

WIENER-BASED DECONVOLUTION METHODS FOR IMPROVING THE ACCURACY OF SPOT SEGMENTATION IN MICROARRAY IMAGES

A. Daskalakis^{*}, C. Argyropoulos^{*}, D. Glotsos^{*}, S. Kostopoulos^{*}, E. Athanasiadis^{*}, D. Cavouras^{**}
and G. Nikoforidis^{*}

^{*}Medical Image Processing and Analysis Group, Laboratory of Medical Physics, University of Patras, 26500 Patras, Greece

^{**}Medical Image and Signal Processing Laboratory, Department of Medical Instruments Technology, Technological Educational Institute of Athens.

daskalakis@med.upatras.gr

Purpose: Microarray experiments are important tools for high throughput gene quantification. Nevertheless, such experiments are confounded by a number of technical factors, which operate at the fabrication, target labelling, and hybridization stages, and result in spatially inhomogeneous noise. Unless these sources of error are addressed, they will propagate throughout the stages of the analysis, leading to inaccurate biological inferences. The aim of this study was to investigate whether image restoration techniques may improve the accuracy of subsequent microarray image analysis steps (i.e. segmentation and gene quantification).

Materials and Methods: A public dataset of seven microarrays obtained from the MicroArray Genome Imaging & Clustering Tool (MAGIC) database were used. Each image contained 6400 spots investigating the diauxic shift of *Saccharomyces cerevisiae*. Restoration was based on the Wiener deconvolution. Subsequently, restored images were processed with the MAGIC tool for semi-automatic gridding and segmentation. The influence of the restoration process on the accuracy of spot segmentation was quantitatively assessed by the information theoretic metric of the Kullback-Liebler divergence.

Results: Pre-processing based on Wiener deconvolution increased the range of divergence (0.04 – 3.01 bits) and consequently improved the accuracy of subsequent spot segmentation.

Conclusion: Information theoretic metrics confirmed the importance of image restoration as a pre-processing step that significantly improved the accuracy of subsequent segmentation, thus leading to more accurate gene quantification.

Introduction

Complementary DNA (cDNA) microarray imaging is considered as an important tool for large-scale gene sequence and gene expression analysis [1, 2]. Molecular biologists and bioinformaticians are using microarray technology not only for identifying a gene in a biological sequence but also for predicting the function

of the identified gene within a larger system, such as the human organism [3].

The basic microarray experimental procedure involves hybridization of complementary nucleic acid molecules, one of which (target) has been immobilized in a solid substrate (e.g. glass) using a robotically controlled device (arrayer). The two main techniques for printing targets are metallic pin and inkjet based systems, which lead to the formation of circular spots of known diameter and cDNA target. Those *spots* are located at the vertices of a rectangular lattice on the solid substrate surface. Each one of them serves as a highly specific and sensitive detector of the corresponding gene [4]. In order to create a genome expression profile of a biological system with microarrays, the messenger RNA from a particular sample is isolated, labelled and hybridized on the microarray. Although labelling and detection of hybridized probes is performed using various protocols i.e. P^{32} , chromogenic systems (i.e. digoxigenin, antidigoxigenin etc [5], fluorescent dyes (e.g. Cy3, Cy5) can be characterized as the most popular. After labelling and hybridization the microarrays are “read”, using methods contingent upon the nature of the labelling reaction i.e. PhosphoImager plates, confocal laser scanners, and Charged Couple Devices [6].

The data output of the microarray experiment are two 16-bit tagged image files, one for each fluorescent dye (Cy3, Cy5). By isolating the spots for each channel via image segmentation, and by analyzing the pixel intensities of each segmented spot, it is possible to accurately quantify gene expression. These three crucial steps, experiment, image processing and gene quantification characterize the microarray analysis pipeline.

Gene quantification is, nevertheless, confounded by a number of technical factors, which operate at the fabrication, target labelling, and hybridization stages, and result, in the microarray output images, not only as spatially inhomogeneous noise but also as irregularities of spot shape, size, and position.[7, 8]. Additive degradation caused by the confocal laser scanner, used as “reading” method, is furthermore complicating gene quantification. Unless these sources of error and

degradation are addressed, they will propagate throughout the stages of the analysis leading to inaccurate biological expression.

In spite of the potential importance of image pre-processing in correcting these error sources, existing software tools [9-12] focus mainly on spot localization and microarray image segmentation. Analysis of microarray images is thereby separated into three sequential steps, namely, gridding, segmentation and intensity extraction [13-16]. Only few studies have examined the impact of image pre-processing upon the steps of spot detection and segmentation, which are crucial intermediaries in the microarray pipeline [13, 17].

The aim of this study was to investigate whether image deconvolution techniques may improve the accuracy of subsequent microarray image analysis steps (i.e. segmentation and gene quantification). Consequently, this study explores the performance of Wiener deconvolution on a public available dataset [18]. In addition, to objectively quantify whether deconvolution improved segmentation, information theoretic measures were applied to the distributions of the signal and the background intensity values.

Methods and Material

The microarray images used for this study comprised a publicly available dataset of seven 16-bit Tiff images obtained from the MicroArray Genome Imaging & Clustering Tool (MAGIC) website [19]. Each image contained 6400 spots investigating the diauxic shift of *Saccharomyces cerevisiae*. The particular dataset was selected because the original authors [18] used a common reference messenger RNA pool to control for biological variability [20-22]. This particular design affords an adequate degree of replication required for the quantitative statistical assessment of the effects of pre-processing on the image segmentation and subsequent gene profiling [20].

Microarray images were restored by implementing the Wiener deconvolution algorithm [23] with Matlab source code [24]. The Wiener deconvolution algorithm requires prior knowledge, both, of the Point Spread Function (PSF) of the imaging apparatus and the inherent noise of the image. Since we did not have access to the confocal microscopy scanner used by the original authors [18], we modelled the PSF by a spatially invariant Gaussian PSF [25, 26] for two different choices of the Full Width Half Maximum (FWHM) parameter. Additionally, noise estimation was done by calculating the standard deviation of sampled regions in the background surrounding of the spots. The estimated noise value was employed in the Wiener deconvolution filter. Subsequently, restored images were processed with the MAGIC tool for semi-automatic gridding and segmentation.

Gridding and segmentation procedures were performed on both original and restored microarray images using two separate, well known algorithms in

microarray spot segmentation namely Fixed Circle (FC) and Seeded Region Growing (SRG) [17, 27, 28]. The FC method is based in segmenting each spot by using constant diameter and is implemented in most of the common commercial software packages [29-31]. The SRG segmentation method requires the specification of starting points, or seeds for the spots. Both are proved to be very effective in the accurate segmentation of microarray spots. Accurate segmentation provides a better estimation of the signal (spot) and background intensities, which will be further used for gene quantification.

After segmentation, foreground (spot), and background intensity values for the common reference channel (green, Cy-3) were extracted for further processing. Histogram evaluation and visualization for the distribution of signal and background intensity values was then performed with Matlab. Furthermore, the influence of the deconvolution process to the accuracy of spot segmentation was quantitatively assessed by the information theoretic metric of the Kullback-Liebler divergence [32]. Images with high value of divergence, which corresponds to well separable distributions of signal and background, are more likely to give accurate segmentation results and, thus, are preferable.

Results

Results of the Wiener deconvolution procedure are shown in Fig.1. One may observe that the image restored using a FWHM of 0.3 pixels (Fig.1b) is virtually indistinguishable from the original (Fig.1a); setting the FWHM parameter to 0.5 pixels, results in a sharpened image with higher contrast (Fig.1c).

Table 1 and Table 2 tabulate the values of the Kullback-Liebler divergence between spot and background log intensity distributions for the FC and the SRG segmentation procedures respectively.

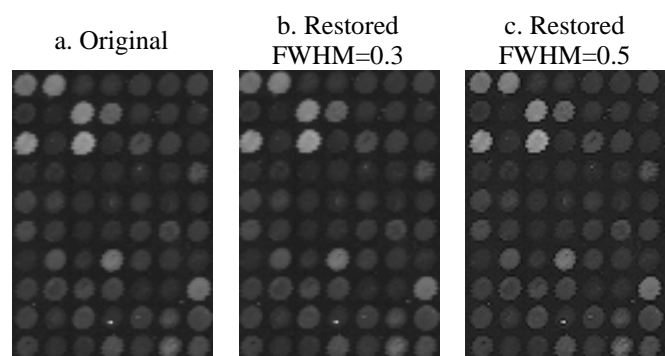


Fig.1. Original and Wiener restored sections of microarray images for two different choices of Gaussian PSF kernel

In all arrays evaluated, pre-processing based on the Wiener deconvolution increased the range of the divergence metric both for the FC and SRG segmentation methods e.g. for the image coded as

“1313_OD730” in the Magic database [19], the increase in the Kullback-Liebler value was 2.97 nits.

Table 1.2. Kullback-Liebler divergence metric between spot (signal) and background intensity (log-space) values for the green (common reference sample) channel for the seven arrays used in the DeRisi publication. The results correspond to the FC and SRG segmentation methods.

FC Segmentation Method			
Images	Original	Wiener 0.3	Wiener 0.5
1302_OD370	1.337	1.374	1.205
1303_OD014	0.825	0.834	0.189
1309_OD690	1.457	1.792	1.937
1310_OD046	0.795	1.324	0.726
1311_OD080	0.574	0.608	0.772
1312_OD180	0.866	0.926	0.982
1313_OD730	1.499	1.569	0.843
SRG Segmentation Method			
Images	Original	Wiener 0.3	Wiener 0.5
1302_OD370	3.743	3.502	4.059
1303_OD014	1.298	1.241	2.222
1309_OD690	1.676	2.278	2.424
1310_OD046	1.773	2.874	1.613
1311_OD080	1.474	1.647	2.725
1312_OD180	4.832	3.999	3.806
1313_OD730	2.211	5.251	3.617

Discussion-Conclusion

Microarray technologies have transformed the field of genomic research by allowing the simultaneous profiling of thousands of genes. The microarray process is based entirely on the extraction of quantitative information from images. Despite the importance of image pre-processing steps in other fields that rely on quantification of image features (e.g. medical and astronomical imaging) there have been very few attempts to apply similar techniques to microarray data analysis. Previous work has focused mainly on microarray image de-noising by stationary wavelet transforms [33] and adaptive fuzzy filters [34]. Both methods were shown to be capable of removing noise while preserving structural information provided by the spot size, shape etc. The resulting images were judged to be of higher quality than the originals by either direct visual inspection or comparison with subjectively validated quality numerical indices [35].

In the present work, the Wiener deconvolution filtering algorithm was applied to the problem of microarray image restoration aiming to improve either spot segmentation or gene profiling. To objectively quantify the beneficial impact (if any) of the deconvolution procedure, information theoretic metric of Kullback-Liebler divergence were evaluated after the segmentation procedure for both the original and the enhanced images. Results confirmed the importance of

image restoration as a pre-processing step that significantly improved the accuracy of subsequent segmentation, thus leading to more accurate gene quantification.

References

- [1.] M. Taniguchi, K. Miura, H. Iwao, and S. Yamanaka, Quantitative assessment of DNA microarrays--comparison with Northern blot analyses, *Genomics*.71 (2001) (1) 34-9
- [2.] A. Alizadeh, M. Eisen, D. Botstein, P.O. Brown, and L.M. Staudt, Probing lymphocyte biology by genomic-scale gene expression analysis, *J Clin Immunol*.18 (1998) (6) 373-9
- [3.] M. Chee, R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart, M.S. Morris, and S.P. Fodor, Accessing genetic information with high-density DNA arrays, *Science*.274 (1996) (5287) 610-4
- [4.] E. Southern, K. Mir, and M. Shchepinov, Molecular interactions on microarrays, *Nat Genet*.21 (1999) (1 Suppl) 5-9
- [5.] Mark Schena, *Microarray Biochip Technology*. 1st ed: Eaton Publishing Company. 2000
- [6.] Worley, A systems approach to Fabricating and Analyzing DNA microarrays, in *Microarray Biochip Technology*, M. Schena, Editor, Eaton Publishing. (2000)
- [7.] Y. Balagurunathan, E.R. Dougherty, Y. Chen, M.L. Bittner, and J.M. Trent, Simulation of cDNA microarrays via a parameterized random signal model, *J Biomed Opt*.7 (2002) (3) 507-23
- [8.] Y. Balagurunathan, N. Wang, E.R. Dougherty, D. Nguyen, Y. Chen, M.L. Bittner, J. Trent, and R. Carroll, Noise factor analysis for cDNA microarrays, *J Biomed Opt*.9 (2004) (4) 663-78
- [9.] M. Steinfath, W. Wruck, H. Seidel, H. Lehrach, U. Radelof, and J. O'Brien, Automated image analysis for array hybridization experiments, *Bioinformatics*.17 (2001) (7) 634-41
- [10.] M.A. Zapala, D.J. Lockhart, D.G. Pankratz, A.J. Garcia, C. Barlow, and D.J. Lockhart, Software and methods for oligonucleotide and cDNA array data analysis, *Genome Biol*.3 (2002) (6) SOFTWARE0001,1-0001,9
- [11.] A.M. White, D.S. Daly, A.R. Willse, M. Protic, and D.P. Chandler, Automated Microarray Image Analysis Toolbox for MATLAB, *Bioinformatics*.21 (2005) (17) 3578-9
- [12.] M.B. Eisen. ScanAlyze. Available from: <http://rana.stanford.edu/software>.
- [13.] V. Barra, Robust segmentation and analysis of DNA microarray spots using an adaptive split and merge algorithm, *Comput Methods Programs Biomed*.81 (2006) (2) 174-80
- [14.] J. Angulo and J. Serra, Automatic analysis of DNA microarray images using mathematical morphology, *Bioinformatics*.19 (2003) (5) 553-62

- [15.] A.N. Jain, T.A. Tokuyasu, A.M. Snijders, R. Segraves, D.G. Albertson, and D. Pinkel, Fully automatic quantification of microarray image data, *Genome Res.*12 (2002) (2) 325-32
- [16.] Y.H. Yang, M.J. Buckley, and T.P. Speed, Analysis of cDNA microarray images, *Brief Bioinform.*2 (2001) (4) 341-9
- [17.] L. Qin, L. Rueda, A. Ali, and A. Ngom, Spot detection and image segmentation in DNA microarray data, *Appl Bioinformatics.*4 (2005) (1) 1-11
- [18.] J.L. DeRisi, V.R. Iyer, and P.O. Brown, Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science.*278 (1997) (5338) 680-6
- [19.] <http://www.bio.davidson.edu/projects/MAGIC/MAGIC.html>.
- [20.] E. Sterrenburg, R. Turk, J.M. Boer, G.B. van Ommen, and J.T. den Dunnen, A common reference for cDNA microarray hybridizations, *Nucleic Acids Res.*30 (2002) (21) e116
- [21.] Y.H. Yang and T. Speed, Design issues for cDNA microarray experiments, *Nat Rev Genet.*3 (2002) (8) 579-88
- [22.] G.A. Churchill, Fundamentals of experimental design for cDNA microarrays, *Nat Genet.*32 Suppl (2002) 490-5
- [23.] R.C. Gonzalez and R.E. Woods, *Digital Image Processing* ed. Addison-Wesley. 1992
- [24.] <http://www.mathworks.com/>.
- [25.] Y. Chen, J.D. Muller, P.T. So, and E. Gratton, The photon counting histogram in fluorescence fluctuation spectroscopy, *Biophys J.*77 (1999) (1) 553-67
- [26.] G.M.P. van Kempen, L.J. van Vliet, and P.J. Verveer. Application of image restoration methods for confocal fluorescence microscopy, in *3-D Microscopy: Image Acquisition and Processing IV*, Proc. SPIE, (1997), 114-124
- [27.] 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 (2005) (12) 2875-82
- [28.] A.A. Ahmed, M. Vias, N.G. Iyer, C. Caldas, and J.D. Brenton, Microarray segmentation methods significantly influence data precision, *Nucleic Acids Res.*32 (2004) (5) e50
- [29.] QuantArray Analysis Software. 1999
- [30.] GenePix4000A User's Guide, 1999)
- [31.] M.B.Eisen. ScanAlyze. Available from: <http://rana.stanford.edu/software>.
- [32.] S. Kullback, *Information Theory and Statistics*. 2nd ed: Dover Publications. 1968
- [33.] X.H. Wang, R.S. Istepanian, and Y.H. Song, Microarray image enhancement by denoising using stationary wavelet transform, *IEEE Trans Nanobioscience.*2 (2003) (4) 184-9
- [34.] R. Lukac, P. K.N., S. B., 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.*2005)
- [35.] Z. Wang and H. Bovic, A universal image quality index, *IEEE Signal Processing Letters.*9 (2002) (3) 81-84