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 The restoration of handwritten degraded historical documents plays a major role in degradation process.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.

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 reduce the noise from the scanned degraded 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:

\[ \text{MSE} = \frac{\text{Mean Square Error}}{\text{Size Of Image}} \]

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.

\[ \text{PSNR} = 20 \log_{10} \left( \frac{255}{\text{MSE}} \right) \]

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 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:
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 is given below.

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

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.

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

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

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


