Proceedings of the 10th conference of the Hellenic Society for Computational Biology & Bioinformatics
HSCBB15 – Biomedical Research Foundation Academy of Athens (BRFAA), Athens, Greece, 09-11/10/2015

A novel approach for accurate 2D-gel image segmentation

Eirini Kostopoulou[1], Stamos Katsigiannis[1], Dimitris Maroulis[1], K.I. Pappa[2], and N.P. Anagnou[2]
[1] Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens, Ilisia, 15784, Greece
[2] School of Medicine, National and Kapodistrian University of Athens, Goudi, 115 27, Greece
Correspondence to: ikostop@di.uoa.gr

Abstract—**This work introduces a novel 2D-gel image spot detection and segmentation method based on multi-thresholding, a custom grow-cut algorithm, region growing and morphological operators. Experimental results on real and synthetic data show that the proposed method outperforms state-of-the-art software solutions and results in more accurate segmentation.**

I. Introduction

The field of proteomics offers key opportunities such as the exploration of various complex biological processes inside the cells [1]. A widely used proteomics technique is the 2D Polyacrylamide Gel Electrophoresis (2D-PAGE) technique which achieves protein separation [2]. 2D-PAGE leads to the creation of a digital image (2D-gel) that may contain up to thousands of protein spots. Spot detection and segmentation are very challenging tasks of 2D-gel image analysis, since these images may exhibit high levels of noise, artifacts, inhomogeneous background and highly overlapped spots [3]. A number of commercial software solutions are available for 2D-gel image analysis including PDQuest [4], Melanie 7 [5], ImageMaster 2D [6], Delta2D [7] and Progenesis Samespots [8], each achieving a different level of success. Moreover, various works are available in the literature. These approaches are based on the detection of local minima [9], the watershed algorithm [10], morphological operations [11], active contours [12], 2D histograms, 3D spot morphology [13], etc. The most common drawbacks of these approaches are the segmentation of overlapping spots as one single spot, the segmentation of a single spots into more, the failure to detect some spots, the detection of artifacts as spots, the incorrect approximation of spot boundaries and the need for manual parameter tuning. To address these problems, the authors present a novel 2D-gel image spot detection and segmentation method.

II. Methodology

A. Spot detection

The 2D-PAGE image is processed in overlapping and tiled windows W_i and W_i' i =1,...,N of size dxd and d'xd' respectively, in order to address the different characteristics along the regions of the image. Then, a multi-level thresholding technique [14] is applied to each W_i, in order to compute $T_{i,j}$, j=1,..l,…,h,..,M thresholds. The pixels of W_i' corresponding to W_i are classified as foreground (F) if their intensity value is higher than $T_{i,h}$ and as background (B) if their intensity is lower than $T_{i,l}$. Then, the pixels that belong to the set $\cup_{i=1}^{N} F_i$, i=1,2,...,N and correspond to local intensity maxima are selected as candidate spot centers. In order to eliminate multiple spot center candidates inside a spot, the Euclidean distance amongst them is determined and if it is less than k then they are merged. Moreover, spot center candidates located near spot boundaries are removed if after applying optimal thresholding technique [15] in a rectangular region of size axa around them, the Euclidean distance among them and the nearest pixel classified as background is less than k.

B. Spot segmentation

A rough segmentation of the image is first obtained by applying a modified grow-cut algorithm [16] using pixels labeled as foreground (F) and background (B) by the spot detection stage as seeds. Initially, each non-seed pixel is labeled as "undefined" and assigned a strength $\theta_p = 0$, whereas the θ_p for seed pixels is set as 1. Each pixel p is labeled by comparing θ_p with the attack force $\lambda(q_{an}) = g(q_{an}) \cdot \theta(q_{an})$ of its already labeled neighbors, an=1,..,8. If $\lambda(q_{an}) > \theta_p$, then p receives the label of q_{an} and θ_p becomes equal to $\lambda(q_{an})$.

$$g(q,p) = \begin{cases} 0, & p,q: ([l(q)=f] \wedge [I(p) < T_{i,l}]) \vee \\ & ([l(q)=b] \wedge [I(p) > T_{i,h}]) \\ 1, & p,q: ([l(q)=f] \wedge [I(p) > T_{i,h}]) \vee \\ & ([l(q)=b] \wedge [I(p) < T_{i,l}]) \\ 1 - \dfrac{T_{i,h} - I(p)}{T_{i,h} - T_{i,l}}, & p,q: [l(q)=f] \wedge [T_{i,l} \le I(p) \le T_{i,h}] \\ \dfrac{T_{i,h} - I(p)}{T_{i,h} - T_{i,l}}, & p,q: [l(q)=b] \wedge [T_{i,l} \le I(p) \le T_{i,h}] \end{cases} \quad (1)$$

The grow-cut g function is replaced as shown in Eq. 1 in order to take into account the label of q_{an}, augmenting and decreasing λ as needed. After the grow-cut algorithm, each region R_a formed by pixels labeled as foreground, is segmented to the spot regions $R_{a,b}$ it contains through a majority voting criterion among labeled neighboring pixels. The final spot boundaries are determined by the application of the optimal thresholding technique [15] to each dilated $R_{a,b}$ region for both the intensity and gradient intensity domains.

III. Experimental Results

Real as well as synthetic 2D-gel images containing a total of ~10 200 spots were utilized for the experimental evaluation against two widely used software packages: Melanie 7 [5] and PDQuest [4]. Spot detection performance was evaluated

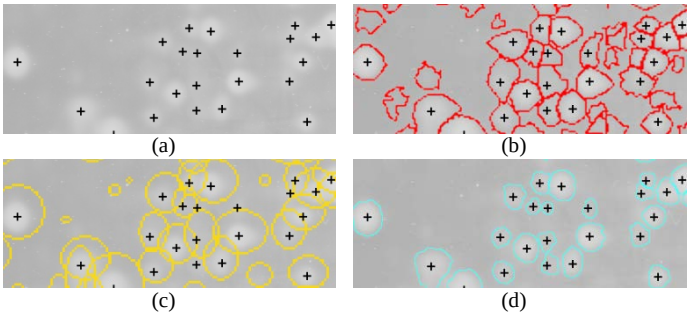
through precision (*P*), sensitivity (*S*), and their weighted harmonic mean, i.e. the F-measure (*F*), as shown in Table I.

TABLE I. SPOT DETECTION RESULTS

Method	***P***	***S***	***F***
Melanie 7	85.2 %	94.1 %	89.4 %
PDQuest	86.3 %	93.9 %	89.9 %
Proposed	96.7 %	93.5 %	95.1 %

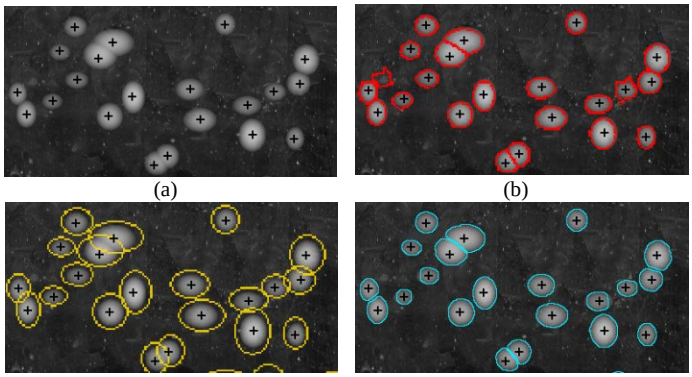
TABLE II. SPOT SEGMENTATION RESULTS

Method	***VO***	***VE***	***VOE***
Melanie 7	98.1 %	25.6 %	21.6 %
PDQuest	99.4 %	82.1 %	44.8 %
Proposed	97.3 %	5.6 %	8.1 %



(a) (b) (c) (d)

Figure 1. (a) Region of a real 2D-GE image. Segmentation results obtained by (b) Melanie 7, (c) PDQuest, and (d) the proposed approach. Black crosses indicate spot locations (ground truth).



(a) (b) (c) (d)

Figure 2. (a) Region of a synthetic 2D-GE image. Segmentation results obtained by (b) Melanie 7, (c) PDQuest, and (d) the proposed approach. Black crosses indicate spot locations determined by the ground truth.

A region of a real 2D-PAGE image and the results obtained by Melanie 7, PDQuest and the proposed approach are shown on Fig. 1. The proposed approach detected all protein spots and no spurious spots, in contrast to both Melanie 7 and PDQuest that detected a large number of spurious spots.

To evaluate the segmentation performance, the volume of pixels characterized as "Actual Spot" (*ASV*), "False Spot" (*FSV*) and "False Background" (*FBV*) is computed in order to calculate the Volumetric Overlap (*VO*), Volumetric Error (*VE*) and Volumetric Overlap Error (*VOE*) defined as follows:

$$VO = \frac{ASV}{ASV+FBV}, VE = \frac{FSV}{ASV+FBV}, VOE = 1 - \frac{ASV}{ASV+FBV+FSV} \quad (2)$$

The proposed approach created more accurate spot boundaries and achieved a very high *VO*, as well as the lowest *VE* and *VOE*, as shown in Table II and Fig. 1 and 2. Both Melanie 7 and PDQuest include background in spot boundaries, thus resulting to higher *VO* but also to high *VE* and *VOE*.

IV. CONCLUSIONS

Spot detection and segmentation on 2D-gel images is a challenging task that is hindered by the inherent characteristics of these images. In this work, the authors present a novel approach for spot detection and segmentation on 2D-gel images. Experiments on real and synthetic 2D-gel images show that the proposed approach is robust and effective and outperforms state-of-the-art alternatives, achieving significantly lower segmentation errors, thus providing more accurate results.

REFERENCES

[1] J. Lee, J. Han, G. Altwerger, and E. Kohn, "Proteomics and biomarkers in clinical trials for drug development," J. Proteomics, vol. 18, pp. 2632–2641, 2011.

[2] J. Lopez, "Two-dimensional electrophoresis in proteome expression analysis, J. Chromatograph. B, vol. 849, no. 1-2, pp. 190–202, 2007.

[3] A. Görg, W.Weiss, and M. J. Dunn, "Current two - dimensional electrophoresis technology for proteomics," Proteomics, vol. 4, no. 12, pp. 3665-3685, 2004.

[4] 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.

[5] Melanie 7, Bio-Rad, Geneva, Switzerland, http://www.genebio.com

[6] ImageMaster 2D, GE Healthcare, Munich, Germany, http://www.gelifesciences.com

[7] Delta2D, DECODON GmbH, BioTechnikum Greifswald, Germany, http://www.decodon.com

[8] Progenesis SameSpots, Nonlinear Dynamics, UK, http://www.totallab.com

[9] J. Morris, B. Clark, and H.Gutstein, "Pinnacle: A fast, automatic and accurate method for detecting and quantifying protein spots in 2-dimensional gel electrophoresis data," Bioinformatics, vol. 24, no. 4, pp. 529–536, 2008.

[10] A. dos Anjos, A. Moller, B. Ersboll, C. Finnie, and H. Shahbazkia, "New approach for segmentation and quantification of two-dimensional gel electrophoresis images, Bioinformatics, vol. 27, no. 3, pp. 368–375, 2011.

[11] E. Mylona, M. Savelonas, D. Maroulis, and S. Kossida, "A computer based technique for automated spot detection in proteomics images," IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 4, pp. 661–667, 2011.

[12] M. Savelonas, E. Mylona, and D. Maroulis, "Unsupervised 2D-gel electrophoresis image segmentation based on active contours," Pattern Recog., vol. 45, no. 2, pp. 720–731, 2012.

[13] E. Kostopoulou, E. Zacharia, and D. Maroulis, "An Effective Approach for Detection and Segmentation of Protein Spots on 2D-gel Images," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 1, pp. 67-76, 2014.

[14] P.S. Liao, and P.C. Chung, "A fast algorithm for multilevel thresholding," Journal of Information Science and Engineering vol. 17, no. 5, pp. 713-727, 2001.

[15] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision. Florence, KY, USA: Cengage-Engineering, 2007.

[16] V. Vezhnevets, V. Konouchine, "GrowCut - Interactive Multi-Label N-D Image Segmentation By Cellular Automata," in: Proceedings of the 15th International Conf. on Computer Graphics and Applications, 2005.