

# 3-D Spot Modeling for Automatic Segmentation of cDNA Microarray Images

Eleni Zacharia and Dimitris Maroulis\*, *Member, IEEE*

**Abstract**—Spot segmentation—the second essential stage of cDNA microarray image analysis—constitutes a challenging process. At present, several up-to-date spot-segmentation techniques or software programs—proposed in the literature—are often characterized as “automatic.” On the contrary, they are in effect not fully automatic since they require human intervention in order to specify mandatory input parameters or to correct their results. Human intervention, however, can inevitably modify the actual results of the cDNA microarray experiment and lead to erroneous biological conclusions. Therefore, the development of an automated spot-segmentation process becomes of exceptional interest. In this paper, an original and fully automatic approach to accurately segmenting the spots in a cDNA microarray image is presented. In order for the segmentation to be accomplished, each real spot of the cDNA microarray image is represented in a three-dimensional (3-D) space by a 3-D spot model. Each 3-D spot model is determined via an optimization problem, which is solved by using a genetic algorithm. The segmentation of real spots is conducted by drawing the contours of their 3-D spot models. The proposed method has been compared with various published and established techniques, using several synthetic and real cDNA microarray images that contain thousands of spots. The outcome has shown that the proposed method outperforms prevalent existing techniques. It is also noise resistant and yields excellent results even under adverse conditions such as the appearance of spots of various sizes and shapes.

**Index Terms**—cDNA microarrays, genetic algorithm, image analysis, spot segmentation, 3-D spot modeling.

## I. INTRODUCTION

**T**HE DISCOVERY of cDNA microarrays in 1995 has fundamentally altered the way scientists monitor the expression levels of genes. Instead of analyzing the expression level of one gene at a time, scientists can simultaneously analyze the expression levels of thousands of genes over different samples [1]. Due to this revolutionary feature, during the last decade,

Manuscript received May 21, 2009; revised February 05, 2010; accepted May 03, 2010. Date of publication June 01, 2010; date of current version September 01, 2010. This work was supported by the Greek General Secretariat of Research and Technology (25%), the European Social Fund (75%), and the private sector, under the framework of Measure 8.3 of E.P. Antagonistikotita-3rd European Support Framework-PENED 2003 under Grant 03-ED-332. *Asterisk indicates corresponding author.*

E. Zacharia is with the Department of Informatics and Telecommunications, University of Athens, GR 15784 Panepistimiopolis, Ilisia, Greece (e-mail: ezacharia@gmail.com).

\*D. Maroulis is with the Department of Informatics and Telecommunications, University of Athens, GR 15784 Panepistimiopolis, Ilisia, Greece (e-mail: dmaroulis@di.uoa.gr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNB.2010.2050900

cDNA microarrays have been broadly used in many biomedical application areas such as: i) cancer research, infectious disease diagnosis, and treatment (i.e., determination of molecular differences between normal and abnormal cells, classification of tumors, determination of risk factors, and monitoring of treatment during different disease stages) [2]; ii) pharmacology research (i.e., determination of correlations between the genetic profiles of patients and their therapeutic responses to drugs) [3]; iii) toxicology research (i.e., determination of correlations between toxic responses to toxicants and changes in the genetic profiles of objects exposed to such toxicants) [4]; and iv) agricultural development [5]. In current-day bioinformatics, even if other techniques (such as proteomics) increasingly gain ground in the aforementioned biomedical applications, cDNA microarrays continue to be used.

Image analysis is an essential aspect of a cDNA microarray experiment. The process of analyzing a cDNA microarray image can be divided into three main stages [6]: i) gridding: the cDNA microarray image is cut into numerous compartments, each containing one individual spot and background; ii) spot segmentation: each compartment is individually segmented into a spot area and a background area; and iii) intensity extraction: the expression levels of the genes in these spot areas are calculated.

Although cDNA microarray image analysis may appear as relatively straightforward, it is in fact a rather complicated procedure. Amongst its stages, the segmentation of cDNA spots is the most challenging one. This is attributed to the nature of cDNA microarray images. The quality of images is often degraded due to the existence of noise and/or artifacts as well as due to uneven background [7]. Additionally, many spots are rather different from the ideal ones (circles with fixed diameters) as they vary in size and shape. Furthermore, some spots are so poorly contrasted that are not clearly visible. A considerable number of spot-segmentation techniques have therefore been proposed to date, some of which have been incorporated into commercial software programs that nevertheless fail to unequivocally resolve the aforementioned problems. According to [6], the segmentation techniques can be classified into: i) shape-segmentation approaches and ii) shape-independent approaches.

Shape-segmentation approaches assume that the spot has a certain two-dimensional (2-D) shape. Amongst them, the fixed circle segmentation algorithm (implemented by the ScanAlyze software program [8]) and the adaptive circle segmentation algorithm (implemented by the Dapple software program [9]) match circular templates to spots. Recently, another shape-segmentation method capable of segmenting noisy microarray spots has been proposed [10]. In this method, each spot is represented by a parametric circle with one or two elliptical

center holes. A weakness of the preceding methods is that they can segment only circular microarray spots. To deal with the various shapes of spots, the seeded region growing algorithm, adopted by the Spot software program [11] has been suggested. Although this algorithm can segment regions of irregular shapes by implementing a watershed algorithm, its performance is based on the appropriate selection of starting points.

Shape-independent segmentation approaches regard pixel intensities in a compartment as one-dimensional real values and classify pixels into a spot area or a background area. Amongst them, there are clustering algorithms such as K-means and, hybrid K-means, which use cluster intensity representation [12]–[15]. A weakness of these methods is that they do not perform well in the cases when the spots are poorly contrasted or when the spots are very close to each other. The model-based segmentation algorithm, based on the clustering of pixels' values, has been also suggested [16]. This method uses a removal technique of the components that are spatially connected in order to exclude small disconnected clusters which are assumed to be artifacts. An obvious weakness is that this removal technique is based on a threshold that defines the minimum size of the connected components in a cluster. Finally, there are segmentation methods based on active contours and multiple snakes [17]–[20]. Although these methods are very effective in segmenting typical microarray spots, they usually need a preprocessing stage in order to deal with large variations of spots sizes. Furthermore, in the case of the compartment being contaminated with noise and artifacts, these methods occasionally yield inaccurate results.

Hybrid approaches, which have recently been presented, utilize prior knowledge of the spot's shape and location as well as of the expected intensity information. For example, the Markov random fields method (MRF) [21], [22] utilizes neighboring information, along with intensity information based on an MRF modeling of the compartment. Although this combination of intensity and spatial information results in a more accurate pixel classification process, it requires an initial classification of the pixels, which in turn affects the final results [23]. Nagarajan [24] has also proposed a spot-segmentation method based on the correlation of the pixels comprising a spot. Although the correlation statistics is useful in flagging low-intensity spots, this method can perform well only on good-quality images. Gottardo *et al.* [25] has applied a Markov random field approach to segmentation. In their method the red and green intensities for both foreground and background were represented using an uncorrelated bivariate *t* distribution. Another hybrid approach is the segmentation method included in the Matarray toolbox of Matlab [26], which also combines both spatial and intensity information. A disadvantage of this method is that it requires input parameters in order to segment the spots.

As a result, the up-to-date spot-segmentation techniques or software programs lack automation; they require human intervention in order to specify mandatory input parameters or to correct their results. Automating this part of the process is essential because: i) it allows the acceleration of the analysis of the expression levels of thousands of genes and ii) it prevents variations in the results of gene expression levels due to human intervention as it is reported in the experiment presented in [27].

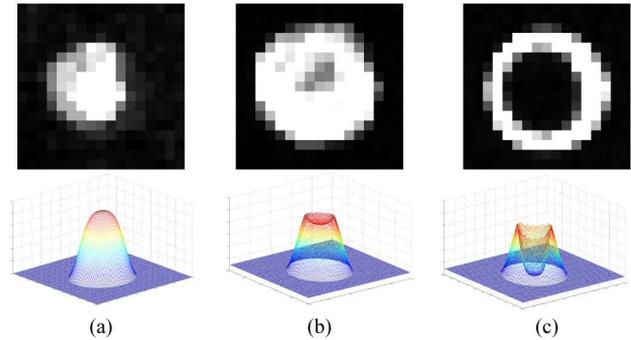


Fig. 1. Three types of real microarray spots in 2-D and 3-D representations. (a) Peak-shaped spot. (b) Volcano-shaped spot. (c) Doughnut-shaped spot.

In this paper, an original, fully automatic, and unsupervised approach to the segmentation of spots in cDNA microarray images is presented. The proposed approach does not belong to any of the aforementioned groups, as it processes microarray spots in a three-dimensional (3-D) space, in which the third dimension represents the intensity. In particular, the proposed approach relies on a genetic algorithm (GA) which represents the real spots of the cDNA microarray image with spot models, in a 3-D space. The segmentation of the real spots is conducted by drawing the contours of the spot models. Earlier versions of our proposed method on spot segmentation, published in conferences [28]–[31], utilize the diffusion model in order to represent the real spots in a 3-D space. The main disadvantage of these earlier versions of our method is that the diffusion model cannot represent all types of microarray spots (Fig. 1). It can only deal with peak-shaped spots. Therefore, the segmentation results are inefficient in the cases of volcano or doughnut-shaped spots. In the proposed approach an original spot model is presented which can be used for the representation of all types of real microarray spots. Consequently, the proposed method can segment all possible types of microarray spots. Moreover, the genetic algorithm has been further developed in order to be noise-resistant and yield more accurate results. It adopts fuzzy logic so as to take into account the uncertainties that exist in the pixels' intensities due to noise, artifacts, and uneven background. The proposed approach can conclusively deliver under the following adverse conditions: i) the appearance of various spot shapes, such as peak-shaped, volcano-shaped, and doughnut-shaped spots; ii) the appearance of spots of diverse intensities, such as low-intensity spots or saturated spots; and iii) the appearance of various spot sizes. Last but not least, the accuracy of the proposed method is higher than that of various well-known and broadly used segmentation techniques, as is evident in the comparison results in Section V.

The remainder of this paper is structured in five sections as follows: In Section II, a typical cDNA microarray spot is portrayed and a brief description of fuzzy logic and genetic algorithms is provided. In Section III, the proposed modeling of a microarray compartment is presented, while in Section IV the segmentation method is described. Section V illustrates experiments on the proposed method and offers its comparison with nine existing techniques. To this extent, several artificial cDNA microarray images and real cDNA microarray images were used. Our conclusions are contained in Section VI.

## II. BACKGROUND MATERIAL

### A. Typical cDNA Microarray Spots

Based on empirical observations [32], the following three types of microarray spots, in relation to their shapes, can be identified in a cDNA microarray image.

- 1) Peak-shaped spot [Fig. 1(a)]; this type of spot has an intensity that peaks at its central region and declines at regions further from the center. In the case when the peak is narrow, the spot resembles a 3-D-Gaussian function. In the case when the peak is wide, the spot resembles a plateau.
- 2) Volcano-shaped spot [Fig. 1(b)]; this type of spot is defined as the peak-shaped spot having a small hole in the area of its peak. It therefore resembles a volcano.
- 3) Doughnut-shaped spot [Fig. 1(c)]; this type of spot has a narrow rim of high intensity and a large hole of very low intensity at its central region.

### B. A Conventional Genetic Algorithm

Genetic Algorithm is a powerful optimization search methodology and a promising alternative to conventional heuristic methods as it exhibits a number of advantages: it can optimize a large number of variables of extremely complex cost surfaces and solve problems for which there is little or no *a priori* knowledge of the underlying processes. Furthermore, it permits the implementation of our method automatically without human intervention which in turn has the following benefits: i) acceleration of the process of microarray images and ii) prevention of eventual variations of the results.

A conventional genetic algorithm emulates the principles of natural selection and evolution [33]. Its search begins by constructing a finite number of potential solutions encoded as alpha-numerical sequences called chromosomes. These chromosomes, which constitute an initial population  $\text{Pop}_1$ , are evaluated using a fitness function. Subsequently, the population  $\text{Pop}_1$  evolves into a new population  $\text{Pop}_2$  using the following three genetic operators: reproduction, crossover, and mutation. This evolutionary cycle of a current population  $\text{Pop}_n$  (where  $n$  stands for the consecutive number of populations) to its next  $\text{Pop}_{n+1}$  continues until a specific termination criterion is satisfied. In conclusion, the following four basic elements are deemed necessary for the determination of a particular problem: chromosome representation, chromosome evaluation, evolutionary cycle, and termination criteria.

### C. Fuzzy Logic

The fundamental assumption upon which the two-valued logic (or classical logic) is based is that every proposition is either *true* or *false*. However, it is now well acknowledged that many propositions can simultaneously be partially true and partially false. In order to accommodate the concept of uncertainty appearing in propositions, fuzzy logic [34] broadens the two-valued logic by permitting intermediate values to be defined between the two conventional ones. As a result, the “truth value” of a proposition is represented by a value between

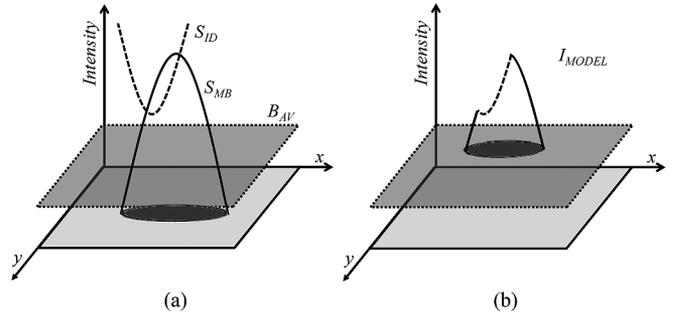


Fig. 2. The compartment model  $I_{MODEL}$  and its components. (a) The solid curve represents the main body  $S_{MB}$  of the spot model, the dashed curve represents the inner dip  $S_{ID}$  of the spot model, while the dotted surface represents the average background intensity  $B_{AV}$  of the compartment model. (b) The corresponding compartment model.

0 and 1 according to the amount of certainty that the proposition is truth.

In 1965, Zadeh [34] described mathematically the fuzzy logic using the concept of fuzzy subsets. More precisely, let  $X = \{x\}$  be a nonempty set, called the “universe of discourse,” consisting of all the possible elements of concern in a particular context. A fuzzy subset  $A$  of  $X$  is defined as a set of ordered pairs as follows:

$$A \equiv \{\langle x, \mu_A(x) \rangle \mid x \in X\} \quad (1)$$

where  $x$  is an element of Universe  $X$  and  $\mu_A(x)$  is defined as the membership function of  $x$  in the fuzzy set  $A$ , which assigns to each element  $x$  a value between 0 and 1 ( $\mu_A(x) \in [0, 1]$ ). The value of  $\mu_A(x)$  is called the “membership degree” of  $x$  in the fuzzy set  $A$ , and it expresses the degree in which the element  $x$  belongs to the fuzzy set  $A$ . If the value of  $\mu_A(x)$  equals 0, then  $x$  does not belong to the fuzzy set  $A$ , or in other words, the proposition “ $x$  belongs to  $A$ ” is completely false. Respectively, if the value of  $\mu_A(x)$  equals 1, then  $x$  belongs to the fuzzy set  $A$ , or in other words, the proposition “ $x$  belongs to  $A$ ” is completely true.

## III. PROPOSED MODELING OF A MICROARRAY COMPARTMENT

Due to the aforementioned common spots’ characteristics, a microarray compartment can be represented by a 3-D compartment model, in which the third dimension represents the intensity. More precisely, let  $I_{REAL}$  be a compartment of a real microarray image containing one individual spot and its background. A compartment model  $I_{MODEL}$  can represent, in a 3-D space, the real one  $I_{REAL}$  by modeling the latter’s intensity values. The  $I_{MODEL}$  can be expressed as a 3-D curve constituted of two components: i) a surface representing the average background intensity  $B_{AV}$  of the compartment model and ii) a 3-D curve representing the spot model  $S_{MODEL}$ . In turn, the spot model  $S_{MODEL}$  can be expressed as a 3-D curve made of two components: The first component  $S_{MB}$  represents the 3-D curve of the main body of the spot model, while the second one  $S_{ID}$  represents the 3-D curve of the inner dip of the spot model (Fig. 2).

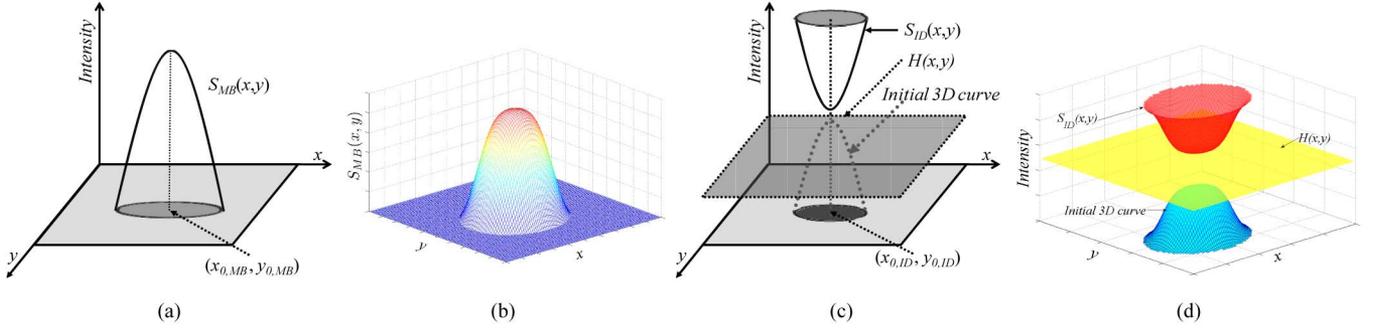


Fig. 3. Components of the spot model. (a), (b) The main body  $S_{MB}(x, y)$  of the spot model. (c), (d) 3-D representation of the  $S_{ID}(x, y)$  3-D curve, the initial 3-D curve, and the horizontal surface  $H(x, y)$ .

### A. The Spot Model and Its Components

Let us consider a curve  $C(x, y)$  whose 3-D shape resembles the 3-D Gaussian or plateau curve. The  $C(x, y)$  3-D curve can be defined by the following equation:

$$C(x, y) = h \cdot [\text{erf}(a + r(x, y)) + \text{erf}(a - r(x, y))] \quad (2)$$

where

$$r(x, y) = \sqrt{(x - x_0)^2/D_x + (y - y_0)^2/D_y}. \quad (3)$$

$h$  controls the height of the 3-D curve.  $\text{erf}(z)$  denotes the error function encountered in integrating the normal distribution.  $(x_0, y_0)$  are the center coordinates of the 3-D curve.  $D_x, D_y$  control the slope of the 3-D curve on the two main directions  $(x, y)$  of the 2-D plane.  $a$  controls the shape of the 3-D curve. For  $a \rightarrow 0$ ,  $C(x, y)$  resembles a 3-D Gaussian function, while for  $a \rightarrow \infty$ ,  $C(x, y)$  resembles a plateau or a saturated spot.

In order for the  $C(x, y)$  3-D curve to have any possible direction on the 2-D plane, a rotation transformation is applied; the  $C(x, y)$  3-D curve is rotated by an angle  $\theta$  in respect to the vertical axis passing through its center  $(x_0, y_0)$ .

The main body  $S_{MB}(x, y)$  of the spot model can be expressed as a 3-D curve deriving from (2) [Fig. 3(a), (b)]. Correspondingly, the inner dip of the spot model  $S_{ID}(x, y)$  can be defined as a symmetrical 3-D curve to an ‘‘initial 3-D curve’’ (deriving from (2)), in respect to an horizontal surface  $H(x, y)$  [Fig. 3(c), (d)].

The total spot model  $S_{Model}(x, y)$  is constructed by combining the  $S_{ID}(x, y)$  and  $S_{MB}(x, y)$  3-D curves, and is defined as

$$S_{MODEL}(x, y) = \text{Min} [S_{MB}(x, y), S_{ID}(x, y)]. \quad (4)$$

Fig. 4 depicts 2-D illustrations of two spot models by portraying the cross sections of their  $S_{MB}(x, y)$  and  $S_{ID}(x, y)$  3-D curves. The resulting spot models  $S_{MODEL}(x, y)$  are the surfaces colored in grey. It is worth pointing out that the resulting total model in Fig. 4(b) is a peaked-shaped spot. The cross section of its 3-D curve is identical to the cross section of the 3-D curve of its main body  $S_{MB}(x, y)$ .

### B. The Compartment Model

The compartment model  $I_{MODEL}(x, y)$  corresponds to a 3-D curve whose function is defined as

$$I_{MODEL}(x, y) = \text{Max} [B_{AV}, S_{MODEL}(x, y)] \quad (5)$$

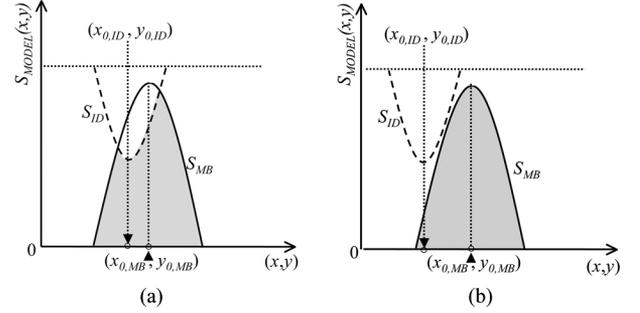


Fig. 4. 2-D illustrations of spot models.

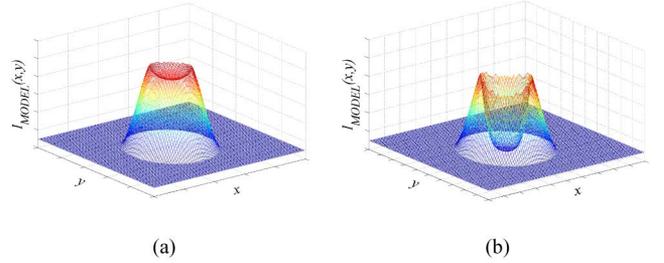


Fig. 5. Examples of the 3-D compartment models containing: (a) a volcano-shaped spot and (b) a doughnut-shaped spot.

where  $B_{AV}$  denotes the average background intensity of the compartment model  $I_{MODEL}$ .  $B_{AV}$  corresponds to a threshold of the lowest values of the  $S_{MODEL}(x, y)$ . Pixels whose values are lower than  $B_{AV}$  belong to the background and their values are set equal to  $B_{AV}$ .

Two examples of compartment models  $I_{MODEL}(x, y)$  are depicted in Fig. 5. The first compartment model contains a volcano-shaped spot model, while the second compartment model contains a doughnut-shaped spot model.

## IV. PROPOSED GENETIC ALGORITHM TO SEGMENTING MICROARRAY IMAGES VIA OUR SPOT MODEL

Given that  $I_{REAL}$  is one of the compartments of a real microarray image containing one individual spot  $S_{REAL}$  and background  $B_{REAL}$ , the segmentation procedure aims at the delimitation of the boundaries of the spot  $S_{REAL}$ . The segmentation procedure is divided into two stages.

Stage 1: The compartment  $I_{REAL}$  of the microarray image is optimally represented by a 3-D compartment model  $I_{MODEL}$ .

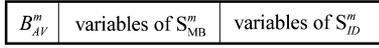
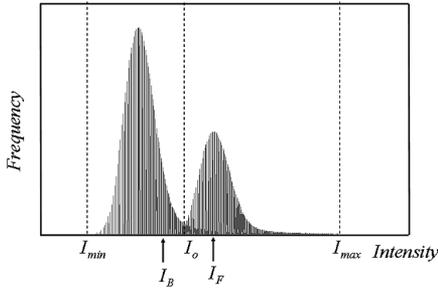
Fig. 6. The three segments of a chromosome  $m$ .

Fig. 7. A typical histogram of a real microarray compartment. The left curve corresponds to background pixels while the right corresponds to spot pixels.

Stage 2: The boundaries of the microarray spot  $S_{REAL}$  are depicted by drawing the contour of the spot model  $S_{MODEL}$ .

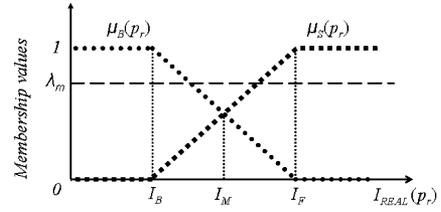
The first stage of the segmentation procedure is regarded as an optimization problem of modeling a microarray compartment, and it is tackled by using the proposed genetic algorithm. The genetic algorithm determines the values of the variables of the compartment model (5) so that the resulting compartment model represents optimally the real one.

#### A. Chromosome Representation

A chromosome  $m$  represents a specific compartment model  $I_{MODEL}^m$  in a 3-D space, where  $m$  stands for a specific chromosome. It is therefore a simple numerical sequence which encodes the values of the variables defining the specific compartment model. It consists of three segments (Fig. 6). The first segment encodes the value of the average background intensity of the compartment model  $B_{AV}^m$ . The second segment encodes the values of the variables of the main body  $S_{MB}^m$  of the spot model  $S_{MODEL}^m$ , while the third segment encodes the values of the variables of the inner dip  $S_{ID}^m$  of the spot model  $S_{MODEL}^m$ . The use of a numerical sequence instead of a binary sequence is preferred in the present application of genetic algorithms. The reason lies in the fact that a real-coded genetic algorithm reveals numerous benefits over the binary-coded genetic algorithms [35].

#### B. Fuzzy Logic for Determining the Error Estimation Between the Model Compartment and the Real One

The aim of the genetic algorithm is the maximization of the resemblance between the compartment model  $I_{MODEL}^m$  and the real one  $I_{REAL}$ . However, since the real compartment is contaminated with noise and artifacts, its intensity values are noticeably fluctuated—even between two consecutive pixels—resulting in a scabrous 3-D curve that contains many peaks. As a result, pixels belonging to the spot area  $S_{REAL}$  may have lower intensity values than the pixels belonging to the background area  $B_{REAL}$ . Correspondingly, pixels belonging to the background area  $B_{REAL}$  may have higher intensity values than the pixels belonging to the spot area  $S_{REAL}$ . To deal with the ambiguity and vagueness of the intensity values of pixels—due to noise,

Fig. 8. The membership functions  $\mu_B(p_r)$  and  $\mu_S(p_r)$ .

artifacts, and uneven background—we set the *membership degree* of a pixel  $p_r$  to belong to the background area or to the spot area according to the following two rules of fuzzy logic theory:

- 1) The smaller the intensity's value  $I_{REAL}(p_r)$  is, the greater the *membership degree* that  $p_r$  belongs to the background area becomes.
- 2) The higher the intensity's value  $I_{REAL}(p_r)$  is, the greater the *membership degree* that  $p_r$  belongs to the spot area becomes.

Based on the two aforementioned rules, the membership function  $\mu_B(p_r)$  of a pixel  $p_r$  so that it belongs to the background area, and the membership function  $\mu_S(p_r)$  of a pixel  $p_r$  so that it belongs to the spot area, are defined by the following equations:

$$\mu_B(p_r) = \begin{cases} 1, & \text{if } I_{REAL}(p_r) \leq I_B \\ \frac{I_F - I_{REAL}(p_r)}{I_F - I_B}, & \text{if } I_B < I_{REAL}(p_r) < I_F \\ 0, & \text{if } I_{REAL}(p_r) \geq I_F \end{cases} \quad (6)$$

and

$$\mu_S(p_r) = 1 - \mu_B(p_r) \quad (7)$$

where  $I_B$  and  $I_F$  are two intensity values. More precisely, let  $I_o$  be the intensity corresponding to the minimum between the maxima of those two normal distributions which represent the distributions of background pixels and spot pixels (Fig. 7).  $I_{min}$  and  $I_{max}$  are the minimum and maximum intensity values that appear in the  $I_{REAL}$ . Let  $N_1, N_2$  be the number of pixels whose intensity values are less or higher than  $I_o$ , respectively.  $I_B$  is chosen so that  $k \cdot N_1$  number of pixels has intensity lower than or equal to  $I_B$ , where  $k$  is a constant ( $0 \leq k \leq 1$ ).  $I_F$  is chosen so that  $k \cdot N_2$  number of pixels has intensity higher than or equal to  $I_F$ .

Fig. 8 represents the membership functions  $\mu_B(p_r)$  and  $\mu_S(p_r)$ . Pixels with intensity lower than or equal to  $I_B$  belong to the background area ( $\mu_B(p_r) = 1$  and  $\mu_S(p_r) = 0$ ), while pixels with intensity higher than or equal to  $I_F$  belong to the spot area ( $\mu_B(p_r) = 0$  and  $\mu_S(p_r) = 1$ ). Pixels with intensity between  $I_B$  and  $I_F$  have a *membership degree*  $\mu_B(p_r)$  so as to belong to the background area and a *membership degree*  $\mu_S(p_r)$  so as to belong to the spot area. In this case, the values of  $\mu_B(p_r)$  and  $\mu_S(p_r)$  range between (0,1).

Contrary to the scabrous 3-D curve of the real compartment, the compartment model has a smooth 3-D curve. Consequently, some of the points of the 3-D curve of the compartment model are identical to the points of the real one, while some others interpolate the points of the real one. The identical points should belong mostly to the region near the spot's contour, while the

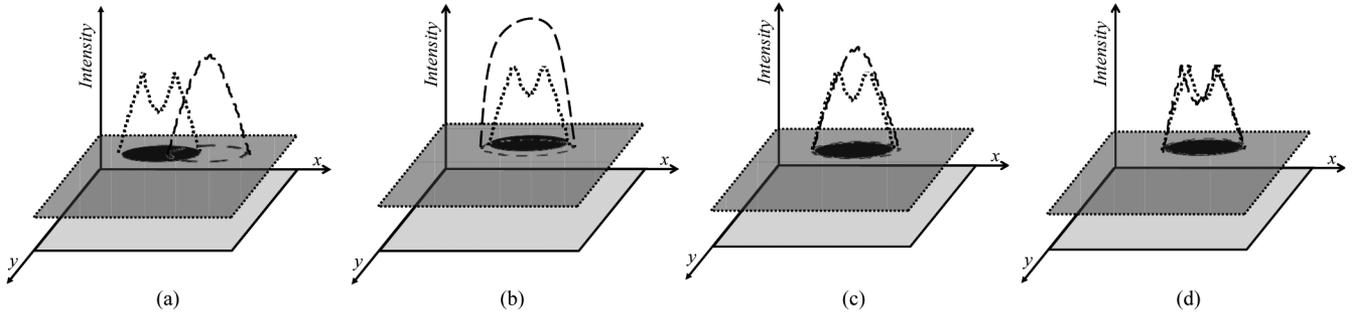


Fig. 9. Overlapping of the real compartment  $I_{\text{REAL}}$  and compartment model  $I_{\text{MODEL}}^m$ . The dotted curve represents the real spot  $S_{\text{REAL}}$ , while the dashed curve represents the spot model  $S_{\text{MODEL}}^m$ . The chromosome  $m$  representing the  $I_{\text{MODEL}}^m$  in (d) should have progressively higher fitness value than that of (c), (b), and (a).

interpolated points should belong mostly to spot areas or background areas. Therefore, the *relative intensity error*  $E^m(p_m, p_r)$  between a model-compartment's pixel  $p_m$  and its corresponding real compartment's pixel  $p_r$  is defined according to the pixels' locations, as follows:

$$E^m(p_m, p_r) = \begin{cases} E_I^m(p_m, p_r), & \text{if } p_m, p_r \in \text{spot background area} \\ E_R^m(p_m, p_r), & \text{if } p_m, p_r \in \text{near spot's contour area} \end{cases} \quad (8)$$

where the  $E_I^m(p_m, p_r)$  function calculates the *relative intensity error* in the case of the  $p_m$  and  $p_r$  pixels belonging to the spot area or the background area. Likewise,  $E_R^m(p_m, p_r)$  function calculates the *relative intensity error* in the case of the  $p_m$  and  $p_r$  pixels being situated near the spot's contour area.

More precisely, in the case when the membership value  $\mu_B(p_r)$  (or  $\mu_S(p_r)$ ) of the pixel  $p_r$  is higher than or equal to a constant  $\lambda_m$ , the  $p_r$  and its corresponding  $p_m$  pixel are considered to be located in the background area (or the spot area, respectively) (Fig. 8). In this case, the optimal intensity value of the compartment model on the  $p_m$  pixel may interpolate a set of intensity values near the  $p_r$  pixel of the real one. Therefore, the  $E_I^m(p_m, p_r)$  *relative intensity error* is defined as the minimum relative error between the intensity of the model-compartment's pixel  $p_m$ , and the intensity values of the  $p_i$  pixels located in a  $3 \times 3$  window around its corresponding  $p_r$  pixel of the real compartment

$$E_I^m(p_m, p_r) = \text{Min} \left[ \frac{|I_{\text{MODEL}}^m(p_m) - I_{\text{REAL}}(p_i)|}{I_{\text{REAL}}(p_i)}, i = 0, \dots, 8 \right]. \quad (9)$$

Otherwise, according to (8), the  $p_r$  pixel and its corresponding  $p_m$  pixel are considered to be located in the region near the spot's contour. In this case, the optimal intensity value of the compartment model at  $p_m$  pixel should equal to the intensity value of the real one at  $p_r$  pixel. Therefore, the  $E_R^m(p_m, p_r)$  *relative intensity error* is defined as the relative error between the intensity of the model compartment's pixel  $p_m$ , and the intensity value of its corresponding real compartment's pixel  $p_r$ .

$$E_R^m(p_m, p_r) = \frac{|I_{\text{MODEL}}^m(p_m) - I_{\text{REAL}}(p_r)|}{I_{\text{REAL}}(p_r)}. \quad (10)$$

### C. Chromosome Evaluation

The aim of the genetic algorithm is satisfied by the appropriate evaluation of the chromosomes. Each chromosome  $m$  is evaluated using a fitness function  $F(m)$ , which assigns to it a degree of how appropriate a solution to the optimization problem it is. As far as the specific optimization problem of modeling a microarray compartment, the higher the resemblance of a compartment model  $I_{\text{MODEL}}^m$  to the real one  $I_{\text{REAL}}$  is, the higher the value of the fitness function of the chromosome  $m$  becomes. During evaluation, a chromosome  $m$  can be classified into one of the following two categories.

- 1) An *inefficient* chromosome: This is a chromosome which represents a compartment model  $I_{\text{MODEL}}^m$  that hardly resembles the real one  $I_{\text{REAL}}$ . In this case, the spot model  $S_{\text{MODEL}}^m$  satisfies one of the following attributes: i) its  $S_{\text{MB}}^m$  component is mostly or entirely located in the background area  $B_{\text{REAL}}$  [Fig. 9(a)] or ii) its  $S_{\text{MB}}^m$  component approximates deficiently the real spot  $S_{\text{REAL}}$  [Fig. 9(b)].
- 2) An *acceptable* chromosome: This is a chromosome which represents a compartment model  $I_{\text{MODEL}}^m$  that resembles—to a degree—the real one  $I_{\text{REAL}}$ . In this case, the spot model  $S_{\text{MODEL}}^m$  satisfies one of the following attributes: i) its  $S_{\text{MB}}^m$  component approximates—to a degree—the real spot  $S_{\text{REAL}}$ , but its  $S_{\text{ID}}^m$  component does not [Fig. 9(c)] or ii) both its  $S_{\text{MB}}^m$  and  $S_{\text{ID}}^m$  components of the spot model  $S_{\text{MODEL}}^m$  resemble—to a degree—the real spot  $S_{\text{REAL}}$  [Fig. 9(d)].

The main difference between an *inefficient* chromosome and an *acceptable* one is that the former's  $S_{\text{MB}}^m$  component approximates deficiently the real spot  $S_{\text{REAL}}$ . Therefore, there is no need for a precise estimation of the fitness value in the former case. The fitness function  $F(m)$  of a chromosome  $m$  is defined by the following equation:

$$F(m) = \begin{cases} R_R(m), & \text{if } m \equiv \text{"inefficient"} \\ R_P(m), & \text{else if } m \equiv \text{"acceptable"} \end{cases} \quad (11)$$

where  $R_R(m)$  denotes a rough estimation of the resemblance degree between the real compartment  $I_{\text{REAL}}$  and the model compartment  $I_{\text{MODEL}}^m$ , while  $R_P(m)$  denotes a precise estimation of the above resemblance degree.

More precisely, the  $R_R(m)$  is defined as the product of the following two percentages:

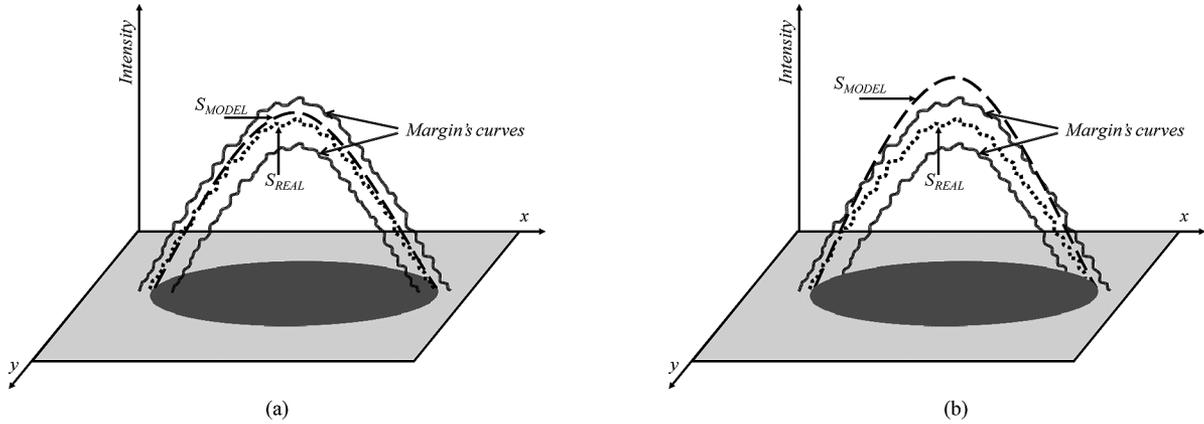


Fig. 10. The dotted curve represents the 3-D curve of the real spot while the dashed curve represents the 3-D curve of the spot model. Spot models whose values fall within the margin, defined by the solid curves, efficiently represent the real microarray spot. (a) An efficient spot model. (b) An inefficient spot model.

- i) the portion of spot pixels (their intensity's value is higher than  $I_M$ —Fig. 8) whose intensity value has been *efficiently represented* by the main body  $S_{MB}^m$  of the spot model, and
- ii) the portion of background-pixels (their intensity's value is less or equal to  $I_M$ —Fig. 8) whose intensity value has been *efficiently represented* by the average background intensity  $B_{AV}^m$ .

A  $p_r$  pixel is considered to be *efficiently represented* by its corresponding pixel  $p_m$  if its *relative intensity error* is less or equal to a constant  $E_{MAX}$  ( $E^m(p_m, p_r) \leq E_{MAX}$ ).  $E_{MAX}$  is a positive constant which controls an acceptable margin of the error existing between the intensity values of the compartment model and the intensity values of the real compartment. For example, Fig. 10 depicts the margin in the area of the real microarray spot. Spot models whose values fall within this margin represent efficiently the real microarray spot.

The precise estimation of the resemblance's degree  $R_P(m)$  is defined as

$$R_P(m) = q \cdot |f_S(m)| \cdot |f_B(m)| \cdot f_O(m) \quad (12)$$

where

$$f_S(m) = \sum_{\forall p_r \wedge p_m} \mu_S(p_r) \cdot (1 - w \cdot E^m(p_m, p_r)), \quad (13)$$

$$f_B(m) = \sum_{\forall p_r \wedge p_m} \mu_B(p_r) \cdot (1 - w \cdot E^m(p_m, p_r)), \quad (14)$$

$$f_O(m) = TP(m) \cdot TN(m). \quad (15)$$

$$q = \begin{cases} -1, & \text{if } f_S(m) < 0 \vee f_B(m) < 0 \\ +1, & \text{otherwise.} \end{cases} \quad (16)$$

$f_S(m)$  denotes the degree of the real spot's pixels whose intensity values have been *efficiently represented* by the spot model  $S_{MODEL}^m$ .  $f_B(m)$  denotes the degree of the background's pixels whose intensity values have been *efficiently represented* by the average background intensity  $B_{AV}^m$ .  $w$  is a weight coefficient which is set equal to 0.1 or to 1 according to the value of the  $E_R^m(p_m, p_r)$  error. If  $E_R^m(p_m, p_r) \leq E_{MAX}$ , the *relative intensity error* between the compartment model  $I_{MODEL}^m$  and the real one  $I_{REAL}$ , on the corresponding pixels, is considered negligibly small and thus insignificant. Consequently, the

$E^m(p_m, p_r)$  is multiplied by 0.1 ( $w = 0.1$ ). Otherwise  $w$  is set equal to 1.

$q$  is a weight coefficient which is set equal to +1 or to -1 according to the values of the  $f_S(m)$  or  $f_B(m)$ . If one of the  $f_S(m)$  or  $f_B(m)$  value is negative,  $q$  is set equal to -1 so that the resemblance  $R_P(m)$  also becomes negative. Otherwise, it is set equal to +1.

$f_O(m)$  denotes the degree to which the area containing the spot model overlaps the area containing the real spot.  $TP(m)$  denotes the true positive rate: the portion of the compartment model's pixels which have been correctly identified as spot pixels; in other words the portion of pixels which have been identified as spot pixels and their intensity value is higher than  $I_M$  (Fig. 8).  $TN(m)$  denotes the true negative rate: the portion of the compartment model's pixels which have been correctly identified as background pixels; in other words the portion of pixels which have been identified as background-pixels and their intensity value is lower or equal to  $I_M$  (Fig. 8). Due to the uncertainty existing in the pixel's intensities, the pixels contributing to the calculations of  $TP(m)$  and  $TN(m)$  are weighted: For the calculation of  $TP(m)$ , the weight coefficient of a pixel  $p_r$  equals to its corresponding *membership degree*  $\mu_S(p_r)$ , while for the calculation of  $TN(m)$  the weight coefficient of a pixel  $p_r$  equals to its corresponding *membership degree*  $\mu_B(p_r)$ . The higher the  $TP(m)$  and the  $TN(m)$  are, the higher the overlapping of the spot model with the real one is.

Last but not least, an *inefficient* chromosome in (11) represents a model-compartment having one of the following attributes: i) its resemblance  $R_R(m)$  to the real compartment is less than a threshold  $Th$  or ii) its resemblance  $R_R(m)$  to the real compartment is higher than the threshold  $Th$  but its resemblance  $R_P(m)$  to the real compartment is negative. Otherwise, a chromosome is *acceptable*.

It is worth mentioning that by using this fitness function the genetic algorithm can progressively assign—from left to right—a higher fitness value to the chromosomes representing the compartment models in Fig. 9.

#### D. Evolutionary Circle Termination Criteria

A new population  $Pop_{n+1}$  is created from the current  $Pop_n$  by applying the following stages: i) Reproduction stage:  $P_r\%$  of

TABLE I  
PROPOSED METHOD VERSUS ESTABLISHED TECHNIQUES

Algorithm	Probability of error		Discrepancy distance	
	GQI	LQI	GQI	LQI
Fixed Circle [8]	0.049	0.049	0.027	0.027
Adaptive Circle [9]	0.019	0.192	0.017	0.074
Seeded region growing [6]	0.099	0.114	0.037	0.048
Mann-Whitney [37]	0.165	0.162	0.066	0.074
Hybrid k-means [13]	0.017	0.020	0.016	0.029
Markov random field [21]	0.154	0.053	0.063	0.039
Matararray [26]	0.004	0.031	0.008	0.068
Model-based segmentation [16]	0.094	0.101	0.052	0.067
k-means [12]	0.000	0.025	0.000	0.041
<b>Proposed method</b>	<b>0.000</b>	<b>0.012</b>	<b>0.000</b>	<b>0.018</b>

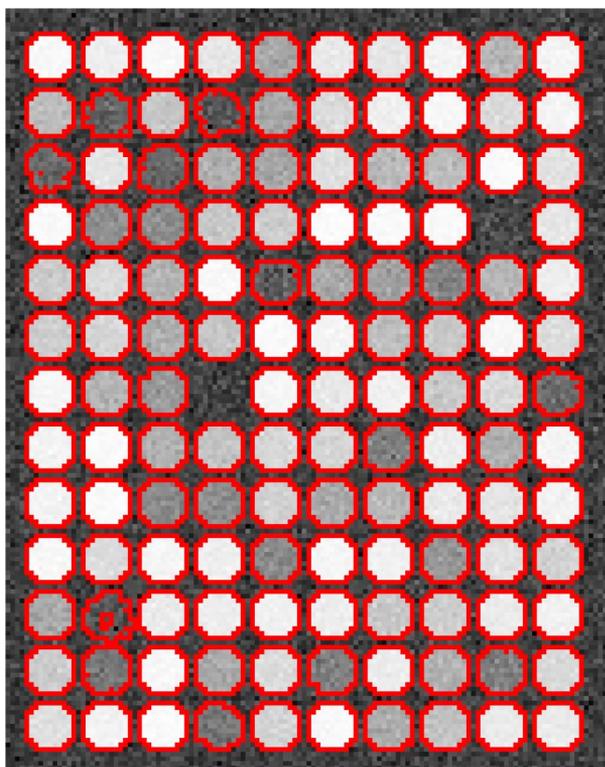


Fig. 11. Spot-segmentation result of a block in a good-quality artificial microarray image.

the best chromosomes of the current population  $\text{Pop}_n$  are carried over to the new population  $\text{Pop}_{n+1}$ . ii) Crossover-mutation stage: The chromosomes needed to complete the new population  $\text{Pop}_{n+1}$  are produced through iterations of the following: Four chromosomes of the population  $\text{Pop}_n$  are selected using the tournament selection method; these chromosomes are subsequently subjected in turn to the joint application of the BLX-a, and the dynamic heuristic crossover operator (according to a  $P_c\%$  probability) and then to the wavelet mutation operator (according to a  $P_m\%$  probability). The best two of the four resulting chromosomes proceed to the new population  $\text{Pop}_{n+1}$ . New populations are thus produced until the genetic algorithm is executed up to a maximum number of populations  $G_{Fit}$  for which the best fitness value has remained unchanged. Due to

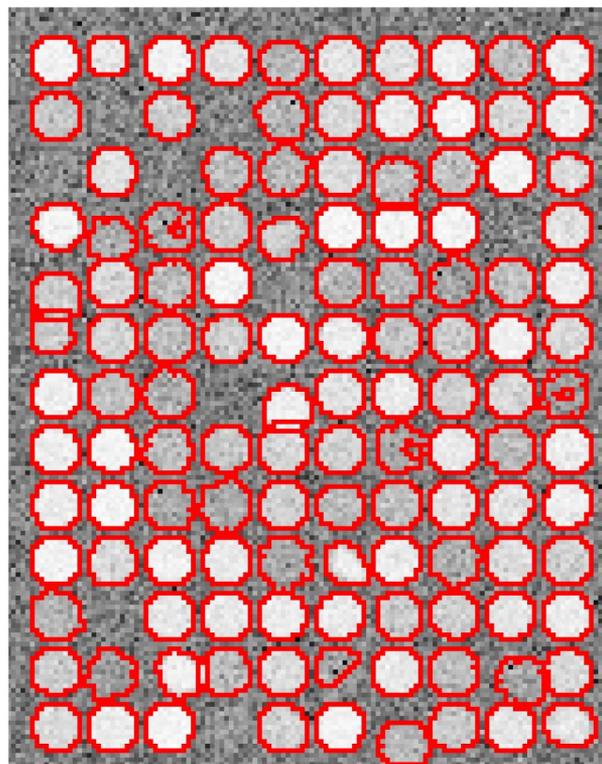


Fig. 12. Spot-segmentation result of a block in a low-quality artificial microarray image.

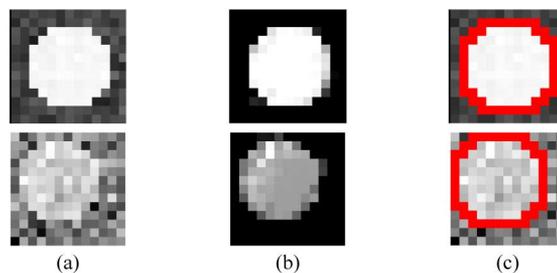


Fig. 13. Segmentation results of two magnified artificial spots. Two spots from a GQI (a), with their spot models (b) and their segmentation results (c).

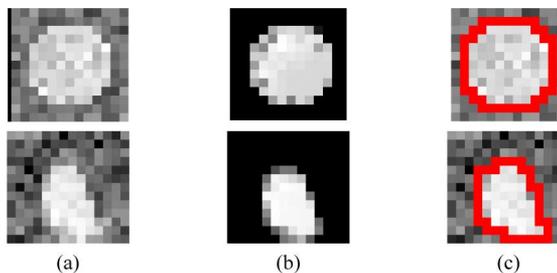


Fig. 14. Segmentation results of two magnified artificial spots. Two spots from a LQI (a), with their spot models (b) and their segmentation results (c).

the evolutionary circles of the genetic algorithm, it is obvious that its execution time is higher than the ones of relevant techniques (see results section). This drawback is compensated by: i) the high accuracy of the proposed method's segmentation results, which, as demonstrated in the results section, is higher than the ones provided by established techniques and software programs and ii) the lack of variations in the results of gene expression levels due to human intervention, since the proposed

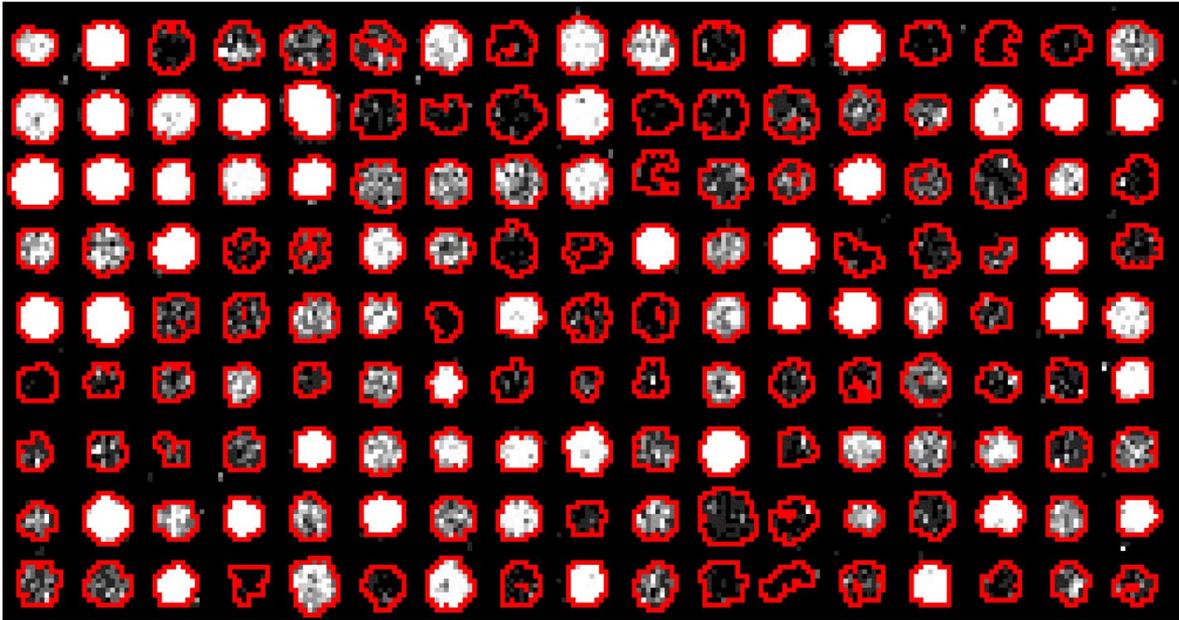


Fig. 15. Segmentation results of a region of a real microarray block obtained from the Stanford microarray database.

method is fully automatic. These attributes are very significant since erroneous segmentation results can detrimentally affect the final gene-expression results [27]. Moreover, the computational speed of modern microcomputers is ever increasing, while the execution time of the proposed method can be decreased by parallelizing the genetic algorithm.

## V. RESULTS

Several experiments on various images, including synthetic and real cDNA microarray images of different manufacturing techniques, were conducted so as to evaluate the performance of our proposed method in spot segmentation. Given that spot segmentation is the second stage in the process of microarray image analysis as the gridding stage precedes the spot-segmentation one [6], gridding is necessary to be executed before applying the spot-segmentation algorithm. Although any gridding technique could be used in combination with our spot-segmentation approach, in our conducted experiments, gridding has been performed as described in our past publication [36]. More precisely, in this preceding stage of gridding, a genetic algorithm determines near optimally the line segments constituting the borders of each compartment. The main reason for choosing our past method is that its performance achieves the high accuracy value of 95%, and it outperforms other methods.

Moreover, in all the experiments, the values of the  $N_{\text{pop}}$ ,  $P_r$ ,  $P_c$ ,  $P_m$ , and  $G_{\text{Fit}}$  parameters of the genetic algorithm have been set equal to their corresponding values of the genetic algorithm described in [36]. This can be explained by the fact that these parameters do not depend on the optimization problem at hand. The parameters mentioned in Sections IV-B and IV-C have been experimentally adjusted once. Actually, the constant  $k$  of 0.6 was adopted as the most appropriate for the definition of the  $\mu_B(p_r)$  and  $\mu_S(p_r)$  membership functions of a  $p_r$  pixel. The constant  $\lambda_m$  of 0.7 was adopted as the most suitable for the calculation of the model compartment's error on a specific pixel.

A constant  $E_{\text{MAX}}$  of 0.2 was adopted so as to control the maximum acceptable error of a model compartment on a specific pixel. A threshold  $\text{Th}$  of 0.15 was adopted in order to define the minimum rough estimation  $R_R(m)$  of an *acceptable* chromosome  $m$ . Since all of the above parameters were adjusted once, they remained constant during all experiments. Thus, the whole experimental procedure on the synthetic and real cDNA microarray images took place without any human intervention.

### A. Comparison of the Proposed Segmentation Approach With Established Methods Using Synthetic Microarray Images

In order to objectively compare the proposed segmentation method with established methods, a dataset of synthetic cDNA microarray images was used, which is available on the Internet [38]. This dataset contains 50 good-quality images (GQI) and 50 low-quality images (LQI) for which the ground truth is known. Each image is digitized at  $330 \times 750$  pixels, it contains 1000 spots, and it has been produced by the microarray simulator of Nykter, which generates synthetic cDNA microarray images with realistic characteristics [39]. The good-quality images have low variability in spot sizes and shapes, and their noise level is reasonably low. On the contrary, the low-quality images contain spots whose shape and size vary significantly. In addition, their noise level is significantly higher in the low-quality images.

The aforementioned dataset has already been used for the comparison of various established segmentation techniques (see Table I, first nine rows), by means of the statistical analysis described in [23]. In order to compare our method with the aforementioned techniques, we evaluated its efficiency by means of the same statistical analysis. According to this statistical analysis, we have examined the pixel-level accuracy of our segmentation's method by using the following two measures.

- 1) The first one is the probability of error PE which measures the missegmented pixels, and it is defined as

$$\text{PE} = P(F) \cdot P(B|F) + P(B) \cdot P(F|B) \quad (17)$$

where  $P(F)$  and  $P(B)$  are the *a priori* probabilities of foreground and background.  $P(F|B)$  denotes the probability of error in classifying background as foreground, while  $P(B|F)$  denotes the probability of error in classifying foreground as background.

- 2) The second one is the discrepancy distance  $D$ , which gives different weights for missegmented pixels, based on how spatially far they are located from the nearest correct segmentation result. It is defined as

$$D = \sqrt{\sum_{i=1}^N d^2(i)} / A \quad (18)$$

where  $N$  is the number of missegmented pixels,  $d(i)$  is the Euclidian distance from the  $i$ th missegmented pixels to the nearest pixel that actually belongs to the missegmented class.  $A$  is the total number of pixels in the image.

The evaluation results of the proposed method are shown in Table I (last row). It is evident that the proposed method can near optimally segment the spots of good-quality images while it can very efficiently segment the spots of low-quality images. In the same table we have included the results of nine established segmentation techniques (first nine rows), as they are reported by Lehmußola *et al.* [23]. By comparing the results, one can conclude that the proposed method is radically more successful than the first eight techniques, indicating its high performance. Compared with the K-means technique, in the case of HQI the proposed method yields the same results. However, in the case of LQI the proposed method is significantly superior. The significant number of spots which are contained in the used dataset additionally supports these arguments. Indeed, the evaluation of all methods has been statistically calculated in 50 000 artificial microarray spots for which the ground truth is given, which means that the correct segmentation result is known.

Fig. 11 illustrates the segmentation result of a microarray block taken from a good-quality synthetic image, while Fig. 12 illustrates the segmentation result of a microarray block taken from a low-quality synthetic image. On these segmentation results, one can observe that the proposed approach has near-optimally segmented all the microarray spots of Fig. 11 and most of the microarray spots of Fig. 12. Moreover, the proposed method has not segmented any spurious spot.

Figs. 13 and 14 illustrate four magnified spots isolated from a good and low synthetic microarray image respectively. They also depict their spot models determined by the genetic algorithm, as well as their segmentation results, produced by our proposed method. It is obvious that the proposed approach has near optimally segmented these microarray spots. Moreover, this example demonstrates that the proposed segmentation algorithm is able to segment not only circular spots but also spots of various shapes (Fig. 14).

### B. Evaluation of the Performance Using Real Microarray Images

In order to evaluate the performance of our segmentation method, on real cDNA microarray images, we used the Stanford Microarray Database (SMD) [40], which is publicly available.

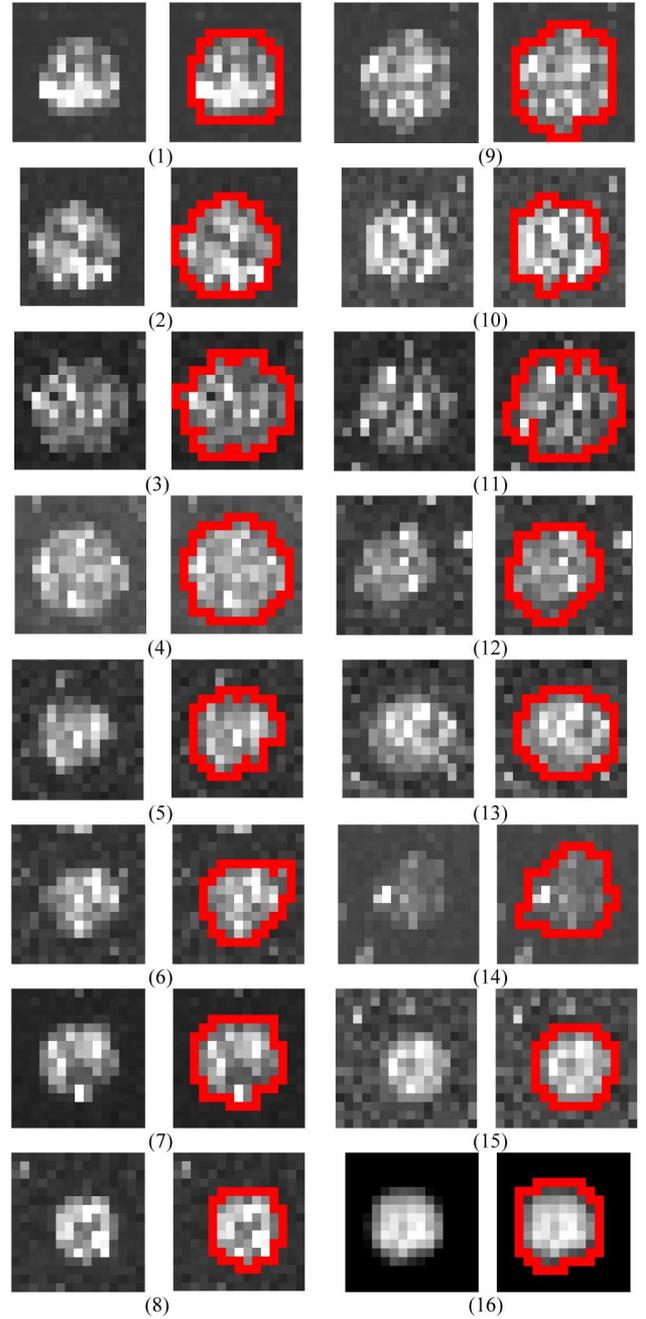


Fig. 16. Spot-segmentation results of several magnified real microarray compartments.

These microarray images have been produced by comprehensively analyzing the gene expression profiles in 54 specimens of acute lymphoblastic leukemia, 37 positive and 17 negative to BCR-ABL [2]. BCR-ABL is a fusion gene product resulting from translocation between the 9th and the 22nd chromosomes. From these images we have arbitrary selected ten microarray blocks. Each block is digitized at  $\sim 450 \times 450$  pixels at 16-bit gray level depth, and it contains 864 spots. Thus, these blocks contain 8640 spots in total.

Fig. 15 illustrates the segmentation results of a region of a poor quality image obtained from the Stanford Microarray Database. This image contains several spots whose intensity value is low while the background level and noise is high enough. In

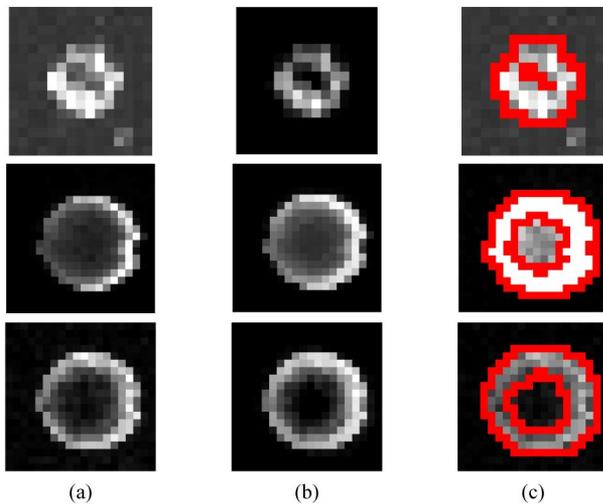


Fig. 17. Segmentation results of three magnified real spots whose shape resembles a volcano (first line) or a doughnut curve (2nd and 3rd lines). (a) Two real spots, with (b) their spot models and (c) their segmentation results.

order for the reader to be able to observe the results even for the low-intensity spots, we have increased the brightness. On this image, it is apparent that the proposed method has very efficiently segmented the real microarrays spots. It should be noted that the proposed method segmented all the low-intensity spots, even those which are not easily distinguished. However, in two cases, the proposed method segmented two spurious spots.

Fig. 16 illustrates 16 magnified microarray compartments which have been isolated from real microarray blocks obtained from the Stanford Microarray Database. The first 15 compartments contain very low intensity spots, while the last one contains a high-intensity spot. Although the ground truth is not known, it is clear that the proposed method has very efficiently segmented the real microarrays spots.

Fig. 17 illustrates the spot models and the segmentation results of three real magnified spots of different shapes. The real spots have been isolated from a real microarray image and their shapes resemble a volcano (1st line) or a doughnut 3-D curve (2nd and 3rd lines). The proposed method has very efficiently represented the real spots with the spot models and therefore has very efficiently segmented the real microarrays spots.

## VI. CONCLUSION

Spot segmentation in microarray images is undoubtedly one of the most challenging stages of the microarray image analysis sequence. In this paper, the segmentation procedure is a result of an optimization problem which is tackled by using a genetic algorithm, which represents, in a 3-D space, the real spots of the microarray image with spot models. To this end, fuzzy logic is adopted in order to take into account the uncertainty existing in the pixels' intensities, which can be caused by noise, artifacts, and/or uneven background. The segmentation of the real spots is conducted by drawing the contours of the spot models.

The proposed approach is robust, very accurate, and efficient under the following adverse conditions: i) the appearance of various spot shapes, such as peak-shaped spots, volcano-shaped spots and doughnut-shaped spots; ii) the appearance of spots

of diverse intensities, such as low-intensity spots which are not clearly visible or saturated spots; and iii) the appearance of various spot sizes. Additionally, it is fully automatic since all needed parameters have been adjusted once and they have been kept fixed during all the experiments. Consequently, there is no requirement of any input parameter or human intervention in order to determine properly the contours of microarray spots. The experimental results over synthetic and real images firmly confirm the validity of our method, as well as its robustness and effectiveness.

## REFERENCES

- [1] Y. F. Leung and D. Cavalieri, "Fundamentals of cDNA microarray data analysis," *Trends Genetics*, vol. 19, no. 11, pp. 649–659, Nov. 2003.
- [2] D. Juric *et al.*, "Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias," *J. Clin. Oncol.*, vol. 25, no. 11, pp. 1341–1349, Apr. 2007.
- [3] Y. A. Su *et al.*, "Dysregulated mitochondrial genes and networks with drug targets in postmortem brain of patients with Posttraumatic Stress Disorder (PTSD) revealed by human mitochondria-focused cDNA microarrays," *Int. J. Biol. Sci.*, vol. 4, pp. 223–235, 2008.
- [4] S. L. Byrne, K. Durandea, I. Nagy, and S. Barth, "Identification of ABC transporters from *Lolium perenne* L. that are regulated by toxic levels of selenium," *Planta*, 2010.
- [5] E. K. Lobenhofer, P. R. Bushel, C. A. Afshari, and H. K. Hamadeh, "Progress in the application of DNA microarrays," *Environ. Health Perspectives*, vol. 109, no. 9, pp. 881–891, Sep. 2001.
- [6] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. Speed, "Comparison of methods for image analysis on cDNA microarray data," *J. Comp. Graph. Stat.*, vol. 11, no. 1, pp. 108–136, 2002.
- [7] X. H. Wang and R. S. H. Istepanian, "Microarray image enhancement by denoising using stationary wavelet transform," *IEEE Trans. NanoBiosci.*, vol. 2, no. 4, pp. 184–189, Dec. 2003.
- [8] M. B. Eisen, 1999, ScanAlyze [Online]. Available: <http://rana.lbl.gov/EisenSoftware.htm>
- [9] J. Buhler, T. Ideker, and D. Haynor, Dapple: Improved techniques for finding spots on DNA microarrays, UW CSE Tech. Rep. UWTR 2000-08-05, Aug. 2000, pp. 1–12.
- [10] P. Sarder, A. Nehorai, P. H. Davis, and S. L. Stanley, "Estimating gene signals from noisy images," *IEEE Trans. NanoBiosci.*, vol. 7, no. 2, pp. 142–153, June 2008.
- [11] M. J. Buckley, 2000, The Spot User's Guide, CSIRO Mathematical and Information Sciences [Online]. Available: <http://www.cmis.csiro.au/IAP/Spot/spotmanual.htm>
- [12] D. Bozinov and J. Rahnenfuhrer, "Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering," *Bioinformatics*, vol. 18, no. 5, pp. 747–756, 2002.
- [13] J. Rahnenfuhrer and D. Bozinov, "Hybrid clustering for microarray image analysis combining intensity and shape features," *BMC Bioinformatics*, vol. 5, no. 5, pp. 47–58, Apr. 2004.
- [14] R. Nagarajan, "Intensity-based segmentation of microarray images," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 882–889, Jul. 2003.
- [15] R. Nagarajan and C. A. Peterson, "Identifying spots in microarray images," *IEEE Trans. NanoBiosci.*, vol. 1, no. 2, pp. 78–84, 2002.
- [16] Q. Li, C. Fraley, R. E. Bumgarner, K. Y. Yeung, and A. E. Raftery, "Donuts, scratches and blanks: Robust model-based segmentation of microarray images," *Bioinformatics*, vol. 21, no. 12, pp. 2875–2882, Apr. 2005.
- [17] J. Ho and W. L. Hwang, "Automatic microarray spot segmentation using a snake-fisher model," *IEEE Trans. Med. Imag.*, vol. 27, no. 6, pp. 847–857, Jun. 2008.
- [18] T. Srinark and C. Kambhamettu, "A microarray image analysis system based on multiple snakes," *J. Biol. Syst.*, vol. 12, no. 25, Jun. 2004.
- [19] T. Srinark and C. Kambhamettu, "A framework for multiple snakes," *Proc. Comp. Vis. Pattern Recogn.*, vol. 2, pp. 202–209, 2001.
- [20] M. Katzer, F. Kummert, and G. Sagerer, "Methods for automatic microarray image segmentation," *IEEE Trans. NanoBiosci.*, vol. 2, no. 4, pp. 202–212, 2003.
- [21] O. Demirkaya, M. H. Asyali, and M. M. Shoukri, "Segmentation of cDNA microarray spots using markov random field modeling," *Bioinformatics*, vol. 21, no. 13, pp. 2994–3000, Apr. 2005.

- [22] R. Lukac, K. N. Plataniotis, B. Smolka, and A. N. Venetsanopoulos, "A multichannel order-statistics technique for cDNA microarray image processing," *IEEE Trans. NanoBiosci.*, vol. 3, no. 4, pp. 202–214, Dec. 2004.
- [23] A. Lehmussola, P. Ruusuvaari, and O. Yli-Harja, "Evaluating the performance of microarray segmentation algorithms," *Bioinformatics*, vol. 22, no. 23, pp. 2910–2917, Oct. 2006.
- [24] R. Nagarajan and M. Upreti, "Correlation statistics for cDNA microarray image analysis," *IEEE Trans. NanoBiosci.*, vol. 3, no. 3, pp. 232–238, Sep. 2006.
- [25] R. Gottardo *et al.*, "Probabilistic segmentation and intensity estimation for microarray images," *Biostatistics*, vol. 7, no. 1, pp. 85–99, 2006.
- [26] X. Wang, S. Ghosh, and S.-W. Guo, "Quantitative quality control in microarray image processing and data acquisition," *Nucleic Acids Res.*, vol. 29, no. 15, 2001.
- [27] A. A. Ahmed, M. Vias, N. G. Iyer, C. Caldas, and J. D. Brenton, "Microarray segmentation methods significantly influence data precision," *Nucleic Acids Res.*, vol. 32, no. 5, 2004.
- [28] D. K. Iakovidis, D. Maroulis, E. Zacharia, and S. Kossida, "A genetic approach to spot detection in two-dimensional gel electrophoresis images," presented at the Int. Conf. Inf. Technol. Biomed., Ioannina, Greece, Oct. 2006.
- [29] E. Zacharia and D. Maroulis, "A genetic approach to cDNA Microarray image Analysis," in *Proc. 11th Panhellenic Conf. Informatics*, Patras, Greece, 2007, pp. 383–392.
- [30] E. Zacharia and D. Maroulis, "An unsupervised and fully-automated image analysis method for cDNA microarrays," in *Proc. IEEE Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Maribor, Slovenia, 2007, pp. 389–394.
- [31] E. Zacharia and D. Maroulis, "Microarray image analysis based on an evolutionary approach," presented at the 19th Int. Conf. Pattern Recogn. (ICPR), Tampa, FL, 2008.
- [32] H. Y. Kim *et al.*, "Characterization and simulation of cDNA microarray spots using a novel mathematical model," *BMC Bioinformatics*, vol. 8, pp. 485–496, March 2007.
- [33] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning*. Reading, MA: Addison-Wesley, 1989, ch. 1.
- [34] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.
- [35] F. Herrera, M. Lozano, and J. L. Verdegay, "Tackling real coded genetic algorithms: Operators and tools for behavioural analysis," *Artif. Intell. Rev.*, vol. 12, no. 4, pp. 265–319, Nov. 1998.
- [36] E. Zacharia and D. Maroulis, "An original genetic approach to the fully-automatic gridding of microarray images," *IEEE Trans. Med. Imag.*, vol. 27, no. 6, pp. 805–813, Jun. 2008.
- [37] Y. Chen *et al.*, "Ratio-based decisions and the quantitative analysis of cDNA microarray images," *J. Biomed. Opt.*, vol. 2, pp. 364–374, 1997.
- [38] [Online]. Available: <http://www.cs.tut.fi/sgn/csb/spotseg/>
- [39] M. Nykter *et al.*, "Simulation of microarray data with realistic characteristics," *BMC Bioinformatics*, vol. 7, pp. 349–366, Jul. 2006.
- [40] Stanford Microarray Database [Online]. Available: <http://genome-www5.stanford.edu/>

**Dimitris Maroulis** received the B.Sc. degree in physics, the M.Sc. degree in radioelectricity, the M.Sc. degree in cybernetics with honors, and the Ph.D. degree in computer science from the University of Athens, Greece.

He was a Research Fellow for three years at the Meudon Observatory, Department of Spatial Sciences, France. Currently, he is an Associate Professor in the Department of Informatics and Telecommunications, University of Athens. He has coauthored more than 130 research papers on image/signal processing and analysis, real-time systems, and biomedical applications, which have been cited more than 500 times. He is a reviewer for more than 10 international journals, including the IEEE TRANSACTIONS ON MEDICAL IMAGING and the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE. He has been actively involved in more than 15 European and national R&D projects. His research interests include data acquisition and real-time systems, image/signal processing and analysis, and pattern recognition with applications on biomedical systems and bioinformatics.

**Eleni Zacharia** received the B.Sc. degree in informatics and telecommunications, the M.Sc. in signal processing and multimedia with honors, and the Ph.D. degree in computer science from the Department of Informatics and Telecommunications, University of Athens, Greece, in 2004, 2006, and 2009, respectively.

She is currently a Research Fellow at the University of Athens. She is a regular speaker at prestigious international bioinformatics conferences. Her articles have been published in leading international journals, including the IEEE TRANSACTIONS ON MEDICAL IMAGING. Her research interests include image/signal processing and analysis with applications on biomedical systems and bioinformatics.