

Automatic Segmentation of cDNA Microarray Images using different methods

Islam Fouad¹, Mai Mabrouk² and Amr Sharawy³

¹College of Applied Medical Sciences, SALMAN Bin ABDUL-AZIZ University, Kharj, KSA

²Biomedical Engineering, MUST University, 6th of October, Egypt

³Biomedical Engineering, Cairo University, Giza, Egypt

ABSTRACT

Due to the vast success of bioengineering techniques, a series of large scale analysis tools has been developed to discover the functional organization of cells. Among them, cDNA microarray has emerged as a powerful technology that enables biologists to study thousands of genes simultaneously within an entire organism, and thus obtain a better understanding of the gene interaction and regulation mechanisms involved. The analysis of DNA microarray image consists of several steps; gridding, segmentation, and quantification that can significantly deteriorate the quality of gene expression information, and hence decrease our confidence in any derived research results. Thus, microarray data processing steps become critical for performing optimal microarray data analysis and deriving meaningful biological information from microarray images. Segmentation is the process, by which each individual cell in the grid must be selected to determine the spot signal and to estimate the background hybridization. In this paper, four segmentation methods are explored; "fixed circle", "adaptive circle", "thresholding", and "adaptive shape" segmentation. By comparing the results, it was found that the "adaptive shape segmentation method" can segment noisy microarray images correctly, gives high accuracy results and minimal processing time, and can be applied to various types of noisy microarray images.

Keywords: Noisy Microarray Image, Gene Expression, Analysis of DNA Microarray Image, Segmentation.

1 Introduction

Microarray technology came on time to cover the need to monitor in parallel all the DNA sequences and to have the adequate sensibility to detect the variation of gene expression. There may be tens of thousands of spots on an array. Each spot contains tens of millions of identical DNA molecules with lengths from tens to hundreds of nucleotides. Afterwards, the microarray slide is exposed to a set of labeled cDNA samples, which are derived from tissue of interest. With the completion of hybridization reaction, the amount of the target that bounds to each sample is measured with the aid of image capturing devices and computer technology. The measurement is based on the intensity of the spot. There are three basic steps in the processing of microarray image [1]; gridding, segmentation, and quantification.

Image segmentation is the process of distinguishing objects from their background [2]. It is usually the first step in vision systems, and is the basis for further processing such as description or recognition. The goal of segmentation is to extract important features from images. Segmentation of an image can also be seen, in practice, as the classification of each image pixel to be assigned to one of the image compositions.

Different segmentation methods have been presented include the dynamic system modeling based approach [3] performs pixel clustering operations in a parallel manner to speed-up the segmentation process. The cellular neural network scheme [4, 5] segments the spots by performing a number of operations such as background clean-up, grid analysis, irregular spot elimination, and intensity analysis. The morphology based approach [6] uses a series of optional steps to segment the microarray image. The combination of Markov random field based grid segmentation and active contour modeling constitutes an approach suitable for spot detection and segmentation [7]. The two-stage clustering based approach [8] is comprised of spots" boundaries adjusting and intensity-based partitioning operations. The use of adaptive thresholding and statistical intensity modeling is the base for some segmentation schemes [9], whereas another approach [10] uses a seeded region growing algorithm to identify spots of different shapes and sizes. Histogram and thresholding operations were used to classify microarray image samples into either foreground (spots) or background pixels [11].

One of the key steps in extracting information from a microarray image is the segmentation whose aim is to identify which pixels within an image represent which gene. This task is greatly complicated by noise within the image and a wide degree of variation in the values of the pixels belonging to a typical spot. A comparison between four different segmentation methods are presented on a noisy microarray image, with high accuracy. The paper is organized as follows: a brief introduction is presented in this section, section 2 presents the used materials, section 3 summarizes the proposed segmentation methods for various cDNA microarray noisy images and section 4 discusses the results of the applied algorithms on microarray data set image. Conclusions are presented in section 5.

2 Materials

Different images have been selected from two different data sets, to test the performance of the proposed methods. They have different scanning resolutions, and different noise types, in order to study the flexibility of the proposed methods to detect spots with different sizes and features.

The images are stored in TIFF files with 16-bit gray level depth. A chosen microarray image with various kinds of artifacts or noises is drawn from Princeton University Microarray (PUMA) database [12].

The image includes thirty two sub-grids presenting acute lymphoblastic leukemia tissues (PUMA Experiment ID: 10223). The slide name is (shae082) and it is a cDNA microarrays spotted by a total of 24192 genes.

MATLAB [13] is used for data analysis and technical computing, as it is a high performance and powerful tool. MATLAB version is 6.5 on a windows XP platform. The P.C used has a processor: Intel Core r2 4300 CPU with 2GB of memory.

3 Methods

Once grids have been placed, discrimination between areas that are considered the spot signal and areas that are considered the background signal must be carried out. The process, by which each individual cell in the grid must be selected to determine the spot signal and to estimate the background hybridization, is called segmentation. That information will be put towards a quantitative measurement at each cell. There are four presented approaches for segmentation; fixed circle segmentation, adaptive circle segmentation, thresholding segmentation, and adaptive shape segmentation.

3.1 Fixed Circle Segmentation

Fixed circle segmentation method assigns all the spots the same size and shape. It uses a constant-diameter circle as the shape of all the spots in the image. The proposed fixed circle segmentation technique is presented as in the following steps:

1. *Input the gridded image.*
2. *Get the region of interest (R.O.I.).*
3. *Let: $l = i^{th}$ row in the R.O.I.*
4. *$l = 1$ to end of R.O.I.*
5. *For each l , we calculate the center of its corresponding boundary box.*
6. *Draw a circle with the obtained center and the obtained estimated period (distance between two adjacent spots).*

3.2 Adaptive Circle Segmentation

Adaptive circle segmentation considers the shape of each spot as a circle, where the center and diameter of the circle are estimated for each spot. It involves two steps. First, the center of each spot needs to be estimated. Second, the diameter of the circle has to be adjusted. Adaptive Circle Segmentation method is shown in figure

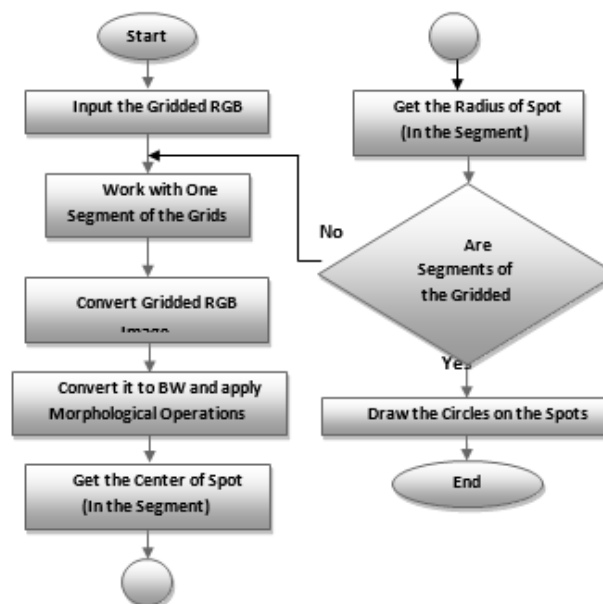


Figure 1 Flowchart of Adaptive Circle Segmentation

3.3 Thresholding Segmentation

The first step to start this technique is to apply a single threshold level to the whole image, so all spots are detected equally. However, it doesn't work so well due to the large differences in the spot brightness. One way to equalize large variations in magnitude is by transforming intensity values to logarithmic space. This works much better, but some weak spots are still missed. Alternatively, the bounding boxes can be used to determine local threshold values for each spot. For an overview of the Thresholding segmentation algorithm the reader is referred to figure 2

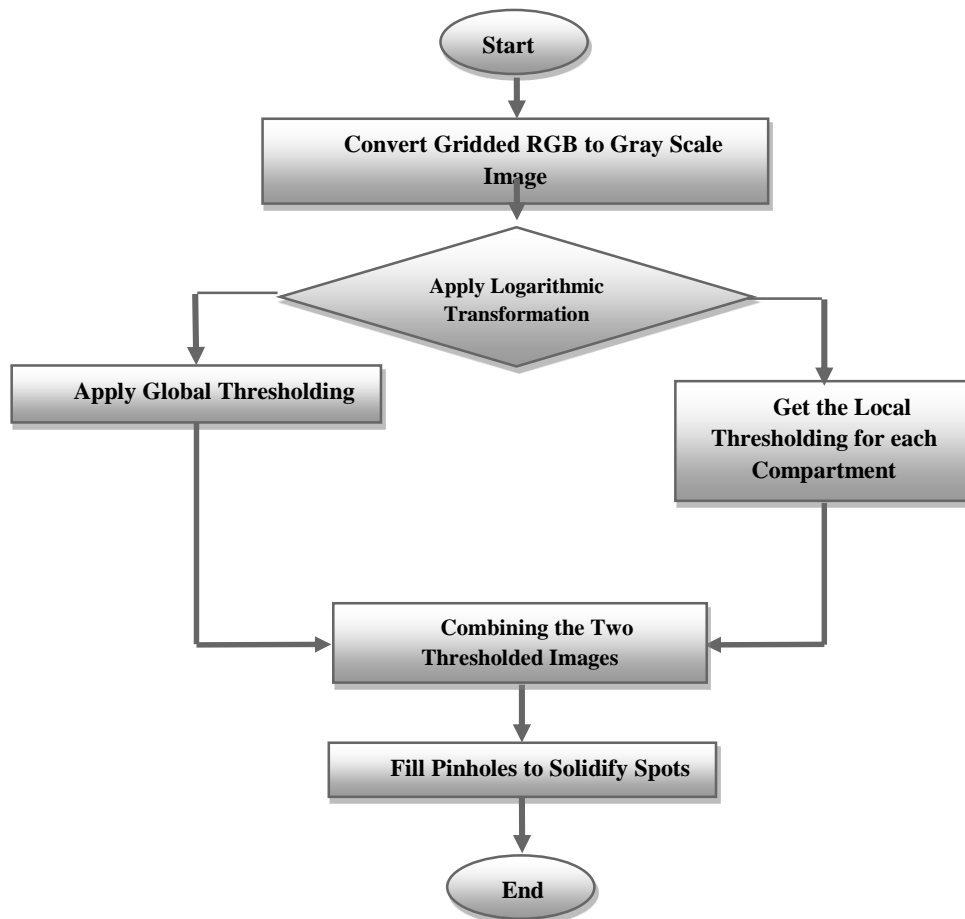


Figure 2 Flowchart of Thresholding Segmentation.

Unfortunately, results reveal that weak spots showed up well but spots with bright perimeters were as bad as the original global threshold before log space transformation. Since each of the local and the global thresholding has a special advantage, we combined the best of both approaches. These spot segmentation results were indeed much better.

The silhouettes of some spots still contained pinholes. The whole image could be filled using image filling operation, but this may not be a good idea. Notice that some spots run together. If four mutually adjacent spots (sharing a common corner) were all joined at their edges then a single function call would incorrectly fill in the common corner as well. To avoid that possibility, it is safer to fill each spot; take each bounding box region at a time by looping. Indeed, the spot segmentation now looks quite good [14].

3.4 Adaptive Shape Segmentation

Seeded region growing (SRG) is a common technique that deals with different shapes in image segmentation. In SRG, the regions grow outwards from the seed points, preferentially, based on the difference between the pixel value and the running mean of values in an adjoining region. This method requires an initial point to be known, which is called the *seed*. After obtaining the seeds, the process is repeated simultaneously for both foreground and background regions until all the pixels are assigned to either foreground or background. Those pixels that are adjacent to a region are assigned first according to its intensity. Figure 3 shows the steps of applying the method.

An important advantage of SRG method is that it can be applied to microarray images containing spots of any shape and size.

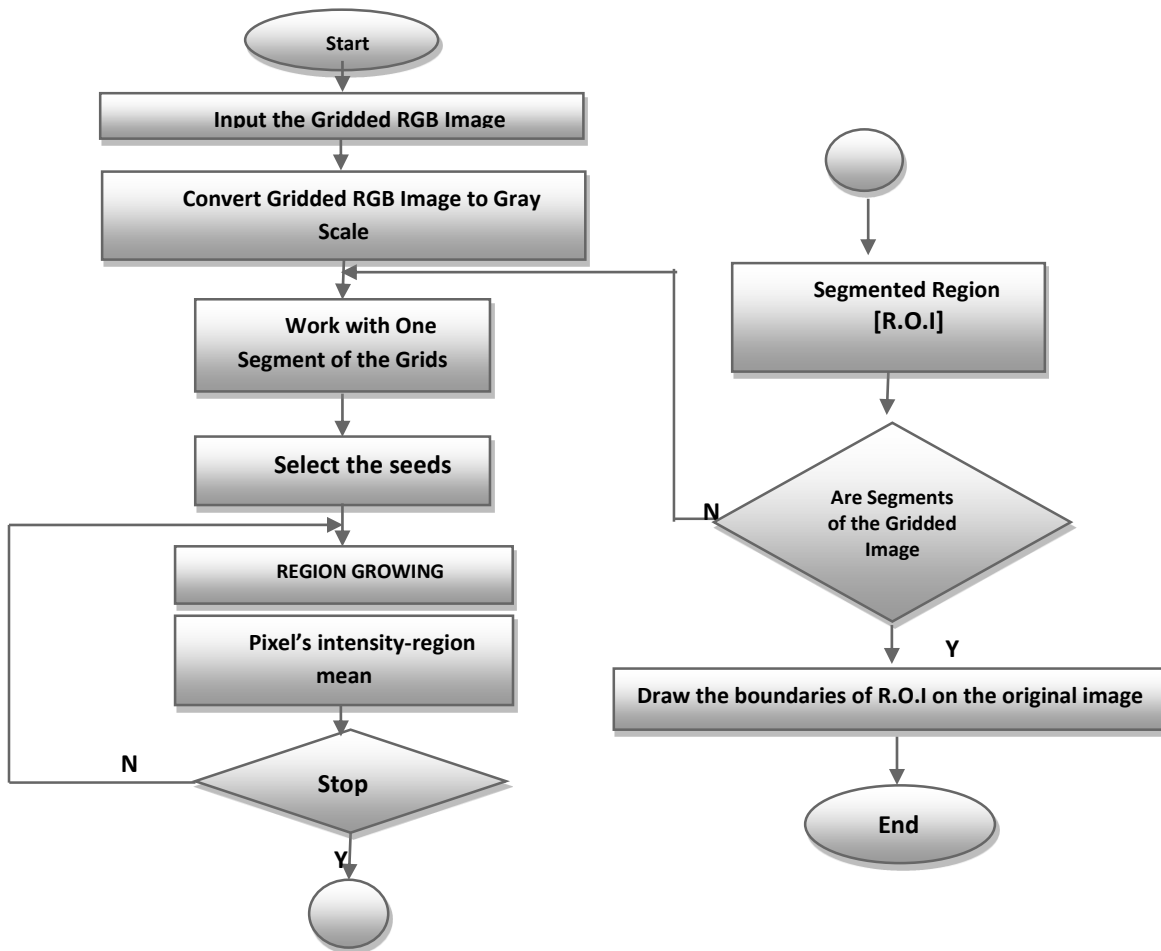


Figure 3 Flowchart of S.R.G. Segmentation

The seed; which is the initial point to start applying the method is determined automatically. The pixel which has the maximum intensity in the spot window: as that point should most probably one pixel of the inspected spot.

4 Results and Discussion

Four different segmentation methods were presented; Fixed Circle, Adaptive Circle, Thresholding, and Adaptive Shape Segmentation.

These presented segmentation methods are tested on a number of noisy microarray images drawn from Princeton University Microarray (PUMA) database. The results of applying these segmentation methods on a microarray image are shown in this section. The chosen noisy microarray image includes thirty two sub-grids (PUMA Experiment ID: 10223) and it is spotted by a total of 24192 genes. To test the efficiency of the proposed methods, a sub-array in the fourth row and the second column was cropped as shown in figure 4.

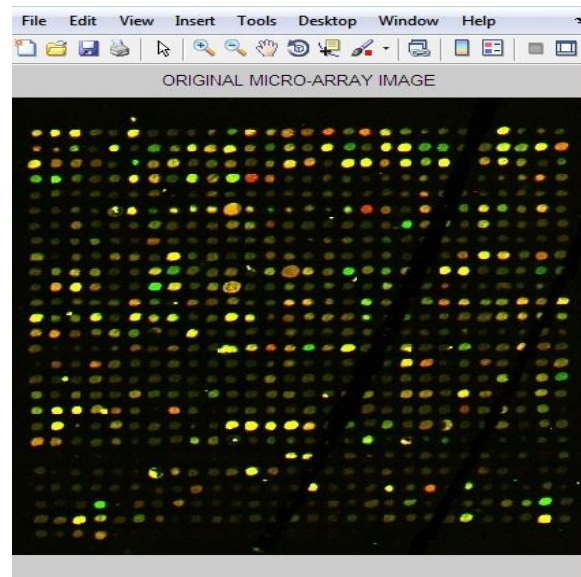


Figure 4: The Original Microarray Image

The accuracy of the proposed segmentation methods were analyzed by means of a statistical analysis [15]. A spot was “very efficiently segmented” if at least 90% of the entire spot area was enclosed in the contour of that spot.

Figure 5 shows the effect of applying the fixed circle segmentation method the microarray image, figure 6 shows the effect of applying the adaptive circle segmentation method on the microarray image, figure 7 to figure 12 show the effect of applying the thresholding segmentation method, and figure 13 shows the effect of applying the adaptive shape segmentation method.

4.1 Fixed Circle Segmentation

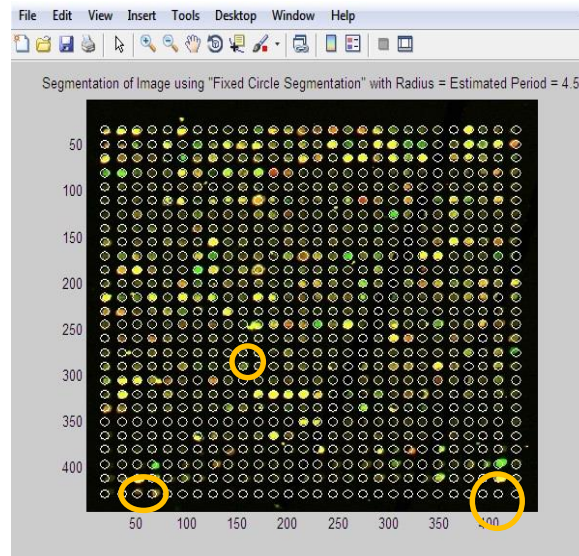


Figure 5: Image after Applying “Fixed Circle Segmentation” Method.

It was obviously shown that the fixed circle method clearly cannot satisfy the needs. Figure 5 shows the resulting image after applying the fixed circle approach. We can see that some regions within the high intensity areas (spots) are left out of the foreground, and some regions within the low intensity areas (background) are included in the foreground regions. That is because of the fixed diameter of all the drawn circles despite the variation in the diameters of the spots in the same image.

4.2 Adaptive Circle Segmentation

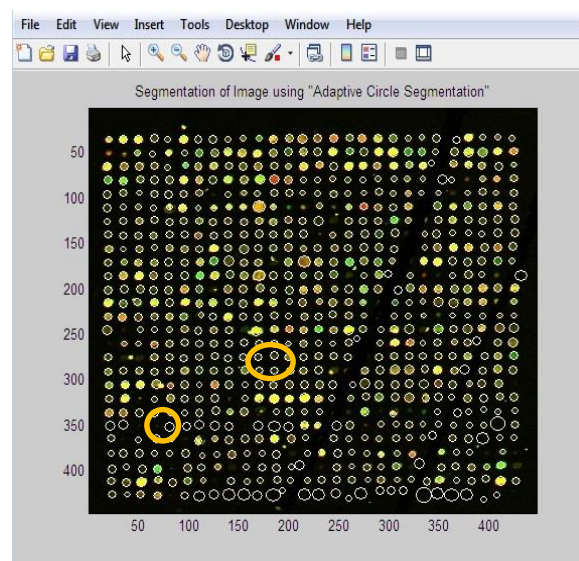


Figure 6: Image after Applying “Adaptive Circle Segmentation” Method.

The adaptive circle segmentation method achieves better results for circle-shaped spots. However, the spots in a microarray image can take shapes including ellipses. That is the main drawback of the adaptive circle segmentation method is that it restricts the shape, it's obvious in figure 6.

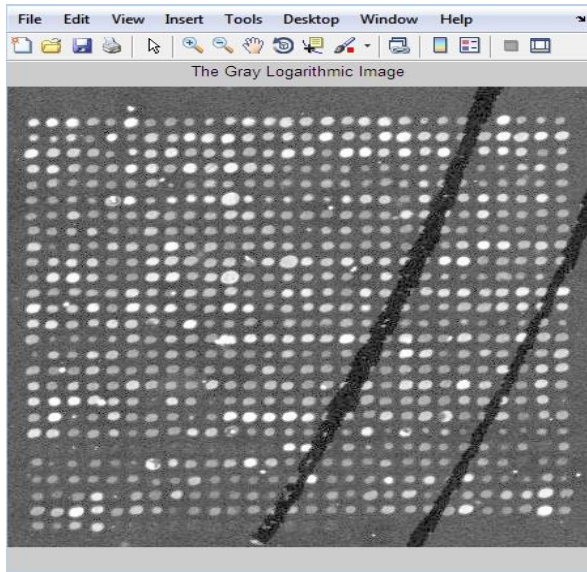


Figure 7: Gray Logarithmic Image

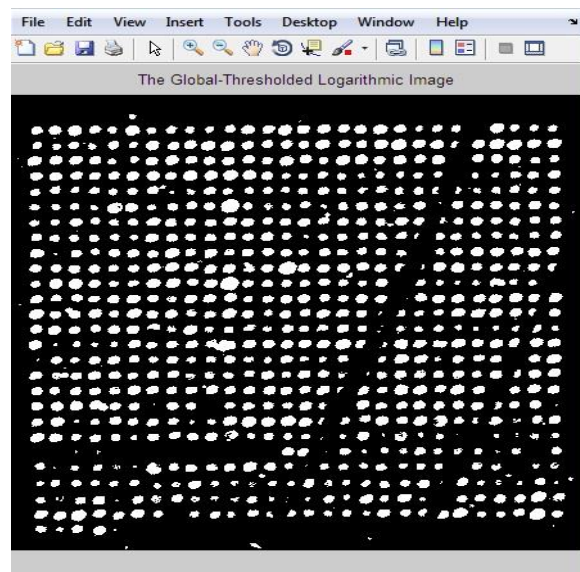


Figure 8: Image after Applying Global Thresholding

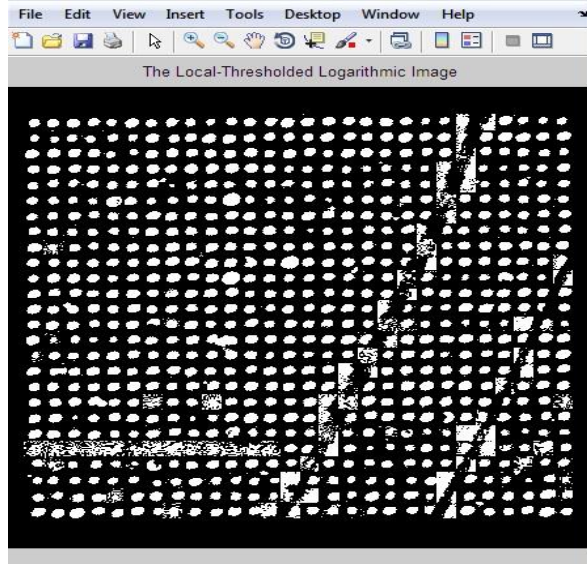


Figure 9: Image after Applying Local Thresholding

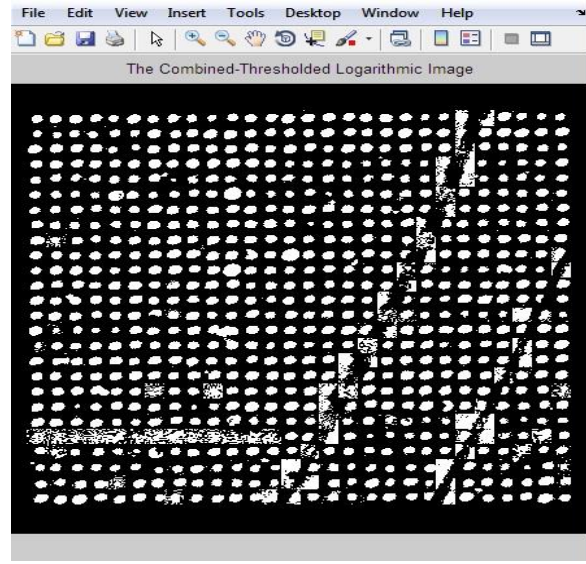


Figure 10: Image after Applying Combined- Thresholding

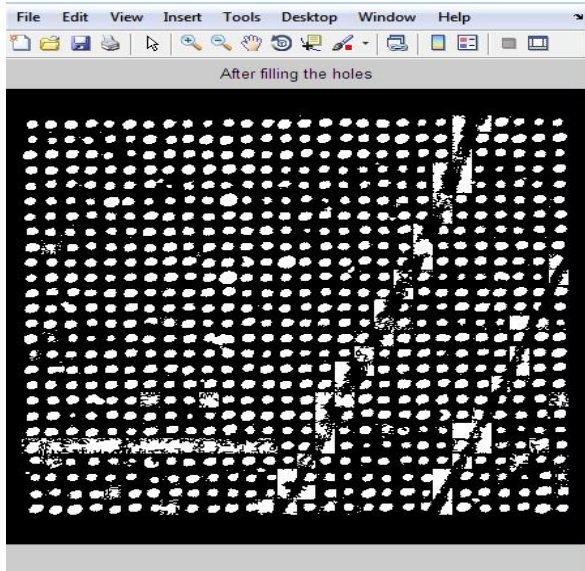


Figure 11: Image after Filling the Pinholes Method

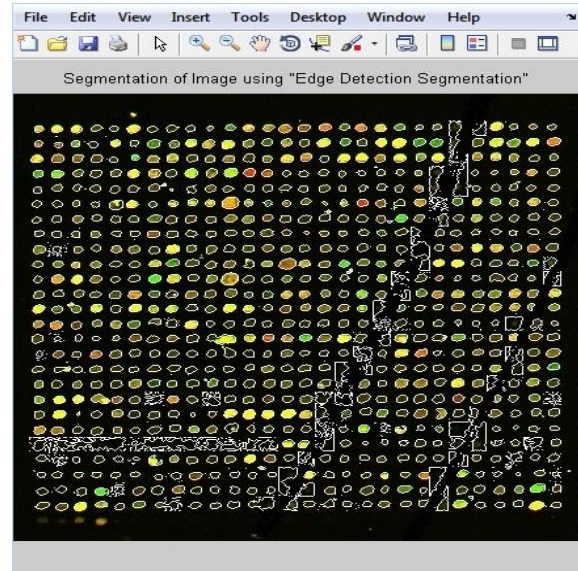


Figure 12: Image after Applying "Thresholding Segmentation" Method

Figure 7 to figure 12 show the results of applying thresholding segmentation on the microarray image. It is clearly observed that when using the current segmentation method to separate background from foreground in microarray images, it gives good results in the images that have clear spots. But it cannot deal well enough with weak spots.

4.3 Adaptive Shape Segmentation

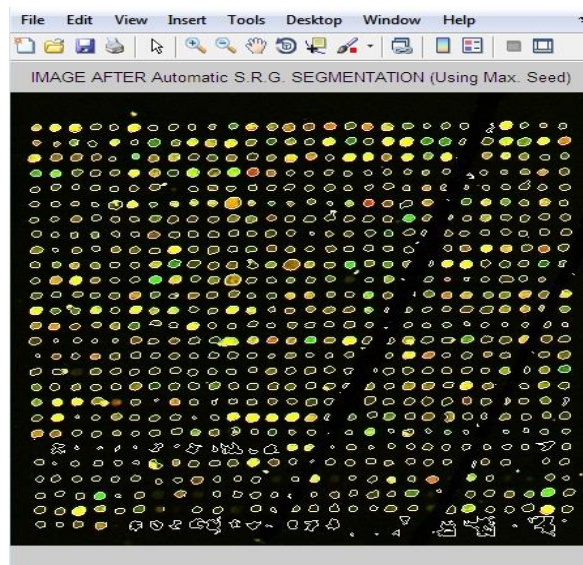


Figure 13: Image after Applying "Adaptive Shape Segmentation" Method.

Figure 13 shows the result of applying the adaptive shape segmentation method using a maximum seed point. This method requires an initial point to be known, which is called the seed. An advantage of using

SRG in microarray image segmentation is that the location of foreground pixels and background pixels can be estimated. It also deals with different shapes and sizes of spots.

The accuracy of the proposed methods 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 [15]. By comparing the results of applying the four presented segmentation methods; Fixed Circle Segmentation, Adaptive Circle Segmentation, Thresholding Segmentation, Adaptive Shape Segmentation methods, It is clearly obvious that the Adaptive Shape Segmentation method can segment noisy microarray images correctly despite of the degree of noise and the shape and size of the spots. That is shown in table 1.

Table 1: Comparison between the Four Proposed Segmentation Methods

Method	Accuracy	Segmented Percentage
Fixed Circle Segmentation	Good	79%
Adaptive Circle Segmentation	Very Good	86%
Thresholding Segmentation	Very Good	90%
Adaptive Shape Segmentation	Excellent	98%

5 Conclusion

The development of biomedical research has been led by the increasing knowledge as well as new advances in technology. Traditionally, researchers were able to investigate a small number of genes at a time by using the available techniques back then. DNA microarray technology has enabled biologists to study all the genes within an entire organism to obtain a global view of genes’ interaction and regulation. This technology has a great potential in obtaining a deep understanding of the functional organization of cells. The emergence of this technology allows the researchers to tackle difficult problems and reveal promising solutions in many fields

In this respect, our research based heavily on the requirement for a reliable yet time efficient automated method. One of the key steps in extracting information from a microarray image is the segmentation whose aim is to identify which pixels within an image represent which gene. This task is greatly complicated by noise within the image and a wide degree of variation in the values of the pixels belonging to a typical spot.

This work emphasizes the impact of microarray image segmentation. In the segmentation stage, four segmentation methods have been presented; “fixed circle segmentation”, “adaptive circle segmentation”, “thresholding segmentation”, and “adaptive shape segmentation” methods. When the four proposed methods were compared, it was found that the first two methods are shape-based segmentation methods. Fixed circle segmentation cannot segment clearly all types of images; when the spots size varies in the image, it will work bad. While the adaptive circle segmentation method achieves better results for circle-shaped spots. The other two segmentation methods; “adaptive shape segmentation” and “thresholding segmentation” gives better results. When using “thresholding segmentation” to separate background from foreground in microarray images, the current methods

cannot deal well enough with weak spots. So, it was very clear that “adaptive shape segmentation” gives better results using a maximum seed.

REFERENCES

- [1]. Istepanian, R., Microarray image processing: Current status and future directions. *IEEE Tran. Nanobioscience*, 2: 4, 2003
- [2]. T. Asano, D. Chen, N. Katoh, T. Tokuyama. Polynomial-time solutions to image segmentation, *Proc. of the 7th Ann. SIAM-ACM Conference on Discrete Algorithms*, 104-113, 1996.
- [3]. A. P. G. Damiance, L. Zhao, and A. C. P. L. F. Carvalho, A dynamic model with adaptive pixel moving for microarray images segmentation, *Real-Time Imaging* 10 (2004), 189-195.
- [4]. P. Arena, M. Bucolo, L. Fortuna, and L. Occhipinty, Cellular neural networks for real-time DNA microarray analysis, *IEEE Eng Med Bio* 21 (2002), 17-25.
- [5]. X. Y. Zhang, F. Chen, Y. T. Zhang, S. G. Agner, M. Akay, Z. H. Lu, M. M. Y. Waye, and S. K. W. Tsui, Signal Processing techniques in genomic engineering, *Proc IEEE* 90 (2002), 1822-1833
- [6]. R. Hirata Jr, J. Barrera, R. F. Hashimoto, D. O. Dantas, and G. H. Esteves, Segmentation of microarray images by mathematical morphology, *Real-Time Imaging* 8 (2002), 491-505.
- [7]. Katzer, M., Kummert, F. & Sagerer, G. (2003). Methods for automatic microarray image segmentation. *IEEE Transactions on Nanobioscience*, 2, 202-214.
- [8]. R. Nagarajan, Intensity-based segmentation of microarrays images, *IEEE Trans Med Imaging* 22 (2003), 882-889.
- [9]. A. W. C. Liew, H. Yana, and M. Yang, Robust adaptive spot segmentation of DNA microarray images, *Pattern Recognit* 36 (2003), 1251-1254.
- [10]. Yang, Y.H., Buckley, M.J., Dudoit, S. & Speed, T.P. (2002). Comparison of Methods for Image Analysis on cDNA Microarray Data. *Journal of Computational & Graphical Statistics*, 11, 108-136.v
- [11]. Y. Chen, E. Dougherty, and M. Bittner, Ratio-based decisions and the quantitative analysis of cDNA microarray images, *J Biomed Opt* 2 (1997), 364-374.
- [12]. Princeton University Microarray Database (PUMA; <http://puma.princeton.edu/>)
- [13]. Matlab (R2012b) Image Processing Toolbox, Signal Processing Toolbox.
- [14]. Islam Fouad, Mai Mabrouk, Amr Sharawy, “A Fully Automated Method for Noisy cDNA Microarray Image Quantification”, *International Journal of Computers and Technology*, Vol.11, No. 3, October 2013.
- [15]. Dimitris Maroulis, Eleni Zacharia, Microarray Image Segmentation using Spot Morphological Model, *IEEE*, July, 2009.