IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A NOVEL METHOD FOR REMOVAL OF NOISE IN DOCUMENT IMAGE

[1]SHASHIDHARA.B

ASSISTANT PROFESSOR

Department of COMPUTER SCIENCE,

IDSG GOVERNMENT COLLEGE, CHIKMAGALURU, INDIA

Abstract: Ancient degraded historical documents carry various important information regarding our culture, economics etc. proper restoration of these documents is very necessary. These documents still contain announces and other poor quality components after digitisation. These have an impact on how the documents look overall. In this work, an innovative approach for improving ancient historical documents is suggested. Digital image processing techniques are used to these documents in digital format to improve them. The efficiency of the techniques is then evaluated visually and quantitatively.

Index Terms - Historical document, restoration, filtering, digital image, image processing**.**

I. INTRODUCTION

Impulse noise often degrades visuals. Generally speaking, this kind of noise may be seen in digital images as a result of scanning issues, video sensor issues, decoding issues, etc. One of the most crucial pre-processing steps before feature extraction is filtering impulsive noise.

Median filters are often used to get rid of impulsive noise. A median filter takes the middle value from all the surrounding neighborhoods and uses it instead of the original central value. The adaptive median filter (AMF), the rank order based adaptive median filter (RAMF), the switching median filter (SMF), the decision based filter (DBF), the hybrid median filter (HMF), etc. are all types of median filters. While median filter versions may often improve picture quality, this is not always the case. These filters not only reduce the amount of noise in a picture, but they also have an effect on the pixels that are free of noise, creating a blurred replica of the original. As a result, a morphological image processing-based alternative is given for dealing with the impulse noise. This method, which makes use of a pixel connection mechanism, has been shown to be effective for eliminating impulsive noise in digital photographs. There are two primary phases to the removal process. The first step is to locate the source of the disturbance, and the second is to restore the picture.

In this study, we suggest a morphological method for suppressing impulse noise in digital photographs that relies on the interconnection of individual pixels. Impulse noise, often known as the salt and pepper noise, lowers the visual quality of the picture by altering the texture and tiny details in both dark and bright areas. To begin, our technique uses 4-connectivity to isolate noisy pixels in the image itself.

The technique relies on the interconnectedness of the noisy pixels inside a given window. When processing a single pixel, a window is selected that contains both noisy and noise-free neighbouring pixels. The processing pixel is classified as noisy or free of noise based on the number of connections made between the noisy pixels alone.

The suggested approach is successful in Even with 80% noise level, the image could be restored using the suggested approach. The corrupted picture is filtered using several iterative techniques, allowing for successful image restoration even in severely damaged images.

II. INVENTED METHOD

The suggested approach divides the picture into black (0), white (255), and gray shade (i.e., intensities other than 0 and 255, respectively) areas for processing. There are two stages to the algorithm. The initial step is locating the problematic pixels. There are two distinct phases within this larger process. Sub stage 1 entails locating and correcting the noisy pixels in the black and white background areas. In this sub stage, we locate the noise source and filter it out.

Stage 2's sub-step focuses on the gray-shade areas where noisy pixels are located through the connection concept. One way in which pixels may be connected to one another is via pixel connection. With this approach, we make advantage of 4-connectivity. A pixel's 4-neighbors are the four neighbouring pixels in both the horizontal and vertical directions.

The second step involves substituting an estimated value for the correct value for the noise found in the second stage. To do this, the median of the data contained inside a moving window is substituted for the noise. Since sub stage 1 handles both noise detection and removal, the second phase is often only relevant to sub stage 2.

III. DETECTION AND RESTORATION OF NOISES

Potential sources of impulse noise are identified at this point. Our technique begins by identifying the processing step at which the pixel is currently located. Here are the detailed measures of the first stage:

The neighbourhood pixels around the processing pixel f(x, y) are first extracted. The amount of noise informs the size of the mask that will be used. Better noise identification is achieved by linearly increasing both the noise density and the mask size. However, the recovered picture becomes blurrier as the mask size is increased.

An 8-bit picture (intensity values between 0 and 255) is used in the suggested technique. A check is made to see whether the central pixel is 0 or 255. If it falls into any of these two categories, it is kept in the running for becoming noise. If not, it is assumed to be noise-free, and the adjacent pixels and their surroundings are used instead.

$$N(x,y) = \begin{cases} 1, f(x,y) = 0 \ or \ 255 \\ 0, else \end{cases}$$

If N(x, y)=1, then the entries 0 and 255 in matrix G are examined. If all the values in G are between 0 and 255, then sub stage 1 is utilized for noise identification; otherwise, sub stage 2 is used.

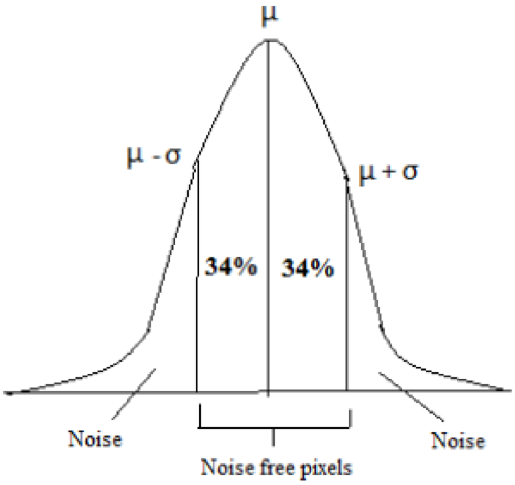
$$G = \begin{cases} sub\ stage\ 1, G(G \neq (0\,||255) = 0) \\ sub\ stage\ 2, else \end{cases}$$

Sub-stage 1: Restoring images with white and black areas

All of the numbers in the extracted window from a noise-free image's white or black section will be either 0 or 255. It important to note that if impulse noise is present, some of the 255 values in the white zone are replaced by noisy 0 values, and vice versa, depending on the noise density level. However, depending on the area, the majority of the pixels in a selected window will be either 0 or 255. Therefore, the greatest occurrence of a value of 0 or 255 in the window is substituted for f(x, y).

Sub-Stage 2: Identification of noise in gray-shaded areas

The distribution of pixels in a given neighbourhood is almost normally distributed, with noise pixels occupying only the extremes of the bell-shaped bell curve of normal distribution, while noise-free pixels are concentrated around the neighbourhood’s mean value.



The distribution of pixels inside an extracted window the variables μ and σ represent the mean and standard deviation, respectively, of the pixel values inside a certain window.

Although it may also include some noisy pixels at higher noise densities, 68% of the curve's overall area is always regarded as noise for processing according to the method. As previously suggested, a noisy pixel is recognized using pixel connectivity. Thus, connection between the pixels found inside the remaining 32% of the curve is established. There could be some noise-free pixels in this area of the curve as well. The ability to identify noise is not much hampered by this, however.

A 4-connectivity is constructed between the foreground pixels once specific requirements are met, depending on which the intensity values in the extracted window are altered to either 0 (background pixel) or 1 (foreground pixel) (1). As a result, a link has been made between the foreground pixels and the noisy pixels in the selected window.

The aforementioned technique is outlined as follows.

1. A subset H of G is identified, where G is a set with its center element removed. Subsequently, the statistical parameters of the mean (μ) and standard deviation (σ) for variable H are determined. As a result, two values, denoted as R1 and R2, are determined. The first equation can be expressed as R1 equals the mean (μ) minus the standard deviation (σ), whereas the second equation can be expressed as R2 equals the It is seen that the majority of pixels without noise are included within this range, whereas the noisy values are situated outside of this range. However, there are certain instances in which photographs exhibit significant noise density or include patches with intensity values that are closer to 0 or 255. In such scenarios, the range of noisy pixels may also encompass such values, since the mean will be centered around either 0 or 255.

2. The values of H are transformed into one of three possible values: 0, 255, or a constant K, depending on the specified circumstances. This statement operates on the premise that the values included by the range R are devoid of any noise, whereas values falling outside the range are considered to be noise.

Circumstances 1: In the scenario when R1 is negative and R2 is positive, all numbers within this range, except zero, are transformed into a constant value denoted as "k." Additionally, all values above R2 are changed to 255.

Circumstances 2: In the scenario when R1 is positive and R2 is positive but less than the maximum intensity value of 255, it is seen that all values falling within this range undergo a conversion to a constant value denoted as K. those below the threshold R1 are transformed to zero, while those over the threshold R2 are transformed to 255.

Circumstances 3: In the scenario where R1 is positive and R2 is positive but exceeds the maximum intensity value of 255, all values within the range, except 255, are uniformly replaced with a constant value denoted as K. Values that are less than R1 are transformed into zero.

3. Currently, the function H just consists of the aforementioned three values. The central pixel of the matrix G, denoted as f(x, y), is combined with matrix H to form a matrix of size (2k+1 x 2k+1). The H values are modified once again, taking into account the function f(x, y).

- Given a function f(x, y) = 0, the process involves modifying all H values that are initially zero to a value of one, followed by changing all remaining values to zero. The sequencing of value changes has significance within this context.

- Given a function f(x, y) = 255, it is seen that all H values equal to 255 undergo a transformation to become 1, whereas all other values undergo a transformation to become 0.

The aforementioned requirements and alterations in values are just implemented to ease the identification of connectedness between pixels. The 4-connectivity of foreground pixels (with a value of 1) in matrix H is determined by counting the number of neighbouring pixels (N) that are linked to the center pixel.

If the value of N exceeds a certain threshold, denoted as A, the processing pixel may be classified as devoid of noise. If a value is below the threshold of A, it might be classified as noise. The value A is contingent upon the size of the mask. In the case of a mask with dimensions (2k+1 x 2k+1), the minimum number of pixels that must be linked in order to ensure that the center pixel is free from noise is equal to (k+1) multiplied by (2k+1).

$$f(x,y) = \begin{Bmatrix} noise\ free, N > A \\ noise, \quad N \leq A \end{Bmatrix}$$

$where\ N\ is\ the\ maximum$ no. $of\ adjacent\ pixels\ connected\ to\ centre\ pixel\ based\ on\ 4 - connectivity\ and\ A = (\ + 1) * (2k + 1)$

IV. NOISE REMOVAL USING MEDIAN FILTER

After the identification of a pixel as noise, it is then substituted with an approximated value representing the appropriate intensity level. The G matrix is utilized to extract certain elements, from which the median value is computed. This median value is subsequently substituted for f(x, y).

1. For a mask of size (2k+1 x 2k+1), the following values are extracted.

$$f' = f(x + s, y + t)$$
$$where\ s, = 0\ to\ 1$$

2. The values other than 0 and 255 are found in f'. Then the median of those values is calculated which in turn is the estimated correct value.

$$f\ x, y = (f'\ (f' \neq 0\ \|255\ ))$$

3. If all the values of f' are either 0 or 255, then the upper limit of sand t is incremented by 1. If still the function f' has no values other than 0 and 255, then the limit is again incremented by 1 and this increment happens till K.

$$f' = f(x + s, y + t)$$
$$where\ s, = 0\ to\ p, p = 0,1, \ldots\ k$$

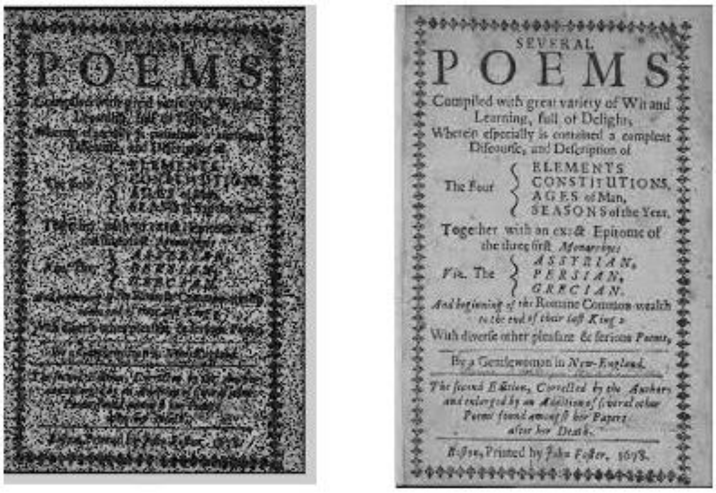
And $f\ (x, y) = (f'(\ f' \neq 0\ \|255))$.

V. EXPERIMENTAL RESULTS

Simulations are conducted on a range of standard images under varying amounts of noise. The evaluation of this scheme is conducted by utilizing the PSNR parameter and comparing it with the PSNR values of alternative approaches. The results demonstrate that our method outperforms the other ways.

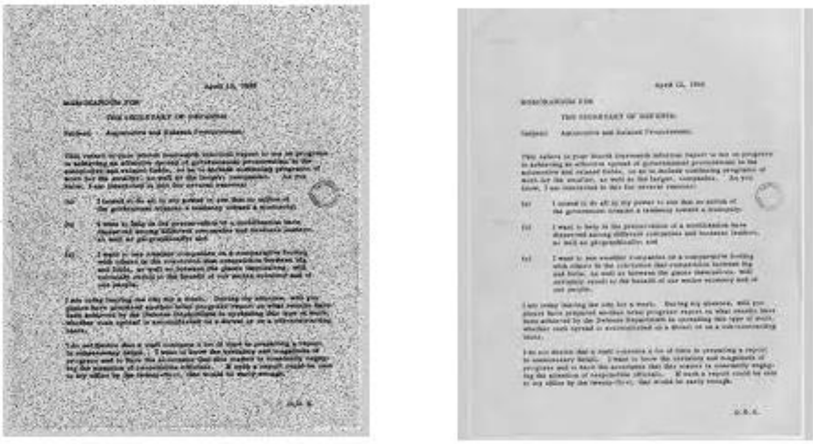
The selected images for testing purposes adhere to standard criteria. These images are grayscale and have a bit depth of 8, with dimensions of 512x512 pixels. The size of the mask used is determined based on the desired noise density level, with a corresponding value of K set to 2. In the second sub stage of identifying noisy pixels, the value of A is assigned as 15. During the processing of a specific pixel in the image, if the number of adjacent pixels connected to the processing pixel, after window modification, exceeds 15, the pixel is classified as noise-free. The quality of the filtered image is assessed using the peak signal-to-noise ratio (PSNR) parameter. In this particular case, the PSNR value obtained is 30.627 dB.

The findings of our experiment are presented in the subsequent graphs and table. The suggested approach demonstrates the ability to effectively eliminate noise while also maintaining the integrity of edges and other intricate picture features. The filtered images exhibit a high level of visual quality.



a) corrupted image b) restored image

Figure 1: simulation results of the image



a) corrupted image b) restored image

figure 2: simulation results of the image

Table 1: PSNR (db) values for image (gray) using MEDIAN FILTER

Noise Density	Window 3*3	Window 5*5	Window 7*7	Window 9*9
20%	15.2213	14.7997	14.9381	14.7296
30%	13.5846	13.1733	13.2755	13.1855
40%	12.5306	12.1409	12.2885	11.9934
50%	11.8135	11.3985	11.4818	11.2967
60%	11.1152	10.8815	10.9387	10.7880
70%	10.7174	10.3916	10.4939	10.2939

VI. CONCLUSION

In addressing the issue of high density salt and pepper noise, a morphological methodology that relies on pixel connection is employed. Connectivity is formed between the central pixel and its nearby pixels, and the determination of noise is dependent on the number of linked pixels. The efficacy of this system is evaluated by conducting tests on a range of standard photographs. The algorithm under consideration yielded favourable outcomes in comparison to alternative approaches. Moreover, it is observed that this approach effectively retains all the texture and edge information included in the image.

REFERENCES

[1] W. K. Pratt, Digital Image Processing. New York: Wiley Interscience, 1991.

[2] Toh, K.K.V. and Isa, N.A.M., 2010. Noise Adaptive Fuzzy Switching Median Filter for Salt-and-Pepper Noise Reduction. IEEE Signal Processing Letters, vol. 17, no. 3, pp. 281-284.

[3] Nair, M.S., Revathy, K., and Tatavarti, R., 2008. Removal of Salt-and-Pepper Noise in Images: A New Decision-Based Algorithm. In Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. I.

[4] T. Nodes and N. Gallagher, “Median filters: Some modifications and their properties,” IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-30, no. 5, pp. 739–746, Oct. 1982.

[5] W. Luo, “An efficient detail-preserving approach for removing impulse noise in images,” IEEE Signal Process. Lett., vol. 13, no. 7, pp.413–416, Jul. 2006.

[6] P. Civicioglu, “Using uncorrupted neighbourhoods of the pixels for impulsive noise suppression with ANFIS,” IEEE Trans. Image Process., vol. 16, no. 3, pp. 759–773, Mar. 2007

[7] Esakkirajan, S., Veerakumar, T., Subramanyam, A.N. and PremChand, C.H., 2011. Removal of High Density Salt and Pepper Noise Through Modified Decision Based Unsymmetric Trimmed Median Filter, IEEE Signal Processing Letters, vol. 18, no. 5, pp. 287-290.

[8] Gonzalez, R.C., and Woods, R.E., 2004. Digital Image processing (2nd edition), Pearson Education.

[9] Xin Zangju, Wang Shoujue, Deng Haojiang, Luo Yujing,“A new filtering algorithm based on extremum and median value”, Journal of Image and Graphics, 2001,6(6),pp.25-28

[10] Gou Zhongkui, Zhang Shaojun, Li Zhongfu, Jin Jian, “New adaptive median filter algorithm based on extreme value”, Infrared and LaserEngineering, 2005, 34(1), pp.98-101.

[11] H. Chan, H. Chung-wa, and M. Mikolova, “Salt and pepper noise removal by median type noise detectors and detail-preserving regularization,” IEEE Transactions on Image Processing, vol. 14, pp.1479-1485, Oct. 2005.

[12] A. K. Jain, “Fundamentals of Digital Image Processing,” Englewood Cliffs, NJ: Prentice-Hall, 1989.