A Review on the Ancient Documents Restoration Methods

[1]Kirandeep, [2]Dr. Harish Kundra
[1,2]Dept. of Computer Science, RIEIT, Ropar, India

Abstract
With the advancement in internet and multimedia technologies a huge amount of multimedia data in the form of documents and images has been used in many areas like medical fields satellite data digital forensics surveillance systems etc. The documents and images of ancient world are available on web media but these are of no use due to lower & degraded quality of images. There is the need to improve the quality of these existing documents to make it use in different fields of medical, science, satellite data etc. Ancient documents may suffer from progressive and various types of degradations. Binarization of these degraded documents is an important step in Document Analysis and Retrieval systems (DAR), because of its impact on the performance of the subsequent applications: OCR (Optical Character Recognition), word-spotting and information retrieval. Document binarization is noise sensitive particularly when the degradations are very complex. There exist various methods and filters to restore these documents and images. In this paper, we are presenting a review on the existing methods of ancient degraded document/images restoration.

Keywords
Image Restoration, Ancient Document, Degraded Images, Binarisation.

I. Introduction

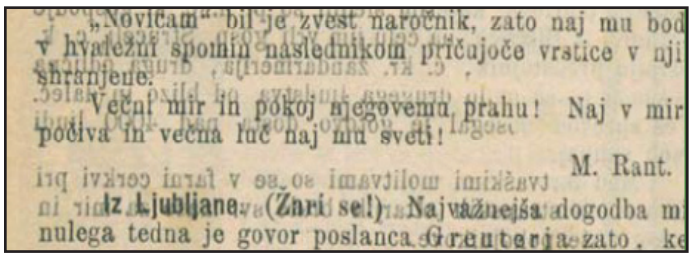
The growing importance of digital image processing stems from two principal application areas and they are (i) Improvement of pictorial information for human interpretation and (ii) Processing of scene data for autonomous machine perception. Digital image processing techniques are now used to solve a variety of problems.

One such important problem in image processing is restoration. The goal of the restoration approach is to improve the given image, so that it is suitable for further processing. Restoration is a technique used to reconstruct or recover an image that has been degraded by using a priori knowledge of the degradation phenomenon.

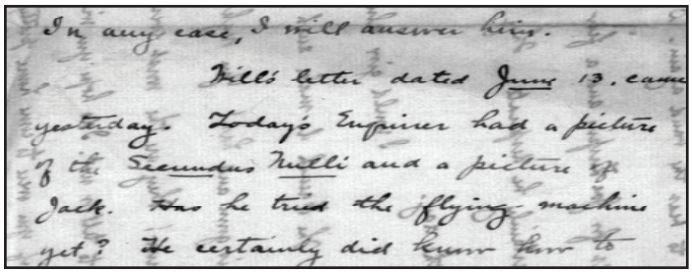
Removal of degradation is becoming increasingly important as image analysis and acquisition system finds more application in society. There are a variety of reasons that could cause degradation of an image, and image restoration is one of the key fields in today's Digital Image Processing due to its wide area of applications. Because of the imperfection of the physical imaging system and due to various physical limitations on every application a recorded image will always be a degraded version of an original image. There exists a large number of possible degradations that an image can suffer [1]. Common degradations are blurring, motion and noise. Blurring can be caused when an object in the image is outside the cameras depth of field during the exposure. Motion blur can be caused when an object moves relative to the camera during an exposure. Photographic defocusing is also a problem in many different imaging situations. This type of blurring is due to effects at the camera aperture, which spreads a point of incoming light across a circle of confusion. Noise is generally a distortion due to the imaging system rather than the scene recorded. Noise results in random variations to pixels in the image.

Whatever be the degrading process, image distortions can fall into two categories, spatially invariant and spatially variant. In a space invariant distortion, all pixels have suffered the same form of distortion. This is generally due to the problems with the imaging system such as distortions in the optical system, global lack of focus or camera motion. In a space variant distortion, the degradation suffered by a pixel in the image depends upon its location in the image. This can be caused by internal factors such as distortions in the optical system or by external factors such as object, motion etc.

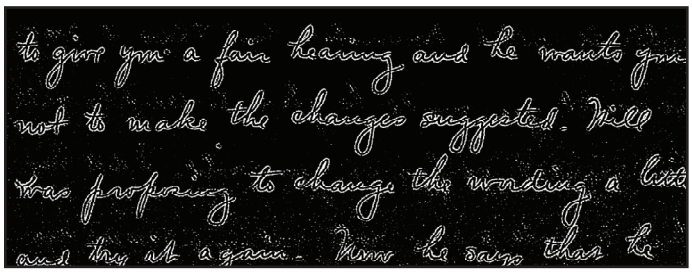
In this paper, we are reviewing on the existing methods for image restoration of ancient document. Documents can be a valuable source of information but often they suffer degradation problems, especially in the case of historical documents, such as strains, background of big variations and uneven illumination, ink seepage, etc. The term of 'ancient document' is document coming from hundreds years ago [2]. Fig. 1(a)-(c) shows some of the ancient documents. The purpose of the study is to preserve the information contained in old documents into a digital form, because the process to rescue the old images using physical approach is too slow. For ancient documents, major degradation factor is noise. In this paper, we have discussed the existing methods for image degradation for the ancient documents.



(a)



(b)



(c)

Fig. 1(a) - (c): Ancient Degraded Documents

Rest of the paper is organized in the following manner. Section II presents the image degradation model. Section III presents the

existing methods of ancient document degradation. Section IV brief out the research gap and Section V concludes the paper.

Image Restoration Model

In digital image processing, the following linear model can accurately represent the degradation of the true image.

$$g(x,y) = f(x,y) * h(x,y) + n(x,y) \quad (1)$$

where f(x,y) is the original image, and g(x,y) is the degraded image which is acquired by the imaging system. In this formulation n(x,y) is the additive noise introduced by the system. Noise term is important because in practical imaging situations, additive noise is not negligible. The function h(x,y) represents the degradation phenomenon and it is usually a Point Spread Function (PSF). The output of restoration filter $\hat{f}(x,y)$ is approximately equal to f(x,y), the original image.

Image restoration is a vital part of many image processing applications. The purpose of image restoration is to recover the original image from a degraded observation. In many instances, the degraded observation g(x,y) can be modeled as the two-dimensional convolution operation of the true image f(x,y) and the PSF (point spread function) h(x,y) of a linear shift invariant system plus an additive noise term n(x,y). A model of the image degradation / restoration process is given in fig. 2.

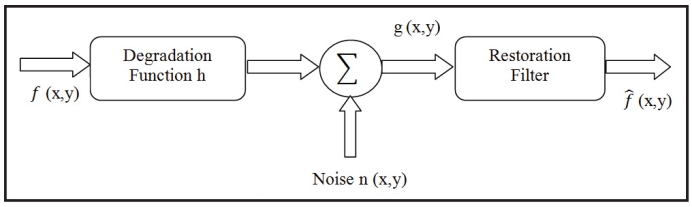


Fig. 2: Image Degradation/Restoration Process

III. Existing Work

Kale et al. [3] proposed a hybrid binarization approach to improve the quality of old historical manuscripts by removing the noise and to adjust the background contrast of the manuscript. A hybrid approach of local and global thresholding is taken into account. Initially the global thresholding technique is applied to denoise the background of manuscript and then local thresholding is applied to each individual portion of the image. The main advantage of the applying the global thresholding technique is that it reduces the computational cost and time to denoise the image as compare to applying on each individual portion of the image. The results obtained from the proposed hybrid concept shows the higher accuracy of the results as deal with the different type of images having background noise. Also, this algorithm does not affect image quality if the image is already in a good condition.

Singh and Sharma [4] have given an overview of the existing image enhancement techniques. They have also tried to find the limitations of the existing image enhancement techniques and suggested some corresponding solutions. They found that in most of the image enhancement techniques image details is ruined after the implementation of the enhancement techniques and at the same time it is not possible to enhance the quality of the entire image. So, a better solution is to use the non-linear image enhancement techniques to remove the blurriness of the image.

Coustaty et al. [5] proposes a new pre-processing that permits to denoise the documents those are degraded by the time and require to be preprocessed before being submitted to image indexing strategy, by using a Aujol and Chambolle algorithm. Aujol and Chambolle algorithm allows to extract meaningful components from image. This process relies on a Aujol and Chambolle algorithm which allow to obtain three layers of information: (1) regularized one which contain shapes, (2) textures one which contain all the oscillating elements and (3) noise one which correspond to the original image less the two first layers. From this decomposition, one can apply specific treatments on one or many layers.Some examples of specific processing applied on each layer are illustrated in this paper.

Cheng et al. [6] has discussed a novel approach for the detection of over- enhancement. Firstly the causes for generating over - enhancement has been analyzed in detail, then an objective and effective criterion has been presented. The experimental results demonstrate that the given approach can locate the over enhanced areas accurately and effectively, and provide a quantitative criterion to assess the over-improvement levels well. The given method will be useful for vigorously monitoring the quality of the improved image, and optimizing the parameter settings of the contrast improvement algorithms.

Rajib et al. [7] has proposed a contrast enhancement technique using scaling of internal noise of a dark image in discrete cosine transform domain. The mechanism of improvement is attributed to noise-induced transition of discrete cosine transform coefficients from a poor state to an improved state. This transition is effected by the inner noise present due to lack of adequate illumination and can be modeled by a general bi-stable system exhibiting dynamic stochastic resonance. The given approach has adopted a limited adaptive processing and significantly improves the image contrast and color information while ascertaining good perceptual quality. When compared with the present improvement approaches such as adaptive histogram equalization, gamma correction etc. the given approach has shown extraordinary performance in terms of relative contrast enhancement, colorfulness and visual quality of improved image.

Cong-ping et al. [8] proposed a new Adaptive Weight Algorithm for Salt and Pepper Noise Removal that consists of two major steps, (i) to detect noise pixels according to the correlations between image pixels, (ii) use adaptive methods based on the various noise levels. For the low noise level, neighbourhood signal pixels mean method is adopted to remove the noise and for the high noise level, an adaptive weight algorithm is used.

Khan et al. [9] presented an efficient Image Noise Removal and Enhancement Method. It presents a new method to remove noise and enhance the image with the help of partial unsharp masking and conservative smoothing. In this method, unsharp masking is applied in partial way for detection of the edges and boundary lines in the image and then a conservative smoothing operation is applied on the selected areas to remove undesirable edges which represents the salt and pepper noise.

Deborah and Arymurthy [10] present the genetic algorithm for the enhancement and restoration of old documents. The experimental results obtained from the performance of genetic algorithm are also compared with and without the use of median filter for the pre-processing of the input image. The results obtained using the genetic algorithm are of 92.9% accurate retrieval of data which was superlative as compare to results of median filter (59.5% data success rate). In this experiment, fitness function for genetic algorithm is used and found to have greater accuracy of the data with number of edges containing only 1 pixel.

Moghaddam & Cheriet [11] proposed an image restoration method to enhance the document quality by taking the virtual diffusion processes and degradation model as a base. The promising results are obtained from the evaluation of both the restoration

and degradation models. In this paper, image restoration and enhancement problems are taken as a diffusion point of view. This model can also be used for the addition defects like ink seepage and aging. It also enhances the quality of double sided documents. This model can also be hybridised with the other well developed model which can simulate the defect originating from imaging processes.

III. Research Gap

One of the major drawbacks of most of the image restoration algorithms discussed above is the computational complexity. The computational complexity is so high that many simplifying assumptions have been made to obtain computationally feasible algorithms. The assumptions that are used in the restoration process are (i) Wide Sense Stationary (WSS) and (ii) availability of second order image statistics and these in turn compromise the accuracy of the restored images. So, it is desirable to develop a restoration algorithm that do not make use of WSS assumption and can be implemented in a reasonable time.

IV. Conclusion

Digital images can be corrupted by noise during the process of acquisition and transmission degrading their quality. Degradation problems are quiet common in these documents. Few factors that impede (in many cases may disable) the legibility of the documents are strains, big variations, uneven illumination, presence of smear, seepage of ink, etc. Specialized processing is required to these document images for removing background noise in order to become more legible. So some image denoising techniques must be needed to restore the quality of image. A major challenge for a denoising algorithm is to improve the visual appearance of an image while preserving relevant features such as edges during the denoising process. In this paper, we have presented the existing methods for digital image denoising. As per the research gap in these existing methods, there is the need of some more efficient approach to enhance the document quality.

References

[1] Banham, Mark R., Aggelos K. Katsaggelos,"Digital image restoration", Signal Processing Magazine, IEEE 14, No. 2, pp. 24-41, 1997.

[2] Moghaddam, Reza Farrahi, Mohamed Cheriet,"A variational approach to degraded document enhancement", Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, No. 8, pp. 1347-1361, 2010.

[3] Kale, P., Phade, G. M., Gandhe, S. T., Dhulekar, P. A., "Enhancement of old images and documents by digital image processing techniques", In Communication, Information & Computing Technology (ICCICT), 2015 International Conference on, pp. 1-5, IEEE, 2015.

[4] Singh, S., Sharma, S.,"A Survey of Image Enhancement Techniques, IJCS, pp-1-5, 2014.

[5] Coustaty, M., Dubois, S., Menard, M., Ogier, J. M., "Ancient documents denoising and decomposition using aujol and chambolle algorithm", In Graphics Recognition. New Trends and Challenges (pp. 15-24). Springer Berlin Heidelberg, 2013.

[6] Cheng, H. D., Yingtao Zhang,"Detecting of contrast over-enhancement", In Image Processing (ICIP), 2012 19th IEEE International Conference on, pp. 961-964. IEEE, 2012.

[7] Jha, Rakesh Kumar, Rajlaxmi Chouhan, Prabir Kumar Biswas, Kiyoharu Aizawa,"Internal noise-induced contrast enhancement of dark images", InImage Processing (ICIP), 2012 19th IEEE International Conference on, pp. 973-976. IEEE, 2012.

[8] Cong-ping, Chen, Wang Jian, Qin Wu, Dong Xiao-gang,"A new adaptive weight algorithm for salt and pepper noise removal", In Consumer Electronics, Communications and Networks (CECNet), 2011 International Conference on, pp. 26-29. IEEE, 2011.

[9] Khan, Nafis Uddin, K. V. Arya, Manisha Pattanaik, "An efficient image noise removal and enhancement method", In Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, pp. 3735-3740. IEEE, 2010.

[10] Deborah, H., Arymurthy, A. M.,"Image enhancement and image restoration for old document image using genetic algorithm. In Advances in Computing, Control and Telecommunication Technologies (ACT)", 2010 Second International Conference on, pp. 108-112, IEEE, 2010.

[11] Moghaddam, R. F., Cheriet, M.,"Low quality document image modeling and enhancement", International Journal of Document Analysis and Recognition (IJDAR), 11(4), pp. 183-201, 2009.