

## ACCURATE SEGMENTATION OF 2D-PAGE IMAGES

*Eirini Kostopoulou, Eleni Zacharia, and Dimitris Maroulis*

Department of Informatics and Telecommunications, University of Athens, Greece  
 ikostop@di.uoa.gr, eezacharia@gmail.com, dmaroulis@di.uoa.gr

### ABSTRACT

Segmentation is an essential and crucial process in 2D-PAGE image analysis. Although several software packages and techniques have been developed and are broadly utilized in biology laboratories, none of them can optimally segment 2D-PAGE images. This paper presents an effective approach for 2D-PAGE image segmentation. This approach extends our previous work by improving the segmentation accuracy. Experiments were conducted on a set of synthetic 16-bit 2D-PAGE images containing ~2000 spots of various intensities. To the best of our knowledge, this is the first attempt to evaluate a spot segmentation method on 2D-PAGE images using a set of synthetic images that provide the challenges of real ones; inhomogeneous background, noise and artifacts, as well as faint, saturated and highly overlapping protein spots. Experimental results demonstrate that the proposed approach outperforms state of the art software packages and techniques since it: 1) generates more accurate boundaries and 2) separates overlapping spots correctly.

**Index Terms**— Proteomics, 2D-PAGE image analysis, spot segmentation

### 1. INTRODUCTION

Proteomics is an established and challenging scientific area for studying the proteins produced in cells, tissues and organs of biological organisms. The key challenge is a better understanding of the strong relation between alterations in protein expression profiles and diseases. By unraveling the disease pathways, the process of identification of new biomarkers and drug targets will be accelerated [1].

Two-Dimensional Polyacrylamide Gel Electrophoresis (2D-PAGE) is a convenient and well-established technique for protein separation [2],[3]. During 2D-PAGE, complex mixtures of proteins are separated in two dimensions according to their molecular weight and their isoelectric point. The output of 2D-PAGE is a digital image which depicts proteins as bright or dark spots over a noisy and inhomogeneous background. Each protein is characterized by its vertical and horizontal position on the image, as well as its brightness, size and shape.

Accurate segmentation of 2D-PAGE images is therefore in need so as to separate spots that overlap and also to distinguish spots from the background. While the segmentation task seems trivial, it is very challenging as it is impeded by the presence of inhomogeneous background, noise and artifacts, as well as the existence of faint, saturated and highly overlapping protein spots.

Several 2D-PAGE image analysis techniques have been proposed in literature such as edge detection [4], stepwise thresholding [5], watersheds [6] and inner-marked watersheds [7]. Edge detection and stepwise thresholding are highly sensitive to noise, artifacts, inhomogeneous background and overlapping spots. Watersheds cope with noise, artifacts, and inhomogeneous background. However, they result in over-segmentation [6] and require manual selection of inner markers [7]. Recent approaches in spot detection and segmentation include Anjos et. al [8] and Savelonas et. al [9] which are based on watersheds and active contours, respectively. The former, named *Scimo*, is noise resistant, but doesn't generate accurate boundaries. The latter is topologically adaptable, but fails to segment overlapping spots.

Several 2D-PAGE image analysis software packages have been developed and are widely utilized by expert biologists in biology laboratories. Among them, Melanie 7.0 [10] and Delta 2D [11] offer a flexible graphical user interface, but require extensive human intervention. The user adjusts input parameters in order to correct the output results, making the analysis of 2D-PAGE images a time-consuming and subjective process.

In our previous work [12], [13] a method for spot detection and segmentation has been introduced. Our present work extends the previous one by improving the segmentation process. The proposed approach explores 2D histograms and 3D morphology of the spot surface as well as the intensity and gradient-intensity values of an area surrounding the spot. Experiments were conducted on a set of synthetic 2D-PAGE images containing ~2000 spots of various intensities, sizes and shapes. Furthermore, these images contain inhomogeneous background and are contaminated with noise and artifacts. An example of a synthetic 2D-PAGE sub-image is depicted on Fig. 1. The main advantage of using synthetic 2D-PAGE images is that the ground truth -in segmentation level- is available. To the

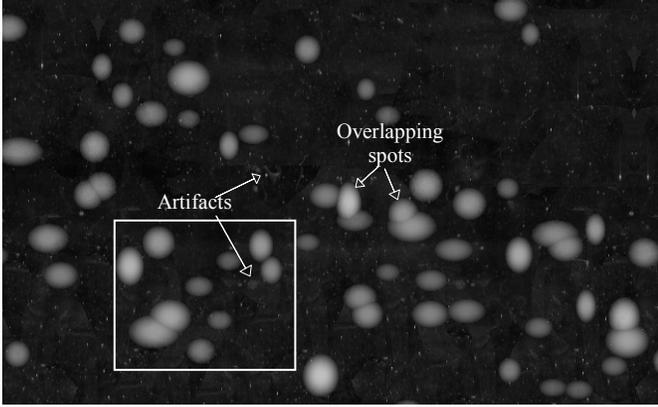


Fig. 1. Synthetic 2D-PAGE sub-image.

best of our knowledge, this is the first attempt of evaluating a spot segmentation method on 2D-PAGE images using a set of synthetic images. The obtained results were compared to the results produced by Melanie 7 and Delta 2D software programs, as well as by Scimo and Savelonas et. al. techniques. The experimental evaluation provides evidence that the proposed approach outperforms the aforementioned software packages and techniques in terms of spot segmentation.

The rest of this paper is organized in three sections. Section 2 describes the proposed approach for spot segmentation on 2D-PAGE images, while Section 3 presents the experimental results obtained by the proposed approach as well as by Melanie 7, Delta 2D, Scimo and Savelonas et. al. Finally, conclusions are summarized in Section 4.

## 2. METHODOLOGY

The main aim of the proposed approach is to enhance the spot segmentation accuracy, so as to generate accurate spot boundaries. Our previous approach [12] was optimized for spot detection while providing adequate performance for spot segmentation. To improve on the generated boundaries, the proposed approach exploits the intensity and gradient intensity values within a dilated sub-region -surrounding each spot- that includes high and low intensity spot pixels.

### 2.1. Main stages of our earlier method

According to [12], a 2D-PAGE image is roughly segmented into a set of  $R_i$  regions containing mostly spots. Subsequently, each distinct  $R_i$  region is further segmented into sub-regions  $R_{ij}$ , each containing a single spot and background. Finally, each spot inside a  $R_{ij}$  sub-region is segmented based on the optimal thresholding technique [14].

Figure 2a depicts a synthetic 2D-PAGE sub-image. Fig. 2b depicts the two regions ( $R_1, R_2$ ) containing mostly spots, whilst Fig. 2c depicts the sub-regions ( $R_{11}, R_{12}, R_{21}$ ) of  $R_1$  and  $R_2$  regions.  $R_{11}, R_{12}$  contain a single spot surrounded by background while  $R_{21}$  contains a single spot. It is worth mentioning that the segmentation of each spot located inside

a sub-region  $R_{ij}$  using our earlier approach is not the optimal one. This can be easily observed in Fig. 3. Indeed, the segmentation results using our previous method (fourth column) differ substantially from the ground truth (second column).



Fig. 2. (a) Synthetic 2D-PAGE sub-image. (b) the set of  $R_i$  regions, (c) the sub-regions of Fig. 2b, (d) dilated sub-regions of Fig. 2c utilized by the proposed approach.

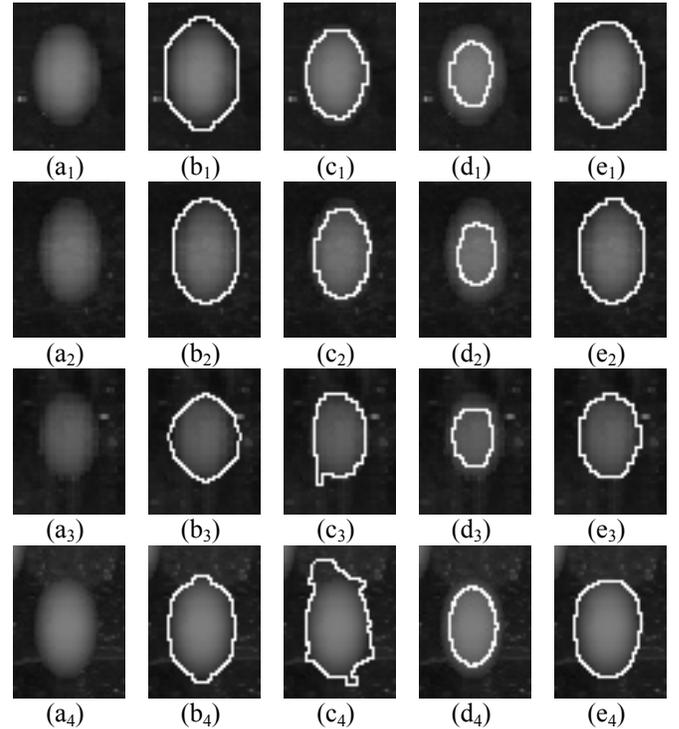


Fig. 3. (a<sub>1</sub>-a<sub>4</sub>) Synthetic 2D-PAGE sub-images containing a single spot, (b<sub>1</sub>-b<sub>4</sub>) ground truth boundaries, (c<sub>1</sub>-c<sub>2</sub>) roughly segmented sub-regions containing spots, (c<sub>3</sub>-c<sub>4</sub>) roughly segmented sub-regions containing spots surrounded by background, (d<sub>1</sub>-d<sub>4</sub>) spot boundaries obtained by our previous approach, (e<sub>1</sub>-e<sub>4</sub>) spot boundaries obtained by the proposed approach.

### 2.2. Accurate segmentation

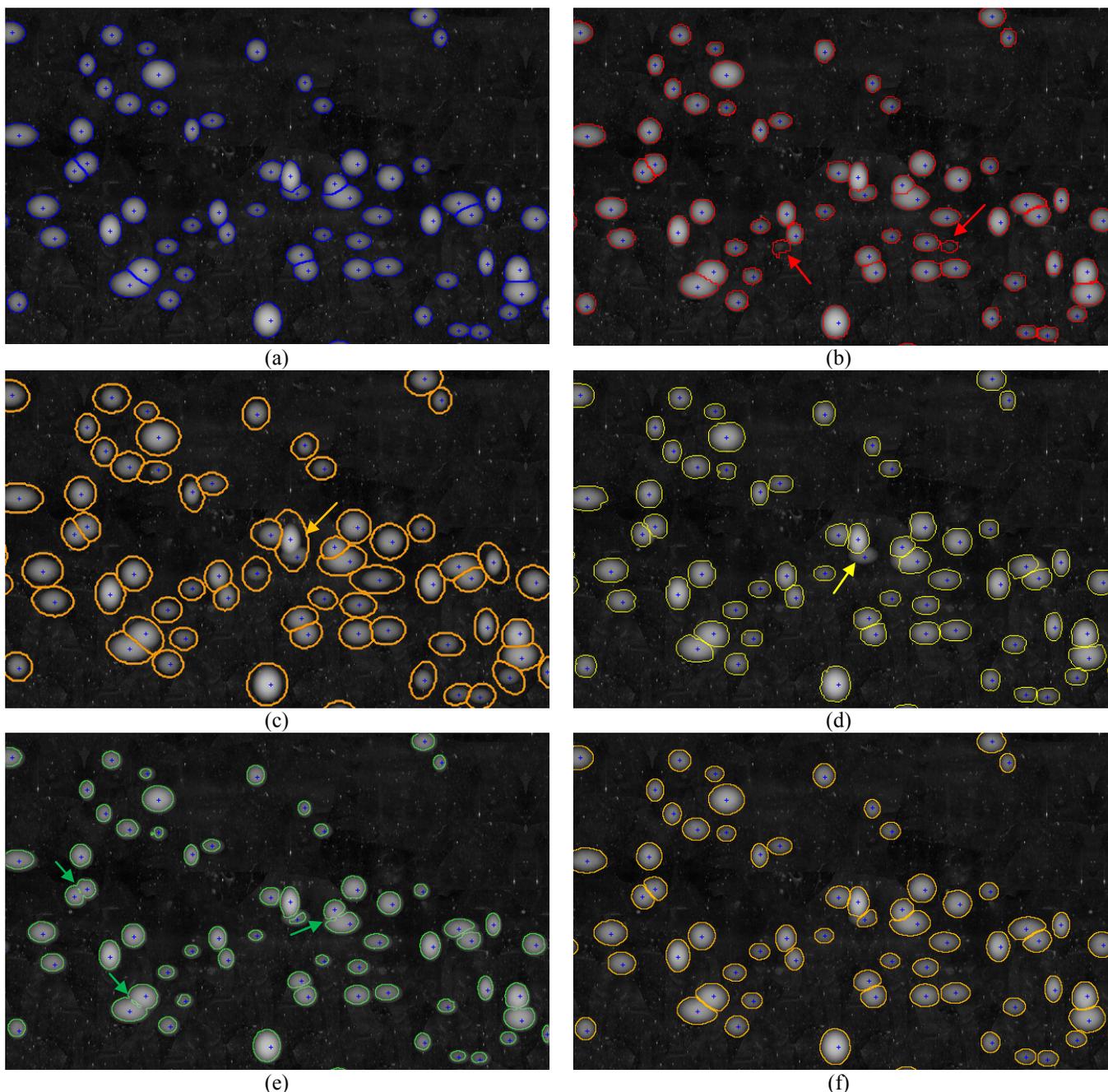
In this approach, the segmented sub-regions  $R_{ij}$  are dilated using a disk-shaped structuring element [14] in order to include spot pixels which might not have been embedded in them. Let  $d_{ij}$  be the dilated sub-region of  $R_{ij}$ . Figure 2d depicts the dilated sub-regions of Fig. 2c. The optimal thresholding technique is applied to: 1) the dilated sub-regions ( $d_{ij}$ ) of the image  $I$ , as well as to 2) the dilated sub-regions ( $d_{ij}$ ) of the gradient  $I_G$  of the image  $I$ . By applying

the optimal thresholding twice in the aforementioned dilated sub-regions, the detection of spot areas with high and low intensity is accomplished. Finally, spot pixels are defined as the pixels  $p$  which fulfill the following:

$$SP = \{p: p \in d_{ij} \text{ and } (I(p) > T_1 \text{ or } I_G(p) > T_2)\} \quad (1)$$

where  $T_1$  and  $T_2$  are two automatically estimated thresholds.

Figure 3d illustrates the spot boundaries of Fig. 3a generated by our previous approach, whilst Fig. 3e depicts the spot boundaries generated by the proposed approach. As one may observe, the proposed approach has generated more accurate boundaries relatively to the ground truth (Fig. 3b) in contrast to our previous approach.



**Fig. 4.** (a) Ground truth of the 2D-PAGE sub-image of Fig. 1. Detection as well as segmentation results obtained by: (b) Melanie 7, (c) Delta 2D, (d) Scimo, (e) Savelonas et. al, and (f) the proposed approach. Positions of 62 spots are indicated with blue crosses.

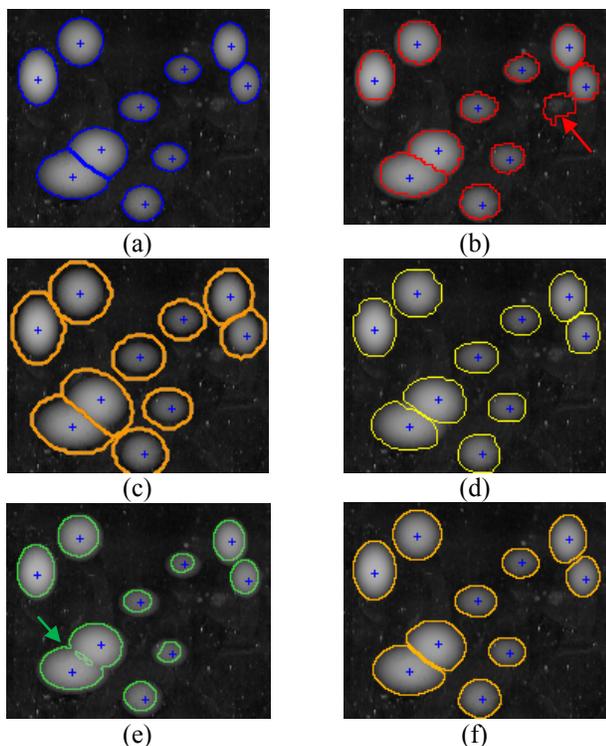
### 3. RESULTS

Several experiments were conducted in order to evaluate the performance of the proposed approach and compare it with state of the art software packages and techniques. Our dataset consisted of synthetic 2D-PAGE images containing ~2000 spots in total. These images have been generated by the ‘Proteus’ software package developed by our research team. The images are of low quality; they contain inhomogeneous background extracted from real 2D-PAGE images, they are contaminated with noise and artifacts and they contain numerous overlapping spots of various intensities, sizes and shapes. To the best of our knowledge, this is the first attempt of evaluating a spot segmentation method on 2D-PAGE images using a set of synthetic images which adopt the challenges of real ones. The main advantage of using synthetic 2D-PAGE images is that the ground truth -in segmentation level- is available and therefore, a quantitative evaluation can be performed.

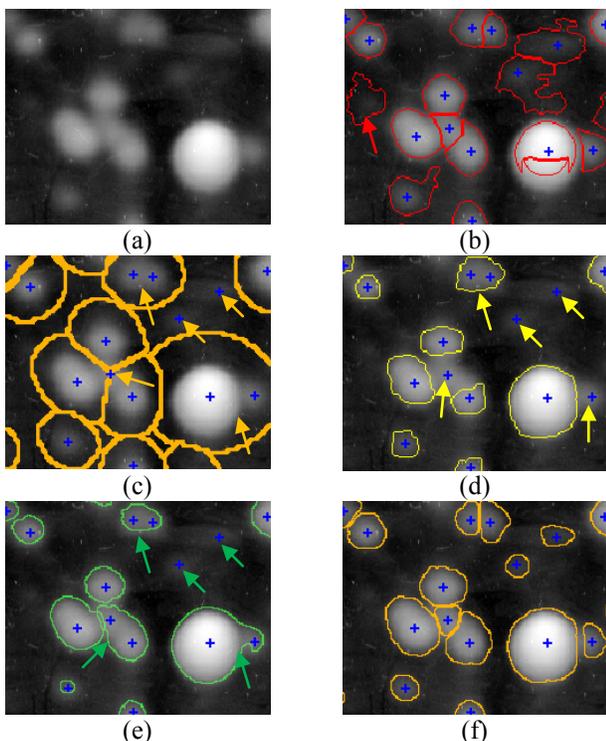
The results were statistically evaluated by computing the normalized error measure  $E$  [15] as well as the Dice similarity coefficient measure [16]. The former estimates the error between an actual spot quantity ( $Q_s$ ) – based on the ground truth – and the spot quantity – measured by a specific software program or technique ( $CQ_s$ ). It is defined as  $E_s = |CQ_s - Q_s| / Q_s$ , where  $CQ_s$  is computed by summing the pixel intensity values  $I(x,y)$  inside the spot boundaries. The Dice measure estimates the overlapping as well as the difference between the segmented spot region ( $S$ ) and the ground truth spot region ( $G$ ). It is defined as  $D(S,G) = 2|S \cap G| / (|S| + |G|)$ . The  $D$  value always lies between 0 and 1, where higher values indicating better segmentation.

Table 1 presents the statistical results. The error value of the proposed approach is limited to 8.9%. Melanie 7 and Scimo have the second and third lowest error values that are ~50% (14.6%) and ~100% (17.7%) greater than the proposed approach, respectively. The proposed approach has the highest Dice value (93.2%), Melanie 7 the second highest (88.1%) and Scimo the third highest (87.8%). Delta 2D and Savalelonas et. al, have Dice values of 75.3% and 73.2%, that are ~25% lower than the Dice value of the proposed approach. The lowest error value as well as the highest Dice value achieved by our proposed approach proves that it accurately segments each individual protein. Therefore, it accurately estimates the amount of each individual protein.

Figure 4 presents the segmentation results of the 2D-PAGE sub-image of Fig. 1 as well as the ground truth in segmentation. The arrows depicted on these images indicate either merged, not segmented or spurious spots. In these images, it is evident that the proposed approach has efficiently segmented the 62 spots of Fig. 1 and has created more accurate boundaries, contrary to the compared software packages and techniques which either include background pixels within the spots regions or exclude spot pixels.



**Fig. 5.** (a) Ground truth overlaid on the enlarged rectangular area of Fig. 1. Segmentation results obtained by: (b) Melanie 7, (c) Delta 2D, (d) Scimo, (e) Savelonas et. al, and (f) the proposed approach.



**Fig. 6.** (a) A real 2D-PAGE sub-image. Segmentation results obtained by: (b) Melanie 7, (c) Delta 2D, (d) Scimo, (e) Savelonas et. al, and (f) the proposed approach.

Table 1. Average estimated error of quantification results and Dice similarity coefficient

Methods	E %	Dice %
Melanie 7	14.6 ± 5.3	88.1 ± 5.6
Delta 2D	32.8 ± 9.1	75.3 ± 9.6
Scimo	17.7 ± 5.2	87.8 ± 7.4
Savelonas et. al	35.2 ± 14.1	73.2 ± 16.4
Proposed Approach	<b>8.9 ± 3.3</b>	<b>93.2 ± 6.3</b>

Figure 5 depicts the segmentation results obtained by Melanie 7, Delta 2D, Scimo, Savelonas et al. and the proposed approach on the rectangular area of Fig.1. It is easily observable that Melanie 7 has segmented an artifact as a spot (see arrow). In the segmentation results of Delta 2D a high number of background pixels is embedded into the spot areas. Scimo and Savelonas et. al, have excluded more spot-pixels from the spot-region. Moreover, Savelonas et. al, has failed to segment two spots that overlap (see arrow). It is obvious that the proposed approach has effectively segmented all the protein spots, by generating boundaries similar to those of the ground truth.

Preliminary experiments were also conducted on real 2D-PAGE images in order to ascertain the effectiveness of the proposed approach on such images. Segmentation results of a real 2D-PAGE sub-image utilized in [12] are illustrated in Fig 6. The arrows indicate the spurious, merged and not segmented spots, which emerged from the software programs and techniques. According to Fig. 6, one may observe that the proposed approach has effectively segmented the protein spots contrary to the compared software packages and techniques.

#### 4. CONCLUSIONS

In this paper, an original method for accurate spot segmentation of 2D-PAGE images is presented. The proposed approach is an extension of our previous one, since low intensity spot areas are encountered into the final spot boundary. Experiments were conducted on synthetic 2D-PAGE images which adopt the challenges of real 2D-PAGE images. To the best of our knowledge this is the first attempt to generate realistic synthetic 2D-PAGE images in order to evaluate qualitative, as well as quantitative segmentation results. The experimental results demonstrate that the proposed approach outperforms state-of-the-art software packages and techniques, by generating more accurate boundaries, as well by separating overlapping spots correctly.

#### 5. ACKNOWLEDGEMENTS

The authors would like to thank Biomedical Research Foundation of the Academy of Athens for the provision of real 2D-PAGE images as well as their detection ground truth.

This research has been 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: THALIS –UOA-CERVI-CAN-PROT.

#### 6. REFERENCES

- [1] G. Rotil and K. Stegmaier, "Genetic and proteomic approaches to identify cancer drug targets," *British Journal of Cancer*, vol. 106, pp. 254–261, 2012.
- [2] M. Berth, F. Moser, M. Kolbe, and 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.
- [3] A.W. Dowsey, M.J. Dunn, and G.Z. Yang, "The role of bioinformatics in two-dimensional gel electrophoresis," *Proteomics J.*, vol. 3, no. 8, pp. 1567–1596, 2003.
- [4] R.D. Appel, R. Vargas, P. Palagi, D. Walther, and De. Hochstrasser, "Melanie 7 II-a third generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms," *Electrophoresis*, vol. 8, no. 15, pp. 2735–2748, 1997.
- [5] 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, pp. 392–401, 2003.
- [6] L. Vincent L, and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.
- [7] Y. Kim, J. Kim, Y. Won, and Y. In, "Segmentation of protein spots in 2-D gel electrophoresis images with watershed using hierarchical threshold", *Computer and Information Sciences*, vol. 2869, pp. 389–396, 2003.
- [8] A dos Anjos, AL Møller, BK Ersbøll, C. Finnie, and H.R. Shahbazkia, "New approach for Segmentation and Quantification of Two-Dimensional Gel Electrophoresis Images," *Bioinformatics*, vol. 27, no.3, pp. 368–375, 2010.
- [9] M. Savelonas, E. Mylona, and D. Maroulis, "Unsupervised 2D Gel Electrophoresis Image Segmentation based on Active Contours," *Pattern Recogn.*, vol. 45, no. 12, pp. 720–731, 2012.
- [10] Melanie 7 [www.genebio.com]
- [11] Delta2D [www.decodon.com]
- [12] E. Kostopoulou, E. Zacharia and D. Maroulis, "Detection and segmentation in 2D gel electrophoresis images," in *Proc. 17th Int. Conf. on Digital Signal Processing*, Corfu, Greece, 2011.
- [13] E. Zacharia, E. Kostopoulou E, D. Maroulis, and S. Kossida, "A Spot Segmentation Approach for 2D Gel Electrophoresis Images Based on 2D Histograms", in *Proc. 20th Int. Conf. on Pattern Recognition*, pp. 2540–2543, Istanbul, Turkey, 2010.
- [14] R. Gonzalez, R. Woods, "Digital Image Processing," Prentice Hall, 3rd Edition, 2008.
- [15] P. Peer, LG Corzo, "Local Pixel Value Collection Algorithm for Spot Segmentation in Two-Dimensional Gel Electrophoresis Research," *Comparative and Functional Genomics*, 2007.
- [16] M. Fradkin, C. Ciofolo, B. Mory, G.Hautvast, and M. Breeuwer, "Comprehensive segmentation of cine cardiac MR images," *Medical Image Computing and Computer-Assisted Intervention*, pp. 178–185, 2008.