Journal of Theoretical and Applied Information Technology
10th July 2014. Vol. 65 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AUTOMATIC GRIDDING METHOD FOR MICROARRAY IMAGES

[1]JONNADULA HARIKIRAN, [2]B.AVINASH, [3]DR.P.V.LAKSHMI, [4]DR.R.KIRANKUMAR
[1]Member IEEE, Assistant professor, Department of IT, GITAM University, Visakhapatnam
[2]Project Fellow, UGC Major Research Project, Department of IT, GIT, GITAM University
[3]Professor and Head, Department of IT, GITAM University, Visakhapatnam
[4]Department of Computer science, Krishna University, Machilipatnam
E-mail: [1] jhari.kiran@gmail.com , [3]pvlakshmi@gitam.edu , [4]kirankreddi@gmail.com

ABSTRACT

A cDNA microarray is a powerful tool in biotechnology providing useful information in analyzing thousands of gene expressions simultaneously. The analysis of microarray images allows the identification of gene expressions to draw biological conclusions for applications ranging from genetic profiling to diagnosis of cancer. The DNA microarray image analysis includes three tasks: gridding, segmentation and intensity extraction. The gridding process is usually divided into two main steps: sub-gridding and spot detection. Most existing gridding algorithms are semiautomatic requiring human intervention and parameter presetting. Each spot is associated with a gene and contains the pixels that indicate the level of expression of that particular gene In addition, improper gridding will affect image analysis in extracting intensity values from each spot that represents gene expression level. In this paper, a fully automatic approach to sub-gridding and spot detection is presented. After gridding, the image is segmented using fuzzy c-means clustering algorithm for separation of spots from the background pixels. The result of the experiment shows that the method presented in this paper is accurate and automatic without human intervention and parameter presetting.

Keywords: *Microarray, Mathematical Morphology, Gridding, Image Processing*

1. INTRODUCTION

Microarrays widely recognized as the next revolution in molecular biology that enable scientists to monitor the expression levels of thousands of genes in parallel [1]. A microarray is a collection of blocks, each of which contains a number of rows and columns of spots. Each of the spot contains multiple copies of single DNA sequence [2]. The intensity of each spot indicates the expression level of the particular gene [3]. The processing of the microarray images [5] usually consists of the following three steps: (i) gridding, which is the process of segmenting the microarray image into compartments, each compartment having only one spot and background (ii) Segmentation, which is the process of segmenting each compartment into one spot and its background area (iii) Intensity extraction, which calculates red and green foreground intensity pairs and background intensities.

Many approaches have been proposed for gridding of microarray images using intensity projection profiles. Hirata [6] presented an automatic sub-array and spot gridding method using the horizontal and vertical profile signal of the image. User assistance was required in this method to fix image rotation and check if the segmentation is correct. This software provides an optional step to enforce a regularization of the griddings and adjust the intervals between horizontal and vertical lines. This is valid if the sub-array sizes are equal. Jain [7] proposed a gridding algorithm based on axis projection of image intensity along the rows and columns of the microarray image. The algorithm requires large number of spots and is not robust to misalignment of different grids. Y.Wang [8] demonstrated a precise and automatic microarray gridding method which will eliminate the need of human intervention and correct the potential alignment and rotation problems. The method is sensitive to contaminations and large number of missing spots. If big chunk of contaminations or noises appear to be the similar size of subarray, misclassification of rotated/nonrotated subarrays will happen. Shuqing Zhao [9] proposed microarray image processing using mathematical morphology. An improved gridding method based on mathematical morphology is proposed, which is characterized by filtering out the block noise and filtering projection plots. Several parameters about the sub-array and spots are required during the gridding and spotting procedure which can be preset in advance or acquired from database. Deepa.J [10] proposed automatic gridding of DNA microarray images using optimum subimage. The approach is based on the selection of optimum subimage and the parameters for gridding are calculated using the intensity projection profile of the sub-image. This process does not consider the intensity projection profile of whole image, which is highly depending on the intensity characteristics of the image.

In this paper, a fully automatic algorithm for sub-gridding and spot detection is presented. This automatic gridding algorithm utilizes the image processing techniques for calculating the parameters like spot centroids and spacing between the spots. Accuracy of gridding method depends on the accuracy in the calculation of the parameters like spot diameter and inter spot distance. After gridding, fuzzy C-means clustering algorithm [4] is used for segmentation of microarray image into spots and image background. The algorithm is automatic and accurate for misalignment of spots in microarray image. Furthermore, when we apply this algorithm on different microarray images, human intervention and parameter presetting is unnecessary. The paper is organized as follows: section II presents sub-gridding algorithm, Section III presents spot detection algorithm, Section IV presents the qualitative and quantitative results, and finally Section V repots conclusions.

2. AUTOMATIC SUB-GRIDDING ALGORITHM

Microarray images are grouped in sub-grids, with each sub-grid containing hundred of spots. To extract each spot from the image, first the image has to segment into sub-grids. The output of the sub-gridding algorithm is a two dimensional matrix G^1 of same size of original image of size MxN. Initially the values of $G^1[i,j]=0$, for i=1,…M and j=1,…N. The steps of the automatic sub-gridding algorithm are as follows:

Step 1: Convert the original microarray image into grayscale image

Step 2: Compute the standard deviation of pixel intensities of every row, producing the profile of horizontal standard deviation

$$S_i = \frac{1}{M-1}\left(I(i,j) - \frac{1}{N}\sum_{j=1}^{N} I(i,j)\right) \quad (1)$$

for i=1,2,…..M and I(i,j) is the intensity of pixel located at row i and column j.

Step 3: Compute the standard deviation of pixel intensities of every column, producing the profile of vertical standard deviation

$$S_j = \frac{1}{N-1}\left(I(i,j) - \frac{1}{M}\sum_{i=1}^{M} I(i,j)\right) \quad (2)$$

for j=1,2…….N and I(i,j) is the intensity of pixel located at row i and column j.

Step 4: Find the minimum value in S_i and S_j.

MIN1= minimum (S_i)
MIN2= minimum(S_j)
(3)

Step 5: These standard deviation profiles have to be thresholded for the estimation of row width and column width.

TS_i = 1, if S_i > MIN1+ Th * MIN1
= 0 otherwise
TS_j = 1, if S_j > MIN2+ Th * MIN2
= 0 otherwise [Threshold (Th) = 0.1]
(4)

Step 6: Calculation or row width and column width: The value in TS_i and TS_j looks like 1's followed by 0's again 1's and so on.

The procedure for row width calculation is as follows:

```
i.  for i = 1 to M,
      Count the number of 0's in $TS_i$ (between the elements in $TS_i$ having 1) = $RW_{yp}$, Where p = 1, 2, 3 …, k.
      Count the number of 1s in $TS_i$ (between the elements in $TS_i$ having 0) = $RW_{xp}$, Where p = 1, 2, 3 …, k.
ii. for p=1,2,…,k
      $RW_{zp}$= $RW_{xp}$ + ($RW_{yp}$ + $RW_{y(p+1)}$)/2,   (5)
iii. Row width ($RW^1$) = median ($RW_z$).
iv. Using the value $RW^1$ draw horizontal grid lines at positions of i, where i=1, 2 ….. M with step increment $RW^1$.
      for i=1to M step $RW^1$
        for j=1 to N
          $G^1[i,j]$ =1;
```

The procedure for column width calculation is as follows:

```
i.  for j = 1 to N,
      Count the number of 0's in $TS_j$ (between the elements in $TS_j$ having 1) = $CW_{yp}$, Where p = 1, 2, 3 …, k.
```

Count the number of 1's in TS_j (between the elements in TS_j having 0) = CW_{xp}, Where p = 1, 2, 3 …, k.

ii. for p=1,2,…,k

$CW_{zp}= CW_{xp} + (CW_{yp} + CW_{y(p+1)})/2$, (6)

iii. Column width (CW^1) = median (W_z).

iv. Using the value CW^1 draw horizontal grid lines at positions of j, where j=1, 2 ….. N with step increment CW^1.

```
for j=1to N step CW^1
    for i=1 to M
        G^1[i,j] =1;
```

Step 7: Map these grid matrix G^1 onto the microarray image producing sub-gridding.

3. AUTOMATIC SPOT DETECTION ALGORITHM

A microarray image contains a number of sub-grids and each sub-grid representing a two dimensional array of spots. Gridding refers to accurately locating each spot within a microarray image namely sub-gridding and spot detection. The sub-grid detection is done by the method presented in Section I. After sub-grid detection, the aim is to separate the sub-grid into spot regions by means of a grid. The output of gridding is to obtain a 2D matrix G of same size of sub-grid. Initially the values of G[i,j]= 0, for i=1,….,M and j=1,….,N.

The steps of the automatic spot detection algorithm are as follows:

Step 1: Consider the sub-gridded microarray image.

Step 2: Perform Edge detection using Bi-dimensional Empirical Mode Decomposition proposed in [11]

Step 3: Perform morphological filling on the edge image obtained from step 2.

Step 4: Perform mean intensity of the pixels on each sub-grid from the morphologically filled sub-gridded image. Consider the sub-grid with maximum mean intensity. That sub-grid is called optimal sub-grid. Perform spot detection on this optimal sub-grid.

Step 5: Calculation of Horizontal and Vertical Intensity profiles

Horizontal and vertical intensity projection profiles of binary image (Morphological Filled Image) are the sum of pixel intensities along each row and column respectively. Let M_b indicates the filled image of size MxN. Then the intensity projection profile along ith row and jth column are computed using (3) and (4). S_i=

$$M_{pi} = \sum_{j=1}^{N} M_b(i,j) \qquad \text{for} \qquad i=1,2.....M$$

(7)

$$S_j = M_{pj} = \sum_{i=1}^{M} M_b(i,j) \quad \text{for } j=1,2........N$$

(8)

Step 6: Calculation of row width (RW) and column width (CW)

The values in S_i are used for identification of row width and values in S_j are used for identification of column width. The values in S_i and S_j looks like zeros followed by nonzero again zeros and so on.

The procedure for row width calculation is as follows:

i. for i = 1 to M,

Count the number of zeroes in S_i (between the elements in S_i having nonzero) = W_{yp}, Where p = 1, 2, 3 …, k.

Count the number of non-zeroes in S_i (between the elements in S_i having zero) = W_{xp}, Where p = 1, 2, 3 …, k.

ii. for p=1,2,…,k

$W_{zp}= W_{xp} + (W_{yp} + W_{y(p+1)})/2$, (9)

iii. Row width (RW) = median (W_z).

iv. Using the value RW draw horizontal grid lines at positions of i, where i=1, 2 ….. M with step increment RW.

```
for i=1to M step RW
    for j=1 to N
        G[i,j] =1;
```

The procedure for column width calculation is as follows:

i. for j = 1 to N,

Count the number of zeroes in S_j (between the elements in S_i having nonzero) = W_{yp}, Where p = 1, 2, 3 …, k.

Count the number of non-zeroes in S_j (between the elements in S_i having zero) = W_{xp}, Where p = 1, 2, 3,……,k.

ii. For p=1,2,…,k

$W_{zp}= W_{xp} + (W_{yp} + W_{y(p+1)})/2$, (10)

iii. Column width (CW) = median (W_z).

iv. Using the value CW draw horizontal grid lines at positions of j, where j=1, 2 ….. N with step increment CW.

```
for j=1to N step RW
    for i=1 to M
        G[i,j] =1;
```

Step 7: Map these grid matrix G onto the optimum sub-grid image.
Step 8: Compute the total number of connected components in the gridded optimum image using 4-adjacency or 8-adjacency. The number of connected components is equal to the number of spots. Assign a label for pixels in each connected component generating a label matrix.
Step 9: The pixels with one particular label represents a region of the single spot. Using region properties, compute the centroid and area for each region (spot). From the centroids of the spots, we can find the distance between the spots.
Step 10: If the number of connected components is not equal to the number of spots, adjust the grid lines in horizontal and vertical direction for the spots having same label.

4. QUALITATIVE AND QUANTITATIVE RESULTS

The proposed automatic gridding algorithm is performed on a two different microarray slides drawn from the Stanford microarray Database corresponds to breast category aCGH tumor tissue. First, automatic sub-gridding algorithm is applied on two microarray images to obtain sub-grids. The output of sub-gridding on two microarray images is shown in figure 1 and figure 2. Then automatic spot detection algorithm is applied on optimum sub-grid for spot detection. The first sub-grid slide taken from figure1 is a 261*289 pixel image that consists of a total of 75429 pixels. The second sub-grid slide taken from figure 2 is a 559*489 pixel image that consists of total 273351 pixels. The output of the proposed automatic spot detection algorithm on two microarray sub-grids is shown in figure3 and figure 4.The accuracy of the gridding algorithm was calculated as

$$\text{Percentage accuracy} = \frac{Number\ of\ spots\ perfectly\ gridded}{Total\ number\ of\ spots} *100 \quad (11)$$

The performance of the proposed gridding algorithm was evaluated by comparing the results with the existing methods are shown in table 1.

Table 1: Percentage Accuracy Of Gridding

Method	Percentage Accuracy Image 1	Percentage Accuracy Image 2
Hirata [6]	87	79
Jain [7]	89	81
Wang [8]	91	82
Shuqing Zhao[9]	90	84
Deepa .J [10]	92	88
Proposed	94	91

After gridding, the segmentation of spots from the background pixels is done by using fuzzy c-means clustering algorithm. The method is implemented in such a way that the intensity value of each pixel and the pixels of the image has been grouped in two clusters. The number of pixels clustered as spot and background for two different microarray images has been presented in Table 2.

Table 2: The Number Of Pixels Clustered As Spots And Background

Fuzzy C-means	Total Number of Pixels	Spots	Background
Image 3	75429	40535	34714
Image 4	273351	111669	161682

5. CONCLUSION

In this paper, a fully automatic gridding method for separating spot centers in microarray images has been proposed. The proposed method automatically locates the individual spots without any input parameters and human intervention. It can be proved that percentage accuracy of gridding is high with the methods that use projection profile of the entire image. After gridding, the image is segmented using fuzzy c-means clustering algorithm. The proposed method is accurate and automatic, which takes a microarray as input image and makes no assumptions about the size of the spots, rows and columns in the grid.

REFERENCES:

[1] M.Schena, D.Shalon, Ronald W.davis and Patrick O.Brown, "Quantitative Monitoring of gene expression patterns with a complementary DNA microarray", Science, 270,199,pp:467-470.

[2] Wei-Bang Chen, Chengcui Zhang and Wen-Lin Liu, "An Automated Gridding and Segmentation method for cDNA Microarray Image Analysis", 19th IEEE Symposium on Computer-Based Medical Systems.

[3] Tsung-Han Tsai Chein-Po Yang, Wei-ChiTsai, Pin-Hua Chen, "Error Reduction on Automatic Segmentation in Microarray Image", IEEE 2007.

[4] Volkan Uslan, Ihsan Omur Bucak, "Clustering based Spot Segmentation of cDNA Microarray Images ", 32nd Annual International Conference of IEEE EMBS , IEEE 2010.

[5] J.Harikiran, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar," Edge detection using mathematical morphology for gridding of Microarray image", International Journal of Advanced Research in Computer Science, vol 3, No 2, pp- 172-175. March 2012.

[6] Roberto Hirata Jr, Junior Barrera, R.F. Hashimoto, O.D.Daniel," Microarray Gridding by mathematical Morphology", 2001 IEEE, PP. 112-119.

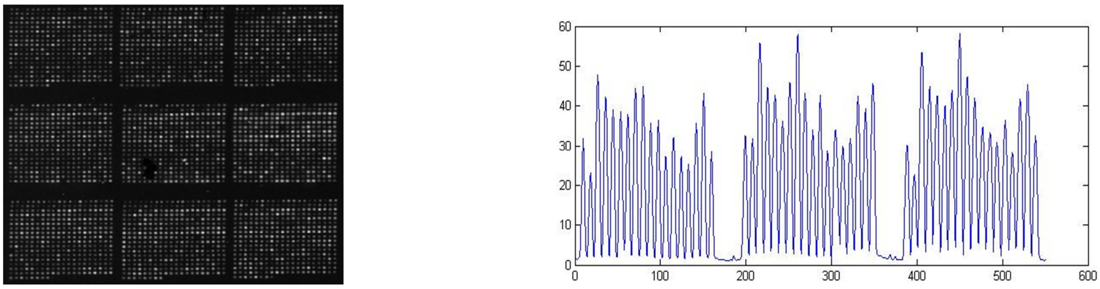
[7] A.N.Jain, Tokuyasu, Snijders," Fully Automatic Quantification of microarray image data", Genome research, 12(2), 32-332.

[8] Y.Wang, F.Y.Shih, M.Ma, "Precise Gridding of microarray images by detecting and correcting rotations in subarrays", in proceedings of sixth International conference on computer vision, Patternrecognition and Image Processing, 2010.

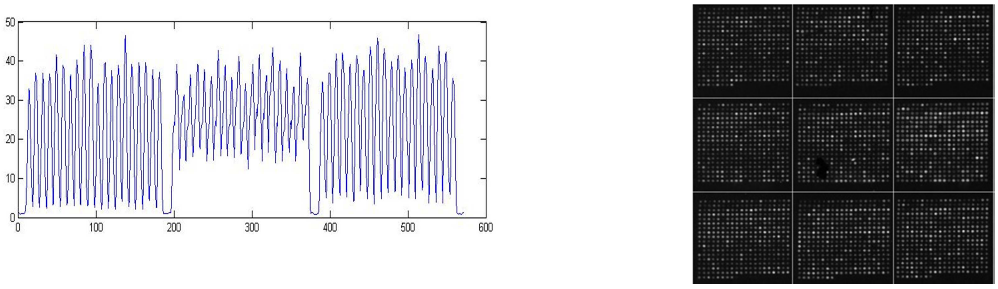
[9] Shuqing Zhao, Qiaoning Yang, "Microarray Image Processing Based on Mathematical Morphology", Proceedings of ICSP 2006, IEEE.

[10] Deepa J, Tessamma Thomas," A New Gridding Technique for High Density Microarray Images Using Intensity Projection Profile of Best Sub Image", Computer Engineering and Intelligent systems, Vol 4 No 1 : pp 7-18.

[11] J.Harikiran, et.al, "Spot Edge Detection in Microarray Images using Bi-dimensional Empirical Mode Decomposition, Proceedings of C3IT 2012, Procedia Technology, Vol 4, pp:227-231

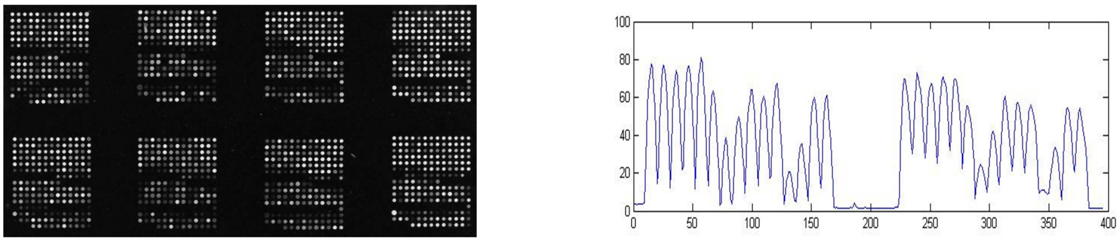


(a) original microarray image (in grayscale) *(b) horizontal standard deviation profile*

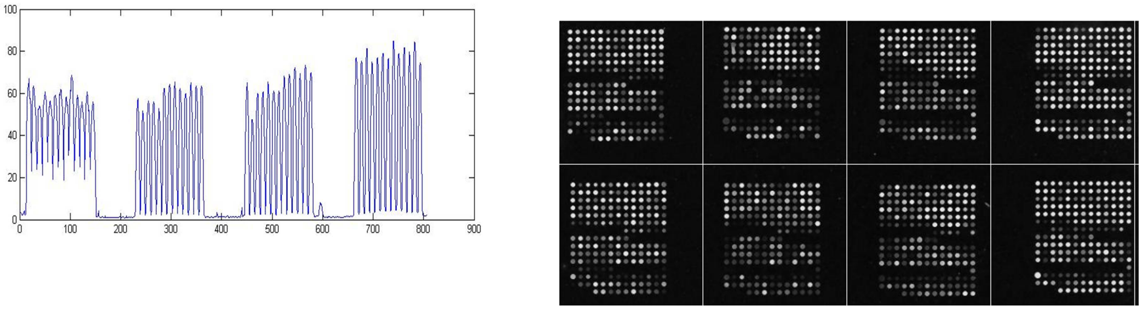


(c) vertical standard deviation profile *(d) Gridded image*

Figure 1

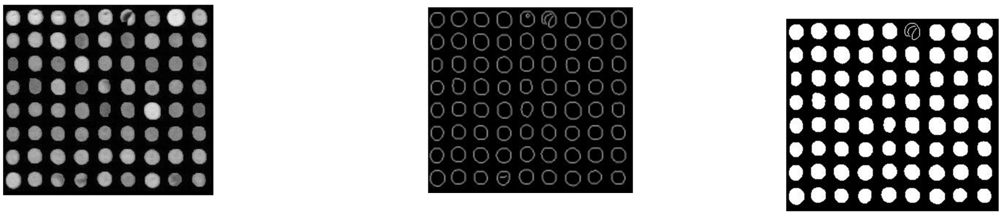


(a) original microarray image (in grayscale) *(b) horizontal standard deviation profile*

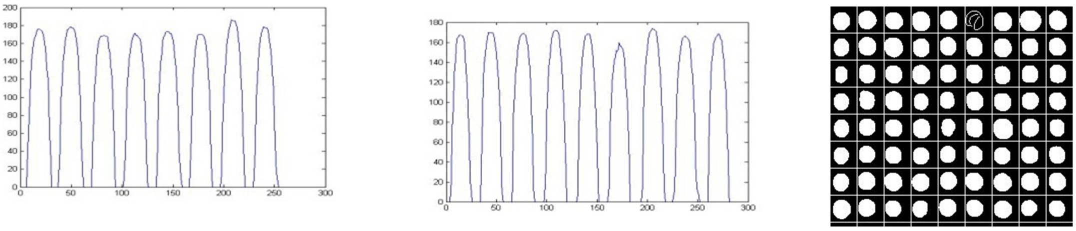


(c) vertical standard deviation profile *(d) Gridded image*

Figure 2



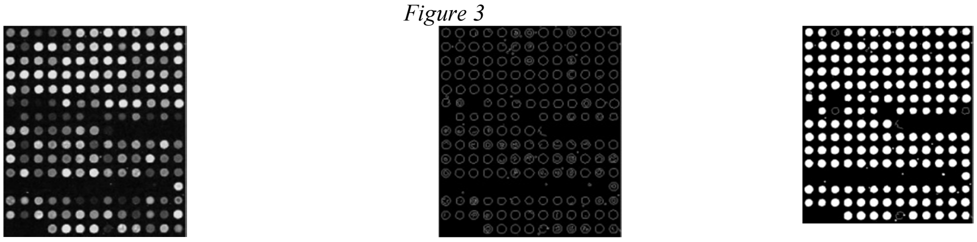
(a) Grey scale Image (b) Edge Detection (c) Morphological Filling



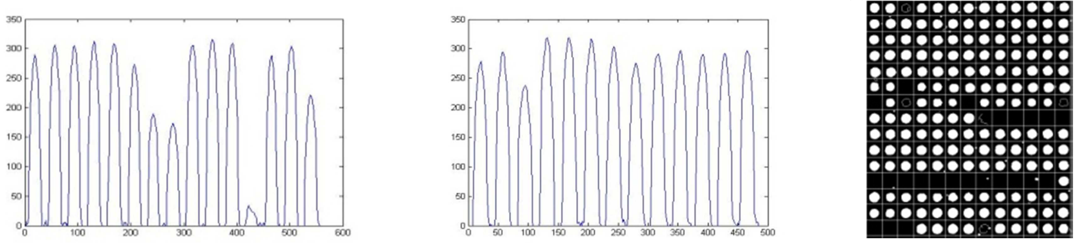
(d) Plot of M_{pi} (for all rows) (S_i) (e) Plot of M_{pj} (for all columns)(S_j) (f) gridded image

(g) Spot centroids (h)Segmented using Fuzzy c-means

Figure 3



(a) Grey scale Image (b) Edge Detection (c) Morphological Filling



(d) Plot of M_{pi} (for all rows) (S_i) (e) Plot of M_{pj} (for all columns)(S_j) (f) gridded image

(g) spot centroids Figure 4 (h) segmented image using fuzzy c-means