Segmentation of Two-Dimensional Gel Electrophoresis Images containing Overlapping Spots

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

Abstract—**This work addresses the segmentation of two-dimensional polyacrylamide gel electrophoresis images containing overlapping protein spots. A novel segmentation approach is proposed, which is capable of detecting spot boundaries within the region of overlap. The proposed approach is based on the observation that the spot boundaries in the overlap region are associated with local intensity minima. The experimental evaluation of the proposed approach demonstrates that it is capable of identifying multiple overlapping spots, as opposed to state of the art segmentation approaches. Moreover, it results in more accurate spot delineations, when compared to Progenesis SameSpots.**

Index Terms—*Segmentation, 2D-PAGE Images, Overlapping Spots*

I. Introduction

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is widely considered as a high-throughput technique in the field of proteomics [1]. It allows for the separation of thousands of complex protein mixtures expressed in a cell, tissue or biological system. The proteins are represented in the gel as spots of various sizes, according to their isoelectric points and their molecular masses. Detection and quantification of such spots may reveal alterations in protein expression within a given biological system. However, their excessive amount, poses the need for utilization of computer assisted image analysis.

2-D PAGE image analysis consists of different operations, including segmentation and protein expression quantification. Segmentation is crucial since its accuracy affects subsequent image analysis operations. However, it is complicated by the presence of noise, the inhomogeneous background and the highly overlapping protein spots.

State-of-the-art 2-D PAGE image segmentation approaches involve stepwise thresholding [2], morphology-based techniques [3], edge detection [4]-[5] and watersheds [6]. Stepwise threshold, morphology-based and edge detection approaches fail to accurately segment protein spots due to the presence of noise and artifacts, whereas watersheds result in over-segmentation and require additional filtering and inner markers [7]-[9]. Level set-based active contour models [10]-[12], which have been widely used in image segmentation, tend to cope with the problems mentioned above. An active contour-based segmentation approach has recently been proposed for the segmentation of 2-D PAGE images [13]. The approach is evaluated with PDQuest, a known commercial 2-D PAGE image analysis software package. The experimental results provided, reveal the efficiency and preciseness of this segmentation approach as far as the identification of non-overlapping protein spots in a gel image is concerned. However, it totally fails to detect the actual boundaries of overlapping protein spots, which is an open issue in 2-D PAGE image segmentation.

In this work, a novel 2-D PAGE active contour-based segmentation approach is proposed, aiming to detect the actual boundaries of protein spots in the presence of overlapping regions. The proposed approach is based on the observation that the spot boundaries in the overlap region are associated with local intensity minima. This association is considered within a novel active-contour based framework so as to allow the separation of the overlapping spots. To the best of our knowledge, a 2-D PAGE active contour-based segmentation approach coping with the presence of overlapping spots has not yet been proposed.

The rest of this paper is organized as follows: Section II describes the Chan-Vese model, Section III describes the proposed segmentation approach in terms of the overlapping protein spots, Section IV presents the obtained results, whereas the conclusions of this study are summarized in Section V.

II. The Chan-Vese Model

The Chan-Vese model as posed in [12] 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 $F(c^+,c^-,C)$:

$$\begin{aligned} F(c^+,c^-,C) = \mu \cdot Length(C) \\ + \lambda_1 \int_{inside(C)} |u_0(x,y)-c^+|^2 \, dxdy \\ + \lambda_2 \int_{outside(C)} |u_0(x,y)-c^-|^2 \, dxdy \end{aligned} \quad (1)$$

where $u_0 : \Omega \to R$ is the input image, $C(s):[0,1] \to R^2$ is a piecewise parameterized curve, c^+ and c^- represent the average value of u_0 inside and outside the curve and parameters $\mu > 0$ and $\lambda_1, \lambda_2 > 0$ are weights for the regularizing term and the fitting terms, respectively.

Manuscript received July 10, 2009. This work was partially funded by National and Kapodestrian University of Athens, Special Account of Research Grants.

M. Savelonas, D. Maroulis and E. Mylona are with Real-Time Systems & Image Analysis Group, Dept. of Informatics and Telecommunications, Univ. of Athens, Panepistimiopolis, Illisia, 15784 Athens, Greece (phone: +30-210-7275307; fax: +30-210-7275333; e-mail: {m.savelonas, d.maroulis, emylona}@ di.uoa.gr).

The energy functional $F(c^+,c^-,C)$ is derived from the general Mumford-Shah functional $F_{MS}(u,C)$ [14]:

$$\begin{aligned} F_{MS}(u,C) &= \mu \cdot Length(C) \\ &+ \lambda \int_{\Omega} |u_0(x,y) - u(x,y)|^2 \, dxdy \\ &+ \int_{\Omega/C} |\nabla u(x,y)|^2 \, dxdy \end{aligned} \tag{2}$$

Following the level set approach [8], $C \subset \Omega$ is represented as the zero level set of a function $\phi : \Omega \to R$, such that:

$$\begin{aligned} &C = \{(x,y) \in \Omega : \phi(x,y) = 0\}, \\ &inside(C) = \{(x,y) \in \Omega : \phi(x,y) > 0\}, \\ &outside(C) = \{(x,y) \in \Omega : \phi(x,y) < 0\} \end{aligned} \tag{3}$$

Using the Heaviside function H and the one-dimensional Dirac measure δ_0 denoted as:

$$H(\phi) = \begin{cases} 1, \phi \geq 0 \\ 0, \phi < 0 \end{cases}, \quad \delta_0(\phi) = \frac{d}{d\phi} H(\phi) \tag{4}$$

the average foreground (inside the contour) and background (outside the contour) intensities c_1 and c_2 are determined by:

$$c_1(\phi) = \frac{\int_{\Omega} u_0(x,y) H(\phi(x,y)) dxdy}{\int_{\Omega} H(\phi(x,y)) dxdy} \tag{5}$$

$$c_2(\phi) = \frac{\int_{\Omega} u_0(x,y)(1 - H(\phi(x,y))) dxdy}{\int_{\Omega} (1 - H(\phi(x,y))) dxdy} \tag{6}$$

where H is the Heaviside function. By keeping c_1 and c_2 fixed, and minimizing F with respect to ϕ, the associated Euler-Langrange equation for ϕ is deduced. Finally, ϕ is determined by parameterizing the descent direction by an artificial time $t \geq 0$, and by solving the following equation:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)[\mu \cdot div(\frac{\nabla \phi}{|\nabla \phi|}) - \lambda^+ (u_0 - c_1)^2 + \lambda^- (u_0 - c_2)^2] = 0 \tag{7}$$

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

As it is evident in Eq. (7), the evolution of the Chan-Vese active contour model is guided by the average intensities c_1 and c_2, inside and outside the contour, respectively. This region-based active contour formulation is appropriate for the detection of the external spot boundaries, since the average intensity of protein spots is higher than the one of the background. However, it is not appropriate in cases of spot overlap, since in such cases, boundaries are generally not defined by differences in average intensity between the overlapping spots.

III. Proposed Segmentation Approach

The proposed segmentation approach is based on the observation that in 2-D PAGE images, spot boundaries within regions of overlap are associated with local intensity minima, with respect to a certain direction. This is evident in the example of Fig. 1, where intensity I is illustrated as a function of the position d, over a line segment, which crosses two overlapping spots in a 2-D PAGE image.

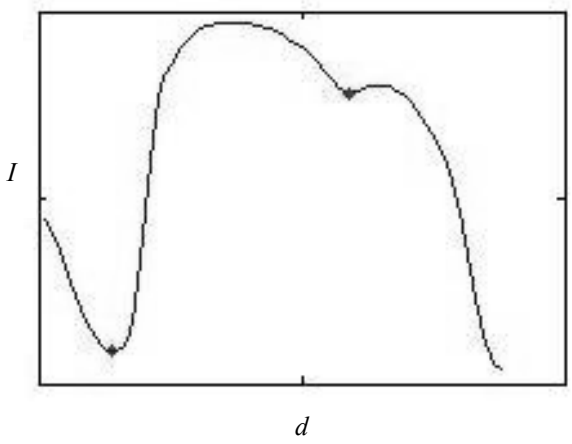

Fig. 1. An example of intensity I as a function of position d, over a line segment, which crosses two overlapping spots in a 2-D PAGE image. Red dots illustrate local minima.

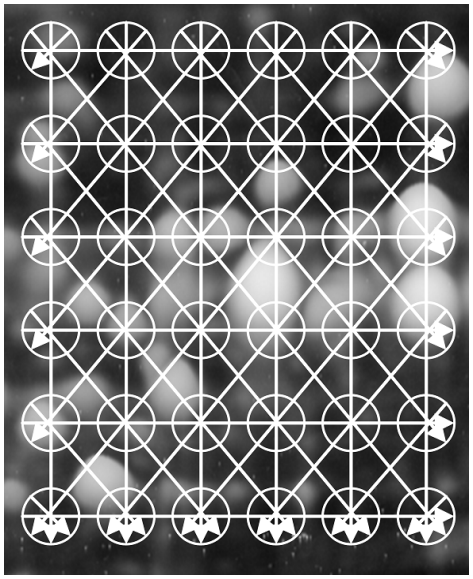

Fig. 2. Multiple directions utilized for local minima detection

The multiple directions considered appropriate for local minima detection in terms of globalization of the problem are illustrated in Fig. 2.

The spot boundary detection algorithm consists of the following steps:

FOR $\theta = 0$ to π step s

• SCAN the 2-D PAGE image with parallel straight line segments of direction θ

• LOCATE the local minima of intensity, for each parallel straight line segment,

• SELECT local minima which:
- are minima over a sub-segment of width exceeding a minimum w
- are associated with points of intensity $I > T$, where T is a threshold value

END FOR

• MARK as white each point of the 2-D PAGE image associated with selected local minima

• APPLY the Chan-Vese active contour model and mark as white the generated contour points

• SUPERSEDE all points marked as white in the previous two steps, so as to generate the end segmentation result.

It should be noted that local minima of intensity are selected if they are minima of a sub-segment with width exceeding a minimum w. Thus, local minima expressed as short-width intensity fluctuations, which can be attributed to noise, are avoided. In addition, a local minimum is selected if it is associated with intensity I higher than a threshold value T, so as to exclude local minima associated with background clutter.

IV. RESULTS

A number of experiments were performed aiming to the assessment of the proposed approach for the segmentation of 2-D PAGE images containing overlapping spots. The 2-D PAGE images were obtained in a digital format at 8-bit gray-level depth. Figure 3 illustrates an example of a 2-D PAGE image used in the experiments.

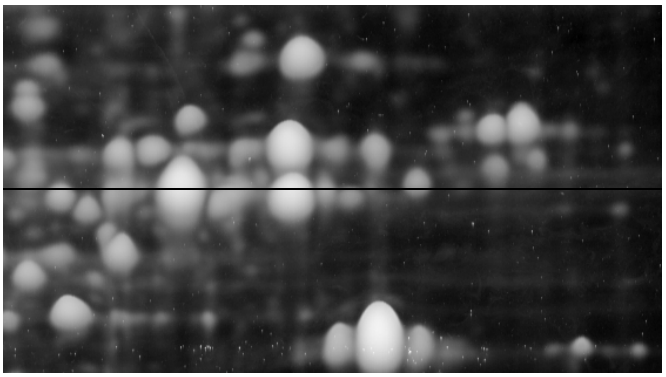

Fig. 3. 2-D PAGE image

The utilized algorithms were implemented in Microsoft Visual C++ and executed on a 3.2 GHz Intel Pentium workstation. A preliminary set of experiments was performed, considering that the variation of model parameters affects the segmentation accuracy, In accordance with the results of these preliminary experiments, parameters w and T were set to 3 and 100 respectively, whereas active contour parameters λ^+, λ^- and μ were set to 1, 2 and 0.01, respectively.

Figure 4 illustrates examples of segmentation results obtained by the application of the standalone Chan-Vese active model and by the proposed segmentation approach on 2-D PAGE images containing overlapping protein spots. Figures 4(a-c) depict three sub-images of such a 2-D PAGE image. Figures 4(d-f) depict the segmentation results of the standalone Chan-Vese active contour model. Figures 4(g-i) depict the segmentation results of the proposed approach.

It is evident that the standalone Chan-Vese active contour is not capable of detecting spot boundaries within regions of overlap. As a result, multiple overlapping spots are identified as one. On the contrary, the proposed segmentation approach is capable of identifying multiple overlapping spots.

The segmentation results of the proposed approach have been qualitatively compared with the results obtained from a renowned 2-D PAGE image analysis software package called Progenesis SameSpots [15]. The results of the application of this software package on the overlapping spots of Fig. 4 are illustrated in Fig. 5. It should be noted that for this application, the 2-D PAGE images had to be inverted. It can be observed that Progenesis SameSpots merely detects overlapping spots without accurately identifying their boundaries, as it is the case with the proposed approach. In addition, Progenesis SameSpots often results in oversegmentation, as in the example of Fig. 5(b), where it has detected four protein spots, although there are only three actual spots.

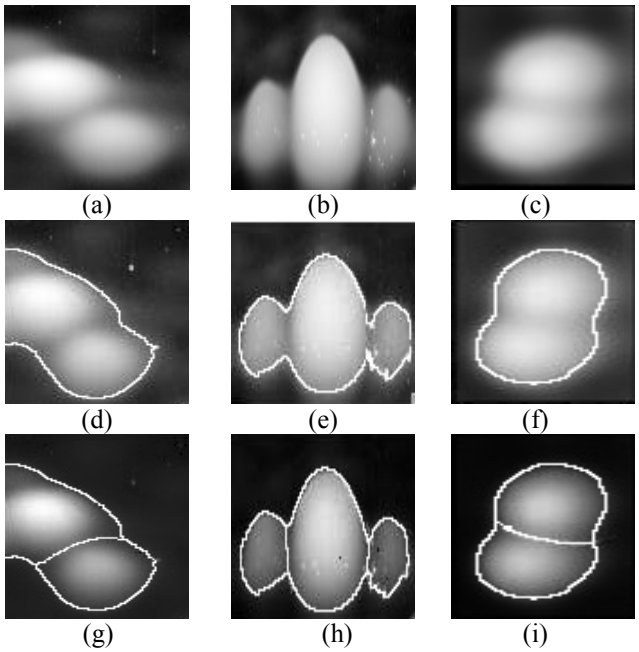

Fig. 4. (a-c) original sub-images of 2-D PAGE images containing overlapping protein spots, (d-f) segmentation results obtained by the standalone Chan-Vese active contour model, and (g-i) segmentation results obtained by the proposed approach.

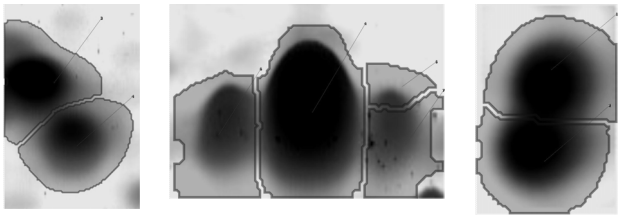

Fig. 5. Example segmentation results obtained from the application of Progenesis SameSpots software package on the overlapping spots of Fig. 4.

V. CONCLUSIONS

In this work, a novel 2-D PAGE image segmentation approach is proposed, aiming to detect the actual boundaries of protein spots in the presence of overlapping spot regions. The proposed approach is based on the observation that the spot boundaries in the overlap region demonstrate local intensity minima. These minima are detected with respect to multiple directions and the extracted boundaries are superseded over the external spot boundaries, as detected by the Chan-Vese active contour.

The experimental results lead to the conclusion that the proposed segmentation approach is capable of identifying overlapping spots, providing an expansion to the standalone Chan-Vese active contour model. Moreover, it results in more accurate spot delineations, when compared to Progenesis SameSpots.

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

[2] 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.

[3] L.Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms", *IEEE Trans. Im. Proc.,* vol. 2, no. 2, 1993, pp. 176-201.

[4] P.F. Lemkin and L.E. Lipkin, "2D electrophoresis gel data base analysis: aspects of data structures and search strategies in GELLAB", *Electrophoresis,* vol. 4, no. 1, 1983, pp. 71-81.

[5] R.D. Appel, J.R. Vargas, P.M. Palagi, D. Walther and D.F. Hochstrasser, "Melanie II-a third generation software package for analysis of two-dimensional electrophoresis images: II.Algorithms", *Electrophoresis,* vol. 8, no. 15, 1997, pp. 2735-2748.

[6] 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.

[7] K. Conradsen, J. Pedersen, "Analysis of two-dimensional electrophoretic gels", *Biometrics*, vol. 48, 1992, pp. 1273-1278.

[8] 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, 1998, pp. 161-172.

[9] Y. Kim, J. Kim, Y. Won, Y. In, "Segmentation of protein spots in 2-D gel electrophoresis images with watershed using hierarchical threshold", LNCS, vol. 2869, 2003, pp. 389-396.

[10] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: active contour models", *Int. J. Comp. Vis.,* vol. 1, 1988, pp. 321-331.

[11] S. Other, J.A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulation", *J. Comp. Phys.,*vol. 79, 1988, pp. 12-49.

[12] T.F. Chan, L.A. Vese, "Active contour without edges", *IEEE Trans., Im., Proc.,* vol. 10, no. 2, 2001, pp.226-277.

[13] P. Tsakanikas, E.S. Manolakos, "Active contours based segmentation of 2DGE proteomics images", *Proc. EUSIPCO,* 2008.

[14] D.Mumford, J.Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Com. Pure. Appl. Math*, vol. 42, 1989, pp. 577-685.

[15] 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, 2003, pp. 1936-1946.