(REVIEW ARTICLE)

Gaussian filter and fuzzy filter based historical document with geodesic morphology

Rinku Kalotra * and Kaushal Kumar

Department of Computer, Applications, Global Group of Institutes, Punjab Technical University, India.

Abstract

The technique of image restoration involves recovering an image from a damaged state, typically a distorted and noisy image. In image processing image restoration is a core issues. This paper based on historical documents and these are discovered by archaeologist. Important information about the tradition, religion science, literature. In this Paper, we will use the geodesic morphological model with the fuzzy filter for the purpose of historical document restoration and compare them on the basis of various performance parameters.

Keywords: Historical documents; Restoration; Geodesic morphological; Gussian Filter; Fuzzy Filter

1. Introduction

1.1. Image restoration

Image restoration is the technique to restore corrupted images to original images which is not degraded. Corruption may be of many forms like camera misfocus and produces many kinds of noise in an image.

Image degradation occur when image is not clear and goes loss of stored information either due to digitization or conversion decreasing visual quality.

1.2. Degradation model

In degradation model, the image is blurred using degradation function and additive noise. The following Figure 1 represents the structure of degradation model.

1.3. Degradation model

The degraded image is described by the following equation:

$$g(x, y) = f(x, y) * h(x, y) + \eta(x, y)$$

In equation (1), we have "g" as the degraded image, "h" as the degradation function, "f" as the original image and "n" is the additive noise.
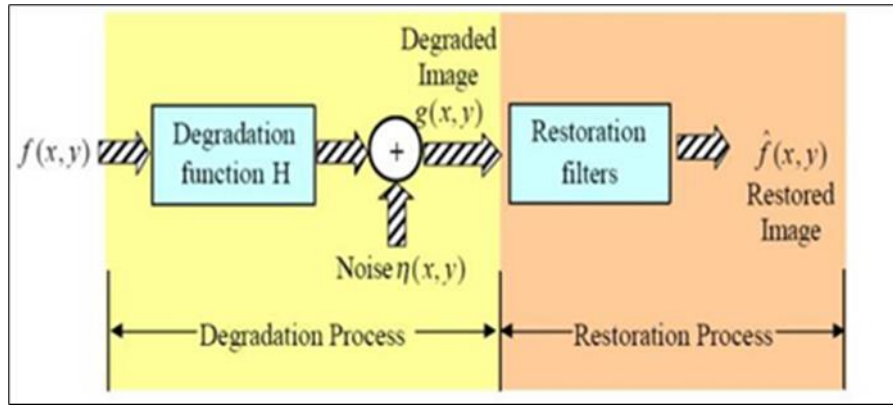
*Corresponding author: Rinku Kalotra

Figure 1 A model of the image degradation/ restoration process

1.4. Historical Documents

Historical documents are the documents which contains information about historical persons, places and incidents.

A collection of rare books and manuscripts, including early copies of works by Aristotle, Dante, Euclid, Homerand Virgil, are available in the Vatican Library in Rome and a beautiful work on Medicine, Religion and Science and Art written by the scholars, called Vedas, is preserved in India. We have the histories of various civilizations stored in libraries and museums around the world, a treasure of excellent literature, which most of people cannot access because of time and travel cost. By providing access to scientific, educational and historical documents and information, digital libraries can create powerful opportunities for revamping the system of education, improving knowledge and providing historical background.

In the last decade the technology has advanced mainly in two areas, computers and communication. This advancement has resulted in creation of a platform called Internet today. The invent and accessibility of internet has provided the opportunity to make this literature available to the people from all around the world. By developing image databases of these manuscripts and documents and making them available on the Internet it will be easier for thepeople to access them. Another significant advantage of image databases is we can save the documents from further degradation by preserving them and at the same time ensures their perpetual accessibility to the users. So, we are able to access images of historical artwork such as old paintings, sculptures, prints etc. through the digital libraries.

Figure 2 Historical documents

2. Literature Survey

The need for efficient image restoration methods have grown with the massive production of digital images of all kinds, often taken in poor conditions owing to the time or other constraints. Even though good cameras are available, but all the images may not be in a condition to be available for direct use for the analysis. From the literature survey, it has been observed that, there are several techniques like median filter, Gaussian filter for noise suppression which effectively suppress the noise but fail to preserve many useful details.

Md. Iqbal Qurashi (2013)

Said that the old degraded historical documents carry various important information regarding our culture, economics etc. Proper restoration of these documents is very necessary. They proposed a novel approach to enhance ancient historical documents and to enhance these digitalformat documents a two way approach is considered. At first stage Particle Swarm Optimization (PSO) and bilateral filter are applied and at second level Non-Linear Enhancement with bilateral filter is applied. Both the approaches are then tested visually and quantitively to show the effectiveness of the approach order to improve the quality of text regions and preserve stroke connectivity

K Shirai 2013

Whichperformanisotropic morphological dilation via implicit smoothing for the purpose of restoring the degraded character shapes of binarized images. Exploiting the idea of geodesic morphology that the binary image and its distance transformed image are interconvertible, they applied a smoothing method not to the binary image but to the distance transformed image, and then reconvert it by binarization. This allows us to apply conventional smoothing methods for continuous intensity, i.e., grayscale, images to the discrete intensity, i.e., binary, image implicitly. For instance, by using anisotropic diffusion together with geodesic dilation, anisotropic dilation along the stroke direction is obtained and brings better results.

Akihito Kitadai (2012)

said that shape features of character patterns on the documents are unstable or missing because most of the documents have been stained and degraded deeply. Digital archives of the documents with accurate character pattern retrieval methods are helpful for archaeologists and historians. They proposed a similarity evaluation method for character patterns with missing shape parts. It collaboratively works with non-linear normalization for such patterns, and modifies the templates for each trial of the retrieval efficiently. In the experiences using 4,911 Kanji (Chinese origin) character patterns from the Japanese historical documents called mokkans, the method shows improvements of the retrieval accuracy. They also presented a simple implementation of gradient presents a new adaptive approach for the binarization and enhancement of degraded documents. The proposed method does not require any parameter by user and deal with degradation which occur due to shadows, non-uniform illumination, low contrast, smear and strain, we follow several steps : a pre-processing procedure using low paasweinerthresholding by combining the calculated background surface with original image while incorporating image up sampling and finally post processing step P

3. Problem formulation

Historical documents are generally in degraded form because of ink bleed, stains, smudge, watermarks, mutilation and cracks etc. So it documents since these historical documents are of great importance, we need to restore these documents. We may restore it manually or digitally. In base paper Particle Swarm Optimisation technique for restoration of these documents has been used. In this paper, The geodesic morphological model will be compared with the fuzzy noise filter and shape restoration model using colour and texture for the purpose of historical document restoration will be compared with each other on the basis of various performance parameter.

3.1. Proposed System

Basically the purpose of this paper is to extract text from the degraded documents by separating background from the foreground by using method called binarization. In this paper the digital image of the document shall be converted into the binary form i.e. 0 & 1. The resultant binary image so obtained shall be having black text and a white background. This text will be more legible and require lesser storage space. First of all we have to identify a threshold level for this binarization.

This will involve the following steps:

- Normal image to grayscale conversion,
- A Gaussian filter to remove the noise,
- Image dilation to estimate the background,
- Estimated background subtraction from grayscale image,
- Global thresholding (OTSU) for modification.

This research is based upon the comparison of the effective historical document restoration techniques in order to analyse their performance on the basis of various performance parameters. The aim of this research is to get the best of techniques available for the purpose of historical document restoration within less time and with higher quality. The geodesic morphological model for character shape recognition for historical document restoration will be compared with the fuzzy noise filter and shape restoration model using colour and texture for the purpose of historical document restoration will be compared with each other on the basis of various performance parameters. The major objective of this research is based upon tackling the different levels and types of the noise in the historical documents under both of the models. To achieve such objectives, the fuzzy based filters are the best options. The proposed model will be using the fuzzy filters for the purpose of the denoising will be designed to remove the various levels of noise from the historical images when they will be available with the various classes of images. The hybrid filter will be using the combination of three filters such as, Histrogram adaptive fuzzy filter (HAF), Weighted fuzzy mean filter and minimum maximum fuzzy filter. The performance parameters will be PSNR, NAE, precision, recall, F-measure etc. for the purpose of the comparison with existing technique.
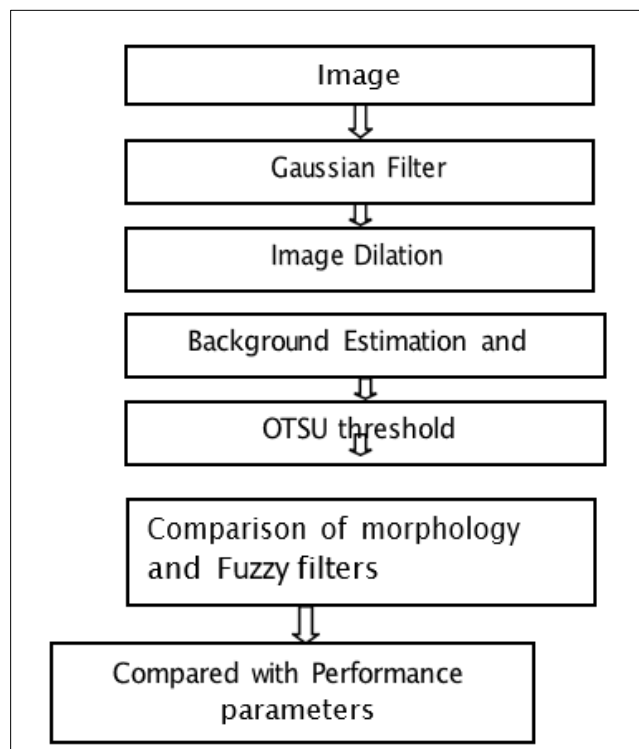

Figure 3 Basic Working of the System

Image acquisition will load the image in our program on which the enhancement has to be applied. Image acquisition will convert the image from its existing digital format in the form of a matrix. Second convert the image into black and white image according to the threshold value. Boundary tracing will select the objects (text in this case). Objects inside the boundaries will be extracted and then the feature enhancement will be performed on the extracted text. The boundary removal will be applied on the image before or after the feature extraction, wherever it will produce the best results.

We will start our research project by conducting a detailed literature review on the historical document and other similar document digital restoration to know the problem in detail. Then, a detailed algorithm designed would be

generated for the restoration of historical documents. The simulation would be implemented using MATLAB. The obtained results would be examined and compared with the existing security mechanism to address the similar issues.

4. Conclusion

In this paper, the performance evaluation of the two effective historical document restoration would be performed. The performance evaluation survey are quite important because they tell us about thestep will contain the feature extraction, which will use image binarization, boundary tracing, background removal. Binarization is an ordinary image processing technique to convert the image matrix values in to 1 and 0, which are called binary values. Binarization willeffectiveness of the existing models, as well as also describes the merits and demerits and research or performance gaps in the existing systems. In this paper, the performance evaluation survey design and the techniques being surveyed have been discussed. The performance evaluation survey would be performed on the fuzzy filters based historical document restoration model with OTSU and Geodesic morphological features based historical document restoration would be conducted. Both of the models would be simulated in the future. The performance of the two models would be then analyzed and compared with each other using the performance parameters like PSNR, MSE, RMSE, etc.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Md. IqbalQuraishi,Mallika De, Krishna Gopal Dhal, SahebMondal, Goutam Das "A NOVEL HYBRID APPROACH TO RESTORE HISTORICAL DEGRADED DOCUMENTS", ISSP, vol. 1,pp. 185-189, IEEE 2013.

[2] K. ShiraniY. Endo, A. Kitadai, S. Inoue,N. Kurushima, "Character Shape Restoration of Binarized Historical Documents by Smoothing via Geodesic Morphology", ICDAR, vol. 12, pp. 1285- 1289, IEEE 2013.

[3] B. Gangamma, Srikanta Murthy K, and ArunVikas Singh," Restoration of Degraded Historical Document Image", Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, No. 5, 2012.

[4] A. Kitadai, M. Nakagawa, H. Baba, andA. Watanabe, "Similarity evaluation and shape feature extraction for characterpattern retrieval to support reading historical documents", in Proc. IAPR Intl. WS.DAS, pp. 359–363, 2012.

[5] Ms.Ketki R. Ingole and Prof. V K. Shandilya, " Image Restoration of Historical Manuscripts", International Journal of Computer Science & Engineering Technolo, Vol. 2 No. 4.

[6] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Adaptive degraded document image binarization," Elsevier Trans. Pattern Recogn., vol. 39, no. 3, pp. 317–327, 2006.

[7] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proc. IEEE ICCV, pp. 839–846, 1998.

[8] Ms.Ketki R. Ingole and Prof. V K. Shandilya, " Image Restoration of Historical Manuscripts", International Journal of Computer Science & Engineering Technology, Vol. 2 No. 4