Morphological Spot Segmentation for Gene Microarray Images

KASHIF I. SIDDIQUI*, ALFRED O. HERO*[#†], AND MATHEEN M. SIDDIQUI[‡]
Department of Electrical Engineering and Computer Science*
Department of Biomedical Engineering[#], Department of Statistics[†]
The University of Michigan, Ann Arbor, MI 48109, U.S.A.
Department of Computer Science[‡], Boston University, Boston, MA 02215, U.S.A.
E-Mail: kisrar@eecs.umich.edu*, hero@eecs.umich.edu*[#†], siddiqui@bu.edu[‡]

Abstract – **DNA microarray technology is one of the most powerful techniques used in modern biology and is extensively used for identification of sequence (gene/gene mutation) and determination of gene expression. A typical microarray image consists of a few hundred to several thousand spots and the extent of hybridization of these spots determines the level of gene expression (abundance) in the sample. A major issue in microarray image analysis is to accurately quantify spot shapes and intensities. In this paper we address this issue by performing accurate spot segmentation of a microarray image, using morphological image analysis techniques. This allows for estimation of the shape of the segmented spots using B-Splines or other parametric shape model.**

Keywords

Morphology, gene microarray images, spot-segmentation, quantification, watershed transform, shape estimation, B-splines.

I. INTRODUCTION

Gene microarray images permit estimation of the relative expression levels of thousands of genes simultaneously. Basically two mRNA (messenger RNA) samples, namely control sample and treatment sample, are reverse transcribed into cDNA (complementary DNA) samples and then tagged with two different dyes [1]. Further these two samples are mixed and scanned to produce a spotted image depicting the variations in fluorescent intensities. The results are color-coded, so that the most active genes (with the greatest degree of hybridization) are colored red, and genes that are repressed (hybridized the least) are colored green [2]. A sample microarray image is shown in Fig.1. The overall intensity within each spot is a measure of the level of gene expression or equivalently the mRNA abundance in the sample. As gene microarrays can suffer from a high-level background noise level, spot segmentation is essential for quantifying this intensity. This paper applies mathematical morphology methods to accomplish this segmentation and to quantify spot shape variability.
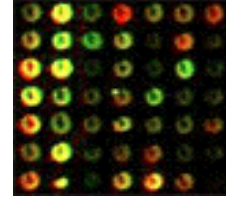

Fig. 1 A sample microarray image

II. MORPHOLOGICAL METHODS FOR SPOT SEGMENTATION

In this section we briefly present the basic principles, definitions and notations used in mathematical morphology. Let f be an image defined on $\mathbb{R}^2$ and B be a planar *structuring element* such that $B \subseteq \mathbb{R}^2$. Also, let ψ be an image operator, which transforms an image f into another image according to some specific task . The *erosion* (dilation) is defined by

$$\psi_{erosion}(f)(x,y) = \bigwedge_{(\varepsilon,\eta)\in B} f(x+\varepsilon, y+\eta)$$
$$= (f \ominus B)(x,y) \tag{1}$$

$$\psi_{dilation}(f)(x,y) = \bigvee_{(\varepsilon,\eta)\in B} f(x-\varepsilon, y-\eta)$$
$$= (f \oplus b)(x,y) \tag{2}$$

where, b is known as the *structuring function* and

$$b(x,y) = \begin{cases} 0, & for (x,y) \in B \\ -\infty, & otherwise \end{cases} \tag{3}$$

for some structuring element B. Erosion replaces the value of the image f at a pixel *(x,y)* by the infimum of the values of f over a structuring element B. Whereas dilation replaces the value of an image f at a pixel *(x,y)* by the supremum of the values of f over a structuring element B^c (reflection of B around the origin) [3], [4].

Structural opening (structural closing) is defined by

$$f \circ b = (f \ominus b) \oplus b \tag{4}$$
$$f \bullet b = (f \oplus b) \ominus b \tag{5}$$

and is used to undo the effect of erosion (dilation) by applying the associated dilation (erosion). It is important to note that openings (closings) are increasing, anti-extensive (extensive) and idempotent [5]. These both are smoothing filters and are used for smoothing contours of an image, suppressing small islands and cutting narrow isthmuses. The amount of smoothing is determined by the size and shape of the structuring function applied. Note that supremum of openings is also an opening and infimum of closings is also a closing. We summarize some of these operations below

Area opening is defined by

$$\psi_{aopn}(f,\alpha)(x,y) = \bigvee_{t \in \mathbb{R}} (x,y) \in \bigcup F_i(t) :| F_i(t) |\geq \alpha \quad (6)$$

where, $F(t)$ is a cross-section of the image intensity f and $\{F_i(t): i = 1,2,3...\}$ are grains of the cross-section and α is the threshold level [10]. Area opening is used to remove grains with area below a given value from the cross sections of a grayscale image. On the other hand *area closing* is defined by

$$\psi_{aclo}(f,\alpha)(x,y) = [\psi_{aopn}(f^*,\alpha)(x,y)]^* \quad (7)$$

and is used to fill in the holes whose area is below a given value in the image cross sections.

Morphological gradient [6] is defined by

$$\psi_{grad}(f) = \frac{1}{2}[f \oplus B - f \ominus B] \quad (8)$$

External morphological gradient (internal morphological gradient) is defined by

$$\psi^+_{grad}(f) = \frac{1}{2}[f \oplus B - f] \quad (9)$$

$$\psi^-_{grad}(f) = \frac{1}{2}[f - f \ominus B] \quad (10)$$

and also note,

$$\psi_{grad}(f) = \psi^+_{grad}(f) + \psi^-_{grad}(f) \quad (11)$$

Opening top-hat operator (closing top-hat operator) is defined by

$$\psi_{opnth}(f) = f - f \circ B \quad (12)$$

$$\psi_{cloth}(f) = f \bullet B - f \quad (13)$$

and is used to produce peaks (hollows) and ridges (ravines) in the topographic model of the image [9].

A morphological operator ψ is said to be a *morphological filter*, if it is increasing and idempotent. The combination of different morphological filters also results in a morphological filter. *Alternating filters* are combination of closings and openings and are defined as

$$\pi_k(f) = (f \circ kb) \bullet kB$$

or, $$\rho_k(f) = (f \bullet kb) \circ kB \quad (14)$$

where

$$kB = \begin{cases} \{origin\} & \text{, for k=0} \\ \underbrace{B \oplus B \oplus ... \oplus B}_{k-1\,dilations} & \text{, for } k \geq 1 \end{cases} \quad (15)$$

We can combine alternating filters to form *alternating sequential filters* (ASF) [10], which are given by

$$\mu_k(f) = \pi_k \pi_{k-1}....\pi_1(f)$$

or, $$\upsilon_k(f) = \rho_k \rho_{k-1}....\rho_1(f) \quad (16)$$

where, k = size or order of the filter.

Distance function $d(\bullet,\bullet)$ is a map from $\mathbb{R}^2 \times \mathbb{R}^2$ into the set $\mathbb{R}^+$ and has the following properties for every $u,v,w \in \mathbb{R}^2$

$$d(u,u) = d(v,v) = 0,$$
$$d(u,v) = d(v,u), \quad (17)$$
$$d(u,w) \leq d(u,v) + d(v,w)$$

where, $d(u,v)$ is distance between points u and v [10]. Using the distance function, we can define the *distance transform $D_f(u)$* of a binary image $f \subseteq \mathbb{R}^2$ at a point $u \in \mathbb{R}^2$ as

$$D_f(u) = \bigwedge_{v \in f} d(u,v) \quad (18)$$

The non-zero values of $D_f(u)$ (distance transform of the foreground) gives the minimum distance of a pixel in background from the foreground boundary, while the non-zero values of $D_f{}^c(u)$ (distance transform of the background) gives the minimum distance of a pixel in foreground from the foreground boundary.

A *regional minimum*, M, of an image f is a connected component of pixels in f with a given value v, such that every pixel in neighborhood of M has a value strictly larger than v. Every regional minimum M has a *catchment basin $C(M)$* associated with it, which is collection of all points of the topographic surface of f, such that a drop of water falling at p slides along the surface until it reaches M [8], [11].

Now, if we flood the topographic surface of an image from its regional minimum and if we prevent the merging of water coming from different sources, we partition the image into two different sets; the catchment basins and the *watershed lines*, where each catchment basin contains one and only one regional minimum M_i [10]. We can define the *watershed transform* as

$$W(f) = D \cap \left(\bigcup_{i \in \mathbb{R}} C(M_i) \right)^c \quad (19)$$

where, D is connected domain of the image f [11]. In the next section we apply these morphological techniques to spot segmentation in gene microarray images.

III. SPOT SEGMENTATION OF MICROARRAY IMAGE

Image Segmentation is defined as the process of isolating objects in the image from the background i.e., partitioning the image into disjointed regions, such that each region is homogeneous with respect to some property [6]. Therefore, *spot segmentation* can be defined as the process of extracting the appropriate homogenous spots, having the desired homogeneity property, from a microarray image.

We will use morphological techniques, discussed briefly above, for spot segmentation of a microarray image [3], [4], [8], [9], [10]. A portion of the original image's grayscale version is shown in Fig. 2.1(a). It can be seen, that there are bright regions inside the spots, which will cause faulty binarization of the image. Applying an area opening (6) solves this problem and the result is depicted in Fig. 2.1(b). *Thresholding* the image in (b) produces the binary image shown in Fig. 2.1(c). The number of spots produced during thresholding is determined by the threshold level we select and thus can be used to filter those spots with weak hybridization levels.

Two iterations of the *alternating sequential filter* (ASF), characterized by a cross structuring element having radius of 1 pixel, are applied to the image in (c) while using the sequence of opening followed by closing operators (14), (16). The topographic model of the resulting image is shown in Fig. 2.1(d). Next we find the *regional maxima* of the image in (d), according to the connectivity defined by cross structuring element. These regional maximums act as markers for each cell, as seen from the resulting image in Fig. 2.1(e).

Now we apply the *watershed transform* (18) to the *negation* of the original image using the markers found previously (shown in (e)) and using the box-structuring element to define connectivity. These watershed lines are used to act as *external markers*, which mark the crest lines of the original image (a). Further we locate the regional minima of the original image and use them as *internal markers*. These external markers and internal marks are united to result in a *combined marker*, which is shown in Fig. 2.1(f) overlaid over the original image. From (f) it is seen that the spot boundaries are constrained between external and internal markers.

The watershed transform is applied to the *gradient* of the image (shown in Fig. 2.1(g)) using the combined marker and the cross structuring element described above. The resulting watershed lines are shown in Fig. 2.1(h) and are overlaid over the original image in Fig. 2.1(i). The final image (i) shows that there are no oversegmentaion problems. Oversegmentation arises when the gradient operator is overly sensitive to grayscale variation and noise, creating a large number of irrelevant catchment basins. The procedure described above produces accurate spot segmentation of the microarray image with very low segmentation noise.
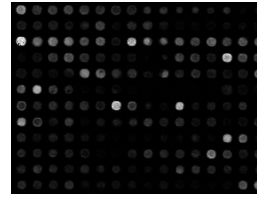

Fig. 2.1(a) grayscale version of original microarray image
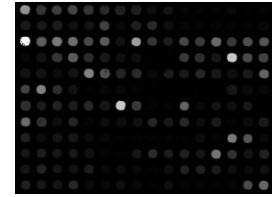

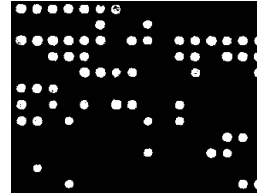Fig. 2.1(b) result after application of area opening

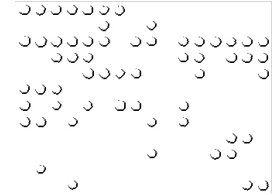Fig. 2.1(c) result after application of thresholding

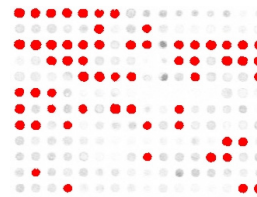Fig. 2.1(d) topographical model of image after apply ASF.

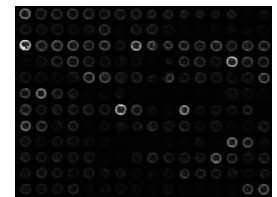Fig. 2.1(e) regional maximums which act as markers for spots

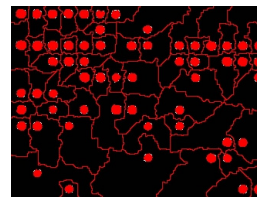Fig. 2.1(g) gradient of the original microarray image

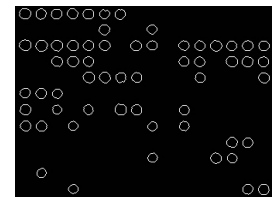Fig. 2.1(f) combined internal and external marker

Fig. 2.1(h) watershed lines resulting after applying watershed transform on gradient of the original image and the combined marker in (f)
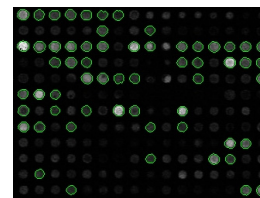

Fig. 2.1(i) watershed lines overlaid over the original image

IV. QUANTIFICATION OF SPOT SHAPE

Our accurate spot segmentation permits quantification of spot shape and other characteristics e.g. noise, background averaging and subtraction. Hence we illustrate the utility of segmentation for shape quantification. After segmentation of the image, the number N of surviving spots in the image is determined (in this case N=60) and then an Nx2 matrix is calculated which contains co-ordinates of the centroid of each spot. Using the centroid of i-th spot we calculate sample values of the boundary, which results in an NxM matrix for the x and y coordinates of sample points of each spot boundary, where M = the number of sample points along the boundary. Using the matrix of Cartesian coordinates of sample points of the boundary, we transform to polar coordinates with respect to centroid of i-th spot.

To achieve a low dimensional parameterization of the spot shape we investigated planar curve model fitting. There are numerous methods available to represent closed boundaries as periodic planar curves, such as Fourier descriptors, Bezier curves, Beta-Splines and B-Splines. In this paper, we adopt B-spline boundary model [12] [13]. A B-spline consists of a set of K fixed positions, called knots, and piecewise smooth curves, called basis functions, connecting each of the knot position. For an m-th order B-spline these curves are specified by polynomial functions of degree m and a vector of parameters, the B-spline co-efficients. To ensure smoothness at each knot, the curve is constrained to have continuous derivatives up to order (m - 1). Using the centroid matrix, which contains centroid of each spot, and the polar coordinates matrix of the sample points of the detected spot boundary we apply B-spline fitting to smoothly parameterize the shape of each spot. A typical result for one spot is show in Fig. 3.
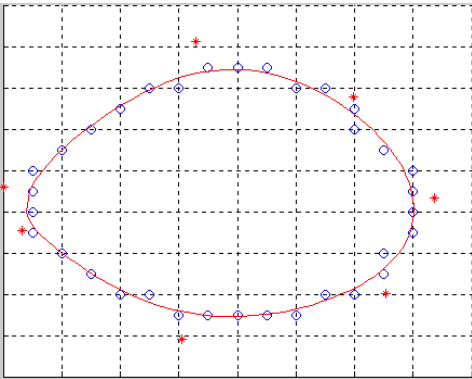

Fig. 3 B-Spline around the boundary of one spot, and here circles are samples of boundary, asterisks are the control points of spline and solid line is the B-spline

After determining the coefficients vector of the B-splines by least squares, we developed a database of spot shapes in the microarray image. This database was used to query for a possible correlation of spot shape to factors such as intensity, background noise and microarray print-head. A spot in a microarray image is classified into a specified intensity changes and the statistics of spot shape can be computed and analyzed as a function of intensity level. Any correlation between intensity and shape can subsequently be used to improve estimates of overall hybridization levels leading to more accurate gene microarray analysis.

V. CONCLUSION

Spot segmentation of a microarray image has been achieved in an efficient manner using morphological techniques for image analysis, which is insensitive to noise problems leading to oversegmentation. The detected boundaries of the valid spots are used to obtain B-spline co-efficients of the shape, which can be further stored in a database for quantification of spot variations. We are in the process of collecting a large database of spot shapes which will be used for statistical spot shape analysis

VI. REFERENCES

[1] Chen Y, Dougherty E, Bittner M, *"Ratio-Based Decisions and the Quantitative Analysis of cDNA Micro-array Images"*, Journal of Biomedical Optics 2: 364 (1997).

[2] Eisen MB. and Brown PO., *"DNA Arrays for Analysis of Gene Expression"*, Methods Enzymol 303, 179-205 (1999).

[3] J. Serra, "*Image Analysis and Mathematical Morphology"*, London, England: Academic Press, (1982).

[4] C. R. Giardina and E. R. Dougherty, "*Morphological Methods in Image and Signal Processing"*, New Jersey: Prentice Hall, (1988).

[5] H. J. A. M. Heijmans, "*Morphological Image Operators"*, Boston: Academic Press, (1994).

[6] J. Serra and P. Soille, "*Mathematical Morphology and its Applications to Image and Signal Processing"*, Dordrecht, The Netherlands: Kluwer, (1994).

[7] L. Vincent, *"Morphological Grayscale reconstruction in image analysis: Applications and efficient algorithms"*, IEEE Transactions on Image Processing, vol. 2, pp 176-201, (1993).

[8] F. Meyer and S. Beucher, *"Morphological segmentation"*, Journal of Visual Communications and Image Processing, vol. 1, pp. 21-46, (1990).

[9] Dougherty, E. R.,*"An Introduction to Morphological Image Processing"*, SPIE Press, Bellingham, (1992).

[10] J. Goutsias and S. Batman, *"Morphological Methods for Biomedical Image Analysis Handbook of Medical Imaging"*, Volume 2, Medical Image Processing and Analysis M. Sonka and J. M. Fitzpatrick (Eds.) pp. 175-272 (2000).

[11] S. Beucher, *"The watershed transformation applied to image segmentation"*, 10th Conf. on Signal and Image Processing in Microscopy and Microanalysis, (1991), Cambridge, UK.

[12] Elaine Cohen, Richard F. Riesenfeld and Gershon Elber, "Geometric Modeling with Splines: An Introduction", Natick Massachusetts, A K Peters (2001).

[13] G. Farin, "Nurbs: From Projective Geometry to Practical Use". A K Peters Ltd, 2nd edition, (1999).