Robust protein microarray image segmentation using improved seeded region growing algorithm

Liqiang Wang (王立强), Xuxiang Ni (倪旭翔), and Zukang Lu (陆祖康)

State Key Lab of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027

Received May 6, 2003

Protein microarray technology has recently emerged as a powerful tool for biomedical research. Before automatic analysis the protein microarray images, protein spots in the images must be determined appropriately by spot segmentation algorithm. In this paper, an improved seeded region growing (ISRG) algorithm for protein microarray segmentation is presented, the seeds are obtained by finding the positions of the printed spots, and the protein spot regions are grown through these seeds. The experiment results show that the presented algorithm is accurate for adaptive shape segmentation and robust for protein microarray images contaminated by noise.

OCIS codes: 170.3880, 100.5010.

The development of protein microarrays has accelerated within the past few years. Protein microarray can be widely used in diagnostics, drug screening and testing, disease monitoring, drug discovery and medical research[1]. Due to the huge amount of data, automatic analysis for protein microarray images has become indispensable[2]. However, automatic analysis for protein microarray images has proved to be difficult due to the poor contrast, and the many artifacts arising from the hybridization procedures such as irregular spot shape and size, dust on the slide, large intensity variation within spots and background, and nonspecific hybridization. In order to extract data from microarray images, it is necessary to correctly identify and segment out each spot[3].

Existing segmentation methods for microarray images can be categorized into four groups, according to the geometry of the spots they produce: 1) fixed circle segmentation, 2) adaptive circle segmentation, 3) adaptive shape segmentation and 4) histogram segmentation[2]. Because of different physical and chemical conditions, the spots of printed protein may not be as regular as one would expect, therefore, adaptive shape segmentation can achieve more accurate results than other segmentation algorithms. There are two commonly used methods for adaptive shape segmentation in protein microarray image, seeded region growing[2] and mathematical morphology[3,4].

Because the number of pixels in a protein microarray image is limited, both smoothing and sharpening filters need to be avoided when processing the image contaminated by noise[5]. In order to get the robust and accurate segmentation resulting from the protein microarray images contaminated by noise, we describe an improved seeded region growing (ISRG) algorithm[6] in this paper for protein microarray segmentation, the algorithm details are described as follows.

Firstly, the starting points, or seeds for ISRG are determined at the gridding stage of protein microarray image. Because of the prior knowledge of the geometric structure of protein microarray, such as the radius of a typical spot in the image, the distance between two neighbor spots in the row and column directions and the approximate location of the spot array in the image etc, the grids are generated after entering these prior data.

Figure 1 is an HCV protein microarray image obtained by GenePix 4000B scanner and Fig. 2 shows the spot seeds of ISRG for the typical part image, they are the intersections of the horizontal lines and the vertical lines. For background seeds, we define them the center pixels of each section divided by the horizontal lines and the vertical lines.

Secondly, we segment the image into regions according as the seeds we found. The process evolves inductively from the seeds, namely, the initial state of the sets A_1, A_2, $\cdots$, A_n. Each step of the algorithm involves the addition of one pixel to one of the above sets. Let T be the set of all the unallocated pixels which border at least one of the sets A_i, $i = 1, 2, \cdots, n$[7],

$$T = \{x \notin \bigcup_{i=1}^{n} A_i : N(x) \bigcap \bigcup_{i=1}^{n} A_i \neq \Phi\}, \qquad (1)$$

where $N(x)$ is the set of immediate neighbors of the pixel x. In this paper, we use a rectangular grid, so the pixel x has 8 immediate neighbors. A single step of the algorithm involves examining the neighbors of each $x \in T$ in turn. If $N(x)$ intersects a region A_j, then $\delta(x)$, measuring the difference between x and the intersected region is calculated. We use the simplest definition for

$$\delta(x) = |g(x) - \mathrm{mean}_{y \in A_j}\{g(y)\}|, \qquad (2)$$

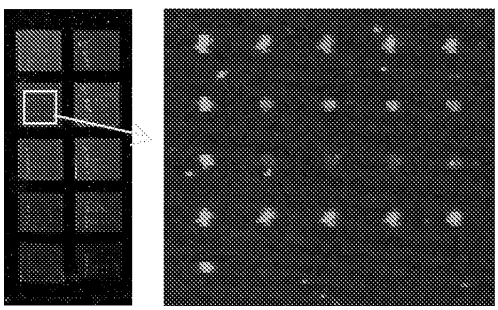

Fig. 1. Left is an HCV protein microarray image, right is a typical part of the image.

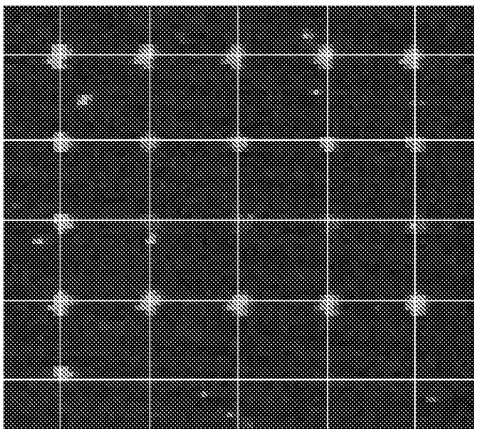

Fig. 2. The intersections of the horizontal lines and the vertical lines are the seeds of the spots, the center pixels of each sections are the seeds of the background.

where $g(x)$ is the intensity of the pixel x. If $N(x)$ intersects more than one region, then A_j is taken to be that region for which $\delta(x)$ is a minimum. In this way a δ value is determined for each $x \in T$. Finally, the pixel $z \in T$ that satisfies

$$\delta(z) = \min_{x \in T}\{\delta(x)\} \tag{3}$$

is appended to the region corresponding to $\delta(z)$. The new state of the regions $\{A_j\}$ then constitutes the input to the next iteration. This process continues until all of the image pixels have been assimilated.

We implement the ISRG algorithm using an ascending priority queue in order to eliminate the pixel order dependence of seeded region growing algorithm[6]. Figure 3 presents the spot segmentation result using ISRG algorithm. The spot region contours are traced out by the gray lines. We also use GenePix Pro 4.1 commercial software to segment the image and Fig. 4 presents the segmentation result, the spot regions are adaptive circles. Circles with vertical bars indicate "spot not found". From the results, it is clear that the adaptive shape segmentation using our ISRG algorithm is better than the adaptive circle segmentation using GenePix Pro 4.1 software.

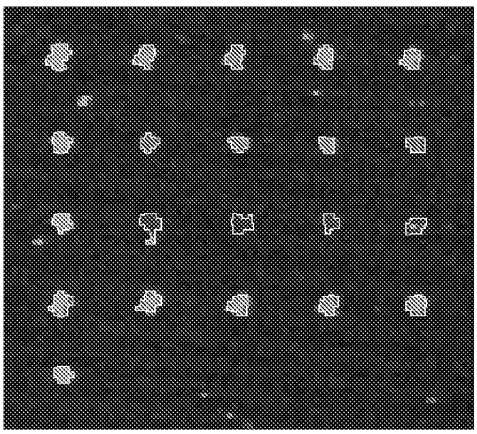

Fig. 3. Protein microarray spot segmentation result using ISRG algorithm.

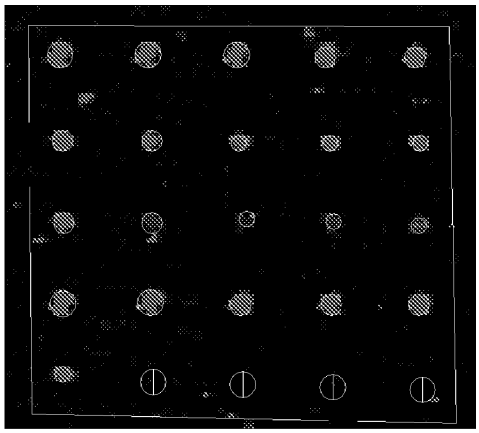

Fig. 4. Protein microarray spot segmentation result using GenePix Pro 4.1.

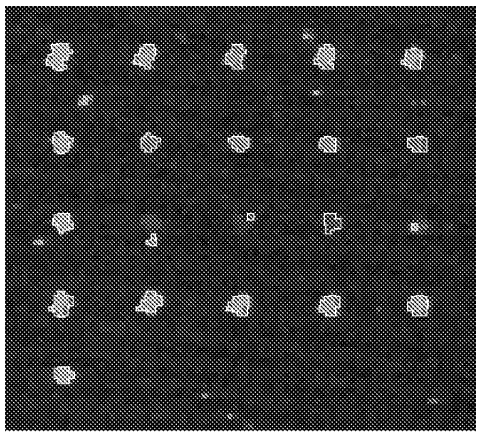

Fig. 5. The segmentation result when choosing spot seeds according to Yang, three spots are not segmented very well due to the noise.

In order to produce a robust segmentation result from the protein microarray image, we simply choose the seeds as the intersections of the horizontal and vertical grid lines. In the paper of Yang *et al.*[2], the seeds are chosen by finding the maximum intensity points over a small region centered at the intersection pixels, however, we find out that although it sometimes may overcome the local irregularities or small errors in the grid estimation, it will produce some inferior segmentation results when the protein microarray is contaminated by noise. Figure 5 is the segmentation result when using the spot seeds according to Yang *et al.*, we can judge that the result is not robust because several spots are infected by the noise.

Because of the noise and contaminations of the protein microarray slide surface, not only the variation of the spot shape and size, but also the high intensity variation of spot signal, robust and accurate segmentation is very important for pre-processing of the microarray image. Although it is becoming increasingly clear that there might never be a "best" approach and that the application of various algorithms will allow different aspects of the results to be explored, choosing the appropriate algorithms for protein microarray image processing is very crucial for protein microarray experimental design[8].

In this paper, we have described an approach for robust

and accurate protein microarray image segmentation. At first, we determine the seeds at the gridding stage of protein microarray image, then an ISRG is performed for image segmentation. By contrasting our algorithm results with other ones, we find that our algorithm is more robust and accurate in protein spot segmentation.

This project was supported by the Key Program of National Natural Science Foundation of China under Grant No. 30227002. L. Wang's e-mail address is zjuwlq@sohu.com.

References

1. J. H. Ng and L. L. Ilag, J. Cellular and Molecular Medicine **6,** 329 (2002).
2. Y. H. Yang, M. J. Buckley, S. Dudoit, and T. P. Speed, J. Computational and Graphical Statistics **11,** 108 (2002).
3. A. W.-C. Liew, H. Yan, and M. Yang, Pattern Recognition **36,** 1251 (2003).
4. R. Hirata Jr., J. Barrera, R. F. Hashimoto, D. O. Dantas, and G. H. Esteves, Real-Time Imaging **8,** 491 (2002).
5. Y. Chen, E. R. Dougherty, and M. L. Bittner, J. Biomedical Optics **2,** 364 (1997).
6. A. Mehnert and P. Jackway, Pattern Recognition Lett. **18,** 1065 (1997).
7. R. Adams and L. Bischof, IEEE Trans. Pattern Analysis and Machine Intelligence **16,** 641 (1994).
8. J. Quackenbush, Nature Reviews Genetics **2,** 418 (2001).