International Journal of Scientific Research in Computer Science and Engineering
Vol.6, Special Issue.1, pp.48-63, January (2018)
Research Paper
E-ISSN: 2320-7639

Document Image Noises and Removal Methods

Gayatri. D. Patil[1] , Lubdha. M. Bendale[2], Roshani. L. Jain[3]

Department of Computer Science and IT, BASPONC College, Bhusawal 425-201 (MS)

Abstract- Document Image processing is the process in which, the important paper document like annual reports , certificates, charter, concept statement, news papers, Xerox papers or many others are first scanned and then stored that scanned images in the hard Disk or any particular location. The stored images will be in compatible electronic format, which images we can access easily. Document image processing (DIP) is management of documents throughout their life cycle from creation to death. It includes the capture of documents as digital images, typically by means of document scanners, and the storage, indexing, retrieval, processing, transmission, and printing of these documents.Document images contaminated by noise during scanning, transmission or conversion to digital form.

Keywords: Noise-Marginal, Max & Min, Median, Wiener, Midpoint

Document Image Noises and Removal Methods

There are several noises on digital document images i.e. Marginal noise, background noise, clutter noise, Edge noise, Rule line noise,stroke like pattern noise, salt and pepper noise.

1) **Marginal Noise**:

In Document digitalization, the old books are translated into digitalized form by scanning the pages of old books. The resulting images contain noise likes dark region, shadow of side page , folds shadow ,handwritten text around the margin of document images is known as the marginal noise.

2) Background Noise in document image:

The old document are scanned, then there is a some background noises in the scan document like interfering strokes , black spots, uneven contrast , background color, shadow of splits of papers , humidity absorbed by paper in different areas.

By using thresholding techniques, background noise is removing

3) salt and pepper noise in document image

Salt-and-pepper noise is a form of noise sometimes seen on images. Saltandpepper noise randomlyscattered black +whitepixels. It is also known as impulse noise. This noise can be caused by sharp and sudden disturbances in the image signal.

- **There are some techniques to remove salt and paper noise.**

1) **Median filter:** Median filtering is a nonlinear process useful in reducing impulsive or salt-and-pepper noise.

2) **Max filter:** Replace the value of a pixel by the maximum of the gray levels(the brightest point) in the neighborhood of the pixel.

$$f(x, y) = \max_{(s, t) \in S_{xy}} \{g(s,t)\}$$

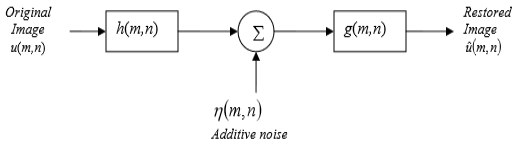
Max filter also known as 100^{th} percentile filter. Max filter helps in removing pepper noise.

3) **Min filter:**Replace the value of a pixel by the minimum of the gray levels (the darkest point) in the neighborhood of that pixel.

$$f(x, y) = \min_{(s, t) \in S_{xy}} \{g(s,t)\}$$

Min filter also known a zeroth percentile filter. Min filter helps in removing salt noise.

4) Wiener filter:

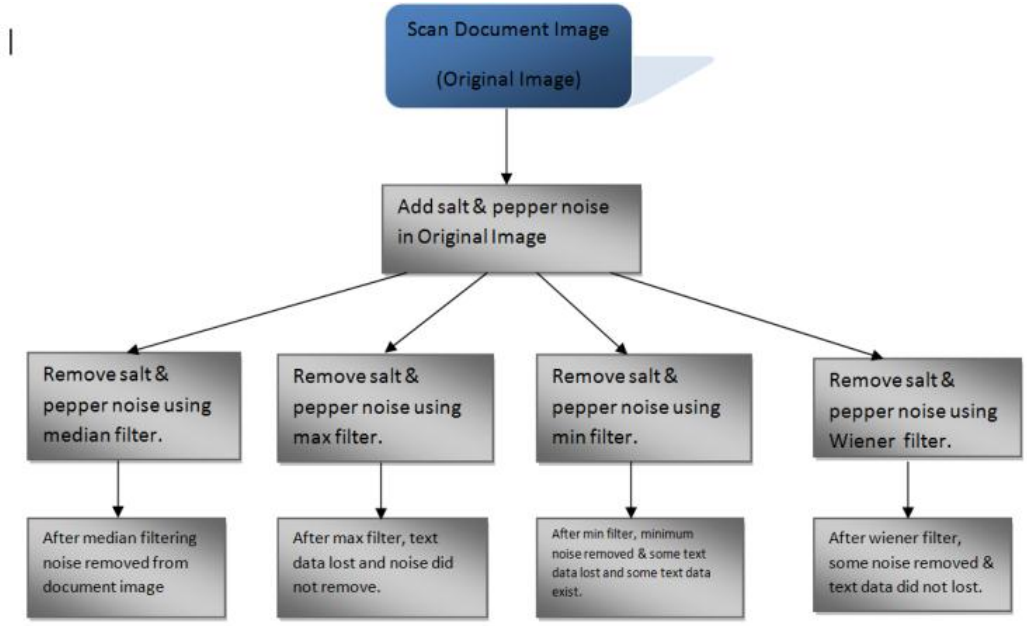


For Wiener Filter: $G(\omega_1,\omega_2)=\dfrac{H^*(\omega_1,\omega_2)S_{uu}(\omega_1,\omega_2)}{|H(\omega_1,\omega_2)|^2 S_{uu}(\omega_1,\omega_2)+S_{\eta\eta}(\omega_1,\omega_2)}$

where S_{uu} and $S_{\eta\eta}$ are Fourier transforms of the auto-correlation functions of $u(m,n)$ and $\eta(m,n)$ respectively.

5) Clutter Noise:

Clutter noise is also type of document image noises in which unwanted foreground content occurs in document image is typically larger that the text in binary images. Foreground content from numerous sources such as punched holes, stapled pins marks, folding marks, text stroke, in scan document image.



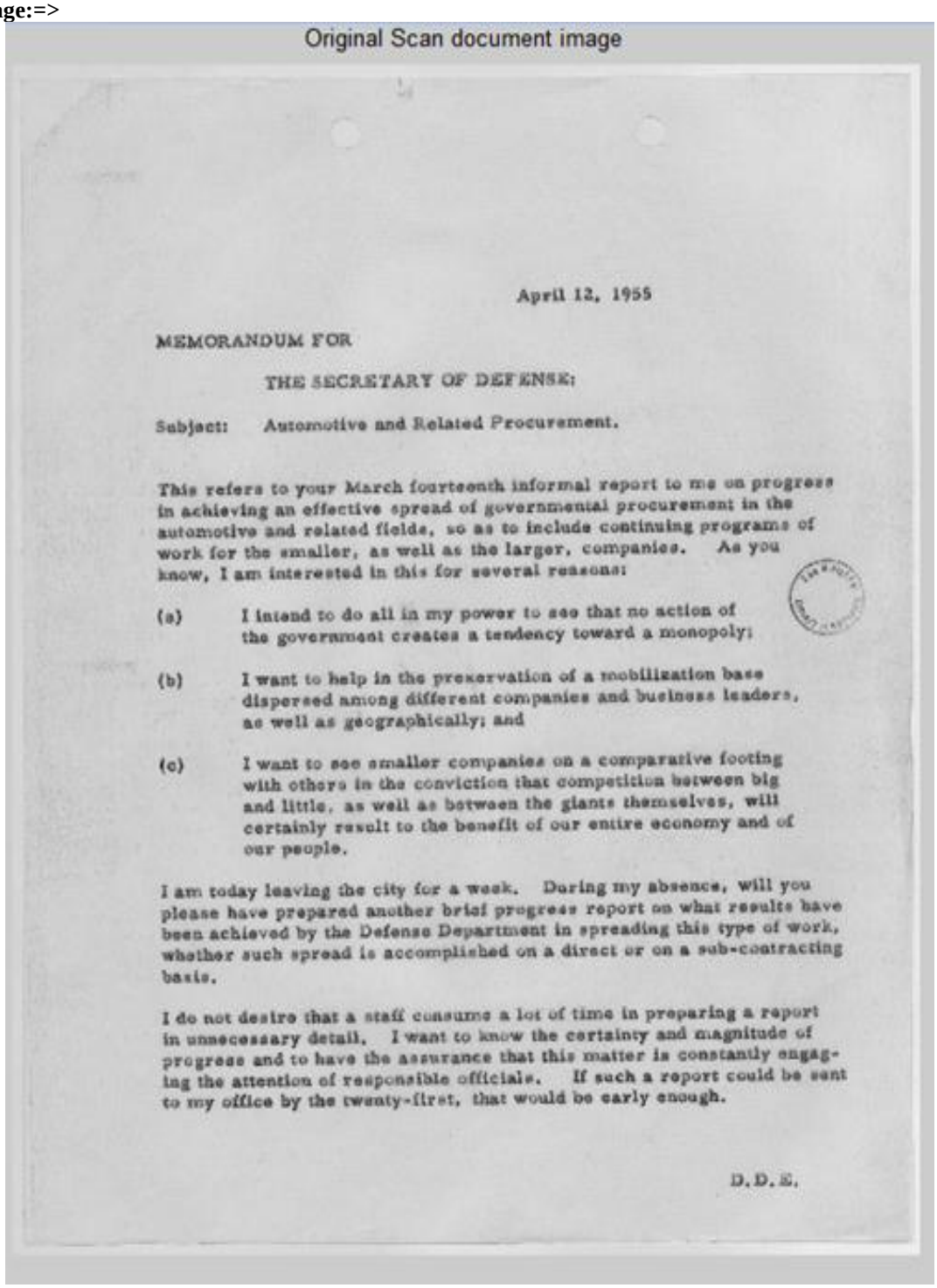
Experiemental Results:

```
>> I = imread('F:\Gayatri\im1.jpg');
>> imshow(I);
>> title('Original Scan document image');
>> b=imnoise(I,'salt & pepper',0.02);
>> imshow(b);
>> title('Salt & Pepper Noise in Scan document image');
>> c=medfilt2(b);
>> imshow(c);
>> title('Remove Salt & Pepper noise using median filter From Scan image');
>> cmax=nlfilter(b,[3 3],'max(x(:))');
>> title('Apply max filter after adding Salt & Pepper noise in Document image');
>> cmin=nlfilter(b,[3 3],'min(x(:))');
>> imshow(cmin);
>> title('Apply min filter after adding Salt & Pepper noise in Document image');
```

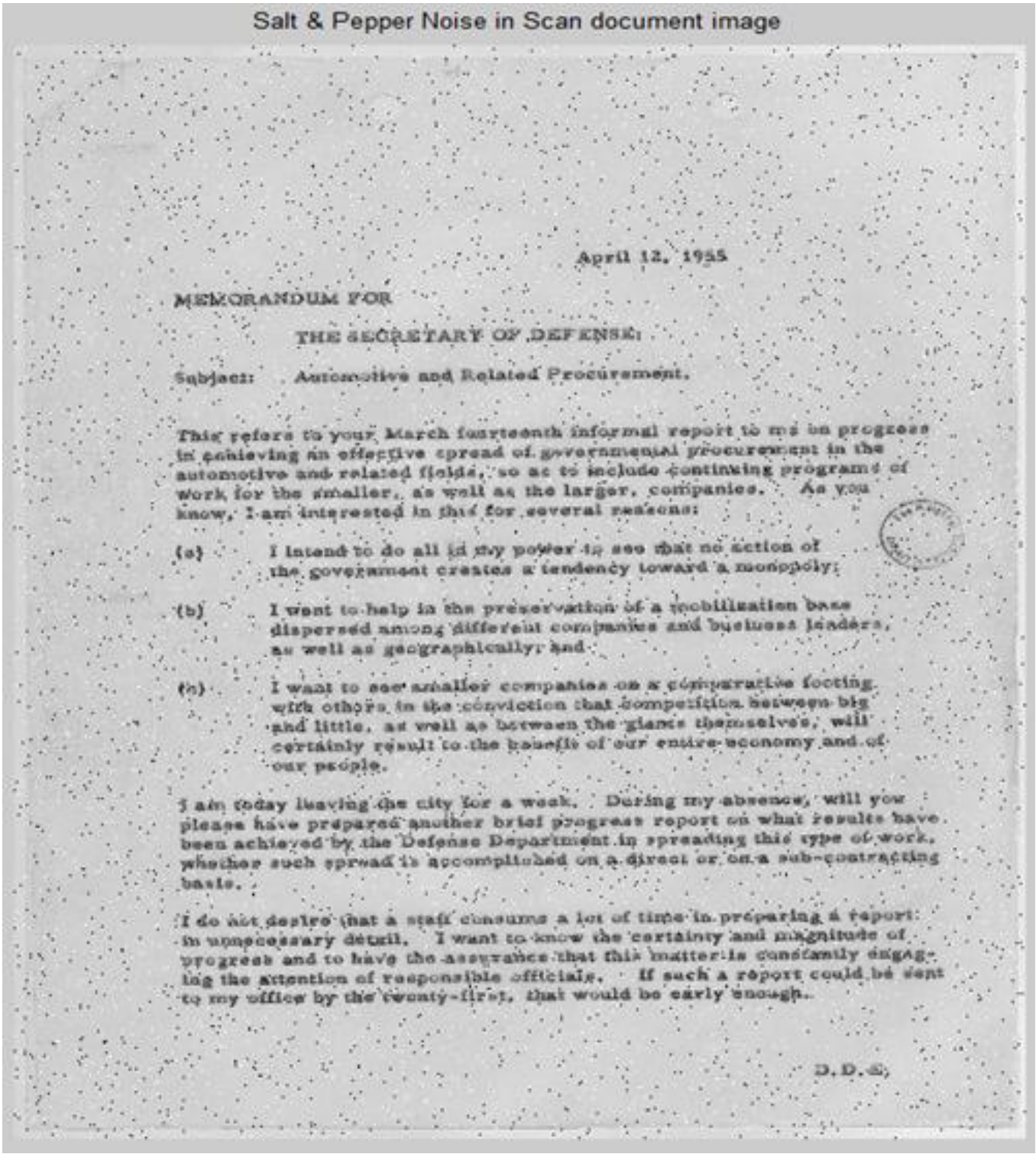


```
>> d=wiener2(b);
>> imshow(d);
>> title('Apply Wiener filter after adding Salt & Pepper noise in Document image');
```

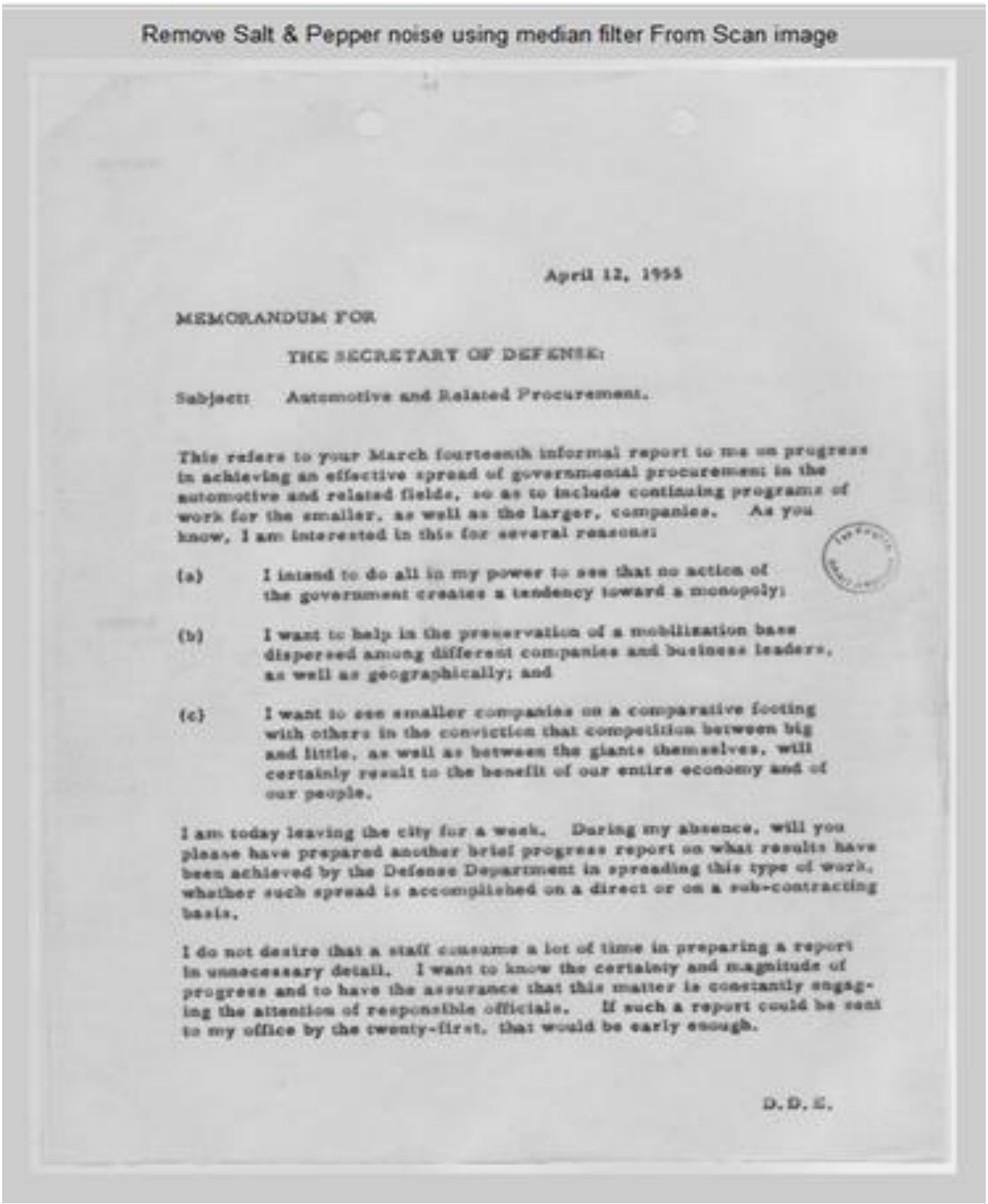
Original Image:=>



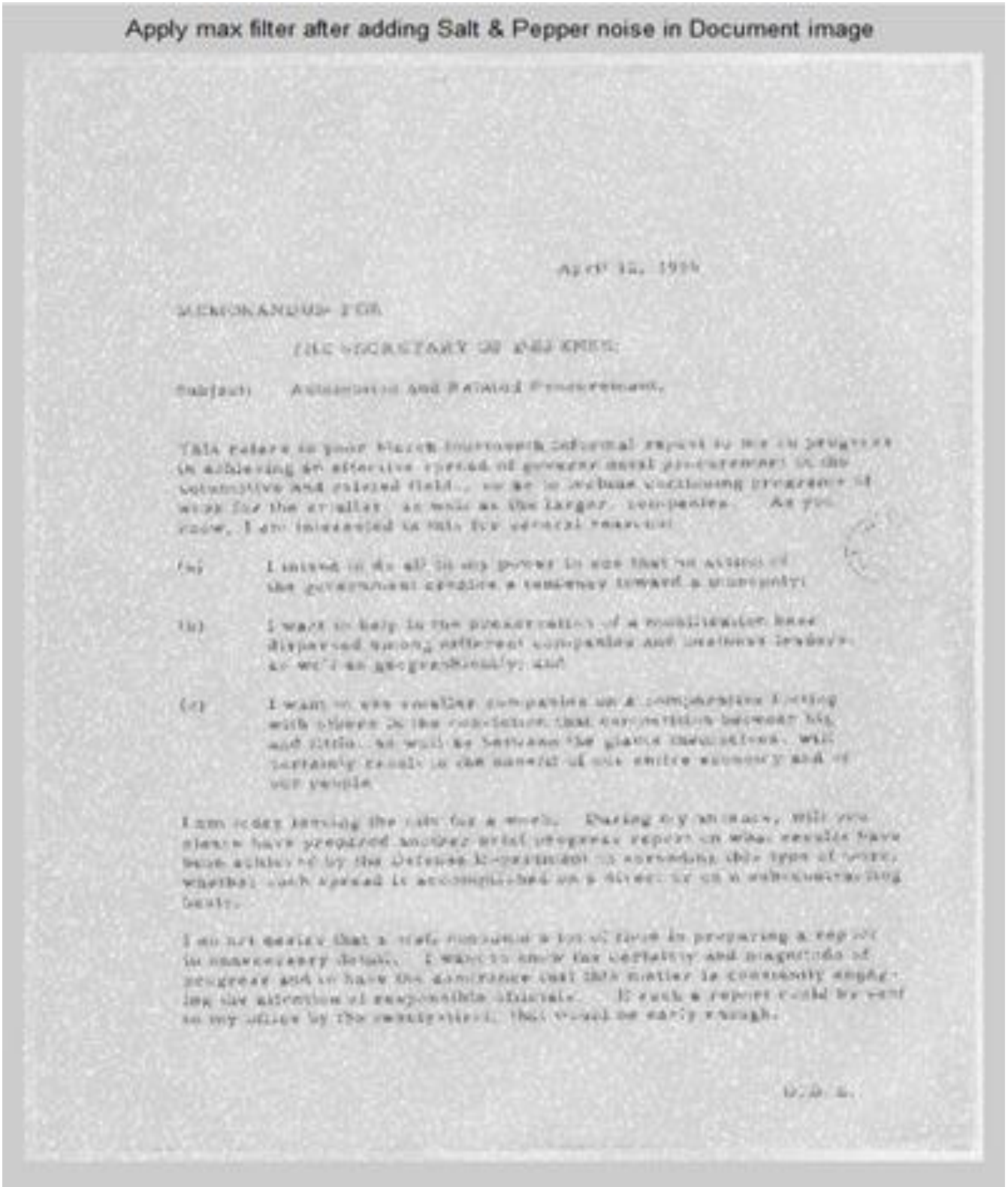
Adding Salt & Pepper noise in scan document image



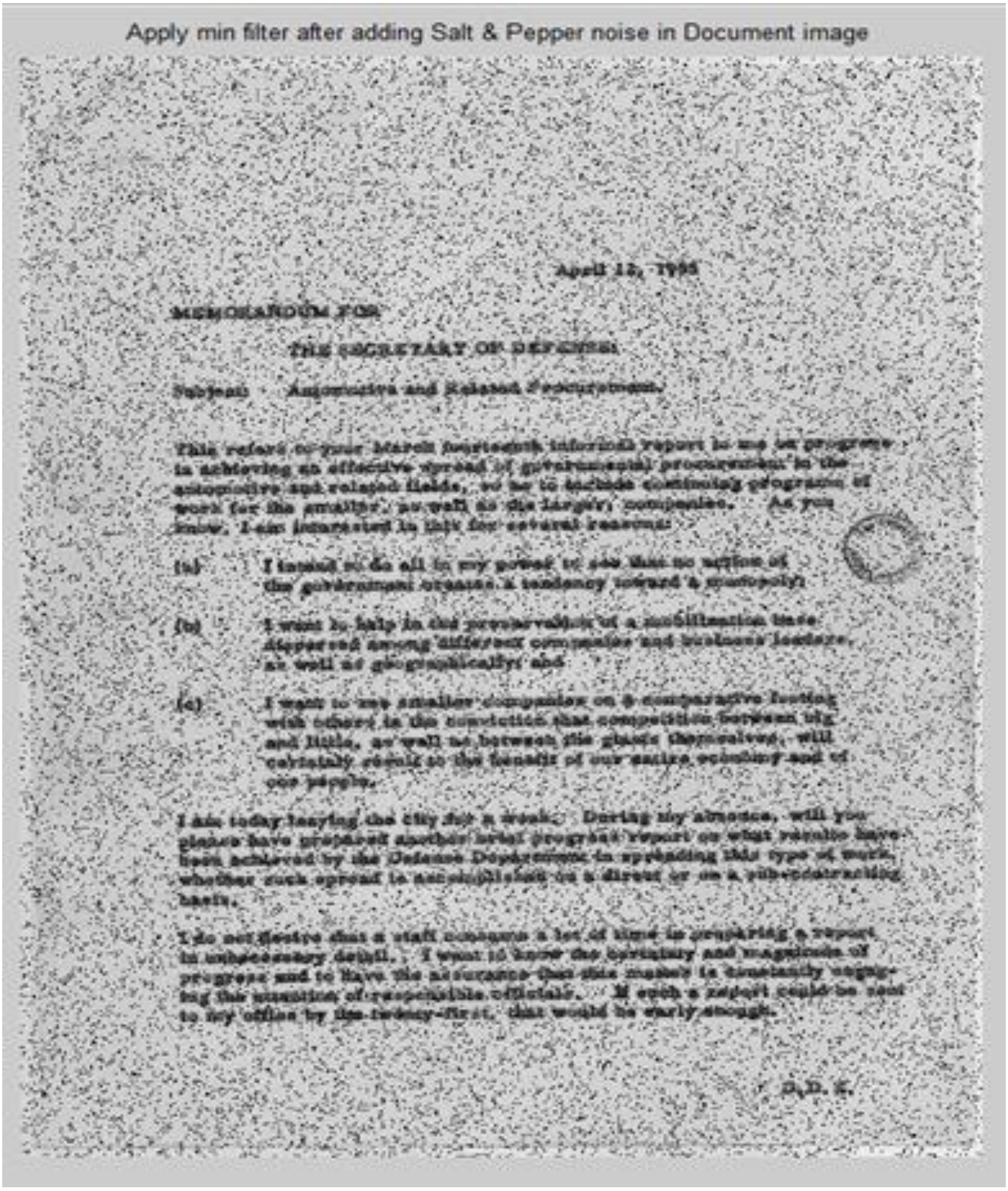
Remove salt & pepper noise using median filter from scan image.



Apply max filter after adding salt & pepper noise in Document image



Apply min filter after adding salt & pepper noise in Document image.



April 15, 1955

MEMORANDUM FOR

THE SECRETARY OF DEFENSE

Subject: Automotive and Related Procurement

This refers to your March fourteenth informal report to me on progress in achieving an effective spread of governmental procurement in the automotive and related fields, so as to include continuing programs of work for the smaller, as well as the larger, companies. As you know, I am interested in this for several reasons:

(a) I intend to do all in my power to see that no action of the government creates a tendency toward a monopoly;

(b) I want to help in the preservation of a mobilization base dispersed among different companies and business leaders, as well as geographically; and

(c) I want to see smaller companies on a comparative footing with others in the conviction that competition between big and little, as well as between the giants themselves, will certainly result to the benefit of our entire economy and of our people.

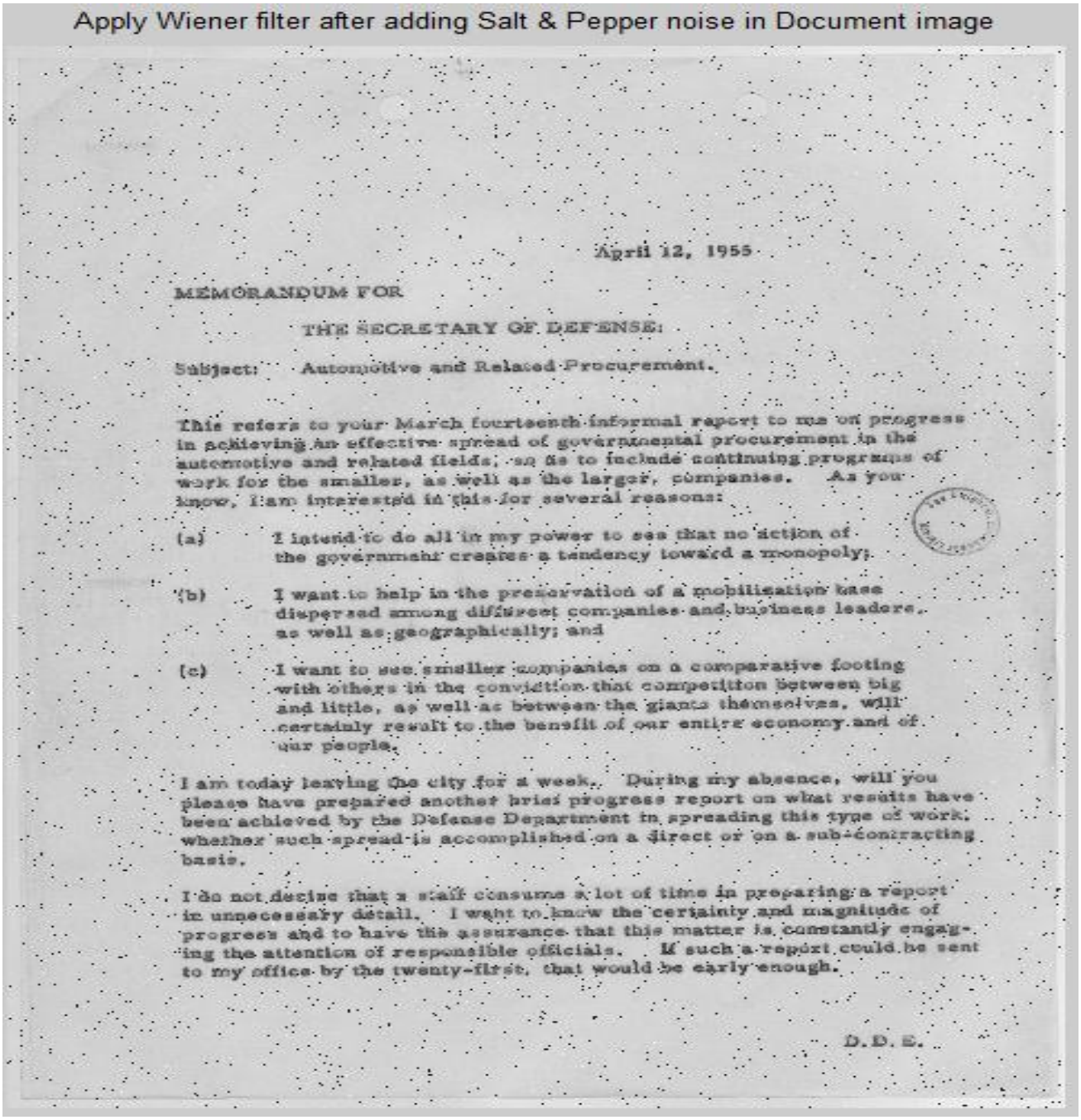
I am today leaving the city for a week. During my absence, will you please have prepared another brief progress report on what results have been achieved by the Defense Department in spreading this type of work, whether such spread is accomplished on a direct or on a subcontracting basis.

I do not desire that a staff consume a lot of time in preparing a report in unnecessary detail. I want to know the certainty and magnitude of progress and to have the assurance that this matter is constantly engaging the attention of responsible officials. If such a report could be sent to my office by the twenty-first, that would be early enough.

D. D. E.

Apply Wiener filter after adding salt & pepper noise in Document image.

Apply Wiener filter after adding Salt & Pepper noise in Document image



April 12, 1955

MEMORANDUM FOR

THE SECRETARY OF DEFENSE:

Subject: Automotive and Related Procurement.

This refers to your March fourteenth informal report to me on progress in achieving an effective spread of governmental procurement in the automotive and related fields, so as to include continuing programs of work for the smaller, as well as the larger, companies. As you know, I am interested in this for several reasons:

(a) I intend to do all in my power to see that no action of the government creates a tendency toward a monopoly;

(b) I want to help in the preservation of a mobilization base dispersed among different companies and business leaders, as well as geographically; and

(c) I want to see smaller companies on a comparative footing with others in the conviction that competition between big and little, as well as between the giants themselves, will certainly result to the benefit of our entire economy and of our people.

I am today leaving the city for a week. During my absence, will you please have prepared another brief progress report on what results have been achieved by the Defense Department in spreading this type of work, whether such spread is accomplished on a direct or on a sub-contracting basis.

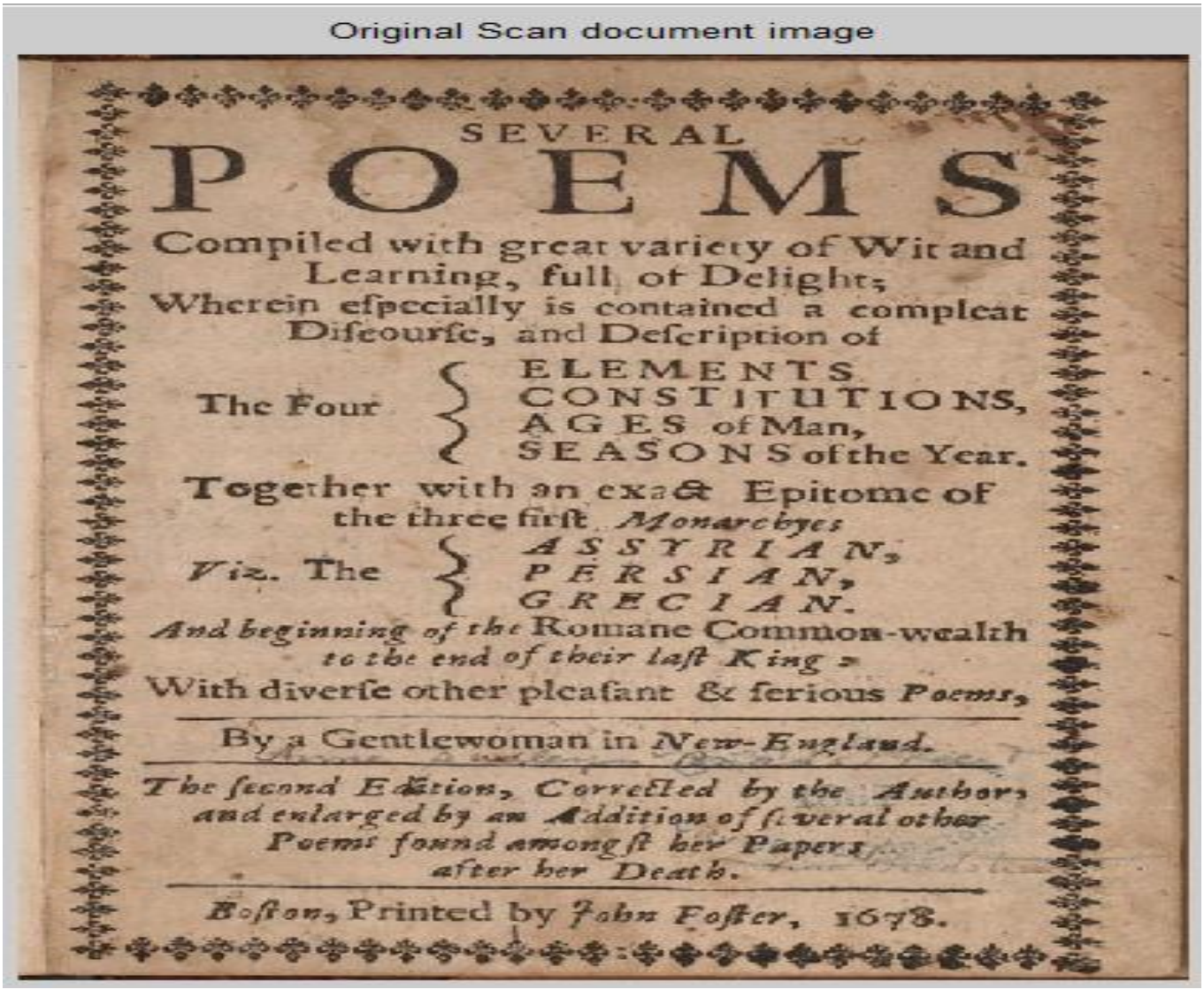
I do not desire that a staff consume a lot of time in preparing a report in unnecessary detail. I want to know the certainty and magnitude of progress and to have this assurance that this matter is constantly engaging the attention of responsible officials. If such a report could be sent to my office by the twenty-first, that would be early enough.

D. D. E.

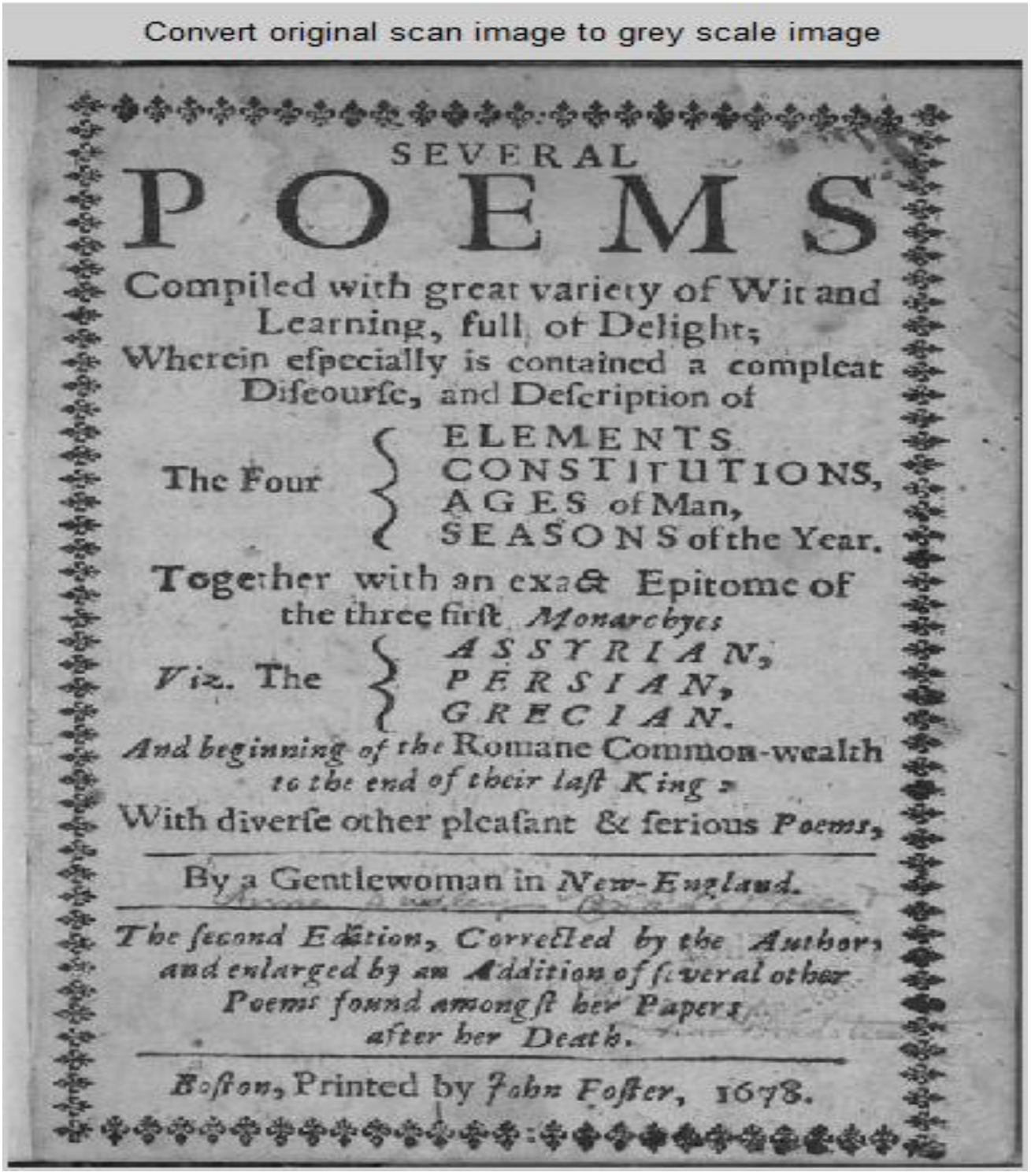
```
>> I = imread('F:\Gayatri\img3.jpg');
>> imshow(I);
>> title('Original Scan document image');
>> a=rgb2gray(I);
>> imshow(a);
>> title('Convert original scan image to grey scale image');
>> b=imnoise(a,'salt & pepper',0.10);
>> imshow(b);
>> title('Salt & Pepper Noise in Scan document image');
```



```
>> c=medfilt2(b);
>> imshow(c);
>> title('Remove Salt & Pepper noise using median filter From Scan image');
>> cmax=nlfilter(b,[3 3],'max(x(:))');
>> title('Apply max filter after adding Salt & Pepper noise in Document image');
>> cmin=nlfilter(b,[3 3],'min(x(:))');
>> imshow(cmin);
>> title('Apply min filter after adding Salt & Pepper noise in Document image');
>> d=wiener2(b);
>> imshow(d);
>> title('Apply Wiener filter after adding Salt & Pepper noise in Document image');
```



Convert original scan image to grey scale image



SEVERAL

POEMS

Compiled with great variety of Wit and Learning, full of Delight;
Wherein especially is contained a compleat Discourse, and Description of

The Four { ELEMENTS, CONSTITUTIONS, AGES of Man, SEASONS of the Year.

Together with an exact Epitome of the three first *Monarchyes*

Viz. The { *ASSYRIAN*, *PERSIAN*, *GRECIAN*.

And beginning of the Romane Common-wealth *to the end of their last King:*

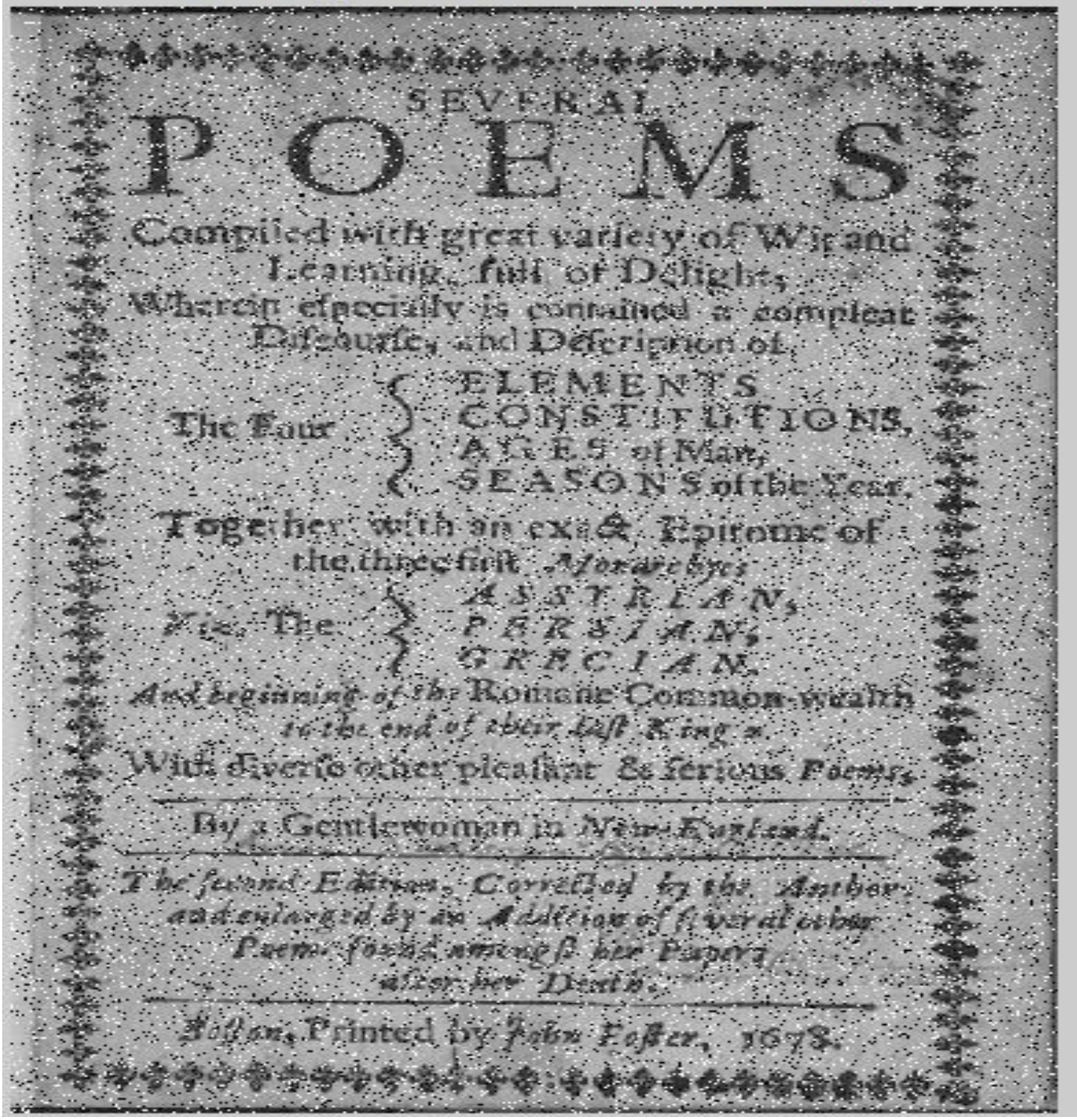
With diverse other pleasant & serious *Poems*,

By a Gentlewoman in *New-England*.

The second Edition, Corrected by the Author, and enlarged by an Addition of several other Poems found amongst her Papers after her Death.

Boston, Printed by *John Foster*, 1678.

Salt & Pepper Noise in Scan document image



SEVERAL

POEMS

Compiled with great variety of Wit and Learning, full of Delights,

Wherein especially is contained a compleat Discourse, and Description of

The Four { ELEMENTS, CONSTITUTIONS, AGES of Man, SEASONS of the Year.

Together with an exact Epitome of the three first Monarchyes

viz. The { ASSYRIAN, PERSIAN, GRECIAN.

And beginning of the Romane Common-wealth to the end of their last King.

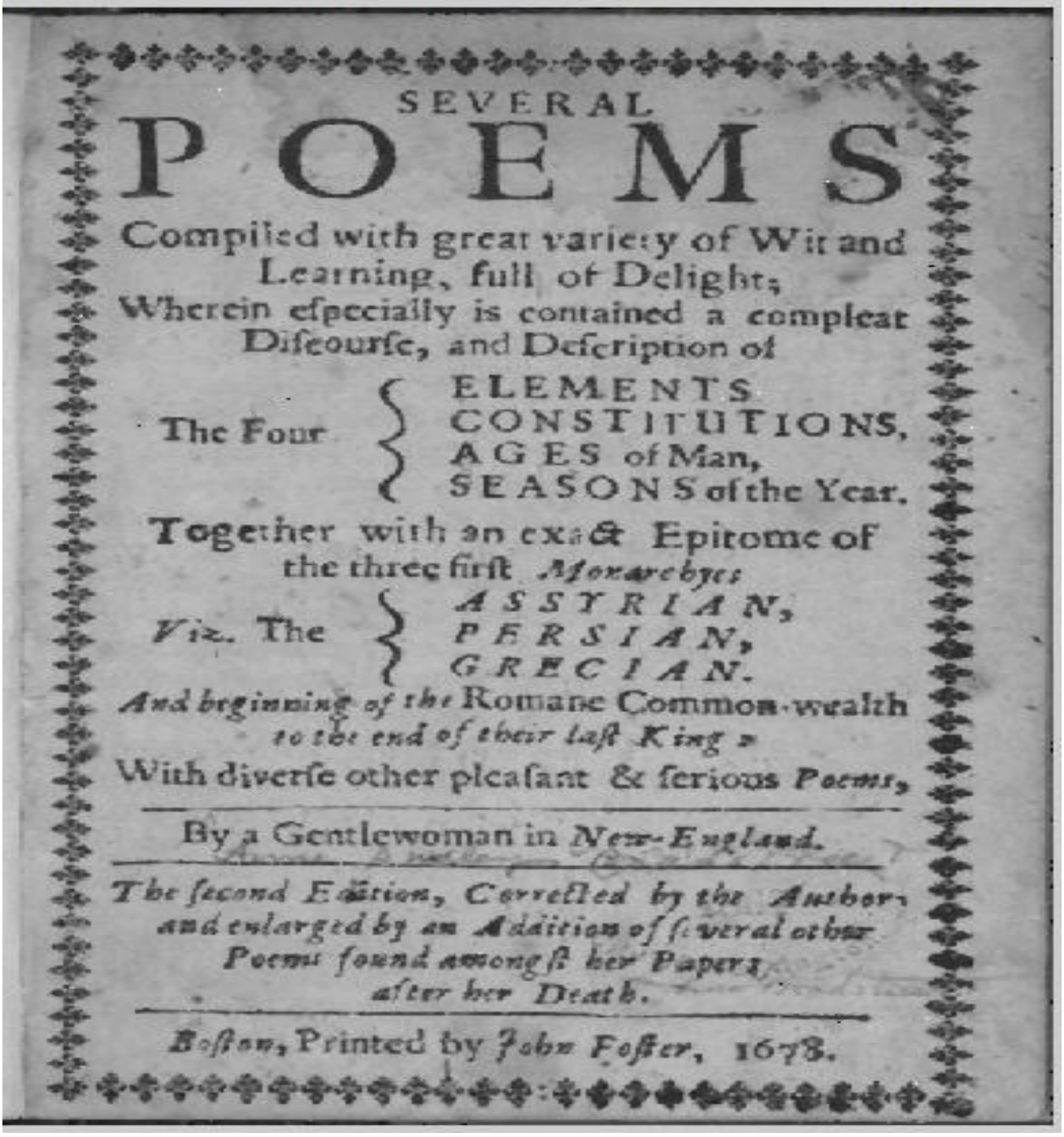
With diverse other pleasant & serious Poems.

By a Gentlewoman in *New-England*.

The second Edition, Corrected by the Author, and enlarged by an Addition of several other Poems found amongst her Papers after her Death.

Boston, Printed by *John Foster*, 1678.

Remove salt & pepper noise from scan image using median filter



SEVERAL

POEMS

Compiled with great variety of Wit and Learning, full of Delight;
Wherein especially is contained a compleat Discourse, and Description of

The Four { ELEMENTS, CONSTITUTIONS, AGES of Man, SEASONS of the Year.

Together with an exact Epitome of the three first *Monarchyes*

Viz. The { *ASSYRIAN*, *PERSIAN*, *GRECIAN*.

And beginning of the Romane Common-wealth *to the end of their last King*:

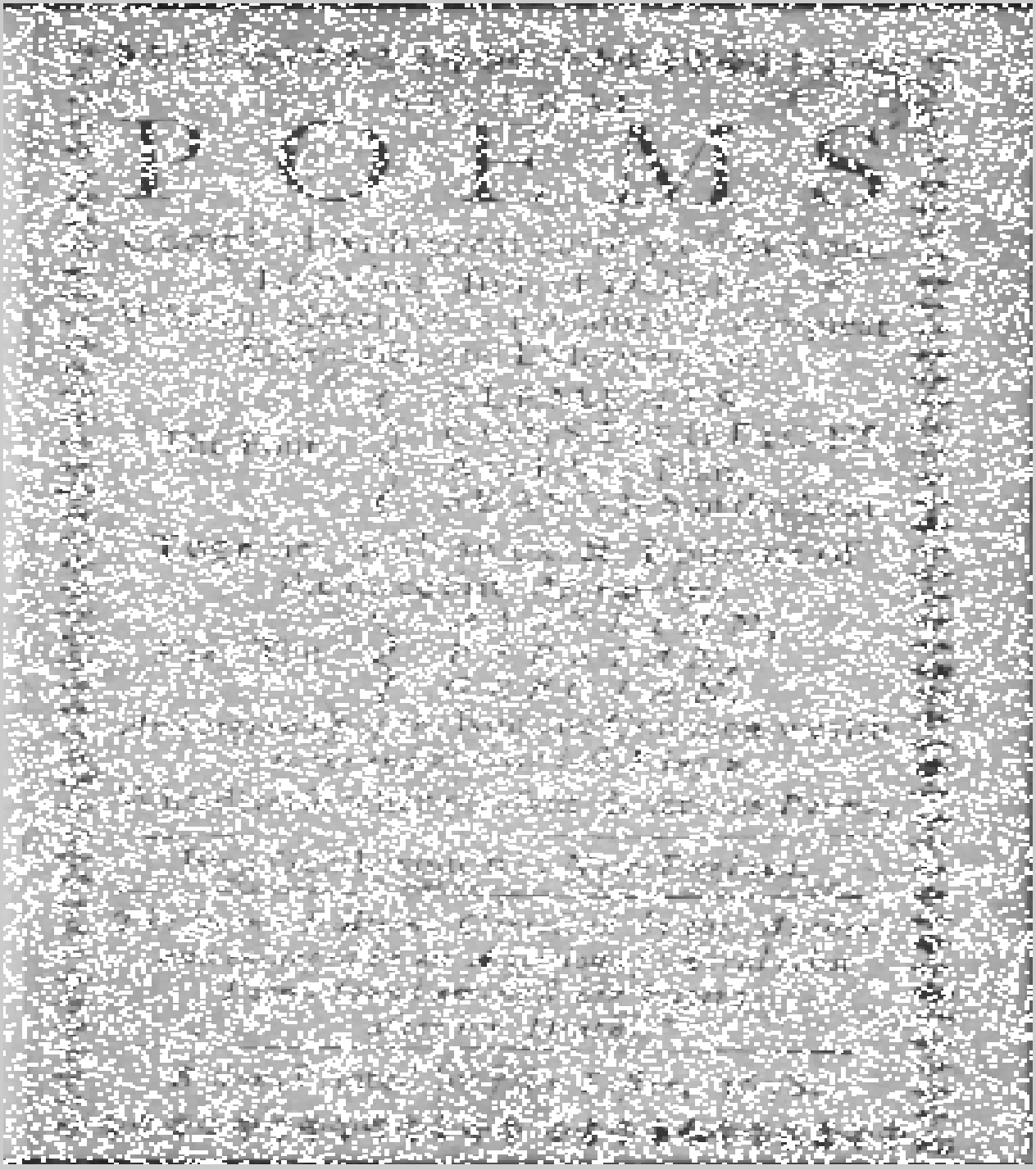
With diverse other pleasant & serious *Poems*,

By a Gentlewoman in *New-England*.

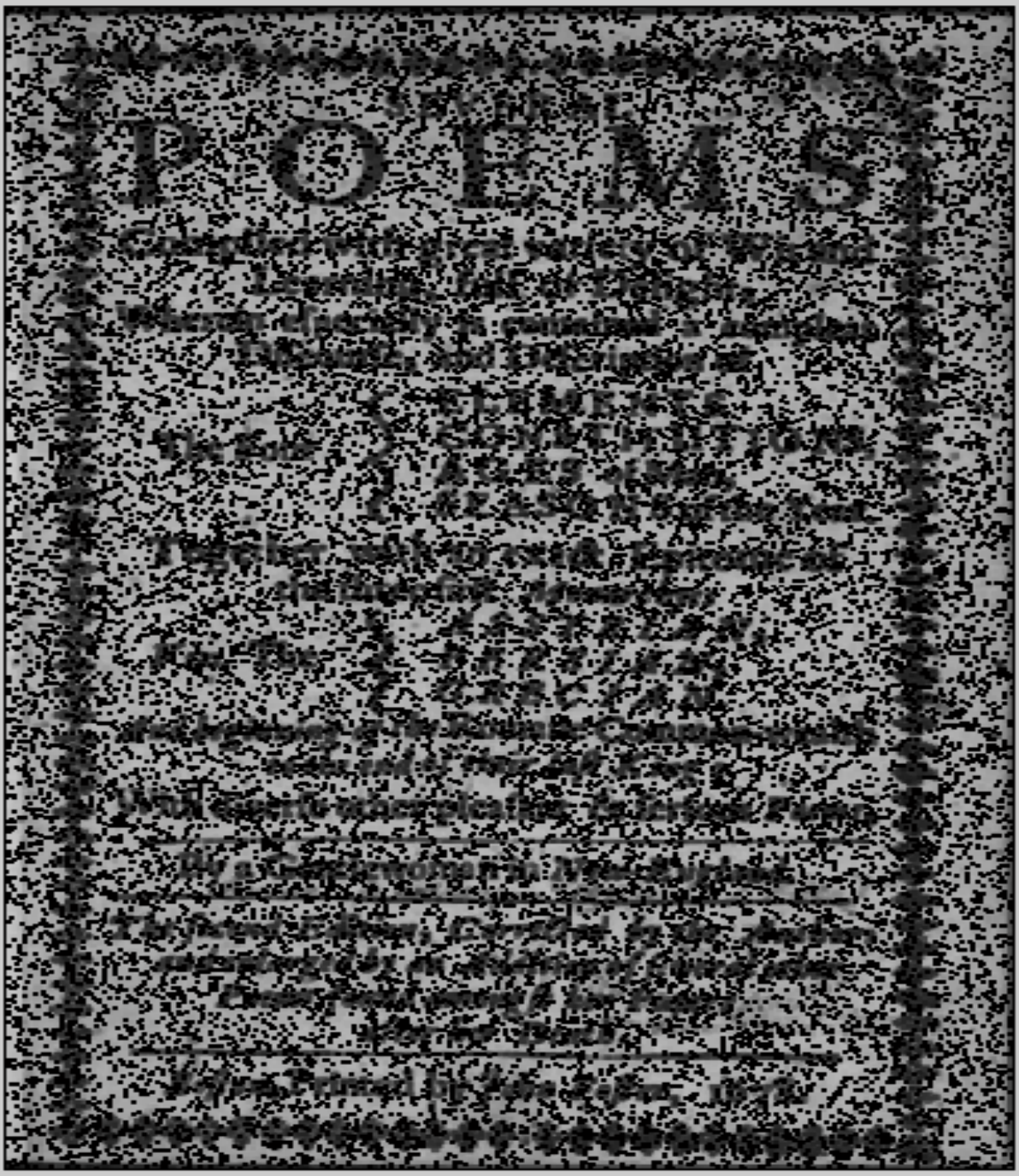
The second Edition, Corrected by the Author, and enlarged by an Addition of several other Poems found amongst her Papers after her Death.

Boston, Printed by *John Foster*, 1678.

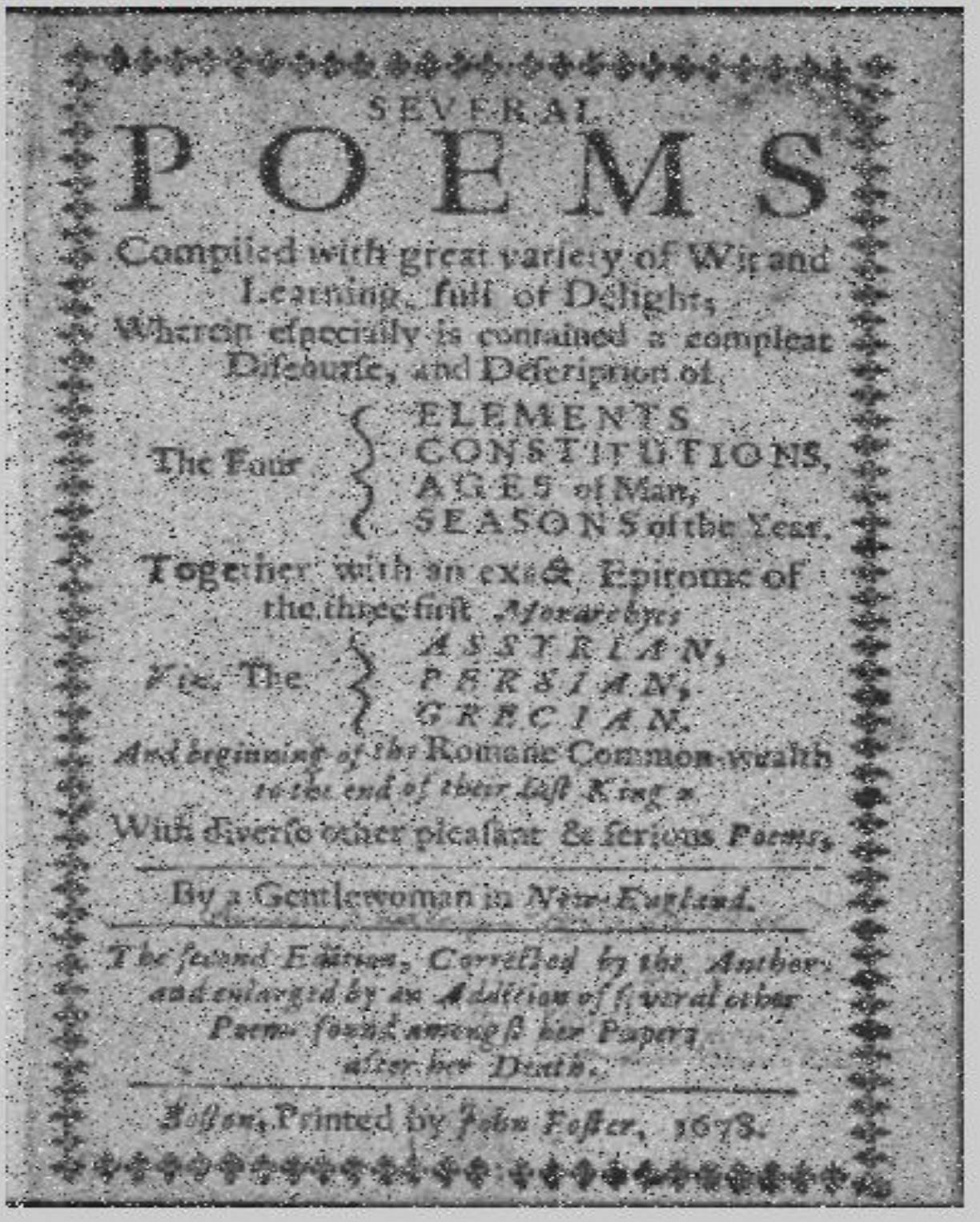
Apply max filter after adding Salt & Pepper noise in Document image



Apply min filter after adding Salt & Pepper noise in Document image



Apply Wiener filter after adding Salt & Pepper noise in document image



SEVERAL
POEMS
Compiled with great variety of Wit and Learning, full of Delight,
Wherein especially is contained a compleat Discourse, and Description of,

The Four { ELEMENTS, CONSTITUTIONS, AGES of Man, SEASONS of the Year,

Together with an exact Epitome of the three first *Monarchyes* viz. The { *ASSYRIAN*, *PERSIAN*, *GRECIAN*.

And beginning of the Romane Common-wealth *to the end of their last King.*

With diverse other pleasant & serious *Poems*.

By a Gentlewoman in *New-England*.

The second Edition, Corrected by the Author, and enlarged by an Addition of several other Poems found amongst her Papers after her Death.

Boston, Printed by *John Foster*, 1678.

Conclusion:

The various methods can use to remove noise from document images. This paper focuses on median, Max & Min, Wiener filters for remove noise from scanned document images . 0.02 salt and pepper noise externally added into scan document image for small text sized data and 0.10 salt and pepper noise for large text size data in scan document image. On resultant image filters applied and results compared.

Finally it concludes that Median filter is better to remove salt and pepper noise from document image than the other filter.

References:

1) Retrieved from https://en.wikipedia.org/wiki/Salt-and-pepper_noise
2) Retrieved from http://nptel.ac.in/courses/117104069/chapter_8/8_16.html
3) Retrieved from https://www.slideshare.net/sardaralam1/filters-for-noise-in-image-processing
4) Fana, K.-C., Wangb, Y.-K., & Laya, T.-R. (2002). Marginal noise removal of document images. *Pattern Recognition, 35*, 2593-2611.
5) Farahmand, A., Sarrafzadeh, ,. A., & Shanbehzadeh, J. (2013). Document Image Noises and Removal Methods. *Proceedings of the International MultiConference of Engineers and Computer Scientists, I*.
6) *www.cvisiontech.com*. (n.d.). Retrieved from http://www.cvisiontech.com/library/pdf/pdf-document/document-image-processing.html
7) https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiZwPCC243XAhWLpY8KHdHPAcAQFggnMAA&url=http%3A%2F%2Fwww.uta.edu%2Ffaculty%2Fkrrao%2Fdip%2FCourses%2FEE5356%2Fproject7.pdf&usg=AOvVaw0N3NYEH_-JSkY5taXiNWQu

