Image Restoration of Historical Manuscripts

Ms. Ketki R. Ingole,*
M.E.(CSc & Engg.) 2nd Sem
Sipna's College of Engineering,
Amravati, Maharashtra, India,

Prof. V K. Shandilya
Assistant Proffessor.,
SSPM's COET
Amravati, Maharashtra, India,

Abstract—The quality of historical documents related to art and architecture, mathematics, astronomy, astrology, architecture, law, music and medicine, written around several hundreds of years, which are still available for reference today and which is preserved by libraries and universities all around the world, is poor due to physical deterioration. Digital image processing techniques are necessary to improve the legibility of the manuscripts.

The deterioration processes often result in uneven background coloration across the image and the darkening of the background which reduces the contrast between the foreground text and the background color of the documents. The problem is introduced during the conversion of manuscripts to their digital image form. The proposed methodology performs the improvement in the quality of historical manuscripts having uneven background and low contrast due to the traditional mode of manufacture and the effect of ageing and degradation.

Keywords— **historical manuscripts, uneven background, low contrast, foreground text**

I. INTRODUCTION

Historical documents related to art and architecture, mathematics, astronomy, astrology, architecture, law, music and medicine, written around several hundreds of years are still available for reference today which is preserved by libraries and universities all around the world. A survey by the Institute of Asian Studies, Chennai, India indicates that there are still about a hundred thousand manuscripts surviving in Indian repositories alone with many more scattered across India, Nepal, Myanmar, Laos, Thailand, Cambodia and other Southeast Asian Countries. In the past, Indian kings, temple authorities, and other concerned individuals ensured that the oldest manuscripts were ritually disposed only after they had been copied. When this age old cycle was broken in the 19th century, the remaining corpus of manuscripts and the knowledge contained in them began a long slide into obscurity and destruction. Most of these documents are almost at the end of their natural lifetime or are facing degradation from elements such as dampness, fungus, ants and cockroaches. These documents always suffered from degradation and contents of the documents are not available clear and useful form. This inspired various communities to develop a preservation tool for future reference of these valuable historical documents.

In the past, Indian kings, temple authorities, and other concerned individuals ensured that the oldest manuscripts were ritually disposed only after they had been copied. When this age old cycle was broken in the 19th century, the remaining corpus of manuscripts and the knowledge contained in them began a long slide into obscurity and destruction. Most of these documents are almost at the end of their natural lifetime or are facing degradation from elements such as dampness, fungus, ants and cockroaches. These documents always suffered from degradation and contents of the documents are not available clear and useful form. This inspired various communities to develop a preservation tool for future reference of these valuable historical documents.

Efforts, funded by many foundations, universities and other institutions, are now underway for recovering and preserving these valuable historical documents. Besides the many programs for preserving the manuscripts in their physical form, scanning and digital photograph imaging have been used to preserve their content and current appearance for future studies.

Despite the availability of advanced photography and scanning equipment, natural aging and deterioration have rendered many documents images unreadable. The original documents are aged, leading to

deterioration of the writing media, with seepage of ink and smearing along cracks, damage to the manuscripts due to the holes used for binding the manuscript leaves and dirt and other discoloration. The process of capturing a digital image of the leaves also presents

some difficulties. In order to best preserve the fragile originals, the digital images are sometimes captured by using digital cameras instead of platen scanners. Manuscripts cannot be forced flat and the light source for digital cameras is usually uneven. These factors lead to a very poor contrast between the foreground text and the background. Digital image processing techniques are necessary to improve the legibility of the manuscripts.

Image processing techniques can help enhance the images of these manuscripts so as to enable retrieval of the written text from these degraded documents. Many methods for historical image enhancement are driven by the goal of improving human readability while maintaining the original "look and feel" of the documents. But these methods do not produce satisfactory results in processing these manuscripts since the color intensity of the background varies throughout the image.

The proposed method for historical image restoration will performed efficiently than the previous method. In the proposed method, first target the problem of uneven background in digital images of manuscripts. Then apply a background normalization algorithm which smoothes out the background. The background normalization enhances the image, making it more legible to the eye as well as facilitating segmentation of the text from the non-text background.

II. Literature Review

Most previous document image enhancement algorithms have been designed primarily for digitization of modern documents. These methods aim to extract text from noisy documents with uneven background. Three popular methods, namely Otsu's thresholding technique[6], entropy techniques proposed by Kapur and al.[4] and the minimal error technique by Kittler and Illingworth,[7] are analyzed and compared. Otsu reference proposed a criterion for maximizing the between-class variance of pixel intensity to perform picture thresholding. However, Otsu’s method for image segmentation is very time consuming because of the inefficient formulation of the between-class variance [6].

Shi and Srirangaraj Setlur ,Venu Govindraju palm leaves were one of the earliest forms of writing media and their use as writing material in South and Southeast Asia has been recorded from as early as the fifth century B.C. until as recently as the late 19th century. Palm leaf manuscripts relating to art and architecture, mathematics, astronomy, astrology, and medicine dating back several hundreds of years are still available for reference today thanks to many ongoing e®orts for preservation of ancient documents by libraries and universities around the world. Palm leaf manuscripts typically last a few centuries but with time the palm leaves degrade and the writing becomes illegible to be useful in any form. Image processing techniques can help enhance the images of these manuscripts so as to enable retrieval of the written text from these degraded documents. They have proposed a set of transform based methods for enhancing digital images of palm leaf manuscripts. The methods first approximate the background of a gray scale image using one of two models (piece-wise linear or nonlinear models). The background approximations are designed to overcome unevenness of document background. Then the background normalization algorithms are applied to the component channel images of a color palm leaf image. They also proposed two local adaptive normalization algorithms for extracting enhanced gray scale images from color palm leaf images. The techniques are tested on a set of palm leaf images from various sources and the preliminary results show significant improvement in readability. The techniques can also be used to enhance images of ancient, historical, degraded papyrus and paper documents [2].

III. Analysis Of Problem

The historical documents, preserved at the National Library, are considered as an important part of cultural heritage. These funds suffer from a progressive degradation and therefore risk disappearing. The automatic processing of this type of documents in order to restore and use, is a definite advantageous. There are number of methods are present and used, as mentioned in literature review section, for preserving the historical documents. But still these methods can’t produce an efficient result as required.

Generally, there are two types of deficiencies in the quality of historical document images. First, the original paper document is aged leading to deterioration of the paper media, ink seeping and smearing, damages and added dirt. The second problem is introduced during conversion of the documents to their digital image form. In order to best preserve the fragile originals, the digital images are usually captured by using digital cameras instead of platen scanners. The paper documents cannot be forced flat and the light source for digital cameras is usually uneven. Due to above deficiencies, the document image background together with the foreground handwritten text are fluctuating. The separation of text from the paper background is often unclear.

The problem of uneven background color intensity across an image is often seen in historical document images. Due to above deficiencies, the document image background together with the foreground handwritten text are fluctuating.

Our method is first targeted towards enhancing images with uneven background. The method works for gray scale images as well as color images. First target the problem of uneven background in digital images of manuscripts. Then apply a background normalization algorithm which smoothes out the background. The background normalization enhances the image, making it more legible to the eye as well as facilitating segmentation of the text from the non-text background.

IV. PROPOSED WORK

The proposed document enhancement methodology permits the improvement of the quality of historical Arabic manuscripts which presented uneven background and low contrast due to the traditional mode of manufacture and the effect of ageing and degradation. It consists of the following steps: foreground extraction, contrast adjustment, foreground-background segmentation, reconstruction of document image with smoothing.

The developed document segmentation method operates with background light intensity normalization algorithm proposed by shi and al [1][2] and applied to palm leaf manuscripts. This technique can be improved with the histogram normalization used in color image manuscript context. The segmentation method proceeds on luminance and distort the contrast. These distortions are corrected with contrast adjustment. The new enhanced foreground image is segmented to foreground-background on the basis of ML estimation. The initial parameters for the ML method are estimated by k-means clustering algorithm. The segmented image is used to produce a final restored document image. Figure 5.1 presents the flowchart of our proposed methodology. The steps below are described in the following sections:

Step 1: Application of an iterative background normalization algorithm for a first foreground-background.

Step 2: Correction of visual distortions of obtained foreground using histogram normalization.

Step 3: The algorithm performs final foreground-background segmentation.

Step 4: Reconstruction of images color space and production of the restored manuscript.

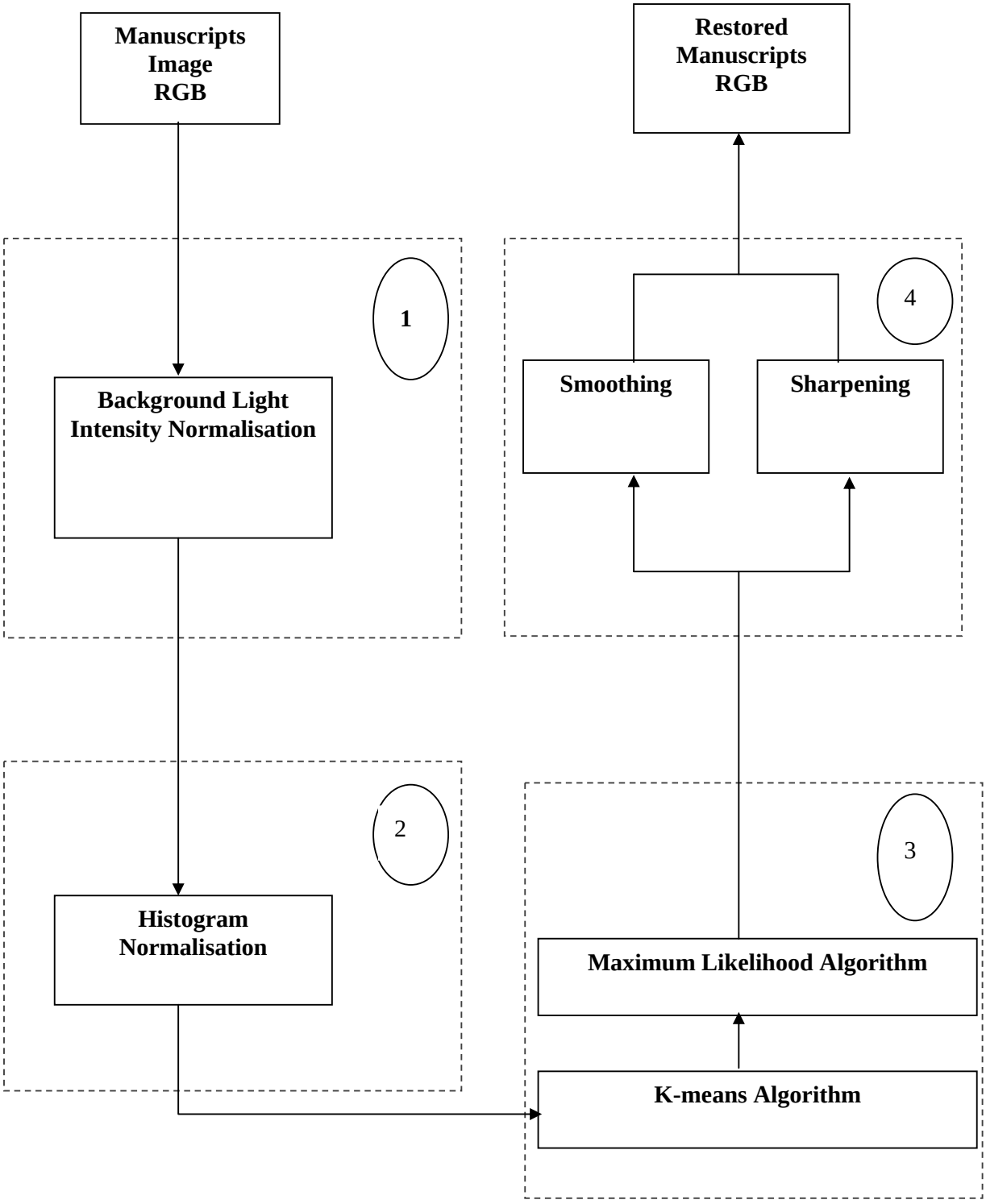

Figure 5.1: Flowchart of proposed methodology

A. Step 1: Application of an iterative background normalisation algorithm

A background normalization algorithm is designed to adaptively adjust the pixel intensity based on an approximation of the background of a document image. The approximation is done along each scanline of an image. Along the scanline, first filter the pixels to leave behind only those that belong to the background.

A moving window is then used to compute a nonlinear curve approximating the background. The background normalization algorithms are designed for normalizing pixel intensities of gray scale document images.

B. Step 2:Enhancing foreground using histogram normalisation

To enhance the contrast, we apply the histogram normalization on the transformed image to elevate the text away from the background.

- *Histogram Normalization*

The original background colors are mapped into a color range very close to white and all other colors darker than the estimated original background colors map to darker grey levels. Since the original background color in many of the aged manuscripts are so dark that the foreground text colors are very close to the background, the transformed grey-scale images -while brighter than the original dark image need further enhancement to increase the contrast between the text and the background. A histogram normalization algorithm is applied to the transformed grey-scale images to effect this contrast enhancement.

C. Step 3:Final foreground-background segmentation

The segmentation method proceeds on luminance and distort the contrast. These distortions are corrected with contrast adjustment. The new enhanced foreground image is segmented to foreground-background. The segmented image is used to produce a final restored document image.

Document image manuscript segmentation can be considered as a statistical classification problem. The estimation of parameters of classification is given by Kmeans algorithm improved by ML method.

D. Step 4:Reconstruction of document with smoothing and sharpening

Foreground/background segmentation is used for the restoration of historical manuscripts. In fact, the restored image is constructed by superposition of the foreground and the average of background. The smoothing is done on the background of the segmented image and sharpening is take place on the foreground of segmented image.

IV. IMPLICATIONS

The large number of the images has obvious uneven background problems and low contrast. Visual inspection of the enhanced images produced by the proposed techniques will show a marked improvement in image quality for human reading. The proposed method successfully will find a better digitized image in all these cases.

The techniques described in this paper will also used to process images of other historical documents such as aged, stained or otherwise discoloured paper documents and it will be used to generate digitized images of very high quality with very little text degradation.

V. APPLICATIONS

Digital Historical library can be created with the help of these applications and such type of historical library is present in south-east Asia which contains various manuscripts. These manuscripts contain precious data which is useful in medical science as well as to know about traditions present and follow at that time.

REFERENCES

[1] Zhixin Shi and Venu Govindaraju ,“Historical Handwritten Document Image Segmentation Using Background Light Intensity Normalization” ,Center of Excellence for Document Analysis and Recognition (CEDAR), State University of New York at Buffalo, Amherst, USA.

[2] Zhixin Shi Srirangaraj Setlur ,Venu Govindaraju “Digital Image Enhancement using Normalization Techniques and their application to Palm Leaf Manuscripts”, Center of Excellence for Document Analysis and Recognition(CEDAR), State University of New York at Buralo, Buralo, NY 14228, U.S.A.February 21, 2005 .

[3] Wafa Boussellaa, “A Methodology for the Separation of Foreground/Background in Arabic Historical Manuscripts using Hybrid Methods”

[4] P. J.N.Kapur and A.K.C.Wong, “A new method for gray- level picture thresholding using the entropy of the histogram," Computer Vision, Graphics, and Image Processing, vol. 29, pp.273{285, 1985}.

[5] G. Leedham, S. Varma, A. Patankar, V. Govindaraju ”Separating Text and Background in Degraded Document Images - A Comparison of Global Thresholding Techniques for Multi-Stage Thresholding” Proceedings Eighth InternationalWorkshop on Frontiers of Handwriting Recognition,September, 2002, pp. 244-249.

[6] N.Otsu, \A threshold selection method from gray level histogram," IEEE Transactions in Sys-tems, Man, and Cybernetics, vol. 9, pp. 62{66, 1979.

[7] J.Kittler and J.Illingworth, \Minimum error thresholding," Pattern Recognition, vol. 19, no. 1,pp. 41{47, 1986.

[8] C.A.B.Mello and R.D.Lins, \Image segmentation of historical documents," in Visual2000, Mex-ico City, Mexico, September 2000.
[9] Y. Leydier, F.L. Bourgeois, and H. Emptoz, Serialized K-Means for Adaptive Color Image Segmentation-Application to Document Images and Others, in 6th.