MICROARRAY IMAGE GRIDDING USING GRID LINE REFINEMENT TECHNIQUE

V.G. Biju¹ and P. Mythili²

School of Engineering, Cochin University of Science and Technology, India E-mail: ¹bvgpillai@gmail.com, ²mythili@cusat.ac.in

Abstract

An important stage in microarray image analysis is gridding. Microarray image gridding is done to locate sub arrays in a microarray image and find co-ordinates of spots within each sub array. For accurate identification of spots, most of the proposed gridding methods require human intervention. In this paper a fully automatic gridding method which enhances spot intensity in the preprocessing step as per a histogram based threshold method is used. The gridding step finds co-ordinates of spots from horizontal and vertical profile of the image. To correct errors due to the grid line placement, a grid line refinement technique is proposed. The algorithm is applied on different image databases and results are compared based on spot detection accuracy and time. An average spot detection accuracy of 95.06% depicts the proposed method's flexibility and accuracy in finding the spot co-ordinates for different database images.

Keywords:

cDNA Microarray, Gridding, Image Processing, Spot Detection, Grid Line Refinement

1. INTRODUCTION

Complementary DNA (cDNA) microarrays are part of a rapidly developing technology in molecular biology, allowing one to measure simultaneously the activity of thousands of biomolecules in the cell under different experimental conditions. Novelty in the field of biotechnology allows monitoring of expression levels of thousands of genes simultaneously. Microarray image is a powerful tool in many biotechnological applications such as cancer research, chromosomal abnormalities, artery diseases, drug discovery and disease diagnosis. Enormous improvement of technology in last decade makes it possible to simultaneously identify and quantify thousands of genes by their gene expression [1]-[4].

Analysis of cDNA microarray images include mainly three steps namely, gridding, segmentation and intensity extraction. The gridding process is usually divided into two main steps, sub array gridding and spot detection. Most of the existing gridding algorithms are semiautomatic requiring human intervention and parameter presetting. Each spot is associated with a gene which contains the pixels that indicate the level of expression of that particular gene. Improper gridding will affect image analysis in extracting intensity values from each spot that represents gene expression level.

Over the past decade several software packages such as ScanAlyze [5], GenePix Instruments [6], Quant Array [7], Spot Finder [8], ImaGene [9] and Dapple [10] were developed. For most of these packages parameters need to be preset manually and sometimes human intervention was necessary to locate the exact spot centers. Jung and Cho [11] proposed a graph based grid approach for gridding of microarray images. A drawback of this algorithm was that it requires input parameters such as number of rows and columns of sub arrays and spots within each sub array in an image. Deng and Duan [12] proposed an axis projection based approach for gridding and it requires user intervention in order to manually adjust the grid location.

A variety of methodologies have been proposed to solve rotation and misalignment problems [13], [14]. Brandle et al. [15] utilized discrete Radon transform for this purpose. Zacharia and Maroulis [16] proposed a Genetic algorithm (GA) based method which accurately determines line segments, constituting borders between adjacent blocks or spots. It does not require any input parameters or human intervention but was time consuming. L. Rueda and I. Rezaeian [17] proposed a fully automatic approach for microarray gridding. The method first finds location of sub arrays of the entire image and then finds coordinates of spots in each sub arrays. The method detects and corrects rotation in the image by applying affine transform.

Gui-Fang S et al. [18] presented a fully automatic gridding technique based on improved Otsu threshold method for locating each spot, co-ordinates in the microarray image. A preprocessing algorithm was applied for noise reduction and then gridding algorithm was applied. In order to reduce the error after the initial grid line placement, the grid lines were optimized by a heuristic technique with the help of estimating the distribution of the spots.

In this paper a fully automatic gridding algorithm which improves spot detection accuracy and computational efficiency irrespective of microarray image database is proposed. The method first detects and corrects image tilt and then enhances image spot intensity in the preprocessing step using a histogram based threshold. Gridding step finds co-ordinates of spots from the horizontal and vertical profile of the image. The proposed grid line refinement technique further improves gridding.

2. THE PROPOSED METHOD

Microarray image analysis involves mainly three steps namely Gridding, Segmentation Intensity extraction and Quantification. The proposed method deals with microarray image gridding. Fig.1 shows the block diagram of proposed method. The method first checks and corrects tilt in the image. This process is important because the tilt in the image makes the whole gridding process ineffective. The second step preprocessing, helps to perform gridding process effectively. The third step gridding finds co-ordinates of sub arrays in an image and also spot co-ordinates within the sub array. Finally, grid line refinement method, corrects the grid line placement errors. Imperfect spot gridding will affect the forthcoming steps such as spot segmentation, gene expression calculation and biological conclusion. So, perfect gridding is a prime step in microarray image analysis.



Fig.1. Block diagram of the proposed method

2.1 IMAGE TILT DETECTION AND CORRECTION

A tilt present in microarray image decreases the gridding efficiency. Hence any tilt in the microarray image has to be rectified before gridding process. Image tilt can be calculated in two stages 1) Tilt identification and 2) Tilt angle calculation. Tilts in the images are observed in two different directions, with respect to the x and y axes. To find the tilt angle in the image, Radon transform is applied. The Radon transform, transforms microarray image into a new image plane with parameters Φ and t. Each point in new image plane accumulates all information corresponding to a line in original image with angle Φ and radius t. When Radon transform localizes a local maximum near an angle Φ_0 and around a slice t_0 , it means the original image has a line in position (Φ_0 , t_0) [19], [20]. Applying Radon transform to an image A(x, y), for a given set of angles can be thought of as computing projection of the image along given angles. The resulting projection is the sum of intensities of the pixels in each direction, i.e. an integral line. Result is a straightened image $R(\rho, \theta).$

$$R(\rho,\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x,y) \delta(\rho - x\cos\theta - y\sin\theta) dxdy \qquad (1)$$

where, ρ is defined as

$$\rho = x\cos\theta + y\sin\theta \tag{2}$$

where, $\delta(0)$ is Dirac delta function and ρ is the perpendicular distance of a line from the origin and θ is angle between line and y-axis.

2.2 PREPROCESSING

Real cDNA microarray image contains a number of sub arrays, and each sub array contains large number of spots arranged in rows and columns.

In the pre-processing stage, the given image is first converted to grayscale. A histogram based threshold is used to replace image background with zero intensity value and enhance spot intensity by using a scaling factor. The edges of spots are detected using canny edge detection. Morphological operations such as dilation and erosion are done after filling spots. Median filtering is done to remove isolated noisy pixels.

2.3 GRIDDING

The aim of gridding is to find the sub array locations in the image and then the spot location co-ordinates within the sub array. The gridding is used to detect each spot location from the image. To find the location of spots, coordinates of each spot is necessary.

Consider a microarray image A having p rows and q columns of sub arrays and each sub array with m rows and n columns of spots. The first step in gridding finds vertical and horizontal sub array location co-ordinates V and H, $V = [V_1, V_2... V_{q+1}]$ and H = $[H_1, H_2...H_{p+1}]$. Second step finds spot location co-ordinates within each sub arrays v and h, where $v = [v_1, v_2... v_{n+1}]$ and h = $[h_1, h_2...h_{m+1}]$, m+1 and n+1 represents total number of horizontal and vertical gridlines required to separate each spot within a sub array. $S = \{S_{ij}\}$, where S is a sub array in image. S_{ij} represents intensity values of each pixel in a sub array and i =1,..., a and j = 1,...,b where $a \times b$ represents the sub array size. The area between adjacent horizontal and vertical vectors h_{i+1} - h_i and v_{i+1} - v_i delimit the area corresponding to a spot. To obtain the co-ordinates, horizontal and vertical intensity profile of the preprocessed image is first computed using Eq.(3) and Eq.(4) respectively.

$$h = \frac{1}{a} \sum_{i=1}^{a} I(i, j) \tag{3}$$

$$v = \frac{1}{b} \sum_{j=1}^{b} I(i, j) \tag{4}$$

The horizontal and vertical intensity profiles are converted to a binary signal, the signals are inverted and autocorrelation of the signals are computed. The peak location coordinates of the auto correlated signals are estimated and are used as the horizontal and vertical co-ordinate of grid lines. The slope of intensity signals, gives the periodicity of spots in the image. The spot layout is calculated from the periodicity of spot in horizontal and vertical directions. The horizontal and vertical start and end lines are computed using a histogram based threshold.

An ideal microarray image has uniform spot size, inter spot distance, and alignment throughout the image. In real microarray images, spot size, inter spot distance and alignment of spots vary with image databases. Noise artifacts and deformation of images, which affect gridding also vary with image databases. The existing gridline optimization or grid line refinement techniques are not able to grid all database images successfully. Hence a more flexible and accurate grid line refinement method is suggested.

2.4 GRID LINE REFINEMENT

In practical images, grid line placement errors are due to redundant lines, missing lines and misplacement of lines. A sub array image with spot layout $m \times n$ requires m+1 horizontal and n+1 vertical grid lines. The occurrence of an increase or decreases of gridline are cases of redundant or missing grid lines respectively.

Grid lines placed inside the spots are cases of misplaced lines, provided the required number of lines in both horizontal and vertical directions are satisfied. A grid line refinement method is proposed in this paper to detect and correct the errors mentioned above. Two parameters need to be computed for detecting the errors.

1) Spacing between horizontal and vertical grid lines $h_d(k)$ and $v_d(l)$.

2) Mean horizontal and vertical grid line spacing distance $\overline{h_d}$ and $\overline{v_d}$.

The parameters, $h_d(k)$ and $v_d(l)$, $\overline{h_d}$ and $\overline{v_d}$ are computed using Eq. (5), Eq.(6), Eq.(7) and Eq.(8).

$$h_d(k) = h_l(k+1) - h_l(k) \ k \in [1, m+1]$$
 (5)

$$v_d(l) = v_l(l+1) - v_l(l) \ l \in [1, n+1]$$
 (6)

$$\overline{h_d} = \frac{1}{m+1} \sum_{k=1}^{m+1} h_d(k)$$
(7)

$$\overline{v_d} = \frac{1}{n+1} \sum_{l=1}^{n+1} v_d(l)$$
(8)

where, h_l and v_l denotes horizontal and vertical grid lines vectors and k and l denote the grid line number.

Redundant line or misplaced grid line is detected in $(k+1)^{th}$ and $(l+1)^{th}$ location if Eq.(9) and Eq.(10) are satisfied respectively.

$$h_d(k) < 0.8h_d \tag{9}$$

$$v_d(l) < 0.8 \overline{v_d} \tag{10}$$

The grid line vectors are updated by deleting the $(k+1)^{th}$ and $(l+1)^{th}$ line as follows:

$$h_l = [h_l(1:k) h_l(k+2:end)]$$

 $v_l = [v_l(1:l) v_l(l+2:end)]$

Missing grid lines are detected in $(k+1)^{\text{th}}$ and $(l+1)^{\text{th}}$ location if Eq.(11) and Eq.(12) are satisfied respectively.

$$h_d(k) < 1.5h_d \tag{11}$$

$$v_d(l) < 1.5 v_d \tag{12}$$

Error can be corrected by updating horizontal and vertical grid line vectors by inserting grid lines in $(k+1)^{th}$ and $(l+1)^{th}$ location as follows.

$$h_{l}(k+1) = h_{l}(k) + 0.9 * h_{d}$$

$$v_{l}(l+1) = v_{l}(l) + 0.9 * \overline{v_{d}}$$

$$tempx = h(k+1:end)$$

$$tempy = v(l+1:end)$$

$$h_{l} = [h_{l}(1:k) h_{l}(k+1) tempx]$$

$$v_{l} = [v_{l}(1:l) v_{l}(l+1) tempy]$$

where, *tempx* and *tempy* are temporary variables used to place the missing line co-ordinate. The above mentioned threshold values are obtained using trial and error method by applying the algorithm on different database images.

3. DATABASE USED

The algorithm is tested with six different image database to validate accuracy and flexibility of proposed method in microarray image gridding. Spots in an image vary in diameter, spacing and alignment in both horizontal and vertical direction. The scanning resolution of an image is also different. Database1 includes thirteen images selected from UNC microarray image database, with experiment IDs 31466 to 31478 under "Cell line" as category and "drug treatment" as sub category. Each image consists of 32 sub arrays and each sub array have 625 spots.

Database2 includes images selected from Stanford microarray data base (SMD), with experiment IDs 20385, 20387, 20391, 20392, and 20395 by selecting "Hormone treatment" as category and "Transcription factors" as subcategory. Each image consists of 32 sub arrays and each sub array have 270 spots.

Database3 includes fourteen microarray images selected from Computational Cancer Genomics (CCG) group of the Swiss Institute of Bioinformatics (SIB) (http://www.isrec.isbsib.ch/). Experiment IDs of images are 661 to 667. Each image includes 4 sub arrays and each sub array with different spot numbers, which varies from 42 to 49.

Database4 includes images downloaded from following website address (http://www.bio.davidson.edu/ projects/magic/magic.html,DeRisi). Database4 includes images from channels 1 and 2 with experiment IDs 1302, 1303, 1309 to 1313. Each image contains four microarray sub arrays and each sub array contains 1,600 spots.

Database5 contains two real images with 36 sub arrays and each sub array with 210 spots. The images can be downloaded from the University of California, San Francisco (UCSF, http://cancer.ucsf.edu/cores/array-sample data).

Database6 includes images from Gene Expression Omnibus. downloaded Images are from website http://ncbi.nlm.nih.gov/geo/GEO. Thirteen images are downloaded corresponding to channels 1 and 2 with experiment IDsGSM15898, GSM16101. GSM16389, GSM16391. GSM17137, GSM17186, GSM17190 and GSM17192. Each sub array has 182 spots.

4. RESULTS AND DISCUSSION

4.1 IMAGE TILT DETECTION AND CORRECTION

To make the whole gridding process effective, microarray image tilt detection and correction are important. To detect microarray image tilt and angle calculation Radon transform is used. To find the accuracy of Radon transform in estimating tilt angle of real microarray images, an artificially tilted microarray sub array image with several angles in both clock wise and counter clockwise directions are used. The tilted angles are estimated by applying Radon transform. The result shows that the angle estimated by the transform is same as that of the artificially tilted angle. So this can be applied successfully for the real microarray tilted images for estimating the unknown angle. Fig.2(a) shows a microarray image tilted by an angle 5 degree and Fig.2(b) shows the corrected image after estimating the angle using Radon transform.

Data Base	Genetic algorithm based method			Maximum between class variance based method			Proposed method		
	Incorrect	Marginal	Perfect	Incorrect	Marginal	Perfect	Incorrect	Marginal	Perfect
UNC	2.31	4.73	92.96	8.12	1.75	90.1	1.6	2.900	95.5
SMD	1.83	5.57	92.6	1.96	2.94	95.1	0.97	2.520	96.51
SIB	6.7	7.8	84.5	2.0	12.9	85.1	5.48	5.414	89.11
Dersi	2.1	6.9	91	5.238	1.33	93.4	4.625	0.563	94.81
UCSF	5.52	0.53	93.95	3.4	1.0	95.6	2.08	1.360	96.56
GEO	6.72	2.43	90.85	3.2	0.8	96	1.271	0.859	97.87

Table.1. Percentage spot detection accuracy obtained for genetic algorithm based method; maximum between class variance based method and proposed method



Fig 2(a). Tilted microarray image by a 5 degree angle, 2(b). Corrected image using radon transform

4.2 PERFORMANCE OF THE PROPOSED METHOD

The performance of the proposed method is tested with six different image database. The images from each database vary in image resolution, sub array layout, spot layout and spot resolution. Proposed method is compared with two existing method such as genetic algorithm based method [16] and maximum between class variance method [18]. Percentage spot detection accuracy obtained for grid lines which separate spots incorrectly, marginally and perfectly for all the six databases are shown in Table.1.

The Table.1 shows that the proposed method has an improvement of 2.54% spot detection accuracy compared to other methods for database1 images. Gridding results obtained for an image from database1 using genetic algorithm based method; maximum between class variance based method and proposed method are shown in Fig.3(a-c) respectively. In genetic algorithm based method the error in gridding is mainly due to

marginal grid line placement. Arrow pointed vertical grid lines in Fig.3(a) shows marginal gridline placement in the method. From Fig.3(a) it is also observed that in this method the bottom edge grid line is missing. Error in maximum between class variance based method is mainly due to incorrect gridline placement. Arrow pointed vertical and horizontal grid lines in Fig.3(b) support this contention.

The proposed method shows an improvement of 1.41% spot detection accuracy for database2 images compared to other two methods. The gridding results of three methods for an image from database2 are shown in Fig.4(a-c). The arrow pointed vertical grid lines in Fig.4(a) and Fig.4(b) show examples of marginal grid line placement obtained using first two methods for an image from database2. From Fig.(4b) it is also observed that in maximum between class variance method the bottom horizontal edge grid line is missing.





Fig.3(b)



Fig.3(c)







Fig.4(b)



Fig.4(a-c). Gridding results obtained using genetic algorithm based method, maximum between class variance method and the proposed method for a sub array image from database2

For database3 images, the perfectly gridded spots, spot detection accuracy is less than 90% for all the three methods. In this database the image spot diameter is less than inter spot distance which result in incorrect and marginal grid line placement errors. As the number of spots in a sub array is less than 50, a single instance of incorrect placement of gridline will considerably affect spot detection accuracy measure. Gridding results from a database3 image using the above three methods are shown in Fig.5(a-c). From Table.1 spot detection accuracy obtained for database3 images shows an improvement of 4 % for the proposed method compared to other methods. The arrow pointed vertical grid lines in Fig.5(a) show examples of incorrect and marginal gridline placement error for genetic algorithm based method.



Fig.5(a-c). Gridding result obtained using genetic algorithm based method, maximum between class variance and the proposed method for a sub array image from database3

Results in Table.1 for database4 images shows an improvement of 1.38 % spot detection accuracy for the proposed method compared to other methods. In this image database image spot layout and spot resolution is 40×40 (1600 spots in a sub array) and 8×8 respectively which result in spots of smaller size in the sub array. So some high noisy images in the database give incorrect gridding and marginal gridding error and hence the average spot detection accuracy is less than 95% for all the methods.

For database5 images the proposed method shows an improvement of 0.96% spot detection accuracy compared to other methods. Incorrect grid line placement error occurs for images in this database due to improperly aligned spots in some rows of the image.



Fig.6(a)



Fig.6(b)



Fig.6(c)

Fig. 6(a-c). Gridding result obtained using genetic algorithm based method, maximum between class variance and the proposed method for a sub array image from database6

The Table.1 shows an improvement of 1.87% spot detection accuracy for the proposed method compared to other methods. Fig.6(a-c) show gridding result of a database6 image using genetic algorithm based method, maximum between class variance based method and the proposed method respectively. Arrow pointed horizontal grid lines in Fig.6(a) shows redundant grid lines in non spot area for genetic algorithm based method.

The average spot detection accuracy obtained for all the six image databases using genetic algorithm based method, maximum between class variance based method and the proposed method are 90.977%, 92.55% and 95.06 % respectively. The proposed method shows an overall improvement of 2.51%.

Performance of all three methods are evaluated on Intel core i3 processor with 2 GB RAM. Sub array images used are from six dataset of different dimensions. Table.2 shows the average time taken for detecting a spot from a set of images in a database. The proposed method is compared with genetic algorithm based method and maximum between class variance based method on a time scale. Time elapsed for spot detection is computed by finding ratio of total time taken for gridding a sub array image to total number of spots in sub array.

In a microarray image, the number of spots in a sub array are large. So a small improvement in time taken in seconds for the proposed method compared to maximum between class variance based method is feasible. Spot detection accuracy and speed of the proposed method compared to other two existing methods shows improved flexibility and accuracy of the proposed method in microarray gridding.

Database	Genetic algorithm based method	Maximum between class variance based method	Proposed method
UNC	0.166	0.01	0.009
SMD	0.315	0.02	0.0197
SIB	0.91	0.04	0.033
Dersi	0.08	0.004	0.003
UCSF	0.292	0.008	0.008
GEO	0.4	0.017	0.011

Table.2. Average processing time for spot detection (seconds)

5. CONCLUSION

In this paper a grid line error correction method based on a refinement technique is proposed. The method is tested with microarray images drawn from six different image datasets of varying resolution. Result shows that the proposed method improves the spot detection accuracy irrespective of microarray image resolution, sub array lay out, spot lay out and inter spot distance without adjusting any other parameters. The spot detection time of the method is also compared with other methods. The overall result obtained on real microarray images shows that the proposed method is a good choice compared to genetic algorithm and maximum between class variance based method in microarray image gridding.

REFERENCES

- [1] Y. H. Yang, M. J. Buckley, S. Dudoit and T. P. Speed, "Comparison of methods for image analysis on cDNA microarray data", *Journal of Computational and Graphical Statistics*, Vol. 11, No. 1, pp. 108-136, 2002.
- [2] A. Lehmussola, P. Ruusuvuori and O. Yli-Harja, "Evaluating the performance of microarray segmentation algorithms", *Bioinformatics*, Vol. 22, No. 23, pp. 2910-2917, 2006.
- [3] A. N. Jain, T. A. Tokuyasu, A. M. Snijders, R. Segraves, D. G. Albertson and D. Pinkel, "Fully automatic quantification of microarray image data", *Genome research*, Vol. 12, No. 2, pp. 325-332, 2002.
- [4] J. Angulo and J. Serra, "Automatic analysis of dna microarray images using mathematical morphology", *Bioinformatics*, Vol. 19, No. 5, pp. 553-562, 2003.
- [5] M. Eisen, "Scanalyze User manual", Technical Report, 1999.
- [6] I. Axon Instruments, "Genepix 4000a Users manual", *Axon Instruments Inc*, 1999.
- [7] G. Luminomics, "Quantarray analysis software, Operators manual", *GSI Luminomics*, 1999.
- [8] P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee and J. Quackenbush, "A concise guide to cDNA microarray analysis", *Biotechniques*, Vol. 29, No. 3, pp. 548-563, 2000.
- [9] I. ImaGene, "ImaGene 6.1 user manual", 2008.
- [10] J. Buhler, T. Ideker and D. Haynor, "Dapple: improved techniques for finding spots on DNA microarrays", Technical Report, pp. 08-05, 2000.
- [11] H. Y. Jung and H.-G. Cho, "An automatic block and spot indexing with k-nearest neighbors graph for microarray

image analysis", *Bioinformatics*, Vol. 18, No. Suppl 2, pp. S141-S151, 2002.

- [12] N. Deng and H. Duan, "The automatic gridding algorithm based on projection for microvirray image", *Proceedings of IEEE International Conference on Intelligent Mechatronics* and Automation, pp. 254-257, 2004.
- [13] P. Bajcsy, "Gridline: automatic grid alignment DNA microarray scans", *IEEE Transactions on Image Processing*, Vol. 13, No. 1, pp. 15-25, 2004.
- [14] M. Steinfath, W. Wruck, H. Seidel, H. Lehrach, U. Radelof and J. OBrien, "Automated image analysis for array hybridization experiments", *Bioinformatics*, Vol. 17, No. 7, pp. 634-641, 2001.
- [15] N. Brandle, H. Bischof and H. Lapp, "Robust DNA microarray image analysis", *Machine Vision and Applications*, Vol. 15, No.1, pp. 11-28, 2001.
- [16] E. Zacharia and D. Maroulis, "An original genetic approach to the fully automatic gridding of microarray images", *IEEE Transactions on Medical Imaging*, Vol. 27, No. 6, pp. 805-813, 2008.
- [17] L. Rueda and I. Rezaeian, "A fully automatic gridding method for cdna microarray images", *BMC Bioinformatics*, Vol. 12, No. 113, p. 1-17, 2011.
- [18] G.-F. Shao, F. Yang, Q. Zhang, Q.-F. Zhou, and L. K. Luo, "Using the maximum between-class variance for automatic gridding of cdna microarray images", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 10, No. 1, pp. 181-192, 2013.
- [19] O. R. Terrades and E. Valveny, "Radon transform for linear symbol representation", *Proceedings Seventh IEEE International Conference on of Document Analysis and Recognition*, pp. 195-199, 2003.
- [20] G. Beylkin, "Discrete radon transform", *IEEE Transactions* on Acoustics, Speech and Signal Processing, Vol. ASSP-35, No. 2, pp.162-172, 1987.