Feature Vector Table Based
Image Binarization for Degraded Document Images

Geetanjali
M.Tech (CSE), Indo Global College of Engineering, Abhipur, Mohali, Punjab, India
Kumud Sachdeva
A.P. (CSE), Indo Global College of Engineering Abhipur Mohali Punjab, India

Abstract:
Image is produced to record or display useful information but due to imperfections in the image capturing process and devices, the recorded image invariably represents a degraded version of the original. These degradations may be as a result of addition of noise, geometrical degradations, illumination and blur. These problems degrade the final quality of the acquired images for display and thus complicate the task of finding the error in object recognition and event detection. The primary objective of this paper is to develop a method for filtering, clearing the historical and degraded documents. 'Wiener filter algorithm' use to clear degraded historical handwritten documents .Experimental results is calculated by using MSE(i.e. mean square error) and PSNR (i.e. peak signal noise ratio).This algorithm helps to get the maximum accuracy and performance.

Keywords: Degraded images, noise, denoising, wiener filter, MSE, PSNR

1. Introduction

The images are used in different field as Medical visualization, Image enhancement/ restoration, Artistic effect and human computer interfaces. Text segmentation is a very challenging task. There is high inter/intra-variation in the document background and foreground text of different document images.[9].Images of outdoor scenes captured in bad weather often suffer from poor contrast. In the 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[3]. Images used to degraded as:-

- **Broken line structures**: Gaps of all sizes in lines were roughly counted. Large gaps were considered worse than small.
- **Broken symbols, text, etc.:** Symbols and text characters with gaps were roughly counted, and the degree of fragmentation was also assessed.
- **Blurring of lines, symbols and text:** Both the number of blurred print objects, and the degree of blurring were assessed.
- **Loss of complete objects**: The number of print objects which were completely lost was roughly counted.
- **Noise in homogeneous areas:** Both the number and the size of noisy spots and false objects in both background and print were estimated.

1.1. Few researchers have studied the noise in degraded documents

Niranjan Damera-Venkata, Thomas D. Kite, Wilson S. Geisler, Brian L. Evans and Alan C. Bovik [1] 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.
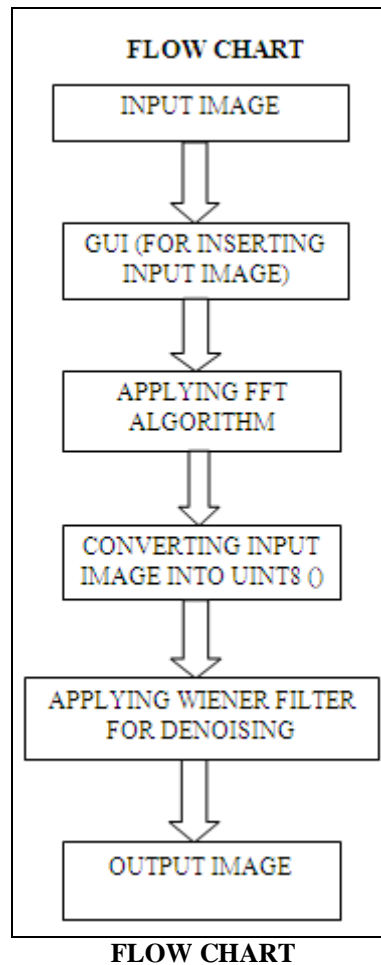
Deepak, Vikas Mittal [8] designed speech recognition system using cross-correlation and FIR Wiener Filter. The algorithm is designed to ask users to record the data in words three times. Starting first and second recorded words are different words which will be used as the input signals. The third recorded 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.

The above methods discussed that the degradation of document could be determined by various methods. Physics-based model that describes the appearances of scenes in uniform bad weather conditions are not sufficient to remove weather effects from images.[3] Distortion measure (DM) of the effect of frequency distortion, and a noise quality measure (NQM) of the effect of additive noise in which one parameter cannot indicate the relative visual impact of the degradations that may occur[1].The adaptive image contrast map

is tolerant to text and background variation rather than whole degraded document[9].It is not easy to clear the degraded documents/images and the contrast varies [4].This paper proposed an wiener filter algorithm to solve the problem of degraded documents which are noisy ones.

2. The Proposed Technique

The proposed methodology for efficient filtering of historical and degraded document images is illustrated in Fig. 1. It consists of many different steps. At the first step, at the run time GUI will execute and preprocessing based on Wiener filtering is applied. At the next step, several binarization results are combined in order to produce a binary (b/w) image taking into account the agreement in the majority of binarization methodologies.

FLOW CHART

INPUT IMAGE

↓

GUI (FOR INSERTING INPUT IMAGE)

↓

APPLYING FFT ALGORITHM

↓

CONVERTING INPUT IMAGE INTO UINT8 ()

↓

APPLYING WIENER FILTER FOR DENOISING

↓

OUTPUT IMAGE

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.

3. Problem Formulation

Image enhancement technique is used to enhance the quality of degraded document images .In the degraded images we generally found the problems like Broken line structures in which the gaps of all sizes in lines were roughly counted and large gaps were considered worse than small. Broken symbols, text, etc in which Symbols and text characters with gaps were roughly counted ,and the degree of fragmentation was also assessed. Blurring of lines, symbols and text in which both the number of blurred print objects, and the degree of blurring were assessed. Loss of complete objects in which the number of print objects which were completely lost was roughly counted. Noise in homogeneous areas in which the number and the size of noisy spots and false objects in both background and print were estimated. IBT using limited number set of rules with2x2 mask which is not able to predict various type of noises, we will try to remove the blurry effect from degraded images using wiener filter algorithm. Wiener filter is used for restoration purpose.

4. Objective for Study

- To reduce the noise in homogenous areas.
- To implement wiener filter algorithm for removing the blurry effect from degraded images.
- Evaluating various parameters for studying percentage of improvement.
- To use GUI for filtering a number of images.
- To calculate execution time for taking output for our final code.

5. Experiment Results

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:
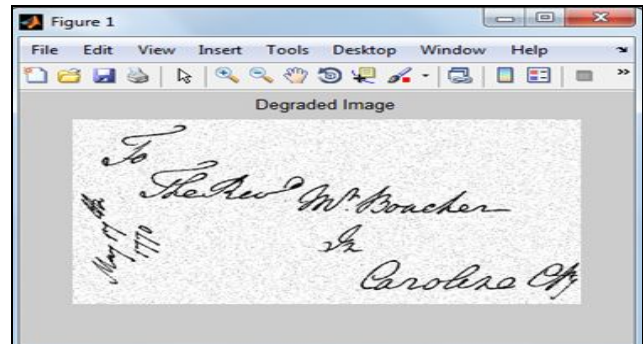

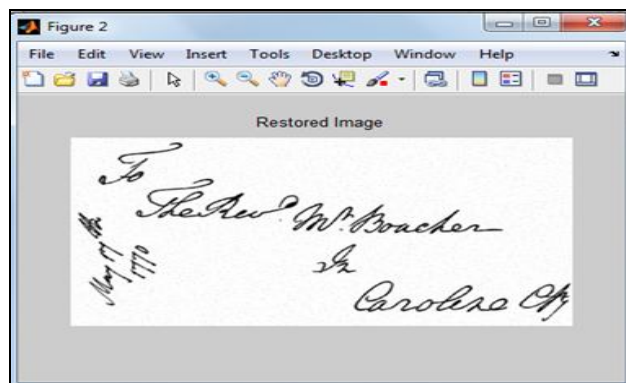*Fig 1: Opening GUI*

Fig 2. Degraded image
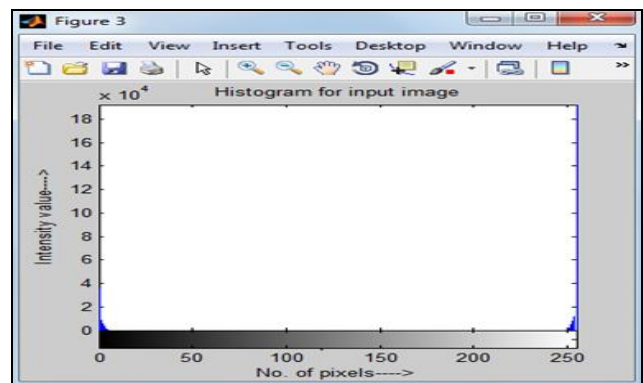

Fig 3: Restored Image
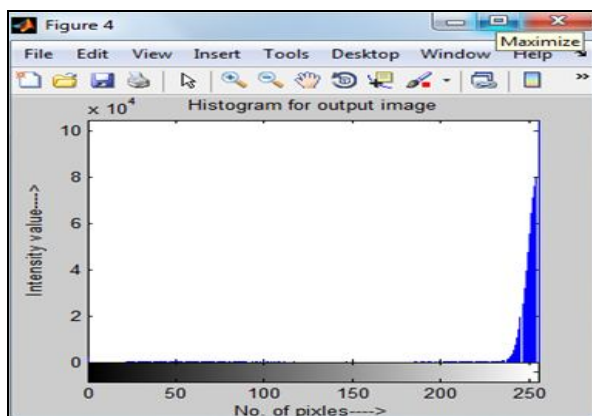

Fig 4 Histogram for input image

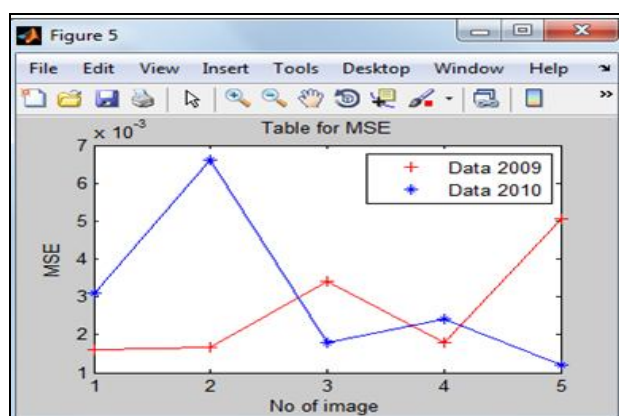*Fig 5: Histogram of Output Image*

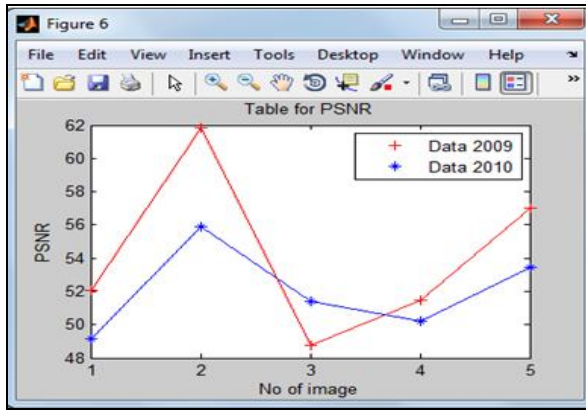*Fig.6 Graph of MSE values for proposed work*

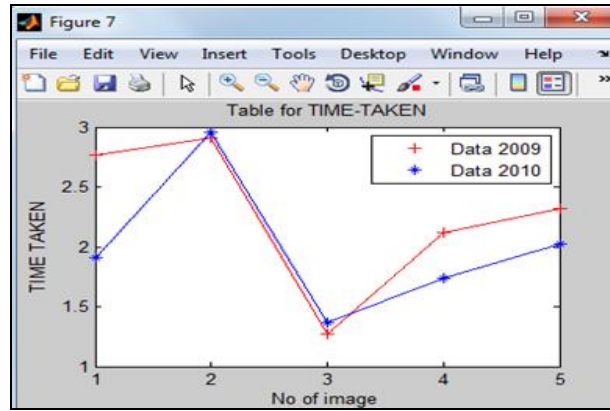*Fig.7 Graph of PNSR values for proposed work*   *Fig 8: Graph of TIME-TAKEN values for proposed work*

In the following tables the various parameters used to calculated by testing on the competition datasets(DIBCO 2009 &H-DIBCO 2010) :-

EVALUATION RESULTS OF THE VARIOUS IMAGES OF THE DIBCO2009

S.NO.	NO.OF PIXELS OF I/P IMAGES	NO. OF PIXELS OF O/P IMAGES	IMAGES SIZE	MSE	PSNR	TIME TAKEN(*10-3)
1	827299	861704	862650	0.0016	52.0497	2.776000
2	1275212	1291893	1292236	0.0016663	61.8478	2.91000
3	269340	285982	286344	0.0034	48.7791	1.279000
4	604542	631549	633871	0.0018	51.4759	2.121000
5	933867	955400	956133	0.00507	57.0137	2.324000

Table 1: Proposed Method (Wiener Method)

EVALUATION RESULTS OF THE VARIOUS IMAGES OF THE H-DIBCO2010

S.NO.	NO.OF PIXELS OF I/P IMAGES	NO. OF PIXELS OF O/P IMAGES	SIZE OF I/O IMAGES	MSE	PSNR	TIME TAKEN(*10-3)
1	528721	560153	565820	0.0031	49.1715	1.919000
2	1283804	1317755	1320370	0.0066117	55.8623	2.964000
3	318176	332435	332478	0.0018	51.4189	1.373000
4	476790	501494	502095	0.0024	50.2258	1.747000
5	650976	673892	674866	0.0012	53.4470	2.028000

Table 2: Proposed Method (Wiener Method)

Following table used to represent the our results are better than the other methods:-

Sl. No.	METHOD	DIBCO-2009 PSNR	H-DIBCO-2010 PSNR
1	OTSU	15.34	17.51
2	SAUV	16.39	15.96
3	NIBL	9.89	15.73
4	BERN	8.89	8.57
5	GATO	16.5	15.12
6	LMM	18.5	17.83
7	BE	18.6	18.14
8	IBT[9]	19.65	20.12
9	PROPOSED	61.8478	55.8623

Table 3: Evaluation Results of the Datasets

6. Conclusion

This paper work is based on removing noise from degraded images (handwritten documents). Our implemented algorithm is Wiener Filter Algorithm. This method includes histogram equalization and debluring. This paper develops a system which is used to clear the

degraded documents. We formulate number of parameters for our output and input images. We used to reduce the noise in homogenous areas, implement wiener filter algorithm for removing the blurry effect from degraded images, Evaluating various parameters for studying percentage of improvement and calculate execution time for taking our final output from our code. We reduce the amount of computation by not including other filters to our algorithm from which the execution time for our code gets very small.

6.1. Future scope
For developing an image technique that will become efficient for clearing degraded images, blur images and other noisy images. In this paper we took number of images for our research work, we calculate MSE, PSNR and Time to implement our design parameters. One can use some other technique to implement same design with reduced time. Someone can also calculate some other parameters and can improve GUI design.

7. References

1. Niranjan Damera-Venkata, Student Member, IEEE, Thomas D. Kite, Wilson S. Geisler, Brian L. Evans, Senior Member, IEEE, and Alan C. Bovik, Fellow, IEEE in APRIL 2000, " Image Quality Assessment Based on a Degradation Model" in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 9, NO. 4.
2. Nikolas P. Galatsanos, Member, IEEE, Vladimir Z. Mesarovic´, Rafael Molina, and Aggelos K. Katsaggelos, Fellow, IEEE in OCTOBER 2000, " Hierarchical Bayesian Image Restoration from Partially Known Blurs" in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 9, NO. 10.
3. Srinivasa G. Narasimhan and Shree K. Nayar in JUNE , " Contrast Restoration of Weather Degraded Images" in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 6.
4. 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.
5. Taeg Sang ChoC. Lawrence Zitnick, Neel Joshi, Sing Bing Kang and Richard Szeliski, in APRIL 2012 ,"Image Restoration by Matching Gradient Distributions" in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 34, NO. 4.
6. 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.
7. G. NAGENDRA, V.SRIDHAR, M.PRAMEELAMMA and M.ZUBAIR in SEPTEMBER 2012 , " FINGERPRINT IMAGE ENHANCEMENT USING FILTERING TECHNIQUES" in International Journal of Computer Science Engineering (IJCSE).
8. Deepak and Vikas Mittal in MAY 2013 , " Speech Recognition using FIR Wiener Filter" in International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 2, Issue 5.
9. 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