International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 08, August-2018

AN EFFICIENT ALGORITHM FOR REMOVING SALT AND PEPPER NOISE FROM DOCUMENT IMAGES BY USING CONNECTIVITY PRINCIPLES

[1]Karasala Surya Prakash, [2]Dr.Amandeep Kaur

[1]*Computer Science and Technology, Central University of Punjab,*
[2]*Computer Science and Technology, Central University of Punjab,*

Abstract—The Proposing method improves the quality of degraded document images which are effected by salt and pepper noise. Noise effects the Document image quality and further processing. Thus to improve the quality of image we need to supress the noise. There are a number of methods to reduce different types of noise in document images. The proposing algorithm focusing on detecting and eliminating salt and pepper noise from infected document image. This proposing method works based on pixel connectivity principles and labelling methods these methods helps to detect salt and pepper noise in the image. If pixels are connected in different label values then those pixels are connected and it is noise free. The results of the proposed technique are improved in terms of PSNR, SSIM, and MAE.

Keywords— PSNR, SSIM, Salt and Pepper noise, connectivity

I. INTRODUCTION

At the present generation, computers, mobiles and electronic devices usage increased in people lifestyle, in old day's people use paper documents for office work, business, etc. maintaining hard copy of documents for long time is very difficult it have chance to lost documents in accident's and transform document also become difficult and time taking etc. Scanning documents is a method of changing printed documents into digital format. There are different kind of problems in process of converting document into document image like skew, blur, noise, brightness etc and one of common problem met when scanning documents is 'noise' which can happen in an image because of paper quality, the typing device used, hardware failure, or it can be formed by scanners during the scanning process. Impulse noise one of the kind of noise which effect document image. Mostly this noise caused by sensor problem. In document Images text should be either handwritten or machine-printed text. These document images corrupted during in the process of transmission, registration, scanning, or storage of the text. Thus, Denoising is an significant action before performing any advance image-processing job such as image compression, coding, and pattern recognition. Document images effected by different kinds of noises like Gaussian noise, speckle noise, salt and pepper noise and periodic noise. This paper concentrating on salt and pepper noise in document images and proposed method detecting noise efficiently and restore image from noise.

II. LITERATURE

The literature specifies many other procedures to denoising document images. Among these procedures two kinds of filtering procedures are there one is linear and another is nonlinear filter. Linear filters such as Wiener, and Kalaman filtering Techniques. (Sanjay Chandra Arya)Wiener filter is very oldest method it was used to reduce the noise in signal and it was proposed by Norbert Wiener during the 1940s and published in 1949. Later this filter can study by so many people for the different research work. The wiener filter had ability of handling both the degradation along with noise. It diminishes mean square error and noise but still it produces some error in image results.
(KALMAN)Kalman algorithm gives the Solution for the Wiener Problem. With the state-transition method .it describing a recursive solution to discrete data linear filtering problems it gives the best results to many tracking and data prediction tasks. Kalman filter is standard method since it generate good results in practice due to optimality and structure. It is Suitable for online and real time processing. Easy to convey and implement given a basic understanding. Measurement equations there is no necessity for inverting. Kalman filters used for only linear state transitions. For nonlinear problems we can use advance one like particle filters (SMCs). Kalman Filter most significantly only with those sensor's output can be used, who outcomes in Gaussian form output. There is an Extended Kalman Filter (EKF) which is used for non-linear classifications, but it failed if the functions are extremely nonlinear (1st order approximation). (Kaur)Even though linear filters are easy to implement, but they blur the edges of the image, equally increasing the noise and the data, and cause a losing of image details. Therefore, many other non-linear filters are there, which are based on optimization principle, median filtering, and morphological technique are proposed to restore the document images .nonlinear Techniques preserves image details.

A new adaptive operator, which forms predicts based on the differences between the current pixel and the results of centre-weighted median filter with various canter weights which formulated by (Tao Chen). Extensive simulations demonstrating that the suggested method reliably works well in removing both types of impulses at different noise ratios. ACWMF faces difficulties to recover image from corrupted version of image. An innovative nonlinear filter that integrates the standard median (SM) filter and the centre weighted median (CWM) filter for detecting noise in corrupted image. This tri state filter generating efficient results but also generating unwanted lines over the image so reading content is difficult. (Hancheng Yu) Proposed rank-ordered relative differences (RORD) statistic impulse detector. Different from ROAD (rank-ordered absolute differences) detector, this impulse detector recovers the impulse noise detection exactness by using a reference image. Then they introduce a simple weighted mean filter to suppress the impulse noise while maintaining image details but performance is low while comparing PSNR and SSIM values with other algorithms. (Yijiang Shen) A novel optimization approach using signomial programing to restored noise binary and grey scale images .it consuming more time for image restoration.

(Dr.V.Radha) Median filter is familiar in nonlinear filters for removing Salt & Pepper noise. Salt and pepper noise is supressed by substituting the window centre value by the median value of centre region in compound images (combination of text, pictures, graphs) .From the performance examines the median filter gives well results for compound document images compared with scanned compound images .Newer techniques like various DWTs (Discrete Wavelet Transforms) have been newly used to remove salt-and-pepper noise from an image. Over the earlier few years wavelet transforms have been used for image de-noising. A discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. An important advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information. The DWT of image signals produces a non-redundant image representation which provides better spatial and spectral localization of image formation compared with other multi scale representations such as Gaussian and Laplacian pyramid. The DWT can be interpreted as signal decomposition in a set of independent, spatially oriented frequency channels. Non-linear filters, in general, produces better results in reducing salt-and-pepper noise. Among linear filters contra -harmonic mean filters give good performance. In non-linear filters, adaptive median filter and decision based median filters produce very good results. Recently, discrete wavelet transforms have also been used to remove salt-and-pepper noise.

(S.Awad) To restoring document images that have been corrupted by salt and pepper noise. This algorithm consists of a number of shells and is based on the connectivity principle; whereby, pixels that have connections with at least one other pixel in each of the consecutive shells are deemed noise-free pixels. The new Denoising method is not only useful in eliminating the noise efficiently; it also keeps the fine details of the image. The proposed algorithm is compared with well-known algorithms such as efficient procedure for removing impulse noise (EPRIN), adaptive centre weight median filter (ACWMF).while comparing author experimented results the peak signal noise ratio and structural similarity the proposed method is giving best performance .few new methods also had some drawbacks discreet wavelet transform (DWT) and computation of block prior probabilities also have problems like which they do not handle properly with text and their time execution very expensive. There are many other proposed filters used mainly to restore random valued impulse noise. The proposed non-linear filter method tackles mainly one of the worst noise types, which is random-valued impulse noise, in addition to salt and pepper noise. Similarity and connectivity principle is one of the best method to restore document image from noise. But still it also had problem at higher level noise it performance like older ones. At lower noise ratios it performs very well but at higher values this algorithm performance is not well.

III. RESEARCH METHODOLOGY AND PROPOSED METHOD

There are several proposed filters used for mainly to restore images from Salt and Pepper noise. The proposed non-linear filter method holds mainly one of the worst noise types, Noise in Document images frequently occurred due to hardware or software or storage problems. Document images effected by different types of noise specifically we are talking about salt and pepper or impulse noise in document images. Here I am using connectivity principle to detect noise in document image.

A. Connectivity:

Relation between pixels is called connectivity. Here we have different kind of classification of connectivity between pixels those are 4-connected, 8-connected, m-connected pixels. In Four connectivity considered only neighbour pixels left, top, bottom, right sides values only it will treated as 4-connected.diagnol pixels not considered

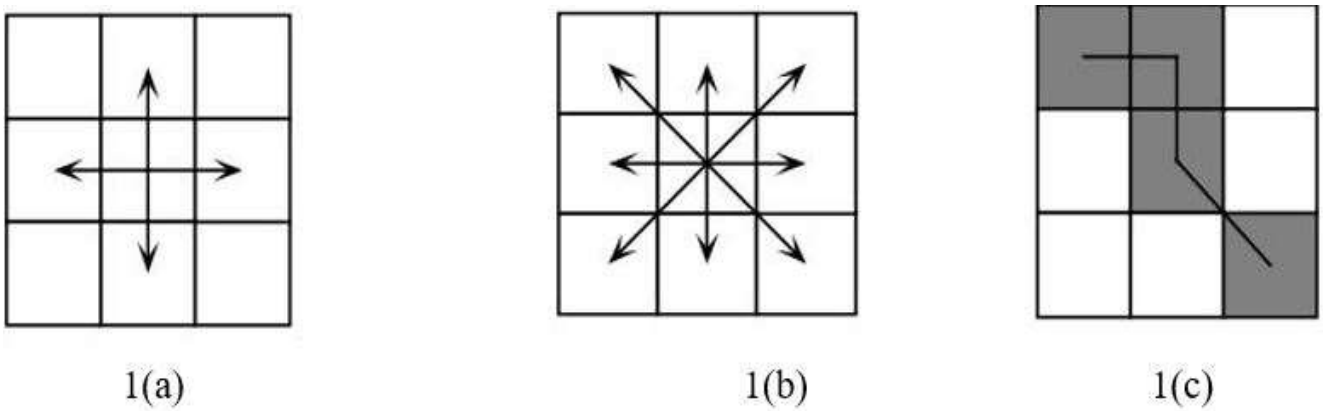

Figure 1: (a) 4-connected pixels, (b) 8-connected pixels, (c) m-connected pixels

Considering 4-connected pixels and diagonal pixels called 8-connected as shown in figure 1(b).Another connectivity mixed connectivity a usage of 8-connectivity that reflects diagonally-adjacent pixels to be connected if no shared 4-connected neighbour exists shown in 1(c).

B. Data Set

Created Document image data set. Image 1 and image 2 are considered for experiment results and Comparative Analysis**.** Images are shown in below .Figure 2 is single string document image .Figure 3 is English hand written page.

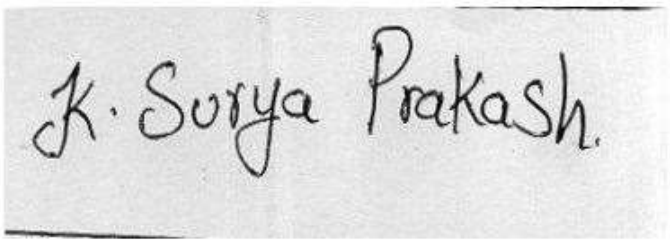

Figure2: Image 1

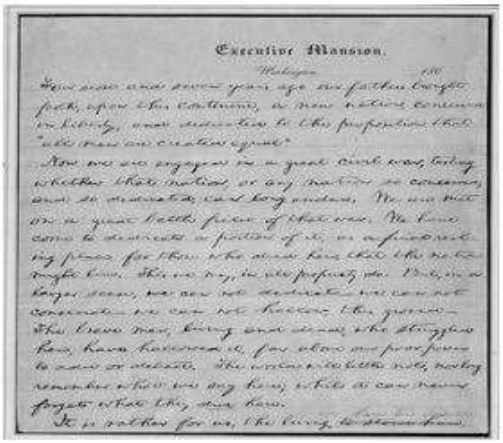

Figure 3: image 2

C. Algorithm

Proposing method is improved quality of image results compare with previous methods. In this new method identifying noise in mage based on connected pixels in image. In this process there are several steps are there. In the process of detecting noise in image first step is dividing image into different portions based on set of threshold (intensity) values and analysing about connected pixels. For dividing image into different pieces we requisite a threshold routine.in threshold function it takes input grayscale image , a minimum intensity value, and maximum intensity value .This function returns output as binary image. Next we need to process label method. Which it will take binary image as input and testing pixel, a labelling image, and label value. The function produces the label image with "true or white" area in a binary image, initially at the testing pixel, filled with the value of label. This function uses the notation of -1 denoting an unvisited pixel in the labelling image. First the testing pixel is filled with value of label and store it in a list. If list is not empty then we continue inspecting neighbours by using 8-connectivity. If it visits a neighbour pixel which is unvisited pixel (-1) if that pixel is true in binary image and update the label of that pixel to the label value and it will add to the list.

After this step three is finding connectivity between pixels .This connected pixel function takes input binary image and label image structure. The structure has three fields: a 2D array which is the label image, an array of labels sizes, and the maximum label value. The sizes of array tell you how big (number of pixels) a certain pixels is, and is indexed by label number. Then connected pixels function works by looping over the pixels of the binary image .if it visits unvisited pixel -1 if that visited pixel achieves is " true or white" for that pixel altered with 1 , label function is called with label value ,connectivity, image label, number of rows and columns. If visited or testing pixel achieves is false when it visits neighbourhood pixels, that pixel is altered with zero in labelled image(no connected pixel).once connected pixels is finished , the labelled image will have no -1 pixels .every pixel either it having 0 or positive integer. This process will run multiple times.

```
set input image  to all -1's // -1 for unvisited pixels

for s=1 to n – 1 // n= total threshold values
   minimum = thresholds(i)+1
   maximum = thresholds(i+1) //maximum label value
   binaryimage = threshold_range(minimum, maximum)
   change all 0's in label_image to -1's
   labeledimage = connectivity(binaryimage, labeledimage)
end
```

Algorithm for finding connectivity between pixels

After finding number of connected pixels then we define threshold value. Connected pixels are noise free pixels which are not connected those pixels are noise pixel. if any labelled pixel is less than threshold value its then get its index value and replace by median value by applying median filter this will go for loop until last pixel of matrices is reached. if it is noise less no need replace that pixel value.

```
set matrices equal to all zeros

for i=0 to row size
   for j=0 to column size
     loop over neighbour pixels  and increment matrices column and  rows
   end
end

for i=0 to row_size
   for j=0 column_size
      if (labelled_pixel_value <threshold_value)
         aplly medianFiletr on that noise pixel
         update label_vlaues  to that noise pixel which
         update matrices
         update pixels_values
      end
   end
end
```

Algorithm for Detecting and eliminating noise

III. RESULTS

The proposed method has been compared with existing methods in terms of visual and quantitative measures. Three images from simulated data set Figure 2, Figure 3 are considered as testing images for experimental usages. These testing images are degraded with noise and quality of restoration as well as quality of restoration as well as quality after noise removal are computed. All experiments were performed using Matlab2017b

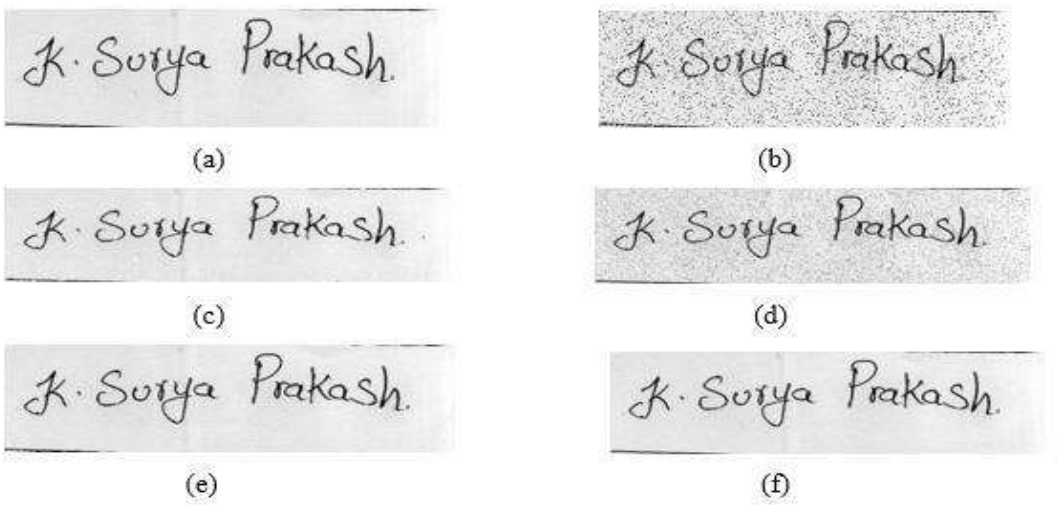

Figure 4: Experimental results for Figure 2 at 10% noise (a) original document Figure (b) Document image with 10 % salt and pepper noise (C) After applying Adaptive centre weighted median filter output (ACWMF) (d) after applying EPRIN (efficient procedure for removing random valued impulse noise in images) method output noise free image showed in above (e) After applying SCP (similarity and connectivity principles based on shells) (f) Applying proposed method

In above observation all most all techniques removed noise efficiently except EPRIN method.at low level of nose even ACWMF also performing very well it gives good output. For concluding best method, we should have to test images in various noise levels which is giving better result in high percentage of noise in images that is efficient method.

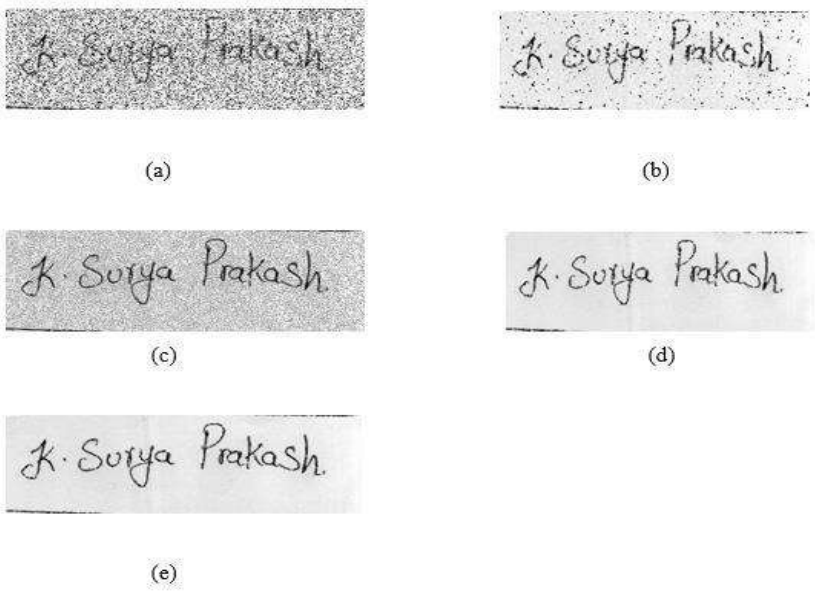

Figure 5: Experimental results for Figure 2 at 40% (a) At 40% noise (b) ACWMF (c) EPRIN (d) SCP (e) Proposed method

In above Figures 16 results ACWMF still had some amount of noise it is giving good results at lower level noise values when noise ratio increases ACWMF performance decreases. EPRIN giving background disturbances. scp (present method) and proposed method still good. Now image1 tested at 60% of noise level.

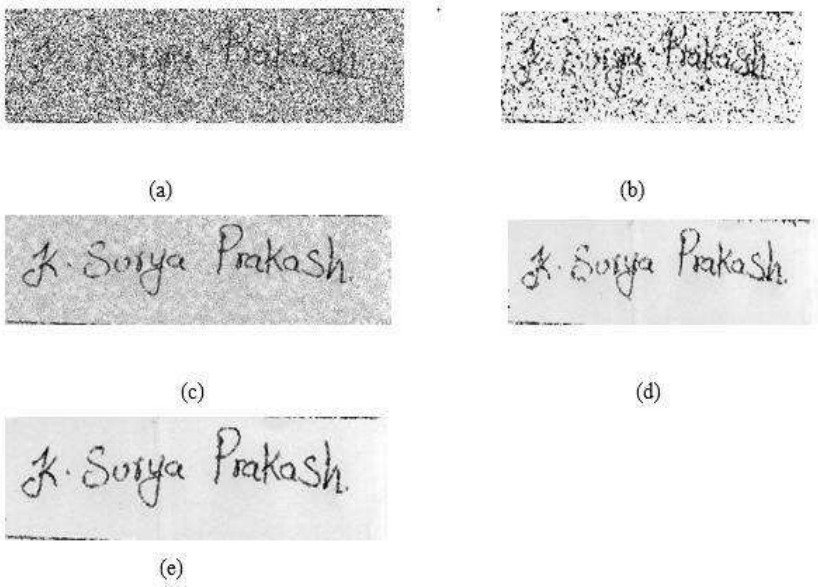

Figure 6: Experimental results for Figure 2 at 60% noise (a) at 60% noise (b) ACWMF (c) EPRIN (d) SCP (e) Proposed method

At 60% noise proposed method given better results than ACWMF, EPRIN and SCP by observing above image results. In similar way to testing proposed method efficiency on multiple text lined image or paragraph image as shown in below. Same experiment conducted on figure 3 to get know about proposed method performance.

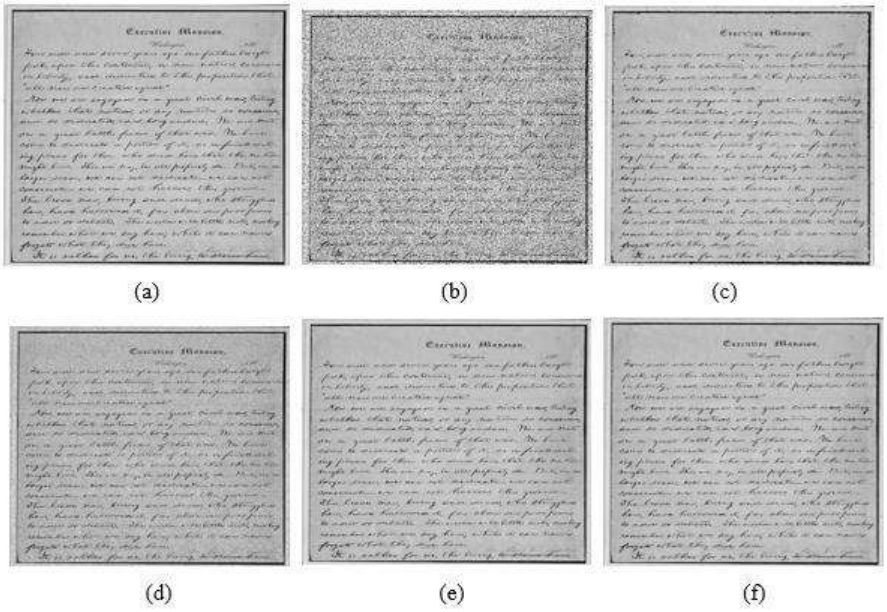

Figure 7: Experimental results for Figure 3 at 20% noise (a) Original Document image (b) at 20 % Salt and Pepper noise (c) ACWMF (d) EPRIN (e) SCP (f) Proposed method

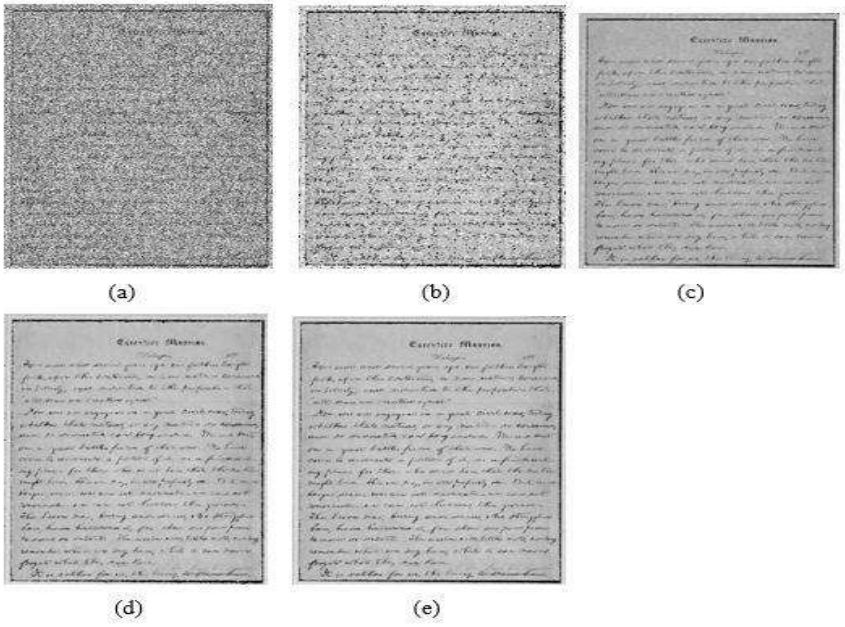

Figure 8: Experimental results for Figure 3 at 50% noise (a) at 50% noise (b) ACWMF (c) EPRIN (d) SCP (e) Proposed

At higher noise levels also proposed method giving good results. Proposed method is improved quality of image compare to SCP algorithm. Compare results with qualitative metrics.

IV.COMPARITIVE ANALYSIS

Comparing each algorithm performance by using PSNR, SSIM values of each experimented image

PSNR: the ratio between power i.e. maximum possible value of signal and the power of corrupting noise which effects the quality.it is expressed in terms of decibel.it is calculated as

PSNR=10*log10(MAX*MAX/MSE)

Where MAX=maximum possible value of the image

MSE: Mean square Error which is the average of square of errors

$$MSE=1/(m*n)\sum_0^{m-1}\sum_0^{n-1}\|f(i,j)-g(i,j)\|^2$$

Where f (i, j) = original image, g (i, j)=degraded image ,**m** represents number of rows and **i** is index of that row and **n** represents no of column and **j** is the index of that column

SSIM: this term is Structural Similarity Index measure. SSIM measures image quality.it used to compare the visual quality of image obtained from proposed image and original image. The measure between two windows x and y of common size M*M is

$$SSIM=\frac{(2\mu_x\,\mu_y+C1)(2\sigma_x\,\sigma_y+c2)}{(\mu_x^2+\mu_y^2+c1)(\sigma_x^2+\sigma_y^2+c2)}$$

The experimented image is Figure 2 and this image tested under 10% noise to 70% in all methods. By observing above psnr data PROPOSED method highest PSNR value it gives good result. But in case of SCP initially it gives good values by increasing noise ratio performance decreased at 70% noise level EPRIN and SCP almost closer values there is no much big difference but proposed one it given better values. By observing chart diagram also we can say proposed method is giving better results

Table 1: PSNR values of Figure 2

Noise level→	10	20	30	40	50	60	70
EPRIN	19.45961	16.56136	14.79273	14.02642	14.14089	15.49191	17.90355
ACWMF	27.21198	23.87924	19.82035	16.29513	13.28046	10.41522	8.349292
SCP	28.62399	26.95598	25.31408	2135726	22.08214	20.40551	18.83593
PROPOSED	35.06318	31.38406	28.63204	26.96147	25.27993	2119848	21.99269

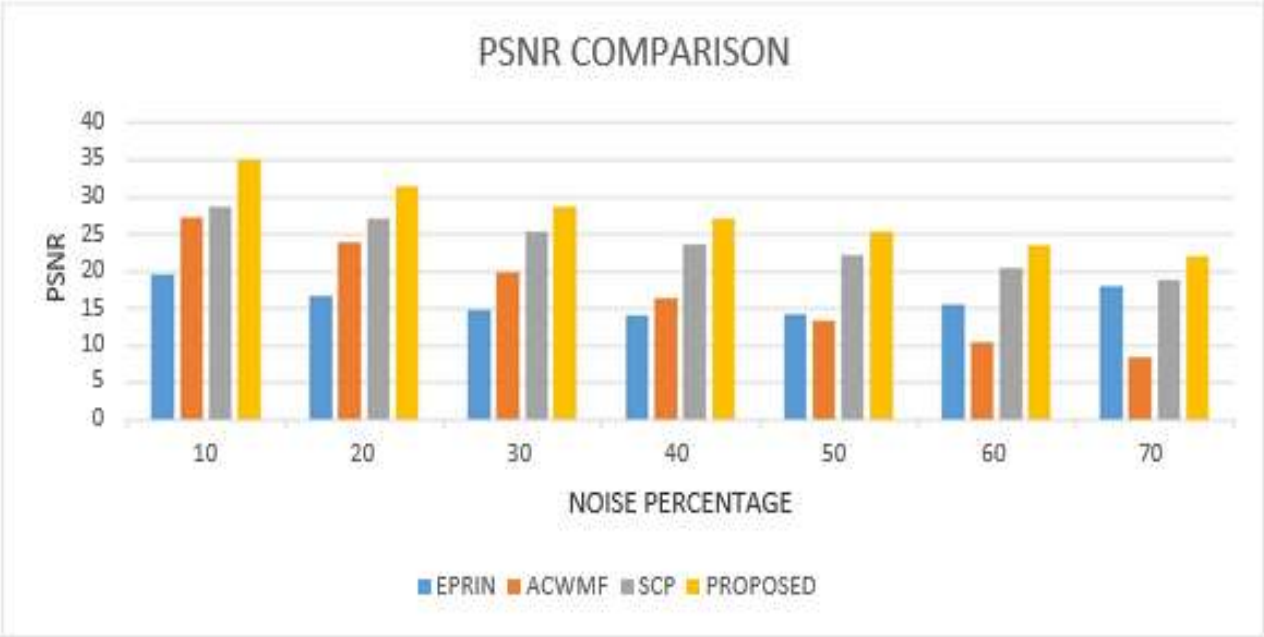

Figure 9: PSNR comparison of Figure 2 at various noise levels between EPRIN, ACWMF, SCP, and Proposed

SSIM is compares the quality of image. If SSIM value is high value means image quality is good .otherwise it is poor quality.

Table 2: SSIM values of Figure 2

Noise	10	20	30	40	50	60	70
EPRIN	0.309469	0.226554	0.194209	0.18025	0.180146	0.201389	0.289074
ACWMF	0.902019	0.777744	0.58332	0.404958	0.250824	0.140753	0.081046
SCP	0.988384	0.973623	0.954722	0.93135	0.90141	0.860973	0.811633
PROPOSED	0.981602	0.96777	0.953552	0.942589	0.923876	0.89926	0.866528

By observing table 2 and figure scp and proposed methods are almost similar at 30% noise ratio but at 70% noise level proposed method giving best results.

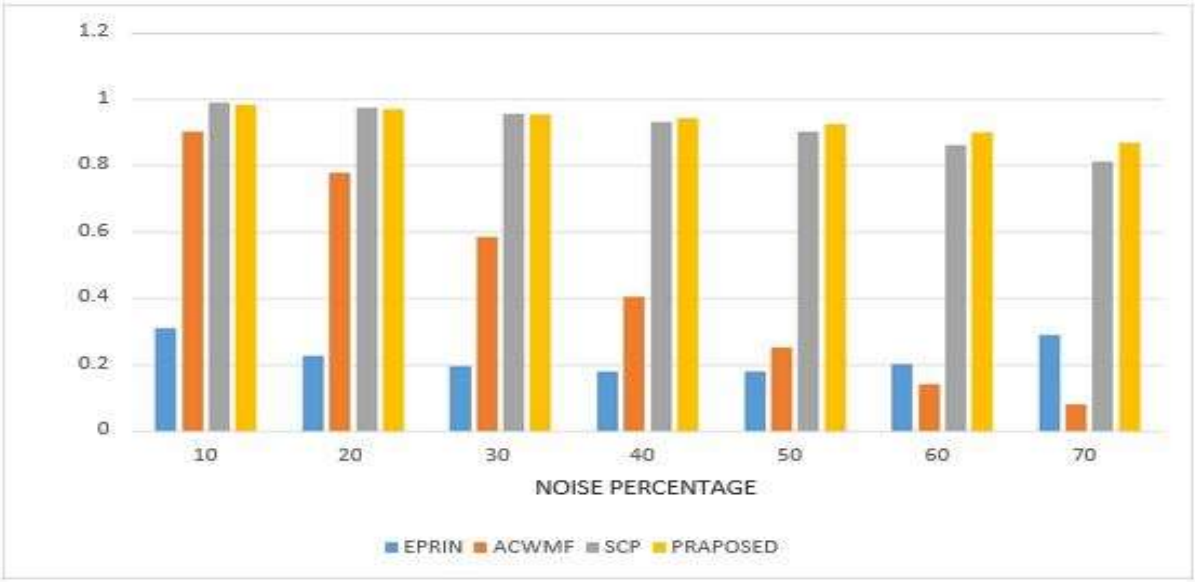

Figure 10: SSIM comparison of Figure 2 at various noise levels between EPRIN, ACWMF, SCP, and Proposed

PSNR values comparison between different algorithms for figure 3 at various levels of noise

Table 3: PSNR values of Figure 3

Noise level→	10	20	30	40	50	60	70
EPRIN	15.5105	15.03377	14.83693	15.18706	16.36568	18.69893	21.03351
ACWMF	31.0565	26.25341	21.08856	16.58141	12.96944	10.12654	7.802782
SCP	33.1639	29.76682	27.17243	25.07757	23.16962	21.1406	19.13435
PROPOSED	33.02108	30.77448	29.00623	27.46857	26.04916	24.60953	22.89453

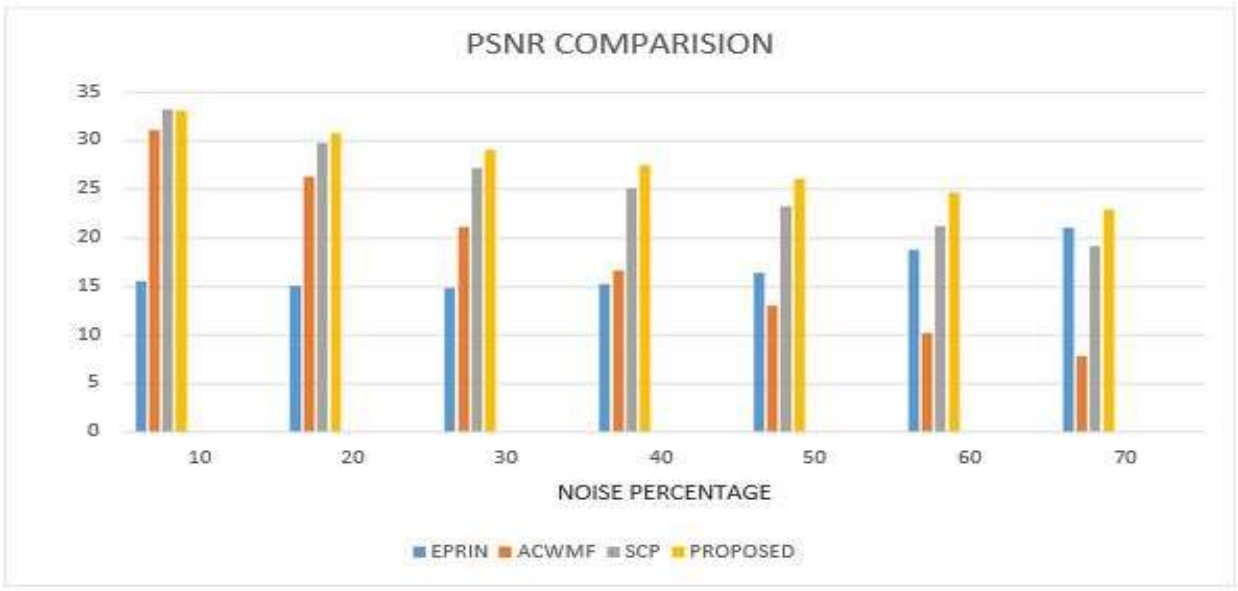

Figure 11: PSNR comparison of Figure 3 at various noise levels between EPRIN, ACWM, SCP, and Proposed

SSIM comparisons for Figure 3 at various noise levels

Table 4: SSIM values of Figure 12

Noise level→	10	20	30	40	50	60	70
EPRIN	0.219382	0.216472	0.217322	0.226143	0.254938	0.33985	0.51944
ACWMF	0.988455	0.950279	0.819424	0.587659	0.34822	0.184588	0.09656
SCP	0.985884	0.985187	0.974704	0.9593	0.934987	0.899242	0.847257
PROPOSED	0.992452	0.979342	0.972051	0.963442	0.952911	0.936705	0.912301

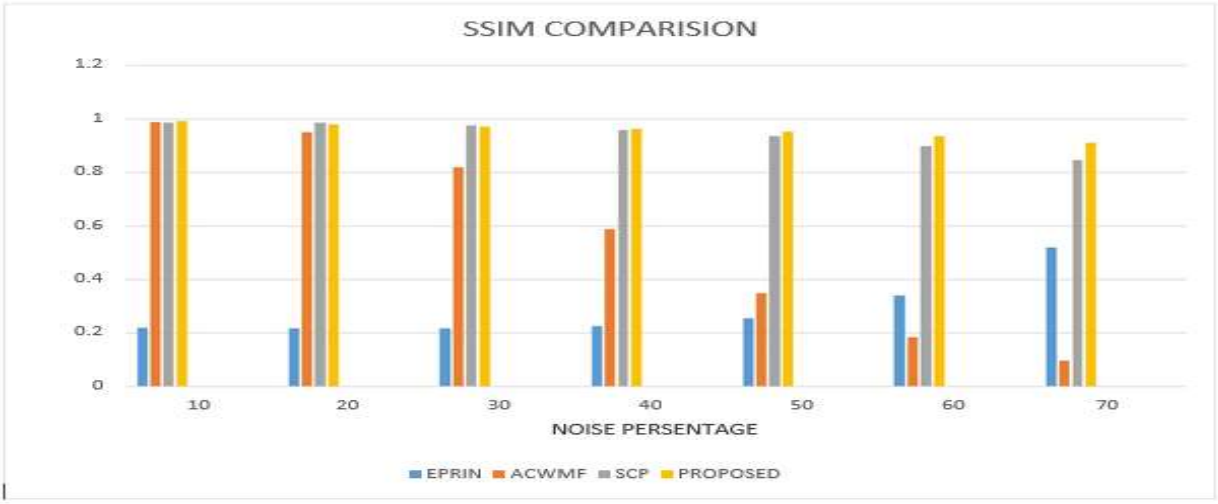

Figure 12: SSIM comparison of Figure 3 at various noise levels between EPRIN, ACWMF, SCP, and Proposed

By observing experimented results for figure 3 and by comparing PSNR, SSIM values of figure 3 it shows that proposing method giving good results even at higher level noise.

VI.CONCLUSION

It is known that letters in any language are collected from one or more paths. Each one has similar pixels connected in series within the path. If one or all the lines are effected by any noise type, then, it is problematic to read the letter. Therefore, an effective new procedure is proposed to restore document images degraded by salt and pepper noise. Usually document images degraded in creating, storage process, photocopying process, during scanning. In proposed method using connecting pixels and labelling method to restore document images from salt &pepper noise .this method tested at various levels of noise and threshold values and it delivered good results

REFERENCES

[1] Arya, S. C., Singh, R. S., & Mandoria, H. L (2015). Image Denoising in Hand Written Document for Degraded Documents using Wiener Filter Algorithm. *International Journal for Research in Emerging Science and Technology*, *2*(7), 50-56.

[2] Yu, H., Zhao, L., & Wang, H. (2008). An efficient procedure for removing random-valued impulse noise in images. *IEEE Signal Processing Letters*, *15*, 922-925.

[3] Chen, T., Ma, K. K., & Chen, L. H. (1999). Tri-state median filter for image denoising. *IEEE Transactions on Image processing*, *8*(12), 1834-1838.Quyoum, A. (2010). Shazia Akram Dr. Mehraj-Ud-Din Dar Aasia Quyoum.

[4] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, *82*(1), 35-45.

[5] Kaur, S. (2015). Noise Types and Various Removal Techniques;". *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume*, *4*.

[6] Chen, T., & Wu, H. R. (2001). Adaptive impulse detection using centre-weighted median filters. *IEEE Signal Processing Letters*, *8*(1), 1-3.

[7] Jayaraman, S., & Esakkirajan, S. (2009). Digital image processing.

[8] Pushpavalli, R., & Sivaradje, G. (2012). New Tristate Switching Median Filter for Image Enhancement. *International Journal of Advanced research and Engineering Technology*, *3*(1), 55-65.

[9] Mehta, R., & Aggarwal, N. K. (2014). Comparative Analysis of Median Filter and Adaptive Filter for Impulse Noise–A Review. *International Journal of Computer Applications*, *4*(11), 29-34.

[10] Premchaiswadi, N., Yimngam, S., & Premchaiswadi, W. (2009, August). A scheme for salt and pepper noise reduction on graylevel and color images. In *Proceedings of the 9th WSEAS international conference on Signal processing, computational geometry and artificial vision* (pp. 57-61).

[11] Pyatykh, S., & Hesser, J. (2014). Salt and pepper noise removal in binary images using image block prior probabilities. *Journal of Visual Communication and Image Representation*, *25*(5), 748-754.

[12] Shen, Y., Lam, E. Y., & Wong, N. (2008). A signomial programming approach for binary image restoration by penalized least squares. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *55*(1), 41-45.

[13] Maheswari, D., & Radha, V. (2010). Noise removal in compound image using median filter. *IJCSE) International Journal on Computer Science and Engineering*, *2*(04), 1359-1362.Awad, A. S. (2014).

[14] Ali S.Awad (2014), Denoising of document images based on similarity and connectivity principles. *Computers & Electrical Engineering*,*40*(8), 79-85.

Karasala Surya Prakash