

A Microarray Image Gridding Method Based on Projection Transformation and Power Spectral Analysis

Yan Jun Feng^{1, 2, a}, Song Kai^{1, 2, b}, Liu Jun^{1, c*}

¹ Shenyang Ligong University, School of Information Science and Engineering, Shenyang 110159, China

² Shenyang Agricultural University, Shenyang 110866, China

^aemail: braverfyj@126.com, ^bemail: ap9351@163.com, ^cemail: lj-mail-sut@163.com

Abstract

Microarray image gridding is one important step of microarray image analysis to determine 2D image coordinates of all array spots in the hybridized gene chip image. Accuracy of image gridding will affect the reliability of gene-chip data extraction and even the final analysis results of gene-chip assays. To promote microarray image gridding accuracy and computation efficiency, we presented a microarray image gridding method based on image projection transformation and power spectral analysis. Firstly we transformed 2D microarray image into vertical and horizontal 1D projection sequences, secondly utilized signal processing methods of low pass filtering, zero mean, FFT, and power spectral estimation by periodogram method to get spots array row and column span information, and finally realized the microarray image gridding according to the local maxima and span information of spots array. The gridding experiments showed that this method had met the requirements of computing accuracy and efficiency of microarray image gridding.

Keywords: Gene Chip; Microarray Image Gridding; Projection Transformation; Power Spectral Analysis

Introduction

Gene chip is also called cDNA microarray or DNA chip. Gene chip assay is a testing method of life sciences, which first makes biological genetic material-nucleic acid molecules amplified and hybridized on one glass slide surface, then detects hybridization signal of nucleic acid molecules on the gene chip and acquires microarray images, and ultimately achieves the aims of biological expression data extracting, handling and analyzing. Gene chip analyzing technology has been applied extensively in such fields as life science research, clinical disease diagnosis, food security and environment monitoring [1]

[2] [3] [12].

Fig.1 shows the typical working process of gene chip analyzing system. The gene chip is prepared by amplifying nucleic acid samples marked with fluorescent dyes and hybridizing amplified nucleic acid molecules with the arrayed probe spots (Oligonucleotide molecules) on the gene chip surface. The hybridized arrayed spots with different fluorescent intensity on the surface of gene chip are detected, imaged and scanned by an IR CCD camera, a fluorescent microscope, and scanning machines. The typical low-density fluorescent spots microarray image we utilized in this project is shown in Fig.2.

Microarray image gridding is one important step for gene chip image analysis [1] [2] [3]. Through microarray image gridding analysis, 2D image coordinates of each fluorescent spot on gene chip can be determined and prepared for the consecutive single spot image segmentation and fluorescent intensity (or gene expression) data extraction steps. Accuracy of gridding will affect the reliability of gene chip data extraction and even the final analysis results of gene chip assays.

In recent years, some microarray image gridding methods presented [4] [5] [10] [11] [12] had more accuracy but more computation complexity than the precedent ones. In this paper, we presented one serialized microarray image gridding method based on image projection sequences power spectral estimations and local maxima searching for projection sequences, which promoted gridding accuracy and decreased computation complexity significantly.

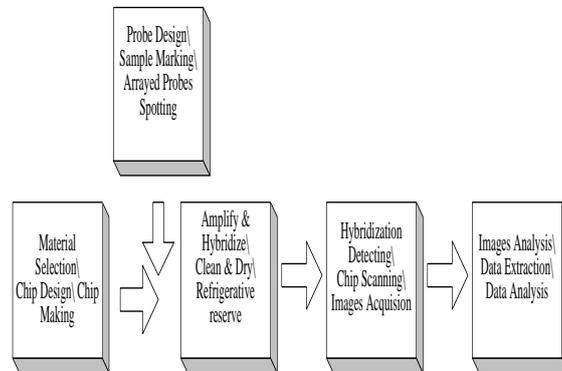


Fig.1 the working process of gene chip analyzing system

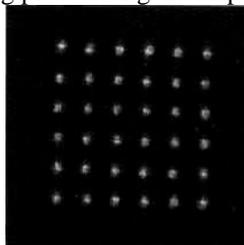


Fig.2 the low-density fluorescent spots microarray image

The principle of proposed method

Microarray image gridding is a serialized image analysis process. Our microarray image gridding method includes the following steps: image enhancement, image segmentation and binarization, binary image morphology filtering, binary image projection transformation, one-dimensional projection sequences compound filtering, FFT-based power spectral estimation and projection sequences crest-trough search.

The basic thought of binary image projection transformation is projecting and accumulating pixels of all the sampled spots with intensity over given threshold on x-y axis of image coordinate system. As shown in the projection results (Fig.3), the 2D signal of microarray image will be transformed into one-dimensional space domain signals. These space domain signals will have wave-crest and wave-trough distribution clearly (as Fig.4), which is consistent with the microarray grid structure. After power spectral estimation to the filtered and smoothed projection signals, we can get row and column span information for the spots array with the known gene chip array parameters. According to the spots array's span information, the locations of every projection signal's crest and trough will be determined and then microarray gridding structure is also recognized. This projection transformation method transformed the 2D signal analysis into one-dimensional signal analysis, which reduced the computation complexity and increased microarray image gridding analysis speed.

Realization of Algorithm

This algorithm is described as the following 6 steps:

1. For the image (as Fig. 2) which has M rows and N columns in pixels number, compute one-dimensional vertical and horizontal projection sequences $COL[j]$ and $ROW[i]$.

$$ROW[i] = \sum_{j=0}^{N-1} f(i, j) \quad (i = 0, 1, \dots, M-1)$$

$$COL[j] = \sum_{i=0}^{M-1} f(i, j) \quad (j = 0, 1, \dots, N-1)$$

2. Make a compound filtering with one-dimensional logical, median, zero mean to the projection sequences $COL[j]$ and $ROW[i]$. The formulas of zero mean are:

$$\bar{R} = \frac{1}{M} \sum_{i=0}^{M-1} ROW(i)$$

$$ROW'[i] = ROW(i) - \bar{R} \quad (i = 0, 1, \dots, M-1)$$

$$\bar{C} = \frac{1}{N} \sum_{j=0}^{N-1} COL(j)$$

$$COL'[j] = COL(j) - \bar{C} \quad (j = 0, 1, \dots, N-1)$$

3. Compute FFT for the filtered projection sequences, and then get power spectral estimation sequences P_{col} and P_{row} by periodogram method. The FFT formulas are as follow: $c(n)$ is filtered vertical projection sequence, $r(n)$ is filtered horizontal projection sequence, and L is transformation length.

$$C(k) = FFT[c(n)] = \sum_{n=0}^{L-1} c(n)W_N^{kn}$$

$$R(k) = FFT[r(n)] = \sum_{n=0}^{L-1} r(n)W_N^{kn}$$

The formula used for computing the power spectral sequences are:

$$P_{col}(\omega) = \frac{1}{N} |C(e^{j\omega})|^2$$

$$P_{row}(\omega) = \frac{1}{N} |R(e^{j\omega})|^2$$

4. Search the locations of crest and trough for power spectral sequences $P_{col}(\omega)$ and $P_{row}(\omega)$, which are represented by $MaxVPSPos$ and $MaxHPSPos$.

So the columns span and rows span estimations are got by this way:

$$ColSpan = N / (MaxVPSPos - 1)$$

$$RowSpan = M / (MaxHPSPos - 1)$$

In above formula, $MaxVPSPos$ and $MaxHPSPos$ must meet the following conditions:

$$P_{col}(\omega) \Big|_{\omega=MaxVPSPos} = MAX_{\omega} \{P_{col}(\omega)\}$$

$$P_{row}(\omega) \Big|_{\omega=MaxHPSPos} = MAX_{\omega} \{P_{row}(\omega)\}$$

5. Select one appropriate initial searching point j_s for the vertical projection sequence $COL'[j]$, and find the locations of local maxima in sequence $COL'[j]$ with length of $ColSpan$. These locations of local maxima are just the horizontal coordinates x_n ($n = 1, 2, \dots, NC$) of each spot center, in which, NC is the known column number in this spots array. Then, select one appropriate initial each point i_s for the horizontal projection sequence $ROW'[i]$, and find the locations of local maxima in sequence $ROW'[i]$ with length of $RowSpan$. These locations of local maxima are just the vertical coordinates y_m ($m = 1, 2, \dots, NR$) of each spot center, in which, NR is the known row number in this spots array.

In this step, x_n ($n = 1, 2, \dots, NC$) and y_m ($m = 1, 2, \dots, NR$) must be satisfied with the following conditions:

$$COL'[j] \Big|_{j=x_n} = MAX_j \{COL'[j]\} \quad (n = 1, 2, \dots, NC)$$

$$ROW'[i] \Big|_{i=y_m} = MAX_i \{ROW'[i]\} \quad (m = 1, 2, \dots, NR)$$

6. Put all the spots' 2D image coordinates $\{(x_n, y_m), n = 1, 2, \dots, NC, m = 1, 2, \dots, NR\}$ into data files or one database in the form of tables.

Experiments and result analysis

We have accomplished some gridding experiments by the method mentioned above for the given low-density microarray image (as Fig.1) with 6×6 array structure, which has been corrected tilting before gridding computation. Firstly, we adopted the adaptive thresholding method to segment the image, and got the binary image; secondly, we had a morphology close operation to the binary image with the disk-structured element, whose radius is 1 pixel. After the first step of the algorithm, we got the vertical and horizontal one-dimensional projection sequence: $COL[j]$ and $ROW[i]$. The waveforms of these sequences are showed as Fig. 3.

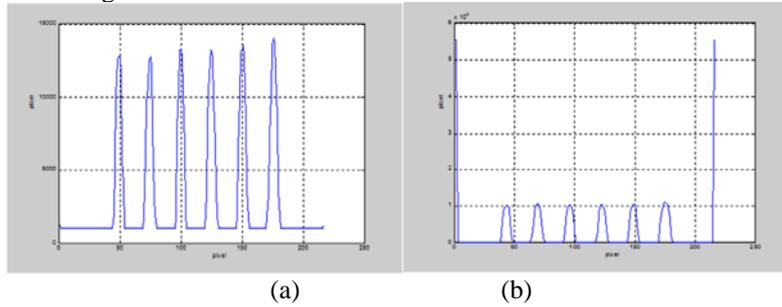


Fig. 3 the vertical (a) and horizontal (b) projection sequence waveform for the 6×6 low density micarray image

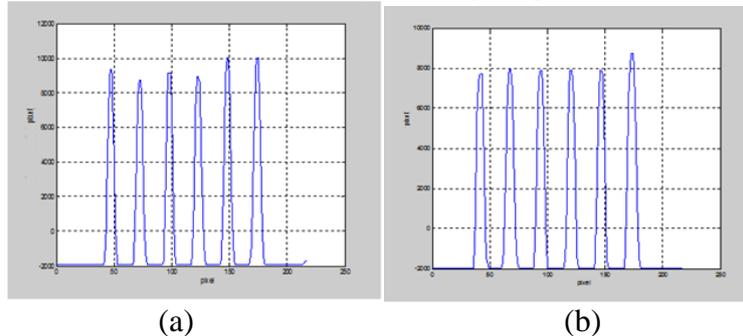


Fig. 4 the one dimensional synthesized filtering result waveform for the vertical

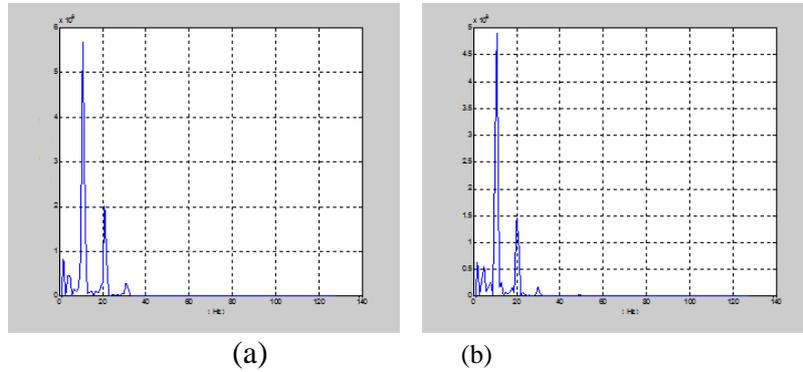


Fig. 5 the power spectral density waveform for the vertical (a) and horizontal (b) projection sequence

In the second step of the algorithm, we adopted the compound filtering methods of one-dimensional logic and median, and finally zero mean to the projection sequences. The filtering results are shown in the Fig. 4.

By computing FFT to the filtered projection sequences in the third step of the algorithm, we get the power spectral estimation waveforms. They are shown in the Fig. 5.

Through searching locations with maximum values in the power spectral waveform for the vertical and horizontal projection sequences in the fourth step of the algorithm, we get two maximum values: $MaxVSPPos=10$, $MaxHSPPos=10$. Furthermore, we get the spots array's column-span and row-span estimations, which are both 24.

In the fifth and sixth step of the algorithm, we make each forward jumping point with minimum value as the initial searching point in the one-dimensional compound filtered vertical and horizontal projection sequences, and search for local maxima points in the subsequences with length of 24 of the above filtered projection sequences, finally we get all the spots centers' approximate pixel coordinates in the microarray image. Fig. 6 shows the microarray gridding result with power spectral analysis and local maxima searching method.

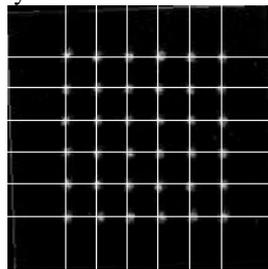


Fig.6. the microarray image gridding result with power spectral analysis and local maxima searching method

Because this algorithm adopted a serialized signal processing method, its computation complexity is not very high. For one microarray image which has m rows and n columns of spots in the spots array, and M rows and

N columns of pixels, this algorithm's computation complexity is $O(mn + MN)$, so to this algorithm presented, computation complexity will increase with spotting density and image pixel resolution. For the low density microarray image in our gridding experiments (Fig.6), $m = 6$, $n = 6$, $M = 256$, and $N = 256$, the computation time of this algorithm is about 23.5 milliseconds on AMD 3000+, 2.3GHz, 1G RAM, and Windows XP SP2 platform.

Conclusion

According to the experiment results, the presented microarray image gridding method has met the requirements of gridding accuracy and computing efficiency for microarray image gridding analysis. However, because this algorithm needs to compute power spectral estimations for the 1D projection sequences, the gridding analysis results are susceptible to interference of image noises which caused some deviations of gridding lines (as Fig. 6). To further promote the gridding accuracy, we will improve this algorithm in the future work.

Acknowledgement

This project was sponsored by Scientific Enterprise Public Research Fund of Liaoning Province (Project No. 2012004002). The authors thank PhD candidate Ma Pingquan and Associate professor Yan jiachao for their good advices and beneficial helps in this research work.

References

- [1] Deng N, Duan HL. Automated Microarray Image Gridding Using Image Projection Vectors Coupled With Power Spectrum Model [J]. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) , Volume: 24, Issue: 4 (2010): 581-600.
- [2] Deng N, Duan HL. The Automatic Gridding Algorithm Based on Projection for Microarray Image [A] . In: Proceeding s of the 2004 International Conference on Intelligent Mechatronics and Automation[C]. Chengdu: IEEE Robotics and Automation Society. 2004. 254.
- [3] Deng N, Duan HL. An Automatic and Power Spectral-based Rotate Correcting Algorithm for Microarray Image [A]. In: Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference[C]. Shanghai: IEEE EMBS, 2005: 898-901.
- [4] Hu Yijun, Weng Guirong. Segmentation of cDNA Microarray Spots Using K-means Clustering Algorithm and Mathematical Morphology [C]. Information Engineering, 2009, ICIE'09, WASE International Conference On, 10-11 July 2009: 110 – 113.
- [5] Luis Rueda, Vidya Vidyadharan. A Hill-Climbing Approach for Automatic

Gridding of cDNA Microarray Images[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), January-March 2006 (vol. 3 no. 1): 72-83.

[6] Angulo J, Sorra J. Automatic analysis of DNA microarray images using mathematical morphology [J]. Bioinformatics, 2003, 19(5): 553-562.

[7] Antonio PG, Damiane J, Liang Z, et al. A Dynamical Model with Adaptive Pixel Moving for Microarray Images Segmentation [J]. Real-Time Imaging, 2004, 10: 189-195.

[8] Nobert B, Horst B, Hilmar L. Robust DNA Microarray Image Analysis [J]. Machine Vision and Applications, 2003. 15: 11-28.

[9] Li QH, Fraley C, Bumgorner RE, et al. Donuts, Scratches and Blanks: Robust Model-based Segmentation of Microarray Images [J]. Bioinformatics, 2005, 21(12): 2875-2882.

[10] Arena P, For tuna L, Occhipinti L. A CNN Algorithm for Real Time Analysis of DNA Microarrays [J]. IEEE Trans on Circuits and Systems I: Fundamental Theory and Applications, 2002, 49(3): 335-340.

[11] Liang Y, George EO, Kelemen A. Bayesian Neural Network for Microarray Data[A]. In : Proceedings of the IEEE International Joint Conference on Neural Networks[c]. Honolulu: IEEE Neural Networks Society, 2002, 1: 193-197.

[12] Lukac R, Plataniotis KN, Smolka B, et al. cDNA Microarray Image Processing using Fuzzy Vector Filtering Framework [J]. Fuzzy Sets and Systems, 2005, 152: 17-35.