A Custom Grow-cut Based Scheme for 2D-gel Image Segmentation

Eirini Kostopoulou, Stamos Katsigiannis, and Dimitris Maroulis, *Member, IEEE*

Abstract— **This work introduces a novel method for the detection and segmentation of protein spots in 2D-gel images. A multi-thresholding approach is utilized for the detection of protein spots, while a custom grow-cult algorithm combined with region growing and morphological operators is used for the segmentation process. The experimental evaluation against four state-of-the-art 2D-gel image segmentation algorithms demonstrates the superiority of the proposed approach and indicates that it constitutes an advantageous and reliable solution for 2D-gel image analysis.**

I. Introduction

The field of proteomics has gained a lot of attention in the past years and has been increasingly utilized for large scale study of proteins in cells [1], in order to assist the evaluation of new drugs, the creation of new biomarkers and the exploration of biological events [2]. The Two-Dimensional Gel Electrophoresis (2D-gel) technique is a powerful proteomics technique that has the ability to separate thousands of proteins on polyacrylamide gels in order to facilitate protein identification [3]. Results are visualized into a digital image that can be further analyzed. The aim of 2D-gel image analysis is the quantification of the expression level of each protein in order to extract biological conclusions. This is achieved by detecting the pixels belonging to protein spots through an image segmentation process. 2D-gel images are processed in sets of samples from the same experiment and segmentation is applied either on each original image or after aligning the images of the set in order to match the corresponding protein spots. Alignment can be applied after or prior to detecting and segmenting the protein spots [4]. Spot detection and segmentation on 2D-gel images can be a very challenging task due to the characteristics of these images, which may contain thousands of spots with varying intensity, size and shape. Moreover, adjacent spots often overlap with each other and the overall quality of these images is affected by artifacts, inhomogeneous background and high levels of noise [5].

Many commercial software programs for 2D-gel image analysis are available, achieving different levels of success [4]: e.g. Melanie 7 [6], PDQuest [7], Delta2D [8], DeCyder 2D [9], and ImageMaster 2D [10]. Their most important drawback is their dependence on manual parameter tuning that can lead to subjective results due to user intervention. Except for the available software, many methods have also been proposed in the literature for 2D-gel image analysis. Approaches were based on spot shape [11], local minima detection on a denoised average image [12], watershed [13] on an average image [14] and on individual images [15], morphological operations [16], active contours [17], 2D histograms and 3D spot morphology [18], and multidirectional texture and spatial intensity information [19].

Nevertheless, the results of both the software programs and the proposed methods may suffer from: 1) segmentation of overlapping spots as a single spot, 2) splitting a single spot into more, 3) failure to detect some spots, 4) detection of artifacts as spots and 5) inaccurate spot boundaries. All these problems affect the extracted protein expression levels and can lead to erroneous biological conclusions. Extensive manual editing of the results is needed in order to address these problems, a process that is time-consuming and leads to subjective and non-reproducible results and to limited throughput of the analysis process.

In this work, a novel approach for detecting and segmenting protein spots on 2D-gel images is presented. A multi-thresholding scheme is locally applied on the image in order to detect the protein spots, while a custom grow-cut algorithm [20], combined with a region growing scheme and morphological operators, is utilized for the segmentation process. The proposed approach is evaluated against three commercial software programs: Melanie 7 [6], PDQuest [7] and Delta2D [8], and against the recently proposed approach "Scimo" [15]. Experimental evaluation on a real and a synthetic 2D-gel image dataset, containing a total of approximately ~10000 protein spots, demonstrated the effectiveness of the proposed method for both detection and segmentation.

The rest of this paper is organized in three sections. The proposed approach is described in Section 2, whereas the experimental evaluation is presented in Section 3. Finally, conclusions are drawn in Section 4.

II. Methodology

A. Spot detection

For the detection of spot centers, the image I is split into overlapping windows W_i of size dxd, and tiled windows W_i' of size of size $d'xd'$, as shown in Fig. 1(a), with d determined as $d = d' + 2\lceil d'/8 \rceil$ and $i = 1,\ldots,N$. The number N of tiled windows is calculated as $N = \lceil \text{Image Width}/d' \rceil \cdot \lceil \text{Image Height}/d' \rceil$. Then, the multiple Otsu thresholding technique [21] is used in order to automatically obtain $T_{i,j}, j=1,\ldots,l,\ldots,h,\ldots,M$ thresholds for each W_i. The pixels inside W_i' with intensity values below $T_{i,l}$ are classified as background (B_i) pixels and above $T_{i,h}$ are classified as

*This work was supported in part by the European Union (European Social Fund—ESF) and in part by Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: THALIS, UOA, CERVI-CAN-PROT.

E. Kostopoulou, S. Katsigiannis and D. Maroulis are with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Panepistimioupolis, Athens, Greece (E. Kostopoulou: +30 210 7275317; fax: +30 210 72753333; e-mail: ikostop@ di.uoa.gr).

foreground (F_i) pixels. Intensities between the two thresholds indicate pixels that may belong to either category, as shown in Fig. 1(b).

The pixels classified as foreground that correspond to local intensity maxima constitute the candidate spot centers. An elimination process is then performed in order to select one spot center per spot. A local thresholding technique [22] is applied on a rectangular region of size a x a around each candidate spot center s. Then, s is excluded for the candidate spot centers if its Euclidean distance from the nearest pixel classified as background inside the a x a region is less than k, with k being the minimum allowed distance. Then, the remaining spot center candidates with a pairwise Euclidean distance less than k are merged. An example of the final spot centers for a window W_i' is shown in Fig. 1(c).

B. Segmentation

The first step of the segmentation process is the selection of pixels that are predefined as spot or background pixels. The pixels p that belong to the sets B and F of the background and spot seeds respectively are determined as:

$$B = \{p: (p \in W_i') \wedge (p \in W_i) \wedge [I(p) < T_{i,l}]\} \quad (1)$$

$$F = \{p: (p \in W_i') \wedge (p \in W_i) \wedge [I(p) > T_{i,h}] \quad (2)$$

$I(p)$ is the intensity p, i=1,...,N and $h = 2 \cdot l$. The automated selection of seed pixels enhances the reproducibility of the segmentation results by avoiding user intervention. Then, the grow-cut [20] segmentation algorithm is applied in order to distinguish between background and spot pixels. Each pixel p is represented by a triplet $(l_p, \theta_p, \vec{C}_p)$, with l_p being the class label of p, θ_p the measure of certainty that l_p corresponds to p and $\vec{C}_p$ is equal to the intensity of p since 2D-gel images are grayscale. At the beginning, l_p is set to "Background" (b) and "Spot" (f) for the background and spot seed pixels respectively and their θ_p is set to 1. The label of the rest pixels is set to "Undefined" and their θ_p to 0. Then, the following process is iterated until all pixels have been labeled: For each pixel p, $g(q_{an}, p)$ is computed for its adjacent neighbors q_{an}, an=1,..,8. The modified $g(q_{an}, p)$ (3) is used instead of the original grow-cut g function [20] in order to take into consideration both the difference in intensity between p and q, and the class q belongs to. The original g function failed to allow the correct labeling of pixels near the spot boundaries due to the minimal difference in their intensity values compared to the adjacent background pixels.

$$g(q,p) = \begin{cases} 0, & ([l_q = f] \wedge [I(p) < T_{i,l}]) \vee ([l_q = b] \wedge [I(p) > T_{i,h}]) \\ 1, & ([l_q = f] \wedge [I(p) > T_{i,h}]) \vee ([l_q = b] \wedge [I(p) < T_{i,l}]) \\ 1 - \frac{T_{i,h} - I(p)}{T_{i,h} - T_{i,l}}, & (l_q = b) \wedge (T_{i,l} \le I(p) \le T_{i,h}) \\ \frac{T_{i,h} - I(p)}{T_{i,h} - T_{i,l}}, & (l_q = b) \wedge (T_{i,l} \le I(p) \le T_{i,h}) \end{cases} \quad (3)$$

Then, $\lambda(q_{an}) = g(q_{an}, p) \cdot \theta(q_{an})$ is computed for all q_{an}

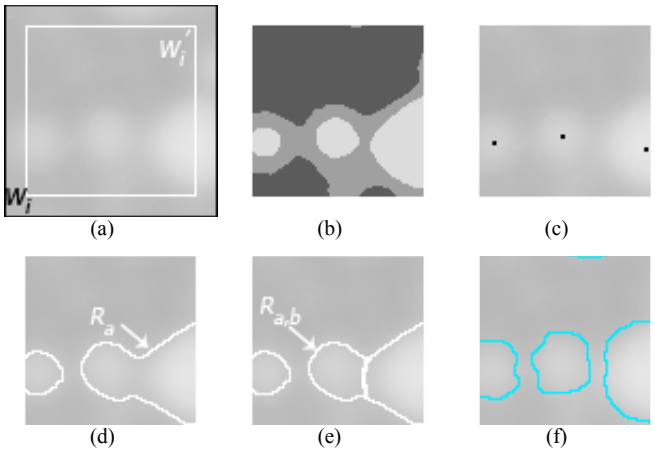


Fig. 1. (a) Overlapping W_i and tiled W_i' window for a region of a real 2D-gel image, (b) foreground, background and middle intensity pixels of W_i', (c) spot centers, (d) modified grow-cut result, (e) region growing result, (f) fusion of optimal thresholding results on intensity and gradient values.

with lq_{an}≠"Undefined". If $\lambda(q_{an})$>θ_p, then l_p becomesequal to $l_{q_{an}}$ and θ_p becomes equal to $\lambda(q_{an})$. After enough iterations, the algorithm will converge and pixel labels will cease to change. Fig. 1(d) shows the region depicted in Fig. 1(a) after grow-cut's convergence.

Then, for each region R_a formed by pixels labeled as spot pixels by the grow-cut algorithm, the pixels that correspond to spot centers are labeled as belonging to a different region $R_{a,b}$, b=1,2,..., and the pixels that are not spot centers are labeled as belonging to an undefined region. The undefined pixels are then sorted into descending intensity order and are separated into z groups GP_e, e=1, ... , z, each containing pixels of x distinct intensities. For each GP_e , pixels p that are not assigned to a region, are assigned to the region of the majority of their neighbors, without considering neighbors with undefined region. This process is iterated until all pixels have been assigned to a region $R_{a,b}$, as shown in Fig. 1(e). Then, the optimal thresholding technique [22] is applied to the intensity values of each region $R'_{a,b}$, which is defined as $R'_{a,b} = \{p: (p \in (R_{a,b} \oplus D)) \wedge (p \notin (R_{a,b}, \oplus D))\}$ where $R_{a,b} \oplus$D denotes the application of the dilation morphological operation on $R_{a,b}$, with the disk (D) of radius r as the structuring element, in order to determine the $R'_{a,i}$ which contains the pixels with "high" intensity values. The optimal thresholding technique is also applied to the gradient intensity values of each region $R'_{a,b}$ in order to determine $R'_{a,o}$ which contains the pixels with "high" gradient intensity values. The final spot pixels for each region $R_{a,b}$ are determined as the pixels: $(p \in R'_{a,o}) \vee (p \in R'_{a,i})$, as illustrated in Fig. 1(f).

III. Experimental Evaluation

Several experiments on real as well as synthetic 2D-gel images were conducted in order to evaluate the performance of the proposed approach against other widely used software packages and techniques; namely Melanie 7, PDQuest, Scimo and Delta2D. It should be noted that these techniques need parameter tuning -by expert biologists- for every single image, while the optimal parameters for the proposed approach were experimentally determined once and then

used for all the experiments. The parameters utilized were: d'=80, M=6, h=2, l=4, k=4, a=31, x=10, r=5.

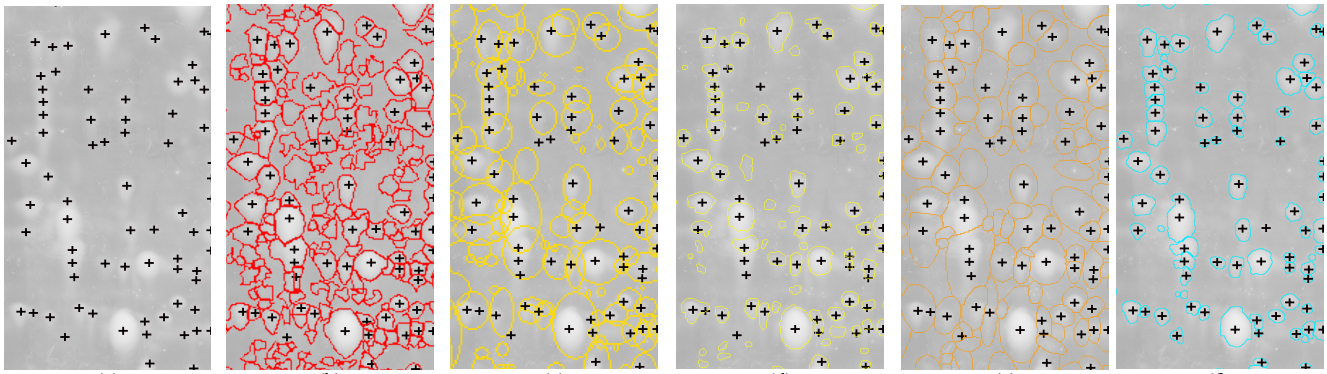
The real 2D-gel image images as well as their detection ground truth were kindly provided by the courtesy of the Biomedical Research Foundation of the Academy of Athens (BRFAA) [23], and their use was approved by the local Ethics Committee of the BRFAA. Table I presents the detection results for the real 2D-gel images in terms of Sensitivity, Precision, and the F-measure [24]. Higher Sensitivity indicates that less real spots are not detected, while higher Precision indicates that fewer spurious spots are detected. The F-measure is a more reliable measure as it takes into consideration both Sensitivity and Precision. As a result, higher values indicate better overall detection performance. It is evident, that the proposed approach outperforms the other methods as it: a) detects almost all protein spots, b) detects very few spurious spots, and c) achieves the highest F-measure value. Fig. 2 depicts the detection and segmentation results of the proposed approach as well as the compared techniques for a region of a real 2D-gel image. The location of each spot -according to the ground truth given by expert biologists- is illustrated with a black cross. It is evident that all the compared techniques detect a large number of spurious spots (Fig. 2(b-e)), contrary to the proposed approach (Fig. 2(f)).

TABLE I. DETECTION RESULTS FOR THE REAL 2D-GEL IMAGES

Measure	Method				
	Melanie 7	***PDQuest***	***Scimo***	***Delta2D***	***Proposed Approach***
Sensitivity	91.2 ± 2.7	91.0 ± 3.0	73.5 ± 3.5	75.0 ± 6.3	90.4 ± 2.8
Precision	83.0 ± 2.9	83.5 ± 2.7	88.6 ± 3.9	72.6 ± 16.1	93.4 ± 2.7
F-measure	86.9 ± 2.8	87.1 ± 3.2	80.3 ± 3.3	73.8 ± 10.7	91.9 ± 2.7

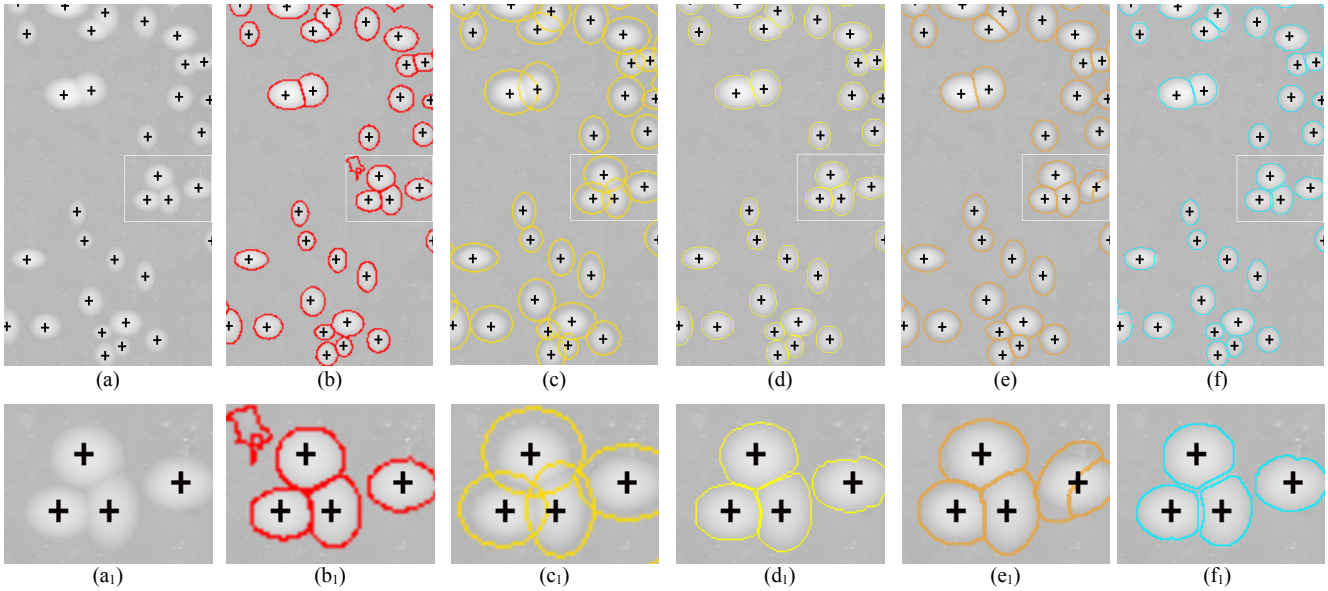
Additionally, the spot boundaries obtained by PDQuest often overlap, and similar to those obtained by Delta2D: a) have elliptical shape, contrary to real spots which have irregular shape and b) may include more than one spot. Moreover, Melanie7, PDQuest and Delta2D include background pixels inside the resulting spot regions, while boundaries obtained by Scimo often exclude parts of spots. On the contrary, the proposed approach effectively detected and segmented almost all of the protein spots and provided more plausible spot boundaries.

Since the segmentation ground truth cannot be available for the real images, synthetic 2D-gel images emulating the real ones were generated by our group in order to allow for a qualitative as well as quantitative comparison of the segmentation results. By comparing the segmentation results with the ground truth, each pixel can be characterized as "Actual Spot" (AS), "Actual Background" (*AB*), "False



(a) (b) (c) (d) (e) (f)

Fig. 2. (a) Region of a real 2D-gel image. Spot boundaries obtained by (b) Melanie 7, (c) PDQuest, (d) Scimo, (e) Delta2D, and (f) the proposed approach.



(a) (b) (c) (d) (e) (f)

(a_1) (b_1) (c_1) (d_1) (e_1) (f_1)

Fig. 3. (a) Region of a synthetic 2D-gel image. Detection and segmentation results for (a) obtained by (b) Melanie 7, (c) PDQuest, (d) Scimo, (e) Delta2D, and (f) the proposed scheme, whereas (a_1) – (f_1) are magnified regions of (a)–(f), respectively.

TABLE II. SEGMENTATION RESULTS FOR THE SYNTHETIC 2D-GEL IMAGES

Measure	Method				
	Melanie 7	*PDQuest*	*Scimo*	*Delta2D*	*Proposed Approach*
VO	95.6 ± 2.1	97.9 ± 1.1	97.3 ± 0.8	98.6 ± 0.8	95.3± 2.0
VE	27.1 ± 9.9	84.9 ± 23.3	14.0 ± 2.9	69.7 ± 20.8	6.9 ± 1.7
VOE	23.6 ± 5.1	47.3 ± 6.5	14.9 ± 2.1	45.2 ± 8.7	9.2 ± 1.1

Spot" (*FS*), and "False Background" (*FB*) pixel. Then, the Volumetric Overlap ($VO = ASV/(\mathrm{ASV} + \mathrm{FBV})$), Volumetric Error ($VE$=FSV/(ASV+FBV)) and Volumetric Overlap Error ($VOE = ASV/(\mathrm{ASV} + \mathrm{FBV} + \mathrm{FSV})$) measures [25] can be estimated utilizing "*ASV*", "*FSV*", "*ABV*", "*FBV*" which correspond to the cumulative volumes of "*AS*", "*FS*", "*AB*", and "*FB*" pixels, respectively. The segmentation performance of each method is presented on Table II. Higher *VO* indicates higher overlap between the segmented spot pixels and the ground truth, while lower *VE* and *VOE* values indicate less falsely segmented pixels. The lowest *VE* and *VOE* as well as the high *VO* achieved by the proposed approach demonstrate its effectiveness in segmenting 2D-gel images compared to the four other evaluated methods. Segmentation results for the evaluated methods for a region of a synthetic image (Fig. 3(a)) are shown in Fig. 3(b-f), while magnified regions of Fig. 3(a-f) are depicted in Fig. 3(a_1-f_1), respectively. It is evident that the proposed approach provides more accurate spot boundaries that contain almost all spot pixels, while avoiding the inclusion of background pixels. On the other hand, the results obtained by PDQuest and Delta2D include many background pixels inside the spot boundaries, while Melanie 7 and Scimo occasionally miss spot pixels. Moreover, PDQuest results in boundaries that overlap.

IV. CONCLUSION

In this paper, the authors proposed an original approach for segmenting 2D-gel images. The proposed methodology combines a custom grow-cut segmentation algorithm with a region growing approach and morphological operators. Evaluation on real and synthetic 2D-gel images demonstrated its robustness and effectiveness for 2D-gel image spot detection and segmentation. The proposed approach outperformed state-of-the-art 2D-gel image analysis software packages and techniques, achieving significantly lower volumetric error (*VE*) and volumetric overlap error (*VOE*). As a result, it provides more accurate results, leading to improved quantification of the protein expression levels and as a consequence to enhanced reliability of the extracted biological conclusions.

ACKNOWLEDGMENT

The authors would like to thank the a) Biomedical Research Foundation of the Academy of Athens for providing the real 2D-gel images, their detection ground truth, as well the segmentation results obtained by Melanie 7, PDQuest and Scimo and b) the Institute of Molecular Biology and Biotechnology (IMBB) of the Foundation for Research and Technology-Hellas (FORTH) for the provision of segmentation results obtained by Delta2D.

REFERENCES

[1] S.L. Havre, M. Singhal, D.A. Payne, M.S. Weir Lipton, B.-J.M Webb-Robertson, "Enabling proteomics discovery through visual analysis," IEEE Engineering in Medicine and Biology Magazine, vol. 24, no. 3, pp. 50-57, May-Jun 2005.

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

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

[4] S. Magdeldin, Y. Zhang, B. Xu, Y. Yoshida, T. Yamamoto, "Two-dimensional polyacrylamide gel electrophoresis—A practical perspective," Gel Electrophor Principles Basics, pp. 91–116, 2012.

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

[6] Melanie 7, Bio-Rad, Geneva, Switzerland, www.genebio.com.

[7] J.I. Garrels, "The QUEST system for quantitative analysis of two-dimensional gels," Journal of Biological Chemistry, vol. 264, no. 9, pp. 5269-5282, Mar. 1989.

[8] Delta2D, DECODON GmbH, BioTechnikum Greifswald, Germany, www.decodon.com.

[9] DeCyder 2D, GE Healthcare, Germany, www.gelifesciences.com.

[10] ImageMaster 2D, GE Healthcare, Germany, www.gelifesciences.com.

[11] J. W. Yoon, S. J. Godsill, C. Kang, T.S. Kim, "Bayesian inference for 2D-gel electrophoresis image analysis," in Proc. 1st Int. Conf. Bioinf. Res. Develop, 2007, pp. 343–356.

[12] 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, Feb. 2008.

[13] F. Li, F. Seillier-Moiseiwitsch, "Differential analysis of 2D-gel images," Methods Enzymol., vol..487, 595–605, 2011.

[14] L. Vincent, P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, no. 6, pp. 583–598, Jun. 1991.

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

[16] E. Mylona, M. Savelonas, D. Maroulis, 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, Apr. 2011.

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

[18] E. Kostopoulou, E. Zacharia, 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, Jan. 2014.

[19] E. Zacharia, E. Kostopoulou, D. Maroulis, N. P. Anagnou, K. I. Pappa, "2D-gE spot detection combining multidirectional texture and spatial intensity cues," in: Proc. IEEE 13th Int. Conf. on Bioinformatics and Bioengineering (BIBE), 2013, pp. 1-4.

[20] V. Vezhnevets, V. Konouchine, "GrowCut-Interactive Multi-Label N-D Image Segmentation By Cellular Automata," in Proc.15th Int. Conf. on Computer Graphics and Applications, 2005.

[21] P.S. Liao, P.C. Chung, "A fast algorithm for multilevel thresholding," J. of Inform. Sc. and Engineering, vol. 17, no. 5, pp. 713-727, 2001.

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

[23] Biomedical Research Foundation of the Academy of Athens, (BRFAA), Athens, Greece, www.bioacademy.gr.

[24] A.T. Azar, H, I. Elshazly, A. E. Hassanien, A. M. Elkorany, "A random forest classifier for lymph diseases," Computer Methods and Programs in Biomedicine, vol. 113, no. 2, pp. 465–473,) Feb. 2014.

[25] T. Heimann, M. Bryan, M. Styner, M. Niethammer, S. Warfield, "Segmentation of knee images: A grand challenge," in Proc. MICCAI Workshop on Medical Image Analysis for the Clinic, Beijing, 2010.