hardie in 2007,December ," A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter " in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 12, DECEMBER 2007

[17] Jingdong Chen, Member, IEEE, Jacob Benesty, Senior Member, IEEE, Yiteng (Arden) Huang, Member, IEEE, and Simon Doclo, Member, IEEE in July 2006 , " New Insights Into the Noise Reduction Wiener Filter " in IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 4, JULY 2006.

[18] Amer Dawoud and Mohamed S. Kamel in September 2004, " Iterative Multimodel Subimage Binarization for Handwritten Character Segmentation " in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 9, SEPTEMBER 2004.

[19] Srinivasa G. Narasimhan and Shree K. Nayar in JUNE 2003, " Contrast Restoration of Weather Degraded Images" in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 6.

[20] Department of Systems Design Engineering, University of Waterloo, Canada in 2003, " Iterative Sub-image Binarization for Document Images " in IEEE 0-7803-7750-8.