31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Microarray Image Segmentation Using Chan-Vese Active Contour Model and Level Set Method

Kaustubha A. Mendhurwar, Rajasekhar Kakumani, and Vijay Devabhaktuni, *Member*, *IEEE*

Abstract—**Microarray technology is considered to be one of the major breakthroughs in bioinformatics for profiling gene-expressions of thousands of genes, simultaneously. Analysis of a microarray image plays an important role in the accurate depiction of gene-expression. Segmentation, the process of separating the foreground from the background, of a microarray image, is one of the key issues in microarray image analysis. Level sets have tremendous potential in the segmentation of images. In this paper, a new approach for segmentation of the microarray images is proposed. In this work, Chan-Vese approximation of the Mumford-Shah model and the level set method are employed for image segmentation. Illustrative examples of the proposed method are presented highlighting its effectiveness.**

Index Terms— **Chan-Vese approximation model, level sets, microarray image, segmentation.**

I. INTRODUCTION

Microarray technology has created a tremendous impact in bioinformatics over the last decade by its unprecedented capacity to monitor expression levels of thousands of genes simultaneously. This technique has evolved multifold from its introduction [1], and has now become a high-throughput technology to simultaneously measure RNA abundances of tens of thousands of mRNAs (messenger ribonucleic acid). As such, accurate analysis of the spots on the microarrays has gained a lot of significance.

The analysis of DNA microarray mainly consists of two steps: image quantization and gene-expression data analysis [1]. The quantization step deals with the extraction of gene expression data. In the second step, the gene expression data is analyzed employing several statistical methods. It is vital in identifying genes that behave differently in different cell populations. It is handy in comparisons like tumor cells *vs* healthy cells, brain cells *vs* liver cells, and/or same tissue in different organisms, *etc*. To achieve accurate results from the second step *i.e.* for better analysis of the microarray data, first step *i.e.* image quantization is very crucial. Microarray image quantization is a multi phase process that consists of steps like denoising, gridding and segmentation.

Microarray image segmentation is one of the key issues in the image quantization process. Numerous approaches are available in the literature. Fixed circle approach is one of the first techniques used for the segmentation. It assigns a circle of a fixed radius to all the spots, and is implemented in the software ScanAlyze [2]. Few other softwares like, GenePix [3] and ScanArray Express [4], also provide the option for fixed circle method. The radius of circle has been treated as a parameter and adaptive circle method, has been proposed. This method is available in GenePix, ScanAlyze, ScanArray Express, Imagene, and Dapple [5]. Subsequently, adaptive ellipse method too has been already proposed [6]. Histogram approach has also been used to classify a pixel into foreground/background. Clustering has also been used owing to the fact that it is not restricted to a particular shape and size of the spots. However, clustering methods are prone to include noisy pixels in the foreground, producing incorrect quantization measures for the spots.

In this paper, Chan-Vese approximation of the Mumford-Shah model [7] and the level set method is employed for the segmentation purpose. The method is illustrated using a few examples. Section II of the paper introduces the concept of microarray image briefly. Section III addresses the proposed segmentation approach. Illustrative examples are presented in section IV, and finally section V contains conclusion.

II. MICROARRAY IMAGE

Microarrays are grids of thousands of dissimilar single-stranded DNA molecules, immobilized on a surface, to serve as probes. Two major kinds of microarrays are those using synthesized oligonucleotides and those using spotted cDNAs (complementary-DNA molecules). Experimental procedure resulting in the formation of microarray image is described in [6]. RNA is first extracted from cells and converted to single-stranded cDNAs. Fluorescent labels are then attached to the cDNAs and afterwards they are allowed to hybridize, on a microarray, with their respective complimentary probes. The resulting fluor-tagged hybrids are detected *via* excitation of the attached fluors, and image formation is carried out using a scanning confocal microscope. Relative RNA abundance is obtained using the measurement of signal intensity from the attached fluors. This intensity is obtained by image processing. The resulting microarray image intensities are subject to statistical analysis with particular attention towards the detection of high/low expressing genes, expression based phenotype classification,

---

Kaustubha A. Mendhurwar and Rajasekhar Kakumani are with the Faculty of Engineering and Computer Science, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, H3G1M8, Quebec, Canada.

Vijay Devabhaktuni is with the Department of Electrical Engineering and Computer Science, University of Toledo, MS 308, 2801 W. Bancroft St., Toledo, OH 43606. (e-mail: Vijay.Devabhaktuni@utoledo.edu).

and the discovery of the multivariate inter-gene predictive relationships. As such, microarrays have truly revolutionized molecular biology research, and genomic clinical diagnosis.

III. PROPOSED SEGMENTATION APPROACH

Many of the segmentation methods employed might not perform well when the image is noisy and/or rotated. In such scenarios, prior to segmentation a denoising algorithm needs to be employed. Moreover, the available software for segmentation require input parameters, such as, number of rows and columns in an image, number of spots in each row and column, and, end-coordinates of the rectangle that is enclosing the spots, *etc*. Gridding is required, for the fixed and/or adaptive circle, and ellipse method. The proposed segmentation approach does not require any of the above mentioned steps, including denoising of the input noisy image and correcting the rotation.

In this section, one of the techniques of active contours, which has gained a lot of importance for a variety of signal processing applications, like, segmentation, motion tracking, *etc.*, is discussed. The basic idea of this methodology is to detect the boundary of the object of interest by starting with a contour around it and moving the contour normally inward until it assumes the required boundary. There are two kinds of active contour models for image segmentation based on the force evolving the respective contours: edge-based and region-based. Edge-based active contours normally use the gradient information, to find the boundaries of sub-regions and to attract the contours to the detected boundaries. Region-based active contours use the statistical information of image intensity within each sub-region instead of searching for geometrical boundaries.

A formulation, to implement these active contours, was first proposed using level set theory [8]. In this framework, a contour is implicitly represented using a two-dimensional Lipschitz function, $\phi(x,y):\Omega \to \Re$, defined on the image plane Ω. The function, $\phi(x,y)$, is called level set function, and the contour C, is normally represented, using the zero-level, of $\phi(x,y)$ *i.e.*

$$C \equiv \{(x,y):\phi(x,y)=0\}, \forall (x,y) \in \Omega. \tag{1}$$

For the region inside a contour, $\phi(x,y) > 0$, and for the region outside, $\phi(x,y) < 0$ (see Fig. 1). The evolution of the contour is equivalent to the evolution of the level set function, *i.e.* $\partial C / \partial t = \partial \phi(x,y)/\partial t$. The evolution of the contour at any time *t* is given by zero-level function $\phi(t,x,y)$. The initial contour C_0 may be given as,

$$C_0 \equiv \{(x,y):\phi(0,x,y)=\phi_0(x,y)=0\}, \forall (x,y) \in \Omega. \tag{2}$$

The evolution of the contour in normal direction with speed F is generally represented as a partial differential equation and is given as,

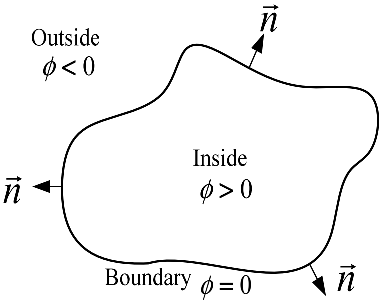


Figure 1. An evolving contour.

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| F, \; \phi(0,x,y) = \phi_0(x,y). \tag{3}$$

A formulation of contour evolution by mean curvature when

$$F = \operatorname{div}\left(\frac{\nabla \phi(x,y)}{\| \nabla \phi(x,y) \|} \right), \tag{4}$$

which is the curvature of the level-curve $\phi(x,y)$, passing through (x,y), becomes

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = |\nabla \phi| \operatorname{div}(\dfrac{\nabla \phi}{\| \nabla \phi \|}), \; t \in (0,\infty), \forall (x,y) \in \Omega \\ \phi(0,x,y) = \phi_0(x,y), \forall (x,y) \in \Omega \end{cases} \tag{5}$$

Another active contour model based on the mean curvature motion is given by

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = g(|\nabla u_0|) |\nabla \phi| \operatorname{div}(\dfrac{\nabla \phi}{\| \nabla \phi \|}), \; t \in (0,\infty), \forall (x,y) \in \Omega \\ \phi(0,x,y) = \phi_0(x,y), \forall (x,y) \in \Omega \end{cases} \tag{6}$$

where, g is an edge detector given by

$$g(|\nabla u_0(x,y)|) = \frac{1}{1 + |\nabla G_\sigma(x,y) * u_0(x,y)|^2}, \tag{7}$$

and $G_\sigma * u_0$ is convolution of the image u_0 with Gaussian pdf $G_\sigma(x,y) = \sigma^{-1/2} e^{-|x^2+y^2|/4\sigma}$. Function, $g(|\nabla u_0(x,y)|)$, is positive in homogeneous regions and is zero at the edges.

Similarly, several other active contour models based on level sets using the image gradient to stop the evolution were proposed [8]-[9]. Most of the active contour models are based on the edge-function, $g(|\nabla u_0(x,y)|)$, to stop the evolution. Hence, such models are limited for detecting edges defined by gradient. Chan and Vese [7] have proposed another contour evolution model based on minimization of an energy-based segmentation similar to the Mumford-Shah segmentation method. An energy functional, $F(c_1, c_2, \phi)$, with c_1, c_2 being constants depending on C, is defined as,

$$
\begin{aligned}
F(c_1, c_2, \phi) = \mu \int_\Omega \delta(\phi(x,y)) \, |\nabla \phi(x,y)| dxdy \\
+ \nu \int_\Omega H(\phi(x,y)) dxdy \\
+ \lambda_1 \int_\Omega |u_0(x,y) - c_1|^2 H(\phi(x,y)) dxdy \\
+ \lambda_2 \int_\Omega |u_0(x,y) - c_2|^2 (1 - H(\phi(x,y))) dxdy,
\end{aligned} \tag{8}
$$

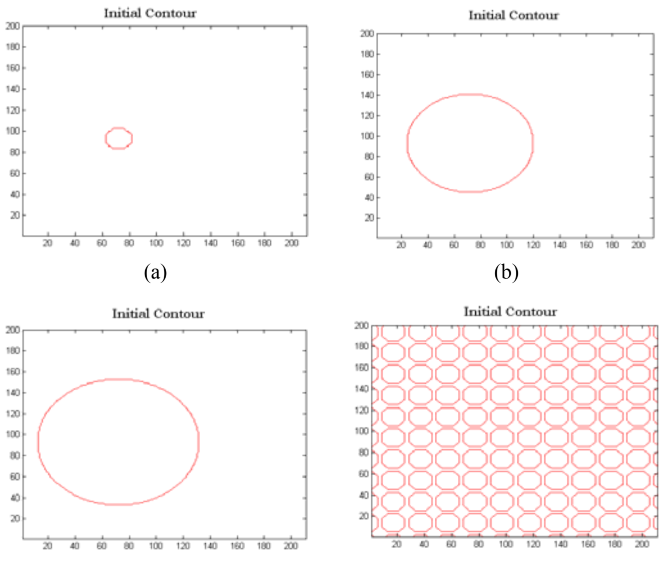
where μ, λ_1 and λ_2 are positive constants. H, the Heaviside function and δ is the Dirac measure given by

$$
H(z) = \begin{cases} 1, \text{ if } z \geq 0 \\ 0, \text{ if } z < 0 \end{cases}, \text{ and } \delta(z) = \frac{d}{dz} H(z). \tag{9}
$$

The associated Euler-Lagrange equation obtained by minimizing the energy functional F with respect to ϕ, and keeping c_1 and c_2 fixed is given by [7]

$$
\begin{aligned}
\frac{\partial \phi(x,y)}{\partial t} = \delta_\varepsilon(x,y)[\mu \operatorname{div}(\frac{\nabla \phi}{|\nabla \phi|}) - \nu - \lambda_1 (u_0 - c_1)^2 \\
+ \lambda_2 (u_0 - c_2)^2] = 0 \text{ in } (0, \infty) \times \Omega, \\
\phi(0, x, y) = \phi_0(x,y) \text{ in } \Omega, \\
\frac{\delta(\phi)}{|\nabla \phi|} \frac{\partial \phi}{\partial \vec{n}} = 0 \text{ on } \partial \Omega
\end{aligned} \tag{10}
$$

where $\vec{n}$ denotes exterior normal to the boundary $\partial\Omega$, and $\partial\phi / \partial\vec{n}$ denotes the normal derivative of ϕ at the boundary.



(a) (b)

(c) (d)

Figure 2. Various initial contours of level set. (a) Small. (b) Medium. (c) Large. (d) Whole.

Four types of initial contours C_0, namely, small, medium, large, and whole are employed in the proposed approach (see Fig. 2). All the four available initial contours are tried on the real microarray data and results are presented in the section IV. Figure 4 depicts the comparison results of the various initial contours employed for segmentation purpose. Convergence speed increases in the order of using small to whole contour, but with slight decrease in the accuracy of spot detection. Small contour took about 32 seconds; medium one took about 23 seconds, large contour took 14 seconds, and whole contour took 7 seconds on a p4-2.7GHz computer. Figure 4 shows that the whole contour provides acceptable results, which is four times faster than the small.

An outstanding characteristic of level set methods is that contours can split or merge as the topology of the level set function changes. Therefore, level set methods can detect more than one boundary simultaneously, and multiple initial contours can be placed (*i.e.* whole contour). This flexibility and convenience provide excellent means for autonomous segmentation of a microarray image by using a predefined set of initial contours.

IV. ILLUSTRATIVE EXAMPLES

Segmentation result obtained from commonly used Spot software are shown (see Fig. 3). Although, this tool is not restricted to a particular shape (circle, ellipse), it has certain disadvantages. The spots generated by the tool are typically smaller compared to the actual size. Also, for more accurate results, user has to identify the grid-points manually.

In this section, the proposed method has been applied on several real-life microarray images, and Fig. 5 shows the obtained segmentation results. Fig. 5(a), (c), (e) are input images, and Fig. 5(b), (d), (f) illustrate the respective segmented images. Especially, noisy microarray images are considered to illustrate the potential of the proposed method. As can be seen, the proposed method is effective and the steps like gridding and rotation correction (which are mandatory for the existing methods) are not required.

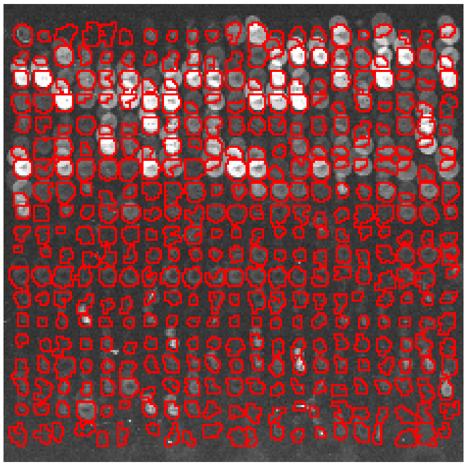
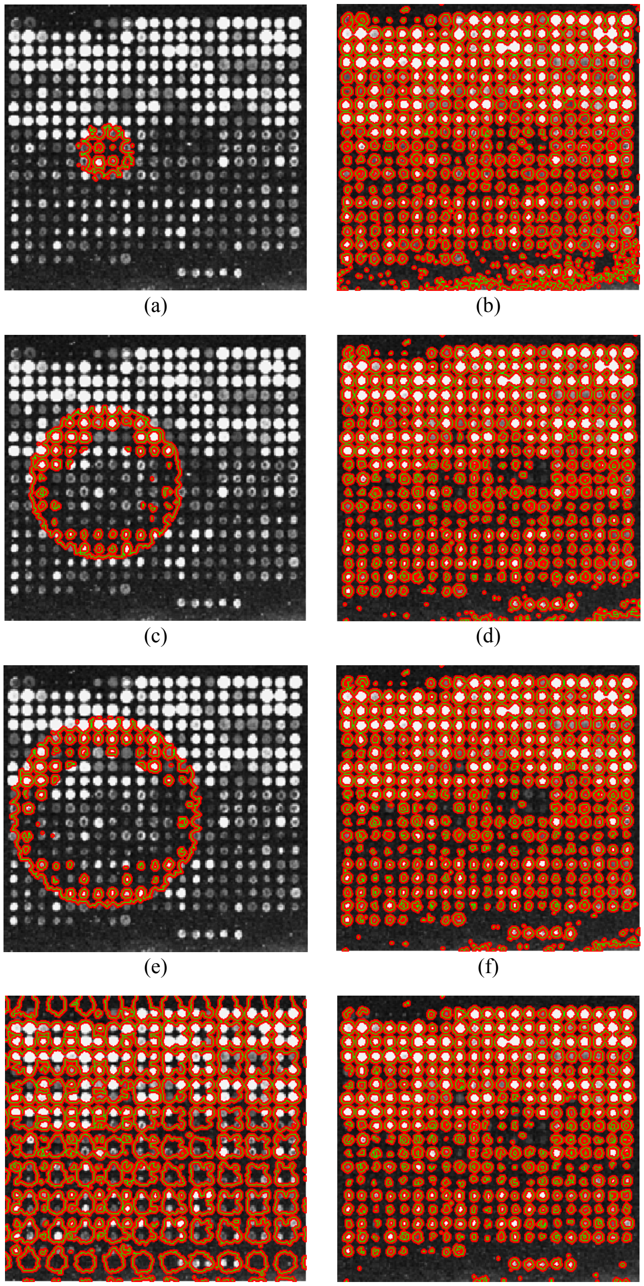


Figure 3. Spot's result of seeded region growing on the microarray data.


(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 4. Comparison of segmentation results using various initial contours. (a), (b) Small. (c), (d) Medium. (e), (f) Large. (g), (h) Whole.

V. CONCLUSION

In this paper, Chan-Vese approximation of the Mumford-Shah model and the level set method has been proposed for the segmentation of microarray images. Unlike the popular segmentation techniques for microarray images, denoising and gridding are not prerequisites for the proposed method. Highly accurate segmentation has been achieved irrespective of the spot shape and noise intensity. Four types of initial contours were tested to assess the best approach and the multiple initial contour approach has shown highest performance in terms of spot detection and time taken. The superiority of the approach is illustrated using several examples implying an automatic, robust and efficient tool for segmentation of microarray images.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank Dr. T. D. Bui of Concordia University for his valuable comments.

REFERENCES

[1] M. Schena, *Microarray Analysis.* New York: John Wiley&Sons, 2002.

[2] M. Eisen. ScanAlyze User's Manual. M. Eisen, 1999.

[3] Genepix 4000A: User's Manual. Axon Instruments Inc., 1999.

[4] QuantArray Analsyis Software: Operator's Manual. 1999.

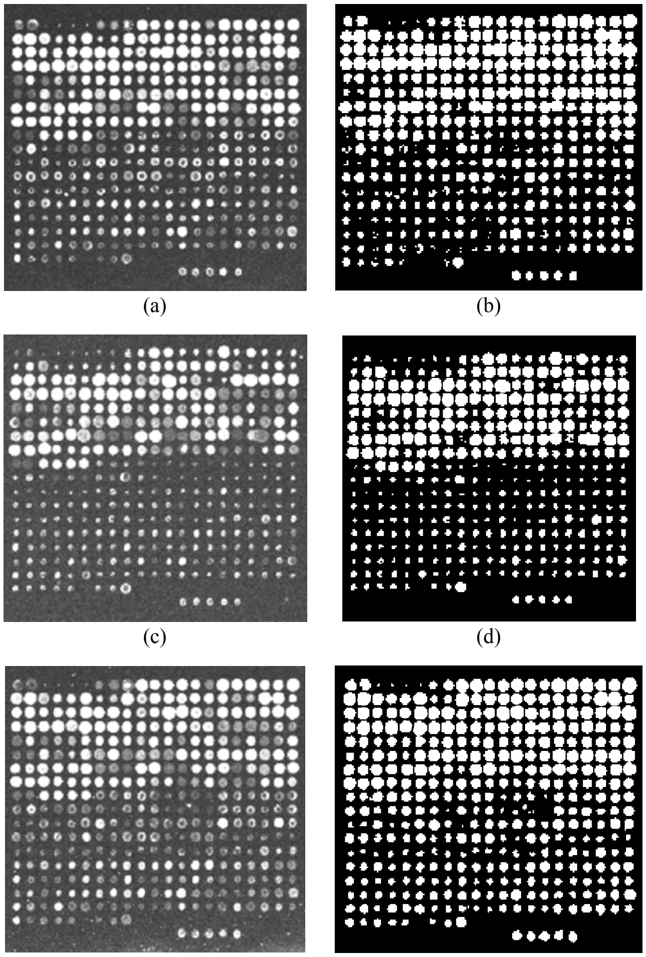
[5] J. Buhler, T. Ideker, and D. Haynor, "Dapple: improved techniques for finding spots on DNA Microarrays," Technical Report UWTR 2000-08-05, University of Washington, 2000.

[6] L. Rueda and Li Qin, "A new method for DNA Microarray image Segmentation," *Springer, Image Analysis and Recognition*, vol. 3656, pp. 886-893, 2005.

[7] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, pp. 266-277, 2001.

[8] J. Sethian, *Level Set Methods and Fast Marching Methods*. Cambridge: Cambridge University Press, 1999.

[9] S. Gao and T. Bui, "Image segmentation and selective smoothing by using Mumford-Shah model," *IEEE Trans. Image Processing*, vol. 14, pp. 1537-1549, 2005.


(a) (b)

(c) (d)

(e) (f)

Figure 5. Segmentation results. (a), (c) and (e) Input microarray images. (b), (d) and (f) Segmented microarray images.