

# A Study on Microarray Image Gridding Techniques for DNA Analysis

Maziidah Mukhtar Ahmad  
School of Microelectronic Eng.  
University Malaysia Perlis  
Perlis, Malaysia  
jjdan8189@gmail.com

Asral Bahari Jambek  
School of Microelectronic Eng.  
University Malaysia Perlis  
Perlis, Malaysia  
asral@unimap.edu.my

Mohd Yusoff bin Mashor  
School of Mechatronic Engineering  
University Malaysia Perlis  
Perlis, Malaysia  
yusoff@unimap.edu.my

**Abstract**—Microarray is one of the most promising tools available for researchers in the life sciences to study gene expression profiles. Through microarray analysis, gene expression levels can be obtained, and the biological information of a disease can be identified. The gene expression information embedded in the microarray is extracted using image-processing techniques. Gridding is one of the important processes used to extract features in DNA microarray, by assigning each spot in the microarray with individual coordinates for further data interpretation. This paper evaluates popular techniques of DNA microarray image gridding in the literature with an emphasis on gridding accuracy, speed, and the ability to remove noise. Based on our evaluation, the Otsu method can provide a better performance in terms of processing speed, accuracy, and ability to remove noise compared to other methods discussed in this paper.

**Keywords** — Image Processing; Microarray; Gridding

## I. INTRODUCTION

Microarray technology was invented in 1995 by M. Schena et al. [1] and has since has been used as an important tool for gene studies. The chip contains abundant levels of deoxyribonucleic acid (DNA) with its own unique location for each spot. This allows estimation of the expression of thousands of genes simultaneously. It presents hidden biological information [2], identifies suitable drugs and their treatment forms [3]; and identifies gene expression for cancers. By analysing and comparing both normal and abnormal microarray gene expression profiling, the genes involved in that particular disease can be identified.

Three steps are used for microarray processing: gridding, segmentation and information extraction. The first step is important because errors at an early stage can lead to incorrect results [4]. Microarray image segmentation is a complex and time-consuming process as it contains image noise, non-homogeneous objects, and features low intensity

and weak contrast [5]. This also affects feature extraction accuracy.

Typically DNA microarray images are represented by 2D images that use two fluorescent dyes—red and green—to label the cDNA samples. In microarray image processing, gridding is important for assigning coordinates for every element of the spot. Spots will be separated from each other by horizontal and vertical grid lines, as shown in Figure 1. This process registers a set of irregularly spaced, parallel and perpendicular lines on the image. Depending on the amount of human intervention required during the process, there are three types of microarray gridding methods: manual, semi-auto and automatic. While each method offers different amounts of gridding accuracy, these methods share common problems such as grid misalignment and high image noise levels. These will impact on image gridding accuracy, and thus on extracting accurate DNA information.

This paper discusses several methods for gridding microarray that have been proposed in the literature. These methods will be analysed and compared to identify the best gridding method in terms of accuracy and speed. The rest of the paper is organized as follows: section two reviews the existing method for gridding microarray, section three discusses the performance comparison between input methods, and finally the conclusion is discussed in section four.

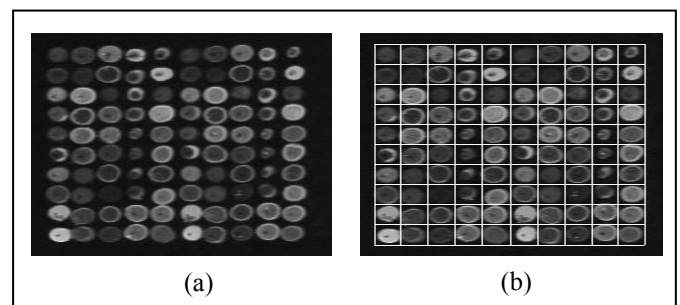


Figure 1: Microarray image gridding (a) without gridding (b) with gridding

## II. MICROARRAY IMAGE GRIDDING

In this section, five existing methods of DNA microarray gridding from the literature will be discussed. Katzer et al. [7] utilize the Markov random field (MRF) model. This method models microarray gridding by the origin, two-axis points, and the regular grid along each axis. Several steps must be completed including MRF site identification, labels, potentials and energy. Next, heuristic modelling grid placement hypotheses are performed. Finally, MRF energy minimization is performed. This method is not suitable for microarray gridding since a complete search for the optimal MRF configuration is not feasible. Furthermore, noise removal cannot be performed in this method because the quality of certain microarray background colours can create difficulties in removing noise. Figure 2 shows a histogram determining the centres of each region as discussed in [7].

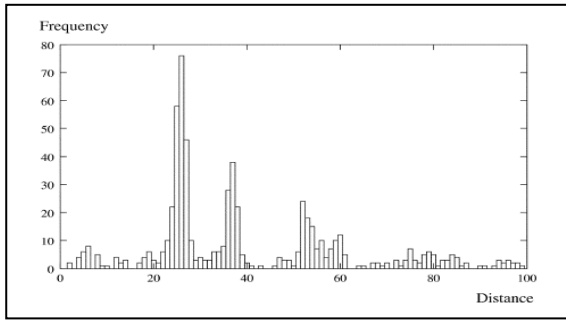


Figure 2: Markov random field [7].

Bajcsy [8] discusses the automatic grid alignment for DNA microarray based on template matching. The block diagram illustrating this method is shown in Figure 3. To calculate the grid, this method first finds a 2D array of rows and columns of the image, together with its rotational and translational offsets. The intersection between two rows and columns will be designated as the spot centre. Due to background clutter, this method did not perform noise removal.

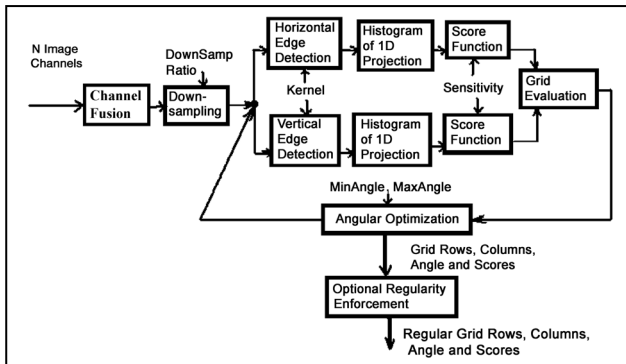


Figure 3: Grid alignment algorithm [8].

Bariamis and Maroulis [10] utilize a support vector machine (SVM), as shown in Figure 4. The method uses a set of soft-margin SVMs to forecast the lines of the DNA microarray grid by maximizing the margin between the lines and the spots. The authors automatically set the separation line between consecutive rows and columns to maximize the margin between the spots and the lines. Then, the spot detection step selects the places that have specific properties before filtering out any anomalies and artefacts. The remaining spots are then automatically split into rows and columns by calculating the distance between two consecutive rows and columns of spots, as well as the image rotation angle.

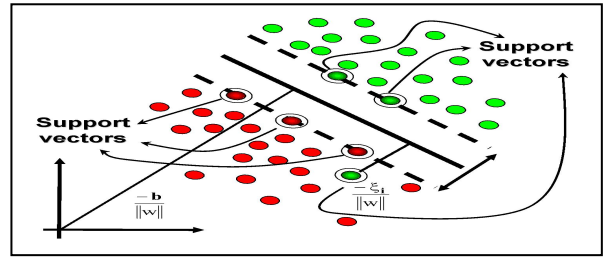


Figure 4: Support vector machine [10].

Other methods of gridding microarray image include clustering-based spot segmentation using the K-means method [11] as shown in Figure 5. The K-means method is one of the simplest techniques compared to other popular cluster methods, and is operated as follows. First, the centroid of spots and background are initialized. Then, the each pixel is assigned to the closest cluster. Lastly, a new centroid is computed after all pixels are clustered. At this stage the centroid of the new cluster is recalculated, as shown in Figure 5.

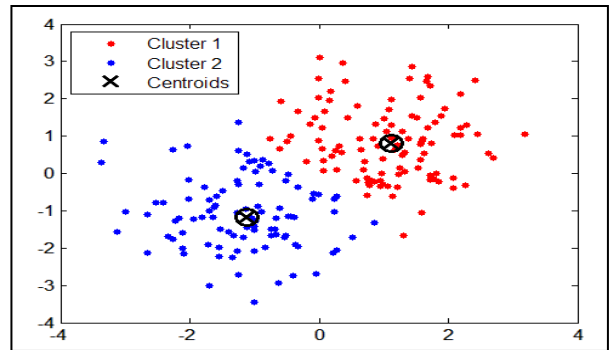


Figure 5: K-means clustering [15].

The Otsu method is discussed in [12]. This method converts a greyscale image into a purely binary image that

separate the pixels into two classes: background and foreground image. An optimum threshold is selected to maximize this separation. In this method, the colour image will first be converted to greyscale image as shown in Figure 6(a). Then the image is sharpened using image edge processing as shown in Figure 6(b). Then, removal of the unwanted part is performed to obtain a black and white image shown in Figure 6(c) before the grid can be calculated correctly.

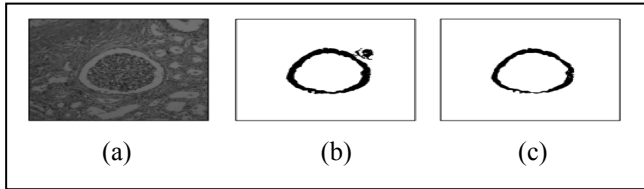


Figure 6: Otsu method [12]: (a) original image (b) edge detection image (c) Otsu image

### III. RESULT AND DISCUSSION

Table 1 shows the comparison of five gridding methods as discussed in Section II. In the table, factors such as accuracy, processing speed and noise removal ability are compared. The data are taken from reference [7, 8, 10, 11, 12]. Since the data are taken from different sources, each piece of literature may have used different benchmark data. However, for comparison purposes, the reported accuracy can give us some indication of the performance of each algorithm.

From Table 1, it is obvious that methods that implement noise removal result in higher gridding accuracy compared to methods without noise removal capability. By implementing noise removal prior the gridding process, the algorithm can extract the gridding information with more than 95% accuracy. Of all the methods discussed in this paper, the Otsu method [12] gives the best gridding performance at 99.5% accuracy. Since this method removes most of the noise that can disturb the gridding process, the method can achieve good gridding accuracy.

To verify our analysis, we have implemented our microarray gridding algorithm using the Otsu method in MATLAB software. In this experiment, the microarray image is first split into green, red and yellow image planes. Then, each image plane is converted into greyscale before it is processed using the Otsu method. The result from each image plane is then combined to form the final result. The experiment is performed using several images from [15] that represent different image qualities: high, moderate and poor, as shown in Figure 7.

Table 1: Comparison of Gridding Microarray

| Method                      | Accuracy       | Speed (in seconds) | Noise Removal |
|-----------------------------|----------------|--------------------|---------------|
| Markov Random Field [7]     | Low 89.4%      | Low (N/A)          | No            |
| Gridline [8]                | Moderate 93.6% | Low 1.75           | No            |
| Support Vector Machine [10] | High 95.1%     | Low (N/A)          | Yes           |
| K-Means Clustering [11]     | High 94.89%    | High (N/A)         | Yes           |
| Otsu [12]                   | High 99.5%     | High 0.94          | Yes           |

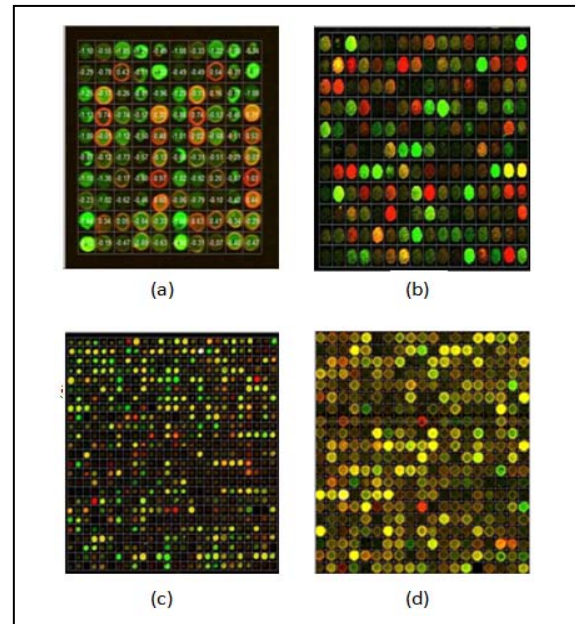


Figure 7: Result of microarray image gridding using the Otsu method on different image qualities: (a) high quality, (b) and (c) moderate quality, (d) poor quality

Table 2 summarizes the results of our experiment. The accuracy of image gridding is calculated using Equation 1 [14], where  $N_{\text{correct spots}}$  and  $N_{\text{total spots}}$  represent the number of spots correctly gridded and the total number of spots in images, respectively. Processing time is defined as time taken for MATLAB to complete the computation. These programs are simulated using a laptop computer (processor: Intel(R) Core (TM) i3- 2.10GHz, 2.00 GB RAM). Based on the table, the Otsu method gives 100% accuracy for high quality images, 94%-97% accuracy for medium quality images, and 89.9% accuracy for poor quality images. This shows that the Otsu method could achieve a good accuracy for a wide range of image qualities.

$$A = (N_{\text{correct spots}} / N_{\text{total spots}}) * 100\% \quad (1)$$

Table 2: Accuracy and processing time

| Images | Number of spot | Accuracy | Processing time(sec) |
|--------|----------------|----------|----------------------|
| a      | 100            | 100%     | 18                   |
| b      | 176            | 96.6%    | 9                    |
| c      | 650            | 94.2%    | 18                   |
| d      | 380            | 89.9%    | 14                   |

#### IV. CONCLUSION

In this paper, the existing microarray gridding methods have been analysed and compared. These were the Markov random field [7], gridline [8], support vector machine [10], K-means clustering [11] and Otsu methods [12]. Based on our analysis, the methods that utilize noise removal before performing image gridding give a higher accuracy (>95%) compared to methods that do not perform noise removal. Among the reviewed methods, the Otsu method [12] gives the best results with 99.5% reported accuracy. Our experiment, using various microarray images, confirm this. In the next phase of this work, we will further analyse the Otsu algorithm to study its effectiveness in hardware implementation targeting for a portable bio-medical application.

#### References

[1] Schena M., Shalon D., Davis R.W., and Brown P.O, 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray'. *Science*, 270, 1995, pp. 467-470.

[2] Guillermo, H. L., R. L. Alajandro, A. S. Jose, P.V. Beatriz and Fernando M. S. 'Microarray and colon cancer in the road for translational medicine'. *Current Bioinformatics*, June 2011.

[3] Debouck, C., and Goodfellow, P.N. 'DNA Microarray in Drug Discovery and Development'. *Nat Genetics*, 1999.

[4] White, A. M., Daly D. S., Willse A.R., Protic M., and Chandler D.P. 'Automated microarray image analysis toolbox for MATLAB'. Oxford University Press, Oxford, 2005.

[5] Solomon, C., and Breckon T. *Fundamentals of digital image processing: a practical approach with examples in MATLAB*. 2011.

[6] Ceccarelli M., and Antoniol G. A deformable grid-matching approach for microarray images. *IEEE Transactions on Image Processing*, 15(10): 3178-3188, 2006.

[7] Katzer M., Kummert F., and Sagerer G. 'A Markov random field model of microarray gridding'. *Proc. 18th ACM Symp. Applied Computing*, pp. 72-77, 2003.

[8] Bajcsy P. 'Gridline: automatic grid alignment in DNA microarray scans'. *IEEE transactions on image processing*, vol. 13, no. 1, January 2004.

[9] Jain A. N., Tokuyasu T. A., Snijders A. M., Segraves R., Albertson D. G., and Pinkel D. 'Fully automated quantification of microarray image data,' *Genome Res.*, vol. 12, no. 2, pp. 325-332, Feb. 2008.

[10] Zacharia E., Maroulis D. 'An original genetic approach to the fully automatic gridding of microarray images'. *IEEE Transactions on Medical Imaging*, vol. 27, no. 6, 2008.

[11] Uslan V., Bucak I.O. Clustering-based spot segmentation of cDNA microarray images. *Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*, pp. 1828-1831, 2010.

[12] Zacharia E., Maroulis D. A spot modeling evolutionary algorithm for segmenting microarray images. *Evolutionary Algorithms*, pp. 459- 480, 2012.

[13] Bishop C. *Pattern Recognition and Machine Learning*. Springer, pp. 383-393, 2006.

[14] Azeem I.A., Mabrouk M.S. 'Developing a New Methodology for Denoising and Gridding cDNA Microarray Images'. *IEEE Cairo International Biomedical Engineering Conference*, pp.145, Dec 2012.

[15] Sherlock G., Kasarkis A. and Brown P.O. 'The Stanford Microarray Database', *Nucleic Acid Research*, vol. 29, no.1, 2001.

