RESEARCH ARTICLE OPEN ACCESS

A Review: Restoring the Images and Remove Blurry Effects from Scanned Document Images Using Various MATLAB Algorithm

Navdeep Kaur [1], Er.Divya Garg [2]
Research Scholar [1], Guide [2]
Universal College of Engineering &Technology
Ballopur

ABSTRACT

Image restoration involves elimination of noise. Filtering techniques were adopted so far to restore images since last five decades. In this research work, i consider the problem of image restoration degraded by a blur function and corrupted by random noise. A method for reducing additive noise in images by explicit analysis of local image statistics is introduced and compared to other noise reduction methods. In the proposed methods, which makes use of an a priori noise model, has been evaluated on various types of images. Inverse filtering and Wiener filter based algorithms and technique of image processing will be implementation for experiment. [1] [2] This article is aimed to provide a basic knowledge of image degradation and restoration process. 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. Inspired by results in cognitive psychology, researchers have begun to focus on holistic word recognition approaches. Here we present some techniques used in past for word recognition 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.[6] [11] The quality of the original documents does not allow us to recognize them with high accuracy - our goal here is to analyze the 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.

Keywordds:- Degraded images, DIBCO, de-noising, Wiener filter.

I. INTRODUCTION

Degradations 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.[6] 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.[3] 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 and canny filters etc.

II. TYPES OF DEGRADED MIAGES

- 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.

Image Restoration

Image Restoration is the operation of taking a corrupt/noisy image and estimating the clean, original image. Corruption may come in many forms such as motion blur, noise and camera mis-focus. Image restoration is performed by reversing the process that blurred the image and such is performed by imaging a point source and use the point source image, which is called the Point Spread Function (PSF) to restore the image information lost to the blurring process.

Image restoration is different from image enhancement in that the latter is designed to emphasize features of the image that make the image more pleasing to the observer, but not necessarily to produce realistic data from a scientific point of view. Image enhancement techniques (like contrast stretching or de-blurring by a nearest neighbor procedure) provided by imaging packages use no a priori model of the process that created the image.

With image enhancement noise can effectively be removed by sacrificing some resolution, but this is not acceptable in many applications. In a fluorescence microscope, resolution in the z-direction is bad as it is. More advanced image processing techniques must be applied to recover the object.

III. ALGORITHM

There are a number of algorithms used in image processing for adding and removing noise from images like photographs, hand-written images, scanned images etc.

I. Wiener filter

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 acquition and transmission. Image de-noising is used to remove the additive noise while retaining as possible as possible the important signal features. 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. These small coefficients can be threshloded without affecting the significant features of the image. [3]

Figure 1: (a) Real, Noisy blurred with var = 0.01(b) Wiener (c) Modified Wiener

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 [5]

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

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. Fundamentally, the image restoration process involves in reversing the distortion effects. [3]

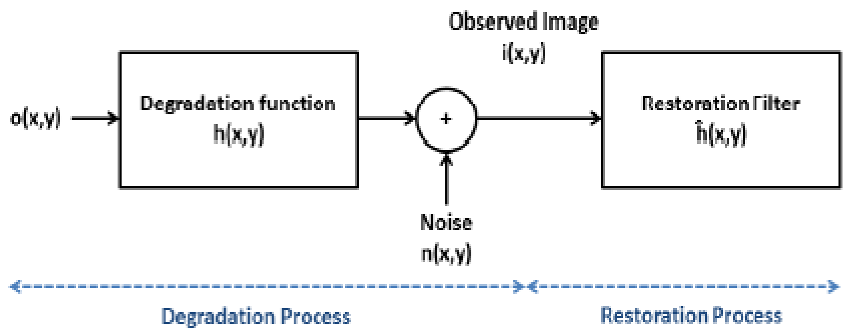
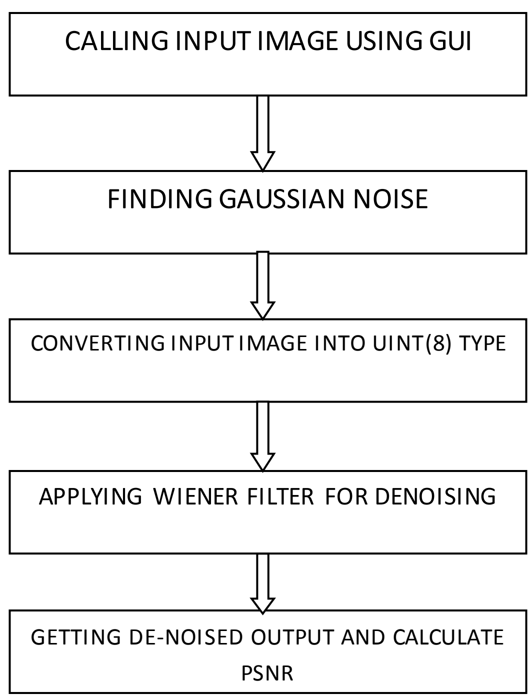


Figure 2: Block diagram of image degradation and restoration process

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).

III. 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.

Matching Gradient Distributions

Matching gradient distributions has been addressed in the texture synthesis literature. Heeger and Bergen synthesize textures by matching wavelet subband histograms to those of the desired texture. Portilla and Simoncelli match joint statistics of wavelet coefficients
to synthesize homogeneous textures. Kopf et al. introduce a nonhomogeneous texture synthesis technique by matching histograms of texels (or elements of textures). Matching gradient distributions in image restoration is not entirely new. Li and Adelson introduce a two-step image restoration algorithm that first reconstructs an image using an exemplar-based technique similar to Freeman et al. and warps the reconstructed image's gradient distribution to a reference gradient distribution using Heeger and Bergen's method.[4] A similarly motivated technique to ours is proposed by Woodford et al. They use a MAP estimation framework called a marginal probability field (MPF) that matches a histogram of low-level features, such as gradients or texels, for computer vision tasks, including denoising. While both Woodford et al.'s and our techniques use a global penalty term to fit the global distribution, MPF requires that one bins features to form a discrete histogram. This may lead to artifacts with small gradients. Our distribution matching method bypasses this binning process using parameterized continuous functions. Also, Woodford et al. use an image prior estimated from a database of images and use the same global prior torec onstruct images with different te xtures. In contrast, we estimate the image prior directly from the degraded image for each textured region. Schmidt et al. match the gradient distribution through sampling, which may be computationally expensive in practice. As with Woodford et al. Schmidt et al. also use a single global prior to reconstruct images with different textures, which causes noisy renditions in smooth regions. HaCohen et al. explicitly integrate texture synthesis to image restoration, specifically for an image up-sampling problem. [4] To restore textures, they segment a degraded image and replace each texture segment with textures in a database of images.

Figure 3: Input image and restored image

IV. OCR TECHNIQUE

The aim of OCR is to turn hard copy text into a soft copy version for use online. The question of acceptable error rates has not been studied much, but a few observations based on common sense are useful in lieu of genuine results: [13] [14]

1. Fixing OCR errors is expensive since a human operator must read every word in both the original and the OCR version. False positives (words incorrectly identified by OCR) are rare but common enough to cause a problem. False negatives (words readable by the operator, but not identified by OCR)

are much more common. The operator can also make errors in attempting the fix.

2. Errors that remain can cause frustration in users of the document, especially when the errors are in key words, or headings.

3. A technique, like that employed in Adobe Capture, where the image of the word is substituted, *in place*, when OCR fails, can alleviate some of this frustration. The downside to this is an inability to put these words (since they are images, not text) into indices or to use them in search.

4. The error rates for degraded documents can be so high as to make it increasingly unlikely that the end result is acceptable, even after manual editing.

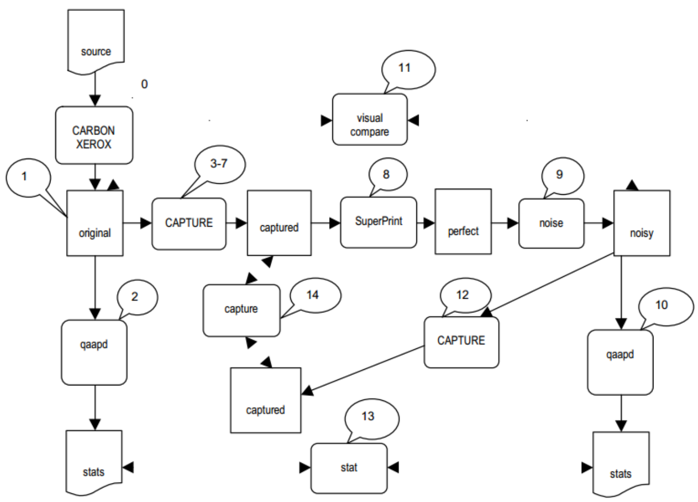


Figure 4: The Degradation Procedure

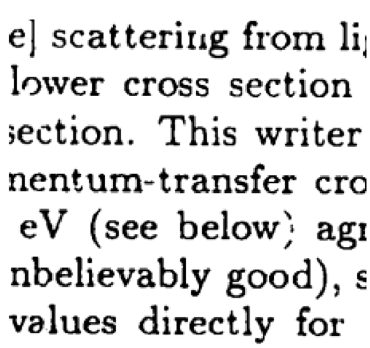

e] scattering from li
lower cross section
section. This writer
nentum-transfer cro
eV (see below) agı
nbelievably good), s
values directly for

Figure 5: A Section of an Original Document

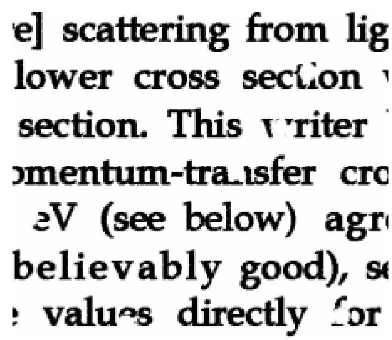

e] scattering from lig
lower cross section
section. This writer
omentum-transfer cro
eV (see below) agr
believably good), s
values directly for

Figure 6: Noisy equivalent of figure 5

VI. NIBLACK ALGORITHM

Niblack's algorithm determines a threshold value to each pixel-wise by sliding a rectangular window over the gray level image. The size of the rectangle window may differ. The threshold is calculated based on the local mean m and the standard deviation S of all the pixels in the window and is given by the following derivation.

$$T_{niblack} = m + K * S$$

$$= m + K\sqrt{\frac{1}{NP}\sum (p_i - m)^2}$$

$$= m + K\sqrt{\frac{\sum P_i^2}{NP} - m^2}$$

Where NP is the total number of pixels presents in the gray image, *T* represents the threshold value, *m* is the average value of the pixels *pi*, and *k* is fixed depends upon the noise still live on the background it may be -0.1 or -0.2.

VII. RELATED WORK

JAGADISH H. PUJAR et al (2010) introduced that in the past two decades, the technique of image processing has made its way into every aspect of today's tech-savvy society. Its applications encompass a wide variety of specialized disciplines including medical imaging, machine vision, remote sensing and astronomy. S.K. Satpathy et al (2010) analyzed that Image restoration involves elimination of noise. Filtering techniques were adopted so far to restore images since last five decades. In this paper, we consider the problem of image restoration degraded by a blur function and corrupted by random noise. Shenbagarajan Anantharajan et al (2012) proposed various types of noise models are subjected to an image and apply the nonlinear filter to reconstruct the original image from degraded image. Image restoration is a technique to attempt of reconstructs the original image by using a degraded phenomenon. In this paper the Lucy-Richardson filter is reconstruct the degraded image which closely resembles the original image. Er. Jyoti Rani et al (2014) given brief introduction of digital image processing is described. Mohini Sharma et al (2014) introduced that Image restoration is a field of image processing which deals with restoring an image that has been degraded by some degradation phenomenon. Degradation may occur due to motion blur, Gaussian blur, noise or camera mismatch. Dr.Salem Saleh Al-amri et al (2014) attempts to undertake the study of Restored Motion Blurred Images. by using four types of techniques of deblurring image as Wiener filter, Regularized filter ,Lucy Richardson deconvlution algorithm and Blind deconvlution algorithm with an information of the Point Spread Function (PSF) corrupted blurred image with Different values of Length and Theta and then corrupted by Gaussian Blurred. A.M.Raid et al (2014) concluded that Image processing including noise suppression, feature extraction, edge detection, image segmentation, shape recognition, texture analysis, image restoration and reconstruction, image compression etc uses mathematical morphology which is a method of nonlinear filters.It is modulated from traditional morphology to order morphology, soft mathematical morphology and fuzzy soft mathematical morphology.

VIII. CONCLUSION

In this article, a number of algorithms developed by researchers are explained. Performance evaluation and practical implementation shows that all the algorithms can be used for the degradation. Wiener filter is the best filter among all the filters for documents degradation. The wiener filter is also used for de-noising of the de graded documents.

REFERENCES

[1] JAGADISH H. PUJAR, 2KIRAN S. KUNNUR, "A NOVEL APPROACH FOR IMAGE RESTORATION VIA NEAREST NEIGHBOUR METHOD" in Journal of Theoretical and Applied Information Technology in 2010.

[2] S.K. Satpathy, S.K. Nayak, K. K. Nagwanshi, S. Panda, C. Ardil, "An Adaptive Model for Blind Image Restoration using Bayesian Approach" International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol:4, No:1 in 2010.

[3] Shenbagarajan Anantharajan1, A. Siva Ganesh2, V. Rajesh Kumar3 and C. Balasubramanian, "IMAGE RESTORATION: DESIGN OF NON-LINEAR FILTER (LR)" in ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING,in NOVEMBER 2012.

[4] Er. Jyoti Rani, Er. Sarabjeet Kaur, "Image Restoration Using Various Methods and Performance Using Various Parameters" in International Journal of Advanced Research in
Computer Science and Software Engineering in 2014.

[5] Mohini Sharma1, Prof. Shilpa Datar, "Image Restoration Using Wavelet based Image Fusion" in International Journal of Engineering Trends and Technology (IJETT) – Volume 15 Number 1 – Sep 2014.

Dr.Salem Saleh Al-amri1 and Dr.Ali Salem Ali, "Restoration and Deblured Motion Blurred Images" in International Journal of

Computer Science Issues, Vol. 11, Issue 1, No 1, January 2014.

[6] A.M.Raid1, W.M.Khedr2, M.A.El-dosuky1 and Mona Aoud, "IMAGE RESTORATION BASED ON MORPHOLOGICAL OPERATIONS" in International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 4, No.3, June 2014.

[7] Aziz Makandar and Anita Patrot, "Computation Pre-Processing Techniques for Image Restoration" in International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 4, March 2015.

[8] Gabriel Scarmana, " DIGITAL IMAGE DEBLURRING DE-DEBLURRING PROCESS FOR IMPROVED STORAGE AND TRANSMISSION " in IEEE in 2015.