A LEVEL SET APPROACH FOR PROTEOMICS IMAGE ANALYSIS

Michalis Savelonas, Eleftheria Mylona, Dimitris Maroulis

Department of Informatics and Telecommunications, University of Athens
Panepistimiopolis. Ilissia, 15784, Athens, Greece
phone: + (210) 7275317, email: {msavel, emylona, dmaroulis}@di.uoa.gr
web: http://rtsimage.di.uoa.gr

ABSTRACT

This paper addresses the detection of protein spots in proteomics images. A novel level set approach is proposed, which is capable of separating spots belonging to multiplets. The proposed approach copes with the presence of noise and the inhomogeneous background, which are major issues in proteomics image analysis. It has been experimentally compared with a recently published method for spot detection. The results of the experiments performed indicate that the previously proposed method tends to falsely identify multiplets as singlets, whereas the proposed approach succeeds in spot separation.

1. INTRODUCTION

The genomics revolution was only the beginning of an enormous metamorphosis of life sciences. The analysis of the proteome has provoked even more radical changes in biological research. Most biological functions are carried out by proteins that interact with each other within a complex biological system. The proteome is defined as a protein complement in a specific cell, tissue or organism. However, it is more complicated than the genome since a gene may encode a number of different proteins. In other words, genes are just the instructions for making proteins whereas proteins make life.

Proteomics are mainly applied in medicine for disease treatment, as well as for drug discovery by analyzing possible protein alterations. For instance, the proteomes of a cancerous and a benign cell are different. The proteins that are missing in the benign cell can be targets for anti-cancer drugs. Thus, protein separation is a major part of proteomics.

The most widely used method for protein separation is two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) [1-3]. The 2-D PAGE process involves the protein separation by two different physical properties: the isoelectric point in the first dimension and the molecular weight in the second. This method can easily detect any alterations in charge and mass due to the fact that it is highly unlikely for two different proteins to resolve to the same position in both dimensions. Moreover, it allows the identification of thousands of proteins with computer assisted software programs.

2-D PAGE results in grey level images illustrating the separated proteins as dark spots on a bright background or vice versa. Common unwanted factors that influence the results emerging from the analysis of 2-D PAGE images are: the presence of noise, dust particles, fingerprints, cracks in the surface and other artefacts such as streaks or tails that are not related to proteins. The main challenges of 2-D PAGE image analysis are: the inhomogeneity of image background, the presence of noise and the existence of overlapping spots.

The detection of overlapping spots has recently drawn the researchers' attention. This trend is motivated by the fact that a large part of the proteins in a gel tend to belong to doublets, triplets or multiplets, which are spot clusters consisting of two, three or multiple proteins, respectively [4]. Most false estimations on the presence or absence of spots stem from the erroneous interpretation of such spot clusters, whereas the challenge is to detect individual proteins. An example of a 2D-PAGE image area containing spot clusters is illustrated in Fig.1.

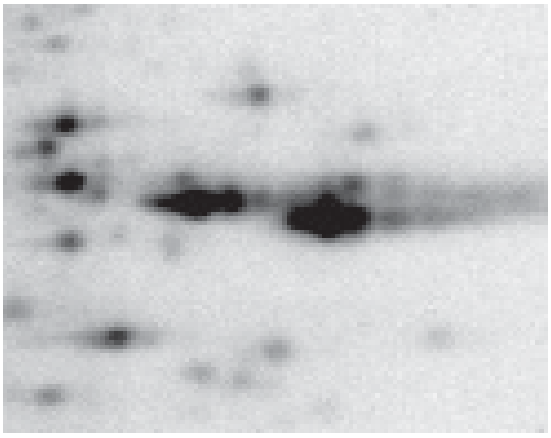

Figure 1 – Complex region containing overlapping spots.

Most common approaches to spot detection include watersheds [5-6], stepwise thresholding [7] and morphology [8-9]. However, watersheds commonly require inner markers so as to confront the issue of over-segmentation [10], whereas stepwise thresholding and morphology fail in the presence of artefacts and noise [11]. In addition, they falsely identify doublets or triplets as singlets.

Level set approaches [12-13], which have been efficaciously used in image analysis, appear well-suited in 2D-PAGE spot detection. They are topologically adaptable facilitating contour merging or splitting, so that multiple boundaries can be detected. In addition, region-based level set approaches, such as [13], are capable of detecting objects

 ISSN 2076-1465

defined by weak edges, like protein spots. Finally, they can be relatively insensitive to noise by involving integral operators, which provide an inherent noise filtering mechanism.

This work introduces a level set approach for the detection of multiple overlapping spots on 2-D PAGE images. The proposed approach is an expansion of the active contour without edges model, originally proposed by Chan and Vese [13], and enables the detection of individual proteins in spot clusters.

The rest of this paper is organised as follows: Section 2 and 3 describe the Active Contour without Edges model and the proposed approach respectively. Section 4 presents the obtained results on real 2-D PAGE images and finally, Section 5 summarizes the conclusions of this study.

2. ACTIVE CONTOUR WITHOUT EDGES

The active contour without edges model is based on the Mumford-Shah functional [14]. The segmentation of an image $u_1 : \Omega \to R$, where Ω is a bounded open subset of R^2 with $\partial\Omega$ its boundary, is formulated as a minimization problem: we seek for the infimum of the energy functional $F(c_1^+, c_1^-, C)$:

$$\begin{aligned} F(c_1^+, c_1^-, C) &= \mu \cdot length(C) \\ &+ \lambda_1^+ \int_{inside(C)} | u_1(x,y) - c_1^+ |^2 \, dxdy \\ &+ \lambda_1^- \int_{outside(C)} | u_1(x,y) - c_1^- |^2 \, dxdy \end{aligned} \tag{1}$$

where $C(s) : [0,1] \to R^2$ is a piecewise parameterized curve, c_1^+ and c_1^- represent the average intensities of u_1 in the foreground and in the background respectively and parameters $\mu > 0$ and $\lambda_1^+, \lambda_1^- > 0$ are weights for the regularizing term and the fitting terms, respectively. This model uses the level set formulation [12], where the curve $C \subset \Omega$ is represented implicitly by the zero level set of a Lipschitz function $\phi_1 : \Omega \to R$, such that:

$$\begin{aligned} &C = \{(x,y) \in \Omega : \phi_1(x,y) = 0\}, \\ &inside(C) = \{(x,y) \in \Omega : \phi_1(x,y) > 0\}, \\ &outside(C) = \{(x,y) \in \Omega : \phi_1(x,y) < 0\} \end{aligned} \tag{2}$$

The average intensities in the foreground (inside the contour) and in the background (outside the contour) c_1^+ and c_1^- are considered for all the pixels in the respective regions and are estimated by:

$$c_1^+(\phi_1) = \frac{\int_\Omega u_1(x,y) H(\phi_1(x,y)) dxdy}{\int_\Omega H(\phi_1(x,y)) dxdy} \tag{3}$$

$$c_1^-(\phi_1) = \frac{\int_\Omega u_1(x,y)(1 - H(\phi_1(x,y))) dxdy}{\int_\Omega (1 - H(\phi_1(x,y))) dxdy} \tag{4}$$

where H is the Heaviside function. To start with, the contour is initialized and at each time step, the average foreground and background intensities are updated. By keeping c_1^+ and c_1^- fixed, and minimizing F with respect to ϕ_1, the associated Euler-Lagrange equation for ϕ_1 is deduced. Parameterizing the descent direction by an artificial time $t \geq 0$, ϕ_1 is determined by solving the following equation:

$$\frac{\partial \phi_1}{\partial t} = \delta(\phi_1)[\mu \cdot div(\frac{\nabla \phi_1}{|\nabla \phi_1|}) - \lambda_1^+ (u_1 - c_1^+)^2 + \lambda_1^- (u_1 - c_1^-)^2] = 0 \tag{5}$$

where $t \in (0, \infty), (x,y) \in \Omega$ and δ is the one-dimensional Dirac function. The contour evolves according to Eq. (5) guided by the average foreground and background intensities. In a 2D-PAGE image, the average protein spot intensity differs from the one of the background. The active contour without edges model is capable of detecting the external spot boundaries, however fails to separate overlapping protein spots.

3. PROPOSED LEVEL SET APPROACH

A novel level set approach is proposed, aiming to overcome the limitations of the active contour without edges model. The contour is initialized and evolves by iteratively solving Eq. (5), until it converges to the external boundaries of the protein spots. Average intensities c_1^+ and c_1^- are iteratively estimated by Eq. (3) and (4) respectively. The results obtained after contour's convergence provide the actual external boundaries of the spot regions. Contour evolution proceeds on the inside of the already identified spot regions and is guided by c_2^+ and c_2^-, which are estimated by the following equations:

$$c_2^+(\phi_2) = \frac{\int_\Omega u_2(x,y) H(\phi_2(x,y)) dxdy}{\int_\Omega H(\phi_2(x,y)) dxdy} \tag{6}$$

$$c_2^-(\phi_2) = \frac{\int_\Omega u_2(x,y)(1 - H(\phi_2(x,y))) dxdy}{\int_\Omega (1 - H(\phi_2(x,y))) dxdy} \tag{7}$$

where ϕ_2, is the new zero level set function and u_2 is the sub-image of u_1 for which $\phi_1 > 0$. The sub-image u_2 consists of all singlets or multiplets. The average intensities c_2^+ and c_2^- correspond to the slightly different intensity distributions of the individual spots, which were initially being

grouped within c_1^+ in cases of overlap. Contour evolution proceeds until convergence, when each individual protein spot is identified, even if it belongs to a spot cluster.

4. RESULTS

The experimental evaluation of the proposed approach included its application on real 2-D PAGE images, as well as on two standard gels used for spot detection analysis and gel matching studies which can be downloaded for free [15]. The 2-D PAGE images were quantized at 8-bit gray level depth, whereas the algorithm has been implemented in Matlab 7.0 and executed on a 3.2 GHz Intel Pentium workstation. Parameters $\lambda_1^+, \lambda_1^- > 0$ and were set to 1, 1 and 0.01 respectively [16]. Figure 2 illustrates a typical 2-D PAGE image which has been inverted so that the average protein spot intensity is higher than the one of the background.

Figure 3 illustrates segmentation results obtained by the application of the active contour without edges model on the 2-D PAGE of Figure 2. Although the contour has identified the external boundaries of the overlapping spot regions, it has failed to separate the overlapping spots. On the contrary, the proposed approach effectively separates the detailed subimages a, b and c of Figure 3, as can be seen in Figure 4.

Figure 5 depicts a part of a standard 2-D PAGE image used in spot detection where the average protein spot intensity is lower than the one of the background [15]. Four image regions associated with multiplets are marked with a-d. This image has been used by Tsakanikas et al. [17] in order to evaluate their method. The detection results obtained by applying this method are illustrated in Figure 6.

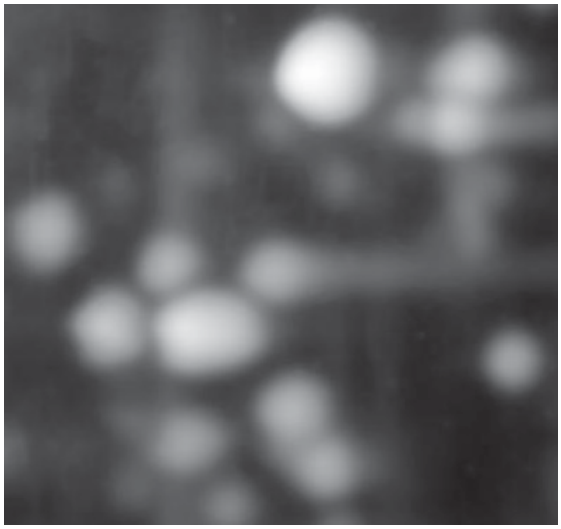

Figure 2 – Part of a 2-D PAGE containing white protein spots.

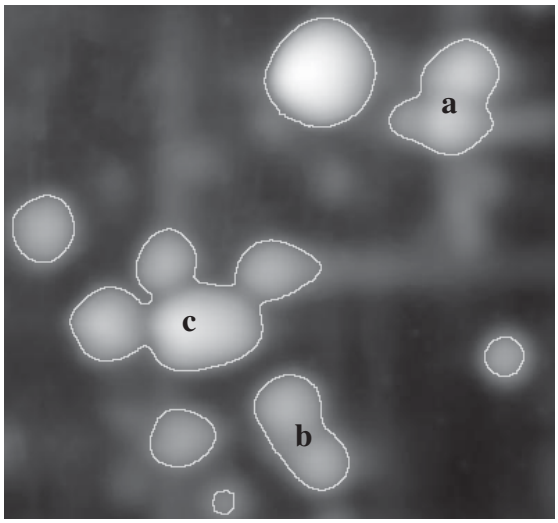

Figure 3 – Segmentation results obtained by the active contour without edges model on the 2-D PAGE image of Figure 2.

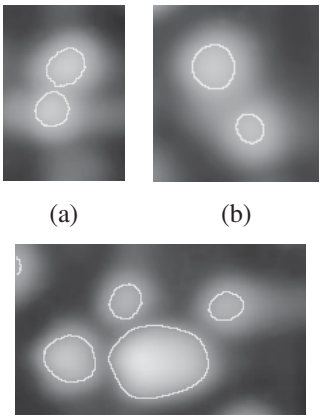

Figure 4 – Detection results obtained by the proposed approach on the 2-D PAGE image of Figure 3.

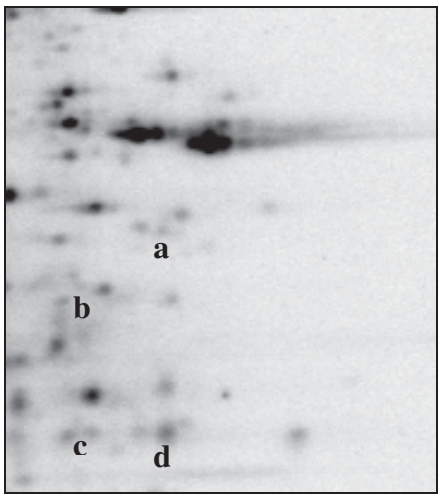

Figure 5 – Part of a standard 2-D PAGE image containing black protein spots. Four regions associated with multiplets are marked with a-d.

Figures 7(a-d) illustrate the detailed sub-images corresponding to the multiplets marked with a-d in Fig. 5. Figures 7(a_1-d_1) illustrate the corresponding sub-images of the ground truth. Figures 7(a_2-d_2) illustrate the detection results obtained by the application of the method of Tsakanikas et al. [17] on each multiplet, whereas Fig. 7(a_3-d_3) illustrate the detection results obtained by the proposed approach on the same multiplets. Considering the ground truth, it is evident in Fig. 7(a_2-d_2) that the multiplets are identified by the method of Tsakanikas et al. as singlets. On the contrary, it is clear in Fig. 7(a_3-d_3) that the proposed approach is capable of separating multiple overlapping protein spots and thus, can be integrated within a segmentation framework.

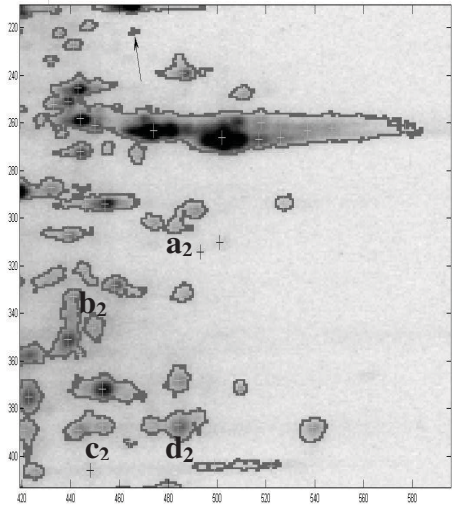

Figure 6 – Detection results obtained by the active contour without edges model on the 2-D PAGE image of Fig. 5 by Tsakanikas et al. [17].

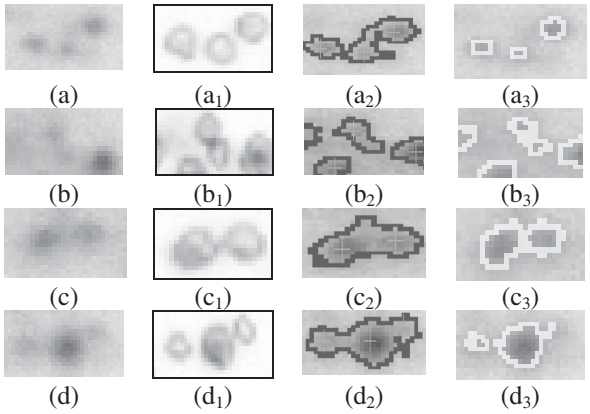

Figure 7 – (a-d) Detailed sub-images of Figure 5, (a_1-d_1) Detailed sub-images of the ground truth, (a_2-d_2) Detection results obtained by Tsakanikas et al. [17] in cases of multiplets, (a_3-d_3) Detection results obtained by the proposed approach in cases of multiplets.

5. CONCLUSIONS

In this paper, a novel level set approach for proteomics image analysis is proposed. The proposed approach enables the separation of protein spots which belong to multiplets. The experimental results demonstrate that it outperforms a recently published proteomics image analysis method [17]. Future perspectives of this work involve the enhancement of the contour initialization process, as well as experimental comparisons with renowned proteomics image analysis software, such as PDQuest and Progenesis [18].

6. ACKNOWLEDGEMENTS

We would like to thank the Biomedical Research Foundation of the Academy of Athens for the provision of 2-D PAGE images. This work was partially supported by the European Social Fund and the Ministry of National Education of Greece through the HRAKLEITOS II program and the National and Kapodestrian University of Athens, Special Account of Research Grants.

REFERENCES

[1] P.H.O' Farell, "High resolution two-dimensional electrophoresis of proteins," *Journal of Biological Chemistry*, vol. 250, pp. 4007–4021, 1975.

[2] R. Westermeier, *Electrophoresis in practice: A guide to theory and practice*. VCH, Weinheim, 1993.

[3] A. W. Dowsey, J. D. Michael, G. Yang, "The role of bioinformatics in two-dimensional gel electrophoresis," *Proteomics*, vol. 3, pp. 1567-1596, 2002.

[4] M. B. Ryke, "Image segmentation and multivariate analysis in two-dimensional gel electrophoresis", *PhD thesis*, Norwegian University of Science and Technology, Trondheim, 2007.

[5] K.P. Pleissner, F. Hoffman, K. Kriegel, C. Wenk, S. Wegner, A. Sahlstrom, H. Oswald, H. Alt and E. Fleck, "New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis databases," *Electrophoresis*, vol. 20, pp. 755-765, 1999.

[6] L. Vincent, P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 13, pp. 583-598, 1991.

[7] 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.

[8] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Im. Proc.*, vol. 2, pp. 176-201, 1993.

[9] G. W. Horgan, C. A. Glasbey, "Uses of digital image analysis in electrophoresis," *Electrophoresis*, vol. 16, pp. 298-305, 1995.

[10] Y. Kim, J. Kim, Y. Won, Y. In, "Segmentation of protein spots in 2-D gel electrophoresis images with watershed using

hierarchical threshold," *Lecture Notes in Computer Science (LNCS)*, vol. 2869, pp. 389-396, 2003.
[11] K. Takahashi, Y. Watanabe, M. Nakazawa, A. Konagaya, "Fully automated spot recognition and matching algorithms for 2-D gel electrophoretogram of genomic DNA," *Genome Inf. Ser. Workshop*, vol. 9, pp. 161-172, 1998.
[12] S. Osher, J.A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulation," *Journal Computer Physics,* vol. 79, pp. 12-49, 1988.
[13] T.F. Chan, L.A. Vese, "Active contour without edges," *IEEE Trans. Im. Proc*., vol. 10, pp. 226-277, 2001.
[14] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, pp. 577–685, 1989.
[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] D.E. Maroulis, M.A. Savelonas, D.K. Iakovidis, S.A. Karkanis, N. Dimitropoulos, "Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images, " *IEEE Trans. Inf. Tech. Biomed.*, vol. 11, no. 5, pp. 537-543, 2007.
[17] P. Tsakanikas, E.S. Manolakos, "Active contours based segmentation of 2DGE proteomics images," in *Proc. European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, 2008.
[18] A.T. Rosengren, J.M. Salmi, T. Aitokallio et al., "Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels," *Proteomics*, vol.3, pp. 1936-1946, 2003.