

A Two-Stage Active Contour-based Scheme for Spot Detection in Proteomics Images

Eleftheria Mylona, Michalis Savelonas, *Member, IEEE* and Dimitris Maroulis, *Member, IEEE*

Abstract— This work introduces a two-stage active contour-based scheme for the detection of protein spots in two-dimensional gel electrophoresis images. The proposed scheme is formulated as an expansion of the Chan-Vese model and is capable of distinguishing overlapping spots. Moreover, it remains unaffected in the presence of noise and the inhomogeneous background that characterize these images. The experimental results demonstrate that the proposed scheme obtains more plausible spot boundaries than PDQuest (BioRad)¹.

I. INTRODUCTION

THE emerging field of proteomics has developed with the publication of the genome sequences of organisms. It specifically deals with the analysis of proteins expressed by one type of genome or by one kind of cell or organ. The primary goal is to assign molecular functions to the predicted gene products, so as to reveal the complex biochemical and physiological phenomena of a living cell. Nevertheless, proteomic research is challenged technically in the need of separation, characterization and quantification of thousands of complex protein mixtures. Aiming to confront these challenges, researchers utilize a powerful technique called two-dimensional polyacrylamide gel electrophoresis (2-D PAGE), which is widely considered to be a workhorse in the field of proteomics [1-3], providing a means for protein identification and visualization.

2-D PAGE enables the separation of protein mixtures utilizing the differences in two independent properties: the isoelectric point in the first dimension by isoelectric focusing and the molecular weight in the second by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Typically, a 2-D PAGE image is an electrophoretogram or a phosphor image (Fig. 1), which contains spots representing proteins. Image intensity variations correspond to variations in spot density, however they can be affected by the gel creation process as well.

Spot detection has to be accurate and robust, so as to quantify every single protein and trace any changes in

expression between gels. Researchers utilize computer-assisted image analysis aiming at protein spot identification and computation of a matching between gel images. Inherent difficulties arising include the inhomogeneous background, the presence of weak and overlapping protein spots as well as the presence of noise.

Most computer-assisted image analysis schemes falsely identify multiple overlapping protein spots as a single spot. Schemes based on watersheds [4-5], stepwise thresholding [6] and morphology [7-8] have been introduced to cope with this issue. However, watersheds lead to over-segmentation [9] whereas stepwise thresholding and morphology fail in the presence of artifacts and noise [10].

Tsai et al. [11] proposed a watershed-based scheme determined by markers so as to tackle the problem of over-segmentation. However, they admit that their scheme fails to detect overlapping spots. This is also the case with the active contours utilized by Tsakanikas et al. [12], as well as with the scheme of linear image search and merge introduced by Śmiateński [13]. The latter works stress the necessity of an algorithm which would efficiently distinguish overlapping spots.

Among the various image analysis approaches, active contour models [14-15] have been gaining increasing interest as they are self-adapting, and lead to continuous, closed, or open curves, without requiring edge-linking operations. Moreover, level-set formulated active contours [16] are topologically adaptable. In addition, region-based active contours, such as [15], are capable of detecting objects defined by weak edges, as it is the case with protein spots. Finally, they can be relatively insensitive to noise by involving integral operators, which provide an inherent noise filtering mechanism. These attributes of active contours motivated us to introduce an active contour-based scheme for the detection of multiple spots on 2-D PAGE images.

This work addresses 2-D PAGE spot detection focusing on the issue of overlapping protein spots. The proposed scheme is an expansion of the Chan-Vese model [15] and comprises two successive stages of contour evolution, which do not require user intervention. To the best of our knowledge, this is the first active contour-based scheme which is capable of detecting overlapping protein spots with very smooth boundaries, despite the inhomogeneous background and the presence of noise.

The outline of this paper is as follows: Section 2 and 3 describe the Chan-Vese model and the proposed scheme respectively. Section 4 presents the results on real images in order to validate the detection performance obtained and

Manuscript received June 15, 2010. This work was partially supported by the European Social Fund and the Ministry of National Education of Greece through the HRACLEITOS II program and the National and Kapodestrian University of Athens, Special Account of Research Grants.

E. Mylona, M. Savelonas, D. Maroulis are with Realtime Systems & Image Analysis Group, Dept. of Informatics and Telecommunications, Univ. of Athens, Panepistimiopolis, Illisia, 15784, Athens, Greece (e-mail: {e.mylona, m.savelonas, d.maroulis}@di.uoa.gr; phone: +30-210-7275317; fax: +30-210-7275333).

¹ <http://www.bio-rad.com>

finally, Section 5 discusses the conclusions of this study.



Fig. 1. 2-D PAGE image containing white protein spots.

II. CHAN-VESE MODEL

The Chan-Vese model as described in [15] has the form of a minimization problem: Let Ω be a bounded open subset of R^2 and $\partial\Omega$ its boundary. We seek for the infimum of the energy functional $F(c_1^+, c_1^-, C)$:

$$F(c_1^+, c_1^-, C) = \mu \cdot \text{Length}(C) + \lambda_1^+ \int_{\text{inside}(C)} |u_1(x, y) - c_1^+|^2 dx dy + \lambda_1^- \int_{\text{outside}(C)} |u_1(x, y) - c_1^-|^2 dx dy \quad (1)$$

where $u_1 : \Omega \rightarrow R$ is the input image, $C(s) : [0, 1] \rightarrow R^2$ is a piecewise parameterized curve, c_1^+ and c_1^- represent the average value of u_1 inside and outside the curve and parameters $\mu > 0$ and $\lambda_1^+, \lambda_1^- > 0$ are weights for the regularizing term and the fitting terms, respectively.

In the level set method [16], $C \subset \Omega$ is represented by the zero level set of a function $\phi_1 : \Omega \rightarrow R$, such that:

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

The average foreground (inside the contour) and background (outside the contour) intensities c_1^+ and c_1^- are determined by:

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

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

where H is the Heaviside function. 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. Finally, ϕ_1 is determined by parameterizing the descent direction by an artificial time $t \geq 0$, and by solving the following equation:

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

where δ is the one-dimensional Dirac measure and $t \in (0, \infty), (x, y) \in \Omega$.

Equation (5) dictates that contour evolution is guided by the average intensities c_1^+ and c_1^- , inside and outside the contour, respectively. Since in a 2D-PAGE image the average protein spot intensity is higher than the one of the background, the Chan-Vese model is capable of detecting the external spot boundaries. Nonetheless, this is not sufficient in overlapping protein spot regions where further separation is required.

III. PROPOSED SCHEME

The proposed scheme acts in two stages: in stage one, the contour is initialized and evolves according to Eq. (5) until it converges to the actual boundaries of the protein spots. This stage separates spot regions from the background. However, different but overlapping spots remain merged. In stage two, which aims at the detection of overlapping protein spots, the contour is not re-initialized, however its evolution is restricted inside the spot regions obtained in stage one. Therefore, the heavy intensity variations attributed to the 2D-PAGE image background are excluded from the average foreground and background intensities of the second stage. These quantities, which in stage two are denoted as c_2^+ and c_2^- , correspond to slightly different intensity distributions of overlapping spots, which were grouped within in stage one.

In stage two c_2^+ and c_2^- are expressed by the following equations:

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

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

where ϕ_2 , is the zero level set function, u_2 is the sub-image

of u_1 for which $\phi_1 > 0$, upon convergence of stage one. The proposed scheme is algorithmically implemented in the following steps:

- 1) Initialize $\phi_1 = \phi_0$,
- 2) Calculate $c_1^+(\phi_1)$ and $c_1^-(\phi_1)$ by Eq. (3) and (4), respectively,
- 3) Solve Eq. (5) to obtain a new ϕ_1 ,
- 4) If ϕ_1 converges then stop else repeat steps 2)-3).
- 5) Mark (x,y) for which $\phi_1(x,y) > 0$,
- 6) Calculate $c_2^+(\phi_2)$ and $c_2^-(\phi_2)$ by Eq. (6) and (7), respectively.
- 7) Solve Eq. (5), for $\phi_2, c_2^+(\phi_2), c_2^-(\phi_2)$, to obtain a new ϕ_2 ,
- 8) If ϕ_2 converges then stop else repeat steps 6)-7).

IV. RESULTS

The proposed scheme has been experimentally evaluated on 11 real 2-D PAGE images provided by the Biomedical Research Foundation of the Academy of Athens. The 2-D PAGE images were obtained in a digital format 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 μ were set to 1, 1 and 0.01 respectively [17].

Figure 2 illustrates segmentation results obtained by the Chan-Vese model on the 2-D PAGE image of Fig. 1. It is evident that the Chan-Vese model is incapable of separating overlapping spots. Figure 3(a₁-d₁) illustrates detailed sub-images of regions a,b,c and d of Fig. 2, whereas Fig. 3(a₂-d₂) illustrates the results of the application of the proposed scheme on the same sub-images. The ground truth images, as provided by an expert biologist, are presented in Fig. 3(a₃-d₃). It is evident that the proposed scheme effectively identifies overlapping spots in most cases. In the case of Fig. 3a₃, the Chan-Vese model misses three faint protein spots (Fig. 3a₂). This affects the proposed scheme as well (Fig. 3a₁), since it employs the Chan-Vese model in the first stage.

The detection results were quantified by means of the detection accuracy (DA) [18], which is defined as:

$$DA = \frac{TP}{FN + TP} \quad (8)$$

where true positives (TP) and false negatives (FN) are defined as correctly detected and falsely detected spots. TP and FN are calculated based on the final ground truth images. Considering the detection results obtained on the utilized dataset of 11 2-D PAGE images, the proposed technique obtains a DA of $84.3 \pm 2.6\%$, whereas the Chan-Vese model obtains a DA of $76.5 \pm 3.6\%$. This improvement in detection accuracy is attributed to the fact that the proposed scheme, as opposed to the Chan-Vese model,

effectively identifies overlapping spots.

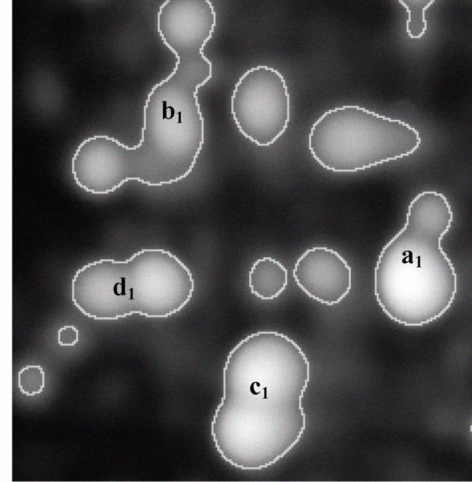


Fig. 2. Segmentation results obtained by the Chan-Vese model on the 2-D PAGE image of Fig.1.

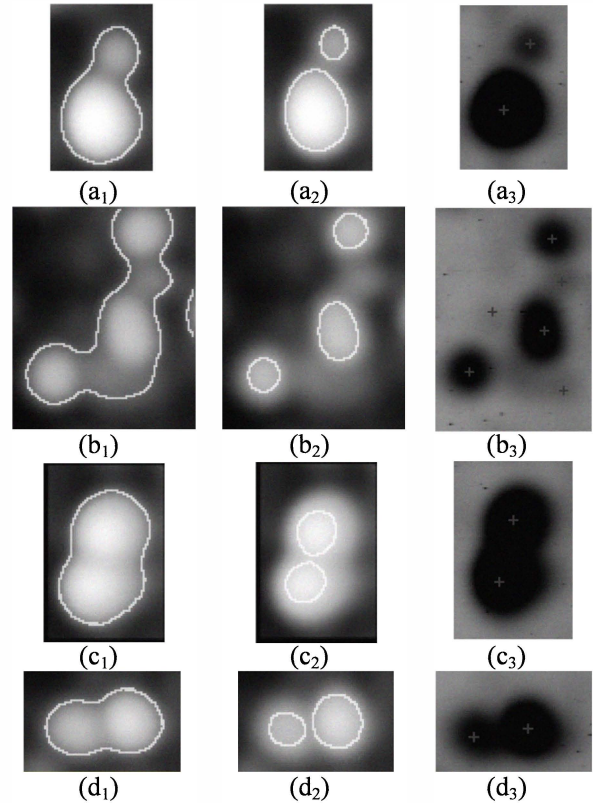


Fig. 3. Detection results obtained by the Chan-Vese model (a₁-d₁) and the proposed scheme (a₂-d₂), on the 2-D PAGE image of Fig. 2. The third column (a₃-d₃) illustrates the ground truth images, as provided by an expert biologist.

Aiming to evaluate the detection results obtained by the proposed scheme, a popular 2-D PAGE image analysis software package, named PDQuest (BioRad) [19], is utilized. Figure 4 depicts the detection results obtained by its application on the 2-D PAGE images of Fig. 3. It is evident that PDQuest fails to detect overlapping spots in the cases of Fig. 4(b-d). In addition, the elliptical boundaries generated by PDQuest are much less plausible than the spot boundaries

detected by the proposed scheme, which indicates that the latter is preferable for integration within a segmentation framework. Furthermore, unlike the proposed scheme, PDQuest, as well as other relevant approaches, such as the watershed-based scheme of Kim et al. [9], require user intervention.

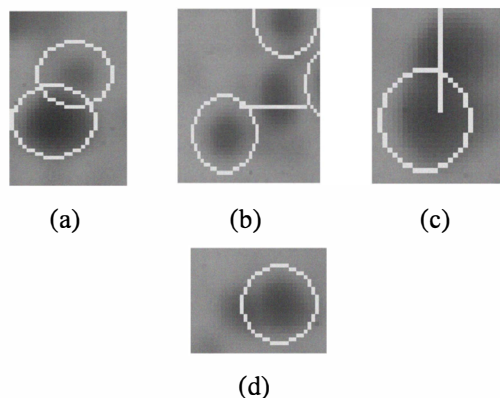


Fig. 4. Detection results obtained by PDQuest on the 2-D PAGE image of Fig. 2.

V. CONCLUSIONS

This work introduces a novel scheme for the detection of protein spots in 2D-PAGE images, which is based on successive active contours. For the test set evaluated and unlike recent relevant works [10-12], as well as a powerful, commercial software gel analysis package called PDQuest, the proposed scheme is capable of separating overlapping protein spots. Moreover, it does not require user intervention, which is the case with the latter [19] and the watershed-based scheme of Kim et al. [9]. This feature of the proposed scheme enhances the objectivity of the obtained spot detection results. Future work involves modifications to cope with the presence of faint protein spots, the extension of the proposed detection scheme for 2-D PAGE image segmentation, as well as evaluation on large repositories of real and synthetic images.

ACKNOWLEDGMENT

We would like to thank the Biomedical Research Foundation of the Academy of Athens for the provision of 2-D PAGE images.

REFERENCES

- [1] P.H.O' Farrell, "High resolution two-dimensional electrophoresis of proteins," *Journal of Biological Chemistry*, vol. 250, 1975, pp. 4007-4021.
- [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, 2002, pp. 1567-1596.

- [4] 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, 1999, pp. 755-765.
- [5] L. Vincent, P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 13, 1991, pp. 583-598.
- [6] 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, 2003, pp. 392-401.
- [7] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Im. Proc.*, vol. 2, 1993, pp. 176-201.
- [8] G. W. Horgan, C. A. Glasbey, "Uses of digital image analysis in electrophoresis," *Electrophoresis*, vol. 16, 1995, pp. 298-305.
- [9] 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*, vol. 2869, 2003, pp. 389-396.
- [10] K. Takahashi, Y. Watanabe, M. Nakazawa, A. Konagaya, "Fully automated spot recognition and matching algorithms for 2-D gel electrophoretogram of genomic DNA," *Proc. Genome Inf. Ser. Workshop*, vol. 9, 1998, pp. 161-172.
- [11] M.H. Tsai, H.H. Hsu, C.C. Cheng, "Watershed-based protein spot detection in 2DGE images," *Int. Computer Symposium*, Taiwan, 2006.
- [12] P. Tsakanikas, E.S. Manolakis, "Active contours based segmentation of 2DGE Proteomics Images," *Proc. European Signal Processing Conference (EUSIPCO)*, 2008.
- [13] J. Śmiateński, "The Scheme for computer analysis of results obtained in two-dimensional electrophoresis of serum protein complexes with congo red," *Bioinformatics*, vol. 2, 2006, pp. 27-33, 2006.
- [14] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active Contour Models", *Int. Journal Comp. Vis.*, vol. 1, 1988, pp. 321-331.
- [15] T.F. Chan, L.A. Vese, "Active contour without edges," *IEEE Trans. Im. Proc.*, vol. 10, 2001, pp. 226-277.
- [16] S. Osher, J.A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulation," *Journal Computer Physics*, vol. 79, 1988, pp. 12-49.
- [17] 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, 2007, pp. 537-543.
- [18] C. Guda, E. Fahy, S. Subramaniam, "MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins," *Bioinformatics*, vol. 20, 2004, pp. 1785-1794.
- [19] 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, 2003, pp. 1936-1946.