

Microarray Image Segmentation using Spot Morphological Model

D. Maroulis, *Member, IEEE*, and E. Zacharia

Abstract—The up-to-date segmentation techniques and software programs for microarray image segmentation require human intervention which in turn may detrimentally affect the biological conclusions reached during microarray experiments. In this paper, an automatic approach for segmenting microarray images, based on the morphological modeling of spots, is presented. The conducted experiments have shown that the proposed approach is very effective even when it is applied to noisy images as well as to images containing spots of various shapes and intensities.

Index Terms— *Microarrays, Image Analysis, Segmentation*

I. INTRODUCTION

IN the last decade, microarray technology has been rendered a revolutionary tool for biomedical research as it enables the simultaneous monitor of the expression levels of thousands of genes [1]. The end product of a microarray experiment is a high resolution digital image, containing thousands of spots, whose intensities are proportional to the expression levels of specific genes. Consequently, the intensity of each microarray spot should be calculated.

The main process for measuring spot intensity includes three stages [2]: In the first stage, called gridding, the position of each spot is assigned; compartments of the image containing one individual spot and background are therefore defined. In the second stage, called spot-segmentation, the contour of each spot is determined. In the third stage, called intensity extraction, the intensity of each spot is calculated.

From the aforementioned stages, the spot-segmentation is the most challenging one. Consequently, several techniques and software programs have been proposed for this stage [3]. Amongst them, there are methods which are based on the assumption that the spot has a circular shape. For example, ScanAlyze [4], and Dapple [5] software programs match circular templates to the spots. Sarder et al [6] have also proposed a method in which each spot is represented by a parametric circle with one or two elliptical center holes. Given that the shape of spots is not always circular, the aforementioned methods fail to segment non-circular spots. To deal with the various shapes of spots, the seeded region growing algorithm [7],[8] which can segment regions of irregular shapes by implementing a watershed algorithm has

been suggested. However, the performance of this method is based on the appropriate selection of the starting points.

Intensity-based segmentation techniques have been also developed. Amongst them, there are clustering techniques such as k-means [9], hybrid k-means [10], and the model-based segmentation algorithm [11]. The former two algorithms may fail to segment poorly contrasted spots, while the latter one is based on a threshold value. More sophisticated methods are based on active contours [12], and multiple snakes [13],[14]. However, these methods give inaccurate results in the case that the image contains spots with large variations or when the compartment is contaminated with noise and artifacts.

Overall, the current microarray segmentation techniques have several limitations whose main cause lies to the poor quality of microarray images [15]. Indeed, these images are contaminated with noise, and artifacts. Moreover, real spots vary significantly from the ideal ones; they have not always a circular shape and their intensity is not always high enough to be clearly visible. As a result, segmentation techniques require human intervention in order to specify mandatory parameters or to correct their results. However, this lack of automation can significantly affect biological conclusions [16] and it is therefore their main disadvantage.

In this paper, a precise and automatic spot segmentation technique for microarray image is presented. The proposed approach optimally represents the morphological shape of a real-spot with a spot-model. Consequently, the contour of a microarray spot is depicted in the image plane by drawing the contour of spot-model. The proposed approach improves the ones reported in [17],[18] since it can effectively segment all morphological types of microarray spots (Fig. 1). For this scope, an original spot-model is presented which can be used for the representation of all types of real microarray compartments. Experiments - over microarray images containing thousands of spots - showed that the proposed approach achieves an accuracy of more than 91% in contrast to the corresponding accuracy of 85.3% found in our previous approaches. Moreover, the experiments have shown that the proposed approach can be very effectively applied to images contaminated with noise as well as to images containing low-intensity spots.

The remainder of this paper is structured as follows: In section II, the proposed methodology for segmenting microarray images is described, while in section III the results of the proposed approach are presented. Finally, in section IV the conclusions of this study are summarized.

Manuscript received July 3, 2009. This research was partially funded by the special account of research grants of the University of Athens, Greece.

D. Maroulis, and E. Zacharia are members of the Department of Informatics and Telecommunications, University of Athens, GR 15784 Panepistimiopolis, Ilisia, Greece (emails: dmaroulis@di.uoa.gr, eezacharia@gmail.com).

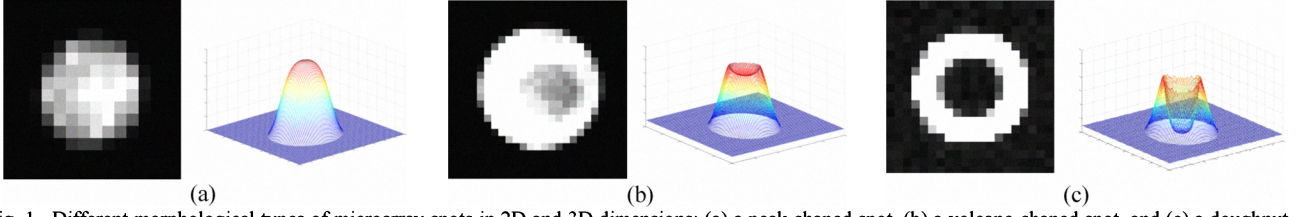


Fig. 1. Different morphological types of microarray spots in 2D and 3D dimensions: (a) a peak-shaped spot, (b) a volcano-shaped spot, and (c) a doughnut-shaped spot.

II. METHODOLOGY

According to Kim et al [19], microarray spots can be classified into three categories based on their shapes: peak-shaped, volcano-shaped, and doughnut-shaped spots (Fig.1).

Based on the above remark, our proposed segmentation method is conducted into two stages: (i) Firstly, the morphological spot shape is represented by a spot-model, and (ii) Secondly, the spot contour is depicted in the image plane by drawing the contour of its spot-model.

A. Morphological Models for a Microarray Spot and its Compartment

All the aforementioned spots categories can be represented using: (i) a 3D-curve representing the main-body S_{MB} of the spot-model, and (ii) a 3D-curve representing the inner-dip S_{ID} of the spot-model.

Both the main-body and the inner-dip 3D curves are expressed as the diffusion function $C(x,y)$ proposed by Bettens et al [20], and therefore resemble the 3D Gaussian or plateau curve. Moreover, their orientation is opposite; the base of the main-body of the spot-model is down and its peak is up, while the base of the inner-dip of the spot-model is up and its peak is down (Fig.2).

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

$$S_{MODEL}(x,y) = \text{Min}[S_{MB}(x,y), S_{ID}(x,y)] . \quad (1)$$

A graphical explanation of eq. 1 is depicted in Fig. 2. The resulting total-models (grey areas) depend on the 3D curves of their corresponding S_{MB} and S_{ID} components. More precisely, in the case of the distance between the S_{MB} and S_{ID}

centers being large, the resulting total-model resembles a peak-shaped spot (Fig.2a). In the case of the distance between the S_{MB} and S_{ID} centers being small, the resulting total-model resembles a volcano-shaped spot (Fig.2b) or a doughnut-shaped spot (Fig.2c), according to the height of the S_{ID} 3D curve.

Likewise, the morphological compartment-model can be defined as:

$$I_{MODEL}(x,y) = \text{Max}[B_{AV}, S_{MODEL}(x,y)] \quad (2)$$

where B_{AV} denotes the average background intensity of the compartment-model and it 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} . A graphical explanation of eq. 2 for a volcano-shaped spot is depicted in Fig. 3.

B. Optimum Spot-Model Representation and Definition of Real-Spot Contour

A genetic algorithm determines the compartment-model which optimally represents the real-one. For this scope, it searches for the optimal values of the parameters of the morphological compartment-model defined by (2). Its search starts by the generation of an initial population (Pop_1) of N chromosomes. Each chromosome m represents a morphological compartment-model I_{MODEL}^m . Consequently, it is encoded as a numerical sequence consisting of three segments: The first segment encodes the value of the average background intensity B_{AV}^m of the compartment-model. The second segment encodes the values of the variables of the main-body S_{MB}^m , while the third segment encodes the values of the variables of the inner-dip S_{ID}^m of

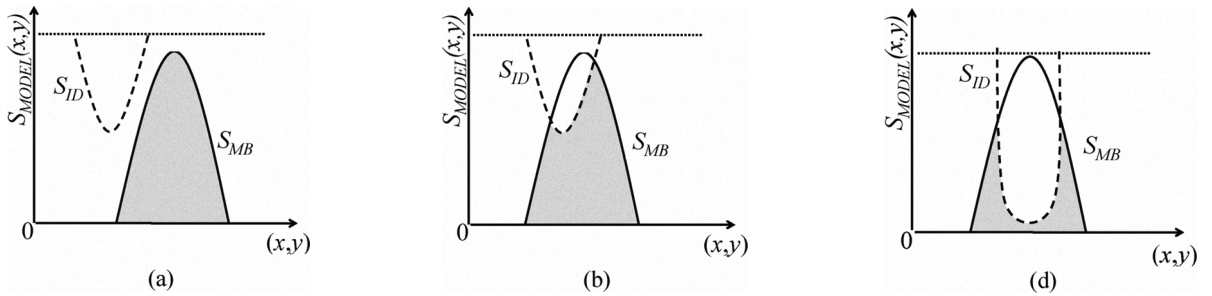


Fig. 2. S_{MB} and S_{ID} components of the morphological models of: (a) a peak-shaped spot, (b) a volcano-shaped spot, (c) a doughnut-shaped spot. The total morphological models are the grey areas.

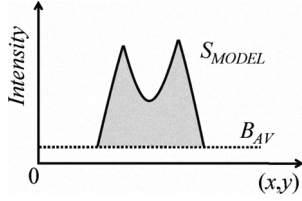


Fig. 3. Morphological model for a compartment containing a volcano-shaped spot.

the spot-model S_{MODEL}^m .

Subsequently, the chromosomes constituting the Pop_1 are evaluated. The higher the resemblance of the morphological compartment-model I_{MODEL}^m (represented by the chromosome m) to the real-compartment I_{REAL} is, the higher the value of the fitness function of a chromosome m becomes. Therefore, the relative intensity error $E(x,y)$ between the intensity of the compartment-model's pixel (x,y) and the corresponding one of the real-compartment's pixel is calculated as follows:

$$E(x,y) = \frac{|I_{MODEL}^m(x,y) - I_{REAL}(x,y)|}{I_{REAL}(x,y)}. \quad (3)$$

The fitness function $F_S(m)$ of a chromosome m is defined by the following equation:

$$F_S(m) = - \iint_{x,y \in \text{Compartment}} E(x,y) \, dx dy. \quad (4)$$

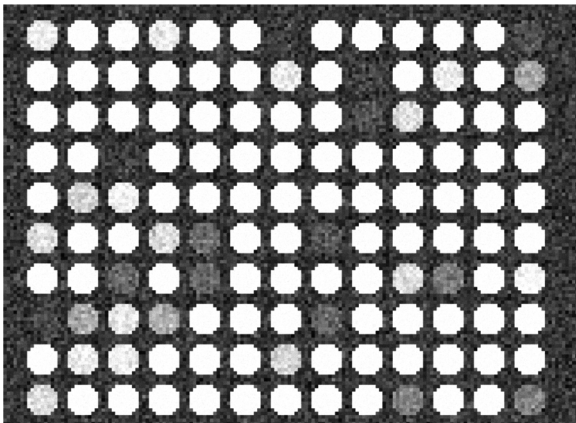
Thereafter, the genetic algorithm makes the population Pop_1 evolve into a new population Pop_2 , the Pop_2 to a new population Pop_3 and so forth as follows: The $P_r\%$ of the best chromosomes of Pop_n is maintained in Pop_{n+1} . The rest are reproduced by applying: i) the joint application of the *BLX- α* crossover and of the dynamic heuristic one and ii) the wavelet mutation. Genetic algorithm finishes when a maximum number of populations is reached, for which the best fitness value has remained unchanged. Thus, the

optimum compartment model with the higher resemblance to the real-one is determined. The real-spot is obtained by drawing the contour of this optimum spot-model.

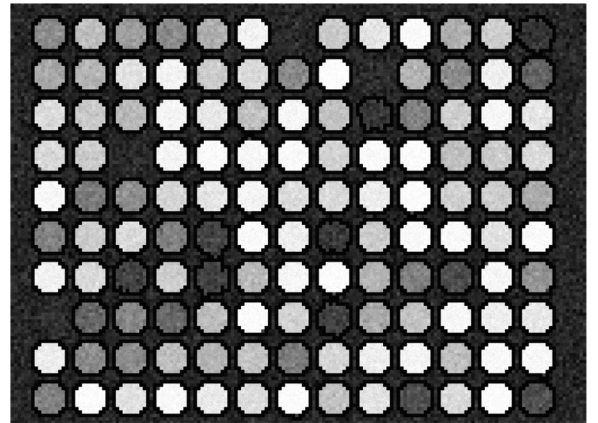
III. RESULTS

Several experiments were performed in order to evaluate our proposed approach for spot-segmentation. The microarray images used for the evaluation were obtained on the Internet. They have been produced by the microarray simulator of Nykter which generates synthetic microarray images with realistic characteristics and varied quality. Half of the images have good quality; their noise level is reasonably low, and they have low variability in spot sizes and shapes. The rest of them are low quality images; their noise level is high enough, and they contain spots whose shape and size vary substantially. Finally, each image is digitized at 330×750 pixels and contains 1000 spots. The parameters of the genetic algorithm had been experimentally adjusted once, and thus they remained unchanged during all the experimental procedure. The population size N was set to 100. The percentage P_r was relatively small ($P_r=10\%$) as the reproduction was used only for the best chromosomes of the population to be preserved in the next population. The termination criterion was satisfied when the best fitness value remained unchanged for 200 populations.

The accuracy of the proposed method was analyzed by means of a statistical analysis. More precisely, using the proposed approach, 91.5% of spots were “very efficiently segmented”, and no spurious spot were detected. A spot was “very efficiently segmented” if at least 90% of the entire spot area was enclosed in the contour of that spot. Fig. 4 illustrates the segmentation result of a microarray block taken from a good-quality synthetic image. In this figure, one can observe that the proposed method has optimally segmented all the microarray spots. On the other hand, Fig. 5 illustrates the segmentation result of a microarray block taken from a low-quality synthetic image. On its segmentation results (Fig. 5b), it becomes obvious that the



(a)



(b)

Fig. 4. Spot segmentation results (b) in a good quality microarray subimage (a).

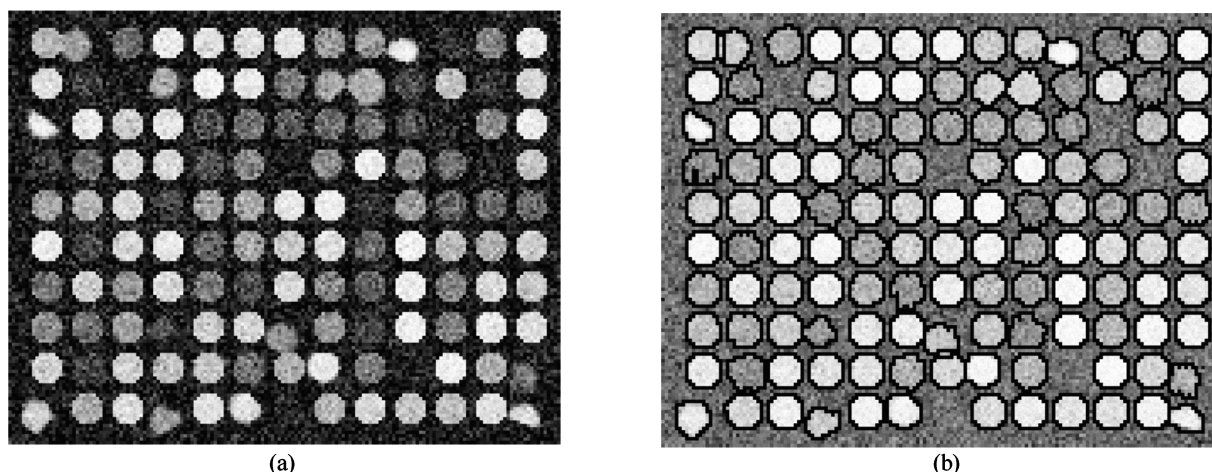


Fig. 5. Spot segmentation results (b) in a low quality microarray image (a).

proposed method has optimally segmented nearly all its microarray spots.

IV. CONCLUSIONS

In this paper an original approach for segmenting microarray images is described. The proposed approach is based on the morphological shape of the microarray spots. Its key concept is based on the idea that each microarray spot can be represented with a morphological spot-model, and therefore its contour can be depicted in the image plane by depicting the contour of its morphological spot-model.

The proposed approach is automatic as all the needed parameters were experimentally adjusted once and kept stable during all experiments. Moreover, it can be applied to images containing missing spots or containing spots of various intensities, sizes and shapes. The experimental results on synthetic microarray images - in which the ground truth is known - confirm the validity and the effectiveness of our method.

REFERENCES

- [1] Y. F. Leung, and D. Cavalieri, "Fundamentals of cDNA microarray data analysis," *Trends in Genetics*, vol. 19, no. 11, pp. 649-659, 2003.
- [2] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. Speed, "Comparison of methods for image analysis on cDNA microarray data," *J. of Comp. & Graphical Statistics*, vol. 11, no. 1, pp. 108-136, 2002.
- [3] P. Bajcy, "An overview of DNA microarray grid alignment and foreground separation approaches," *EURASIP Journal on Applied Signal processing*, vol. 2006, pp. 1-13, 2006.
- [4] M. B. Eisen. (1999). ScanAlyze. [Online]. Available: <http://rana.lbl.gov/EisenSoftware.htm>
- [5] J. Buhler, T. Ideker, and D. Haynor, "Dapple: improved techniques for finding spots on DNA microarrays," UW CSE Technical Report UWTR 2000-08-05, pp. 1-12, Aug. 2000.
- [6] P. Sarder, A. Nehorai, P. H. Davis, and S. L. Stanley, "Estimating Gene Signals From Noisy Images," *IEEE Trans. on Nanobioscience*, vol. 7, no. 2, pp. 142-153, June 2008.
- [7] R. Adams, and L. Bischof, "Seeded Region Growing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641-647, Jun. 1994.
- [8] Buckley MJ (2000) The Spot User's Guide, CSIRO math. and inf. sciences. <http://www.cmis.csiro.au/IAP/Spot/spotmanual.htm>
- [9] D. Bozinov, and J. Rahnenfuhrer, "Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering," *Bioinformatics*, vol. 18, pp. 747-756, 2002.
- [10] J. Rahnenfuhrer, and D. Bozinov, "Hybrid clustering for microarray image analysis combining intensity and shape features," *BMC Bioinformatics*, vol. 5, no. 5, p. 47-58, Apr. 2004.
- [11] 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, pp. 2875-2882, 2005.
- [12] M. Katzer, F. Kummert, and G. Sagerer, "Methods for automatic microarray image segmentation," *IEEE Trans. on Nanobioscience*, vol. 2, no. 4, pp. 202-212, 2003.
- [13] J. Ho, and W. L. Hwang, "Automatic Microarray Spot Segmentation using a snake-fisher model," *IEEE Trans. on Medical Imaging*, vol. 27, no. 6, pp. 847-857, Jun. 2008.
- [14] T. Srinark, and C. Kambhamettu, "A Microarray Image Analysis System Based on Multiple Snakes," *Journal of Biological Systems*, vol. 12, no. 25, Jun. 2004.
- [15] X. H. Wang, and R. S. H. Istepanian, "Microarray image enhancement by denoising using stationary wavelet transform," *IEEE Trans. on Nanobioscience*, vol. 2, no. 4, pp. 184-189, Dec. 2003.
- [16] A. A. Ahmed, M. Vias, N. G. Iyer, C. Caldas, and J. D. Brenton, "Microarray segmentation methods significantly influence data precision," *Nucleic Acids Research*, vol. 32, no. 5, 2004.
- [17] E. Zacharia, and D. Maroulis, "An Unsupervised and Fully-Automated image Analysis Method for cDNA Microarrays," in *Proc. of IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, Maribor, Slovenia, 2007, pp. 389-394.
- [18] E. Zacharia, and D. Maroulis, "Microarray Image Analysis based on an Evolutionary Approach," in *Proc of the 19th International Conference on Pattern recognition (ICPR)*, Tampa, Florida, USA, 2008.
- [19] 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.
- [20] E. Bettens *et al.*, "Computer analysis of two dimensional electrophoresis gels: A new segmentation and modelling algorithm," *Electrophoresis*, vol. 18, no. 5, pp. 792-798, May 1997.