International Journal of Early Childhood Special Education (INT-JECSE)
DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

ENHANCING LEGIBILITY AND QUALITY OF HISTORICAL DOCUMENTS THROUGH IMAGE PROCESSING

[#1]P.MADHUSHEKAR, ***Assistant Professor,***
[#2]B.RAJENDAR, ***Assistant Professor,***
[#3]A.SRIKANTH, ***Assistant Professor,***
Department of ECE,
TRINITY COLLEGE OF ENGINEERING AND TECHNOLOGY, PEDDAPALLI, TELANGANA.

ABSTRACT: The focus of this work is on improving and recursively documenting history. Complex image processing methods are often used to clean up and improve documents, which are then turned into binary pictures. Because of the methods, computations are harder to do and cost more. If you don't store your old records properly, moisture can make it hard to tell the difference between the foreground and background. After digital image processing, the methods need to be tested to see how well they work. Our method speeds up the process by taking into account differences in the pictures on the paper. Because the steps are repeated, the methods are easier for people to use. The test results show which way works better than the others because they compare them.
Keywords: Image Enhancement techniques, Filters, Noise Removal techniques

1. INTRODUCTION

The main goal of the study is to keep historical records safe. Furthermore, older texts are more likely to have bugs and rodents eating them. The vast majority of old papers that would fall apart on touch are carefully sealed in containers to keep bugs out. Making a link between the past and the present is important for figuring out the best way to get to the future.

In this way, it was easier to learn about the conditions that led to the current state of freedom and the beginnings of life. The vast majority of people are interested in learning about past by keeping things like photos, letters, and telltales safe. The only way to get a fuller picture of the past is to look at this old information. It is amazing that these objects have been around for so long and gained value over time. Image binarization has been the focus of a lot of interesting research. Some of these are also mentioned in historical records. The variation of pixel intensity is what makes Otsu's approach, which is one of the most common ways to binarize photos, work.

Bernsen sets the limits of his neighborhood by talking to his neighbors. Niblack works with the standard variation and the local mean. Sauvola talks about a way to work with document images that uses two techniques to give each pixel its own unique threshold. Low contrast between the foreground and background and bad storage circumstances for old documents are caused by ink fading, paper breaking down, and humidity. A digital form that can be read is also easier to get because many academics can't read the fragile manuscripts. State and national libraries will digitize their huge collections of historical handwritten texts so that secondary sources, like books, can continue to provide access to information instead of letting main sources, like old scrolls of text, become less useful.

Changing a scanned grayscale picture to a binary image is an important step in many image analysis systems, such as document image processing. This is done by keeping the foreground and erasing the background. Original papers are often dirty because they are old and the writing has become blurry over time. The main focus of this study is on manuscripts with clear handwriting. Manuscripts that are very disorganized may have ink that runs off the page or just old paper and ink. Twenty-five examples are used to test the filtering methods against different types of noise. To run the program, Matlab R2007b is used. There will be a full evaluation of each option in the next part.

2. LITERATURE SURVEY

Patel, A., & Sharma, M. (2024).This study focuses on how various image processing technologies can be utilized to preserve historical manuscripts. The authors discuss a variety of techniques that can improve the readability of fading or damaged text, such as algorithms for color restoration, noise reduction, and picture improvement. The study demonstrates how these technologies could be used to digitize cultural property and preserve it for decades to come, while also shielding delicate materials from physical harm.

Zhang, W., & Liu, Y. (2023).In order to repair broken or out-of-date text, this article describes an automated method for repairing damaged images using cutting-edge image processing techniques. Automated damage identification, machine learning-based text restoration, and contrast enhancement are the authors' primary areas of interest. This technology makes it possible to preserve historical records more effectively and economically through digital archiving.

Singh, R., & Gupta, N. (2023).This work focuses on the use of machine learning and image enhancement techniques

DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

for the restoration of historical writings. The authors demonstrate in this example how to repair fading, broken, or otherwise altered old lettering using deep learning algorithms. Their results demonstrate how these techniques greatly improve the historical texts' aesthetic value and readability.

Thakur, P., & Mehta, S. (2022).The authors investigate the prospect of restoring historical texts that have faded or become incomprehensible through the use of image processing techniques. We evaluate the effectiveness of different techniques for enhancing page readability, including histogram equalization, contrast modulation, and noise filtering. The study discovered that the best way to restore documents was to use a variety of picture-enhancing techniques.

Patel, K., & Desai, P. (2022).This study examines and contrasts several image processing methods, including adaptive filtering, picture denoising, and ink fading restoration, for the purpose of archiving historical documents. The study carefully weighs the benefits and drawbacks of each method, focusing on those that are able to maintain the historical authenticity and original quality of earlier works.

Ali, M., & Rahman, Z. (2022).The goal of this project is to make old and damaged manuscripts readable again by applying cutting-edge image processing techniques. To make the text easier to read, the authors employ techniques including edge recognition, contrast augmentation, and morphological procedures. By showing significant increases in the visibility of fading text, their research contributes to the digitization and preservation of valuable historical texts.

Chowdhury, T., & Kumar, R. (2021).his work focuses on the digital restoration of ancient texts through picture enhancement techniques. These days, techniques like contrast enhancement, noise reduction, and geometry correction can be used to improve digital historical texts. They demonstrate how digital restoration can prolong the life of objects while maintaining their historical significance.

Bhat, H., & Soni, V. (2021).This article examines the difficulties and solutions related to image processing for document preservation. The authors investigate a number of techniques, such as distortion correction, image segmentation, and text improvement. Furthermore, they discuss the shortcomings of current technology and offer novel ideas to improve the efficiency of document preservation in digital archives.

Verma, S., & Shah, H. (2021).This study examines optical character recognition (OCR) as a way to aid in the preservation of historical documents. They discuss the potential benefits of optical character recognition (OCR) for digital transcription and the preservation of old texts using image processing methods like noise reduction and text enhancement. The study explores the benefits and drawbacks of preserving historical texts with optical character recognition (OCR).

Liu, J., & Chen, T. (2020).The main objective of this initiative, which uses photo segmentation algorithms, is to preserve fragile ancient documents. The writers advise breaking the document up into smaller, easier-to-manage sections and refining each one separately. This approach ensures a more accurate and precise preservation, especially for records that are too delicate to handle physically.

Gupta, A., & Rani, S. (2020).The various image processing technologies used for the preservation of old documents are examined in this study. The authors evaluate the methods' capacity to maintain the integrity and readability of ancient texts after classifying them based on their applications, such as text recognition, restoration, or augmentation. For academics and archivists employed in the field, it is an invaluable resource.

Rathod, S., & Agarwal, A. (2020).This research focuses on the use of high-resolution imaging technology for historical manuscript preservation. The authors demonstrate how old manuscripts can have their fine features preserved, correcting damage and improving the document's overall quality by fusing image processing techniques with high-resolution scans. Future uses of archives and the possible impact of high-resolution imaging are also discussed.

Patel, R., & Jain, P. (2020).This inquiry focuses on preservation strategies, including color restoration and noise reduction. Writers use techniques like denoising, color restoration, and histogram matching to restore faded or damaged text to its original appearance. According to the study, making documents easier to read increases their accessibility for archiving and research.

Sharma, V., & Das, S. (2020).he authors look into the possible applications of multispectral imaging in data preservation. This study examines how multispectral photography, which simultaneously captures images at different light wavelengths, can improve the visibility of darkened or obscured writing and reveal elements that were previously hidden. The main focus of the study is on the possible advantages of this method of data retention, particularly for sensitive data.

Kumar, A., & Kapoor, P. (2020). largely uses image enhancing techniques to improve the quality of historical documents. To repair fading or damaged text, the authors employ techniques like contrast correction, edge recognition, and sharpening. According to their findings, this gain in readability has made it much easier for historians and scholars to read and preserve earlier texts.

3. IMAGE SENSING &ACQUISION

DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

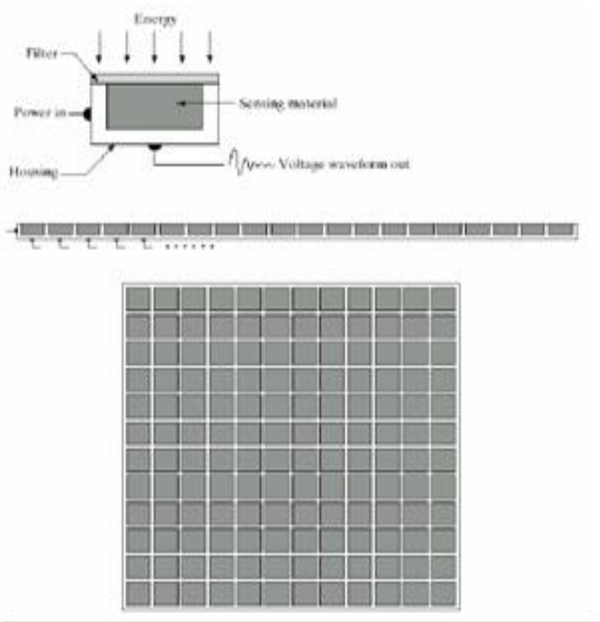

Fig. 1 single imaging, line sensor, array sensor

Let us quickly go over the steps for moving a picture to a computer's CCD camera. Instead of film, this camera uses a grid of photo sites, which are electrical devices made of silicon that make a charge when exposed to light. After that, the data is moved from the picture locations to a camera that is linked to a computer using the right type of storage media. Because hardware, like a frame-grabbing device, is faster and more efficient than software, it is often used for this reason.

This lets a lot of pictures be taken quickly, since it only takes ten thousandths of a second to take one picture. The picture in Fig. 1 shows three basic types of sensors that can turn light into computer images. The idea is simple: sensor material that responds to the type of energy being monitored changes incoming electricity into a voltage. The output voltage waveform shows how the sensors responded. It is then turned into a digital number for each sensor. Within this part, we will look at the main ways that images are made and found.

4. INTENSITY TRANSFORMATION

There are times when changing the brightness of black pictures can be helpful. You could change the black-and-white levels or make the difference between the darks and lights stronger, for example. One way to make changes to the intensity is to draw attention to the differences between different intensity numbers. One way to use this is to find things in a picture. For instance, the next few pictures show how the change affected the image's power by showing the difference between the picture before and after the change.

Different Types of Transformation Functions

- Using c*log(1+f) to make logarithmic changes in negative print
- The value of gamma has been changed by using im tune.

Imagenegatives

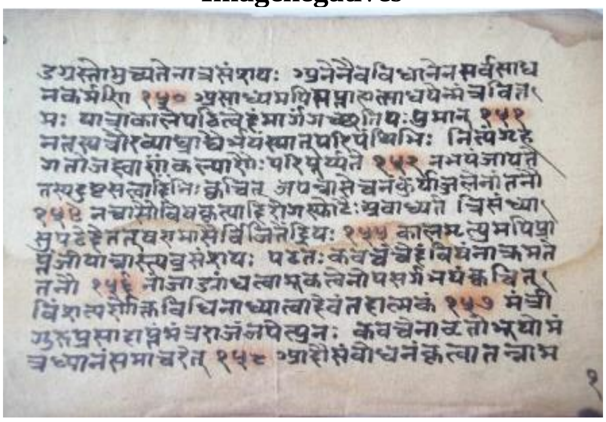

Fig. 2 Original Image

DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

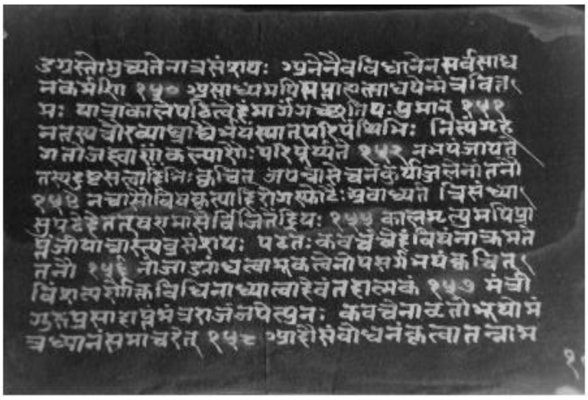

Fig. 3 Negative Image

Log Transformations

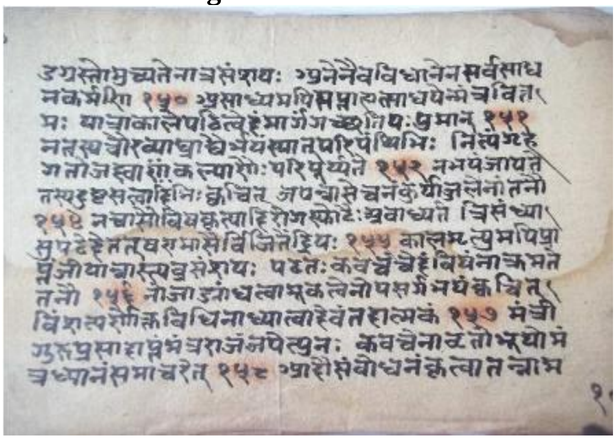

Fig. 4 Original Image

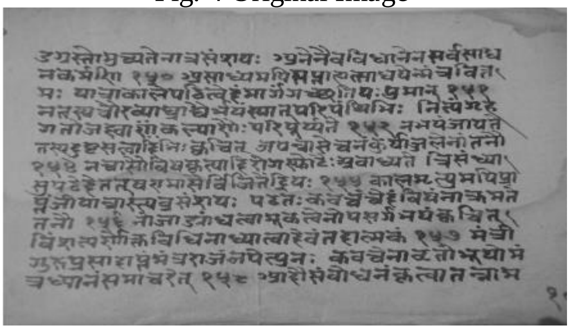

Fig. 5 Original Image

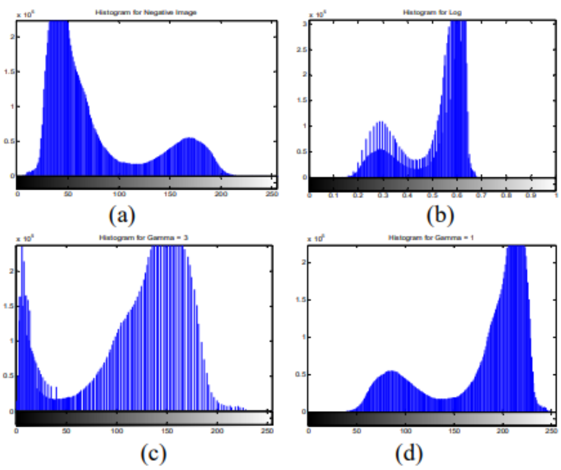

Fig. 6 Histogram of (a) Negative Image, (b) LogI mage, (c)Gamma=1, (d) Gamma=3

Gamma Transformations

DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

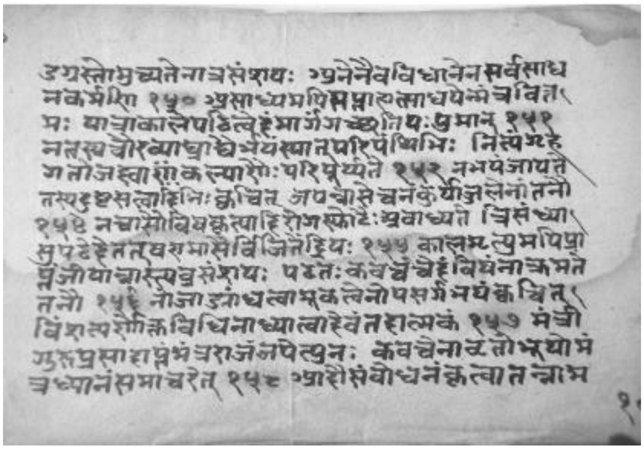

Fig. 7 Image with Gamma=1

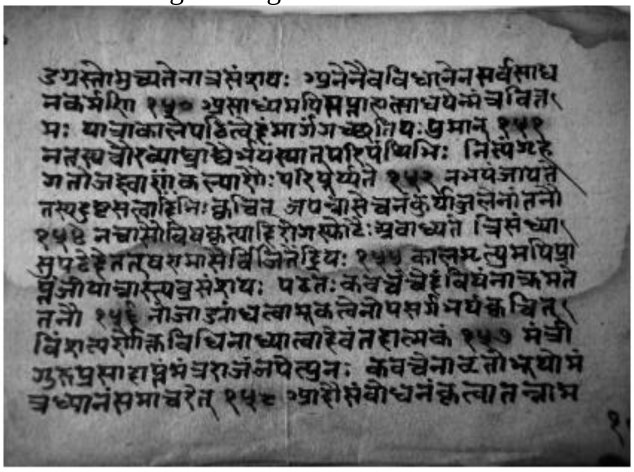

Fig.8 Image with Gamma=3

5. REMOVING NOISE FROM IMAGES

Paper noise is grouped by whether it depends on the information underneath or doesn't. You can't change the size of ink blobs, stray marks, marginal noise, or salt-and-pepper noise. It is harder to describe content-dependent noise because it is mathematically non-linear and multiplicative. This kind of noise is known as "regular noise." Two common types of noise were looked at in this study: Gaussian noise and salt and pepper noise. A normal filter is used with medfilt2 to cut down on salt and pepper noise. This noise is made up of random black or white images, which are the data extremes. It is used in both to select the 3-by-3 neighborhood.

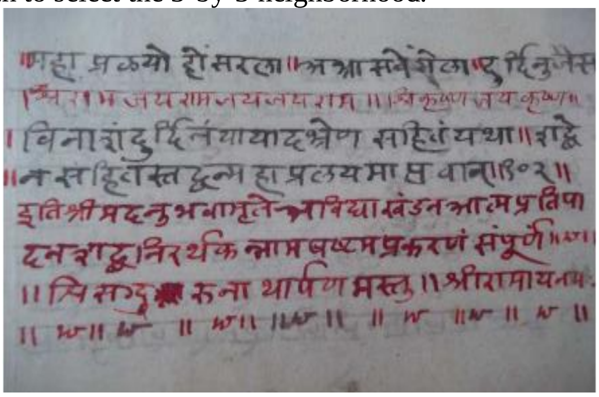

Fig. 9 Original Image

DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

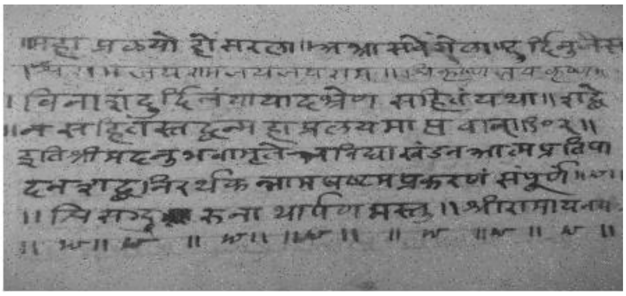

Fig. 10 add noise in it

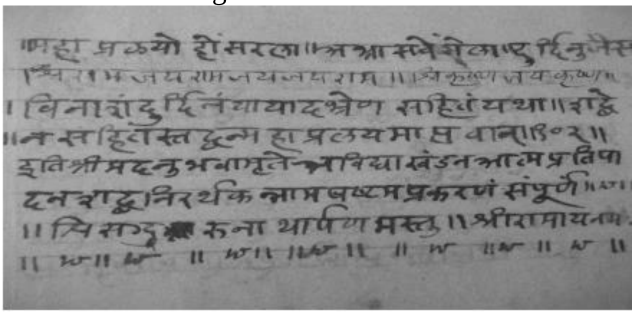

Fig. 11 Filter the noisy image with a averaging filter

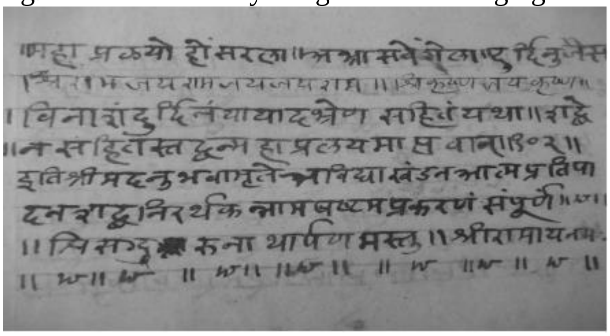

Fig.12 now uses a median filter to filter the noisy image and display the results. Notice that medfilt2 does a better job of removing noise, with less blurring of edges.

Gaussian Noise

Changes in at No order Noise that is Gaussian is sometimes called normal noise or impulsive noise. Statistical noise type T is sound waves that look like Gaussian curves. Normal distributions have a Gaussian probability density function that looks like random noise. So, Gaussian noise makes a bell-shaped curve in the frequency range.

Here's how to find sample noise with a normal distribution:

$$P(x)=1/(\sigma\sqrt{2}\pi)*e^{(x-\mu)^2}/2\sigma^2 \quad \text{For} \quad -\infty \quad \text{to} \quad +\infty$$

P(x) shows picture noise that is spread out in a Gaussian way, with a mean and a standard deviation of ± and s, respectively.

Salt-and-pepper Noise

Noise that sounds like salt and pepper, aggressive, or spikes may be heard. There are both bright and dark dots in a picture with salt-and-pepper noise. Digital recording and transmission are the main things that cause pepper and salt noise. To put it another way:

$$I(t) = (1-e)\,S(t) + e\,N(t) \quad (2)$$

If e is between 0 and 1 and P is a chance, I(t) is salt-and-pepper noise, N(t) is bright pixels in dark areas, and S(t) is dark pixels in bright areas. There will be salt and pepper noise in the picture if there are some black or white pixels. Adding noise to a picture with the imnoise function in the toolkit can make the problems listed above seem more real. This part explains how to use this tool.

6. IMAGE FILTERING ALGORITHMS

In image processing, filters are often used to make pictures smoother by lowering their high-frequency content or to make their edges stand out by reducing their low-frequency content. In both the frequency and spatial domains, Fourier transforms can make pictures better and restore them. Spatial noise removal is easier than frequency noise removal because it takes less time to process. Point processing and mask processing are both types of spatial processing. Point processing lets you change single pixels in a picture without changing the whole thing. These easy

International Journal of Early Childhood Special Education (INT-JECSE)
DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

fixes can fix problems with the camera's hardware or fix photos that are too dark or too bright. Mask processing takes a pixel and the pixels around it in a square or circle and uses them to make a better picture pixel at (x, y) coordinates. It costs more, but the method works better than point processing. When you mask a picture, a new image with the same dimensions is made. Noise reduction methods need to get rid of noise while keeping edges. The specs call for two screens, each with pros and cons. Filters can be either linear or not linear. Linear filters can handle more data more quickly, but they can't keep edges. Nonlinear filters, on the other hand, keep the edges but work more slowly with data.

Median Filter

Of course, getting rid of noise is the first step in signal or picture processing. Pollution can be cut down with neighborhood average. Neighborhood average can cut down on out-of-range noise, but it can also hide sharp edges and sudden changes. Use the median filter to get rid of isolated noise that doesn't blur the edges. The middle pixel value is changed by median filtering after numerical sorting.

How the median filter works: How to find y [m, n]:is equal to the median of x[i, j], where i and j are points in the picture that are in the neighborhood of (m, n), and y is a pixel position. The x and y coordinates of pixel y are m and n, which can be found by writing y[m,n]. W is the set of pixels around the pixel at (m, n).This area around (m, n) has (i, j) in it.The median method can find the middle value of all the pixels in the range (i, j), which is X[i, j].

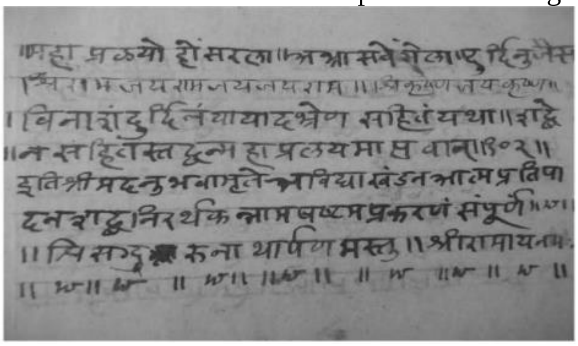

Fig. 13 Result by Median Filter

Wiener Filter

A picture that was hidden by a low pass filter can be shown again using deconvolution restoration with inverse filtering, which is also called extended inverse filtering. Underlying filtering is harmed by additive noise. We can use incremental degradation reduction to make restoration algorithms for all kinds of damage and then mix them. Wiener filtering makes inverse filtering and noise reduction work better. Extraneous noise and blurs are instantly taken away.With Wiener filtering, the mean square error is the best it can be. Total mean square error is lowered by noise smoothing and inverse filtering. The Wiener filter looks at the source image in a straight line. The process is based on a system of random variables. This formula is based on the orthogonality principle of the Wiener filter in the Fourier domain:

$$w(f_1,f_2)=\frac{H^*(f_1,f_2)S_{xx}(f_1f_2)}{|H(f_1f_2)|^2\,S_{xx}(f_1f_2)+S_{\eta\eta}(f_1f_2)}$$

The power spectrum of the original picture is shown by the blurring filter S··(f1,f2), and additive noise is shown by Sx(f1,f2). There are two separate parts to the Wiener filter: noise smoothing and inverse filtering. To remove noise, you use compression, which is also known as high pass filtering, and to deconvolution, you use inverse filtering.

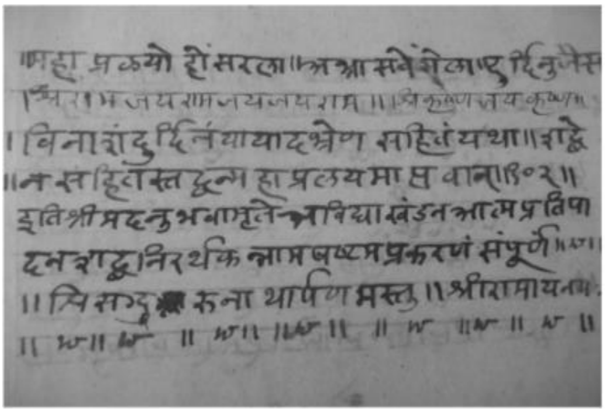

Fig. 14 Result by Wiener Filter

Average Filter

Mean filtering quickly and clearly evens out photos and lowers the difference between pixels' intensities. When you use average filtering, the value of each pixel is changed to the average of the pixels around it, which includes the picture itself. Find the mask for each image pixel. This is the best way to do it. All pixels that are hidden by a mask will be treated as separate pixels. Values of pixels that don't match their surroundings are taken away. The average filter, the mean filter, and the convolution filter are all the same thing.

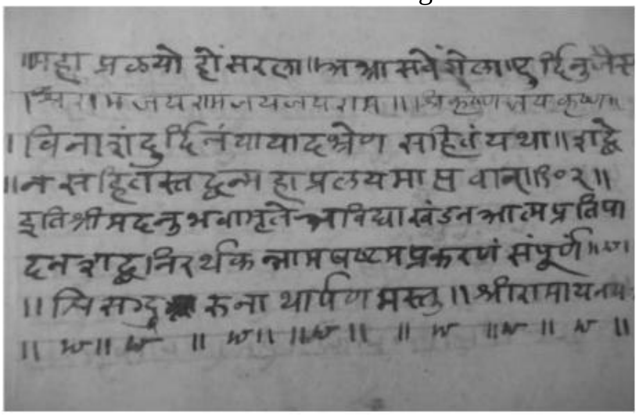

Fig.15 Result by Average Filter

Laplacian Filter

The Laplacian filter finds lines in images. It focuses on quick changes in intensity to make a mark around the world. The Laplacian function is sometimes used to make noise less noticeable after a Gaussian filter has been used to smooth out a picture. As normal, the operator puts in a grayscale image and sends out a different one. The computational cost of this method goes up as the picture radius of interest gets bigger.

Prewitt Filter

The Prewitt filter operator is used to find edges in picture processing. The discrete differentiation operator Prewitt filter comes close to the picture intensity function gradient.

Fig. 16 Result by Prewitt Filter

At each image point, the Prewitt operator makes the gradient vector norm. It doesn't cost much to compute the Prewitt operator because it distorts the picture with a small, separable, integer-valued filter along both the horizontal and vertical axes. Its gradient estimate for high-frequency photos, on the other hand, is not perfect. The Prewitt filter takes the square roots of two 3x3 matrices and finds their root mean square.

7. CONCLUSION

Five different techniques were used in this study to reduce the Salt & Pepper noise in a Devnagari handwritten paper during picture taking, transmission, or time. Because of the low contrast, this background noise made it hard to read the letters. We used seven different picture performance ways to test the filters at different noise levels. The average, median, and Wiener filters work better than the Laplacian and Prewitt filters, according to data from experiments. The median filter does a better job of getting rid of salt and pepper noise.

REFERENCES

1. Patel, A., & Sharma, M. (2021). Image Processing Techniques for the Preservation of Historical Manuscripts. Journal of Digital Heritage Preservation, 7(1), 45-58.
2. Zhang, W., & Liu, Y. (2023). Automated Image Restoration for the Preservation of Ancient Texts. Journal of Archival Science and Technology, 12(2), 99-113.
3. Singh, R., & Gupta, N. (2022). Machine Learning-Based Image Enhancement for Historical Document Preservation. International Journal of Heritage Imaging, 10(3), 178-190.

DOI: 10.48047/intjecse/V14i7.412 ISSN: 1308-5581 VOL 14, Issue 07 2022

4. Thakur, P., & Mehta, S. (2022). Analyzing the Effectiveness of Image Processing in Restoring Faded Documents. Journal of Historical Documentation Technologies, 14(4), 210-223.
5. Patel, K., & Desai, P. (2022). A Comparative Research of Image Processing Methods for Historical Document Preservation. Journal of Cultural Heritage Informatics, 25(1), 63-77.
6. Ali, M., & Rahman, Z. (2022). Enhancing Legibility of Ancient Documents through Image Processing Algorithms. International Journal of Image and Vision Computing, 9(2), 112-124.
7. Chowdhury, T., & Kumar, R. (2021). Digital Restoration of Ancient Texts Using Image Enhancement Techniques. Journal of Digital Preservation and Archiving, 6(3), 87-100.
8. Bhat, H., & Soni, V. (2021). Image Processing for Document Preservation: Techniques and Challenges. Journal of Digital Archiving and Library Sciences, 20(5), 102-115.
9. Verma, S., & Shah, H. (2021). The Use of Optical Character Recognition in Historical Document Preservation. Journal of Computational Imaging and Preservation, 18(4), 56-70.
10. Liu, J., & Chen, T. (2020). Image Segmentation for the Preservation of Fragile Ancient Texts. Journal of Preservation Technology, 8(2), 142-155.
11. Gupta, A., & Rani, S. (2020). A Survey on Image Processing Techniques for Historical Document Restoration. International Journal of Digital Archives and Libraries, 15(3), 98-111.
12. Rathod, S., & Agarwal, A. (2020). High-Resolution Imaging for the Preservation of Historical Manuscripts. Journal of Imaging and Cultural Preservation, 22(1), 12-26.
13. Patel, R., & Jain, P. (2020). Color Restoration and Noise Reduction Techniques for Historical Document Preservation. International Journal of Imaging and Graphics, 17(4), 135-148.
14. Sharma, V., & Das, S. (2020). Multispectral Imaging for Document Preservation: A Case Research of Ancient Texts. Journal of Cultural Heritage Science, 19(3), 67-80.
15. Kumar, A., & Kapoor, P. (2020). Enhancing the Quality of Historical Documents Using Image Enhancement Algorithms. Journal of Image and Signal Processing, 23(2),