2D-GE IMAGE SEGMENTATION BASED ON LEVEL-SETS

E.A. Mylona [a], *M.A. Savelonas* [a], *D. Maroulis* [a], *M. Aivaliotis* [b]

[a]Dept. of Informatics and Telecommunications, University of Athens, Greece
[b]Institute for Molecular Biology and Biotechnology, IMBB-FORTH, Crete, Greece
[a]{emylona,msavelonas,dmaroulis}@di.uoa.gr
[b]{maivaliot}@imbb.forth.gr

ABSTRACT

This work introduces a novel level-set based scheme for the segmentation of protein spots in 2D-GE images. The proposed scheme incorporates a protein spot detection stage based on the identification of local intensity minima in order to allow the separation of overlapping spots. Moreover, histogram adaptation and morphological operators are applied for the identification of faint spots. A segmentation stage follows, where a level-set function is initialized based on local intensity maxima and evolves according to energy terms derived in the detection stage. The experimental evaluation demonstrates that the proposed scheme generates more plausible spot boundaries than PDQuest 8.0.1 and Melanie 7 and outperforms both software packages in terms of segmentation performance.

Index Terms—2D-GE Images, Protein Spot Segmentation, Level-Sets.

1. INTRODUCTION

The growing interest in proteome analysis triggered the development of proteome visualization technologies. Two-dimensional gel electrophoresis (2D-GE) is the highest resolving procedure for protein display prior to post–separation analysis [1]-[2]. 2D-GE results in digital grayscale images containing a few hundred up to several thousands of protein spots. The amount of each migrated protein can be estimated by the cumulative intensity value of each pixel within the spot region.

Protein spot regions can be identified by means of segmentation of 2D-GE images. However, 2D-GE image segmentation is complicated because of the non-uniformity of the background intensity, the presence of faint and saturated spots as well as the complex regions containing overlapping spots.

State-of-the-art commercial image analysis software packages such as: Melanie (GeneBio/GE Healthcare) [3]-[4], and PDQuest (Bio-Rad) [5]-[6] are powerful tools in 2D-GE image analysis. Nevertheless, they are highly parametric and the results obtained require manual editing by biologists through a time-consuming, error-prone and subjective process [7]. These open issues motivate further research on robust and objective computer-assisted methods for 2D-GE image segmentation.

Stepwise thresholding [8], edge detection [9] and watershed methods [10] have been applied on 2D-GE image segmentation. However, stepwise thresholding and edge-detection fail in the presence of background non-uniformity and overlapping spot clustering [11] whereas watershed methods lead to oversegmentation [12].

Level-sets [13] have been widely used in image segmentation since they are topologically adaptable, facilitating contour merging or splitting. This topological adaptability of level-sets is of key importance in the case of 2D-GE image, which may count thousands of protein spots. Moreover, the inherent continuity and smoothness of level-sets can compensate for noise, gaps and other irregularities in target boundaries.

This work introduces a novel level-set based scheme for protein spot segmentation in 2D-GE images. The proposed scheme employs an initial protein spot detection stage based on the identification of local intensity minima, which facilitates the separation of overlapping spots. In addition, histogram adaptation and morphological operators are applied for the identification of faint spots. A segmentation stage follows, which involves initialization of a level-set function based on local intensity maxima and evolution of the level-set surface guided by energy terms derived in the detection stage.

The outline of this paper is organized as follows: Section 2 provides the theoretical background of the proposed segmentation scheme, which includes the Chan-Vese level-set model [14]. Section 3 introduces the proposed scheme whereas Section 4 demonstrates the experimental results. Finally, Section 5 summarizes the conclusions of this study.

2. CHAN-VESE MODEL

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: seeking for the infimum of the energy functional $F(c_1^+, c_1^-, C)$:

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

where $C : [0,1] \to R^2$ 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. In the level set formulation [13], the curve $C \subset \Omega$ is represented implicitly by the zero level-set of a Lipschitz function $\phi_1 : \Omega \to R$. By keeping c_1^+ and c_1^- fixed, and minimizing F with respect to ϕ_1, the associated Euler-Langrange 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:

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

where $t \in (0, \infty), (x,y) \in \Omega$ and δ is the one-dimensional Dirac function. The Chan-Vese level-set model is noise-robust and topologically adaptable; however it fails to segment 2D-GE images containing complex regions of overlapping spots as well as faint spots.

3. PROPOSED SCHEME

The proposed scheme is comprised of an initial protein spot detection stage, followed by a segmentation stage based on level-sets.

3.1. Protein Spot Detection

The proposed scheme aims to cope with a crucial issue in 2D-GE image analysis, which involves the presence of complex regions containing overlapping spots. It can be observed that the region of spot overlap is associated with local intensity minima with respect to a particular direction. Considering this observation, the image is initially scanned with parallel straight-line segments of variable lengths and multiple directions. Local intensity minima are selected according to the following constraints: intensity value a) exceeds a threshold T_1 and b) is a global minimum over a square sub-segment of width, which exceeds a minimum value w. These constraints are imposed so as to exclude local intensity minima associated with background clutter. Figure 1 illustrates: (a) a part of a synthetic 2D-GE image containing overlapping spots and (b) the detection results obtained by the local intensity minima process, with each minimum marked as black. It is evident that regions of spot overlap are correctly identified.

(a) (b)

Fig. 1: (a) Part of a synthetic 2D-GE image containing overlapping spots, (b) detection results obtained by the local intensity minima process.

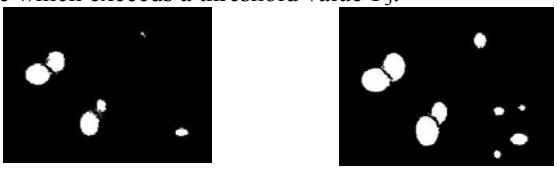
Another challenge, complicating 2D-GE image analysis, is posed by the presence of faint spots. The proposed scheme aims to cope with this issue by employing a histogram equalization variant called contrast-limited adaptive histogram equalization (CLAHE) [15]. Contrary to the original histogram equalization, CLAHE adaptively enhances the image contrast in $q \times q$ non-overlapping image regions called tiles. The resulting contrast is restricted by a parameter h called clip limit so as to avoid unwanted amplifications induced by noise, artifacts and streaks. The resulting neighboring tiles are merged by means of bilinear interpolation so as to reduce artificially induced boundaries.

As a next step, the contrast-enhanced image generated by CLAHE is binarized according to a threshold value T_2. However, the identified spot regions contain "holes", as a result of intensity inhomogeneity. Morphological flood-fill operation [16] is applied so as to eliminate "holes". Figure 2 illustrates results obtained by the morphological flood-fill operation on the image of Fig. 1(b): (a) without the application of CLAHE and (b) with the application of CLAHE. It is evident that the application of CLAHE is essential in order to identify faint spots.

3.2. 2D-GE Image Segmentation

The second stage of the proposed scheme involves initialization and evolution of a level-set based segmentation model. The level set function is initialized so that the zero-levels approximate protein spots. The initialization process is based on the observation that protein spots are associated with regional intensity maxima [16]. Accordingly, initialization starts with the detection of such maxima so as to construct a surface of multiple cones centered at protein spot positions. Regional intensity maxima are selected according to the following constraints: a) intensity value should be the maximum over an $m \times m$ adjacent region and b) every pixel over a $z \times z$ square

neighborhood of each one of them should have intensity value which exceeds a threshold value T_3.



(a) (b)

Fig. 2: Results obtained by the morphological flood-fill operation on Fig. 1(b): (a) without the application of CLAHE, (b) with the application of CLAHE.

The evolution of the level-set surface is restricted in separate $g \times g$ image regions which share the same center position with each level-set cone generated in the initialization process. This separate evolution is computationally more efficient than iterative calculations of the entire level-set surface. The level-set evolution proceeds 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 original 2D-GE image and the output image of the protein detection stage, 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 λ_1^+, λ_2^+ and λ_1^-, λ_2^- are the weights for the regularizing and fitting terms of u_1 and u_2, respectively. Equation (3) encompasses information derived by local intensity minima and morphological processing of the original 2D-GE image, facilitating correct detection of overlapping and faint spot boundaries.

4. RESULTS

The proposed scheme has been experimentally evaluated on a dataset of 20 synthetic 2D-GE images generated by the Real-Time Systems & Image Analysis Lab, so as to allow quantitative evaluation of the segmentation results. The simulation of the 2D-GE images addresses the key attributes of real 2D-GE images: 1) multiplet occurence, 2) spot saturation, 3) background inhomogeneity and 4) presence of streaks and noise [2]. Each synthetic image has size of approximately 1500x2000 pixels, gray level depth of 16-bit and contains about 200 spots distributed according to the beta function. Synthetic spots are generated following a flat top intensity profile in order to emulate the saturation characterizing actual protein spots. Synthetic background is inhomogeneous, noisy and contains streaks, as it is the case with real 2D-GE image background.

The proposed scheme has been implemented in Matlab R2009b and executed on a 3.2 GHz Intel Pentium workstation. Parameters r, T_1, w, q, h, T_2, m, z, T_3, g, λ_1^+, λ_1^-, λ_2^+, λ_2^- and μ were set to 4, 150, 3, 2, 0.01, 160, 3, 3, 75, 100, 1, 1, 1, 1 and $0.006 \cdot 255^2$ respectively, according to preliminary experimentation.

Figure 3 illustrates: (a) a synthetic 2D-GE image, as well as the segmentation results obtained by (b) the proposed scheme, (c) Melanie 7 and (d) PDQuest 8.0.1 software packages. It is evident that the proposed scheme results in more plausible spot boundaries than Melanie 7 and PDQuest 8.0.1. PDQuest 8.0.1 generates elliptical boundaries which do not correspond to the irregular shape of the actual spot boundaries. Furthermore, Melanie 7 identifies false spots and thus, forces the expert biologists to a laborious, error-prone and time-consuming correction process. It should be noted that we conducted further experiments on real 2D-GE images, which lead to similar qualitative conclusions. Due to space constraints these results are not presented in this paper.

The segmentation results are quantified according to the spot volume V [3]:

$$V = \sum_{x,y \in Re\ gion} I(x,y) \quad (4)$$

where $I(x,y)$ is the intensity value of pixel (x,y).
Comparison of the segmentation results with the corresponding ground truth image, as generated by the 2D-GE image simulation software allows the categorization of each pixel in one of the following four region types: "actual spot region (ASR)", "false spot region (FSR)", "false background region (FBR)" and "actual background region (ABR)".

The spot volumes which are calculated according to Eq. (4) for the above four cases of regions, correspond to the "actual spot volume" (ASV), "false spot volume" (FSV), "false background volume" (FBV) and "actual background volume" (ABV), respectively. The segmentation performances are quantitatively evaluated in terms of *sensitivity* and *error*, which are defined as follows:

$$sensitivity = \frac{ASV}{ASV + FBV}, \quad error = \frac{FSV}{ASV + FBV} \quad (5)$$

Table I illustrates the results obtained by the proposed scheme, as well as by Melanie and PDQuest. It is evident that the proposed scheme outperforms Melanie 7 and PDQuest 8.0.1 in terms of *sensitivity* and *error*.

TABLE I: SEGMENTATION RESULTS

	Proposed Scheme	Melanie 7	PDQuest 8.0.1
sensitivity	92,0±1,2%	86,5±3,2%	80,2±4,6%
error	20,0±3,2%	83,1±8,9%	55,0±6,7%

5. CONCLUSIONS

This work introduces, a novel level-set based scheme for protein spot segmentation in 2D-GE images. The proposed

scheme incorporates a protein spot detection stage based on the identification of local intensity minima for the separation of overlapping spots. Histogram adaptation and morphological operators are applied for faint spots identification. A segmentation stage follows, where a level-set function is initialized according to local intensity maxima and evolves guided by energy terms derived in the detection stage. The proposed scheme copes with crucial issues in 2D-GE image analysis, including the presence of noise, overlapping and faint spots.

The experimental evaluation demonstrates that the proposed scheme generates more plausible spot boundaries than PDQuest 8.0.1 and Melanie 7 and outperforms both software packages in terms of sensitivity and error. Future perspectives of this work include further experimentation on real 2D-GE images and integration within a 2D-GE image analysis system.

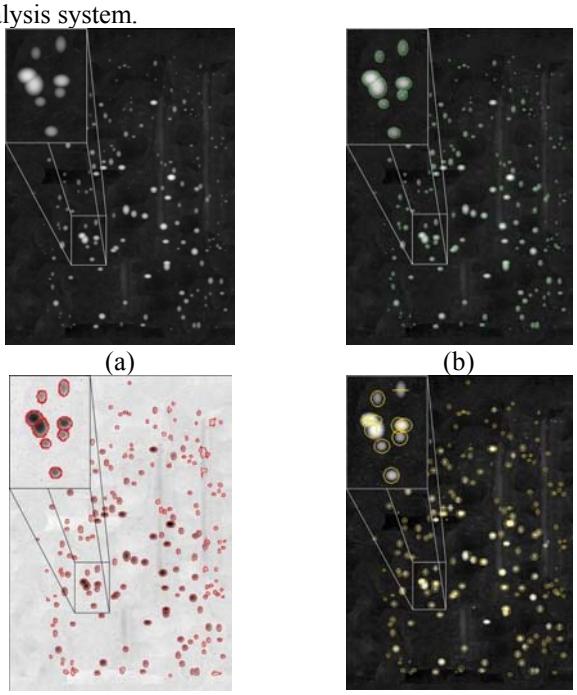


Fig. 3: (a) Synthetic 2D-GE image and segmentation results emerged from: (b) the proposed scheme, (c) Melanie 7 and (d) PDQuest 8.0.1 software packages.

ACKNOWLEDGEMENTS

We would like to thank the expert biologists Dr. S. Kossida and Dr. A. Vlahou of the Biomedical Research Foundation of the Academy of Athens for their fruitful recommendations in the experimental evaluation of the results. This research 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.

6. 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] 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, pp. 2735-2748, 1997.

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

[5] J.I. Garrels, "The Quest System for Quantitative Analysis of Two-Dimensional Gels," *Journal of Biological Chemistry*, vol. 264, no. 9, pp. 5269-5282, 1989.

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

[7] 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.

[8] 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.

[9] P.F. Lemkin and L.E. Lipkin, "2-D Electrophoresis Gel Data-Base Analysis Aspects of Data-Structures and Search Strategies in Gellab," *Electrophoresis,* vol. 4, no. 1, pp. 71-81, 1983.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] T.F. Chan and L.A. Vese, "Active Contour Without Edges," *IEEE Trans. Im. Proc*., vol. 10, no. 2, pp. 226-277, 2001.

[15] 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.

[16] P. Soille, Morphological Image Analysis-Principles and Applications, Springer, Berlin, 1999.