Microarray Image Denoising Using a Two-Stage Multiresolution Technique *

Hara Stefanou[1,2], Thanasis Margaritis[3,4], Dimitris Kafetzopoulos[3], Konstantinos Marias[1]
Panagiotis Tsakalides[1,2]

[1] Institute of Computer Science, Foundation of Research and Technology, Heraklion, Crete, Greece,
hstef@ics.forth.gr, tsakalid@ics.forth.gr, kmarias@ics.forth.gr

[2]University of Crete, Computer Science Department, Heraklion, Greece, hstef@csd.uoc.gr,tsakalid@csd.uoc.gr

[3] Institute of Molecular Biology and Biotechnology, Foundation of Research and Technology,
Heraklion, Crete, Greece, thama@imbb.forth.gr, kafetzo@imbb.forth.gr

[4]University of Crete, Biology Department, Heraklion, Greece

Abstract

DNA microarrays have demonstrated an excellent potential in correlating specific gene expression profiles to specific conditions. However, they are affected by inherent noise. This paper presents a two-stage approach for noise removal that processes the additive and the multiplicative noise component. The proposed approach first decomposes the signal by a multiresolution transform and then accounts for both the multiscale correlation of the subband decompositions and their heavy-tailed statistics. Real microarray images have been processed by the proposed method and its improved performance is shown through quantitative measures and qualitative visual evaluation.

1. Introduction

Over the last decade, a revolution has been witnessed in the biosciences, medical sciences, biotechnology and pharmaceutical industry. The post-genomics revolution is driven mostly by microarray technology [22], [26] as it allows the concurrent observation of the expression of all known genes. Because patterns of gene expression correlate strongly with function [8], [29], microarrays are providing unprecedented information both on basic and applied research. Microarrays can also be used for the identification of alterations in gene sequences, paving the way for a new era of genetic screening, testing and diagnostics.

Microarray experiments, however, involve a large number of error-prone procedures that lead to a high level of noise in the resulting data. The high level of the uncertainty associated with each microarray experiment originates by biological variations and experimental noise. Changes in the measured transcript values in the samples render the clustering of genes into functional groups [8], [29] and the classification of samples difficult [9], [31]. A major challenge in microarray analysis is to eliminate the effect of the noise and recover the gene expression measurements.

Several authors have addressed the noise issue explicitly. Lee *et al.* [14] reached the conclusion that repetition can increase the significance of conclusions from gene array experiments. Some techniques consider only the process statistics using either an additive model [3], [20] or a model incorporating both additive and multiplicative noise [30]. More theoretical approaches were introduced by Hartemink *et al.* [10] and Dror *et al.* [5] for the Affymetrix GeneChip data; the former assumed a log-normal data distribution while the latter developed the Bayesian estimation of array measurements (BEAM) technique. Rocke and Durbin [24] modeled the measurement error in cDNA spotted arrays as a function of the expression level.

The importance of including both additive and multiplicative measurement-specific noise in an error model for gene arrays is already established in the literature [24]. The omission of the measurement-specific additive noise term leads to exaggerated ratio estimates, false identification of significant differences, and understated uncertainty measures when the observations are small. The omission of the multiplicative noise term leads to similar problems when the observations are large.

A major source of noise, often overlooked by researchers, is caused by the microarray image generation process. This, usually, involves the collection of fluorescence of the labeled samples, the amplification of the analog signal and the conversion to digital through dedicated imaging devices. The microarray area is divided into equally sized pixels and the imaging device produces a digital map

*This work has been funded by the Marie Curie ToK-DEV "ASPIRE" grant within the 6th European Community Framework Program.

(image channel) of the fluorescence intensities for each pixel in the form of 16-bit tagged image files.

Previous studies have used methods for image enhancement to address the effect of the microarray image noise [32], [17], [16], [18], [19]. Daskalakis *et al.* [1] introduced a complete framework for microarray image analysis, which takes into account the effect of local spot-image noise in microarray images for improving spot segmentation and subsequently gene quantification.

Microarray images consist mostly of low-intensity features, that are not well distinguishable from the background. This is a result of the great differences on transcript abundance in all eukaryotic cells, which is a common knowledge for over 30 years now [2]. The most interesting genes, including cell cycle and transcriptional regulators have a low level of expression level of one copy per cell [13]. To increase the dynamic range of the measurements and better determine the most important genes, researchers produce multiple images of the same microarray at increasing detection settings [6], [28] and transform the intensity values of the individual genes into one "true" measurement. By increasing the detection settings, the source noise (mostly additive), which includes photon noise and dust on the slides, remains unaffected. On the other hand, the detector noise (mostly multiplicative), which includes features of the amplification and digitization process, is increased [33].

In this paper, we propose an image denoising method which accounts for both noise components and makes the microarray spot area more homogeneous and more distinctive from their local background. The proposed approach consists of two stages: one that processes the additive component of the noise and one that processes the multiplicative component. The method first performs a multiresolution decomposition of the image and then accounts for the heavy-tailed statistical behavior of the representation coefficients as well as for their strong statistical dependence across multiple scales. The utility of this framework is validated with real microarray data through visual evaluation and quantitative performance metrics.

Subband decompositions of an image have significantly high-order statistics that are eluded by the simple thresholding methods. On the other hand, a Bayesian denoising method exploits these higher-order statistics rendering it a more reasonable choice for image processing. Moreover, the correlation between the wavelet coefficients of adjacent scales infers that there is a significant feature at the position that should be passed through the filter. Therefore, a method which exploits this dependence would be an essential denoising method.

The paper is organized as follows: In Section 2 we describe the characteristics of the used microarray images (Subsection 2.1) and analyze, in details, the proposed approach (Subsection 2.2). Section 3, validates the performance of our proposed method through qualitative visual evaluation and quantifies the achieved performance improvement. Finally, conclusions are drawn in Section 4.

2. Materials and Methodology

2.1. Materials

The microarray images, that are processed by the proposed method, are 16-bit grayscale images. They come from the high detection settings of a homotypic hybridization of a leukemic cell line (KARPAS-231, human B cell leukemia) on microarrays containing oligos corresponding to both human genes (Operon Human Oligo Library v3) and external control genes from *Bacillus subtilis*, *E. coli* and *phage P1*.

4 μg of total RNA were amplified, as described in [7] with the following modifications: first strand synthesis was perfomed using an anchored T7-oligo (dT) primer and Superscript III (Invitrogen) for 20min at 44^0C and 1h 45min at 50^0C. The amplified RNA (aRNA) in vitro transcription was done with Ampliscribe T7 transcription kit (Epicentre) at 42^0C following the manufacturer's protocol, using a ratio of UTP/aminoallyl-UTP (Epicentre) of 1. The aRNA was cleaned up using RNeasy columns (Qiagen) and quantified using Nanodrop. Equal quantities of aRNA, supplemented with aRNA of the externalcontrol genes at 1:1, 1:3 and 3:1 ratios, were labelled with Alexa 555 and 647 Succinimidyl ester (Molecular Probes, Invitrogen) according to the manufacturer's protocol at 50^0C.

The hybridization was performed in a Tecan HS4800 hybridization station. The slides were prehybridized for 1.5 hour at 42^0C with 5x SSC, 0.1% SDS, 1% BSA and then hybridized for 16h at 42^0C in a buffer with a final concentration of 5x SSC, 0.1 %SDS, 50% formamide and 1.5μg/ml fragmented salmon sperm DNA. The arrays were subsequently washed with 2x SSC, 0.1% SDS at 42^0C, followed by a wash in 0.1x SSC, 0.1% SDS at 23^0C and a third wash with 0.1x SSC at 23^0C. The slides were dried by nitrogen.

Arrays were read with a ScanArray 5000 scanner (GSI Lumonics) at 5 μm resolution at three different photomultiplier tube voltage settings (high, medium and low). The fluorescence intensity for each fluor and each element on the array was captured using spotSegmentation package [15] written in R [23].

This image analysis software allows the quantitation of microarray measurements without user intervention. It also identifies the presence or absence of a given spot or artifacts by clustering each spot's area in 1 (only background) to 3 (background, spot and artifact) classes and using the appropriate Gaussian mixture model that maximizes the Bayesian Information Criterion (BIC).

2.2. Methodology

In general, the expression level uncertainty in microarray systems, fundamentally originates from the probabilistic characteristics of the detection process, from sample extraction and mRNA purification to hybridization and imaging. In the following, we formulate the microarray image noise removal problem starting with a brief essential overview of the signal model.

Denote by *I(x,y)* a noisy observation of the microarray image *S(x,y)* that has to be recovered and by $n_m(x,y)$ and $n_a(x,y)$ the multiplicative and additive noise components, respectively. The signal model is

$$I(x,y) = S(x,y) \cdot n_m(x,y) + n_a(x,y) \tag{1}$$

To estimate the multiplicative noise component we have to ignore the additive component $n_a(x,y)$. Then, equation (1) becomes

$$I(x,y) = S(x,y) \cdot n_m(x,y). \tag{2}$$

The multiplicative noise model can be transformed into an additive one, by applying the logarithmic function in both sides of (2)

$$\log I(x,y) = \log S(x,y) + \log n_m(x,y). \tag{3}$$

Then, if we replace the logarithms of $I(\cdot)$, $S(\cdot)$, $n_m(\cdot)$ with $f(\cdot)$, $g(\cdot)$, $\epsilon(\cdot)$, respectively, (3) can be rewritten as

$$f(x,y) = g(x,y) + \epsilon(x,y). \tag{4}$$

As the initial processing step, we employ an undecimated wavelet decomposition of the image. The undecimated wavelet transform, called the à trous algorithm [11], is a multiresolution analysis tool with good localization properties in scale-space representation, it is shift-invariant and it decomposes the signal in subbands that are correlated across different levels of analysis. The low pass-filter *h* is said to be à trous if it satisfies

$$h_{2n} = \frac{\delta(n)}{\sqrt{2}}, \tag{5}$$

where δ(n) is the Kronecker delta function.

From the original image $I(x,y)$, we get a smoothed approximation $A_1(x,y)$ and the wavelet plane $W_1(x,y)$ (also termed *detail image*). The decomposition formula is:

$$W_i(x,y) = A_i(x,y) - A_{i-1}(x,y), 0 < i \leq J, A_0(x,y) = I(x,y) \tag{6}$$

while the reconstruction is performed by adding up the images in set *W*:

$$I(x,y) = A_J(x,y) + \sum_{i=1}^{J} W_i(x,y) \tag{7}$$

The wavelet transform is a linear operation. Consequently, after applying the à trous algorithm to (4), we get in each of the three directions, sets of noisy wavelet coefficients written as the sums of the transformations of the signal and noise

$$d^i_{j,k} = s^i_{j,k} + \xi^i_{j,k} \tag{8}$$

where $k = 0, \ldots, 2^{J+j}-1$ and $-1 < j < -J$ refer to the decomposition level or scale and $i = 1, 2, 3$ refers to the three spatial orientations.

After the signal is transformed, the computed coefficients are further processed in order to get an image with better resolution and more distinctive attributes. A number of well-known image processing techniques, including soft and hard thresholding, Bayesian denoising based on Gaussian or Laplacian signal modeling, and multiresolution methods that exploit the correlation between the wavelet coefficients of adjacent scales have been applied to microarray images by ordinarily assuming the presence of either additive or multiplicative noise.

During the first processing stage of the proposed approach, henceforth referred to as the *coring stage*, the multiresolution coefficients in each subband are described by a heavy-tailed Laplacian statistical model. Then, a Bayesian noise removal algorithm is applied, which performs a "coring" operation to the data. The "coring" non-linear noise suppression preserves high-amplitude observations while suppressing low-amplitude values from the high-pass bands of the decomposed signal [27].

In order to estimate the true signal *s* from the noisy observation *d* we use the maximum *a posteriori* estimator (MAP) which is:

$$\hat{s}(d) = \arg\max_s p_{s|d}(s|d). \tag{9}$$

Using the Bayes rule, we get

$$\begin{aligned} \hat{s}(d) &= \arg\max_s p_{d|s}(d|s) \cdot p_s(s) \\ &= \arg\max_s p_\xi(d-s) \cdot p_s(s), \end{aligned} \tag{10}$$

where $\xi = d - s$ is the noise. The Bayesian estimator designed under the assumption of Laplacian signal in zero-mean Gaussian noise results in

$$\hat{s}(d) = sign(d) \left(|d| - \frac{\sqrt{2}\sigma_n^2}{\sigma} \right) \tag{11}$$

which is an optimal soft thresholding operation with a variance-dependent threshold.

At the second processing stage, henceforth referred as the *correlation stage*, we exploit the correlation between the wavelet coefficients of adjacent levels in order to discriminate microarray spots from additive noise. This approach

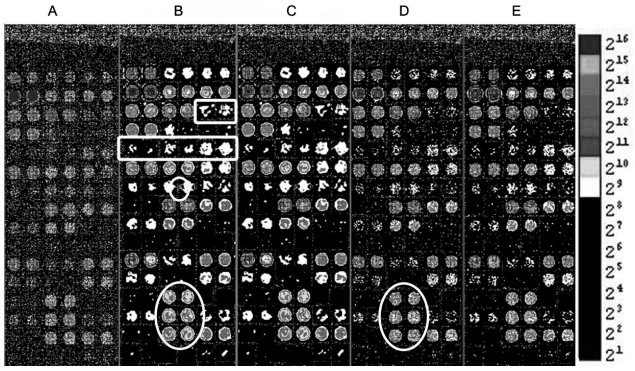

Figure 1. Results of the proposed two-stage approach: correlation stage for additive noise removal and coring stage for multiplicative noise removal. A is the original image and B-E are the images processed with the correlation stage, the correlation stage and then with the coring stage, the coring stage, the coring stage and then with the correlation stage, respectively. Box areas show low-intensity spots enhanced by the correlation method. Elliptic areas show spots processed by the correlation method that have been dilated (Subfigure B) while do not when being processed with the coring stage (Subfigure D).

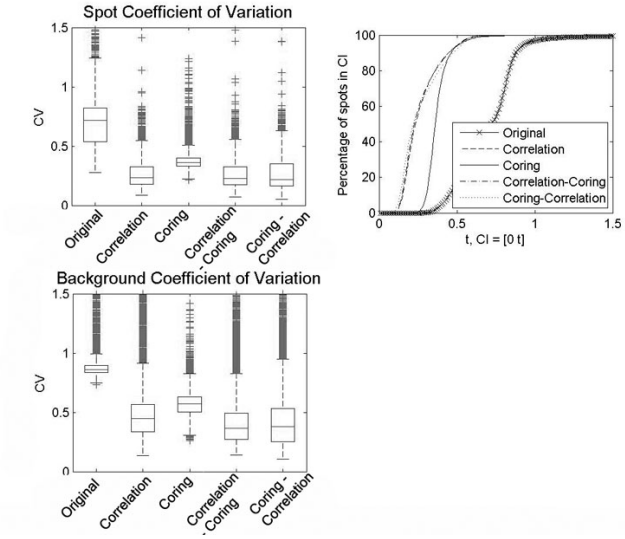

Figure 2. Effect of the two-stage approach on the homogeneity of the microarray spot and background areas. The upper set of boxplots represents the signal standard deviation to signal mean ratio. The lower set of boxplots represents the background standard deviation to background mean ratio. In each boxplot set, the first plot corresponds to the original image, the second corresponds to the image processed by the correlation stage, the third corresponds to the image processed by the coring stage, the fourth corresponds to the image processed by the two-stage correlation followed by coring method and the fifth corresponds to the image processed by the two-stage coring followed by correlation method.

is based on the fact that, when assuming only the additive noise component, local maxima in wavelet planes tend to propagate across scale when they are due to significant discontinuities in the image, while they do not propagate if caused by noise [12].

We compute a correlation image $P_j x, y$ which is defined at each location (x,y) by the direct spatial multiscale product of the wavelet coefficient images at adjacent levels in the à trous representation:

$$P_j(x,y) = \prod_{i=1}^{J} W_i(x,y). \tag{12}$$

Therefore, we exploit the fact that the correlation product of significant coefficients across scales at a specific location has a large value only if the local maxima propagate down to the considered scale.

Before computing correlations, it is desirable to select the most significant wavelet coefficients and to reduce the influence of non-significant noisy coefficients by thresholding [21]. The thresholding strategy is the $k\sigma$ hard thresholding technique

$$t_{hard}(W_i^Y, t_i) = \begin{cases} W_i^Y & W_i^Y \geq t_i \\ 0 & W_i^Y < t_i \end{cases} \tag{13}$$

with $t_i = k\sigma_i$, where σ_i is the standard deviation of the noisy wavelet coefficients at scale *i.* A robust estimation of σ_i is obtained from the MAD estimate [25] and is given by

$$\sigma_i = \frac{\bar{\sigma}}{0.67}, \tag{14}$$

where $\bar{\sigma}$ is the median absolute deviation of the wavelet coefficients at scale *i.*

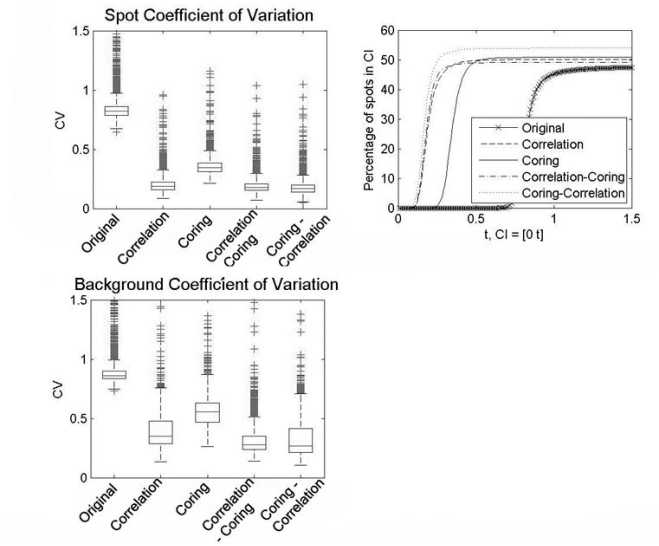

Figure 3. Effect of the two-stage approach on the homogeneity of the microarray spot and background areas that correspond to low-intensity spots.

3. Results

Figures 1-5 show qualitative and quantitative results from the application of this two-stage approach in real high-resolution DNA microarray images. Figure 1 shows the original and processed images for the two processing stages and order thereof. The box areas in Figure 1B show that the correlation stage significantly enhances low-intensity spots therefore discriminating them from the local noisy background. We also note that, in some cases (as highlighted by the elliptic area in Figure 1B) the resulting spots seem to be dilated. Application of only the coring stage does not produce such an effect, as highlighted in Figure 1D. Comparison of Figures 1B and 1D tends to indicate that the correlation stage applied on its own has a better denoising effect for low-intensity microarray spots.

Figures 2 and 3 show a quantitative performance measure, namely the coefficient of variance (CV), that demonstrates the effect of the two stages in the homogeneity of both the background and the microarray spot areas. More specifically, the figures provide boxplots of (i) the signal CV (the lower this metric, the higher the signal homogeneity) and (ii) the background CV (the lower this metric, the higher the background homogeneity) and a plot which presents the percentage of spots whose CV falls into a confidence interval (CI). This CI plot indicates that the proposed method tends to assign a CV close to zero to the major percentage of spots. Figure 3 is the same boxplot as Figure 2 but it accounts only for the low-intensity spots. It illustrates that when we apply first the coring and then the correlation stage we get more low-intensity spots with CV close to zero which implies more homogeneous spot areas corresponding to low-intensity spots.

Table 1 presents the amount of spots that SpotSegmentation is unable to detect. From this table, it appears that the coring stage finds most spots than the proposed approach. But when looking at the images provided by the software package (data not shown), we can discern that SpotSegmentation identifies "spots" in areas where no spot exists. The percentage of 8.3% consists of really detected spots along with a number of false positives. From the images of SpotSegmentation and from Table 1, it is obvious that the proposed method increases the spot detectability of the original image by approximately 40%. This result demonstrates the significance of the proposed two-stage approach.

	Original	Coring	Correlation	Coring - Correlation	Correlation - Coring
Spots NOT detected	42.21%	8.3%	21.2%	18.8%	20.8%

Table 1. Results from SpotSegmentation

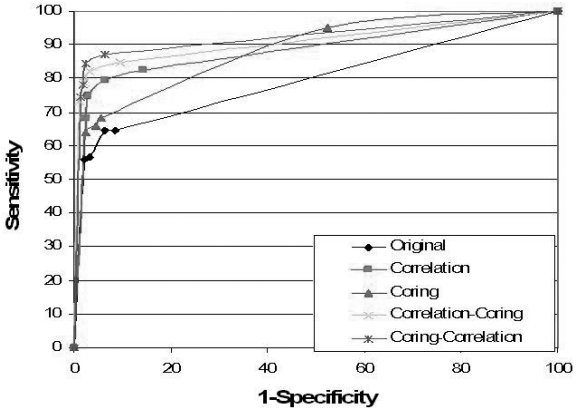

Figure 4. ROC curves for all methods. Threshold1: 0<region<∞, 0<overlap<∞, Threshold2: region = 1, 0<overlap<∞, Threshold3: region = 1, overlap>70%

Figure 4 illustrates the Receiver Operating Characteristic (ROC) curves of the results from SpotSegmentation, which plot the hit rate as a function of the false alarm rate. We used a combination of objective criteria (image enhancement, histogram equalization, our method and others), with the help of the experimentalist (molecular biologist), to determine if any spot signal was present or not. The criteria used for the classification of spots as true positive, false positive, true negative, false negative consist of the number of regions within the mask that is defined as the area of the ex-

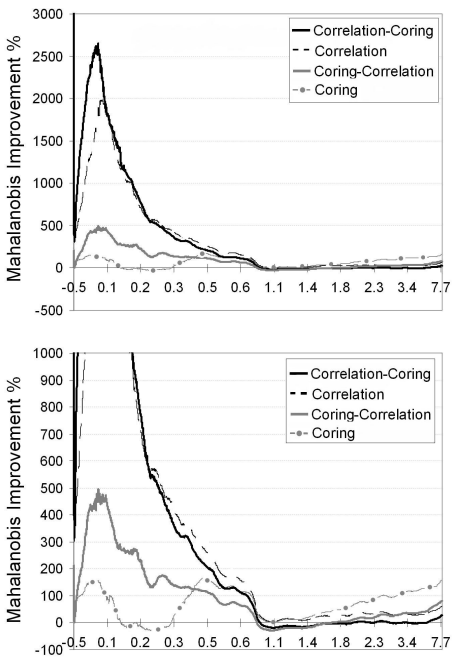

Figure 5. Improvement of the Mahalanobis distance of spots and background between the original and the processed images as a function of the spot-to-background intensity ratio.

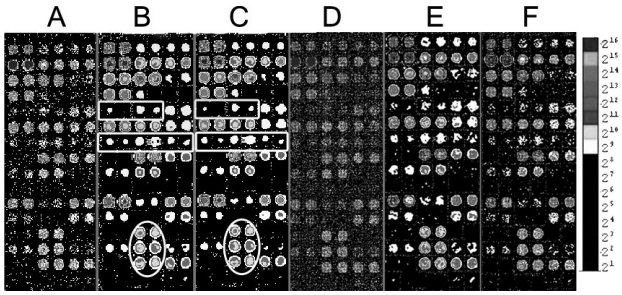

Figure 6. Results of the proposed two-stage approach in a low-resolution image. A is the original low-resolution image and B-C are the images processed with the correlation stage and then with the coring stage and the coring stage and then with the correlation stage, respectively. D is the original high-resolution image and E-F are the images processed as in B-C. Box areas show low-intensity spots that have vanished in low-resolution processed images. Elliptic areas show processed spots that have been dilated while do not when processing by the coring stage and then the correlation in the high-resolution image (Figure E).

pected ideal spot and the overlap of the area of the detected spots and this ideal mask. The area under the ROC curve is a measure of how good the methods are in discriminating between detectable and not detectable spots. The larger this area is the better the method. From all these, it can be concluded that coring followed by correlation demonstrates a great ability in detecting the spots.

Figure 5 illustrates another quantitative performance metric, namely, the Mahalanobis distance improvement of spot and background areas between the processed and original images as a function of the spot-to-background intensity ratio. This Figure contains an approximation of the individual values. While Figure 3 illustrates that the application of the correlation stage followed by the coring stage tends to make the low-intensity spots more homogeneous than they were, Figure 5 shows that these spots become more distinctive from the local background. On the other hand, when combining the two stages vice versa we get a better discrimination for the high-intensity spots but not as good for the low-intensity ones (cf. Figure 5). It is obvious that both orders of the methods application have complementary advantages, therefore we cannot select one of the combinations as the best.

When decreasing the resolution of a microarray image by applying a smoothing mask on it, the results of the proposed method appear to be not that good. In Figure 6, we present the results from the smoothed version of image in Figure 1A. It is obvious (as highlighted by the box areas in Figures 6B and 6C) that some low-intensity spots that were enhanced in the high-resolution image processing (Figures 6E and 6F) vanish in the processing of the low-resolution image. Therefore, low-resolution images, such as smoothed images and images that come from the lowest detection settings of a microarray image production, are unwanted in our method.

4. Conclusions

Microarray images consist mostly of low-intensity spots that are not well distinguishable from the background. These low-intensity spots are affected by inherent additive and multiplicative noise components. Our results suggest that in high throughput whole-genome approaches, applying a two-stage "correlation and coring" approach enhances the dynamic range of existing microarray imaging technology, which is very important in order to identify the most significant genes with increased accuracy and robustness. Our method disencumbers researchers of producing multiple images of the same microarray at different detection set-

tings because it yields better results for the image obtained at high detection settings. Although the proposed method is tested on the 2-color DNA microarray technology, it is performed on the images separately. Consequently, we argue that it is applicable to the single-color microarrays, as well.

References

[1] A. Daskalakis *et al*. Improving gene quantification by adjustable spot-image restoration. *Bioinformatics*, June 28 2007.

[2] J. O. Bishop, J. G. Morton, M. Rosbash, and M. Richardson. Three abundance classes in HeLa cell messenger RNA. *Nature*, 250(463):199–204, 1974.

[3] Y. Chen, E. Dougherty, and M. Bittner. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed.Opt*, 2:364–374.

[4] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[5] R. O. Dror, J. G. Murnick, N. J. Rinaldi, V. D. Marinescu, R. M. Rifkin, and R. A. Young. Bayesian estimation of transcript levels using a general model of array measurement noise. *J. Comp.Biology*, 10(3/4):433–452, 2003.

[6] A. M. Dudley, J. Aach, M. A. Steffen, and G. M. Church. Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc.Nat.Acad.Sci. U.S.A.*, 99:7554–7559, 2002.

[7] J. Eberwine, H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel, and P. Coleman. Analysis of gene expression in single live neurons. *Proc.Nat.Acad.Sci. U.S.A,*, 89(7):3010–3014, 1992.

[8] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc.Nat.Acad.Sci. U.S.A,*, 95:14863–14868, 1998.

[9] T. R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc.Nat.Acad.Sci. U.S.A,*, 98:15149–15154, 2001.

[10] A. J. Hartemink, D. K. Gifford, T. S. Jaakola, and R. A. Young. Maximum likelihood estimation of optimal scaling factors for expression array normalization. *SPIE BiOS*, 4266.

[11] M. Holschneider, R. Kronland-Martinet, J. Morlet, and A. Grossmann. A real-time algorithm for signal analysis with the help of the wavelet transform. In J. M. Combes, A. Grossmann, and P. Tchamitchian, editors, *Wavelets, Time-Frequency Methods and Phase Space*, pages 286–297. Springer-Verlag, 1989.

[12] M. Jansen and A. Bultheel. Multiple wavelet threshold estimation by generalized cross validation for images with correlated noise. *IEEE Trans. Image Proc*, 1998.

[13] J. J. Kang, R. M. Watson, M. E. Fisher, R. Higuchi, D. H. Gelfand, and M. J. Holland. Transcript quantitation in total yeast cellular RNA using kinetic PCR. *Nucleic Acids Res*, 28(2):e2, 2000.

[14] M. T. Lee, F. C. Kuo, G. A. Whitmore, and J. Sklar. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Nat. Acad. Sci., U S A,*, 97(9):9834–9, 2000.

[15] Q. Li, C. Fraley, R. E. Bumgarner, K. Y. Yeung, and A. E. Raftery. Donuts, scratches and blanks: robust model-based segmentation of microarray images. *Bioinformatics*, 21(12):2875–2882, 2005.

[16] R. Lukac, K. N. Plataniotis, B. Smolka, and A. N. Venetsanopoulos. cDNA microarray image processing using fuzzy vector filtering framework. *Journal of Fuzzy Sets and Systems: Special Issue on Fuzzy Sets and Systems in Bioinformatics*, 152(1):17–35, 2005.

[17] R. Lukac and B. Smolka. Application of the adaptive center-weighted vector median framework for the enhancement of cDNA microarray. *Int. J. Appl. Math. Comput. Sci.*, 13:369–383, 2003.

[18] M. Mastriani and A. E. Giraldez. Microarrays denoising via smoothing of coefficients in wavelet domain. *International Journal of Biomedical Sciences*, 1:1306–1216, 2006.

[19] A. Mastrogianni, E. Dermatas, and A. Bezerianos. Robust pre-processing and noise reduction in microarray images. *Biomedical Engineering*, 2007.

[20] O. Ermolaeva *et al*. Data management and analysis for gene expression arrays. *Nat.Genet*, 20(1):19–23, 1998.

[21] J.-C. Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989–1996, 2002.

[22] A. C. Pease, D. Solas, E. J. Sullivan, M. T. Cronin, C. P. Holmes, and S. P. Fodor. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Nat. Acad. Sci., U S A,*, 91:5022–5026, 1994.

[23] R Foundation for Statistical Computing. *R: A language and environment for statistical computing*, 2005.

[24] D. Rocke and B. Durbin. A model for measurement error for gene expression arrays. *J. Comp.Biology*, 8(6):557–569, 2001.

[25] B. M. Sadler and A. Swami. Analysis of multiscale products for step detection and estimation. *IEEE Trans. Information Theory*, 45(3):1043–1051, 1999.

[26] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science, New York, N.Y 270, 1995.

[27] E. P. Simoncelli and E. H. Adelson. Noise removal via Bayesian wavelet coring. *IEEE Proc Third Int Conference Image Processing*, I:379–382, 1996.

[28] D. S. Skibbe, X. Wang, Z. X., L. A. Borsuk, D. Nettleton, and P. S. Schnable. Scanning microarrays at multiple intensities enhances discovery of differentially expressed genes. *Bioinformatics*, 22(15):1863–1870, 2006.

[29] D. Slonim. From patterns to pathways: gene expression data analysis comes of age. *Nat.Genet*, 32 Suppl.:502–508, 2002.

[30] T.R. Hughes *et al*. Functional discovery via a compendium of expression profiles. *Cell*, 102(1):109–126, 2000.

[31] L. J. van't Veer *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415:530–536, 2002.

[32] X. H. Wang, R. S. H. Istepanian, and Y. H. Song. Microarray image enhancement by denoising using stationary wavelet transform. *IEEE Trans Nanobioscience*, 2(4):184 – 189, 2003.

[33] Y. Yang, M. Buckley, S. Dudoit, and T. Speed. Comparison of methods for image analysis on cDNA microarray data. *J. Comp. Graph. Stat.*, 11:108–136, 2000.