



ELSEVIER

journal homepage: [www.intl.elsevierhealth.com/journals/cmpb](http://www.intl.elsevierhealth.com/journals/cmpb)

## 2D-gel spot detection and segmentation based on modified image-aware grow-cut and regional intensity information



E. Kostopoulou\*, S. Katsigiannis, D. Maroulis

Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, 15784 Panepistimioupolis, Athens, Greece

### ARTICLE INFO

#### Article history:

Received 7 January 2015

Received in revised form

15 June 2015

Accepted 16 June 2015

#### Keywords:

Segmentation

Proteomics

2D-gel electrophoresis

### ABSTRACT

**Background:** Proteomics, the study of proteomes, has been increasingly utilized in a wide variety of biological problems. The Two-Dimensional Gel Electrophoresis (2D-PAGE) technique is a powerful proteomics technique aiming at separation of the complex protein mixtures. Spot detection and segmentation are fundamental components of 2D-gel image analysis but remain arduous and difficult tasks. Several software packages and academic approaches are available for 2D-gel image spot detection and segmentation. Each one has its respective advantages and disadvantages and achieves a different level of success in dealing with the challenges of 2D-gel image analysis. A common characteristic of the available methods is their dependency on user intervention in order to achieve optimal results, a process that can lead to subjective and non-reproducible results. In this work, the authors propose a novel spot detection and segmentation methodology for 2D-gel images.

**Methods:** This work introduces a novel spot detection and spot segmentation methodology that is based on a multi-thresholding scheme applied on overlapping regions of the image, a custom grow-cut algorithm, a region growing scheme and morphological operators. The performance of the proposed methodology is evaluated on real as well as synthetic 2D-gel images using well established statistical measures, including precision, sensitivity, and their weighted measure, F-measure, as well as volumetric overlap, volumetric error and volumetric overlap error.

**Results:** Experimental results show that the proposed methodology outperforms state-of-the-art software packages and methods proposed in the literature and results in more plausible spot boundaries and more accurate segmentation. The proposed method achieved the highest F-measure (94.8%) for spot detection and the lowest volumetric overlap error (8.3%) for the segmentation process.

**Conclusions:** Evaluation against state-of-the-art 2D-gel image analysis software packages and techniques proposed in the literature, including Melanie 7, Delta2D, PDQuest and Scimo, demonstrates that the proposed approach outperforms the other methods evaluated in this work and constitutes an advantageous and reliable solution for 2D-gel image analysis.

© 2015 Elsevier Ireland Ltd. All rights reserved.

\* Corresponding author. Tel.: +30 210 7275317; fax: +30 210 72753333.

E-mail addresses: [ikostop@di.uoa.gr](mailto:ikostop@di.uoa.gr) (E. Kostopoulou), [stamos@di.uoa.gr](mailto:stamos@di.uoa.gr) (S. Katsigiannis), [dmaroulis@di.uoa.gr](mailto:dmaroulis@di.uoa.gr) (D. Maroulis).

<http://dx.doi.org/10.1016/j.cmpb.2015.06.007>

0169-2607/© 2015 Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

In recent years, proteomics, i.e. the study of proteomes under different conditions, has been increasingly utilized for revealing the complex processes of cells [1]. Some key opportunities offered by the field of proteomics are the evaluation of new drugs and the exploration of biological events [2–6]. The *Two-Dimensional Polyacrylamide Gel Electrophoresis* (2D-PAGE) technique is a powerful technique in proteomics aiming at protein separation and identification and has been widely used due to its ability to separate thousands of proteins on polyacrylamide gels. This is achieved using isoelectric focusing and sodium dodecyl sulfate polyacrylamide gel electrophoresis that allows for protein separation according to the differences in their net charge and their molecular mass [7–10]. Results are then visualized into a digital image that can contain thousands of protein spots. In a 2D-gel electrophoresis experiment the aim of 2D-gel image analysis is the rapid identification of: (a) proteins located on a single gel and (b) differentially expressed proteins between samples from a series of 2D-gels.

2D-gel image analysis can be summarized into four main stages: (1) spot detection, (2) spot segmentation, (3) spot quantification and (4) image alignment. The objective of the first three stages is to detect the number of pixels that belong to protein spots in order to quantify the protein expression levels. The fourth stage is performed in order to match the corresponding protein spots from different images. Two different workflows for 2D-gel image analysis can be followed [11,12]. In the straightforward workflow, spot detection and segmentation are performed prior to image-alignment [13,14]. In the workflow followed by the Delta2D [15] and Progenesis Samespots [16] software tools, image-alignment is applied prior to spot detection and segmentation. The characteristics of 2D-gel images make spot detection and segmentation a very challenging task. 2D-gel images may contain thousands of spots that exhibit great variety in intensity, size and shape. Spots can be so poorly contrasted that they are not clearly visible and adjacent spots can often be highly overlapped. Furthermore, the overall quality of these images suffers due to artifacts, inhomogeneous background and high levels of noise [17].

Due to the importance of 2D-gel image analysis, many commercial software solutions have been developed [12]: e.g. PDQuest [18], DeCyder 2D [19], Melanie 7 [20], ImageMaster 2D [21], Delta2D [15] and Progenesis Samespots [16]. Each software has its respective advantages and disadvantages and achieves a different level of success in dealing with the challenges of 2D-gel image spot detection and segmentation [12,22]. An important observation is that all these software tools are dependent on manual parameter tuning, thus requiring human intervention that can lead to subjective results. Moreover, the most common problems in their results are that they may: (1) segment overlapping spots as one single spot, (2) over-segment a single spot into more, (3) fail to detect some spots, (4) mistake artifacts for spots and (5) incorrectly approximate the spot boundaries. The accumulation of all these errors alters the extracted protein expression levels, thus leading to erroneous biological conclusions. This

problem is addressed by extensive manual editing of the results, which is a time-consuming process that requires an average of 1 to 4 man-hours per gel [23]. Human intervention leads to subjective and non-reproducible results and limits the throughput of the analysis process. As a result, the development of an automated 2D-gel image analysis method would be of utmost importance since it would allow for high throughput analysis of the expression levels of thousands of proteins, and would lead to safer, objective and reproducible biological conclusions.

Many researchers have tried to address the challenges of 2D-gel image analysis. Based on the assumption that the shape of protein spots follows a Gaussian distribution, Yoon et al. [24] proposed the Reversible Jump Markov Chain Monte Carlo (RJCMCMC) method for separating overlapping spots and enhancing the weaker spots. RJCMCMC can be very time consuming and due to the original assumption it underperforms in cases of images with many non-Gaussian shaped spots. Morris et al. [25] proposed “Pinnacle”, a method for spot detection and quantification. Pinnacle’s spot detection is performed on a denoised average image of a properly aligned 2D-gel image set and is achieved by detecting local minima (“pinnacles”) on the denoised average image and combining them within a defined proximity. Pinnacle’s main advantage lies on detecting overlapping spots [25]. Nevertheless, it occasionally results in detecting spurious spots, i.e. false positive spots [26]. Li et al. [27] proposed “RegStatGel”, a method that is also performed on an average image of a properly aligned 2D-gel set and incorporates the watershed algorithm [28] for spot segmentation. Its main drawback is that it underperforms in splitting overlapping spots [26]. Based on morphological operations, Mylona et al. [29] proposed a method for spot detection that achieves increased performance in detecting overlapping spots at the expense of occasionally missed spots. Dos Anjos et al. [14] proposed “Scimo”, a spot segmentation and quantification method that is also based on the watershed algorithm and achieves a more realistic estimation of close proximity spots, as well as partially overlapping non-saturated spots. However, the authors make the assumption that each basin of the watershed contains only one protein spot. This assumption is not always valid in case of major overlapping. An active contour-based method [13] has also been proposed for 2D-gel image segmentation but requires empirical adjustment for a large number of parameters for different image datasets, a process that is lengthy and time-consuming, as described in [13]. Kostopoulou et al. [30] proposed an effective spot detection and segmentation approach based on 2D histograms and 3D spot morphology. Although effective, this approach suffers from a high computational complexity. Zacharia et al. [31] proposed a spot detection approach based on multidirectional texture and spatial intensity information, but is not capable of segmentation.

In this paper, the authors present a novel approach for 2D-gel image spot detection and segmentation. The proposed detection approach is based on a multi-thresholding scheme applied on overlapping regions of the image. For the segmentation process, the proposed method combines a grow-cut algorithm [32] that utilizes a custom update rule, with a region growing approach and morphological operators. The proposed detection and segmentation methodology is evaluated against

**Table 1 – Algorithm steps for detection.**

<p>Split image into windows</p> <ul style="list-style-type: none"> <li>• Split image <math>I</math> into tiled windows <math>W'_i</math>, <math>i = 1, \dots, N</math> of size <math>d' \times d'</math></li> <li>• Split image <math>I</math> into overlapping windows <math>W_i</math>, <math>i = 1, \dots, N</math> of size <math>d \times d</math>, centered around <math>W'_i</math></li> </ul> <p>Thresholding of each window</p> <ul style="list-style-type: none"> <li>• For each overlapping window <math>W_i</math>: <ul style="list-style-type: none"> <li>• Calculate <math>M</math> thresholds <math>T_{i,j}</math>, <math>j = 1, \dots, l, \dots, h, \dots, M</math> by applying the multiple Otsu Thresholding method [33] to <math>W_i</math></li> </ul> </li> <li>• For each tiled window <math>W'_i</math>: <ul style="list-style-type: none"> <li>• Add the pixels of <math>W'_i</math> with intensity value <math>&lt;T_{i,l}</math> to the set of background pixels <math>B</math></li> <li>• Add the pixels of <math>W'_i</math> with intensity value <math>&gt;T_{i,h}</math> to the set of foreground pixels <math>F</math></li> </ul> </li> </ul> <p>Spot center candidates selection</p> <ul style="list-style-type: none"> <li>• Detect the <i>Local Intensity Maxima</i> of image <math>I</math></li> <li>• Create a list of <i>Spot Center Candidates</i> by selecting the pixels of <math>F</math> that correspond to <i>Local Intensity Maxima</i> of <math>I</math></li> </ul> <p>Elimination of spot center candidates</p> <ul style="list-style-type: none"> <li>• For each pixel <math>s</math> belonging to the <i>Spot Center Candidates</i>: <ul style="list-style-type: none"> <li>• Select the rectangular region <math>V_s</math> of size <math>a \times a</math> around <math>s</math></li> <li>• Apply the local thresholding technique [34] in <math>V_s</math> to determine which pixels belong to the background and the foreground</li> <li>• Remove <math>s</math> from the <i>Spot Center Candidates</i> if its Euclidean distance from the nearest pixel of <math>V_s</math> classified as background is <math>&lt;k</math></li> </ul> </li> </ul> <p>Merging of spot center candidates</p> <ul style="list-style-type: none"> <li>• For each pixel <math>s</math> belonging to the <i>Spot Center Candidates</i>: <ul style="list-style-type: none"> <li>• For each pixel <math>s'</math> (<math>s' \neq s</math>) belonging to the <i>Spot Center Candidates</i>: <ul style="list-style-type: none"> <li>• Calculate the Euclidean distance <math>\delta_{s,s'}</math> between <math>s</math> and <math>s'</math></li> <li>• If <math>\delta_{s,s'} &lt; k</math> then replace <math>s</math> and <math>s'</math> in the <i>Spot Center Candidates</i> list with their midpoint pixel</li> </ul> </li> </ul> </li> <li>• Return the remaining <i>Spot Center Candidates</i> as the final <i>Spot Centers</i></li> </ul>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

three commercial packages, Delta2D [15], PDQuest [18] and Melanie 7 [20] as well as against the recently proposed method: “Scimo” [14]. Two datasets of real and synthetic 2D-gel images containing a total of approximately 20,400 spots was used for the experimental evaluation. Experimental results demonstrate the efficiency of the proposed method in detecting and segmenting spots on 2D-gel images.

The rest of this paper is organized in three sections. A thorough description of the proposed methodology is provided in Section 2. The experimental evaluation procedure and results are presented in Section 3, whereas conclusions are drawn in Section 4.

## 2. Method

The proposed approach is divided in two stages: (1) spot detection and (2) spot segmentation.

### 2.1. Spot detection

The spot detection algorithm’s steps are summarized in Table 1.

#### 2.1.1. Detection of spot center candidates

The initial step for spot center detection is the exclusion of image areas that are not likely to include spot centers.

In order to address the problem of uneven intensity values among spot and background regions and the insufficiency of a global threshold to detect regions containing spot centers, the image  $I$  is processed locally. The algorithm starts by splitting  $I$  into overlapping windows  $W_i$ , and tiled windows  $W'_i$ ,  $i = 1, \dots, N$  of size  $d \times d$  and  $d' \times d'$ , respectively, as illustrated in Fig. 1a. Then  $T_{i,j} = 1, \dots, l, \dots, h, \dots, M$  thresholds are obtained automatically for each  $W_i$  by using the multiple Otsu thresholding technique [33]. The pixels inside  $W'_i$  are divided into clusters based on the automatically estimated local thresholds. More specifically, two local thresholds  $T_{i,l}$  and  $T_{i,h}$  ( $T_{i,l} < T_{i,h} < T_{i,M}$ ) are introduced for each  $W_i$ : A low threshold  $T_{i,l}$  such that pixels whose intensity values are less than  $T_{i,l}$  are classified as belonging to the background ( $B_i$ ) and a high threshold  $T_{i,h}$  such that pixels whose intensity values are larger than  $T_{i,h}$  are classified as pixels belonging to the foreground ( $F_i$ ).  $B_i$  and  $F_i$  are sets containing the background and the spot (foreground) pixels, respectively. The pixels with intensity between the two thresholds may belong to either spot or background regions. An example tiled window of  $I$  is illustrated in Fig. 1b whereas the result of clustering with the  $M + 1$  thresholds and the two thresholds ( $T_{i,l}, T_{i,h}$ ) is illustrated in Fig. 1c and d, respectively. This overlapping-tiled framework minimizes the likelihood of selecting wrong thresholds as it takes into consideration the intensities of pixels located around each tiled window  $W'_i$ , and enables the characterization of pixels that belong to more than one overlapping window.

#### 2.1.2. Spot center selection

The pixels belonging to the set  $U_{i=1}^N F_i$ ,  $i = 1, 2, \dots, N$ , that correspond to local intensity maxima, are selected as candidate spot centers. An example of candidate spot centers is shown in Fig. 2a with black color. As one may observe, some spots may contain multiple candidate spot centers that are located either near their center or near their boundary. In order to select one pixel among the candidates as spot center, the candidate spot centers are (a) filtered by spatial intensity information and (b) merged depending on their distance.

Let  $s$  be a spot center candidate,  $V_s$  a small rectangular region of size  $a \times a$  around it and  $BN_s$  the binary image resulting from the application of a local thresholding technique [34] in  $V_s$ . If the Euclidean distance among  $s$  and its nearest pixel classified as background in  $BN_s$  is less than  $k$  then  $s$  is excluded from the spot center candidates. Then the Euclidean distance among spot center candidates is computed and those with a pairwise proximity less than the minimum allowed distance  $k$  are merged. If the minimum allowed distance  $k$  is set too small then several spot centers might be detected on a single spot. On the other hand, if  $k$  is set too large, then overlapping spots might not be separated. An example of the set  $S_i$  of spot centers for a window  $W_i$  after the two elimination procedures is illustrated in Fig. 2b. The final set  $S$  containing all the spot centers of the image is defined as  $S = U_{i=1}^N S_i$ ,  $i = 1, 2, \dots, N$ .

Typical values for the parameters of the spot detection algorithm are presented in Section 3.1. Those parameter values were experimentally determined using various 2D-GE image datasets and were kept constant in all our experiments. As it is demonstrated in the results section, these parameter

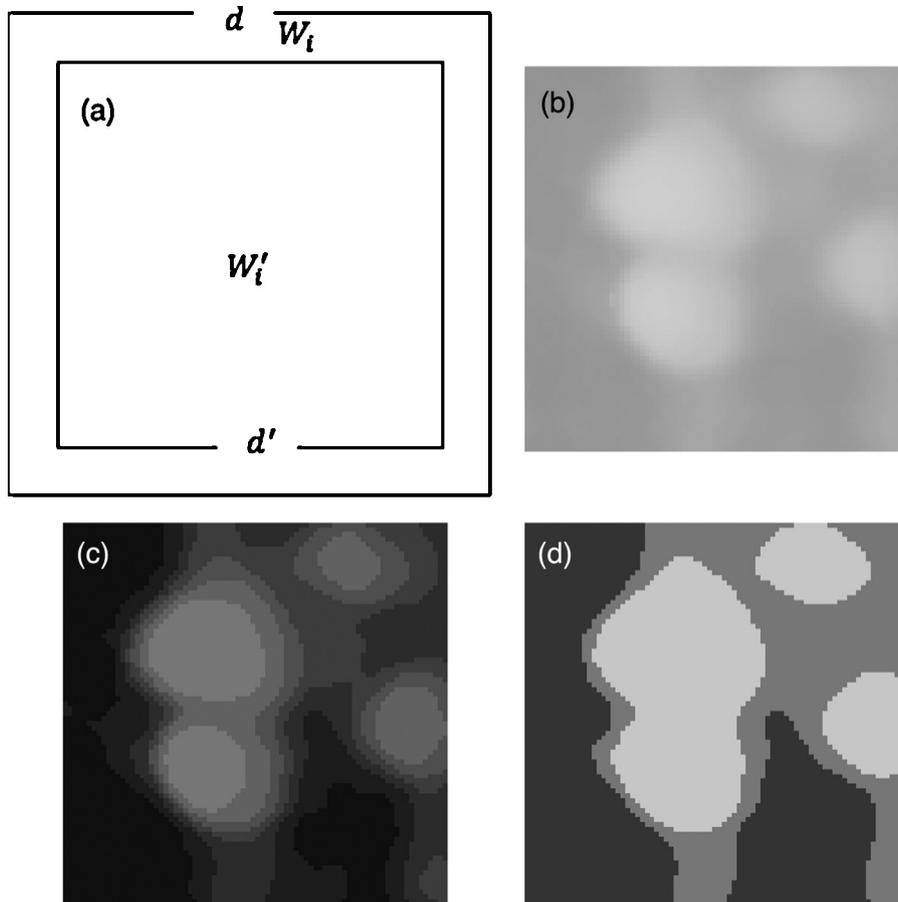


Fig. 1 – (a) Overlapping and tiled windows  $W_i$ ,  $W'_i$ , (b) region of a real 2D-gel image, (c) region (b) after classification in  $(M + 1)$  classes, (d) foreground, background and middle intensity pixels of (b).

values efficiently work on 2D-GE images from a wide variety of sources. The value of  $d$  is dependent on  $d'$  and was experimentally determined as  $d = d' + 2 \cdot \lceil d'/8 \rceil$ . In effect, the larger the tiled windows are, the larger the overlapping windows around them will be. The number of tiled windows  $N$  for an image of size  $\text{Width} \times \text{Height}$  can be determined

from  $d'$  as  $N = \lceil \text{Width}/d' \rceil \cdot \lceil \text{Height}/d' \rceil$ . Moreover, an evaluation of the sensitivity of the proposed method on the number of overlapping windows showed that the proposed approach exhibits significant performance stability to variations of the size of the overlapping windows, as explained in Section 3.5.

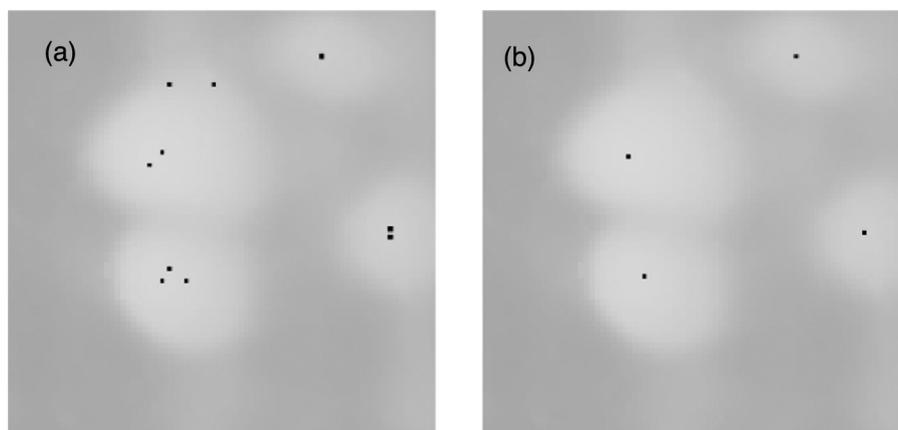


Fig. 2 – Candidate spot centers (a) before and (b) after elimination.

**Table 2 – Algorithm steps for segmentation.**

Rough segmentation using grow-cut

- Set pixels belonging to  $B$  as background seeds
- Set pixels belonging to  $F$  as foreground seeds
- Apply the modified grow-cut algorithm to segment the image  $I$  into foreground and background pixels using the seeds above

Creation and dilation of spot regions

- Select the regions  $R_{\alpha}$  consisting of the pixels labeled as foreground by the grow-cut algorithm
- For each region  $R_{\alpha}$ :
  - Determine the set  $K$  of pixels that are spot centers and have been labeled as foreground by grow-cut, and label them as belonging to a different region  $R_{\alpha,b}$ ,  $b = 1, 2, \dots$
  - Determine the set  $L$  of pixels that are not spot centers and have been labeled as foreground by grow-cut, and label them as belonging to an undefined region
  - Sort the pixels of  $L$  into descending intensity value order
  - Separate the pixels of  $L$  into groups  $GP_e$ ,  $e = 1, \dots, z$  of  $x$  distinct intensity values
  - For each group of pixels  $GP_e$ :
    - Select the pixels that belong to an undefined region
    - Assign them to the region  $R_{\alpha,b}$  of the majority of their neighbors (neighbors belonging to an undefined region are not considered)
  - Apply the dilation operator with the disk of radius  $r$  as the structuring element to each region  $R_{\alpha,b}$  in order to create the regions  $DR_{\alpha,b}$
  - From each region  $DR_{\alpha,b}$  remove the pixels that belong to other regions  $DR_{\alpha,b'}$ ,  $b \neq b'$

Precise segmentation

- Compute the gradient  $GI$  of image  $I$
- For each region  $DR_{\alpha,b}$ :
  - Compute the *Intensity Threshold* by applying the optimal thresholding technique [34] to image  $I$
  - Compute the *Gradient Intensity Threshold* by applying the optimal thresholding technique [34] to gradient image  $GI$ 
    - The pixels of  $DR_{\alpha,b}$  with intensity value  $>$  *Intensity Threshold* are labeled as spot pixels
    - The pixels of  $DR_{\alpha,b}$  with gradient intensity value  $>$  *Gradient Intensity Threshold* are labeled as spot pixels

## 2.2. Spot segmentation

The steps of the segmentation algorithm that are described in the following paragraphs are summarized in Table 2.

### 2.2.1. Seed pixels selection

For the segmentation process, the proposed methodology requires the selection of seed pixels that are predefined as spot or background pixels. Let  $B$  and  $F$  be the sets containing the background and spot seed pixels of image  $I$ , respectively.  $B$  and  $F$  are determined from the previous multithresholding step using the following equations:

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

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

where  $I(p)$  is the intensity value of pixel  $p$ ,  $i = 1, \dots, N$  and  $h = 2-l$ .

### 2.2.2. The original grow-cut algorithm

The next step of the proposed methodology is the application of the grow-cut [32] segmentation algorithm. Grow-cut [32] is a segmentation algorithm based on Cellular Automata

[35], that can segment grayscale or multichannel images into regions belonging to multiple classes. The grow-cut algorithm is a relatively simple algorithm that can efficiently segment a wide range of images and can be initialized with a large number of seed pixels. Moreover, it has been successfully utilized in a wide variety of applications including brain tumor segmentation [36], IVUS lumen segmentation [37], region-based medical image retrieval [38], tumor segmentation [39], cDNA microarray image segmentation [40], etc.

A triplet  $(l_p, \theta_p, \vec{C}_p)$  represents each pixel  $p$  of the image, where  $l_p$  is the class label of  $p$ ,  $\theta_p$  takes values inside  $[0, 1]$  and is a measure of certainty that  $p$  should be labeled as  $l_p$ , and  $\vec{C}_p$  is a multidimensional vector containing the intensity of pixel  $p$  at each channel for multichannel images or  $\vec{C}_p$  is equal to the intensity of  $p$  in case of grayscale images. A  $\theta_p$  equal to 1 denotes that the label assigned to  $p$  cannot be altered, whereas  $\theta_p < 1$  indicates that  $l_p$  may change during the progress of the algorithm. At the end of the grow-cut algorithm, all the pixels of the image should be assigned with one of the possible labels. Initially,  $l_p$  for all pixels is set to “Undefined” and  $\theta_p$  to 0. The user provides the initial pixel seeds by setting their  $l_p$  and  $\theta_p$ . Then, the following steps are iterated until a label has been assigned to all pixels:  $g(q_{an})$  is computed for each pixel  $p$  and its adjacent neighbors  $q_{an}$  ( $an = 1, \dots, 8$ ).  $g(q_{an})$  is a function bounded to  $[0,1]$ , defined as:

$$g(q_{an}) = 1 - \frac{\|\vec{C}_p - \vec{C}_{q_{an}}\|_2}{\max\|\vec{C}\|_2} \quad (3)$$

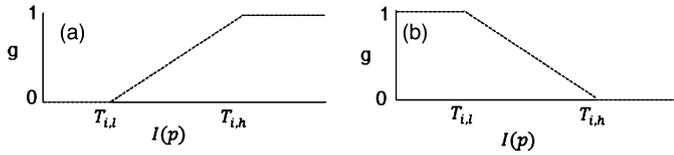
where  $\max\|\vec{C}\|_2$  is the maximum value an element of  $\vec{C}_p$  can take. For example, for an image with three channels and  $L$  bit depth per channel,  $\max\|\vec{C}\|_2 = 2^L - 1$ . Then,  $\lambda(q_{an})$  is computed for all  $q_{an}$  with  $l_{q_{an}} \neq$  “Undefined” and is defined as:

$$\lambda(q_{an}) = g(q_{an}) \cdot \theta(q_{an}) \quad (4)$$

If  $\lambda(q_{an}) > \theta_p$ , then  $p$  receives the label of  $q_{an}$  ( $l_p = l_{q_{an}}$ ) and  $\theta_p$  becomes equal to  $\lambda(q_{an})$  ( $\theta_p = \lambda(q_{an})$ ). After a number of iterations, pixel labels cease to change and the algorithm reaches a state of convergence where all pixels have been assigned with a label.

### 2.2.3. Proposed application of the grow-cut algorithm

The original grow-cut algorithm when applied to 2D-gel images results in image boundaries that do not contain the whole object of interest, i.e. the spot pixels in our case. The reason for this lies in the fact that spot pixels near the spot boundary have intensity values similar to the background. To effectively segment the 2D-gel images, a modified grow-cut update rule that takes into consideration the characteristics of 2D-gel images is proposed. The grow-cut algorithm is applied to  $I$  in order to label the pixels of the image  $I$  as spot (foreground  $-f$ ) or background ( $b$ ). For the proposed methodology, seed pixels are estimated automatically as explained in Section 2.2.1. The use of automatically selected seed pixels enhances the reproducibility of the segmentation results and reduces user intervention. The use of the original  $g$  function for the grow-cut algorithm leads to the expansion of background seeds over spot boundaries due to the minimal difference in their



**Fig. 3** –  $g(q,p)$  function when  $l(q)$  is (a) spot and (b) background.

intensity values that result to high  $g$  values. To address this problem, the  $g$  function has been modified in order to take into consideration both the difference in intensity between  $P$  and  $q$ , and the class  $q$  belongs to. The incentive behind this modified  $g$  function is to enhance or reduce the strength of an attacking pixel based on the intensity difference between the pixel and the seed pixels of the class the attacking pixel belongs to. If the pixel  $q$  is labeled as spot, then if  $I(p)$  is close to  $T_{i,h}$  then the attack force is enhanced. Otherwise, the attack force is diminished. On the other hand, if the attacker pixel  $q$  is labeled as background, then if  $I(p)$  is close to  $T_{i,l}$  then the attack force is enhanced. Otherwise, the attack force is diminished. Eq. (5) shows the modified  $g$  function, while the plots for  $g$  are illustrated in Fig. 3.

$$g(q,p) = \begin{cases} 0, & p,q : ([l(q) = f] \wedge [I(p) < T_{i,l}]) \vee ([l(q) = b] \wedge [I(p) > T_{i,h}]) \\ 1, & p,q : ([l(q) = f] \wedge [I(p) > T_{i,h}]) \vee ([l(q) = b] \wedge [I(p) < T_{i,l}]) \\ 1 - \frac{T_{i,h} - I(p)}{T_{i,h} - T_{i,l}}, & p,q : [l(q) = f] \wedge [T_{i,l} \leq I(p) \leq T_{i,h}] \\ \frac{T_{i,h} - I(p)}{T_{i,h} - T_{i,l}}, & p,q : [l(q) = b] \wedge [T_{i,l} \leq I(p) \leq T_{i,h}] \end{cases} \quad (5)$$

#### 2.2.4. Post grow-cut segmentation

After the application of the grow-cut algorithm, each region  $R_a$  formed by pixels labeled as spot pixels, is further processed in order to segment it in the spot regions  $R_{a,b}$  it contains, as shown in Fig. 4a and b. The pixels belonging to the subset  $K$ , with  $K = \{p : (p \in S) \wedge (p \in R_a)\}$ , with  $S$  being the set containing the spot centers of the image, are selected as seed points for spot regions inside  $R_a$ , with each seed pixel being the seed for a different  $R_{a,b}$  region. The rest of the pixels  $L = \{p : (p \notin K) \wedge (p \in R_a)\}$  are labeled based on their neighborhood pixel labels by the following procedure: At each iteration of the procedure, the unique intensities  $V_a$  of the non-seed pixels of  $L$  are determined in descending intensity order and then groups of similar intensity values – each containing  $x$  distinct intensities – are formed. Let  $u$  be the number of different intensities of  $L$  and  $z = u/x$  be the number of groups. Each group of pixels  $GP_e$ ,  $e = 1, \dots, z$ , contains the pixels of  $L$  whose intensities are within the range of  $[V_{x,(e-1)+1}, V_{x,e}]$ . The pixels of each  $GP_e$  are then labeled based on their neighborhood pixels. For each pixel  $p$ , its neighborhood pixels are examined and  $p$  is labeled by the application of a majority voting criterion on the labels of its neighbors. Neighbors with undefined labels are not considered for the majority voting process. If a pixel is not assigned with a label then it will be reexamined at the next iteration of the procedure, whereas the assignment of a label to a pixel is considered as final and the pixel will not be reexamined. The processing of pixels in groups of intensities enables the

speedup of the algorithm, since pixels with similar high intensity values are examined first and then impose their label on the neighborhood pixels of lower intensity values. As a result, the algorithm can cope with fluctuations in the intensity of the image that may be caused by noise or uneven background. An example of the application of the majority voting process is shown in Fig. 4b.

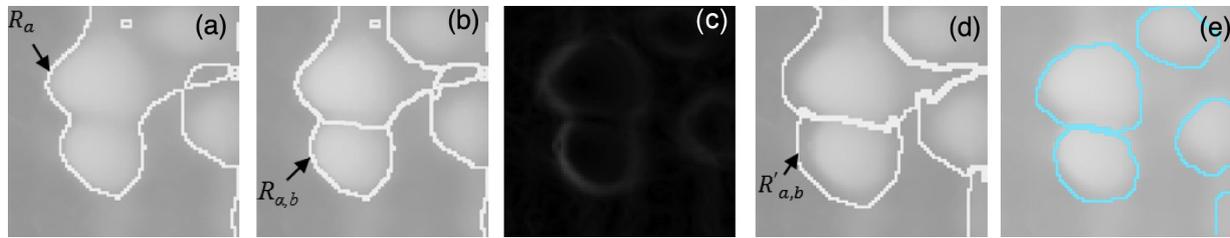
As it can be observed, in some cases the  $R_{a,b}$  regions may contain background pixels, especially in the case of clusters of spots. These pixels are then excluded by applying the optimal thresholding technique [34] in each region  $R'_{a,b}$ , which corresponds to an extended region of  $R_{a,b}$  (see Fig. 4d) and 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 with structuring element  $D$  ( $D =$  disk of radius  $r$  in this work) on  $R_{a,b}$  and  $R'_{a,b}$  is a region other than  $R_{a,b}$ . The optimal thresholding technique is applied twice: (a) once on the gradient intensity values of pixels belonging to  $R'_{a,b}$  in order to detect the pixels near the spot boundary ( $R'_{a,o}$ ), whose gradient intensity value is high (see Fig. 4c) and (b) on the intensity values so as to detect the pixels far from the spot boundary ( $R'_{a,i}$ ) and whose intensity value is high (see Fig. 4d). The final spot pixels are the pixels belonging to the set FS, with

$FS = \{p : (p \in R'_{a,o}) \vee (p \in R'_{a,i})\}$ . Fig. 4e depicts an example of the final segmentation results for the region depicted in Fig. 1b.

## 3. Experimental evaluation

### 3.1. Experimental setting

The proposed algorithm has been evaluated on two datasets consisting of real (Re) and synthetic (Sy) 2D-gel images, respectively, in order to allow for a qualitative and quantitative comparison with three established commercial packages (Delta2D [15], Melanie 7 [20] and PDQuest [18]) as well as the recently published method Scimo [14]. Each dataset (Re, Sy) consists of 2D-gel images containing a total of  $\sim 10,200$  spots and the bit depth of the images for both datasets is 16 bits. The real dataset has been provided through the courtesy of the Biomedical Research Foundation of the Academy of Athens (BRFAA) [41] whereas the synthetic 2D-gel images have been generated by our research group. The latter images have been created by combining spots produced using the 2-D Gaussian flat top function with background extracted from real images in order to imitate the real ones more accurately and retain their characteristics, i.e. overlapping spots, poorly contrasted spots, inhomogeneous background, artifacts and streaks. An example of a real and a synthetic image from the datasets utilized in the evaluation is shown in Fig. 5. It should



**Fig. 4** – (a) Spot regions after application of the modified grow-cut, and (b) rough segmentation (c) gradient, (d) dilation of spot regions, (e) precise segmentation.

**Table 3** – Parameter values.

$d = 100$	$k = 4$
$d' = 80$	$a = 31$
$M = 6$	$x = 10$
$l = 2$	$D = \text{disk with radius } 5$
$h = 4$	

be noted that the parameters needed for the proposed algorithm were experimentally adjusted once and then remained constant during all experiments performed on both datasets, in contrast with the parameters of the state-of-art programs which need tuning – by expert biologists – for each particular image. The selected parameters are expected to provide satisfactory performance for 2D-gel images obtained from different sources and are shown on Table 3. In particular, the constants were chosen in order to best detect the real spots, split overlapping spots, avoid detection of spurious spots and create more accurate spot boundaries that do not include background pixels. The parameters for the software packages compared with the proposed method were determined by expert biologists from the Biomedical Research Foundation of the Academy of Athens (BRFAA) [41] for the Melanie 7, Scimo and PDQuest, and from the Institute of Molecular Biology and Biotechnology (IMBB) of the Foundation for Research and Technology-Hellas (FORTH) [42] for Delta2D.

### 3.2. Real images

For spot detection, the ground truth information for the real images was provided by expert biologists of the BRFAA, who manually determined the locations of protein spots and marked them by drawing a cross inside each unique spot region. Three ground truth replicates were created for each

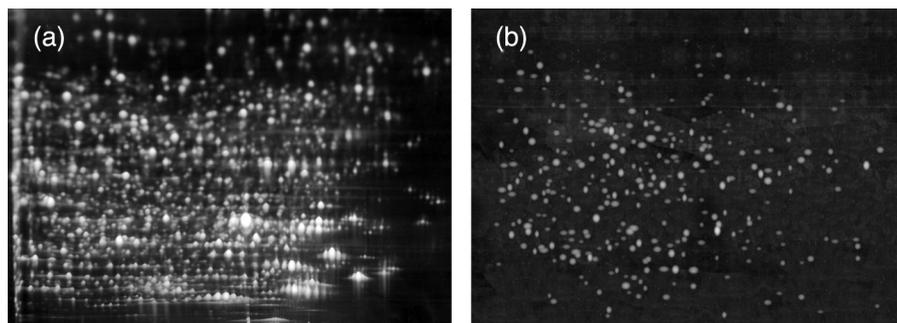
image and the final ground truth was determined by majority voting. Comparing the detection results with the corresponding ground truth allows the characterization of each spot as “correctly detected” (true positive – TP), “not detected” (false negative – FN) and “falsely detected” as spot (false positive – FP). Then, a quantifiable comparison with the other methods is provided by means of the precision (P), sensitivity (S), and weighted harmonic mean (F-measure) measures [43–45], which are defined as:

$$P = \frac{TP}{TP + FP} \quad (6)$$

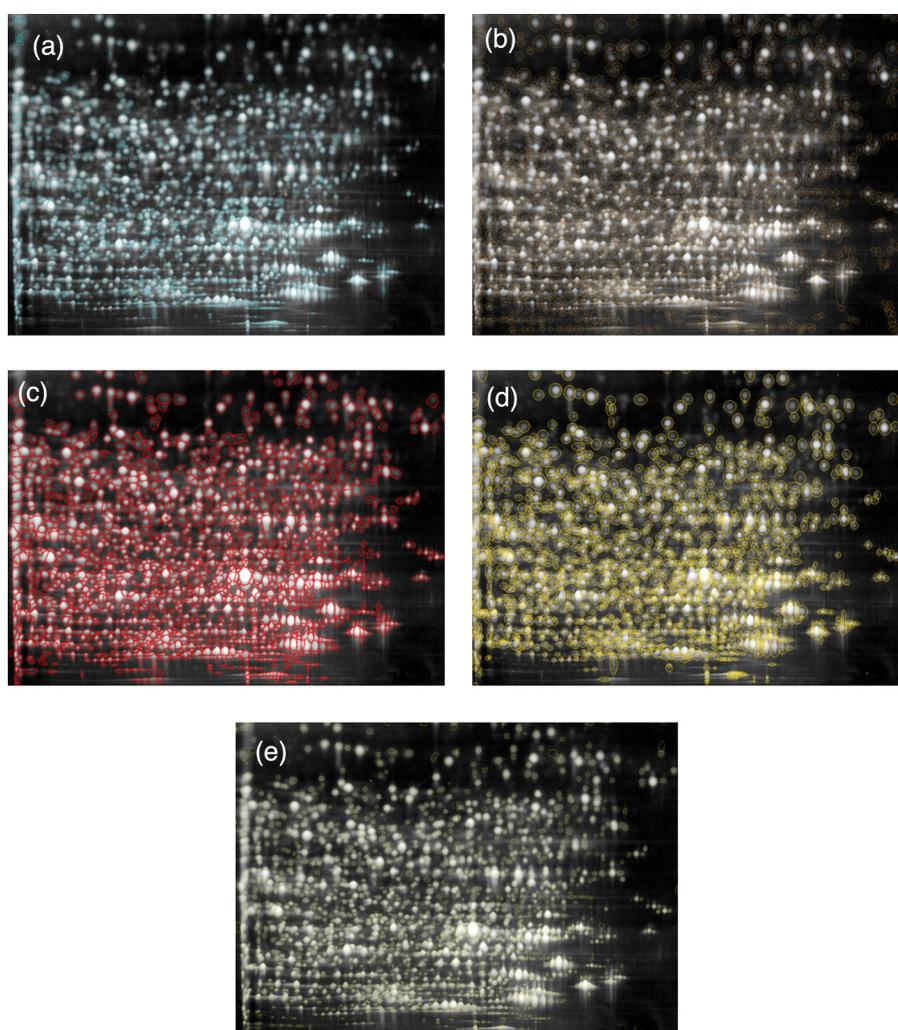
$$S = \frac{TP}{TP + FN} \quad (7)$$

$$F\text{-measure} = 2 * \frac{P * S}{P + S} \quad (8)$$

The ground truth information for the segmentation of the real images cannot be available, thus spot boundaries are evaluated through visual inspection. Fig. 6 and Fig. 7 present the segmentation results obtained by the proposed approach and four other commercial packages and published methods for a real 2D-gel image as well as for magnified regions of this image. It is evident that the proposed approach (Fig. 6a and Fig. 7a<sub>1–3</sub>) is more successful in detecting and segmenting the protein spots than all the three image analysis software packages, namely Delta2D (Fig. 6b and Fig. 7b<sub>1–3</sub>), Melanie 7 (Fig. 6c and Fig. 7c<sub>1–3</sub>) and PDQuest (Fig. 6d and Fig. 7d<sub>1–3</sub>), as well as the recently published approach Scimo (Fig. 6e and Fig. 7e<sub>1–3</sub>). Moreover, the proposed approach provides more accurate and smoother spot boundaries as shown in Fig. 6a and Fig. 7a<sub>1–3</sub>.



**Fig. 5** – 2D-gel images from (a) real and (b) synthetic datasets utilized in the experiments.



**Fig. 6** – Segmentation results for the real 2D-gel image of Fig. 5a obtained by (a) the proposed approach, (b) Delta2D, (c) Melanie 7, (d) PDQuest, and (e) Scimo.

In Fig. 7a<sub>1</sub>–e<sub>3</sub>, TPs are depicted as spots with a cross and boundaries, FNs as spots with a cross and no boundaries, and FPs as spots with boundaries but no cross inside. The sensitivity and precision values are both important, complementary measures for the detection performance. However, the F-measure can be considered as a more reliable measure, as it takes into account both the number of detected protein spots as well as the number of falsely detected (spurious) spots. Table 4 presents the statistical results obtained by the proposed approach and the four other methods, using the real 2D-gel image dataset. The effectiveness of the proposed approach is proved by the high precision (96.5%), sensitivity (93.2%), and F-measure (94.8%)

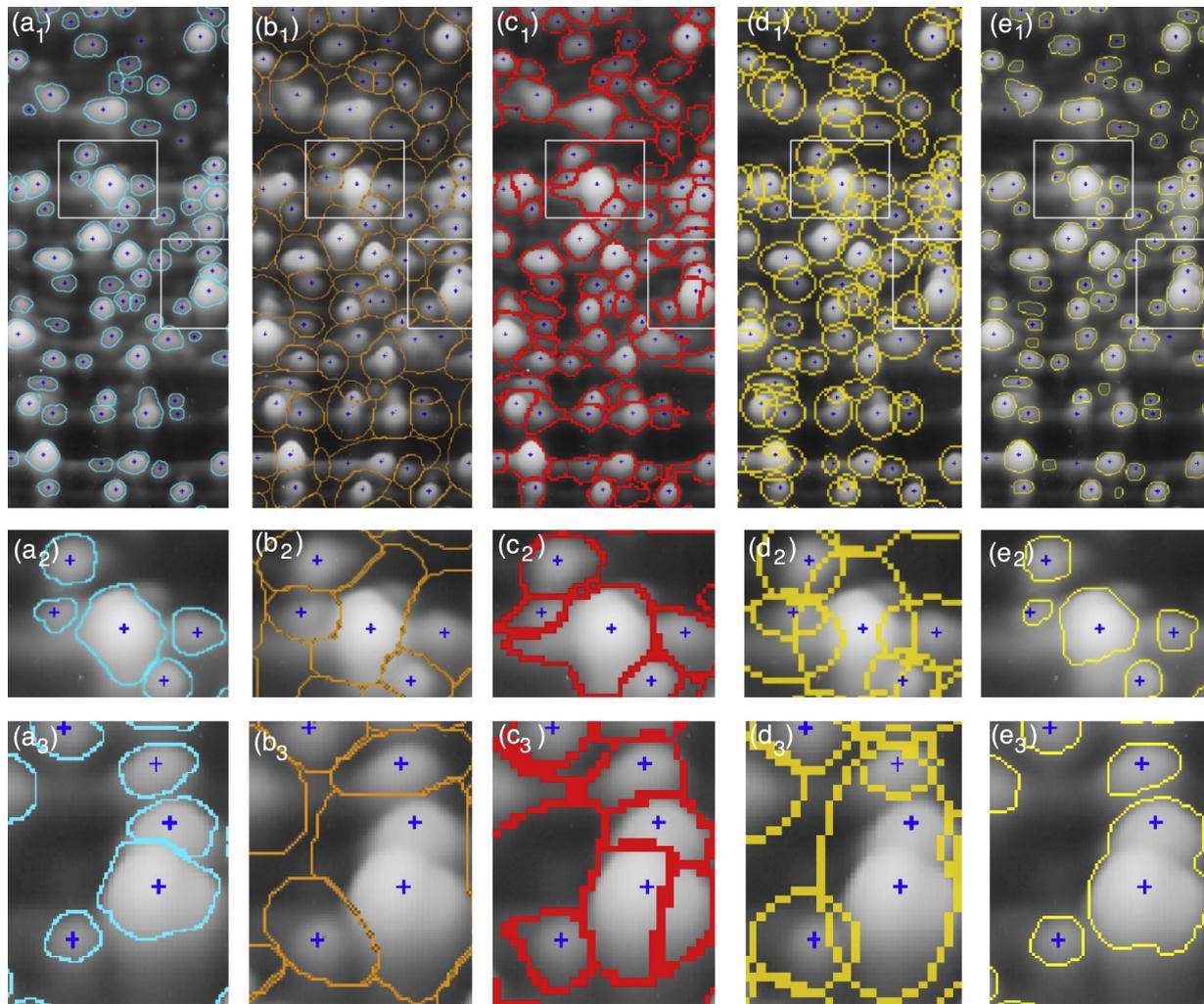
value it achieves. The better spot detection performance of the proposed approach can lead to more accurate overall results since it results to a larger number of correctly detected spots, less false-positive (spurious) spots, and can assist the segmentation algorithm for the separation of overlapping spots.

### 3.3. Synthetic images

Fig. 8 presents the segmentation results obtained by the proposed algorithm and the four other methods for the synthetic image illustrated in Fig. 5b. More detailed regions of these images are shown in Fig. 9. From these images it can be

**Table 4** – Comparison of detection results on real 2D-gel images.

Methodology	Precision (%)	Sensitivity (%)	F-measure (%)
Proposed approach	96.5 ± 2.9	93.2 ± 2.6	94.8 ± 2.8
Delta2D	75.7 ± 15.9	78.1 ± 6.2	76.9 ± 10.6
Melanie 7	85.2 ± 2.8	94.1 ± 2.4	89.4 ± 2.4
PDQuest	86.3 ± 2.9	93.9 ± 3.1	89.9 ± 3.0
Scimo	91.9 ± 3.7	76.4 ± 3.6	83.4 ± 3.3



**Fig. 7 – Segmentation results for a region of a real 2D-gel image obtained by (a<sub>1-3</sub>) the proposed approach, (b<sub>1-3</sub>) Delta2D, (c<sub>1-3</sub>) Melanie 7, (d<sub>1-3</sub>) PDQuest, and (e<sub>1-3</sub>) Scimo. (x<sub>2</sub>) and (x<sub>3</sub>) depict magnified versions of the annotated regions of image (x<sub>1</sub>), with x = a, . . . , e.**

easily derived that the proposed approach (Fig. 8a and Fig. 9a<sub>1-3</sub>) is more efficient in spot segmentation than the other methods, namely Delta2D (Fig. 8b and Fig. 9b<sub>1-3</sub>), Melanie 7 (Fig. 8c and Fig. 9c<sub>1-3</sub>), PDQuest (Fig. 8d and Fig. 9d<sub>1-3</sub>), and Scimo (Fig. 8e and Fig. 9e<sub>1-3</sub>), as it creates more plausible boundaries, separates overlapping spots, and includes almost all spot pixels inside the created spot boundaries while avoiding the inclusion of background pixels. It is evident that the proposed approach can effectively segment the protein spots and provide spot boundaries that look plausible and do not contain background pixels. The efficiency of the proposed method is also confirmed by the following statistical evaluation.

Since the ground truth information for the synthetic dataset is available, by comparing it with the segmentation results, each pixel can be characterized as “Actual Spot” (AS) pixel, “False Spot” (FS) pixel, “Actual Background” (AB) pixel and “False Background” (FB) pixel. This pixel characterization is then utilized in order to provide a quantifiable comparison between the examined methods through the Volumetric

Overlap (VO), Volumetric Error (VE), and Volumetric Overlap Error (VOE) measures as proposed in Ref. [46], which are defined as:

$$VO = \frac{ASV}{ASV + FBV} \quad (9)$$

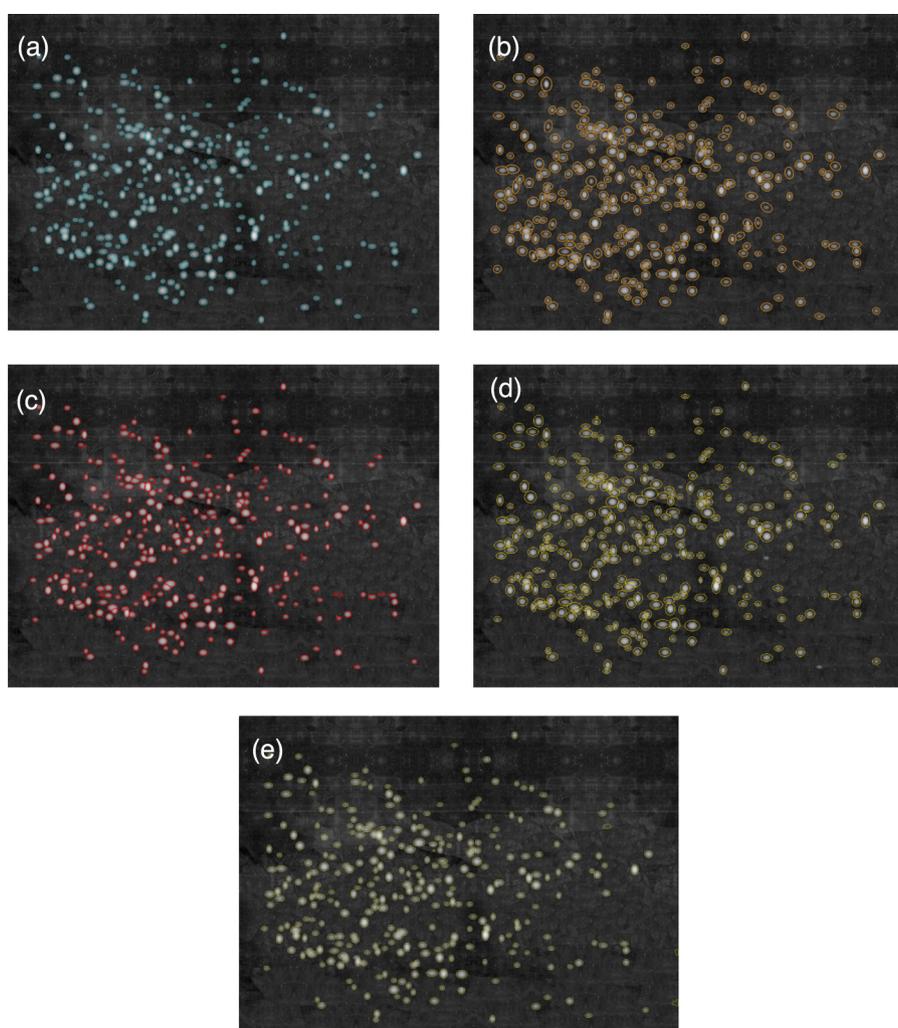
$$VE = \frac{FSV}{ASV + FBV} \quad (10)$$

$$VOE = 1 - \frac{ASV}{ASV + FBV + FSV} \quad (11)$$

where ASV, FBV, and FSV denote the volume of each pixel group defined as:

$$V = \sum_{x,y} I(x,y) \quad (12)$$

Table 5 and Fig. 10 present the results obtained by the proposed approach and the four other methods, using the



**Fig. 8 – Segmentation results for the synthetic 2D-gel image depicted in Fig. 5b obtained by (a) the proposed approach, (b) Delta2D, (c) Melanie 7, (d) PDQuest, and (e) Scimo.**

synthetic 2D-gel image dataset. Based on these results, it is evident that the proposed approach outperforms the four compared software packages and techniques for the synthetic image dataset. In particular, it achieves a very high VO value (97.0%) while having a very low VE value (5.7%). Additionally, it achieves the lowest VOE value (8.3%) which shows that the proposed approach outperforms the compared techniques. The VO, VE, VOE values indicate that the proposed approach manages to: (a) segment the majority of spot pixels (high VO), (b) avoid segmenting background pixels as spot pixels (low VE)

and achieves the best balance between VO and VE among the compared software packages and techniques (having the lowest VOE). The latter remark is significant since both high VO and low VE minimize the effort and time needed for validation and correction of the results by expert biologists. While most methods achieve slightly higher or comparable VO, they perform much worse in terms of VE, leading to worse VOE than the proposed method. Delta2D, Melanie 7, PDQuest and Scimo achieve elevated VO because they extend the spot boundaries outside of the spot edges, resulting to the inclusion of background pixels along the spot pixels, that as a result leads also to high VE.

Considering the high VO, low VE and low VOE achieved by the proposed approach, it is evident that it misses only few spot pixels (high VO), while falsely segmenting only few background pixels as foreground pixels (low number of FS pixels, low VE). The low VOE achieved by the proposed approach indicates its superiority over the compared software packages and techniques. For example, PDQuest achieved a slightly higher VO value (99.4% compared to 97.0% for the proposed method) but has simultaneously segmented as spot pixels a much

**Table 5 – Comparison of segmentation results on synthetic 2D-gel images.**

Methodology	VO (%)	VE (%)	VOE (%)
Proposed approach	97.0 ± 1.8	5.7 ± 1.6	8.3 ± 0.5
Delta2D	99.9 ± 0.1	68.0 ± 20.4	43.0 ± 8.3
Melanie 7	98.1 ± 1.8	25.6 ± 9.8	21.6 ± 5.0
PDQuest	99.4 ± 0.8	82.1 ± 23.5	44.8 ± 6.1
Scimo	98.6 ± 0.7	12.9 ± 2.6	12.6 ± 1.5

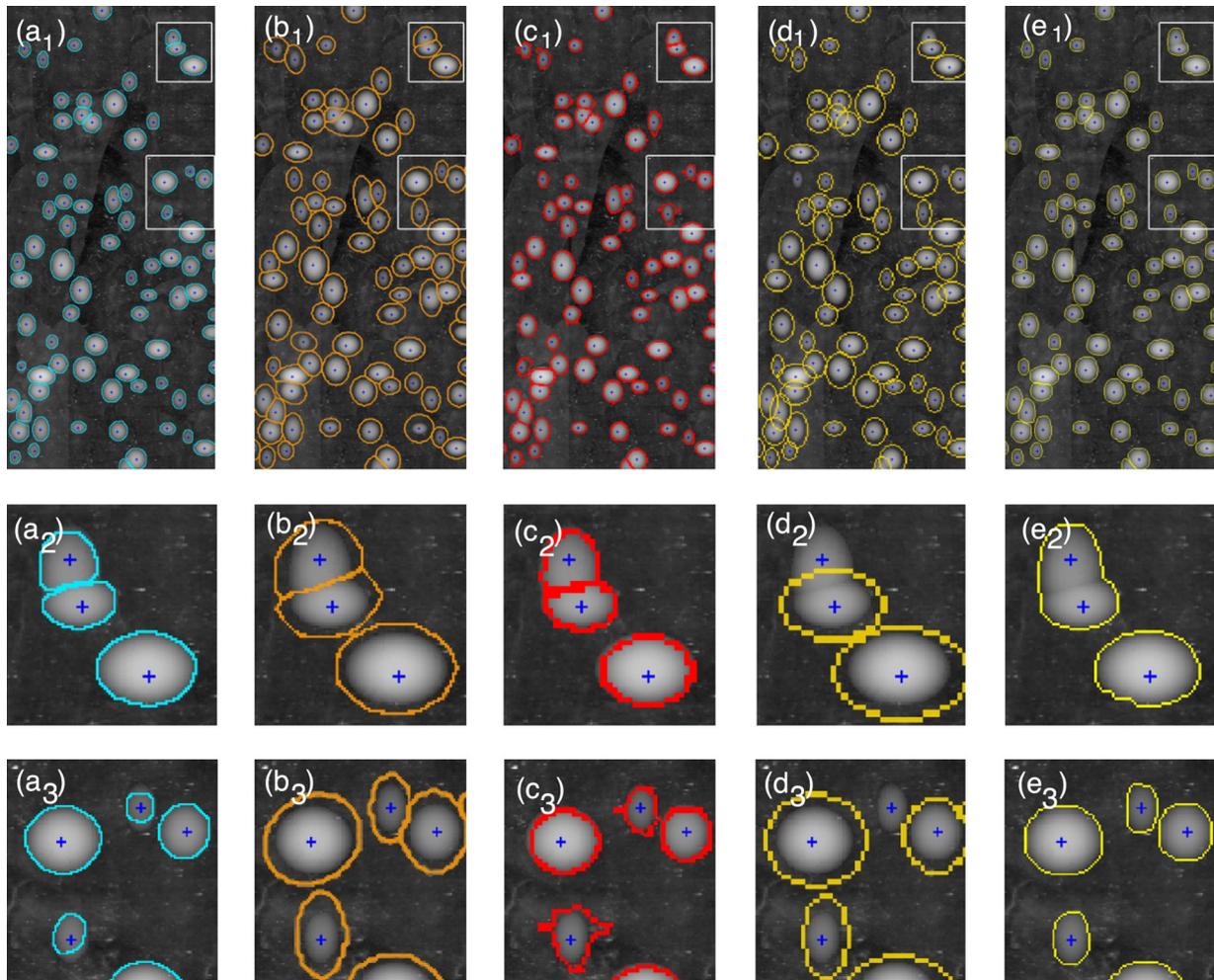


Fig. 9 – Segmentation results for a region of a synthetic 2D-gel image obtained by (a<sub>1-3</sub>) the proposed approach, (b<sub>1-3</sub>) Delta2D, (c<sub>1-3</sub>) Melanie 7, (d) PDQuest, and (e<sub>1-3</sub>) Scimo. (x<sub>2</sub>) and (x<sub>3</sub>) depict magnified versions of the annotated regions of image (x<sub>1</sub>), with x = a, . . . , e.

larger number of background pixels (VE value of 82.1% compared to 5.7% for the proposed method). The same happens for Delta2D, Melanie 7 and Scimo. The Volumetric Overlap Error is utilized for perfect segmentation measuring. Its value is 0 when the perfect segmentation is achieved and 100 when

there is no overlap among the segmentation result and the ground truth. The VOE achieved by the proposed approach is limited to 8.3% compared to the second and third best VOE values achieved by Scimo and Melanie which are 12.6% and 21.6%, respectively. This result indicates that the proposed approach

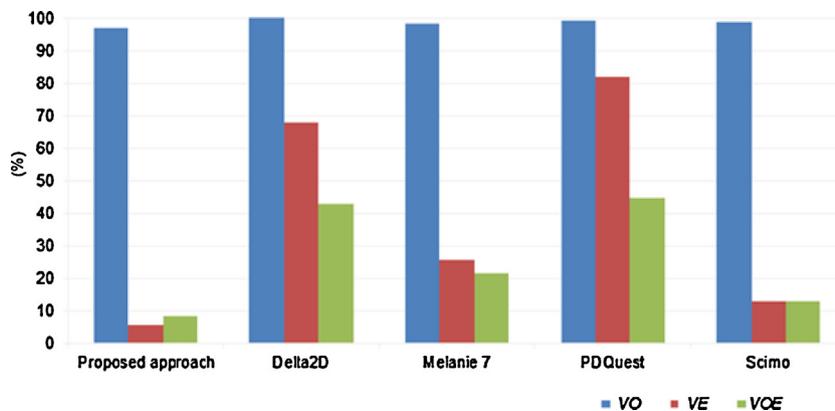


Fig. 10 – Overall segmentation results in terms of VO, VE and VOE obtained by the proposed approach, and the compared state-of-the-art techniques on the synthetic 2D-gel image dataset.

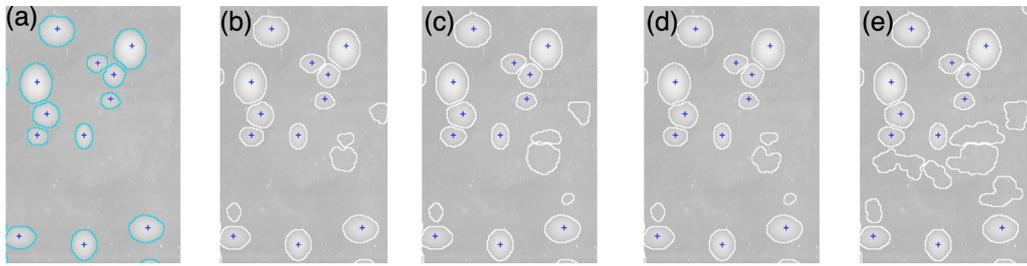


Fig. 11 – Segmentation results for a region of a synthetic 2D-gel image obtained by (a) the proposed approach, and after replacing the custom grow-cut algorithm with (b) the original grow-cut, (c) the fast marching method, (d) segmentation via adaptive weighted distances, and (e) watershed.

provides a more balanced and robust solution for 2D-gel image segmentation.

### 3.4. Evaluation of grow-cut performance against other seeded segmentation methods

Experiments were conducted on the synthetic image dataset in order to evaluate the use of the proposed custom grow-cut method against other seeded segmentation methods. The performance of the proposed 2D-gel image segmentation methodology was evaluated by replacing the proposed custom grow-cut algorithm with: (a) the original grow-cut algorithm [32], the Fast Marching Method [47], segmentation via adaptive weighted distances [48] and watershed [49]. The same background and foreground (spot) pixel seeds were used for all the examined methods. Additionally, all the other steps of the proposed methodology remained the same. Performance results for the examined approaches are shown in Table 6, in terms of the VO, VE and VOE measures. It is evident that the proposed custom grow-cut based method achieves more accurate segmentation, as it provides the lowest VE and VOE and comparable VO. The marginally higher VO achieved by the fast marching, adaptive weighted distances (A. W. D.) and watershed methods is attributed to the fact that they extend the spot

Table 6 – Comparison of final segmentation results on synthetic 2D-gel images using various seeded segmentation methods.

Method	VO (%)	VE (%)	VOE (%)
Proposed approach	97.0 ± 1.8	5.7 ± 1.6	8.3 ± 0.5
Original grow-cut	96.6 ± 2.1	7.9 ± 3.7	9.8 ± 0.6
Fast marching	98.5 ± 1.3	9.9 ± 7.8	11.1 ± 2.3
A. W. D. (*)	97.5 ± 1.9	8.9 ± 4.5	10.8 ± 2.4
Watershed	99.5 ± 0.7	49.0 ± 2.6	50.3 ± 2.5

A.W.D.: Adaptive weighted distances.

boundaries outside of the spot edges, resulting to the inclusion of background pixels to the spot area. The higher VE and VOE provided by these methods support this conclusion. Moreover, contrary to the proposed approach, all the alternative methods provide a large number of false positive (spurious) spots, as shown in an example of the final segmentation results for a region of an image, presented in Fig. 11. Nevertheless, the adequate performance of most of the examined methods shows that the proposed seed selection and post initial segmentation processing methods are able to enhance the segmentation results of various methods for the task of 2D-gel image segmentation.

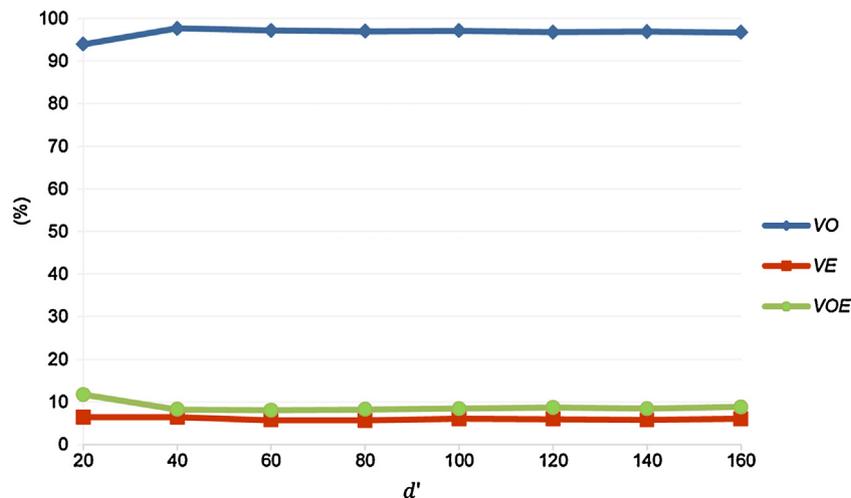


Fig. 12 – Overall segmentation results in terms of VO, VE and VOE obtained by the proposed approach for various values of  $d'$ .

### 3.5. Effect of the number $N$ of tiled windows on performance

Experiments were conducted on the synthetic image dataset in order to evaluate the effect of the number  $N$  of tiled windows on the performance of the proposed approach. As explained in Section 2.1.2,  $N$  depends on the dimension  $d'$  of the tiled windows. The proposed approach was applied on the synthetic image dataset for various values of  $d'$ , spanning from 20 to 160 pixels. The other parameters were kept as shown in Table 3 and performance was measured in terms of VO, VE and VOE. Results for the various values of  $d'$  are presented in Fig. 12. It is evident that the proposed approach exhibits significant performance stability to variations of  $d'$ . As a result, the effect of  $d'$  and consequently of the number of tiled windows  $N$  on the performance of the proposed methodology is minimal. Nevertheless, performance suffers for very small values of  $d'$ , as shown in Fig. 12.

## 4. Conclusion

The detection and segmentation of protein spots in 2D-gel images can be very challenging and arduous tasks due to the characteristics of the images. In this paper, an original approach for the detection and segmentation of 2D-gel spots is presented. The proposed approach is based on the grow-cut segmentation algorithm with a custom update rule that takes into consideration the inherent characteristics of 2D-gel images, along a region growing technique and morphological operators. The proposed methodology was evaluated on real as well as synthetic 2D-gel images. The high  $F$ -measure as well as the low volumetric overlap error (VOE) value achieved by the proposed approach indicate that it is a robust and effective method for 2D-gel image spot detection and segmentation. The experimental evaluation demonstrated that the proposed approach outperforms state-of-the-art 2D-gel image analysis software packages and techniques including Melanie 7, Delta2D, PDQuest, and Scimo, and as a result constitutes an advantageous and reliable solution for 2D-gel image analysis. The significantly lower volumetric error (VE) and volumetric overlap error (VOE) compared to the other methods shows that the proposed methodology provides more accurate results, enhancing the ability to extract more reliable biological conclusions at the next steps of 2D-gel image analysis. Additionally, the increased accuracy of the proposed spot detection method, in terms of the more reliable  $F$ -measure, reduces the number of false-positive (spurious) spots compared to the other methods and assists the segmentation process to effectively separate overlapping spots.

### Conflicts of interest

There are no conflicts of interest.

### Acknowledgments

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. The authors would like to extend their sincere appreciation to the expert biologists of the Biomedical Research Foundation of the Academy of Athens and in particular to Dr. S. Kossida and Dr. A. Vlahou for the provision of the real 2D-gel images and their ground truth and the segmentation results obtained by the Melanie 7 software program and Scimo, and to Dr. M. Makridakis for the segmentation results obtained by PDQuest. Furthermore, the authors would like to thank expert biologist Dr. M. Aivaliotis from the Institute of Molecular Biology and Biotechnology (IMBB) of the Foundation for Research and Technology-Hellas (FORTH) for providing the segmentation results obtained by the Delta2D software package.

### 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 Eng. Med. Biol. Mag.* 24 (3) (2005) 50–57.
- [2] G. Klaiman, T. Petzke, J. Hammond, A. LeBlanc, Targets of caspase-6 activity in human neurons and Alzheimer disease, *Mol. Cell Proteomics* 7 (8) (2008) 1541–1555.
- [3] K. Kultima, B. Scholz, H. Alm, K. Sköld, M. Svensson, A. Crossman, E. Bezar, P. Andrén, I. Lönnstedt, Normalization and expression changes in predefined sets of proteins using 2D-gel electrophoresis: a proteomic study of l-DOPA induced dyskinesia in an animal model of Parkinson's disease using DIGE, *BMC Bioinf.* 7 (1) (2006) 475–501.
- [4] M. Fuchs, T. Beißbarth, E. Wingender, K. Jung, Connecting high-dimensional mRNA and miRNA expression data for binary medical classification problems, *Comput. Methods Programs Biomed.* 111 (3) (2013) 592–601.
- [5] J. Lee, J. Han, G. Altwerger, E. Kohn, Proteomics and biomarkers in clinical trials for drug development, *J. Proteomics* 18 (2011) 2632–2641.
- [6] M. Natale, D. Bonino, P. Consoli, T. Alberio, R. Ravid, M. Fasano, E. Bucci, A meta-analysis of two-dimensional electrophoresis pattern of the parkinson's disease-related protein DJ-1, *Bioinformatics* 26 (7) (2010) 946–952.
- [7] I. Bajla, I. Holländer, S. Fluch, K. Burg, M. Kollár, An alternative method for electrophoretic gel image analysis in the GelMaster software, *Comput. Methods Programs Biomed.* 77 (3) (2005) 209–231.
- [8] A. Dowsey, M. Dunn, G. Yang, The role of bioinformatics in two dimensional gel electrophoresis, *Proteomics* 3 (8) (2003) 1567–1596.
- [9] J. Lopez, Two-dimensional electrophoresis in proteome expression analysis, *J. Chromatogr. B* 849 (1–2) (2007) 190–202.
- [10] W. Van Belle, N. Ånensen, I. Haaland, O. Bruserud, K. Hogda, B. Gjertsen, Correlation analysis of two-dimensional gel electrophoretic protein patterns and biological variables, *BMC Bioinf.* 7 (2006).
- [11] T. Aittokallio, J. Salmi, T. Nyman, O. Nevalainen, Geometrical distortions in two-dimensional gels: applicable correction methods, *J. Chromatogr. B: Analyt. Technol. Biomed. Life Sci.* 815 (1–2) (2005) 25–37.
- [12] S. Magdeldin, Y. Zhang, B. Xu, Y. Yoshida, T. Yamamoto, Two-dimensional polyacrylamide gel electrophoresis – a practical perspective, in: *Gel Electrophor.– Principles Basics*, 2012, pp. 91–116.

- [13] M. Savelonas, E. Mylona, D. Maroulis, Unsupervised 2D-gel electrophoresis image segmentation based on active contours, *Pattern Recognit.* 45 (2) (2012) 720–731.
- [14] A. dos Anjos, A. Moller, B. Ersboll, C. Finnie, H. Shahbazkia, New approach for segmentation and quantification of two-dimensional gel electrophoresis images, *Bioinformatics* 27 (3) (2011) 368–375.
- [15] Delta2D, DECODON, GmbH, BioTechnikum Greifswald, Germany, <http://www.decodon.com>
- [16] Progenesis SameSpots, Nonlinear Dynamics, UK, <http://www.totallab.com>
- [17] A. Görg, W. Weiss, M.J. Dunn, Current two-dimensional electrophoresis technology for proteomics, *Proteomics* 4 (12) (2004) 3665–3685.
- [18] J.I. Garrels, The QUEST system for quantitative analysis of two-dimensional gels, *J. Biol. Chem.* 264 (9) (1989) 5269–5282.
- [19] DeCyder 2D, GE, Healthcare, Munich, Germany, <http://www.gelifesciences.com/>
- [20] Melanie 7, Bio-Rad, Geneva, Switzerland, <http://www.genebio.com>
- [21] ImageMaster 2D, GE, Healthcare, Munich, Germany, <http://www.gelifesciences.com/>
- [22] B. Clark, H. Gutstein, The myth of automated, high-throughput two dimensional gel analysis, *Proteomics* 8 (6) (2008) 1197–1203.
- [23] P. Cutler, G. Heald, I. White, J. Ruan, A novel approach to spot detection for two-dimensional gel electrophoresis images using pixel value collection, *Proteomics* 3 (4) (2003) 392–401.
- [24] J.W. Yoon, S.J. Godsill, C. Kang, T.S. Kim, Bayesian inference for 2D-gel electrophoresis image analysis, in: *Proceedings of the 1st Int. Conf. Bioinf. Res. Develop.*, 2007, pp. 343–356.
- [25] J. Morris, B. Clark, H. Gutstein, Pinnacle, A fast, automatic and accurate method for detecting and quantifying protein spots in 2-dimensional gel electrophoresis data, *Bioinformatics* 24 (4) (2008) 529–536.
- [26] Y. Wu, L. Zhang, Comparison of two academic software packages for analyzing two-dimensional gel images, *J. Bioinf. Comput. Biol.* 9 (6) (2011) 775–794.
- [27] F. Li, F. Seillier-Moisewitsch, Differential analysis of 2D-gel images, *Methods Enzymol.* 487 (2011) 595–605.
- [28] L. Vincent, P. Soille, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (6) (1991) 583–598.
- [29] E. Mylona, M. Savelonas, D. Maroulis, S. Kossida, A computer based technique for automated spot detection in proteomics images, *IEEE Trans. Inf. Technol. Biomed.* 15 (4) (2011) 661–667.
- [30] E. Kostopoulou, E. Zacharia, D. Maroulis, An effective approach for detection and segmentation of protein spots on 2D-gel images, *IEEE J. Biomed. Health Inform.* 18 (1) (2014) 67–76.
- [31] E. Zacharia, E. Kostopoulou, D. Maroulis, N.P. Anagnostou, K.I. Pappa, 2D-GE spot detection combining multidirectional texture and spatial intensity cues, in: *Proceedings of the IEEE 13th International Conference on In Bioinformatics and Bioengineering (BIBE)*, 2013, pp. 1–4.
- [32] V. Vezhnevets, V. Konouchine, GrowCut-interactive multi-label N-D image segmentation by cellular automata, in: *Proceedings of the 15th International Conference on Computer Graphics and Applications*, 2005.
- [33] P.S. Liao, P.C. Chung, A fast algorithm for multilevel thresholding, *J. Inf. Sci. Eng.* 17 (5) (2001) 713–727.
- [34] M. Sonka, V. Hlavac, R. Boyle, *Image Processing, Analysis and Machine Vision*, Cengage-Engineering, Florence, KY, USA, 2007.
- [35] J.V. Neumann, in: A.W. Burks (Ed.), *Theory of self-reproducing automata*, Univ. Illinois Press, Champaign, IL, 1966.
- [36] R. Kikinis, S. Pieper, 3D slicer as a tool for interactive brain tumor segmentation, in: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 6982–6984.
- [37] S. Balocco, C. Gatta, F. Ciompi, O. Pujol, X. Carrillo, J. Mauri, P. Radeva, Combining growcut temporal correlation for IVUS lumen segmentation, pattern recognition and image analysis, *Lect. Notes Comput. Sci.* 6669 (2011) 556–563.
- [38] P. Ghosh, S.K. Antani, L.R. Long, G.R. Thoma, Unsupervised grow-cut cellular automata-based medical image segmentation, in: *Proceedings of the 1st IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB)*, 2011, pp. 40–47.
- [39] T. Yamasaki, T. Chen, M. Yagi, T. Hirai, R. Murakami, GrowCut-based fast tumor segmentation for 3D magnetic resonance images, in: *Proc. SPIE 8314, Medical Imaging: Image Processing*, 2012, p. 831434.
- [40] S. Katsigiannis, E. Zacharia, D. Maroulis, Grow-cut based automatic cDNA microarray image segmentation, *IEEE Trans. NanoBiosci.* 14 (1) (2015) 138–145.
- [41] Biomedical Research Foundation of the Academy of Athens (BRFAA), <http://www.bioacademy.gr>
- [42] Institute of Molecular Biology and Biotechnology (IMBB) of the Foundation for Research and Technology-Hellas (FORTH), <http://www.imbb.forth.gr>
- [43] M. Hayat, Discriminating protein structure classes by incorporating pseudo average chemical shift to Chou's general PseAAC and support vector machine, *Comput. Methods Programs Biomed.* 116 (3) (2014) 184–192.
- [44] B. Park, J. Im, N. Tuvshinjargal, W. Lee, K. Han, Sequence-based prediction of protein-binding sites in DNA: comparative study of two SVM models, *Comput. Methods Programs Biomed.* 117 (2) (2014) 158–167.
- [45] A.T. Azar, H.I. Elshazly, A.E. Hassanien, A.M. Elkorany, A random forest classifier for lymph diseases, *Comput. Methods Programs Biomed.* 113 (2) (2014) 465–473.
- [46] T. Heimann, M. Bryan, M. Styner, M.S. Niethammer, Warfield Segmentation of knee images: a grand challenge, in: *Proceedings of the MICCAI Workshop on Medical Image Analysis for the Clinic*, 2010.
- [47] J.A. Sethian, *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Materials Science*, 2nd ed., Cambridge University Press, 1999.
- [48] A. Protiere, G. Sapiro, Interactive image segmentation via adaptive weighted distances, *IEEE Trans. Image Process.* 16 (4) (2007) 1046–1057.
- [49] F. Meyer, Topographic distance and watershed lines, *Signal Process.* 38 (1994) 113–125.