Grenze International Journal of Computer Theory and Engineering , Special issue

Efficient Pre-Processing of 2D Gel Electrophoresis Images using Fast Independent Component Analysis

Sweekar Sudhakara[1] and Sagar Sudhakara[2]
[1]Dept. of Electronics and Communication, JSSATE, Bengaluru, Karnataka-560060
Email: sweekar.top@gmail.com
[2]Dept. of Electrical Engineering, IIT Bombay, Maharashtra-400076
Email: sagarsudhakara@gmail.com

***Abstract*—Proteomics is a field which studies large scale analysis of proteins. Two dimensional gel electrophoresis (2DGE) is a technique used to study protein expression in this field. Large number of protein spots are separated on the basis of their molecular weight and isoelectric point. 2DGE images have non-linear background and are noisy. A complete automated system is required for the analysis of 2DGE images. Efficient Image pre-processing is the major challenge involved. We use the Fast Independent Component analysis for denoising the two dimensional Gel Electrophoresis Images. We have used Fast Independent component analysis method for denoising when there are different kind of noise present in 2D gel Electrophoresis Images. The results are compared with the performance of other denoising method like Wavelet based denoising of images.**

***Index Terms*— Proteomics, Wavelets, Independent Component Analysis, FastICA & Otsu Method.**

I. INTRODUCTION

2D Gel Electrophoresis is a technique used for Proteomic analysis. Electrophoresis technique is used to separate proteins in a biological sample on a gel. The protein spots are separated on the basis of molecular weight and iso-electric point after the application of electric field and a gradient of pH. The protein spots get separated on the basis of their isoelectric point at all pH level [10]. After the protein spots attain a neutral state imaging of 2DGE is done. The resulting 2DGE images are captured as digital images by imaging system. Efficient pre-processing is required to find out the new proteins which are responsible for making new biomarkers to find out specific diseases like cancer. Separating thousands of protein spots is a very tedious and laborious job. Protein spots are highly irregular in terms of their shape, size and intensity which make their automated detection very challenging. Hence, a complete automated system for the analysis of 2DGE images is increasingly in demand. Image Denoising is one of the major challenge involved in the automation process because 2D Gel Electrophoresis images consists of faint spots, spots in exposed background, streaks in some area, artifacts, overlapped and saturated spots. This kind of spots makes the task of spot detection difficult. Image denoising with respect to 2DGE images has the following challenges:

- Overlapped spots & faint spots,
- background noise & spots on streaks,
- horizontal and vertical streaks,
- background noise & spots in gel with geometric distortion.

Grenze ID: 01.GIJCTE.3.4.69

II. Denoising of 2D Gel Electrophoresis Images

Noise is an undesirable or unwanted signal component in an image. It can occur because of improper acquisition by the digital camera or by poor illumination. It is well known that 2DGE images are inherently noisy due to imperfect image acquisition. Presence of noise in 2D Gel Electrophoresis images might lead to false detection & affect the further process by introducing outliers. It reduces the accuracy of a completely automated system. Denoising is an important preprocessing step of analysis of 2D Gel Electrophoresis images. Noises which affect 2DGE images can be broadly classified as :

- Additive noise - Gaussian noise, uniform noise, salt & pepper noise,
- Multiplicative noise - Non-isometric noise & shot noise.

Additive noise is independent of the spatial distribution of the image, whereas multiplicative noise depends on the intensity levels & spectral distribution of the image [12]. Multiplicative noise, such as non-isometric noise & shot noise, is a byproduct of the stochastic nature of image acquisition system. Since the noise is dependent on the intensity levels of the image pixels being corrupted, it is not easy to form a statistical model for this nonstationary process. Thus, making it very difficult to remove. Spatial filtering is not a solution as we need to preserve the edges along with denoising. A tradeoff must be maintained between denoising and preservation of singularities in 2DGE images.

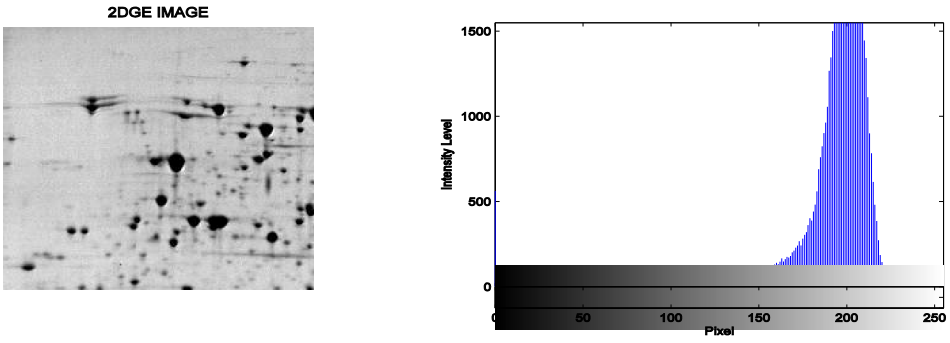

Fig. 1: 2D Gel Electrophoresis image & its histogram

A. Different noise which affects 2DGE images

It is well known that 2DGE images are inherently noisy due to the nonlinear electrophoresis process which gets susceptible to dust and imperfect image acquisition. Due to inherent nonlinear background variation, denoising of 2DGE images is not a trivial task.

1) *Gaussian noise*: The main source for Gaussian noise in digital images is poor image acquisition technique. Gaussian noise is an additive noise, which is independent at each pixel & independent of the signal intensity levels.

2) *Salt & Pepper noise*: Impulsive noise or spike noise is sometimes called salt and pepper noise. An image containing salt-and-pepper noise will have dark pixels in bright regions and bright pixels in dark regions.

3) *Poisson noise*: Poisson noise is also called as shot noise. It occurs mainly because of improper illumination while taking the digital image. The noise which affects the digital image has a poisson distribution.

4) *Speckle noise*: Speckle noise is a multiplicative noise which affects the spectral distribution of 2D Gel Electrophoresis images. It is in direct proportion to the local grey level in any area and it is difficult to alleviate this noise by simple spatial filtering.

III. Denoising Methods

The main aim of any denoising method is to suppress the noise, remove artifacts and correct the background. There are several denoising algorithms in the literature, which can be broadly classified as spatial domain filtering and transform domain filtering. Spatial domain filtering includes mean filter, Weiner filter & median filter. Whereas, the transform domain filtering includes Spatial frequency filtering, Wavelet domain filtering etc.

A. Spatial domain filtering

It is the most fundamental way of denoising any image. It employs spatial filters to denoise the image. Spatial domain filtering can be broadly classified as linear and non-linear filters [1].

Linear filters: The main example of a linear filter is a mean filter which acts on an image by smoothing the image. It is a simple window sliding spatial filter which replaces the center value of the window with the average values of all its neighbour including its pixel value. Linear mean filters are implemented using a convolutional mask. Another important linear filter is the wiener filter which requires the information about the spectra of noise and original signal. Wiener filter does not work out well for 2D Gel Electrophoresis images as it requires the signal to be smooth. Choosing the window size is an overhead [1].

Non-Linear filters: Median filtering is one of the major non-linear filter used for denoising of images. It follows the window sliding principle, where the intensity value of center pixel in a window is replaced by the median of its intensity value and its neighbour intensity values [4]. Generally spatial domain filtering methods alleviate noise upto a reasonable extent, but at the cost of blurring the images. Spatial filtering methods do not preserve edges and produce distorted images. Hence Denoising is an important preprocessing operation and if done accurately it prevents the over estimation of the image background and formation of misleading spots. In image Denoising process, a trade off must be made between noise suppression and preservation of singularities in the image. A good denoising process must remove uneven background distortions, high frequency noise, salt & pepper noise introduced due to dust and horizontal-vertical streaks without introducing any significant distortion in an image.

B. Transform domain filtering

The transform domain filtering technique can be broadly classified as Spatial-frequency filtering and Wavelet domain approach of denoising. Transform domain filtering works well compared to spatial domain filtering in terms of preserving the edges.

Spatial frequency filtering: Usually noise is present in the high frequency part of the image spectrum. In the spatial frequency filtering method, low pass filters are employed to remove noise. It is employed using Fast Fourier transform with a particular cutoff frequency. Fixing a particular cutoff frequency is a tedious process. They may produce frequency artifacts in the resulting denoised image. Using high pass filtering will enhance the edges but will also amplify the noisy component of the image, which is the least thing that we require.

Wavelet domain approach for denoising: Wavelet means small wave (localized waves) which are functions that are generated from basic function called as prototype or mother wavelet. the complete function is obtained by scaling & translating mother wavelet. In wavelet analysis the signal of interest is multiplied with wavelet function & then transform computed for each generated segment. In signal analysis using discrete wavelet we often speak about approximation and detail. The detail components are low scale, high frequency part & the approximation components are high scale, low frequency part of the signal. Wavelet Domain Denoising Technique is one of the method which exploits the relationship between Denoising and Compression. Non-Linear Thresholding on wavelet coefficients is a standard way of denoising natural images. It is a non-linear process and a powerful tool for decorrelation and hence compression. The decorrelating of wavelet transform creates sparse signal i.e. most of the coefficients become insignificant. Most of the significant information is represented by a very small number of coefficients. This sparse nature of wavelet coefficients suggests that even if we remove most of the small coefficients by making them zero below a certain threshold and take inverse wavelet transform, then reconstructed signal will have most of the information & hence will be noise free. The threshold can be hard threshold or soft threshold depending on the statistical property of the image. But the state of the art algorithm for denoising 2D Gel Electrophoresis images was the Wavelet based Modelling approach [11]. This method models the spot of the gel image as gaussian curves and use MMSE estimator to denoise the image. Horizontal and vertical streaks are not removed which affects the further steps of 2DGE images analysis. We use the Fast Independent Component analysis to alleviate the noise.

IV. Independent Component Analysis

It is a statistical method for transforming an observed multidimensional random vector into components that are statistically independent [2]. Transformation of data to reveal its structure (feature extraction). To explain the problem statement in simple words or simple mathematical representation of ICA is as follows Let X be the observed random vector (mixture signals) and S be the underlying source vector

$$X = AS \quad (1)$$

where the mixing matrix A and source vector is unknown. Finding the independent source vector by solving the cost function which either maximizes the non-gaussianity of the calculated independent source components or minimize the mutual information. Apriori knowledge of the source probability distribution can be used in the cost function. Kurtosis which is a measure of non-gaussianity can be used as an objective function. Before denoising of 2D Gel Electrophoresis images by Independent Component analysis we need to follow two preprocessing steps. These steps are:

Centering: Centering the variables to simplify the ICA algorithm, we remove the mean from random vector X and after estimating the demixing matrix, we add back the mean to the estimated independent vector S so as to recentre back the data vector.

$$X' = X - E[X] \quad (2)$$

Whitening: This preprocessing step ensures that all components are treated equally apriori before using FastICA

$$E[X'X'^T] = EDE^T \quad (3)$$

$$Whitening\ matrix, V = ED^{-0.5}E^T \quad (4)$$

$$Z = VX' \quad (5)$$

The new vector Z obtained is white i.e. its components are uncorrelated & the variances is equal to one. FastICA is the most frequently used method for calculating the demixing matrix and estimating the independent source vector.

A. FastICA

FastICA method is an iterative fixed-point algorithm which was proposed by Hyvarinen et al [3]. It is used for maximizing non-Gaussianity & an alternative to gradient based method. This method has fast convergence and can be used for optimizing different types of cost functions, such as negentropy or kurtosis. The greatest advantage of FastICA method over gradient methods is that it does not have a learning rate or other parameters which may affect the convergence.

V. Results

We first denoise the 2D Gel Electrophoresis images corrupted by Gaussian noise, Salt and pepper noise, poisson noise & speckle noise by wavelet based denoising technique. The results obtained after denoising by Fast Independent Component Analysis are compared with other state of the art methods. Using Wavelet based denoising technique when the 2DGE image is corrupted by gaussian noise. The denoised image looks distorted and the vertical & horizontal streaks are not alleviated. The background is not corrected which might lead to false detection in the future steps.

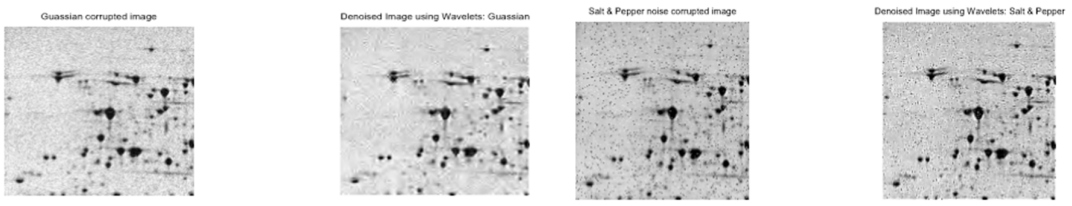

Fig 2: 2D Gel Electrophoresis Images corrupted by Gaussian and Salt & Pepper noise denoised using Wavelets

The image which is corrupted by salt and pepper noise after denoising by wavelet based method does not change much, even though the PSNR improves, horizontal & vertical streaks still exist in the image.
Poisson noise is a multiplicative noise which is difficult to eradicate by spatial filtering. Wavelet domain denoising improves the PSNR but fails to correct the background and remove outlier points from the image.
Wavelet Domain filtering fails to eradicate speckle noise form 2D Gel Electrophoresis images, even though Peak signal to noise ratio improves the resulting image is distorted.

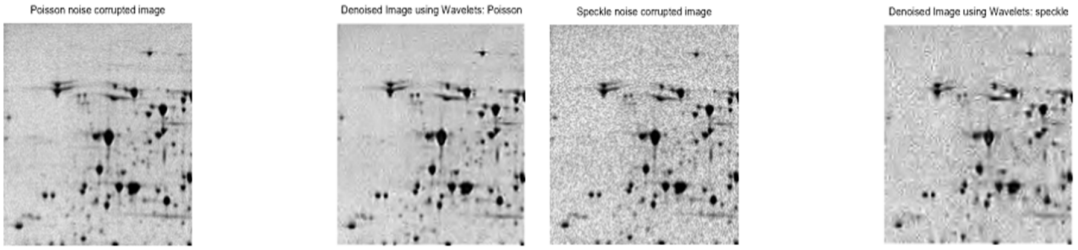

Fig 3: 2D Gel Electrophoresis Images corrupted by Poisson & Speckle noise denoised using Wavelets

We find that PSNR is not a good metric in our case for comparison of efficient preprocessing. Independent component analysis works well even when the image is corrupted by multiplicative noise.

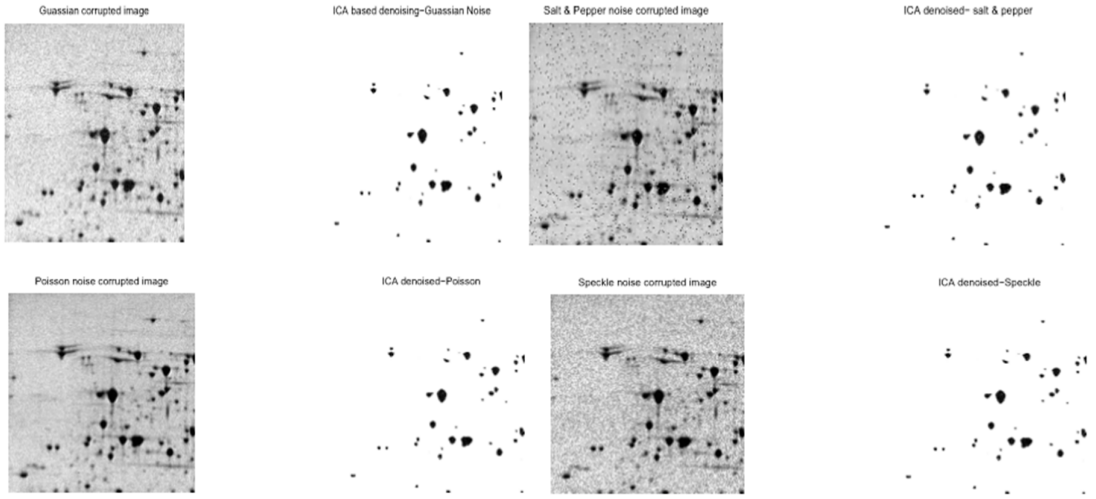

Fig 4: 2DGE Images corrupted by gaussian, salt & pepper, poisson & speckle noise denoised using ICA

The main of the preprocessing step was to correct the background and remove vertical & horizontal streaks without distorting the protein spots. Independent component analysis does a fine job with respect to this and edges are preserved.

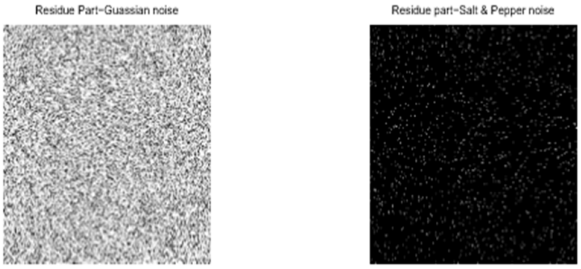

Fig 5: Residual part remaining after 2DGE image is denoised using Independent Component Analysis

The residual images after denoising 2D Gel Electrophoresis images using Independent component analysis is shown in the figure above. It plays an important role in separating out only the noisy part from the image without distorting the image.

VI. Conclusion

Most of the denoising methods fail to remove horizontal and vertical streaks from 2D Gel Electrophoresis images. Wavelet based method improves the PSNR metric but distorts the image finally. We have seen in the results above that PSNR is not a better metric in this case. Independent Component analysis method is robust to outliers and noise. Efficient pre-processing using Independent component analysis will make further steps like Segmentation error free. All the above simulation results are obtained after programming in Matlab Software.

Acknowledgment

I would like to thank Mr Rohit Sharma, scientist at BARC for providing us the data and I am also thankful to Prof. V.M. Gadre for giving me valuable inputs.

References

[1] Motwani, Mukesh C., Mukesh C. Gadiya, Rakhi C. Motwani, and Frederick C. Harris. "Survey of image denoising techniques." In Proceedings of GSPX, pp. 27-30. 2004.
[2] Hyvrinen, Aapo, Juha Karhunen, and Erkki Oja. Independent component analysis. Vol. 46. John Wiley & Sons, 2004.
[3] Hyvrinen, A., J. Karhunen, and E. Oja. "The FastICA package for MATLAB." Helsinki Univ. of Technology, Espoo, Finland (2005).
[4] Stark, Hans-Georg. Wavelets and signal processing: an application-based introduction. Springer Science & Business Media, 2005
[5] Ruikar, Sachin D., and Dharmpal D. Doye. "Wavelet based image denoising technique." IJACSA) International Journal of Advanced Computer Science and Applications 2.3 (2011).
[6] Hames, B. David, ed. Gel electrophoresis of proteins: a practical approach. Vol. 197. OUP Oxford, 1998.
[7] da Cunha, Arthur L., Jianping Zhou, and Minh N. Do. "Nonsubsampled contourlet transform: filter design and applications in denoising." IEEE International Conference on Image Processing 2005. Vol. 1. IEEE, 2005.
[8] Gonzalez, R. S., and Paul Wintz. "Digital image processing." (1977).
[9] Portilla, Javier, et al. "Image denoising using scale mixtures of Gaussians in the wavelet domain." IEEE Transactions on Image processing 12.11 (2003): 1338-1351.
[10] Ong, Shao-En, and Akhilesh Pandey. "An evaluation of the use of two dimensional gel electrophoresis in proteomics." Biomolecular engineering 18.5 (2001): 195-205.
[11] Rangarajan, Raghuram, Ramji Venkataramanan, and Siddharth Shah. "Image denoising using wavelets." Wavelet and Time Frequencies (2002).
[12] Jain, Anil K. Fundamentals of digital image processing. Prentice-Hall, Inc., 1989.