

A New Gridding Technique for High Density Microarray Images Using Intensity Projection Profile of Best Sub Image

Deepa J^{1*} Tessamma Thomas²

1. Department of Electronics, College of Engineering, Chengannur, Kerala, India.
2. Department of Electronics, Cochin University of Science and Technology, Kochi, India.

* E-mail of the corresponding author: deepaj@ceconline.edu , deepaj@cusat.ac.in

Abstract

As the technologies for the fabrication of high quality microarray advances rapidly, quantification of microarray data becomes a major task. Gridding is the first step in the analysis of microarray images for locating the subarrays and individual spots within each subarray. For accurate gridding of high-density microarray images, in the presence of contamination and background noise, precise calculation of parameters is essential. This paper presents an accurate fully automatic gridding method for locating subarrays and individual spots using the intensity projection profile of the most suitable subimage. The method is capable of processing the image without any user intervention and does not demand any input parameters as many other commercial and academic packages. According to results obtained, the accuracy of our algorithm is between 95-100% for microarray images with coefficient of variation less than two. Experimental results show that the method is capable of gridding microarray images with irregular spots, varying surface intensity distribution and with more than 50% contamination.

Keywords: microarray, gridding, image processing, gridding accuracy

1. Introduction

Microarrays have become a standard tool for research in molecular biology and clinical diagnostics for parallel analysis of genomic information. In microarray experiments, thousands of single stranded genomic sequences are printed (spotted) on a substrate. They are allowed to hybridize with the fluorescently labeled complementary part. The surface of the hybridized array is then scanned to produce the microarray image. Microarray image consists of spots having varying intensity representing the level of hybridization. Image analysis is the first stage in a chain of analytical procedures for extraction of information from the microarray. Spots within the array are aligned horizontally and vertically in blocks called subarrays. Ideally, all subarrays are of the same size with spacing between subarray being regular, shape of the spots being same for all the spots within the subarray. Also, there is no dust or contamination on the slide and background intensity across the image is uniform. However, in the real microarray experiments, the exact position and size of the spots may vary due to the several reasons, mainly related to mechanical constraints and hybridization inconsistencies. In addition, spot intensity levels are highly variable and weak spots are often difficult to be detected (Dov Stekel, 2003). Hence, the first step of microarray image processing is the identification of the position of subarrays and then locate each spot within the subarray. This is known as gridding. It is important that addressing procedure be accurate, to ensure precision in the subsequent steps of image analysis (David. Bowtell, Joseph Sambrook, 2003).

2. Existing Gridding Techniques

In the last few years, different Gridding algorithms have been developed; most of them require input parameters and at times manual intervention for locating exact spot center. This impose big burden for the biologists who use microarrays in their research. Automatic gridding algorithm utilizes the image processing techniques for automatically calculating the parameters like spot diameter, spacing between spots and between subarrays automatically. However, even for these algorithms there are always limitations due to unpredictable deviations from the assumed array design, high contamination level, and

large number of missing spots (Novikov. E *et al.*, 2006). Jain *et al.* (2002) proposed a gridding algorithm based on axis projection of image intensity. The algorithm requires a large number of bright spots and is not robust to misalignment of different grids. Y. Wang *et al.* (2005) demonstrated a fully automatic gridding methodology using intensity projection profile of microarray image. The method is sensitive to contaminations and large number of missing spots. Novikov E *et al.* (2006) has implemented a noise resistant grid finding algorithm which also uses intensity projection profile by transforming the fluctuations of the intensity of each row or column to special parameter and systematically penalizing the irregular regions. Rueda L& Vidyadharan V (2006) have developed a method for finding spot locations uses a hill-climbing approach to maximize the energy, seen as the intensities of the spots, which are fit to different probabilistic models .A Radon-transform-based method that separates the sub-grids in a cDNA microarray image has been developed by Rueda *et al.*(2007). The method applies Radon transform to find possible rotations of the image and then finds the subgrids by smoothing the row or column sums of pixel intensities; however, that method does not automatically find the correct number of sub-grids, and the process is subject to data-dependent parameters. Gridding method by Wang *et al.* (2007) performs a series of steps including rotation detection and compares the row or column sums of the top-most and bottom-most parts of the image. The method for gridding cDNA microarray images suggested by (Zacharia E & Maroulis D, 2008) uses an evolutionary algorithm to separate subarrays and detect the positions of the spots. Using maximum margin is a method for automatic gridding of cDNA microarray images based on maximizing the margin between rows and columns of spots developed by (Bariamis D *et al.*, 2010). The following section describes a fully automatic grid alignment technique, which is a modified version of the method proposed by (Deepa J & Tessamma T, 2009). In the previous work, an optimum subimage is identified by searching the block for maximum mean intensity. The method performs well for images with bright spots. However, the accuracy was found decreasing as the contamination and noise increased. The proposed method is more robust against different contamination as well as large background intensity fluctuations.

3. New Gridding method

The different steps involved in this gridding process are (1) Preprocessing (2) Global parameter estimation for locating subarrays (3) Local parameter estimation for locating individual spots within a subarray.

3.1 Pre-Processing

The pre- processing steps used, to construct a binary reference image from the input image so as to keep the input image intact for further analysis are as follows:

Step1: Convert RGB image to gray level image

Step2: Perform contrast enhancement using contrast-limited adaptive histogram equalization and then intensity rescaling.

Step3: Edge detection using Canny method

Step4: Morphological filling

Step5: Morphological sequential filtering

The adaptive histogram equalization of step 2 operates on small data regions for contrast enhancement and provides better results than global histogram equalization method. Canny edge detection method. (J.F. Canny, 1986) has been applied to detect the boundaries of the spot. Using morphological region filling, the region inside the spot boundary is filled. Sometimes the adjacent spots are joined together at their edges incorrectly and fill in the common corners. Morphological sequential filters are used to remove such islands as well as the noise that has been emphasized by step3. The binary image generated after the preprocessing is suitable for estimation of parameters necessary for gridding. Fig.1 (a) shows Coriell cell line BAC array CGH data analyzed by Snijders *et al.* (2001). The Coriell cell line data is widely regarded as a "gold standard" data set. The preprocessing step is applied in the whole array, consists of approximately 7500 spots (2460 BAC and P1 clones in triplicate in a 12 mm square HumArray 1.14). Here each subarray consists of 462 spots. Fig.1. (b) is the reference binary image generated.

3.2 Global gridding

Global gridding refers to the process of locating each sub arrays within a microarray image. The global parameters required for accurately locating subarrays are width and height of each subarray as well as spacing between them. These parameters have been estimated using the intensity projection profiles of the binary reference image generated in the preprocessing step. Horizontal and vertical intensity projection profiles of binary reference image are the sum of pixel intensities along each row and column respectively.

Let I_b indicates the binary reference image of size $M \times N$. Then, the intensity projection profile along r^{th} row (I_{pr}) and c^{th} column (I_{pc}) are computed as shown below:

$$I_{pr} = \sum_{j=1}^N I_b(r, j) \quad (1)$$

$$I_{pc} = \sum_{i=1}^M I_b(i, c) \quad (2)$$

Fig.2 (a) and (b) shows intensity projection profile of the reference image in Fig.1 (b). These intensity projection profiles have to be thresholded for the estimation of the global parameters. Let T_r and T_c be the threshold values for row and column profiles respectively. Then thresholded values of I_{pr} and I_{pc} are denoted by $I_{pr}(T)$ and $I_{pc}(T)$ respectively.

$$I_{pr}(T) = 1, \text{ if } I_{pr} \geq T_r \quad (3)$$

$$= 0 \text{ otherwise}$$

$$I_{pc}(T) = 1 \quad \text{if } I_{pc} \geq T_c \quad (4)$$

$$= 0 \text{ otherwise}$$

In the present work, the threshold values (T_r , T_c) are selected as of 10% of the maximum value of the intensity projection profile along the row and column. Fig.2(c, d) shows the thresholded projection profile.

Here ' W_r ' and ' W_c ' denotes the row and column width of the subarrays respectively. ' S_r ' and ' S_c ' are the row and column spacing between the subarrays. Regions with contaminations or other artifacts show large variation in these parameters values from their mean value. Accuracy of the gridding parameters can be increased by eliminating these irregular parameters. The procedure is as follows:

Let, W_R is the set of all w_r and W_C is the set of all w_c

$$W_R = \{w_{r1}, w_{r2}, \dots, w_{rk1}\} \quad (5)$$

$$W_C = \{w_{c1}, w_{c2}, \dots, w_{ck2}\} \quad (6)$$

The median values of W_R and W_C are evaluated as \tilde{w}_r and \tilde{w}_c respectively. Any row width w_{ri} (for $i=1, 2, \dots, k_1$) and column width w_{cj} (for $j=1, 2, \dots, k_2$) is considered for the determination of final parameters of the subarray if the following condition is satisfied.

$$0.5\tilde{W}_r \leq W_{ri} \leq 1.5\tilde{W}_r \quad (7)$$

$$0.5\tilde{W}_c \leq W_{cj} \leq 1.5\tilde{W}_c \quad (8)$$

Thus the irregular elements in W_R and W_C are rejected. Using the selected W_{ri} and W_{cj} new median values \tilde{W}_{rn} and \tilde{W}_{cn} are estimated. The same procedure is applied for the spacing parameters S_{ri} and S_{cj} also to reject irregular spacing variables. New median spacing values are denoted by \tilde{S}_{rn} and \tilde{S}_{cn} . The global gridding parameters, rowwidth (G_R) and column width (G_C) are estimated as

$$G_R = \tilde{W}_{rn} + \tilde{S}_{rn} \quad (9)$$

$$G_C = \tilde{W}_{cn} + \tilde{S}_{cn} \quad (10)$$

Here row width indicates the width and column width indicates the length of the grid. Thus the subarray grid size can be defined using rectangular window of size $G_R \times G_C$. The resultant image after applying the gridding algorithm is shown in Fig.3.

3.3 Local gridding

A typical microarray slide consists of rectangular subarrays of spots. There are variations among the individual subarray due to non-uniformity in the hybridization, artifacts on the surface of the array and gaps or dark areas where little or no hybridization has occurred. Once each subarray has been located correctly, the next step is locating each spot within the subarray. This process is called local gridding. The preprocessing steps are applied to each subarray for generating binary reference subarray. Fig.4 illustrates different preprocessing steps applied to a subarray consisting of 196 spots. Parameters for locating individual spots are estimated from the most suitable subimage within the reference image.

3.3.1 Identification of the best subimage

Block processing method is used to identify the best subimage in the reference image. The different steps involved in this process are explained below:

Step 1: Identify the optimal block size for block processing.

Let, the size of the subarray be $[m \ n]$. Define the maximum row and column dimension of the block (subimage) using a scalar K. For example, $K = \text{maximum}(m/2, n/2)$. The algorithm for determining block size ($b1 \times b2$) is as follows:

$$b1 = m \text{ if } m \leq K$$

Otherwise, consider all values between minimum ($m/10, K/2$) and K. Choose the value that minimizes the padding required. The same algorithm is then repeated for 'n' also to identify b2.

Step2: Once the block size has been identified as $b1 \times b2$ using the sliding window approach, calculate the mean intensity value of each block, as window slides from the top left to bottom right of the binary reference image.

Step4. Select the block with maximum mean intensity I_{\max}

Step5. Find thresholded intensity projection profile of this image block as per equation (1) to (4)

Step 6: Let, w_{rb} and w_{cb} be the row and column widths of the thresholded projection profiles (shown in Fig.6) and Let W_{RB} is the set of all w_{rb} and W_{CB} is the set of all w_{cb} with their median values \tilde{w}_{rb} and \tilde{w}_{cb} and standard deviation $\tilde{\sigma}_{rb}$ and $\tilde{\sigma}_{cb}$.

$$W_{RB} = \{w_{rb1}, w_{rb2}, \dots, w_{rb1l}\} \quad (11)$$

$$W_{CB} = \{w_{cb1}, w_{cb2}, \dots, w_{cb1l}\} \quad (12)$$

Then the subimage is selected as the best sub image if both $\tilde{\sigma}_{rb}$ and $\tilde{\sigma}_{cb}$ are less than 50% of \tilde{w}_{rb} and \tilde{w}_{cb} and respectively. Otherwise, the selected subimage is rejected .Then search next subimage with lower mean intensity and repeat steps 5 and 6 until the optimum sub image has been identified. The median value of all the row spacing (\tilde{s}_{rb}) and column spacing (\tilde{s}_{cb}) in the selected subimage are estimated from the thresholded projection profile. Fig. (5) shows reference image and the optimum subimage identified for image in Fig.4.Fig (6) is the projection profiles of the subimage before and after thresholding.

3.3.2 Parameter Estimation

Parameters required for exactly locating the spot are spot diameter, row, and column spacing between spots. Using these parameters grid size is evaluated. The spot diameter (D) has estimated as

$$D = \frac{(\tilde{w}_{rb} + \tilde{w}_{cb})}{2} \quad (13)$$

Now row distance (L_R) and column distance (L_C) of each local grid are defined as

$$L_R = D + \tilde{s}_{rb} \quad (14)$$

$$L_C = D + \tilde{s}_{cb} \quad (15)$$

The local grid size for each spot is $L_R \times L_C$. Fig. (7) shows the resultant image after applying the gridding algorithm.

4. Performance Analysis

The validity of the algorithm has been tested and confirmed using various microarray images with different spot sizes, shape, experimental design and levels of contamination. The composite microarray images available from Stanford microarray database (SMD) were used for testing the gridding algorithm. Subarrays from 25 different microarray images were analyzed. Fig.8 (a) is a subarray from microarray Experiment_ ID 11712 with good quality spots having uniform shape and size. Fig.8 (b) is a subarray from Experiment _ID 27746 with nonuniform background. Results shows high gridding accuracy when compared with the method that uses the intensity projection profile of the whole subarray. Fig.8 (c) shows of a portion of 1024×1024 pixel microarray image with 7392 spots (demo images provided by microarray image analysis software package MAIA).It has been shown that the algorithm was able to grid correctly a subarray with more than 50%contamination. The accuracy of the gridding algorithm was calculated as:

$$\text{Percentage Accuracy} = \frac{\text{Number of spots perfectly gridded}}{\text{Total number of spots}} \times 100 \quad (16)$$

The performance of the spot gridding algorithm was evaluated by comparing the results with the method

demonstrated by Y.Wang *et al* (2003), which was also implemented for comparison. Graph in Fig 9. illustrates the gridding accuracy of the two methods for different microarray images. Here the coefficient of variation CV (ratio of standard deviation to mean intensity) has been plotted against the gridding accuracy. Result indicates the superior performance of the present work especially for subarrays with large amount of variation from the mean intensity (large values of CV).

To study the influence of various noises that commonly occur during microarray image acquisition, artificial images were generated with known parameters. Different noises that commonly occur in microarrays were added to these images and the gridding accuracy was evaluated. Fig.8 (d) shows the gridded noisy image when Gaussian noise with zero mean and variance 0.04 added with the microarray image. The morphological filtering applied in the preprocessing step is capable of removing most of the noise present in the image. Since we are estimating parameters using the binary reference image after the preprocessing step, the effect of this noise can be greatly reduced. Table.1 shows accuracy of the algorithm for different noisy images. Comparison with the existing software MAGIC has shown similar performance for images with good quality spots. While locating the spots in arrays with low mean intensity as well as large contaminations the new method shows superior performance. Even though searching the best subimage is a time consuming task, especially if the contamination is high, using the block processing capability of the MATLAB software the computation efficiency has been improved. Table.2 shows the execution time required for gridding subarrays with different number of spots using Pentium (R), Dual Core Processor 3GHz with 2GB of RAM system. The method takes less than 3 seconds for gridding a microarray image with 1024×1024 pixels make it suitable for gridding high-density arrays.

5. Conclusion

In this paper, a new method of automatic gridding of microarray images based on intensity projection profile of best subimage has been introduced. The method involves various task like preprocessing, identification of a subimage and parameter estimation. The most suitable subimage with maximum mean intensity and regular profile has been used to determine the parameters. It has been proved that accuracy is very high when compared with the methods that use projection profile of the entire image. It can automatically locate both subarrays and individual spots without any input parameters and human intervention. The method is robust with respect to different type of contamination and can tolerate a high percentage of missing spots make it a suitable for gridding high density microarray images.

References

- Angulo, Jesus, and Jean Serra. (2003), "Automatic analysis of DNA microarray images using mathematical morphology." *Bioinformatics* ,vol.19, No 5,pp.553-562
- Bariamis, Dimitris, Dimitris K. Iakovidis, and Dimitris Maroulis. (2010), "M3G: Maximum margin microarray gridding." *BMC Bioinformatics*, 11; 49.
- David Bowtell, Joseph Sambrook, (2003), *DNA microarrays: A molecular cloning manual*. Cold Spring Harbor Laboratory Press.
- Deepa J, Tessamma T,(2009), "Automatic Gridding of DNA Microarray Images using Optimum Subimage," in *International Journal in Recent Trends in Electrical & Electronics*, vol.1,no.4,pp.36-40.
- Dov. Stekel, (2003), *Microarray bioinformatics*:. Cambridge: Cambridge University Press.
- J.F. Canny, (1986), "A computational approach to edge detection," *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698.
- Jain, A. N., Tokuyasu, T. A., Snijders, A. M., Segraves, R., Albertson, D. G., & Pinkel, D.(2002) "Fully automatic quantification of microarray image data.", *Genome research*, 12(2), 325-332..
- L. J. Heyer, D. Z. Moskowitz, J. A. Abele, P. Karnik, D. Choi, A. M. Campbell, E. E. Oldham, and B. K. Akin, "MAGIC Tool: integrated microarray data analysis" (2005), available:

<http://www.bio.davidson.edu/projects/magic/magic.html>

Novikov E, Barillot E, (2006) "A noise-resistant algorithm for grid finding in microarray image analysis". *Machine Vision and Applications*; 17:337-345.

Novikov, E., & Barillot, E. (2007). Software package for automatic microarray image analysis (MAIA). *Bioinformatics*, 23(5), 639-640.

Rueda L, "Sub-grid Detection in DNA Microarray Images." (2007), *Proceedings of the IEEE Pacific-RIM Symposium on Image and Video Technology*, 248-259.

Rueda L, "A fully automatic gridding method for cDNA microarray images", (2011), *BMC Bioinformatics*, 12:113.

Rueda L, Vidyadharan V: (2006), "A Hill-climbing Approach for Automatic Gridding of cDNA Microarray Images" *IEEE Transactions on Computational Biology and Bioinformatics*, 3:72-83.

Snijders,A.M., Nowak,N., Segreaves,R., Blackwood,S., Brown,N., Conroy,J., Hamilton,G., Hindle,A.K., Huey,B., Kimura,K. et al. (2001), "Assembly of microarrays for genome-wide measurement of DNA copy number." *Nature Genetics*. 29, 263–264,

Sorin, Draghici, *Data analysis tool for DNA Microarrays*, Chapman&Hall/CRC, Mathematical biology and medicine series, London NewYork, 2003.

Wang Y, Ma M, Zhang K, Shih F,(2007), "A Hierarchical Refinement Algorithm for Fully Automatic Gridding in Spotted DNA Microarray Image Processing." *Information Sciences* 177(4):1123-1135.

Y. Wang, F. Y. Shih, and M. Ma, (2005), "Precise gridding of microarray images by detecting and correcting rotations in subarrays," in *proceedings of Sixth Inter. Conf. on Computer Vision, Pattern Recognition and Image Processing*, Salt Lake City, UT.

Zacharia E, Maroulis D: "Microarray Image Gridding via an Evolutionary Algorithm" (2008), *IEEE International Conference on Image Processing* 2008, 1444-1447.

Deepa J received her M.Tech degree (Electronics) with specialization in Digital Electronics from Cochin University of Science and Technology (CUSAT) and B.Tech from Model Engineering college ,Kerala. She is currently serving College of Engineering, Chengannur as Associate Professor in Electronics Department and peruse towards her doctoral degree under the guidance of Prof. (Dr.) Tessamma Thomas at CUSAT. Her areas of interest include Signal /Image processing, Bioinformatics. She has to her credit more than 10 research papers published in International and National journals and conferences.

Dr.Tessamma Thomas received her M.Tech and Ph.D from Cochin University of Science and Technology, Kochi, India. At present, she is working as Professor in the Department of Electronics, Cochin University of Science and Technology. She has to her credit more than 90 research papers, in various research fields, published in International and National journals and conferences. Her areas of interest include digital signal / image processing, bio medical image processing, super resolution, content based image retrieval, genomic signal processing .

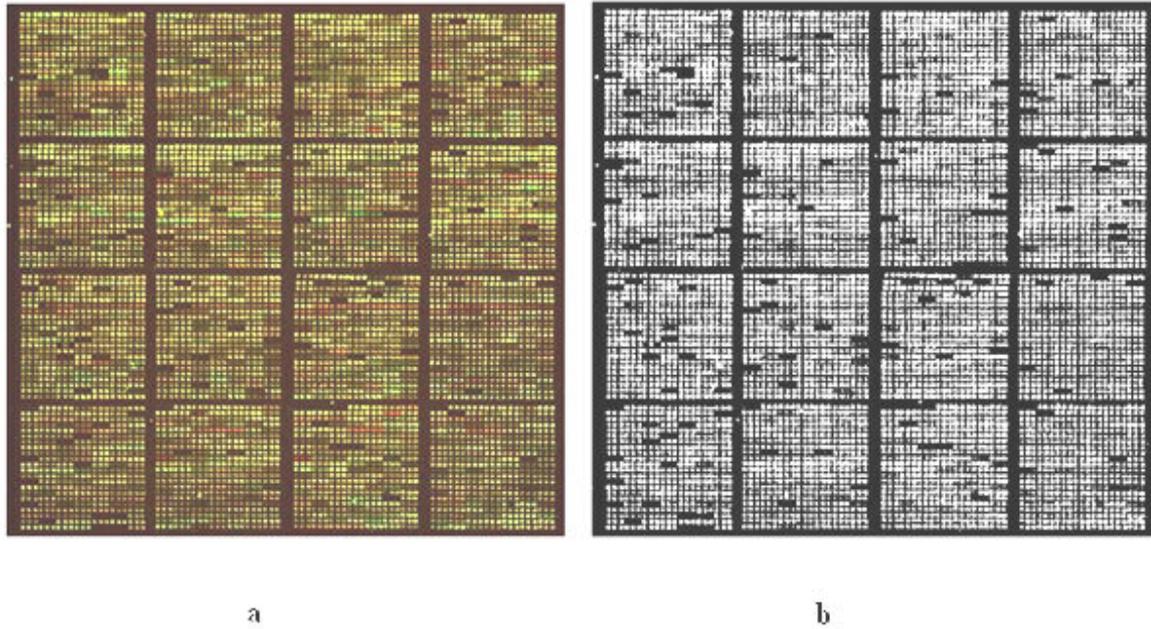


Figure 1. (a) Microarray image with 16 subarrays (b) Binary reference image after preprocessing

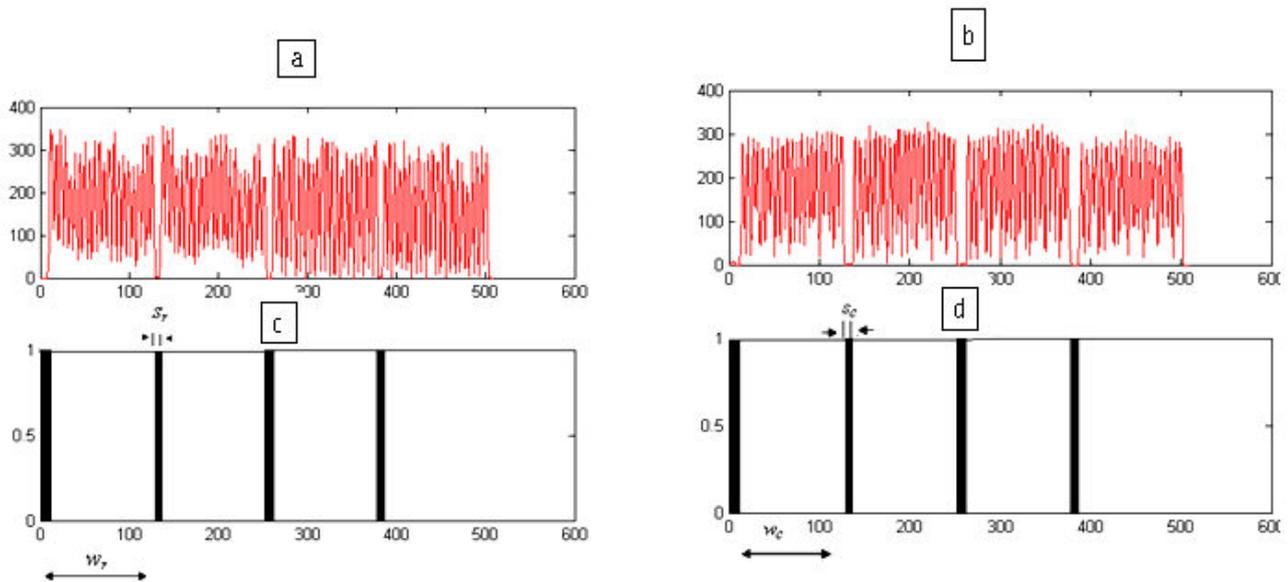


Figure2. i)Intensity Projection profiles of binary image in Fig 1(b). (a)row (b)column (c)Thresholded profiles row (d)column .

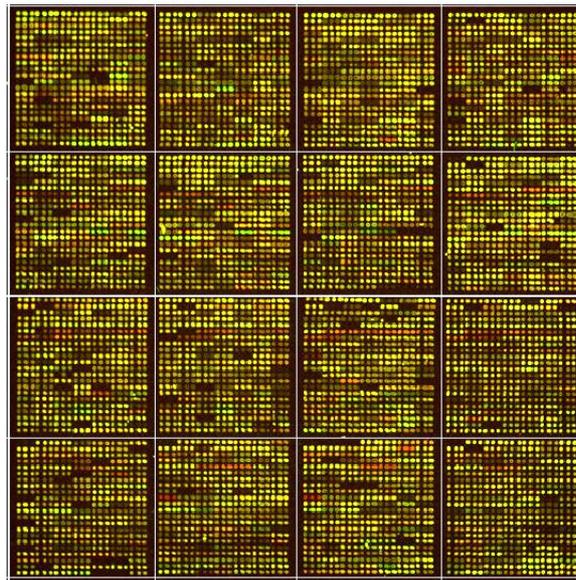


Figure 3. Microarray image after global gridding

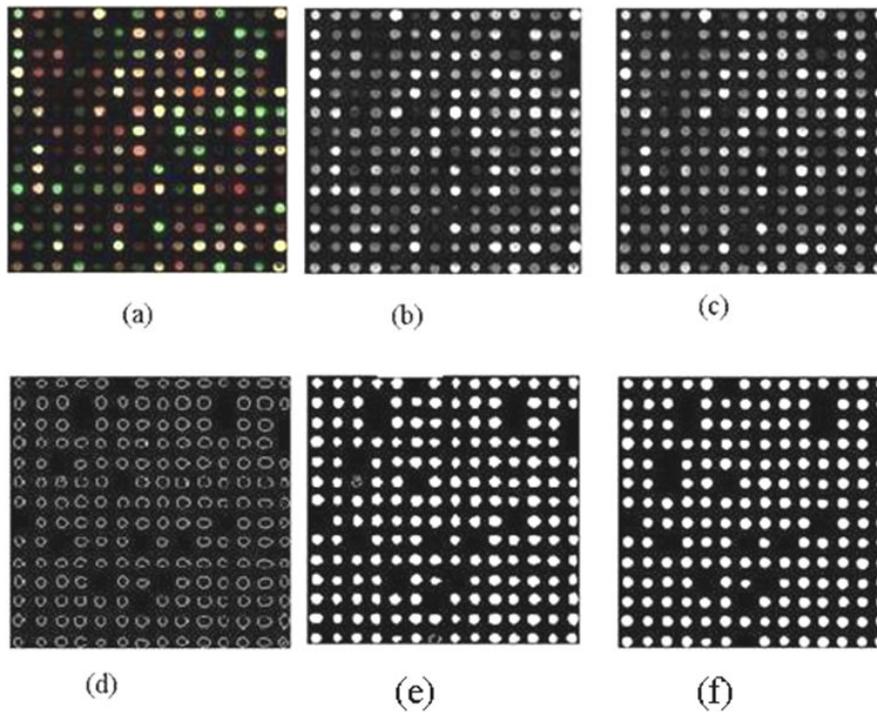


Figure 4. (a) A subarray of microarray image (b) Grayscale converted image (c) After Adaptive histogram equalization (d) Canny edge detection (e) Morphological filling (f) Binary reference image After applying Morphological sequential filter.

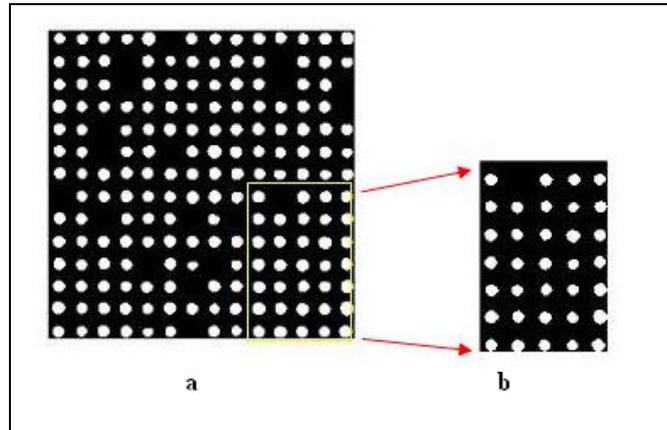


Figure 5. Identified optimum sub image (b) from the binary reference image (a)

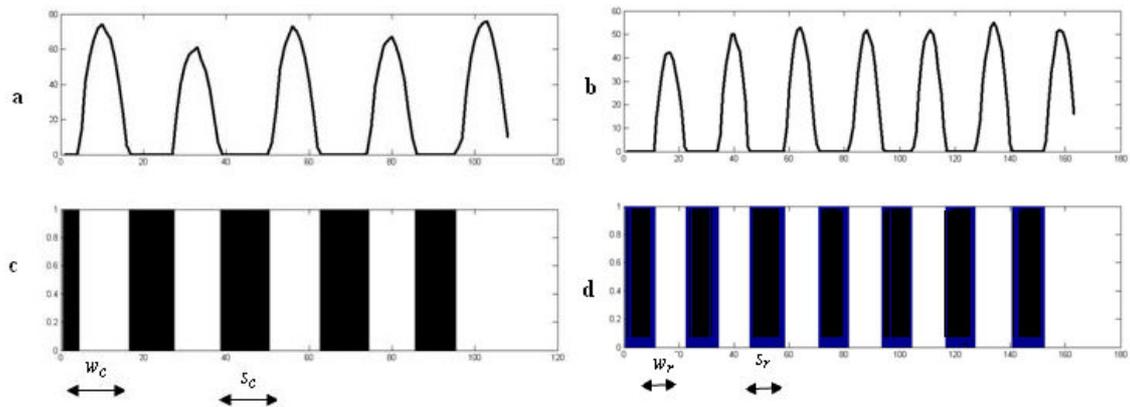


Figure 6. Intensity projection profile of optimum subimage (a) column (b) row. Thresholded column profile (c) and row profile (d)

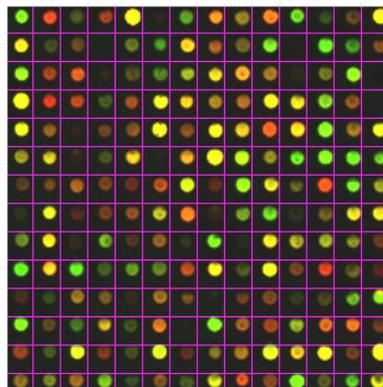
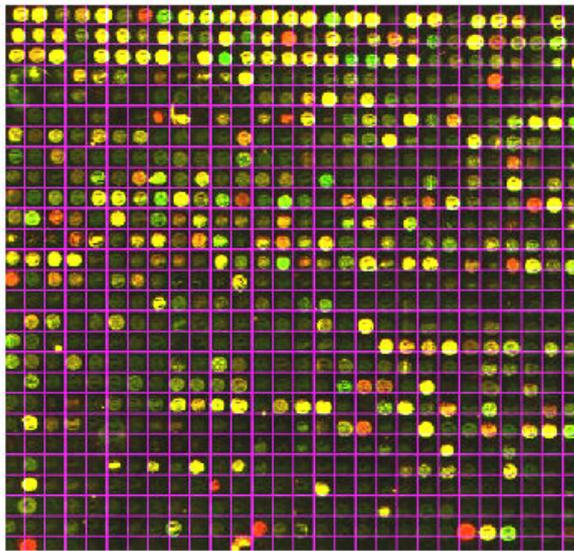
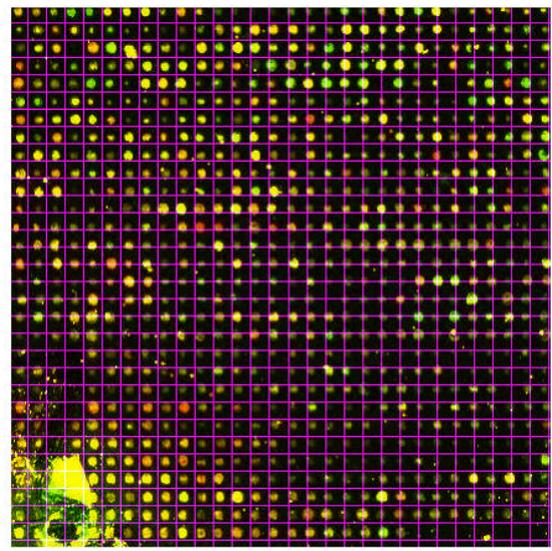


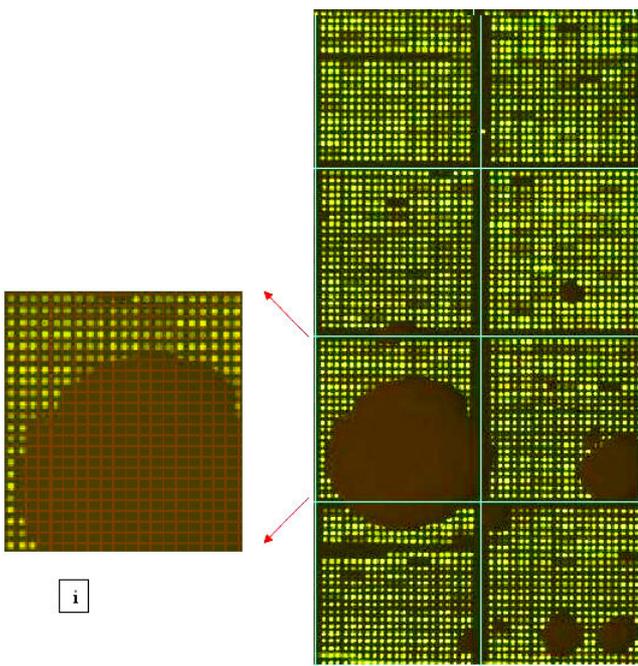
Figure 7. Gridded Image



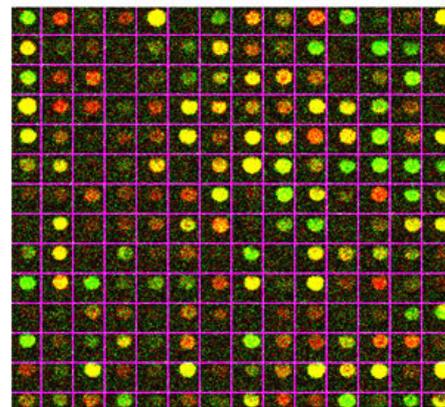
(a)



(b)



(c)



(d)

Figure 8. Gridded subarrays (a,b) images from SMD. (c) (i) A subarray with more than 50% contamination after gridding (ii) Microarray image (d) Gaussian noise added image after gridding

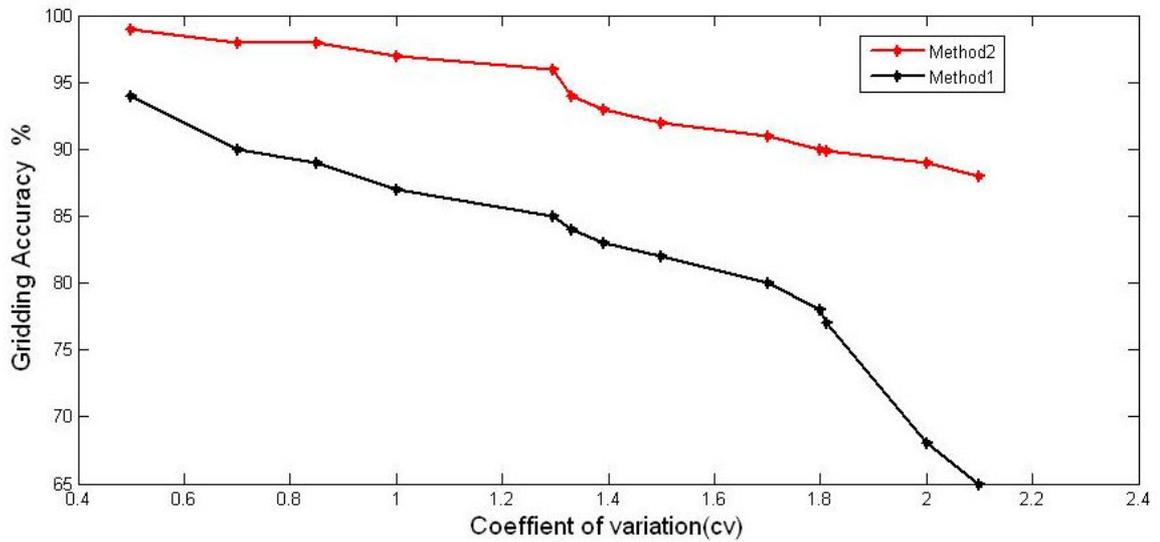


Figure 9. Plot of gridding accuracy vs. coefficient of variation (cv) of subarray images -Comparison between Method 1 (Yu.Wang *et al.*), Method 2 (proposed method)

Table 1. Gridding accuracy for noisy microarray images

Gaussian Noise		Salt&pepper Noise	
Variance (v)	Gridding Accuracy (%)	Noise density(D)	Gridding Accuracy (%)
.01	100	.01	100%
.02	100	.02	100%
.03	100	.05	100%
0.04	95	.09	100%
.05	88	.1	100%

Table 2. Execution time for gridding subarrays

Number of spots in the subarray	Time required (sec)
100	0.832361
196	1.432831
552	2.294093
756	2.586965
900	2.7523924

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage:

<http://www.iiste.org>

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** <http://www.iiste.org/Journals/>

The IISTE editorial team promises to review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library, NewJour, Google Scholar

