DIGITAL ENHANCEMENT OF INDIAN MANUSCRIPT, YASHODHAR CHARITRA

Sai Siddharth Kota, Raja Massand, Abhinaya Agrawal and
Preety Singh

Computer Science and Engineering Department,
The LNM Institute of Information Technology, Jaipur, India

siddharthhp@gmail.com, rajamassand@gmail.com,
abc.abhi99@gmail.com, prtysingh@gmail.com

ABSTRACT

Over the years, many of our ancient manuscripts have been damaged by natural elements or intentionally erased and re-used to record other information. While manual preservation techniques are being carried out for the conservation of our ancient texts, digital image processing is an alternative for the archival storage of the invaluable text contained in them. To successfully recover text from such documents, it is important to understand the nature of the writing and materials on which they are written. Different imaging and processing techniques are needed, depending on the the condition of the manuscript. In recent years, modern imaging techniques have been applied to ancient manuscripts to recover writings that are not visible to the naked eye or not recognizable due to various factors. In this paper, we apply imaging techniques on an ancient manuscript, Yashodhar Charitra, and restore it digitally.

KEYWORDS

*Manuscript Restoration, Noise Removal, Gaussian Bandpass filter, Thresholding, Image Enhancement.*

1. INTRODUCTION

Ancient manuscripts form an integral part of our rich heritage. A number of these texts have been written on parchment. If stored under very dry conditions, parchment can last for thousands of years. For writing on them, early inks were made with carbon black suspended in water. This ink provided, and continues to provide, well-defined and high-contrast writing. Later manuscripts were written with iron gall ink, which was easier to make and harder to remove from the surface.

Natural and artificial agents have resulted in the deterioration of these manuscripts. Image processing techniques can be applied to images of these texts to restore them and store them digitally. However, for application of these techniques knowledge of the ink and parchment used is essential. Damaged parchment is often very dark, making any surviving text characters hard to read.

In this paper, we apply digital image processing techniques for enhancement and text restoration of an ancient manuscript, *Yashodhar Charitra*, which was written using carbon black suspended in water on parchment made up of goat or sheep skin which is very durable. Digitization of the
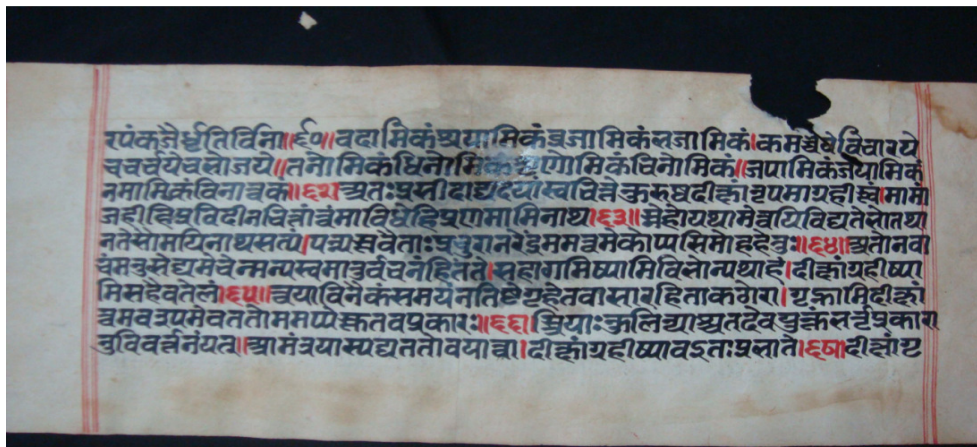
restored manuscript will ensure wider reach of the text to readers, economy of storage and safety from damage inflicted by nature.
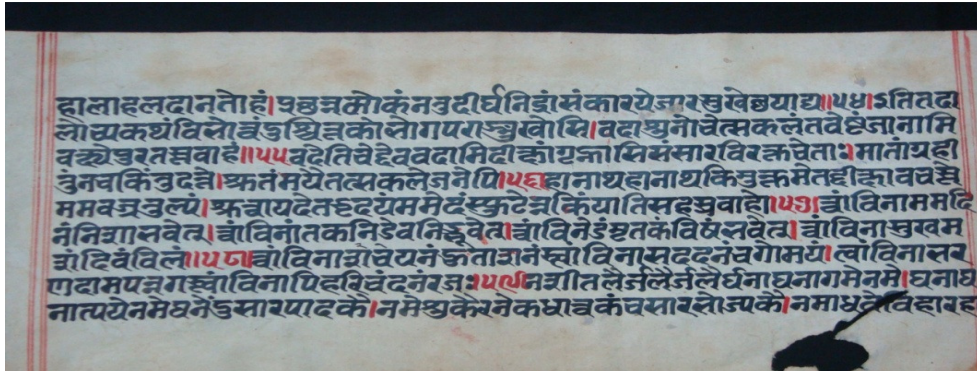
(a)   Missing text due to page tear
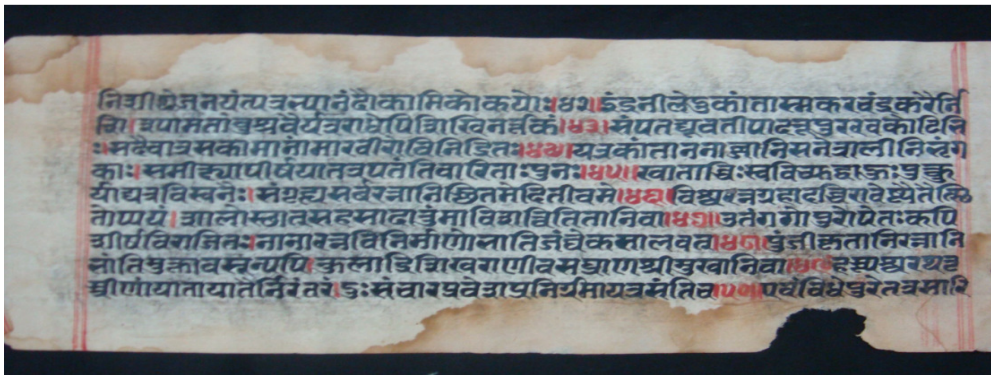
(b)   Blotches

(c)   Damage due to moisture

Figure 1: Various types of damages inflicted

2. RELATED WORK

In [**Error! Reference source not found.**], Alirezaee et al. have developed a restoration algorithm for the Pahlavi or middle-age Persian manuscripts. The algorithm uses mathematical morphology and analysis of connected components to overlapped lines, words and characters. The algorithm was tested on 200 pages of Pahlavi document. The noise and destructive effects are removed.

In [**Error! Reference source not found.**], Chaudhuri et al. have separated text from non-text doodles of poet Rabindranath Tagore in Indian manuscripts. The approach generates connected components and classifies them as text and non-text based on a comparison between the total number of pixels and the number of boundary pixels. Further separation is done based on the stroke width computed for each window.

In [**Error! Reference source not found.**], Hedjam and Cheriet proposed a data representation for text extraction from multispectral historical document images. They performed foreground pattern extraction using region-of-interest (ROI) analysis and a maximum likelihood classifier. Two new features containing spectral components are introduced.

3. PROPOSED METHODOLOGY

Manuscript restoration techniques involves various steps. First of all, we create a database of the damaged manuscript. Each page of the manuscript is photographed using a high-defintion digital camera. The manuscript is analyzed for the different damages. Each page of the manuscript is subjected to noise removal and enhancement technique. The various steps involved in our approcah are :

- Digitization of images: The manuscript was digitized using a standard digital camera.
- Noise removal: The image is filtered to remove the unwanted noise from the digitized image. This methodology uses a filtration technique which uses Gaussian Bandpass filter and the difference in colour intensities of the damaged part to the non-damaged part.
- Thresholding: It was done in order to remove the damaged segments of the image.
- Image enhacement: This helps us in improving the various aspects of an image to make it visually better.
- Text restoration: Some missing areas of the text were recovered using text restoration.

4. EXPERIMENT

4.1 Database

For experimental purpose, we obtained our database from *Digambar Jain Manuscript Conservation Centre* under *The National Mission for Manuscripts (NAMAMI)* which is an autonomous organisation under Ministry of Culture, Government of India, established to conserve Indian manuscripts and create a national resource base. We used a manuscript called *Yashodhar Charitra*, written in the year 1661, by *Gyankirti* based on the fascinating story of *Yashodhara*. It is written in *Sanskrit* with a touch of *Varhadi* dialect. It contains 69 pages. The manuscript was digitized using a digital camera. The database was analysed for various types of damages. It was observed that there were a number of factors like dark areas, blotches, degradation due to moisture and missing text due to page tear.

4.2 Preprocessing

Image preprocessing is very important as it can significantly increase the performance of subsequent enhancement steps. Features of digitized images along the database can vary due to different factors like luminance level, wearing out due to time etc .
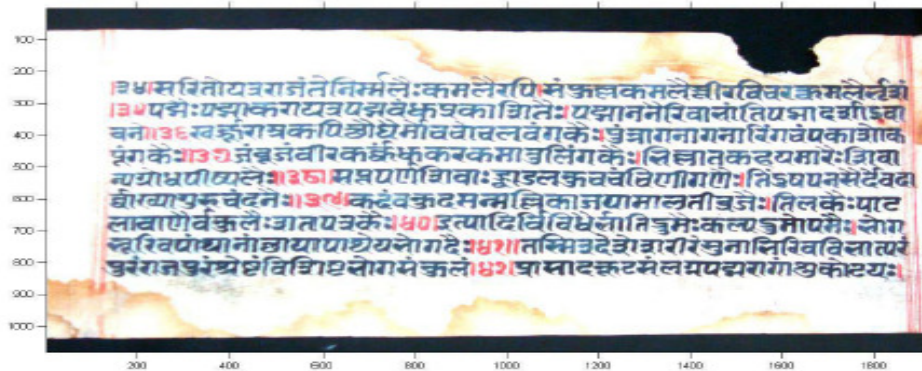
4.2.1 Noise Removal

Noise removal basically acts as a tool to remove the unwanted noise from our digitized image which may have occured during image acquisition. It uses the Gaussian bandpass filter which removes noise contained in a certain range.

Gaussian bandpass filtering is done in the frequency domain. The function makes use of the simple principle that a bandpass filter can be obtained by multiplying a lowpass filter with a highpass filter where the lowpass filter has a higher cut off frequency than the high pass filter. This process is shown in Figure 2.
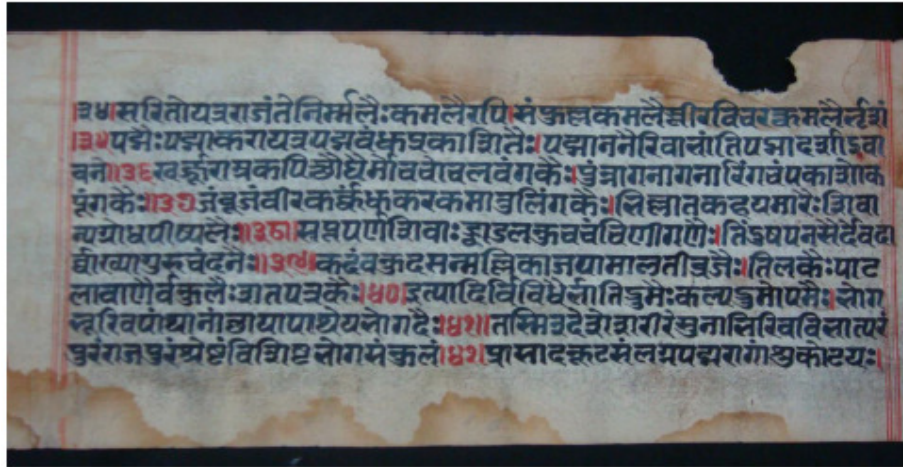
(a) Image before noise removal


(b) Image after noise removal

Figure 2: Noise Removal

4.2.2 Thresholding

Thresholding is the simplest method of image segmentation. We used it to convert our preprocessed RGB image to various other colour formats like YCbCr, HSV, CMYK based on the type of damage inflicted.

- YCbCr was used when we had light blotches in the digital image.
- HSV was used for images degraded by moisture.
- CMYk was used for images having dark blotches.

The thresholded image is shown in Figure 3.
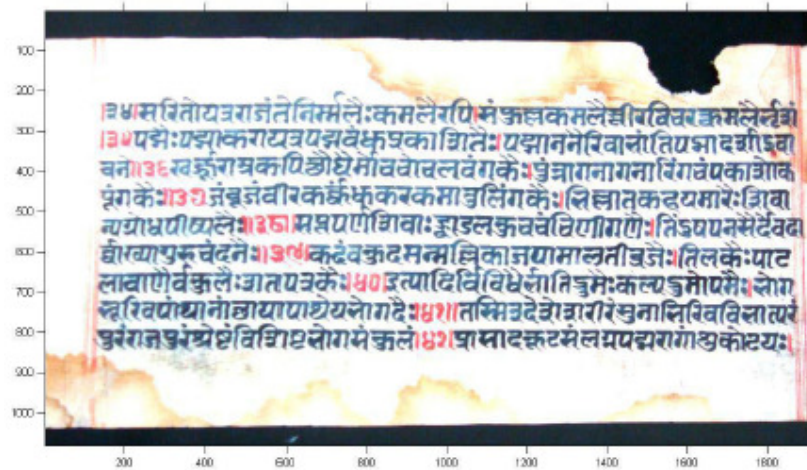
(a) Image before thresholding
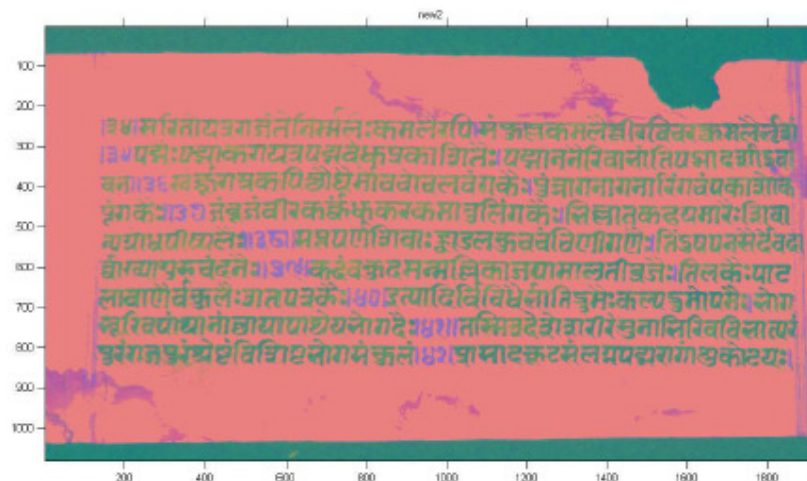
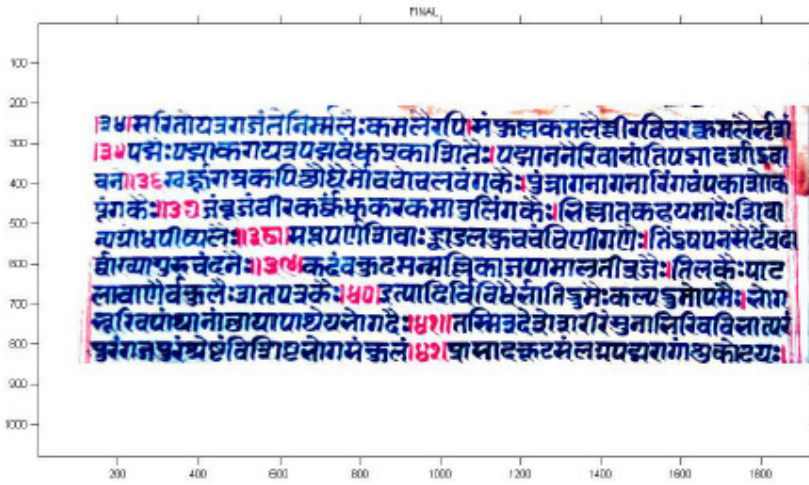
(b) Image after thresholding in YCbCr

Figure 3: Thresholding

4.3 Image Enhancement

After converting the image to a specific color format based on the type of damage, it is enhaced for various aspects like brightness, contrast and sharpness to make it more visually presentable. The enhanced image is shown in Figure 4.
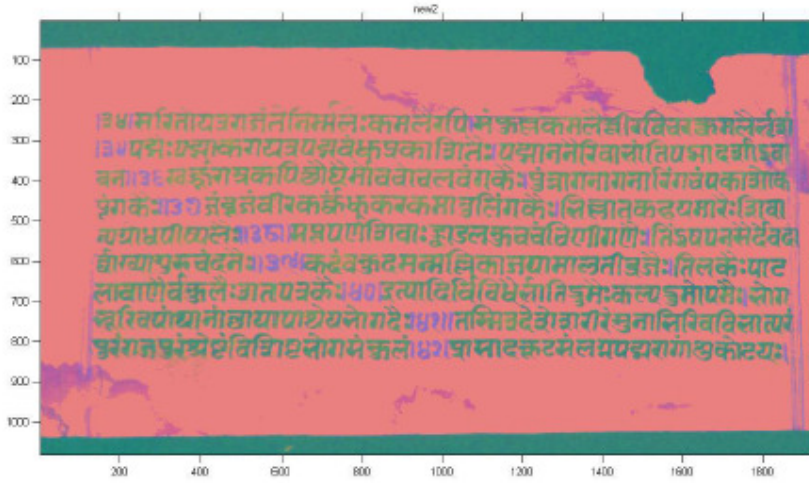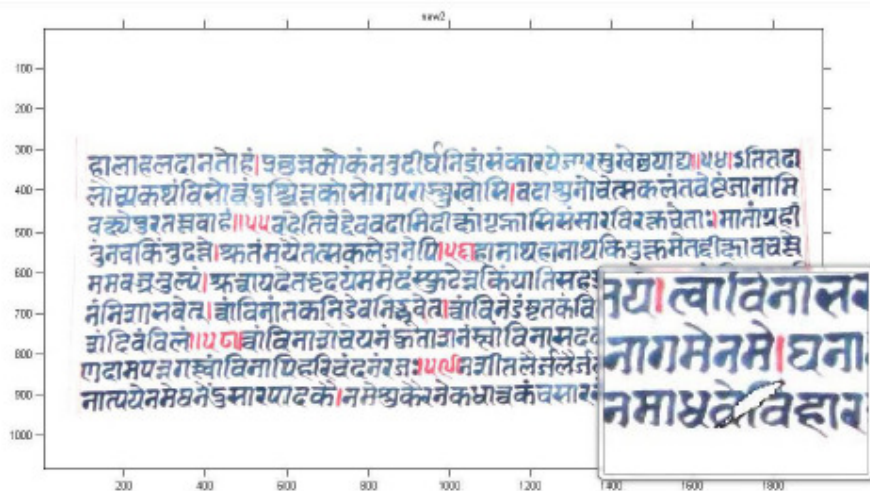

(a) Image before enhancement
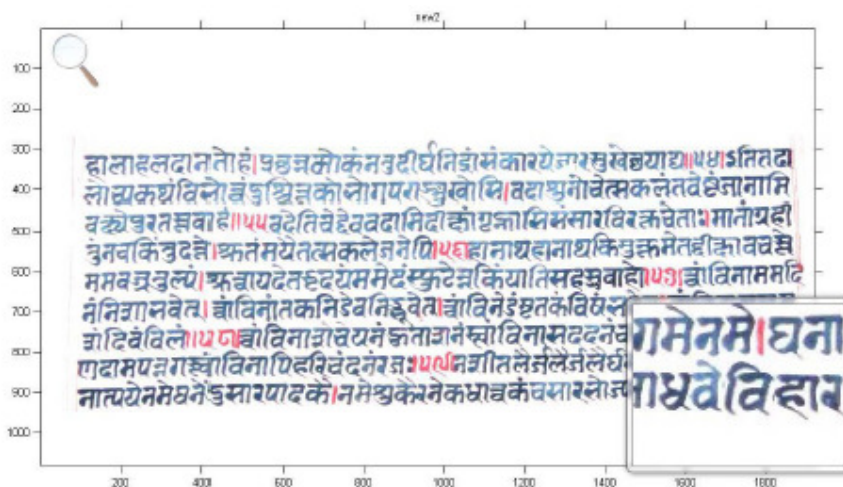

(b) Image after enhancement

Figure 4: Image enhancement

4.4 Text Restoration

Some segments of the manuscript were severely damaged and the text in those areas was missing. For such cases, we consulted the language expert and manually replaced the missing text. Similar text is located elsewhere in the manuscript and digitally picked up and pasted where the text is missing. This can be seen more clearly in Figure 5.
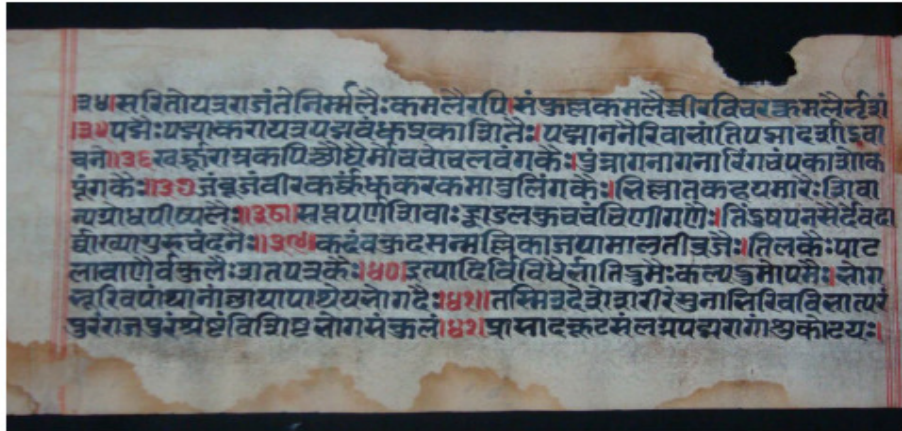
(a) Image before text restoration

(b) Image after text restoration

Figure 5: Text restoration

5. RESULT AND ANALYSIS

The damaged manuscript is restored using noise removal and image enhancement techniques. Text restoration for missing text is done by digitally cropping equivalent text from elsewhere and pasting it where it is missing. The results of the whole process are shown in Figure 6.

It is seen that major spoilage of the manuscript is removed. The manuscript becomes noise free. The blotches caused due to exposure to moisture, air and other pollutants are visibly removed. The text which is missing due to page tear is replaced accordingly. The end result of the restored manuscript is a clean, easy to read and defect-free digital manuscript.
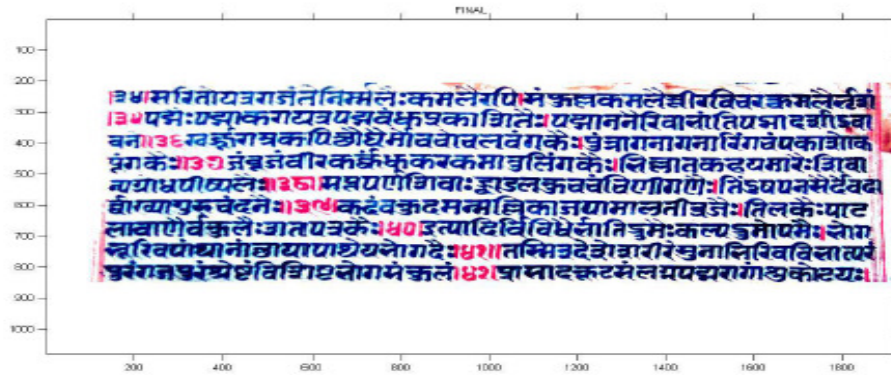
(a)   Image before noise removal

(b) Image after noise removal
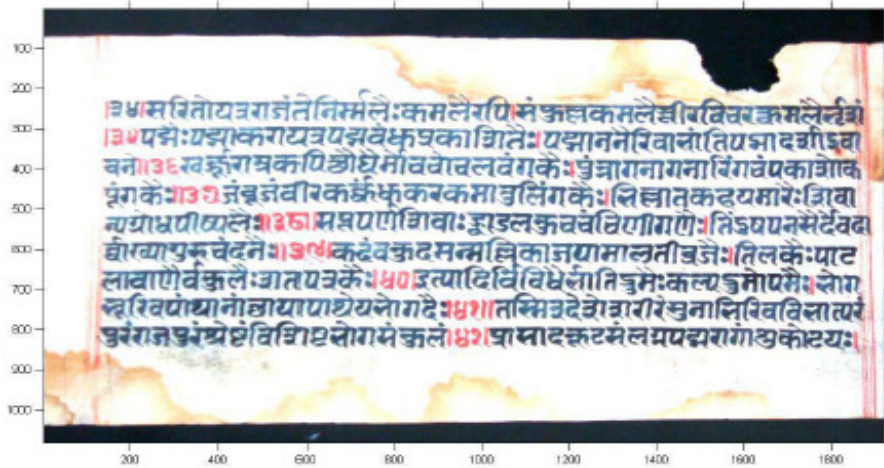
(c) Image after thresholding

(d) Image after enhancement

Figure 6: Various stages of Manuscript restoration

6. CONCLUSION

Image processing techniques were applied for digital restoration of an ancient manuscript, *Yashodhar Charitra*, authored by *Gyankirti* . The deteriorated manuscript was inspected and treated for various defects. The noise free and enhanced manuscript can now be digitally archived for wider reach to the readers and free from effects of natural pollutants. In fututre, the techniques will be applied to other ancient manuscripts available with conservation centers.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Keith T. Knox, Roger L. Easton Jr. , and William Christens-Barry "Image restoration of damaged or erased manuscripts,". in 16th European Signal Processing Conference (EUSIPCO 2008), pp. 1–5, 2008.

[2]   Angelika Garz, Markus Diem and Robert Sablatnig "Detecting Text Areas and Decorative Elements in Ancient Manuscripts,". in 12th International Conference on Frontiers in Handwriting Recognition, pp. 1–6, 2010.

[3]   Avekash Gupta; Sunil Kumar; Rajat Gupta; Santanu ChaudhuryÎ³;Shiv Dutt Joshi "Enhancement of Old Manuscript Images," in IBM India Research Lab, pp. 1–5, 2007.

[4]   Aaron Greenblatt; Karen Panetta "Restoration of Semi-Transparent Blotches in" Damaged Texts, Manuscripts, and Images Through Localized, Logarithmic Image Enhancement," in ISCCSP 2008, pp. 1–6, 2008.

[5]   Umbaugh Scot E "Computer Vision and Image Processing", in Hall, NJ, 1998, ISBN 0-13-264599-8

[6]   B. B. Chaudhuri, A. Saraf, A. Kumari, S. Borah and A. Goyal "Separation of text from non-text doodles of poet Rabrindranath Tagore's manuscripts", in National Conference on Computing and Communication Systems, pp. 1–5, 2012.

[7]   S. Alirezaee, H. Aghaeinaia, M. Ahmadi and K. Faez "An efficient restoration algorithm for the historic middle-age Persian (Pahlavi) manuscripts", in IEEE International Conference on Systems, Man and Cybernetics, Volume 3, pp. 2114–2120, 2005.

[8]   R. Hedjam and M. Cheriet "Novel data representation for text extraction from multispectral historical document image", in IEEE International Conference on Document Analysis and Recognition, pp.172–176, 2011.