

Preprocessing of Microarray by Integrated OSR and SDF Approach for Effective Denoising and Quantification

J.K.Meher¹, P.K.Meher² and G.N.Dash³

¹ Department of Electronics & Telecom Engineering, SITE, Balangir, Odisha, India -767002

²Department of Embedded Systems, Institute for Infocomm Research, Singapore – 138632

³School of Physical Sciences, Sambalpur University, Jyoti Vihar, Odisha, India – 768019,

jk_meher@yahoo.co.in, pkmeher@i2r.astar.edu.sg, gndash@ieee.org

Abstract. Microarray data is inherently noisy due to the noise introduced during the preparation of microarray slide. Hence preprocessing of microarray image is an essential to eliminate the background noise in order to enhance the image quality and effective quantification. The objective of this paper is to introduce novel preprocessing techniques such as optimized spatial resolution (OSR) and spatial domain filtering (SDF) for reduction of noise from microarray data and reduction of error during quantification process for estimating the microarray spots accurately to determine expression level of genes. Besides an integrated optimized spatial resolution and spatial filtering is proposed and found improved denoising of microarray data with effective quantification of spots.

Keywords: Denoising, Microarray, Preprocessing, Quantification, Spatial domain filtering, Optimized spatial resolution.

1. Introduction

The DNA microarray technology enables to monitor the whole genome on a single chip so that researchers can have a big picture of the interactions among thousands of genes simultaneously [1]. For a microarray project, the image quantification makes the transition in the work flow from wet lab procedures to computational ones. This technology is widely used by biologists in the analysis of disease classification, genome expression in cancer, other genetic diseases, viral immunology and so on. Different technologies are in use to design microarrays, like microarrays of cDNA, oligonucleotides, Ink-jet/Bubble jet [2, 3]. The microarray quantification process deals with extraction of spot intensity for each and every probe by estimating the intensity of foreground and background values. The images are usually supplied in pairs where one image shows the intensity of green dye and the other shows the intensity of a red dye. One of the couple of most commonly used quantification methods is histogram method, and the other based on pixel intensity [4-7]. Spot detection and quantification based on histogram method is comprised of three major steps: gridding, segmentation and intensity extraction.

One of the major problems encountered in microarray image processing is due to the non-uniform spacing between the spots. Finding optimal grid is a time consuming process which requires human intervention [8]. For automatic analysis of cDNA microarrays, a method using genetic algorithm approach was implemented to determine optimal gridding and spot segmentation [9]. Dapple has suggested an intelligent technique to determine the location of spots using morphological information which is robust to both variation and artifacts [10]. The authors in [11] present a fully automatic system for microarray image quantification, which locates both subarray grids and individual spots, without user identification of any image coordinates. The analysis of the scanned images is not straightforward process since the quality of microarray images suffer due to noise, artifacts and uneven background. Discrete wavelet transform (DWT) has been used in image denoising. The denoising capabilities of DWT is demonstrated in [12] for removal of

noise that is introduced during the preparation of microarray data. The stationary wavelet transform (SWT) was introduced in 1996 to make the wavelet decomposition time invariant [13]. A two-stage approach for noise removal that processes the additive and the multiplicative noise component is presented in [14]. The denoising capabilities of decimated and undecimated multiwavelet transforms, DMWT and UMWT respectively are demonstrated in [15].

Microarray images consist mostly of low-intensity features that are not well distinguishable from the background. These problems lead to errors that propagate to all the stages of statistical analysis. Thus there is a need of developing new methods for upgradation of image quality. In this paper we present two novel preprocessing techniques, namely optimized spatial resolution and spatial domain filtering to eliminate the noise in microarray that helps in more accurate estimate of the intensity of spots to determine the expression level. Spatial filtering is used for denoising of microarray image and spatial resolution optimization is used to enhance the image for accurate quantification of the spots. In order to improve the quantification result an integrated spatial domain filtering and optimized spatial resolution could be used.

The remainder of this paper is organized as follows. Section 2 presents the proposed preprocessing techniques of microarray image. In section 3 we have discussed the simulation result and performance analysis. Conclusion is drawn in section 4.

2. Proposed Preprocessing Techniques of Microarray Image

The objective of preprocessing is to eliminate the noise in microarray image that helps in reducing the error during quantification process for appropriate interpretation of microarray data. We discuss here the proposed methods for spatial domain filtering, spatial resolution optimization technique and an integrated approach of the former two for preprocessing the microarray image by denoising and image enhancement.

Optimized Spatial Resolution

Spatial resolution is the density of pixels over the image. The greater the spatial resolution, the more pixels are used to display the image. The microarray image consists of large number of pixels. For evaluation of the gene expression data extensive computation is required because of the fact that the number of spots is very large and the data within each spot is also substantial. It is found that pixel intensities of the microarray are appeared in a particular order in alternate rows. The spatial resolution can be optimized for the entire slide by selecting optimal intensity of two consecutive rows without altering the size of the image in both red plane and green plane. In this process the red plane and green plane are separated from the microarray. Each plane is optimized by selecting the higher values of consecutive rows in red plane and smaller values of two consecutive rows of green plane. This property has been selected from the fact that the signal intensity is the ratio of red foreground intensity to the green foreground intensity as discussed in section 5. Blue plane has no role on the signal estimation. Hence it is left untouched. These modified planes are restored in a new image resulting in optimized spatial resolution. This process results in enhanced image quality and signal intensity. The histogram plot (Fig.1) shows pixel intensity having higher magnitude of the effective modified resolution as compared to that of original image. When the optimized image is postprocessed to quantify the spot intensity, it results better estimation of spot as compared to processing original raw image.

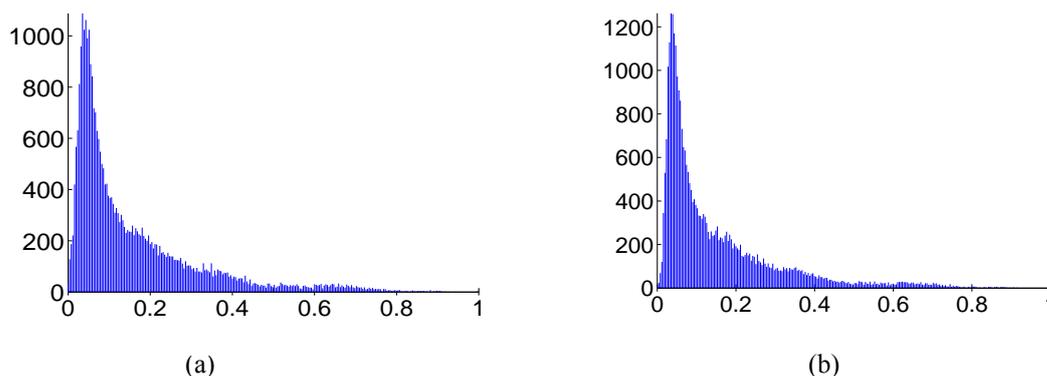


Fig. 1. Histogram plot of (a) Original Image and (b) optimized image.

Spatial domain filtering

In spatial filtering approach, we move a rectangular mask of the order $m \times n$ over the given microarray image. This process creates a new image with gray values calculated by the eqn(1).

$$\sum_{s=-(m-1)/2}^{(m-1)/2} \sum_{t=-(n-1)/2}^{(n-1)/2} w(s,t) \cdot p(i+s, j+t) \quad (1)$$

The mask of the order $m \times n$ represented as a matrix is called a filter. A linear filter can be implemented by multiplying all the elements in the mask by corresponding elements in the neighborhood pixels in the area spanned by the filter mask and adding together all these products [16]. This process is repeated for every pixel in the microarray image. For a 3×3 mask with mask values $w(i,j)$ and that corresponding pixel values $p(i,j)$ as shown in Fig.2, the new pixel value is computed by the eqn(2).

$$\sum_{s=-1}^1 \sum_{t=-1}^1 w(s,t) \cdot p(i+s, j+t) \quad (2)$$

m(-1,-1)	m(-1,0)	m(-1,1)
m(0,-1)	m(0,0)	m(0,1)
m(1,-1)	m(1,0)	m(1,1)

(a)

p(i-1,j-1)	p(i-1,j)	p(i-1,j+1)
p(i,j-1)	p(i,j)	p(i,j+1)
p(i+1,j-1)	p(i+1,j)	p(i+1,j+1)

(b)

Fig. 2. (a) mask values, (b) pixel values.

The microarray image obtained by spatial filtering has higher intensity of signals and thus background noise is suppressed thereby improving the signal intensity. This results higher spot intensity in the quantification process. The intensity of the spot depends on the design of the mask. The mask representing Lowpass Spatial filters based on neighbourhood pixels denoted as averaging filter. A mask consisting neighborhood pixels of size 3×3 acts as linear filter. The average of all the nine values within the mask becomes the gray value of the corresponding pixel in the new microarray image. Each pixel is processed by its neighborhood pixels of mask 3×3 and the new pixel is obtained by

$$P = \frac{1}{9} (a + b + c + d + e + f + g + h + i) \quad (3)$$

The process is continued for the whole image. The new image obtained is the denoised image. This is called lowpass spatial filter. Spot intensity is quantified effectively when this image is processed. 3×3 mask shows better result in compared to 5×5 mask, 3×5 mask. Allied to spatial filtering is spatial convolution. The method for performing a convolution is the same as that for filtering, except that the filter is rotated by 180° before multiplying and adding. Using the $w(i,j)$ and $p(i,j)$ notation, the output of a convolution with a 3×3 mask for a single pixel is given by eqn(4) or eqn(5).

$$\sum_{s=-1}^1 \sum_{t=-1}^1 w(-s,-t) \cdot p(i+s, j+t) \quad (4)$$

$$\sum_{s=-1}^1 \sum_{t=-1}^1 w(s,t) \cdot p(i-s, j-t) \quad (5)$$

Here we have rotated the image pixels by 180° ; this does not affect the result. In practice, most filter masks are rotationally symmetric, so that spatial filtering and spatial convolution will produce the same output. Similarly a 3×3 Gaussian filter can effectively eliminate background noise and image blurring.

Gaussian filter of mask 3×3 shows better result than 5×5 mask and 3×5 mask.

Integrated Approach

In this paper we have presented two new preprocessing methods such as the spatial domain filtering and the optimized spatial resolution to deal with noise reduction in microarray image. Both the methods have respective advantages of denoising and image enhancement. Now an integrated spatial domain filtering and optimized spatial resolution approach is proposed on microarray images to combine the advantages of both that enables effective quantification. The integrated approach is summarized in Fig. 3 that consists of cascaded OSR and SDF stages. The preprocessing stage encompasses procedures for optimizing the spatial resolution followed by spatial lowpass filtering. The combined process not only helps in denoising but also performs image enhancement resulting in high signal to noise ratio. The resultant image is now suitable for post processing. The postprocessing entails computing the resulting image for estimation of spot intensity which follows three steps such as gridding, segmentation and information extraction. It is found that the integrated approach shows much higher spot intensity compared to traditional methods as shown in Table 2.

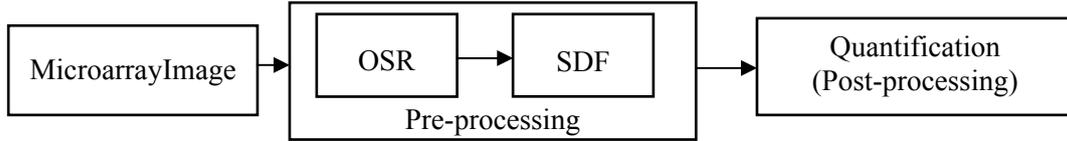


Fig. 3. Basic steps for Integrated Approach.

3. Simulation Result and Performance Analysis

Extensive study has been made in eliminating noise and ensures better gene expression by estimating the spot intensity with the help of novel proposed pre-processing techniques on microarray image. The estimation of microarray spot is discussed using quantification measures known as spot intensity (SI) or spot quality [3]. On the other hand, the assessment parameters that are used to evaluate the performance of noise reduction by different methods employed in this paper are signal to noise ratio (SNR), peak signal to noise ratio (PSNR), image fidelity (IF), structural content (SC), absolute average difference (AAD) and Correlation Quality (CQ) [17].

$$\text{Spot Intensity: } S_I = \frac{F_I}{B_I}, \quad F_I = \frac{R_{fg}}{G_{fg}}, \quad B_I = \frac{R_{bg}}{G_{bg}} \quad (6)$$

where SI is Spot Intensity, FI is the foreground intensity and BI is the background intensity. Information is extracted for each spot on the array in terms of signal intensity which is the mean or median of foreground pixel intensities within a spot. Spot foreground intensities are red foreground Rfg and green foreground Gfg. Similarly spot background intensities are red background Rbg and green background Gbg intensities. Background information is extracted by the existing morphological opening which is a non-linear filter that generates an image of the estimated background intensity for the entire slide. Spot quality is the ratio of foreground to the background ratio. The Spot intensity is computed by several preprocessing techniques such as spatial domain filtering based on neighborhood mask processing, Gaussian mask, median filter, ordered filter, Wiener filter, frequency domain filter and wavelet transform filter on different microarray slides and found that the proposed integrated approach produces high spot intensity that results better gene expression.

$$\text{Signal to Noise Ratio: } SNR = \frac{\sum_{x,y} I(x,y)^2}{\sum_{x,y} (I(x,y) - I_d(x,y))^2} \quad (7)$$

$$\text{Peak Signal to Noise Ratio: } PSNR = \frac{M * N * \max_{x,y} I(x,y)^2}{\sum_{x,y} (I(x,y) - I_d(x,y))^2} \quad (8)$$

$$\text{Image Fidelity: } IF = 1 - \frac{1}{SNR} \quad (9)$$

$$\text{Structural Content: } SC = \frac{\sum_{x,y} I(x,y)^2}{\sum_{x,y} I_d(x,y)^2} \quad (10)$$

Absolute Average Difference:

$$AAD = \frac{\sum_{x,y} |I(x,y) - I_d(x,y)|}{M * N} \quad (11)$$

Correlation Quality:

$$CQ = \frac{\sum_{x,y} I(x,y) * I_d(x,y)}{\sum_{x,y} I(x,y)} \quad (12)$$

In general, higher SNR and PSNR are better; the signal is cleaner and indicates a smaller difference between the original and denoised image. If IF nears 1, we will obtain an image of better quality. Similarly if SC spread at 1, we will obtain image of better quality. A lower AAD gives a cleaner image as more noise is reduced. A larger value of CQ usually corresponds to a better quantitative performance. In this work, we have simulated all the proposed methods in Matlab 7.0 for preprocessing and postprocessing operations. Table 1 summarizes the assessment parameters for denoising methods. Table 2 summarizes spot intensity of one spot in the microarray slide. It is found that the proposed spatial domain filtering based on neighborhood mask filtering and Gaussian mask filtering play vital role in reducing noise in microarray image and thus helps in quantification effectively. Further optimized spatial resolution enhances the image quality and thus improves the signal intensity. Hence the optimized image produces improved estimation of spot intensity when postprocessed. Finally an integrated optimized spatial resolution and spatial lowpass filtering produces still much improved result than that has been done by raw image and other preprocessing methods. The quantitative results of table 1 and 2 show the optimum quantity assessment parameters for the proposed method. Thus the enhanced the image quality and improved denoising capability are observed especially in integrated approach and hence it helps in quantification effectively.

TABLE I
NUMERICAL RESULTS ON ASSESSMENT PARAMETERS FOR DENOISING METHODS

Denoising Methods	SNR	PSNR	IF	AAD	SC	CQ
OSR	48.7588	29.9655	0.9795	1.3812	1.0732	76.9673
Spatial LPF	29.1622	27.7332	0.9657	0.0072	1.0975	75.5819
Gaussian filter	246.2552	36.9988	0.9959	0.0267	1.0386	78.8787
Wiener filter	81.6738	32.2058	0.9878	-0.0619	1.0475	75.1124
Median filter	42.5303	29.3720	0.9765	1.0065	1.0802	76.6032
Ordered filter	3.5297	18.5623	0.7167	-19.6239	0.5658	100.0381
Frequency domain filter	4.4206	19.5398	0.7738	-3.3937	1.4905	58.1798

TABLE II
ESTIMATION OF SPOT INTENSITY BY DIFFERENT METHODS

Preprocessing Methods	Spot Intensity
Processing raw image	1.43
Frequency domain LPF	1.53
Gaussian mask	1.57
Median filter	1.56
Ordered filter	1.57
Wiener filter	1.56
Optimized spatial resolution	1.59
Spatial LPF (neighbourhood mask)	1.58
Integrated approach (OSR+SDF)	2.07

4. Conclusion

In this paper a new approach using integrated optimized spatial resolution and spatial domain filtering is presented to deal with enhancement of image quality and reduction of noise for effective quantification of microarray image. Results computed using proposed optimized spatial resolution and spatial lowpass

filtering show improved interpretation in contrast to one derived from the raw data analysis and other preprocessing methods. Various microarray images were examined to validate the performance of our methods. Experimental results show that the proposed algorithm provides better performance than the other methods from quantitative analysis. The information extracted on down and up regulation of genes under diseased condition can be further analyzed in pharmacogenomics for drug targeting and drug development for the disease.

5. References

- [1] T. K. Attwood and D.J. Parry-Smith, "Introduction to bioinformatics", Addison Wesley Longman Limited, Harlow, England, 1999.
- [2] Joaquin Dopazo, "Microarray data processing and analysis", Bioinformatics Unit, Spanish National Cancer Center (CNIO), Melchor Fernandez Almagro 3, 28029, Madrid, Spain.
- [3] Pre-processing in DNA microarray experiments, Sandrine Dudoit, PH 296, Section 33, 13/09/2001.
- [4] Deepa Vijayan, M Phil Scholar, Supervisors, Dr. Achuthsankar. S. Nair, Hon. Director, Centre for Bioinformatics, University of Kerala, India Microarray image processing: spot detection, quantization and clustering.
- [5] Basim Alhadidi, Hussam Nawwaf Fakhouri and Omar S. AlMousa, "cDNA Microarray Genome Image Processing Using Fixed Spot Position", American Journal of Applied Sciences 3 (2): 1730-1734, 2006, ISSN 1546-9239.
- [6] Shenghua, Pan, Mihaela P, Weizhong and Andrei P, "Spotted cDNA microarray image segmentation using ACWE", Romanian Journal of Information Science and Technology Volume 12, Number 2, 2009, 249(263).
- [7] Rastislav Lukac, Konstantinos N. Plataniotis, "cDNA Microarray Image Segmentation Using Root Signals", Multimedia Laboratory, The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada, Vol. 16, 5164, 2006.
- [8] Ajay N. Jain, Taku A. Tokuyasu, Antoine M. Snijders, Richard Segraves, Donna G. Albertson, and Daniel Pinkell, "Fully Automatic Quantification of Microarray Image Data", 12:325332, 2002 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/02
- [9] A.Sreedevi, D.S.Jangamashetti, "Automatically locating Spots in DNA Microarray Image Using Genetic Algorithm without Gridding," IACSIT spring conference, 978-0-7695-3653-8/09 2009 IEEE.
- [10] Buhler.J., T. Ideker , D. Haynor, "Dapple: Improved Techniques for finding spots on DNA Microarrays", UVCSE Technical Report UWTR 2000-08-05.
- [11] Jain A. N., T. A. Tokuyasu, A. M. Snijders, R. Segraves, D. G. Albertson and D.Pinkel, "Fully Automated Quantification of Microarray Image Data," GenomeResearch, Vol. 12, Issue 2, February 2002, pp. 325-332.
- [12] Qazi M. Ali and Omar Farooq Qazi M. Ali and Omar Farooq, "Wavelet transform for denoising and quantification of microarray data.", AMU Aligarh, India
- [13] X. H. Wang, Robert S. H. Istepanian and Yong Hua Song, "Microarray Image Enhancement by Denoising Using Stationary Wavelet Transform", Transactions On Nanobioscience, Vol. 2, No. 4, December 2003
- [14] Hara Stefanou1, Thanasis Margaritis, Dimitris Kafetzopoulos, Konstantinos Marias, Panagiotis Tsakalides., "Microarray Image Denoising Using a Two-Stage Multiresolution Technique", This work has been funded by the Marie Curie ToK-DEV "ASPIRE" grant within the 6th European Community Framework Program.
- [15] Ali Zifan, Mohammad Hassan Moradi, Shahriar Gharibzadeh, "Microarray image enhancement by denoising using decimated and undecimated multiwavelet transforms", SIViP DOI 10.1007 s11760-009- 0109-4
- [16] R.C. Gonzalez, R.E. Wood, and S.L. Eddins, "Digital Image Processing using Matlab", 2004.
- [17] Mario Mastriani, and Alberto E. Giraldez, "Microarrays Denoising via Smoothing of Coefficients in Wavelet Domain", International Journal of Biomedical Sciences Volume 1 Number .1