Improved NL-means Algorithm for Two-dimensional Gel Electrophoresis Images Denoising

Hua-Mei Xin, Fengping Zhao

College of Physics and Electronics, Shandong Normal University, 250014, Jinan, China
E-mail: huamei.xin@gmail.com

Abstract

Denoising is a fundamental stage of two-dimensional gel electrophoresis (2DGE) images analysis, which strongly influences spot detection, spot matching and other pixel-based methods. It is critical to effective noise suppression. This paper proposed a new method based on the Non-Local-means (NL-means) algorithm, which is simple and can protect the structure well. Compared with the Translation Invariant Wavelet Transform (TIWT), the Total Variation (TV) and the NL-means, the performance of the improved NL-means algorithm outperforms them in PSNR values and the visual quality of the image. Experiments also verified that the improved NL-means algorithm is a more suitable method for 2DGE images denoising.

Keywords: Image Denoising, Non-Local-Means, Two-Dimensional Gel Electrophoresis

1. Introduction

Proteomics involves identification of proteins and determination of their role in physiological and pathophysiological functions. The first step in the proteomic study is to separate and analyze proteins. Although promising progress has been made in alternative protein separation techniques in proteomics, such as isotope-coded affinity tag methodology and two-dimensional liquid chromatography-tandem mass spectrometry, there is still no generally applicable method that can replace two-dimensional gel electrophoresis (2DGE) in its ability to separate highly complex protein mixtures derived from microorganisms, whole cells or tissues [1]. The basic principle in 2DGE is to separate proteins on a gel in two dimensions, isoelectric point being in the first dimension and molecular mass in the second dimension. To make the proteins visible on the gel, staining is applied to the gel after separation. This creates a two-dimensional image, where proteins appear as scattered dark spots throughout the gel based on their individual charge and molecular mass [2]. With thousands of protein spots potentially displayed on each gel, computer-assisted analysis becomes necessary [3]. Preprocessing for noise removal is one typical step of image analysis [4]. 2DGE images need to be filtered to reduce the inherent noise introduced by the experimental preparation procedures and acquisition process of the gel images [5].

Noise suppression is a fundamental stage in 2DGE images analysis, and the effect of denoising may have great impact on the results of other analysis steps [6]. Because the noise hampers spot detection with methods based on signal derivatives. That is, signal derivatives magnify the noise, causing identification of false peaks (spots) and incorrect determination of the borders of spots [7]. The main advantages of a good denoising lead to more accurate estimation of spot properties leading to improved spot differential analysis which is key for reliable biomarkers identification [6,8]. Therefore, we ought to focus on the image denoising for the further enhancement of image processing.

Currently, there are many commercially available 2DGE image analysis tools, such as ImageMaster, Melanie [9,10], PDQuest [11], PreteinMine and Z3 [12]. The principles for 2DGE image analysis denoising in commercial software are based on spatial filtering. Common spatial filters include median, mean, Gaussian and adaptive, and their algorithm is simple and computationally efficient. But these filters tend to severely distort spot edges and alter the intensity values of spot pixels [13]. Median modified wiener filter (MMWF), which is the modification of the Wiener filter, was recently introduced by Carlo V. Cannistraci et al. for 2DGE images [15]. The MMWF was not the best filter in all tests, but it had a good performance in all varying contexts. In 2004, Kaczmarek et al. introduced translation invariant
denoising to 2DGE images, and it was used by a number of publications [7,16,17]. Translation invariant denoising can solve the Pseudo-Gibbs phenomenon caused by wavelet transform. It outperforms spatial filtering and orthogonal wavelet transform in terms of S/N and spot edge-preserving. The contourlet transform was also introduced for 2DGE images denoising as an extension to the wavelet transform recently [6,14,27]. The effect of the contourlet is efficiently improved spot detection, especially in low-intensity spots, and less introduction of artifacts compared to conventional filters and wavelets [3]. But the improvement implies the costs of more complex setting strategies and an increased number of required parameters. As we all know, 2DGE images are typical examples of nonstationary signals due to the large and unstructured variations in spot intensities and size, and it is impossible to distinguish signal from noise in the space or frequency domain alone. Well the spatial-frequency domain approaches need accurate choice and setting of several strategies as well as tuning of at least three parameter values to denoise the images [15]. In this paper a new method is proposed, which is simple and requires fewer parameters. This proposed algorithm is based on the improved Non-Local-means (NL-means) and can effectively remove noise while protecting the structures of gel images. In the following part, we will verify the superiority of this algorithm compared to wavelet transform and total variation.

2. Methods

2.1 Translation Invariant Wavelet Transform (TIWT)

Wavelet denoising process will produce two kinds of artifacts: a bad approximation of the signal and a pseudo-Gibbs phenomena. Translation invariant denoising allows reduction of such artifacts by shifting the signal before noise reduction, followed by the back-shifting after reconstruction and averaging of all signals obtained in that way [13,18]. It is defined as,

$$\bar{T}(I_o; (S_h)_h = Ave(S_{-h}(T(S_h I_o))))_{h \in H}$$

(1)

Where $I_o$ is a two-dimensional noisy signal with the size of $n \times n (n = 2^j)$, $h$ is the cycle shift and $H = \{ \emptyset \leq h < n \}$, where $h \in Z$, $S_h$ is the operator with shifts $h$, $T$ means wavelet thresholding.

For the thresholding method, soft threshold is used because the discontinuities of the hard threshold will cause the pseudo-Gibbs phenomena. Here we use Bayes Threshold to calculate the threshold value,

$$t = \sigma^2 / \sigma_{sj}$$

(2)

where

$$\sigma = (1/0.6745) \text{median} \{ |C_{HH_1}| \}$$

(3)

$$\sigma_{sj} = \sqrt{\text{max}(\sigma^2 - \sigma^2, 0)} \quad \text{for} \quad i = 1, 2, ..., L$$

(4)

$\sigma^2$ is a variance of the coefficients from the i-th subband. $C_{HH_1}$ is the wavelet coefficients from the $HH_1$ subband. In equation (2), when $\sigma_{sj} = 0$, $t = \text{max} \{ |C_{ij}| \}$, and $C_y$ is the fine subband. Then all the detail coefficients reduce to zero.

2.2 Total Variation (TV) Transform

The approaches based on linear time-frequency analysis severely ruin important singular features because both the noise and singularities share high frequencies. So nonlinear operators began to be exploited. In 1992, Rudin, Osher and Fetami proposed a classical TV denoising method: ROF model [19], that is,
\[
\min\left(\int_\Omega |I| d\Omega + \lambda/2 \int_\Omega (I - I_0)^2 d\Omega \right) 
\]

(5)

Here \(I_0\) is the noisy image, and \(I\) is the restored image. \(\Omega\) denotes the image domain. \(\lambda\) is the fitting parameter and it is used to adjust the filter so as to balance the denoising and smoothing, computed by [20],

\[
\lambda = (1/\sigma^2) \int_\Omega \text{div}(\nabla I/|\nabla I|)(I - I_0) d\Omega 
\]

(6)

where \(\sigma\) is the standard deviation of noise, and \(\Omega\) is the area of image.

### 2.3 NL-means Algorithm

Given a discrete noisy image \(v = \{v(i)|i \in I\}\), for a pixel \(i\), the estimated value \(\text{NL}[v](i)\) is computed as a weighted average of all the pixels in the image by

\[
\text{NL}[v](i) = \sum_{j \in I} w(i, j) v(j) 
\]

(7)

where the family of weights \(\{w(i, j)\}\) depend on the similarity between the pixels \(i\) and \(j\), and satisfy the usual conditions \(0 \leq w(i, j) \leq 1\) and \(\sum_j w(i, j) = 1\) [21,22].

The weights are defined by

\[
w(i, j) = 1/Z(i) e^{-\|v(N_i) - v(N_j)\|_2^2/h^2} 
\]

(8)

where \(Z(i) = \sum_j e^{-\|v(N_i) - v(N_j)\|_2^2/h^2}\) is the normalizing factor and the parameter \(h\) controls the decay of the exponential function [21,22]. \(\|v(N_i) - v(N_j)\|_2^2\) is Gaussian weighted Euclidean distance.

The Euclidean distance preserves the order of similarity between pixels as following,

\[
E\|v(N_i) - v(N_j)\|_2^2 = \|u(N_i) - u(N_j)\|_2^2 + 2\sigma^2 
\]

(9)

where \(u\) and \(v\) are the original and noisy images, respectively. \(\sigma\) is the standard deviation of noise and \(a > 0\) is the standard deviation of the Gaussian kernel. Equation (9) shows that the Euclidean distance preserves the order of similarity between pixels, therefore the most similar pixels of \(i\) in image \(v\) also are expected to be the most similar pixels of \(i\) in image \(u\).

### 2.4 Improved NL-means Algorithm

From the equation (7) we can see that the NL-means algorithm can correct the noisy image rather than try to separate the noise from the true image. The restored gray value of each pixel is obtained by the weighted average of the gray values of all pixels in the image. This algorithm has a great advantage in the preservation of the fine structure, details and texture. At the same time, the noise will be preserved. In addition, this algorithm will change the gray values of the image. To solve these problems Gaussian filter is used for smoothing the image first [23,24,25] by

\[
I(x, y) = I_0(x, y)^* G(x, y) 
\]

(10)
where $I_i(x,y)$ is the noisy image, $I(x,y)$ is the smoothed image, and $G(x,y)$ is the Gaussian kernel. Then the NL-means algorithm is carried out on the smoothed image $I(x,y)$. Gaussian smoothing does not change the characteristics of the image and will not affect the robustness of the NL-means algorithm. So equation (9) is still valid.

3. Experiments and Discussion

The proposed denoising schemes were evaluated using both simulated and real 2D gel electrophoresis data of the Escherichia coli proteome.

From equation (7) we can see that the key factor of the NL-means algorithm is the weight $w(i,j)$, which is determined by the similarity of the neighborhoods $v(N_i)$ and $v(N_j)$, where $N_i$ denotes a square neighborhood with fixed size of the pixel $i$. The determination of the neighborhood needs two parameters: the search window $t$ and the similar window $f$. The search window determines whether to find all the neighborhoods of the pixel $i$, and the similar window determines the complex of the computation. In addition, the filtering parameter $h$ also affects the value of the weight. Next, the effect of the above-mentioned three parameters on the denoising will be discussed, and then their appropriate values will be determined.

3.1 Similar Window

The size of similar window determines the size of the image expansion, and thus determines the complexity of the algorithm. So the selection of the window should take into account both the image quality and the computational complexity. In addition, the search window and the filtering parameter should be fixed before the selection of the similar window. Here a search window of size $21 \times 21$ is chosen and the filtering parameter $h$ is equal to the standard deviation of the noise. And the corresponding peak signal-noise ratio (PSNR) values are shown in Fig.1 and we can see that when the similar window is $7 \times 7$ the PSNR value is the highest. So in this paper we choose $7 \times 7$ as the size of the similar window.

3.2 Search Window

The size of search window determines whether denoising overwrites the entire image. That is to say, whether the weight $w(i,j)$ contains all the neighborhood of the pixel $i$. If the weight contains all the neighborhood of the pixel $i$, the details, the texture and the structure of the image will be preserved well. So a bigger size of search window is better. However, as the size of search window increases the computational complexity also increases. Therefore it is necessary to ensure the quality of the image and the amount of computation should been taken into account when the search window is chosen. Here a similar window of size $7 \times 7$ is taken and the filtering parameter $h$ still equal to the standard deviation of the noise. And the corresponding PSNR values are shown in Fig.2.

From the corresponding PSNR values we know that the bigger search window gives more effective denoising. At the same time, more details are protected and less intensities are altered. The bigger the search window is, the more accurate the weights are. While the amount of the computation also increases. Therefore, a search window of $85 \times 85$ is chosen to balance the quality of the image and the complexity of the computation.
3.3 Filtering Parameter

The filtering parameter controls the decay of the exponential function and then the decay of the weights. So it is important to select an appropriate filtering parameter $h$. From the formula

$$Z(i) = \sum_{j \in N(i)} e^{-\gamma |i-j|} I_j$$

we can see that the larger the $h$ is, the more slowly that the exponential function attenuates, and thereby the decay of the weights becomes weaken. So that more details and textures are preserved. But if the filtering parameter is too large, the noise will be retained as well. Through experiments just like the choice of the similar window and the search window, the filtering parameter is taken as the standard deviation of the noise.

3.4 Experiments on Synthetic 2DGE Images

In this section, the different algorithms: TIWT, TV, NL-means algorithm and improved NL-means algorithm are compared. For the real images, when all the information about the analyzed protein mixture comes from the gel image only and there is no alternative method of assessment of the studied mixture’s quantitative and qualitative composition, it is impossible to determine the real number of mixture components and their concentrations in real gel images. All such information comes from a better or worse estimation only. In order to give a quantitatively comparison of the above mentioned methods, we generate 120 synthetic gel images added with standard deviation noise of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 and 60. These simulated images have characteristics similar to the real gel images, and results of the synthetic gel images are also valid for the real gel images. The most common protein spot model is based on the 2-D Gaussian function [26],

$$S(x, y) = B + I \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2}\right)\exp\left(-\frac{(y-y_0)^2}{2\sigma_y^2}\right)$$

(11)
where $B$ is background intensity and $I$ is spot intensity, $x_0$ and $y_0$ are coordinates of control spots, $\sigma_x$ and $\sigma_y$ are standard deviation controlling the spread of the Gaussian independently in $x$ and $y$ directions. Having the ideal image and its version with an added noise, it was very simple to describe the quality of noise reduction as the difference between the filtered image and its ideal counterpart. The most commonly used measure of the noise reduction quality is the peak signal-noise ratio (PSNR),

$$PSNR = 10 \times \log \left( \frac{255^2}{MSE} \right)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \| I(i, j) - K(i, j) \|^2$$

where $I(i, j)$ is the original image and $K(i, j)$ is the denoised image. $m \times n$ is the size of the image. The performance of noise reduction with different denoising methods is presented by PSNR, as shown in Fig. 3.

![Figure 3. Comparative PSNR values for synthetic images using TIWT, TV, NL-means and improved NL-means.](image)

Fig. 3 shows that PSNR values of the improved NL-means are much higher than the other three methods on the whole. When the noise is relatively small the PSNR values of the improved NL-means are smaller than NL-means. That’s because WT is limited in capturing the directional information and TV alters seriously the pixel intensities. NL-means algorithm has the advantage of preservation of structure, details and texture, but the noise also is also protected.

### 3.5 Experiments on Real 2DGE Images

For the real gel images it is impossible to estimate the quality of denoising by means of PSNR because the ideal image is unknown and the synthetic images have similar characteristics with the real images. The best suitable method for the synthetic image is also the best suitable method for the real image. To verify this we present the profiles of the real images denoised by the four methods (Fig. 4).

Fig. 4 clearly shows that the improved NL-means algorithm outperforms the TIWT, TV and the NL-means algorithm.
Improved NL-means Algorithm for Two-dimensional Gel Electrophoresis Images Denoising
Hua-Mei Xin, Fengping Zhao

We can also see from Fig.4 (b) that the profile of the TIWT denoised image still has many oscillations, especially in the discontinuities. And the profiles of the TIWT denoised image has a bad approximation of the original profiles. Fig.4 (c) is the profile of TV denoised image. Compared with the profile of TIWT denoised image, it preserved the details effectively, especially when the horizontal axis is 300. However, TV algorithm retained the noise at the same time, which can be seen when the horizontal axis is 400. There is no doubt that it will interfere with protein detection and protein matching. Fig.4 (d) is the profile of NL-means algorithm denoised image. It protected the details very well and did not produce new ups and downs that cause the false protein spot, while the profile has many small oscillations in the flat region. In order to solve this problem we use Gaussian filter first. The profile of the improved NL-means algorithm denoised image is shown in Fig.4 (e). We can see that the image is more smooth and the details are preserved very well. Therefore, the improved NL-means algorithm is the more suitable method for the 2DGE images denoising.

The denoising effects on the real image using TIWT, TV, NL-means and the improved NL-means can be seen from Fig.5. Fig.5 (a) is the original image. Fig.5 (b) is the TIWT denoised image. It still has much noise. Fig.5 (c) is the TV denoised image and we can see clearly that it has many cartoon phenomenon. Fig.5 (d) is the NL-means denoised image. It is better than the TIWT denoised image and the TV denoised image. However, some noise still retained in the image. Fig.5 (e) is the improved NL-means denoised image. Compared with Fig.5 (b) ~ (d), it has the best visual quality. The image is smooth and the faint spots are preserved simultaneously.
Figure 5. Denoised image using the different methods: (a) original image (b) TIWT denoised image (c) TV denoised image (d) NL-means denoised image (e) improved NL-means denoised image.

4. Conclusions

Denoising is the first and important step in 2DGE images analysis and it has a profound impact on the spot detection and the spot matching. In this paper, we compared four methods for 2DGE images denoising: Translation Invariancy Wavelet Transform, Total Variation, NL-means algorithm and the improved NL-means algorithm. The WT can capture only limited directional information, i.e. vertical, horizontal and diagonal, and it may cause the pseudo-Gibbs phenomenon. Although the TIWT overcomes this phenomenon, there are still some oscillations. TV takes great advantage in the protection of the edges and the extraction of details, while it will alter the gray value during the denoising process. The NL-means algorithm can preserve the detail and texture well, but it also has oscillations, and this phenomenon is particularly evident in the flat areas. The improved NL-means algorithm overcomes these problems. It can preserve the structure, detail and texture, but also suppress the oscillations effectively. Experiment results show that the improved NL-means outperforms the other three methods in the PSNR values and the quality of the images. Therefore, the improved NL-means algorithm is the more suitable method for 2DGE images denoising, and has promising use in practical gel image analysis systems.

5. Acknowledgement

This work is supported in part by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (No.2009-36), Excellent Youth Scientist Award Foundation of Shan Dong Province (No. BS2011DX036), Scientific and Technological Developing Project of Shan Dong Province (No. 2010GGX10136).

6. References

Improved NL-means Algorithm for Two-dimensional Gel Electrophoresis Images Denoising
Hua-Mei Xin, Fengping Zhao


