An Automatically Initialized Level-Set Approach for the Segmentation of Proteomics Images

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

***Abstract*—This work introduces an automatically initialized level-set approach for the segmentation of proteomics images. Level-set methods are natural candidates for this segmentation task, since their topological adaptability is a prerequisite to identify the boundaries of thousands of protein spots present. However, they require manual initialization, whereas their convergence is highly sensitive to the position of the initial contour. The proposed initialization process is based on the formation of a level-set surface of multiple cones, which are centered at regional intensity maxima positions, associated with protein spots. This initialization process, as is the case with the subsequent contour evolution, is aided by means of histogram equalization and morphological processing. The experimental results indicate that the proposed level-set approach facilitates quicker convergence than the one obtained by the straightforward application of the Chan-Vese model. In addition, the identified spot boundaries are more plausible than the ones obtained by the application of state-of-the-art proteomics image analysis software.**

I. Introduction

THE field of proteomics has progressed enormously in the past few years. The complexity of the proteome intrigues researchers to study all the post-transcriptional control and post-translational modification maps which decode the fundamental biological mechanisms. The goal is the quantitative assessment of protein expression.

One of the major techniques in proteomics is two-dimensional gel electrophoresis (2D-GE) [1],[2]. 2D-GE resolves proteins by size and charge by means of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and isoelectric focusing, respectively.

Segmentation of proteomics images is impeded by the presence of noise, artifacts, inhomogeneous background, saturated and faint spots, as well as of highly overlapping protein spots. 2D-GE image analysis software packages such as PDQuest 8.0.1 (Bio-Rad) [3] and Delta 2D (Decodon) [4] aim to confront the aforementioned issues. However, the segmentation results obtained lack objectivity [5] and require manual editing which is a cumbersome and tedious process, even for experienced biologists. Fig. 1 depicts a part of a real proteomics image which contains faint and overlapping spots over an inhomogeneous background.

Fig. 1. Part of a real proteomics image containing faint and overlapping spots over an inhomogeneous background.

Stepwise thresholding methods are capable of resolving intensity variations into distinct connected areas associated with protein spots [6]. Nevertheless, they are highly sensitive to noise and inhomogeneous background [7]. Edge detection methods aim to identify discontinuities in image intensity associated with protein spot boundaries. However, they are susceptible to spurious segmentation results in the presence of noise and breaks in spot boundaries [8]. Watersheds cope well with noise, artifacts and inhomogeneous background but often result in over-segmentation, posing the need for additional post-processing [9], [10].

The level-set paradigm [11] provides a powerful alternative to the aforementioned segmentation methods. Level-set methods are topologically adaptable, whereas their inherent continuity and smoothness can compensate for noise, gaps and other irregularities in target boundaries. These attractive properties are vital for the segmentation of proteomics images since the latter contain thousands of protein spots with faint boundaries. However, level-sets suffer in the sense that they are sensitive to the initial level-set surface, which is often manually determined [12].

In this work, an automatically initialized level-set approach is introduced for the segmentation of protein spots in proteomics images. The level set surface is initialized as multiple cones centered at regional intensity maxima positions, associated with protein spots. Histogram equalization and morphological processing are incorporated to support this initialization process, as well as the subsequent contour evolution. The proposed level-set approach facilitates quicker convergence than the one obtained by the straightforward application of the Chan-

Manuscript received July 5, 2011. This work has been co-financed by the European Union (European Social Fund – ESF) and Greek National Funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) – Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

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

978-1-4577-0554-0/11/$26.00 ©2011 IEEE

Vese model. Most importantly, it outperforms state-of-the-art proteomics image analysis software in terms of segmentation quality.

The remainder of this paper is organized in four sections. Section II briefly presents the theoretical background whereas Section III describes the main components of the proposed segmentation approach. Section IV demonstrates the experimental results on real and synthetic proteomics images whereas Section V discusses the conclusions and future perspectives of this work.

II. Theoretical Background

A. Chan –Vese Model

The Chan-Vese model [13] is implemented with an iterative algorithm based on a second order system defined in the spatial domain. The mathematical formulation of the Chan-Vese model adopts the reduced case of the Mumford-Shah problem [14], resulting in the following evolution equation:

$$\frac{\partial \varphi_1}{\partial t} = \delta\ (\varphi_1)[\mu \cdot div(\frac{\nabla \varphi_1}{|\nabla \varphi_1|}) - \lambda_1^+ (u_1 - c_1^+)^2 + \lambda_1^- (u_1 - c_1^-)^2] \quad (1)$$

where u_1 is a two dimensional image, ϕ_1 is a Lipschitz function c_1^+ and c_1^- are the respective average intensities, δ is the Dirac delta function, t is the artificial time parameterizing the descent direction and $\mu, \lambda_1^+, \lambda_1^- > 0$ are weighting parameters.

Fig. 2 illustrates segmentation results obtained from the straightforward application of the Chan-Vese model where the contour is initialized: a) inside a protein spot, b) to embed a single protein spot and c) to embed multiple protein spots. It is evident that the convergence state is subject to initial contour positioning. Furthermore, the Chan-Vese model fails to split so as to identify the boundaries of individual protein spots within "multiplets" (i.e. complex regions of overlapping protein spots).

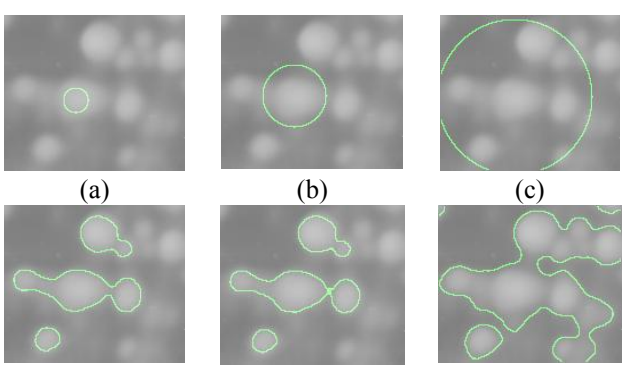

Fig. 2. The initial contour: a) is positioned inside a protein spot, b) embeds a protein spot and c) embeds the majority of protein spots, (d)-(f) segmentation results emerged from the application of Chan-Vese model on Fig. 1 according to (a)-(c) respectively after 2000 iterations.

B. Mathematical Morphology

Let E be an Euclidean space, A a grayscale image in E and B a structuring element (SE) [15]. The SE plays the exact same role with the convolution kernel in the linear filter theory. The shape and size of the SE are selected according to the shape and size of the objects of interest, respectively [16]. In cases of proteomics images, the SE is selected disk-shaped with a radius smaller than the one of the smallest protein spot, aiming to detect regional intensity maxima associated with protein spots.

The dilation image operator, denoted as $A \oplus B$, is defined, in the general case, by:

$$D(A,B) = (A \oplus B)(x) = \sup_{y \in E} [A(y) + B(x-y)] \quad (2)$$

where sup denotes the supremum and x, y are elements of E.

III. Proposed Segmentation Approach

A. Image Enhancement and Morphological Reconstruction

The original proteomics image is reconstructed by means of the processes described in this sub-section, so as to cope with the presence of noise, overlapping protein spots, faint spots and intensity inhomogeneity.

The original image is pre-processed with a median filter so as to cope with the presence of noise. In addition, it has been demonstrated in our previous work [17] that local intensity minima are localized on regions of overlap. The latter provide essential information regarding the boundaries of individual protein spots within multiplets, which are erroneously identified as one by the Chan-Vese model. In order to detect such minima, the filtered image is scanned with parallel straight-line segments of variable lengths and multiple directions. Those minima which appear as short-width intensity fluctuations correspond to background noise and are filtered out. The locations of the surviving minima, which indicate regions of overlap, are marked.

As a next step, a histogram equalization variant called contrast-limited adaptive histogram equalization (CLAHE) [18] is applied, aiming to highlight faint spots. CLAHE enhances the proteomics image contrast in non-overlapping image regions called tiles. A constraint called clip limit is imposed to the height of the local histogram in order to avoid unwanted amplifications of noise. The neighboring transformed tiles are merged using bilinear interpolation to reduce artificially induced boundaries. In the enhanced image resulting by the application of CLAHE, the intensity differences between faint spots and background are stressed.

Finally, the enhanced image resulting by the application of CLAHE is binarized according to a threshold value T. However, due to the intensity inhomogeneity, the faint spot regions in the resulting binary image contain "holes". Accordingly, the binary image is morphologically

reconstructed by means of the flood-fill operation [19], which alternates the connected background pixels of each "hole" to foreground.

The binary image resulting from the application of CLAHE and morphological reconstruction represents protein spots, including faint ones, as well as the boundaries of spot overlap in regions occupied by multiplets. The latter is achieved by assigning to the background the locations of local intensity minima which were marked in the local intensity minima detection stage (Fig. 3). The valuable information of the separated overlapping spots as well as the presence of faint spots, provided by this binary image, is incorporated in contour evolution, as described in III.B.

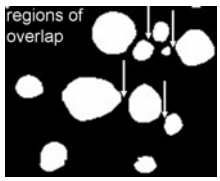

Fig. 3. The binary image resulting from the application of CLAHE and morphological reconstruction. The regions of overlap, indicated by the previously marked local intensity minima, are assigned to background.

B. *Level-set Initialization*

The initialization of the level-set surface is based on regional intensity maxima associated with protein spots. The detection of such maxima is achieved by means of dilation of the proteomics image with a disk-shaped SE, according to Eq. (2). Spurious maxima associated with background noise peaks are avoided. The initial level-set surface is formed as a surface of multiple cones, which are centered at the locations of the regional intensity maxima detected. The position of each such regional intensity maximum is used as the center of an individual cone. The initial level-set surface can be visualized as a spot-targeted level-set function so that the associated zero-level regions approximate the actual protein spots.

SE radius r is set according to the results of preliminary experimentation, which indicate that a certain radius value minimizes the detection of false negative protein spots whereas it allows the detection of regional intensity maxima associated with small spots, even in cases where they overlap with larger spots in complex regions. This initialization process is particularly important within the context of the proposed segmentation approach, since it provides the capability of unsupervised segmentation.

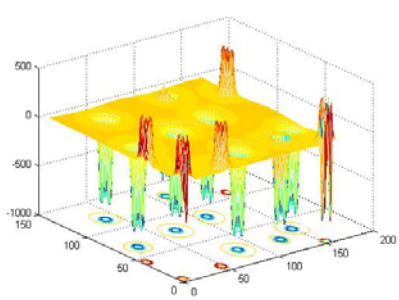

Fig. 4. 3D representation of a level-set surface of multiple cones.

Fig. 4 illustrates a three-dimensional representation of the level-set surface obtained by the application of the proposed initialization process on the proteomics image of Fig. 1, whereas Fig. 5 depicts the detection results of regional intensity maxima on the same image.

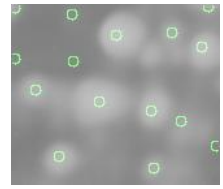

Fig. 5. Detection results of regional intensity maxima on Fig. 1.

Contour evolution is restricted in regions centered at the cones centers of the level-set surface and converges according to the following equation:

$$\frac{\partial\varphi}{\partial t}=\delta(\varphi)[\mu\cdot div(\frac{\nabla\varphi}{|\nabla\varphi|})-\lambda_1^+(u_1-c_1^+)^2+\lambda_1^-(u_1-c_1^-)^2-\lambda_2^+(u_2-c_2^+)+\lambda_2^-(u_2-c_2^-)]=0 \quad (3)$$

where u_1, u_2 are the proteomics image and the binary image obtained by the application of CLAHE and morphological reconstruction, respectively. In addition, c_1^+, c_2^+ and c_1^-, c_2^- are the average foreground and background intensities of u_1 and u_2, whereas μ is the weight for the regularizing term and λ_1^+, λ_2^+ and λ_1^-, λ_2^- are the weights for the fitting terms of u_1 and u_2, respectively. Equation (3) extends Eq. (1) since it incorporates information derived by the binary image u_2. The latter information is essential to identify the presence of faint spots as well as the regions of spot overlap in multiplets.

IV. Results

The proposed segmentation approach has been experimentally evaluated on a dataset of 12 real proteomics images provided by the Biomedical Research Foundation of the Academy of Athens, so as to qualitatively compare the obtained segmentation results with the ones obtained by PDQuest 8.0.1 and Delta2D software packages. The software packages were applied on inverted versions of the proteomics images, whereas parameter settings and calibrations involved were performed by expert biologists, following their experience. Parameter values were set after preliminary experimentation on a single proteomics image. Threshold value T, SE radius r and the weight μ of the regularizing term were set to 160, 4 and $0.006 \cdot 255^2$, respectively, whereas the weights λ_1^+, λ_2^+ and λ_1^-, λ_2^- of the fitting terms were all set to 1.

Fig. 6 illustrates segmentation results obtained by the application of: (a)-(b) the proposed segmentation approach, (c)-(d) PDQuest 8.0.1 and (e)-(f) Delta2D on the proteomics image of Fig. 1. It is evident that the proposed segmentation

approach results in more plausible spot boundaries than the ones identified by PDQuest 8.0.1 and Delta2D software packages. PDQuest 8.0.1 results in elliptical boundaries which do not correspond to the irregular shape of the actual spot boundaries, whereas both software packages tend to include background regions. In the case of Delta2D, the segmentation results obtained suffer from over-segmentation and are subject to laborious, error-prone and time-consuming correction process by the expert biologists.

It should be pointed out that the level-set function of the proposed initialization process converges after approximately 5 iterations. On the contrary, the Chan-Vese model converges after approximately 2000 iterations and still fails to accurately segment proteomics images containing considerable amounts of overlapping or faint protein spots (Fig. 2).

V. Conclusions

In this work, an automatically initialized level-set approach is introduced for the segmentation of proteomics images. The initial level-set surface is formed as a surface of multiple cones, which are centered at the locations of the regional intensity maxima. The initialization process is particularly important since it provides the capability of unsupervised segmentation, as opposed to the Chan-Vese model which requires manual initialization. Histogram equalization and morphological processing are incorporated to support this initialization process, as well as the subsequent contour evolution.

The experimental results demonstrate that the proposed segmentation approach converges faster than the Chan-Vese model, whereas it generates more plausible spot boundaries than PDQuest 8.0.1 and Delta2D software packages. Future perspectives of this work include integration of the proposed segmentation approach within a proteomics image analysis system.

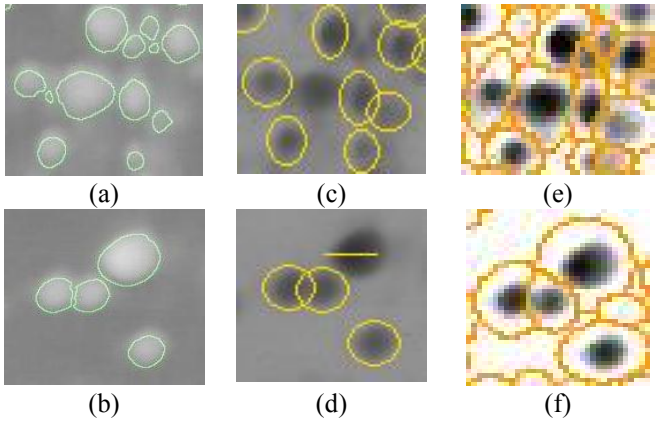

Fig. 6. Segmentation results obtained by the application of: (a)-(b) the proposed segmentation approach, (c)-(d) PDQuest 8.0.1 and (e)-(f) Delta 2D on the proteomics image of Fig. 1.

Acknowledgment

We would like to thank the Biomedical Research Foundation of the Academy of Athens for the provision of real proteomics images as well as M. Makridakis and M. Aivaliotis for the provision of segmentation results obtained by the application of PDQuest 8.0.1 and Delta 2D image analysis software packages on real proteomics images respectively. We would particularly like to thank expert biologists Dr. S. Kossida and Dr. A. Vlahou for their constructive recommendations in the experimental evaluation of the results.

References

[1] M. Berth, F. Moser, M. Kolbe, J. Bernhardt, "The state of the art in the analysis of two-dimensional gel electrophoresis images,'' *Applied Microbiology and Biotechnology*, no. 76, pp. 1223-1243, 2007.

[2] A.W. Dowsey, M.J. Dunn and G.Z. Yang, "The role of bioinformatics in two-dimensional gel electrophoresis," *Proteomics*, vol. 3, no. 8, pp. 1567-1596, 2003.

[3] http://www.bio-rad.com

[4] http://www.decodon.com

[5] B.N. Clark and H.B. Gutstein, "The myth of automated, high-throughput two-dimensional gel analysis," *Proteomics,* vol. 8, no. 6, pp. 1197-1203, 2008.

[6] P. Cutler, G. Heald, I.R. White and J. Ruan, "A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection," *Proteomics*, vol. 3, no. 4, pp. 392-401, 2003.

[7] K. Takahashi, Y. Watanabe, M. Nakazawa and 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.

[8] J.M. Walker, "The proteomics protocols handbook," Humana Press, New Jersey, 2005.

[9] 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, no. (4-5), pp. 755-765, 1999.

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

[11] S. Osher and J.A. Sethian, "Fronts propagating with curvature-dependent speed – algorithms based on Hamilton-Jacobi formulations," *J. Comp. Phys*., vol. 79, no. 1, pp. 12-49, 1988.

[12] S. H. Lee, J.K. Seo, "Level set-based bimodal segmentation with stationary global minimum," *IEEE Trans. Im. Proc.,* vol. 15, no. 9, pp. 2843-2852, 2006.

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

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

[15] J. Serra, "Image analysis and mathematical morphology," 2nd ed. Academic, New York, 1982.

[16] E.R. Urbach, M.H.F. Wilkinson, "Efficient 2-D grayscale morphological transformations with arbitrary flat structuring elements," *IEEE Trans. Im. Proc*. vol. 17, pp. 1-8, 2008.

[17] M.A. Savelonas, D. Maroulis, E. Mylona, "Segmentation of two-dimensional gel electrophoresis images containing overlapping spots", *in Proc. IEEE Int. Conf. Inf. Tech. Appl. Biomed. (ITAB)*, Larnaca, Cyprus, 2009.

[18] S.M. Pizer, E.P. Amburn, J.D. Austin, "Adaptive histogram equalization and its variations," *Comp. Vis. Graph. Im. Proc*., vol. 39, pp. 355-368, 1987.

[19] P. Soille, "Morphological image analysis-principles and applications," Springer, Berlin, 1999.