2nd International Conference "From Scientific Computing to Computational Engineering"
2nd IC-SCCE
Athens, 5-8 July, 2006

SEGMENTATION OF MICROARRAY IMAGES USING GAUSSIAN MIXTURE MODELS AND WAVELET BASED PREPROCESSING FILTERS

Emmanouil Athanasiadis [1], Panagiota Spyridonos [1], Antonis Daskalakis [1], Dimitris Glotsos [1], Spyros Kostopoulos [1], Ioannis Kalatzis [2], Dionisis Cavouras [2] and George Nikiforidis [1]

[1] Medical Image Processing and Analysis Group, Laboratory of Medical Physics,
University of Patras, 26500 Patras, Greece.
e-mail: mathan@upatras.gr , web page: http://mipa.med.upatras.gr

[2] Medical Image and Signal Processing Laboratory, Department of Medical Instruments Technology,
Technological Institute of Athens, Greece.
e-mail: cavouras@teiath.gr , web page: http://medisp.bme.teiath.gr

Keywords: Microarray, wavelet, Gaussian Mixture Models.

Abstract. *The objective of this work was to investigate the segmentation capability of Gaussian Mixture Models (GMMs) on microarray images, using wavelet based image enhancement techniques. A Simulated Microarray image of 200 spots was produced using a Microarray Scan Simulator. The image was pre-processed using 3 different daub4 wavelet based filters, developed in MATLAB. The detail coefficients of the wavelet transform were processed up to scale 3 using Hard, Soft and Sigmoidal Thresholding. An automatic gridding process was developed and applied on the preprocessed microarray image with the purpose of identifying the spots. The GMMs algorithm was then applied to each spot in order to discriminate foreground from background. The segmentation capability of the GMMs was evaluated by calculating the segmentation matching factor for each spot. Optimal segmentation results were obtained by pre-processing the microarray image by all the wavelet-based filters. Wavelet based preprocessing was found to improve the GMMs segmentability, in discriminating the spot's foreground from background.*

1 INTRODUCTION

Microarray imaging is used for the identification of thousands of genes simultaneously in the field of bioinformatics [1]. Gene identification is closely associated with spot identification. By locating the spot in a complementary DNA (cDNA) microarray, measurements of the mean fluorescence intensity value are obtained. These measurements are related to the abundance of the messenger RNA (mRNA).

For the task of measuring spot intensity values, three major steps are taken [1,2]: First, the gridding step, where the exact location of each spot is located. Second, the segmentation step, in which foreground and background recognition of each spots is accomplished, and, finally, the intensity extraction step, where the mean fluorescence value is calculated for each spot.

In the past, several algorithms and software packets have been developed for the task of processing of microarray images [3, 4, 5, 6, 7]. In the ScanAlyze [3] packet, a *fixed circle segmentation* method is used, where all spots are considered circular with a fixed predefined radius. In the GenePix [4] packet, an *adaptive circle segmentation* technique is employed. In that method, the radius of each spot is suitably. In the Spot [6] packet, an *adaptive shape segmentation* technique is followed. The most representative algorithms employed for that technique are watershed [8] and seeded region growing [9]. In the ImaGene [7] packet, a *histogram based segmentation* method is applied, in which the 80th and 95th percentiles contribute to the calculation of the mean intensity value. In all those techniques, the major disadvantages are either that spots are not circular, or a-priori knowledge of the precise position of centers is not accurately defined [10].

In the present study, a Gaussian Mixture Model clustering algorithm [11] was developed in MATLAB$^{@}$. Microarray images were firstly pre-processed using three different types of wavelet based noise suppression filters. Evaluation of the results was achieved by calculating the segmentation matching factor.

2 MATERIALS AND METHODS

2.1 Material

On complementary DNA (cDNA) Microarray experiments [12], two messenger RNA (mRNA) samples are first reverse transcripted into cDNA. Next, the two samples are labeled using two different fluorescence dyers, Cyanine Cy3 (red channel) and Cy5 (green channel) respectively. The two samples are hybridized [1] and after the scanning procedure, two colored fluorescence Tagged Image File Format (TIFF) images are produced for each channel [1]. The whole procedure is illustrated in Figure 1.

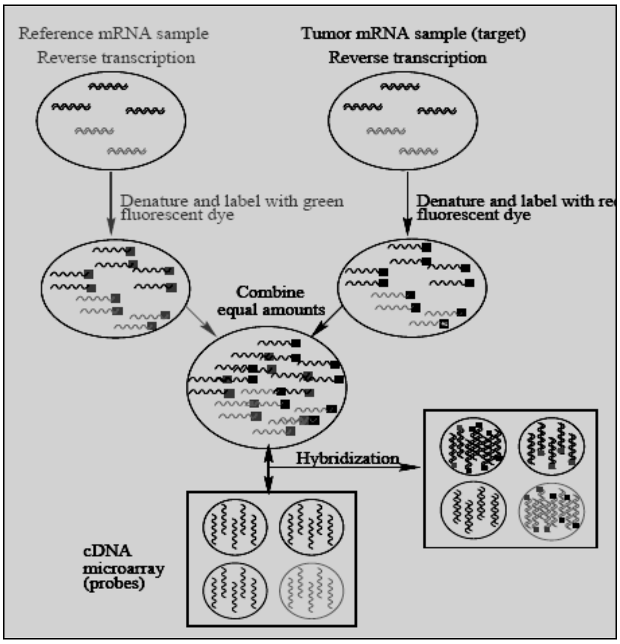


Figure 1. Hybridization of cDNA Microarray images. [13]

The fluorescence intensity value of each spot is related to the expression abundance of the corresponding DNA sequence. In the present study, a microarray image of 200 spots (16-bit colored TIFF image) was created using the *Microarray Scan Simulator* [14].

2.3 Wavelet-based filters

The wavelet-based enhancement involves three steps [15, 16, 17, 18, 19]. In the first step, the DAUB4 DWT (Discrete Wavelet Transform) [18] was applied in two scales. In the second step, the detail coefficients of the DWT were processed in both scales, using in each occasion one of the three different enhancement functions illustrated in sections 2.3.1 to 2.3.3. In the final step, the processed mammograms were reconstructed using the Inverse DWT (IDWT) [18].

2.3.1 Soft -Threshold Function (STF).

In the STF [20, 21, 22 and 23], the multi-scale coefficients of the DWT were modified according to equation 3.

$$W_{out} = \begin{Bmatrix} W_{in} + T \cdot (G-1) \text{ if } W_{in} > T \\ W_{in} - T \cdot (G-1) \text{ if } W_{in} < -T \\ G \cdot W_{in} \qquad \text{otherwise} \end{Bmatrix} \tag{3}$$

were *Wout* denotes the output and *Win* the input coefficient values of the Detail Matrix. *T* and *G* are threshold and gain values respectively.

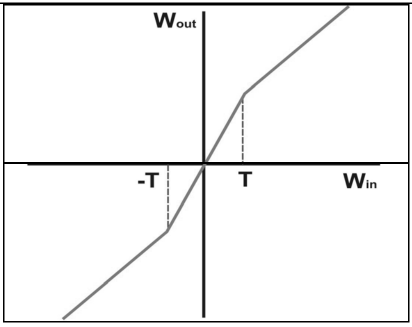


Figure 2: Simple Piece-Wise Linear Mapping Function

2.3.2 Hard-Threshold Function (HTF).

HTF is illustrated in equation 4 and figure 3 respectively.

$$W_{out} = \begin{Bmatrix} W_{in} + T \cdot (G-1) \text{ if } W_{in} > T \\ W_{in} - T \cdot (G-1) \text{ if } W_{in} < -T \\ 0 \quad \text{otherwise} \end{Bmatrix} \tag{4}$$

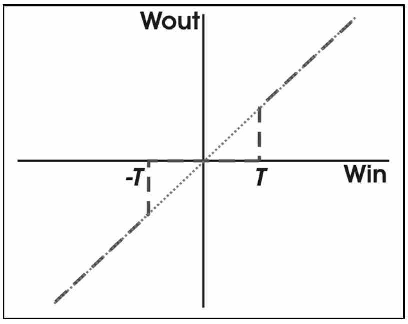


Figure 3: Hard-Threshold Function unction

2.3.3 Sigmoidal non-Linear Enhancement (SNLEF) Function.

In the SNLEF filter, wavelet coefficients in between the threshold value $|T|$ were modified according to a sigmoid function illustrated in equation 5 and figure 4.

$$W_{out} = a\left[sigm\left(G\left(W_{in} - T\right)\right) - sigm\left(-G\left(W_{in} + T\right)\right)\right]$$

$$a = \frac{1}{sigm(G(1-T)) - sigm(-G(1+T))} \tag{4}$$

$$sigm(y) = \frac{1}{1+e^{-y}}$$

W_{out} denotes the output coefficients, W_{in} are the input coefficients and parameters T and G control the threshold and the gain respectively.

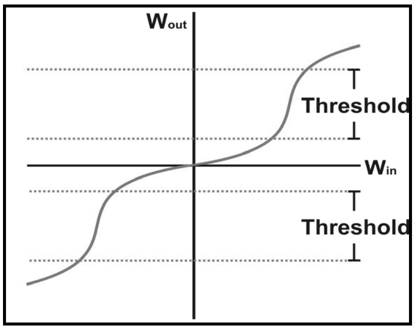


Figure 4: Sigmoidal non-Linear Enhancement Function

2.4 Evaluation.

The accuracy of segmentation was numerically calculated by using the segmentation matching factor (Equation 5) [24].

$$accuracy = \frac{A_{calc} \cap A_{real}}{A_{calc} \cup A_{real}} \times 100 \qquad \text{Equation 5}$$

where A_{calc} is the area of the spot as determined by the proposed algorithm and A_{real} is the actual spot area. A perfect match is indicated by a 100% score, any score higher than 50% indicates reasonable segmentation [24] while a score of less than 50% indicates poor segmentation [24].

3 RESULTS AND DISCUSSION

According to our results, the wavelet-based filters were found to assist the GMM's accuracy by delineating spots sharper. Significant results of the Red (R) channel are illustrated in Table 1.

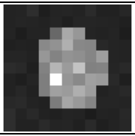
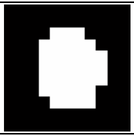
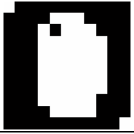
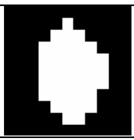
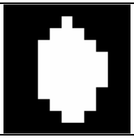
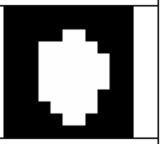
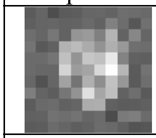
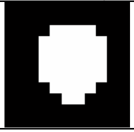
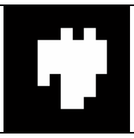
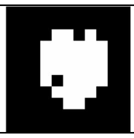
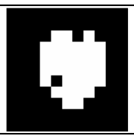
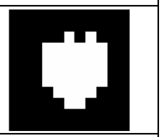
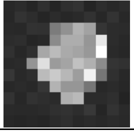
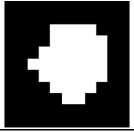
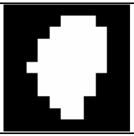
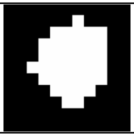
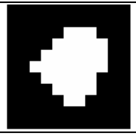
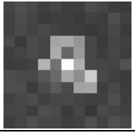
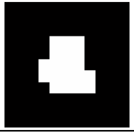
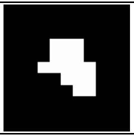
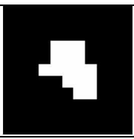
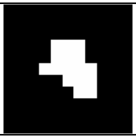
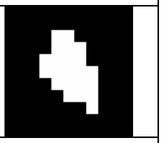
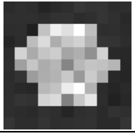
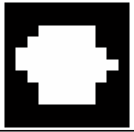
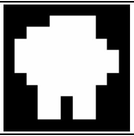
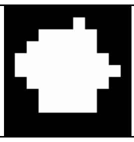
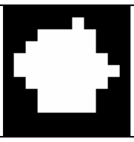
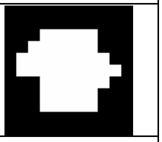
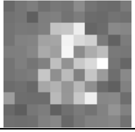
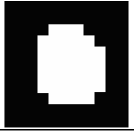
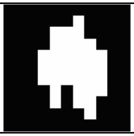
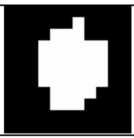
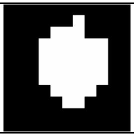
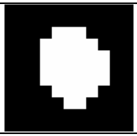
Original Cells	Actual Boundaries	Without preprocess	Preprocessed with STF	Preprocessed with HTF	Preprocessed with SNLEF
					
spot 1	Accuracy:	87.6033	97.5207	97.5207	98.3471
					
spot 2	Accuracy:	95.8678	98.3471	98.3471	98.3471
					
spot 3	Accuracy:	91.7355	97.5207	97.5207	98.3471
					
spot 4	Accuracy:	95.8678	95.8678	95.8678	94.2149
					
spot 5	Accuracy:	89.2562	99.1736	99.1736	99.1736
					
spot 6	Accuracy:	95.0413	96.6942	95.8678	97.5207

Table 1: Comparative results for 6 different spots for the R channel. The first column indicates the simulated spot, the second column indicates the actual boundaries of the spot and the third, forth, and fifth columns present the results of the GMM with no wavelet-based preprocessing, GMM preprocessed with STF, GMM preprocessed with HTF, GMM preprocessed with SNLEF, as well as, the corresponding matching factors.

It is clear that in cases where the foreground intensity is close to the background, wavelet preprocessing was found to boost the GMM algorithm. For instance, the accuracy achieved by the GMM with no preprocessing in spot 1, in Table 1, was 87.6033%. However, by preprocessing the image by the SNLEF, the accuracy of the GMM was 98,347%. Similar results were obtained by using the STF and HTF on the same spot

(97.52%).

4 CONCLUSIONS

Wavelet-based filters performance was found to enhance microarray images, by increasing the ability of GMM to discriminate foreground from background. Processing time using the wavelet-based filter was less than 3 seconds for 512x512 sample images. The addressing stage, as well as the segmentation process of the 200 spots was less than 10 seconds.

5 ACKNOWLEDGEMENTS

We would like to thank the Greek State Scholarships Foundation (I.K.Y.) for funding the above work.

6 REFERENCES

[1] Y.H. Yang, M. J. Buckley, S. Duboit and T.P.Speed (2002), "Comparison of methods for Image Analysis on cDNA Microarray Data", Journal of Computational and Graphical Statistics, vol. 11, pp 108-136.

[2] Y.H. Yang, M. J. Buckley, S. Duboit and T.P.Speed (2001), "Analysis of cDNA Microarray images", Briefing in Bioinformatics, vol.2, No.4, pp. 341-349.

[3] M.B. Eisen, ScanAlyze (1999). http://rana.lbl.gov/EisenSoftware.htm

[4] Axon Instruments, Inc. (1999): GenPix 4000A User's guide

[5] GeneSifter data center: http://www.genesifter.net/web/dataCenter.html

[6] M.J. Buckley (2000), The Spot user's guide. CSIRO Mathematical and Information Science. http://www.cmis.csiro.au/IAP/Spot/spotmanual.htm

[7] ImaGene, ImaGene 6.1 User Manual http://www.biodiscovery.com/index/papps-webfiles-action.

[8] S. Beucher, F. Meyer (1993). "The morphological approach to segmentation: The watershed transformation", Optical Engineering, Vol. 34, pp. 433-481.

[9] R. Adams and L. Bischof (1994). "*Seeded Region Growing*", IEEE Trans. Pattern Anal. Machine Intell., vol 16, pp 641-647.

[10] D. Bozinov and J. Rahenfuhrer (2002), "Unsupervised technique for robust target separation and analysis of DNA Microarray spots through adaptive pixel clustering", Journal of Bioinformatics, vol 18, pp 747-756.

[11] K. Blekas, N.P. Galatsanos and I. Georgiou (2003), "*An unsupervised Artifact Correction Approach for the Analysis of DNA Microarray Images*", in Proc. IEEE International Conf. on Image Processing (ICIP), vol 2, pp 165-168.

[12] M. Schena, D. Shalon, R.W. Davis and P. O. Brown (1995), "Quantitative monitoring of gene expression patterns with a complementary DNA microarrray", Science 270, pp. 467-470.

[13] Sandrine Dudoit, Jane Fridlyand and Terry Speed (2002), "Comparison of discrimination methods for the classification of tumours using gene expression data". Journal of the American Statistical Association, 97(457):77-87

[14] Bill Martin and Robert M. Horton (2004), "A Java Program to Create Simulated Microarray Images", Proceedings on the 2004 IEEE Computational Systems Bioinformatics Conference (CSB 2004).

[15] Pratt W. K. (1991): *Digital Image Processing (Second Edition)*, John Wiley & Sons, New York, pp. 275-285.

[16] Gonzales R. C., Woods R. E. (1992): *Digital Image Processing*, Addison-Wesley, pp. 167-190.

[17] Weeks A. A. (1996): *Fundamentals of Electronic Image Processing*, Wiley-IEEE Press, pp. 109-120.

[18] Daubechies I. (1988): “Orthonormal Bases of Compactly Supported Wavelets”, Communication of Pure and Applied Math. 41, pp. 909-996.

[19] Mekle R, Laine A., Smith S., Singer C., Koenigsberg T., Brown M.(2000): “*Evaluation of a multiscale enhancement protocol for digital mammography*”, Wavelet Applications in Signal and Image Processing VIII, Akram Aldroubi; Andrew F. Laine; Michael A. Unser; Eds. Proc. SPIE 4119, pp. 1038-1049.

[20] Brown T. J. (2000): “*An adaptive strategy for wavelet based image enhancement*”, Proceeding of IMVIP 2000 - Irish Machine Vision and Image Processing Conference, Dublin, Ireland, pp. 67-81.

[21] Brown T. J. (2001): “*Differential Image Enhancement via Wavelet Foveation*”, Proceeding of IMVIP 2001-Irish Machine Vision and Image Processing Conference, Dublin, Ireland, pp. 179-192.

[22] E. Athanasiadis, N. Piliouras, K. Sidiropoulos, H. Georgiou, C. Makris, N. Dimitropoulos, D. Cavouras, (2004)“*Mammographic Image Enhancement Using Wavelet-Based Processing And Histogram Equalization*”, 1st International Conference “From Scientific Computing to Computational Engineering” (1st IC-SCCE), Athens, Greece, September 10-18.

[23] Konstantinos P. Sidiropoulos, Nikos G. Piliouras, Emmanouil I. Athanasiadis, Harris V. Georgiou, Christos S. Makris, Nikos Dimitropoulos, Dionisis A. Cavouras, (2004) “*Statistical Versus Wavelet-Based despeckling techniques for enhancing medical ultrasound images*”, 1st International Conference “From Scientific Computing to Computational Engineering” (1st IC-SCCE), Athens, Greece, September 10-18.