16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

ACTIVE CONTOURS BASED SEGMENTATION OF 2DGE PROTEOMICS IMAGES

Panagiotis Tsakanikas and Elias S. Manolakos

Department of Informatics and Telecommunications, University of Athens
Panepistimiopolis, Ilissia, 15784, Athens, Greece
phone: + (30) 2107275312, email: {tsakanik,eliasm}@di.uoa.gr

ABSTRACT

The detection of protein spots in 2DGE images is one of the most important tasks in a proteomics data analysis workflow. All subsequent steps of differential expression analysis for biomarkers discovery depend on its effectiveness. In this study, we introduce the use of Active Contours without Edges coupled with Contourlet Transform - based image enhancement for extracting accurately the gel image foreground (regions with spots) from the background. We demonstrate, using both synthetic and real gel images, that the proposed approach extracts tight spot regions which do not include background areas but include almost all spots detected by PDQuest a popular commercial 2DGE image analysis package. Furthermore, our method does not require manual calibration for every new image in order to detect weak but often important "faint" spots. The method can be used as a first segmentation stage and/or to validate results of other spot detection algorithms.

1. INTRODUCTION

Proteomics is the field concerned with the large-scale study of proteins, particularly their structures and functions. Proteins are essential components of living organisms participating in physiological metabolic pathways of cells. Differential proteomics is especially useful in discovering biomarkers i.e. proteins that their presence or absence indicates a particular disease. Current research in proteomics requires that proteins be resolved first. To achieve this goal, proteins need to be separated. Protein separation can be performed using two-dimensional gel electrophoresis (2DGE), which discriminates proteins, first by their isoelectric point and then by their molecular weight. This process gives rise to protein spots of irregular shape and size on the gel. Once proteins are separated and quantified, they can be identified. Individual spots are cut out of the gel and cleaved into peptides with proteolytic enzymes. These peptides can then be identified using mass spectrometry methods and database search.

The dynamic range between the smallest and the largest concentrations of proteins can be from 10^6 to 10^{12} while the corresponding range of 2D electrophoresis is only 10^4. At the lower end of the dynamic range, many proteins are hardy visible in the gel image ("faint spots") while at the higher end, many proteins are saturated. The faint spots are well hidden due to noise and background variations across the gel image. Nevertheless, they are very important in differential protein expression analysis. So, missing faint spots leads to erroneous matching of spots across gels, and this may result to unreliable differential expression data.

Gel image analysis by computer software requires applying to every gel image of a proteomics experiment, a pipeline of operations which includes: image preprocessing (noise suppression, artifact removal, and background correction), segmentation (spot boundary detection) and protein expression quantification (spot volume estimation). In general, gel image processing, 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.

Segmentation of protein spots from 2DE gel images is probably the most important task in a gel-based proteomic study, since all the tasks before it, aim at its efficient application, and the tasks following it depend highly on its results. So, while the preprocessing tasks intend to enhance the performance of spot segmentation, a reliable and accurate segmentation makes the downstream proteomics data processing steps trustworthy and useful to the researchers.

Segmentation of 2DGE images requires partitioning them into areas of foreground (include protein spots) and background (with no protein spots). In this paper, we propose a new method based on Active Contours [1] that separates effectively those two areas in a way that: (i) reduces the number of missed faint spots, (ii) finds correct and tight borders for areas with spots, (iii) avoids over-segmentation. We demonstrate that the combined use of Contourlet Transform based image enhancement followed by foreground extraction using Active Contours without edges can meet these objectives. We also show, using both synthetic and real images that the estimated foreground includes the large majority of the spots detected by PDQuest [2], a popular commercial 2DGE image analysis software package. Our results are significant because the presented method can be used: (i) as a preliminary segmentation to extract the foreground accurately and facilitate spot detection and quatitation by spot modeling methods, (ii) to validate and establish a level of confidence on the results of other segmentation methods, especially for faint spots.

Previous work in 2DGE image segmentation includes methods using stepwise thresholding [3], second derivatives [4], the Watershed transform [5], and statistical spot modeling methods [6]. The stepwise thresholding approach is extremely sensitive to noise and artifacts, where additional criteria must hold in order to accept or reject the final connected areas. The second derivatives approach gives acceptable results only when proper noise suppression has been applied. Furthermore, it places the borders at the inside of the spots since the zero crossings of the second de-

rivative are associated with the steepest part of the spot rather than its beginning. The Watershed transform based method has the major disadvantage of over-segmentation. Although this can be addressed using marker controlled watersheds, the selection of a good set of markers is not a trivial task. Finally, the approaches using statistical spot modeling are difficult to apply without prior knowledge of spot shapes and sizes and it is known that they perform poorly in areas with overlapping spots especially if these foreground areas are not accurately estimated (usually this estimation is performed by mathematical morphology).

The rest of the manuscript is organized as follows: Section 2 describes the methodology we used and its justification. In Section 3 we present the experimental evaluation procedure and the data we used. In Section 4 we present and discuss the obtained results, and finally in Section 5 we summarize our work and point to future directions.

2. METHODOLOGY

Finding correct and as tight as possible regions containing protein spots will lead to more accurate spot detection and quantification. To segment the foreground pixels from the background we propose here the use of *Active Contours* (AC) and especially of Active Contours without Edges.

2.1 Active Contours

Active contours, initially introduced by Kass et al [7], are a very powerful tool for image segmentation and object tracking. The key idea is the evolution of a curve, or curves, also called "snakes", subject to constraints from the input image. The main disadvantage of Kass' models is that they are highly sensitive to the initial curve used (it must be known in advance). Also they are not applicable in our case, since it is impossible to manually set initial curves for every spot in a gel image. So, we need another approach that allows automatic topological changes of the curve. An AC algorithm that has such properties was introduced in [8] where the curve is modelled as a specific level set function of time in a higher dimensional surface. So, as time progresses, the curve changes according to the image topology in order to take the desired shape.

The classical approach for active contour models is to rely on an edge function *g,* depending on the gradient $|\nabla u_0|$ of the image u_0, in order to stop the evolving curve at the boundary of the desired objects. These models can detect only objects with edges that are well-defined by their gradient. In practice, the discrete gradients are bounded and the stopping function is never zero at the edges. So, the curve is likely to pass through the object boundary. This is especially true for faint spots that have very diffused and smooth boundary (they look like blurred objects), which means that their gradient is very small (see Figure 1(b)). To address these limitations of snakes with a gradient based stopping term, we adopted the AC models without edges introduced in [1] which use a different form of energy function to be minimized. Their main advantages are that they can detect

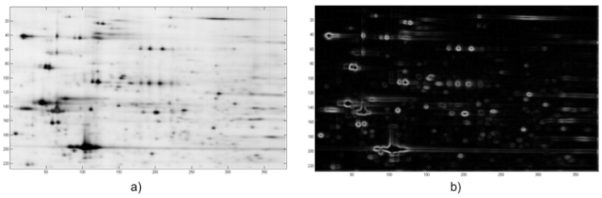

Figure 1 – a) Part of a 2DGE image, b) Gradient image of (a). We observe that many faint spots are missing in the gradient image.

objects with and without gradient, they can detect automatically topological changes due to the level set formulation, and the initial curve can be placed anywhere in the image. This kind of function is described below.

Let u_0 be the image to be segmented and C the initial evolving curve (the boundary of the object). Then consider the fitting term:

$$F_1(C)+F_2(C)=\int_{inside(C)}|u_0(x,y)-c_1|dxdy+\int_{outside(C)}|u_0(x,y)-c_2|dxdy \quad (1)$$

where c_1, c_2 are the average intensities of u_0 inside and outside of the snake C respectively. If the curve C is outside the object, then $F_1(C)>0$, $F_2(C)\approx 0$. If it is inside then, $F_1(C)\approx 0$, $F_2(C)>0$. If the curve is partially inside and partially outside the object then $F_1(C)>0$, $F_2(C)>0$. Finally, if the curve is exactly on the borders of the object then it holds that: $F_1(C)\approx 0$, $F_2(C)\approx 0$ and the fitting term is minimized. In addition, some regularization terms are included, as in the Mumfold-Shah model [1]. So, we will also try to minimize the total length of the curve and the area of the region inside the curve. Therefore, the energy function to be minimized is becoming:

$$E(C,c_1,c_2)=\mu\cdot Length(C)+\nu\cdot Area(C)+\lambda_1 F_1(C)+\lambda_2 F_2(C) \quad (2)$$

where $\mu\geq 0$, $\nu\geq 0$, $\lambda_1>0$, $\lambda_1>0$ are parameters to be determined. So, the goal is to find C,c_1,c_2 such that $E(C,c_1,c_2)$ is minimized.

2.1.1 Active Contour Parameters

Parameter μ controls the importance of the length of the curve C in the minimization process (if $\mu>>0$ then a few large closed curves will remain in the steady state, compared to many small ones). After experimentation we chose the value $\mu=0.006*255^2$. This intermediate value enables the detection of more faint spots, as we relax the constraint of length and so image areas with relative small values of average intensities are included in the final snakes. Also, it promotes tightness i.e. the separation of areas with distinct spots, rather than their inclusion (if they are close to each other) into large regions with a large percentage of background pixels.

Parameter ν controls the importance of the area inside C (if $\nu>>0$ then the curve C is forced to move strictly inwards). In our case due to the fact that faint spots, and not

only them, look like blurred objects, we found that in order to completely enclose a spot (without letting the snake curve crash inwards) a value of $v=0$ is the best choice.

The relationship of λ_1 to λ_2 determines which side of the snake (inside or outside) has higher importance in minimizing the integrals of Eq.2. This is useful in segmenting blurred images [1]. So, if we wish to completely enclose the blurred object e.g. a faint spot, we should choose $\lambda_1 >> \lambda_2$. After experimentation, we chose $\lambda_1=10$ and $\lambda_2=1$. These values let the snakes enclose areas of spots which are very close to each other while also separating distinct spots.

2.2 Image Enhancement

Image enhancement using the Contourlet Transform was applied to reveal faint spots to the extent that after applying AC we can separate a good number of them and avoid their inclusion into large regions with a lot of spots.

The Contourlet Transform (CT) is a very powerful multiresolution technique that, as we have demonstrated in [9], can be used to denoise 2DGE images effectively. In this study we used the Non-Subsampled Contourlet Transform (NSCT) [10], which is also a shift-invariant transform. NSCT provides not only multiresolution analysis but also geometric and directional representation. Since weak edges (such as those of faint spots) are geometric structures, while noise is not, we can exploit this property of NSCT to enhance them.

We, therefore, analyze the 2DGE images with the NSCT using 3 decomposition levels with (4,8,8) directions respectively from coarser to finest level and the *maxflat* filter. Then we apply an enhancement function to the resulting coefficients of the "middle" frequency scales. We do not apply it at the coarser and the finest scale, because we do not want to enhance any background structures or high frequency components (noise or high frequency edges). Afterwards, the coefficients that represent the spots are mainly concentrated in the middle frequency scales.

Throughout the literature there are numerous enhancement functions proposed. Our objective is to enhance weak edges of blurred objects. So, our enhancement function should amplify certain mid-range coefficients only. Thus we chose the simple function given below (proposed in [11]) which also eliminates very small coefficients, corresponding to tiny objects - artefacts:

$$E(u) = u \cdot sign(u) \cdot \tanh(b \cdot n) \cdot (1 + c \cdot \exp(-n^2)) \quad (3)$$

where $n = 2.5 \cdot u/(t \cdot M)$, with u being the coefficient amplitude in the transform domain, M the amplitude of the maximum coefficient. The term $(t \cdot M)$ is a threshold so that coefficients exceeding it are linearly amplified. Parameters b and c, determine the gain in each amplitude interval. The main advantages of this method are its simplicity and that it does not alter the locations of discontinuities. For further details the interested reader is referred to [11]. We should also mention that enhancement is a common practice in 2DGE image analysis. Most faint spots are missed by commercial software packages and the user has to adjust the contrast of each image manually in order to get those spots included in the analysis. In our method the same enhancement function is applied automatically to all images.

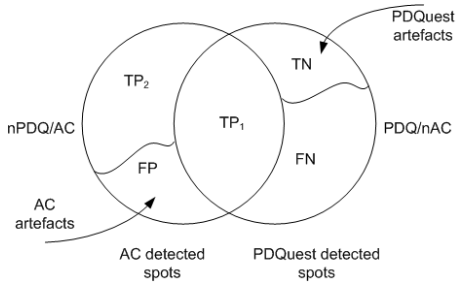

Figure 2 - Diagram of spot subsets in the spot detection experiment.

3. EVALUATION PROCEDURE

3.1 Materials

In order to evaluate our AC based segmentation approach we have used six synthetic images of size 1024x1024 pixels and 8 bit gray scale intensities. They have been created by Rogers et al. [12] so as to highly resemble real 2DGE images. These images can be freely downloaded from the authors' site [13] and have been used in software package evaluations [14]. In those images, we added white Gaussian noise with $\sigma_n = 10$ (usually inherent in 2DGE images). In addition, we used two real 2DGE images (8-bit) described in [15] which are also freely available and have been used as benchmarks in [15, 16].

3.2 Evaluation method

We describe next the experimental evaluation procedure that we have followed. First we denoised all images using our CT-based approach [9] and then we performed background correction using mathematical morphology. In particular we applied the opening morphological operation and subtract the resulting image from the original one. Then we enhanced the images by applying the NSCT followed by the function of Eq. 3. Finally we applied the AC without edges on the resulting image; the resulting spot region boundaries were applied to the original images. In this way, we obtain at the end of the process an image where the foreground (protein spot regions) is separated from the background.

We compared the results obtained by applying the proposed method with those obtained using PDQuest v. 7.2.0. In Figure 2 we show pictorially the different possible spot subsets. Spots that AC missed but were found by PDQuest (denoted as PDQ/nAC) are partitioned into two subsets; existing spots that our AC base method missed (false negatives, FN) and extraneous spots (true negatives, TN = PDQuest artefacts) that AC correctly ignored. For those spots that AC detect but PDQuest missed (denoted as nPDQ/AC) we also consider two subsets. The ones that AC correctly found because they do exist (true positives, TP_2), and those that AC found but do not exist (false positive, FP = AC artefacts).

In order to evaluate quantitatively our results we calculated two metrics, sensitivity and confidence, which are defined as [17, 18]:

Image	PDQ	PDQ/AC	% PDQ/AC	PDQ/nAC	% PDQ/nAC	nPDQ/AC	%nPDQ/AC
1a	1112	1090	98,02%	22	1,98%	13	1,19%
2a	1315	1283	97,57%	32	2,43%	7	0,55%
MP1	262	256	97,71%	6	2,29%	13	5,08%
MP2	265	242	91,32%	23	8,68%	11	4,55%
MP3	227	223	98,24%	4	1,76%	24	10,76%
Rj1	146	123	84,25%	23	15,75%	4	3,25%
RGA	948	919	96,94%	29	3,06%	9	0,98%
RGB	1040	1018	97,88%	19	1,83%	40	3,93%

Table 1 – Evaluation results. PDQuest spots in AC extracted foreground regions were all real spots (PDQ/AC = TP_1). We also report spots detected by PDQuest and not included in our foreground areas (PDQ/nAC = (TN+FN)) and spots in our foreground areas not detected by PDQuest (nPDQ/AC = (TP_2+FP)).

Image	PDQ	AC/PDQ (TP_1)	PDQ/nAC		nPDQ/AC		S	C
			FN	TN	FP	TP_2		
1a	1112	1090	5	17	5	8	99,55%	99,55%
2a	1315	1283	9	23	5	2	99,30%	99,61%
MP1	262	256	2	4	10	3	99,23%	96,28%
MP2	265	242	14	9	4	7	94,68%	98,42%
MP3	227	223	1	3	6	18	99,59%	97,57%
Rj1	146	123	11	12	1	3	91,97%	99,21%
RGA	948	919	20	9	6	3	97,88%	99,35%
RGB	1040	1018	12	7	13	27	98,86%	98,77%

Table 2 – Evaluation results and the categories shown in Fig. 2. Sensitivity is above 91% and Confidence above 96% for all images.

$$S = \frac{TP_1 + TP_2}{(TP_1 + TP_2) + FN} \quad (4)$$

$$C = \frac{TP_1 + TP_2}{(TP_1 + TP_2) + FP} \quad (5)$$

Sensitivity assesses how effective we are in identifying true spots on a gel image. Confidence measures the percentage of the true spots detected by our approach relatively to the total number of detected spots.

4. RESULTS & DISCUSSION

Since we use AC for accurate foreground estimation, in our evaluation we focus on whether the real spots are indeed included in the extracted image foreground and whether the foreground areas extracted are not larger than necessary (i.e. they are tight). The results obtained are summarized in Tables 1 and 2.

Knowing that PDQuest is performing very well in segmenting these images we conclude that our approach correctly reports the foreground regions in a wide collection of different of 2DGE images. We also see that we succeed in avoiding some PDQuest detected artefacts (TN) but also fail to detect very few spots that should have been detected (FN).

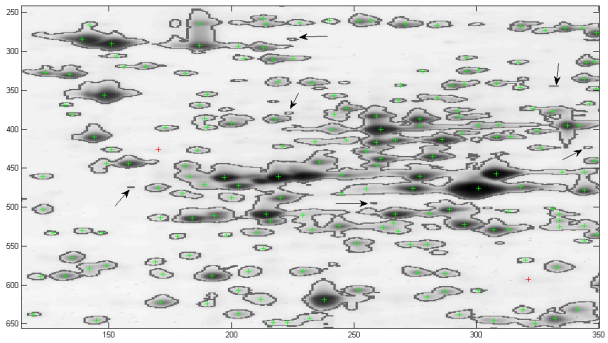

Figure 3 – Part of Raman's Gel B. Crosses indicate the spots detected by PDQuest.

On the other hand, we detect areas that contain spots missed by PDQuest (TP_2), but also where some artefacts (FP) were introduced. All of those cases can be visualized in Figures 3 and 4, where we present representative image areas (boundaries depict AC contours and crosses the centre of protein spots detected by PDQuest). In Table 1 we can see that the ratio of spots missed by our approach compared to PDQuest is pretty small (<3%, except for the Rj1 image) which also indicates that the proposed method results are highly reliable. In addition, as we can see from Table 2, the proposed approach achieves sensitivity over ~91% for all images and a confidence above ~96%. These results indicate that ACs can be very effective in confining protein spots into tightly bounded spot areas.

We should mention that in order to make PDQuest perform well enough we had to manually specify for every image a typical faint spot and fix the sensitivity parameter. We were also getting very different segmentation results depending on the choice of the typical faint spot. So, after a lot of trials we selected for each image among the typical faint spots we tried the one that leads to detecting the maximum number of faint spots. On the opposite tuning of the AC parameters was performed once using a single image and then the same parameters set was applied to all other images. So, we conclude that the proposed approach requires minimal user intervention, while achieving consistent results with different images. This is very useful when we want to segment technical replicate gels produced from the same biological samples; we do not have to change the parameters in order to get similar results among them.

In Figures 3 and 4 we present some typical examples from the real 2DGE images (RGA, RGB) in Raman's set. In those figures we can see the borders separating the foreground from the background generated by the proposed AC based approach and the spots' centres detected by PDQuest (green crosses). We observe that the borders generated by the proposed approach include the large majority of the spot regions. Furthermore the resulting borders are tight enough as needed to accurate detect and quantify the spots using spot modelling. In addition, we can see that there exist: (i) Some cases detected by AC but missed by PDQuest (TP_2, bordered areas without crosses inside them). (ii) Some artefacts introduced by the proposed approach (FP) indicated by arrows. Most of them can be discarded during the spot modelling

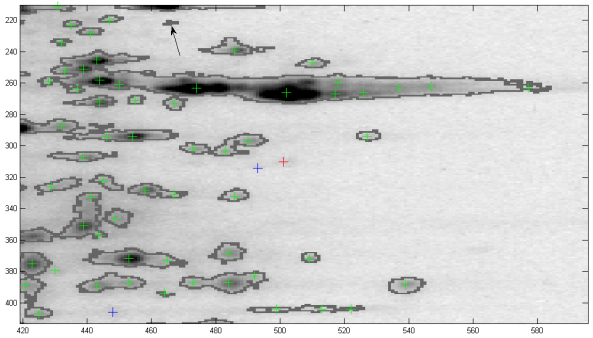

Figure 4 – Part of Raman Gel A. Crosses indicate the spots detected by PDQuest.

operation which normally follows after spot regions segmentation. (iii) Some missed true spots (FN) depicted by the red crosses with no borders surrounding them, and (iv) PDQuest introduced artefacts (TN) marked as blue crosses with no surrounding borders.

The visual inspection of these image regions suggests that our AC based segmentation performs on average equally well compared to PDQuest. An interesting application is that someone can use the proposed method to validate the results obtained and calibrate a commercial image analysis software that may be available. For example, we may compare the results (as we did for PDQuest) and check if all the interesting areas of the gel have been identified by the software package. If not, the user may change the parameters of the software so as to include interesting areas rather than missing them (e.g. selects a different typical faint spot and repeat segmentation).

The process of partitioning foreground areas into individual spots is a known challenging problem. Overlapping spots in some cases are so difficult to separate even by visual inspection. This problem is generally confronted by fitting a parametric spot shape model into each foreground segment of a 2DGE image. However, this is not a trivial task mainly due to the need for choosing the proper model and the number of models to use. The work presented here makes it easier by providing tight and truthful boundaries for overlapping spot regions.

5. CONCLUSIONS

To the best of our knowledge, this is the first attempt to use Active Contours in 2DGE image analysis. Accurate and correct spot areas segmentation is a prerequisite for spot detection and quantification. We have shown that the proposed AC based segmentation achieves comparable results with a mature commercial tool for 2DGE image analysis (PDQuest) but with much less user intervention. We are currently investigating the benefits of using multiphase Active Contours. We also believe that clustering the different kinds of spots (faint, oversaturated etc) as different image objects will result to increased accuracy and detection efficiency.

Acknowledgements: This paper is part of the 03ED306 research project, implemented within the framework of the "Reinforcement Program of Human Research Manpower" (PENED) and co-financed by National and Community Funds (75% from E.U.-European Social Fund and 25% from the Greek Ministry of Development-General Secretariat of Research and Technology). The authors would like to thank the funding agency for their support.

REFERENCES

[1] T. F. Chan, L. A. Vese, "Active Contours without Edges", *IEEE Trans. on Image Processing*, vol. 10, pp. 266-277, Feb. 2001.

[2] J. I. Garrels, "The Quest system for Quantitative analysis of two-dimensional gels", *Journal of Biological Chemistry*, vol. 264, pp. 5269-5282, 1989.

[3] P. Cutler, G. Heald, I. R. White, J. Ruan, "A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection", *Proteomics*, vol. 3, pp. 392-401, 2003.

[4] R. D. Appel, J. R. Vargas, P. M. Palagi, D. Walter, D. F. Hochstrasser, "Melanie II – a third generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms", *Electrophoresis*, vol. 18, pp. 2735-2748, 1997.

[5] K. P. Pleissner, F. Hoffman, K. Kriegel, C. Wenk, S. Wegner,et al., "New algorithmic approaches to protein spot detection and pat tern matching in two-dimensional electrophoresis databa*ses", Electrophoresis*, vol. 20, pp. 755-765, 1999.

[6] E. Bettens, P. Scheunders, J. Sijbers, D. Van Dyck, L. Moens, "Automatic segmentation and modeling of two-dimensional electrophoresis gels", in *Proc. ICIP 1996*, vol. 2, pp. 665-668.

[7] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active Contour Models", in *Proc. of the 1st Int. Conf. on Computer Vision 1987*, pp. 259-268, 1987.

[8] S. Osher, J. A. Sethian, "Front Propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi Formulation", *J. Comput. Phys.*, vol. 79, pp. 12-49, 1998.

[9] P. Tsakanikas, I. Manolakos, "Effective Denoising of 2D Gel Proteomics Images using Contourlets", in *Proc. ICIP 2007*, San Antonio, Texas, USA, September 16-19,2007, pp. VI: 269-272.

[10] A. L. Cunha, J. Zhou, M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications", *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089-3101, Oct. 2006.

[11] E. Nezhadarya, M. B. Shamsollahi, "Image Contrast Enhancement by Contourlet Transform", in *48th International Symposium ELMAR 2006*, Zadar, Croatia, June 7-9, 2006, pp. 81-84.

[12] M. Rogers, J. Graham, R. P. Tonge, "Statistical models of shape for the analysis of protein spots in two-dimensional electrophoresis gel images", *Proteomics*, vol. 3, pp. 887-896, 2003.

[13] www.isbe.man.ac.uk/~mdr/content.php?f=electrophoresis.

[14] M. Rogers, J. Graham, R. P. Tonge, "Using statistical image models for objective evaluation of spot detection in two-dimensional gels", *Proteomics*, vol. 3, pp. 879-886, 2003.

[15] B. Raman, A. Cheung, M. R. Marten, "Quantitative comparison and evaluation of two commercially available, two dimensional electrophoresis image analysis software packages, Z3 and Melanie", *Electrophoresis,* vol. 23, pp. 2194-2202, 2002.

[16] A. T. Rosengren, J. M. Salmi, T. Aittokallio, et al., "Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels", *Proteomics*, vol. *3*, pp. 1936–1946, 2003.

[17] Y. Kim, J. Kim, Y. Won, Y. In, "Segmentation of Protein Spots in 2D Gel Electrophoresis Images with Watersheds Using Hierarchical Threshold", *Lecture Notes in Computer Science*, ISSU 2869, pp. 389-396, 2003.

[18] http://en.wikipedia.org/wiki/Sensitivity_%28tests%29.