I J C T A, 9(34) 2016, pp. 71-76

Remove Noise and Reduce Blur Effect of Scanned Handwritten Degraded Document Image Using Matlab

Anjali[1] and Bhupinder Kaur[2]

ABSTRACT

The segmentation of the text from badly degraded document images is a very challenging task due to the high inter/ intra variation between the document background and the document foreground text of different document images. The poor quality of original documents does not allows us to recognize them with the higher accuracy. Most offline handwriting recognition approaches proceed by segmenting the characters into the smaller pieces which are recognized separately. The restoration of handwritten degraded historical documents plays a important roles in the degradation process. Old document images often suffer from the degradation that render their binarization such a challenging task. Image restoration is a process which is used to restore the badly degraded documents back to the original image. The algorithms used in past few years were Ni-blacks algorithm, the canny edge map algorithm and the OTSU method, Back ground estimation, dynamic thresholding and image binarization etc. The performance of all the algorithms was measured by the Peak Signal To Noise Ratio. The Peak Signal to Noise Ratio is calculated by calculating Mean Square Error first. According to previous results Peak Signal to Noise Ratio calculated by the different approaches is, PSNR by OTSU method was 15.34, by Canny edge map was 19.65, by back-ground estimation was 18.6 and by dynamic thresholding was 16.5. PSNR should be maximum for the better results. The results are obtained by the previously implemented algorithms were not sufficient. So, in this research work Ni-Black's Algorithm will be implemented.

Index Terms: Binarization, Image Restoration Niblack's Algorithm, and Thresholding.

BINARIZATION

The binarization methods are applied to the document images for discriminating the text from the background based on a pure thresholding and filtering combined with image processing algorithms. Binarization refers to the conversion of a grayscale image to binary. Binarization (thresholding) converts the grayscale document image to binary, by changing the foreground pixels (text characters)in to black and background pixels into white. A binarization method of the binarizing an image by extracting the lightness (brightness, density) as a feature amount from the image. When a pixel is selected an image, the sensitivity is added to and/or subtracted from the value concerning the Y values of the selected pixels to set a threshold value range. Next, when a another pixel is selected, the sensitivity is added or subtracted from the value concerning the Y value of the selected pixels and a new threshold value range is set containing the calculation results and the already setup the threshold value range. The pixels of the value concerning the Y value of any pixels in the images within the threshold value range is extracted from the same brightness as the selected pixel and the extraction result is displayed.

IMAGE RESTORATION

Image restoration is an process of taking corrupted/noisy image and producing clean original image. The corruption may come in many form like camera misfocus, motion blur, noise. Image degradation occur

[1] (ME(CSE) Chandigarh University, Gharuan, Mohali
[2] (Associate Professor, CSE)

when image undergoes loss of the stored information either due to the digitization or conversion decreasing visual quality. Most of the image restoration methods are based on the convolution applied globally to the whole image. Degradation of images have many causes: defects of the optical lenses, non-linearity of the electro-optical sensor, graininess of the film material, relative motion between of an object and camera, wrong focus, atmospheric turbulence in remote sensing or astronomy, scanning of photographs, etc..

Image restoration techniques can be classified into two groups: deterministic and stochastic. Deterministic methods are applicable to the images with little noise and a known degradation function. The original image is obtained from the degraded by a transformation inverse to the degradation. Stochastic techniques try to find the best restoration according to the particular stochastic criterion, e.g. a least-squares method. In some of the cases the degradation transformation must be estimated firstly. It is advantageous to know the degradation function explicitly. This knowledge provides, the better results of the restoration. There are three typical degradations with a simple function: the relative constants speed movement of the objects with respect to the camera, wrong lens focus, and atmospheric turbulence. In most of the practical cases, there is insufficient knowledge about the degradation, and it must be estimated and modeled.

The estimation is classified into two groups according to the information available: a prior and a posterior. If degradation type or parameters need to be estimated, this step is the most crucial one, being responsible for the image restoration success or for the failure. It is also the most difficult part of image restoration. A prior knowledge about the degradation which is either known in advance or can be obtained before restoration. For example, if it is known in advance that the images was degraded by relative motion of an objects with respect to the sensor and then the modeling determines only the speed and the direction of the motion. The objective of image restoration techniques is to reduce the noise and recover the resolution loss. Image processing techniques are performed either in the images domain or the frequency domain.

NI-BLACK ALGORITHM

The Ni-Black's algorithm calculation involves a pixel wise threshold by sliding a rectangular window over the grey level image.The Ni-Black algorithm is a local thresholding algorithm that adapts the threshold according to the local mean and the local standard deviation over the specific window size around each pixel location. The local threshold at any pixel (i, j) is calculated as:

$$T(i,j)\ \ m(i,j)+k.(i,j)$$

Where m (i,j> and (i , j) are the local sample mean and the variance, respectively. Size of the local region (window) is dependent upon the application. The value of the weight 'k' is used to control and adjust the effect of the standard deviation due to objects features.The Ni-Black algorithm suggests the value of 'k' to be -0.2.The document images binarized using the Ni-Black algorithm provides the most satisfactory results, rather than all the other algorithms available at the time of writing.However, the Ni-Black's algorithm suffers from the basic problems of local thresholding, i.e. providing the unnecessary details in the binarized images that may not be required in the processing. The Ni-Black fails to adapt the large variation in illumination, especially in the document images.

A local region analysis using Ni-Black does not provide any kind of information about the global attributes of the image that may be helpful in the binarization process of badly illuminated images. So, the gray level variations in the document image make it impossible to the adapt threshold as will be shown in the results. Another problem it Faces is the optimum selection of the weight k. Ni-Black algorithm Uses fix value of this weight. The fix given value of 'k' may work for the document images but for the gray-level images with a lot of variations of gray values and the value of the weight should not be fixed but to change from images to images depending upon their gray-level distributions, therefore, the value of 'k' should be calculated at the run-time.

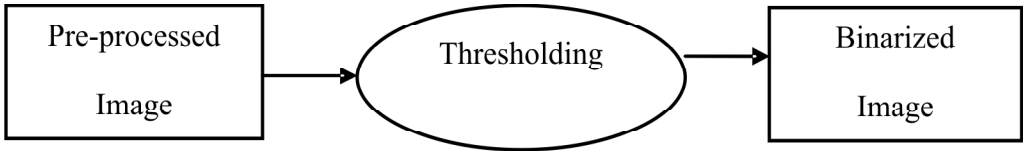

Figure 1: Block diagram of Thresholding

PROBLEM FORMULATION

In degraded images the problems like broken line structure, unwanted noise on background and broken text were found. Different types of noises that occur in different types of images are (1) Marginal noise, appears as dark regions around the margin of the scanned documents. (2) Edge Noise, causes the spreading of the edges and thus changing the edge width, (3) Background noise, appears as additive components in the background, (4) Circular noise, is caused by scanning documents with punched holes, (5) Stroke-like Pattern Noise, (6) Clutter Noise, (7) Broken line structures, in which the gaps of all the sizes in lines were roughly counted and large gaps were considered worse than small. There are many causes for degradation of documents and also a number of problems in their de-noising. Some noise particles were also available in the output images. The main limitation of existing work is found to be noisy and low intensity images. To overcome these problems Ni-black algorithms can be used. The Peak signal to noise ratio and Mean square errors can be improved to check the quality of work done by Ni-Black algorithm.

OBJECTIVE FOR STUDY

- To implement Ni-Black's algorithm.
- Calculate size of input/output images
- Calculate PSNR (Peak Signal to Noise Ratio).
- Calculate MSE (Mean Square Error).

IMPLEMENTED ALGORITHM

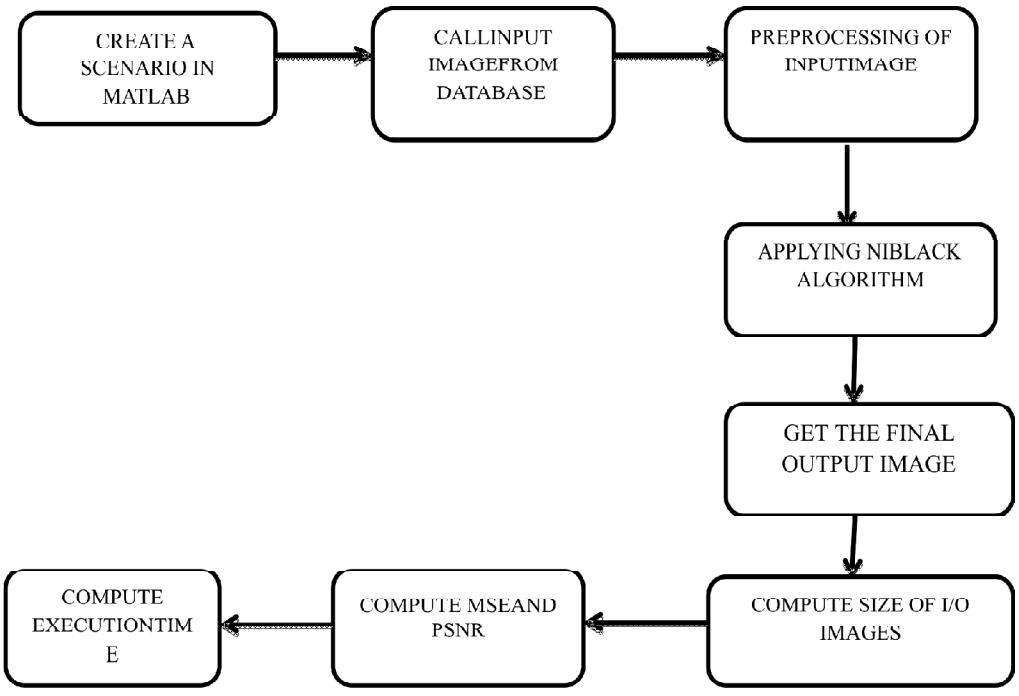

Figure 2: Flow-Chart

RESULTS AND DISCUSSION

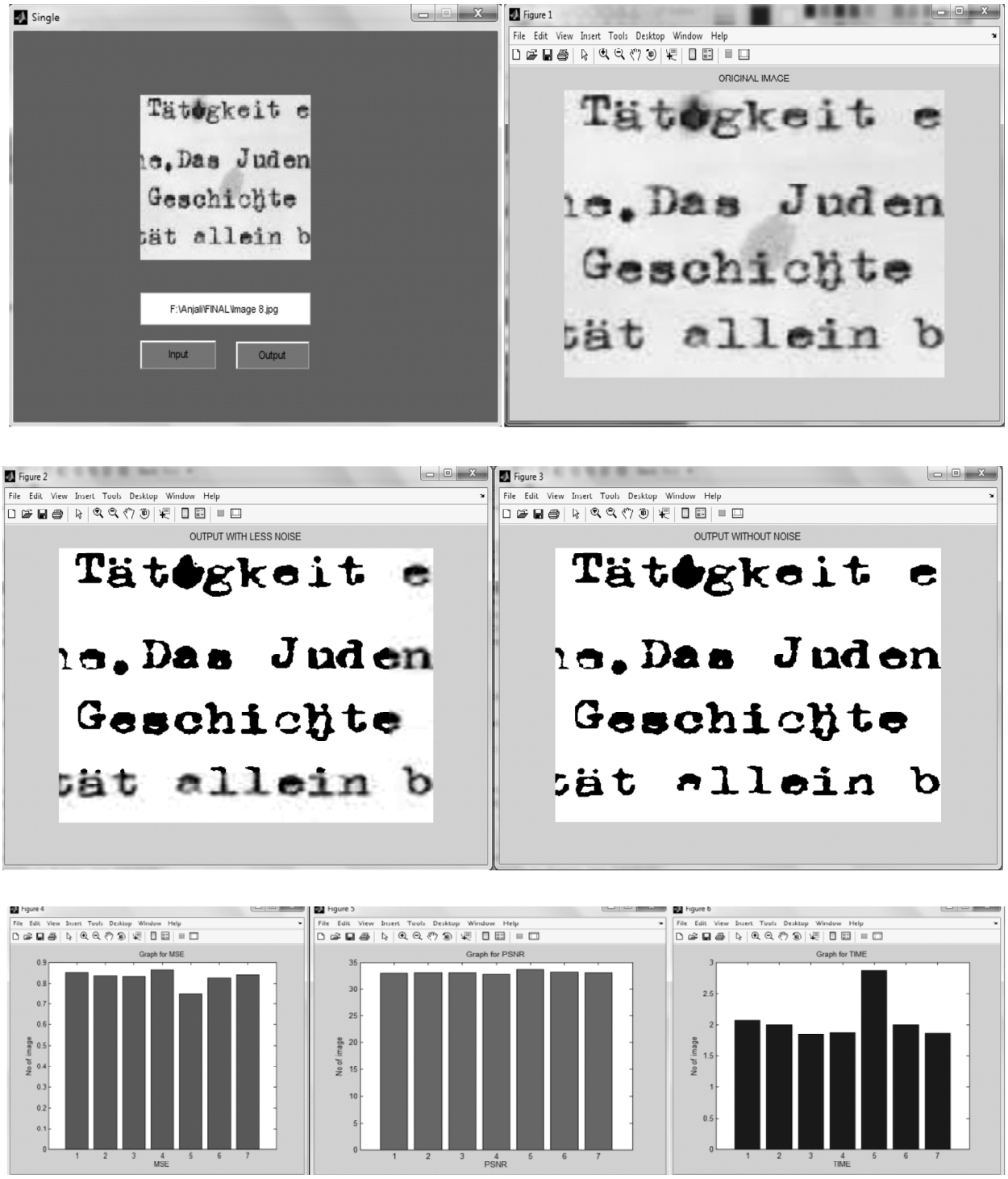

Figure 3: Outputs from the Code

The figure 3 shows all the outputs from the code. First figure shows the GUI developed and second figure shows the input image. When push button 1 will be pressed than the input image will be called. The second push button is for generating outputs. The next two figures show the output with noise and final output without noise respectively. Last three figures show the Peak signal to noise ratio, Mean square error and Elapsed time. The parameters are calculated and shown in the table.

Table

Sr.No.	*Input Image*	*No. of Pixels of Input Image*	*No.of Pixels of Output Image*	*MSE*	*PSNR*	*Elapsed Time*
1.	Image 1	1401797	66672	0.8156	33.2675	2.766598
2.	Image 2	1402500	53207	0.8330	33.1452	2.906401
3.	Image 3	1402500	56916	0.8284	33.1771	3.080248
4.	Image 4	1398283	29536	0.8572	32.9794	3.206844
5.	Image 5	1402454	41317	0.8477	33.0440	3.185473
6.	Image 6	1402500	30236	0.8616	32.9497	6.137955
7.	Image 7	1402500	27118	0.8655	32.9234	3.007985
8.	Image 8	1402496	122945	0.7491	33.7598	2.863961
9.	Image 9	1402500	50650	0.8362	33.1233	2.963282
10.	Image 10	1402500	59484	0.8253	33.1992	2.966103

CONCLUSION AND FUTURE WORK

This research work for developing an image technique that will become the efficient for clearing the degraded images, blur images and noisy images. In this research Ni-Black algorithm has been applied to the degraded images and the collected output is with less noise. All coding part is written in MATLAB, the code is written and tested on a number of images from different DIBCO datasets. This research work develop a system basically which is used to clear the degraded documents. Parameters like Peak Signal to Noise Ratio, Image size, MSE etc. are calculated to show the improvement for our work. In future some other algorithm can be implemented to perform the same operation with adjoining some denoising technique for proving more improvement in Mean Square Error (MSE) and Peak Signal To Noise Ratio (PSNR). The design of GUI can also be improved by adding more options in GUI.Same work out on the implementation time can also be done.

REFERENCES

[1] B. Su, S. Lu, and C. L. Tan, “Robust document image binarization technique fordegraded document images”, IEEE, vol. 22, no. 4, pp. 1408–1417, 2013.

[2] Galatsanos, N.P., Mesarovic, V.Z., Molina, R. andKatsaggelos, A.K., “Hierarchical Bayesian Image Restoration from Partially Known Blurs”, IEEE , vol 10, pp.1784-1797,2000.

[3] Dawoud, A. and Kamel, M.S., “Iterative Multimodel Sub-image Binarization for Handwritten Character Segmentation”, IEEE ,vol 13, 9, pp.1223—1230,2004.

[4] K. Ntirogiannis, B. Gatos, and I. Pratikakis, “Performance evaluation methodology for historical document image binarization”, IEEE, vol. 22, no. 2, pp. 595—609, 2013.

[5] K. M. Amin, M. AbdElfattah, A. E. Hassanien, and G. Schaefer, “A binarization algorithm for historical arabic manuscript images using a neutrosophicapproach”,IEEE, vol. 6, no. 14, pp. 4799–6594, 2014.

[6] Z. Hadjadj, A. Meziane, M. Cheriet, and Y. Cherfa, “An active contour based method for image binarization”, IEEE, vol. 2, no. 14, pp. 4799–7508, 2014.

[7] Feng, M. L. and Tan, Y. P., “Adaptive Binarization Method For Document Image Analysis”, IEEE , pp.339-342,2004.

[8] Gatos, B., Pratikakis, I. and Perantonis, S.J., “Efficient Binarization of Historical and Degraded Document Images”, IEEE, pp. 447-454,2008.

[9] Gangamma, B., Murthy K. S., Chandra, G. C. P., Kaushik, S. and Kumar, S., “A Combined Approach for Degraded Historical Documents De-noising Using Curvelet and Mathematical Morphology “, IEEE, pp. 1-6,2010.

[10] Hoang, T. V., Smith, E. H. and Tabbone, S. , “Edge Noise Removal in Bievel Graphical Document Images Using Sparse Representation”, IEEE , pp.3549—3552,2011.

[11] Bataineh, B. S., Abdullah, N. H. S. and Omar, N. H. S. “An Adaptive local binarization method for document images based on a novel thresholding method and dynamic windows,”, International Journal of Computer Engineering and Technology (IJCET), vol 32,14, pp. 1805-1813, 2011.

[12] M. Paul, “Effective process to remove broken lines effect from degraded document images using matlab algorithm”, International Journal of Computer ScienceTrends and Technology, vol. 3, no. 3, pp. 2347–8578, 2015.

[13] N. Kundal and Anantdeep, “Performance evaluation of novel historical documentsrestoration algorithm”, International Journal of Computer Science EngineeringTechnology, vol. 5, no. 7, pp. 278–282, 2015.

[14] S. Saini and R. Dogra, “A novel in to approach for image binarization for degradeddocument images”, Global Journal of Computers Technology, vol. 3, no. 2, pp. 2394–501, 2015.

[15] S. Jain and P. Parihar, “Remove noise and reduce blurry effect from degradeddocument images using matlab algorithm”, International Journal of Engineering Research and General Science, vol. 3, no. 1, pp. 2091–2730, 2015.