EFFECTIVE DENOISING OF 2D GEL PROTEOMICS IMAGES USING CONTOURLETS

P. Tsakanikas ∗ I. Manolakos

Department of Informatics and Telecommunications
University of Athens, Greece
{tsakanik,eliasm}@di.uoa.gr

ABSTRACT

2D gel electrophoresis (2DGE) is the most commonly used method for protein separation. After gel scanning, images with a plethora of spot features are generated. In this paper we propose the use of the Contourlet Transform (CT) for 2D gels image denoising and compare it to the Wavelet Transform (WT). We show that contourlets not only achieve better average PSNR performance, but also preserve better, relatively to wavelets, the spot boundaries and alter less the intensities of significant spot features. Proper denoising of 2DGE images is essential in order to extract reliable spot features in proteomics workflows for biomarkers discovery.

Index Terms— Biomedical image processing denoising, Wavelets, Contourlets, 2D gel images, Proteomics.

1. INTRODUCTION

Proteomics is the field that studies multiprotein systems, focusing on the interplay of multiple proteins as functional components in a biological system. The first step in a typical proteomics analysis workflow is proteins separation, followed by quantification and differential expression analysis. Despite its limitations, 2D gel electrophoresis (2DGE) remains the most widely used protein separation method. Using 2DGE, individual proteins in a mixture are resolved in the first gel dimension according to their molecular weight and in the second dimension according to their isoelectric point. After gel scanning, protein species are depicted as spots of varying size and positions in the resulting gel image. An example of a typical 2DGE image is shown in Fig. 1.

A very important task in a proteomics study is the correct analysis and interpretation of 2D gel images through image analysis. This task aims at (i) the accurate detection and quantification of protein spots in a gel, followed by (ii) the matching of corresponding spots in sets of gels, as needed to identify proteins that can discriminate reliably between two states of a biological system (biomarkers discovery). 2DGE image analysis typically includes image preprocessing (noise suppression, artifacts removal, and background correction), segmentation (spot boundary detection) and protein expression quantification (spot volume estimation).

It is well known that 2DE gel images are inherently noisy due the gel’s susceptibility to dust and the imperfect image acquisition process [1]. The objective of this work is the effective denoising, i.e. increasing the SNR without inserting significant distortions to the image. Since denoising is at the very beginning of the preprocessing operations pipeline, if successful it may impact greatly on the results of downstream processing steps: (I) It prevents the over-estimation of the image background and helps extracting faint, yet significant, spots [2]. (II) it prevents the formation of misleading spots (artifacts), thus resulting in more truthful spot matching and more accurate determination of the significant spots to be further analyzed by mass spectrometry methods, (III) it leads to more accurate estimation of spot properties (e.g. spot volume) leading to improved spot differential analysis which is key for reliable biomarkers identification [3].

The noise suppression methods used in commercially available image analysis software packages are based on spatial filtering [4]. Despite their simplicity, these filters tend to distort severely spot edges and alter the intensity values of spot pixels. A comprehensive study [5] has recently shown that the Wavelet Transform (WT) outperforms spatial filtering, both in terms of PSNR and in terms of minimizing spot edge distortions. This is not surprising since 2DGE images are typical examples of non-stationary signals due to the large and unstructured variations in spot intensities and size, so it is impossible to distinguish signal from noise in the space or frequency domain alone. In this paper we show that the recently introduced Contourlet Transform (CT) [6] can do better than the Wavelet Transform in denoising 2DGE images. We will show that using the CT for 2DE gel image denoising not only improves PSNR but also better preserves the informative image details, relatively to the WT.

The rest of the paper is organized as follows: In Section 2
we justify the use of the CT over the WT for the problem at hand. Section 3 describes the methodology and datasets we used in the CT vs. WT evaluation for 2DGE image denoising. In Section 4 we present and discuss the results of the evaluation. Finally in Section 5 we summarize our findings and point to future work.

2. WHY USING CONTOURLETS?

Multirate signal analysis provides a natural way to represent images, starting from a coarse approximation and gradually adding details as we move towards finer scales. Image denoising in the space-frequency domain is a three-step procedure: 1) image decomposition, 2) coefficients thresholding, 3) inverse transformation to the original domain.

Despite its many advantages, the WT has also some known disadvantages: (i) wavelets are limited in capturing the geometry of image edges; after all, it is a separable extension of a 1-D transform, (ii) although wavelets are good at isolating the discontinuities at edge-points they do not exploit the smoothness along the edges, (iii) wavelets can capture only limited directional information (vertical, horizontal and diagonal) [6].

Recently, a new multirate transform that overcomes these limitations was introduced, called the Contourlet Transform (CT) [6]. The CT is a flexible multiresolution, local, and directional image decomposition method using contour segments. By construction it involves two filter bank stages: a Laplacian Pyramid (LP) followed by Directional Filter Banks (DFBs). The LP stage decomposes the image into frequency bands, while the DFBs decompose each detail band into several (but power of 2) directions.

The CT not only enjoys the multiscale and space-frequency localization properties of the WT, but also offers a high degree of directionality and anisotropy. Specifically, the CT uses basis functions that may be oriented at any power of 2 directions with flexible aspect ratios. With such a rich set of basis functions, contourlets can represent a smooth contour with fewer coefficients than wavelets. Only contourlets that match in both location and direction with image contours produce significant coefficients. The CT effectively explores the fact that image edges are fixed both in location and direction. Therefore, the CT can represent effectively images exhibiting anisotropic information, such as the 2DGE images.

3. EVALUATION METHODOLOGY

The steps we have followed in the CT vs. WT evaluation for 2DGE image denoising are summarized below: First we find the best set of parameters (basis function, number of decomposition levels, and number of directions at each frequency level) for each transform. This choice is crucial since it affects signal approximation and a wrong selection will lead to loss of information. Next, we use these parameters and compare the two transforms using two of the best known coefficient shrinkage methods. The comparative evaluation was done, first in terms of PSNR (the most commonly used noise reduction measure) and then in term of introduced image distortions. Finally, we applied Watershed based segmentation in order to substantiate the expected improvement in spot detection performance.

There are several coefficient shrinkage methods proposed in the wavelet literature. We selected to apply two popular methods, namely the BayesThres [7] and Bivariate shrinkage with local variance estimation (using on a 7x7 window) [8]. BayesThres, in conjunction with the WT, has been shown to perform very well in 2DGE image denoising [5]. Bivariate shrinkage has not been used for this problem before, but it has been shown to perform well with natural images [8].

For a proper evaluation we need a large number of images with the “ground truth” known. Therefore, we have created 100 synthetic, noise free 2DGE like images (to be called from now on Dataset1). Each image in Dataset1 has 512x512 pixels, 8-bits per pixel, and contains a randomly selected number of spots, ranging from 50 to 1000. Every spot is modeled as a 2D Gaussian function with a full covariance matrix. This spot modeling assumption is considered realistic and is used by most commercially available gel image analysis software packages [9]. Finally, we have added white Gaussian noise with standard deviation values \( \sigma_n = 10, 20, \) and 30 to each synthetic image.

For extra validation purposes, we have also used another set with 8 synthetic images (to be called from now on Dataset2) generated by M. Roger’s group and downloaded from [10]. Those images are of larger size (1024x1024) and have been created so as to exhibit the same statistical characteristics as real 2D gel images [9]. This property has justified their use in a comprehensive software packages comparison study [4]. Again, we have added noise with the same three variance levels as for Dataset1.

4. RESULTS AND DISCUSSION

Our first goal was to determine the most appropriate filters for the two CT stages. To do so we have tried the following filters for the LP stage: pkva, 9/7, 5/3, coiflet (10 vanishing points), Burt, haar, and for the DFBs: pkva, cd, 5/3, haar. Furthermore, we have considered 2 to 7 decomposition levels and 2 to 64 directions. As we move towards finer scales (levels), we double the number of directions for DFBs in every scale or in
Table 1. CT and WT best parameter set.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>WT</th>
<th>CT</th>
<th>WT</th>
<th>CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>LP filter</td>
<td>coiflet</td>
<td>coiflet</td>
<td>pkva</td>
<td>pkva</td>
</tr>
<tr>
<td>DFB filter</td>
<td>cd</td>
<td>cd</td>
<td>cd</td>
<td>cd</td>
</tr>
<tr>
<td>Directions</td>
<td>4,8,8</td>
<td>4,4,8,8</td>
<td>4,4,8,8</td>
<td>4,4,8,8</td>
</tr>
</tbody>
</table>

Table 2. Mean PSNR values obtained for 100 synthetic 2DE gel images (Dataset1).

<table>
<thead>
<tr>
<th>σ₀</th>
<th>WT-Bayes</th>
<th>CT-Bayes</th>
<th>WT-Biv</th>
<th>CT-Biv</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>40.07</td>
<td>39.58</td>
<td>39.82</td>
<td>39.02</td>
</tr>
<tr>
<td>20</td>
<td>35.59</td>
<td>34.89</td>
<td>34.98</td>
<td>34.09</td>
</tr>
<tr>
<td>30</td>
<td>33.04</td>
<td>32.30</td>
<td>32.36</td>
<td>33.36</td>
</tr>
</tbody>
</table>

Table 3. Mean PSNR values obtained for 8 synthetic 2DE gel images (Dataset2) [10].

<table>
<thead>
<tr>
<th>σ₀</th>
<th>WT-Bayes</th>
<th>CT-Bayes</th>
<th>WT-Biv</th>
<th>CT-Biv</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>41.38</td>
<td>41.77</td>
<td>40.39</td>
<td>42.33</td>
</tr>
<tr>
<td>20</td>
<td>37.32</td>
<td>37.85</td>
<td>36.11</td>
<td>38.27</td>
</tr>
<tr>
<td>30</td>
<td>35.00</td>
<td>35.42</td>
<td>35.70</td>
<td>35.99</td>
</tr>
</tbody>
</table>

VI - 271
details (spot borders), and alters less than the WT the intensities of true spot features. Denoising is a critical step applied at the beginning of the long pipeline of proteomics image analysis operations. Improving this step has a very positive effect on the quality of subsequent operations, such as spot detection, spot modeling and volume estimation, thus contributing significantly towards the important goals of correct spot matching and accurate spot quantification. Work in progress includes investigating CT based methods for multiplicative matching and accurate spot quantification. Improving this step has a very positive effect on the quality of subsequent operations, such as spot detection, spot modeling and volume estimation, thus contributing significantly towards the important goals of correct spot matching and accurate spot quantification. Work in progress includes investigating CT based methods for multiplicative and impulsive noise removal from gel images.

6. REFERENCES