



## Efficient Technique for Improving Quality of Degraded Images [DIBCO]

<sup>1</sup>Anita Rana, <sup>2</sup>Dr. V. K. Banga

<sup>1</sup>M-Tech Scholar, Department of Electronics & Communication Engineering,  
Amritsar Collage of Engineering & Technology, Amritsar, Punjab, India

<sup>2</sup>Prof. & Principal, Amritsar Collage of Engineering & Technology, Amritsar, Punjab, India

*Abstract: Digital images are subjected to blurring due to many hardware limitations such as atmospheric disturbance, device noise and poor focus quality. Most offline handwriting recognition approaches proceed by segmenting characters into smaller pieces which were recognized separately. This proposed technique aimed to provide a basic knowledge of image degradation and restoration process. The recognition result of a word is then the composition of the individually recognized parts [1]. Inspired by results in cognitive psychology, researchers have begun to focus on holistic word recognition approaches. Here we present a holistic word recognition approach for degraded documents, which was motivated by the fact that for severely degraded documents a segmentation of words into characters will produce very poor results. The quality of the original documents did not allow us to recognize them with high accuracy - our goal here to produce transcriptions that allowed successful retrieval of images, which has been shown to be feasible even in such noisy environments.*

*Index Terms- Degraded images, DIBCO, de-noising, Wiener filter*

### I. INTRODUCTION

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. 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.[2] The Wiener filter is a linear filter for filtering images degraded by additive noise and blurring. Calculation of the Wiener filter requires the assumption that the signal and noise processes are second-order stationary. Wiener filters are often applied in the frequency domain. An image is often corrupted by noise in its acquisition and transmission. Image de-noising is used to remove the additive noise while retaining as possible as the important signal features [3]. In the recent years there has been a fair amount of research on wavelet thresholding and threshold selection for signal de-noising, because wavelet provides an appropriate basis for separating noisy signal from the image signal. The motivation is that as the wavelet transform is good at energy compaction, the small coefficient is more likely due to noise and large coefficient due to important signal features [4]. These small coefficients can be thresholded without affecting the significant features of the image.

#### *Degraded Images*

Degradations in document images result from poor quality of paper, the printing process, ink blot and fading, document aging, extraneous marks, noise from scanning, etc...Degraded document images are shown in fig (1, 2). Many sources of degradation are present in imaging systems. Some types affect only the gray levels at the individual picture points without introducing spatial blur. They are called point degradation. Other types which do involve blur are spatial degradations. Blurring is a form of bandwidth reduction of an ideal image caused by imperfect image formation process. Point and spatial degradation occur in a variety of applications e.g., in aerial mapping, astronomy, remote sensing, pictures obtained through atmosphere turbulence, aberration of optical systems relative motion between camera and object [5]. 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, Wiener filtering, etc.

Degraded images are: - Broken line structures, Broken symbols, text, Blurring of lines, symbols and text, Loss of complete objects,, Noise in homogeneous areas.[6]

der Natur, und diese sind die Ursanfänge der Dinge.  
Ihre Vereinigung ist der Grund aller Urstoffe, oder  
die Fähigkeit, die die Ursanfänge erlangen zur Bild-  
dung der Urstoffe in der Körperwelt.

Fig 1: Degraded document

*and then. Repairing is pretty good. We expect to build our wheels for next year. I think it will pay us, and give us employment during the winter.*  
*Yours aff. Orville.*

Fig 2: Degraded document

**Image Degradation Model**

For a linear invariant system, the observed/distorted image  $i(x,y)$  can be modeled as a convolution of the object function  $o(x,y)$ , which is the actual object in the scene, with the image degradation function  $h(x,y)$ , which is also commonly known as the point spread function.

$$I(x,y) = o(x,y) * h(x,y) + n(x,y)$$

where  $n(x,y)$  is an additive noise function that describes the random variation of the pixel intensity. According to the convolution theorem, a convolution of two spatial functions can be expressed as product of their respective Fourier transform in frequency domain. Thus, the image degradation model can be written as

$$I(u,v) = O(u,v) H(u,v) + N(u,v)$$

In a simplest image degradation model, the degradation function is modeled as a low pass filter, which resulted in a blurry effect. Fig. shows the block diagram of image degradation and restoration process.[7] Fundamentally, the image restoration process involves in reversing the distortion effects.

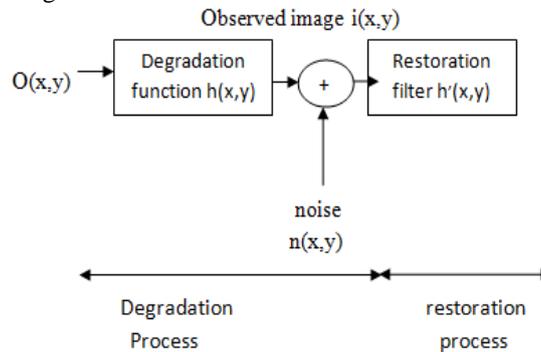
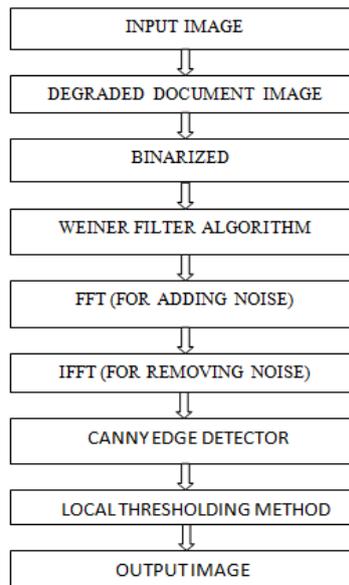


Fig. 3 Block diagram of image degradation and restoration process

The availability of such a uniform collection of documents for learning allows us to:

- To reduce the noise from the DIBCO dataset images by using wiener filter algorithm.
- To improve Peak Signal to Noise Ratio (PSNR).

**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.

## II. EVALUATION MEASURES

*Size of I/O image.* By calculating the row and column pixels, it is used to find the total size of original and restored image. *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.

$$PSNR=20 \log_{10}(255 / \sqrt{MSE})$$

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.[9]

## III. RESULTS AND DISCUSSION

Results includes following steps:

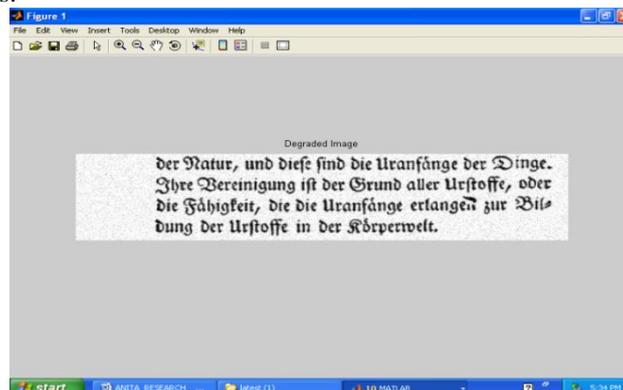


Fig 4: Degraded document

The input image for de-graded document is taken from IEEE data set. Wiener filter is applied on the input image to remove its blurry effects. The design of wiener filter task a different approach, one is assumed to have knowledge of the spectral properties of original signals and the noise and one seeks the linear time invariant filter whose output would comes as close to the original signal as possible. The output of the wiener filter is shown below:-

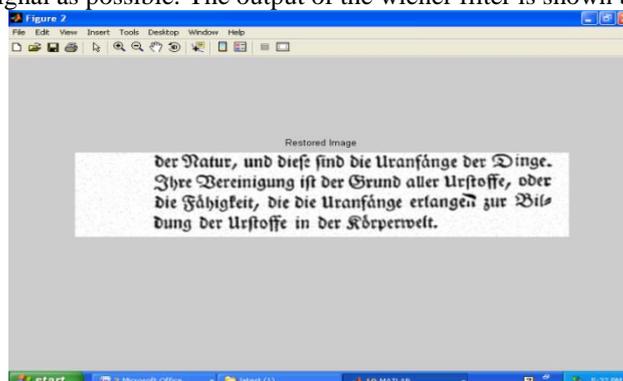


Fig 5: Restored image

Pre-processing methods that aim to suppress degradation using knowledge about its nature are called image restoration .The restored image after applying wiener filter is given above, wiener filter clears the input image by removing the homogenous noise. In the above figure graphical representation of input image is shown. A histogram is a display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size[10]. In the most common form of histogram, the independent variable is plotted along the horizontal axis and the dependent variable is plotted along the vertical axis.

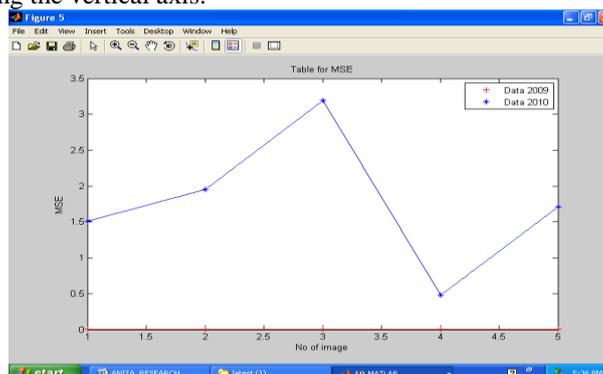


Fig 6: Graph for Mean Square Error

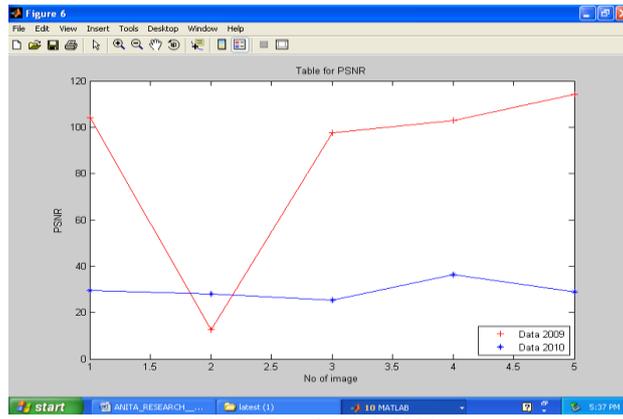


Fig 7: Graph for PSNR

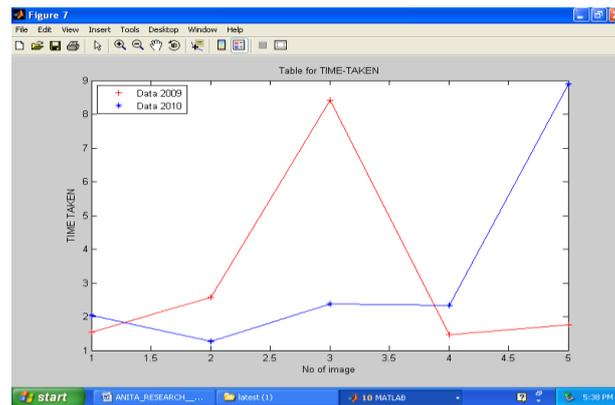


Fig 8: Graph for Time-Taken

Different graphs for various values calculated are shown above. MSE, PSNR and Time Taken are calculated and shown in the form of graphs

Table 1:-Evaluation Results of the Various Images of the Dibco2009

Sr.no	No. of input (units) images	No. of o/p images	Images sizes	MSE	PSNR	Time taken (sec)
1	615269	655479	660093	3.5	19.5	5.5
2	285279	314690	315462	8.2	15.6	2.6
3	211983	221508	221508	1.7	22.7	2.8
4	380057	398846	398940	2.1	21.9	3.3
5	106725	110600	110600	1.1	24.6	1.8

Table 2:-Evaluation Results of the Datasets

S.No	METHOD	PSNR DIBCO2009
1.	OTSU	15.34
2.	SAUV	16.39
3	NIBL	9.89
4.	BERN	8.89
5.	GATO	16.5

6.	LMM	18.5
7.	BE	18.6
8.	IBT	19.65
9.	PROPOSED	20.85

#### IV. CONCLUSION

This research work is based on removing noise from degraded images in both scanned and handwritten documents. All the images are taken from the DIBCO dataset. This proposed method includes deblurring or de noising of degraded documents. It is simple, robust and involves minimum parameter tuning. It has been tested on one public datasets that are used in recent document image binarization contest (DIBCO). This research develops a system which is used to clear the degraded documents. The PSNR is also improved to 20.85..as shown in the tables that has superior performance of our proposed method.

#### REFERENCES

- [1] Bolan Su,Chew Lim Tan ,“A Robust Document Image Binarization Technique for Degraded Document Images,”IEEE Transaction on image processing, vol.22,no.3,pp.1263-841, March 2013
- [2] J.Bharathi, Dr.P.Chandraseakar Reddy,”variation background modeling using grid point sampling for Degraded Image Binarization,” vol 2,pp2250-2459,nov 2012
- [3] Su, Bolan, Shijian Lu, and Chew Lim Tan. "A learning framework for degraded document image binarization using Markov random field." Pattern Recognition (ICPR), 21st International Conference on.IEEE, 2012.
- [4] H-DIBCO handwritten document image binarization competition,” International Conference on Frontiers in Handwriting-Recognition,pp.727-732, Nov 2010
- [5] Mohammed M. Siddeq and Dr. Sadar Pirkhider Yaba,” Using Discrete Wavelet Transform and Wiener filter for Image De-nosing” Wasit Journal for Science & Medicine no 2,pp 18 – 30, 2009
- [6] Huaigu Cao ,and Venu Govindaraju ," 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
- [7] B. Gatos, K. Ntirogiannis, and I. Pratikakis, “ICDAR 2009 document image binarization contest (DIBCO 2009),” International Conference on Document Analysis and Recognition, , pp. 1375–1382. July 2009
- [8] T. Lelore and F. Bouchara, “Document image binarisation using markov field model”, ,International Conference on Document Analysis and Recognition, pp. 551–555, July 2009.
- [9] E. Saund, J. Lin, and P. Sarkar, Pixlabeler: “user interface for pixel-level labeling of elements in document images”, International Conference on Document Analysis and Recognition,. pp. 646–650, July 2009
- [10] B. Gatos, I. Pratikakis and S.J. Perantonis, “Efficient Binarization of Historical and Degraded Document Images” The Eighth IAPR Workshop on Document Analysis Systems. 2008
- [11] Q. Chen, Q. Sun, H. Pheng Ann, D. Xia, “A double-threshold image binarization method based on edge detector,” Pattern Recognition, vol. 41, no. 4, pp. 1254–1267, 2008