

Segmentation of Complementary DNA Microarray Images by Wavelet-based Markov Random Field Model

Journal:	<i>Transactions on Information Technology in BioMedicine</i>
Manuscript ID:	TITB-00018-2008.R1
Manuscript Type:	Paper
Date Submitted by the Author:	11-May-2009
Complete List of Authors:	<p>Athanasiadis, Emmanouil; University of Patras, Medical Image Processing and Analysis (M.I.P.A.) Group, Laboratory of Medical Physics, School of Medical Science</p> <p>Cavouras, Dionisis; Technological Educational Institute of Athens, Medical Image and Signal Processing (MED.I.S.P.) Laboratory, Department of Medical Instruments Technology</p> <p>Glotsos, Dimitris; Technological Educational Institute of Athens, Medical Image and Signal Processing (MED.I.S.P.) Laboratory, Department of Medical Instruments Technology</p> <p>Georgiadis, Pantelis; University of Patras, Medical Image Processing and Analysis (M.I.P.A.) Group, Laboratory of Medical Physics, School of Medical Science</p> <p>Kalatzis, Ioannis; Technological Educational Institute of Athens, Medical Image and Signal Processing (MED.I.S.P.) Laboratory, Department of Medical Instruments Technology</p> <p>Nikiforidis, George; University of Patras, Medical Image Processing and Analysis (M.I.P.A.) Group, Laboratory of Medical Physics, School of Medical Science</p>
TIPS:	INFORMATION SCIENCE & TECHNOLOGY, BIO- INFORMATICS, Knowledge Discovery, Neural Nets & Expert Systems

Segmentation of Complementary DNA Microarray Images by Wavelet-based Markov Random Field Model

Emmanouil I. Athanasiadis, Dionisis A. Cavouras, *Member, IEEE*, Dimitris Th. Glotsos, Pantelis V. Georgiadis, *Member, IEEE*, Ioannis K. Kalatzis, and George C. Nikiforidis

Abstract—A wavelet based modification of the Markov Random Field (*WMRF*) model is proposed for segmenting complementary DNA (*cDNA*) microarray images. For evaluation purposes, five simulated and a set of five real microarray images were used. The 1-level Stationary Wavelet Transform (*SWT*) of each microarray image was used to form two images, a denoised image, using hard thresholding filter, and a magnitude image, from the amplitudes of the *SWT*'s horizontal and vertical components. Elements from these two images were suitably combined to form the *WMRF* model for segmenting spots from their background. The *WMRF* was compared against the conventional *MRF* and the Fuzzy C Means (*FCM*) algorithms on simulated and real microarray images and their performances were evaluated by means of the segmentation matching factor (*SMF*) and the Coefficient of Determination (r^2). Additionally, the *WMRF* was compared against the *SPOT* and *SCANALYZE*, and performances were evaluated by the Mean Absolute Error (*MAE*) and the Coefficient of Variation (*CV*). The *WMRF* performed more accurately than the *MRF* and *FCM* (*SMF*: 92.66, 92.15 and 89.22, r^2 : 0.92, 0.90 and 0.84 respectively) and achieved higher reproducibility than the *MRF*, *SPOT* and *SCANALYZE* (*MAE*: 497, 1215, 1180 and 503, *CV*: 0.88, 1.15, 0.93 and 0.90 respectively).

Index Terms—*cDNA* Microarray, Markov Random Field (*MRF*), image segmentation, Wavelet.

I. INTRODUCTION

Complementary DNA (*cDNA*) microarray chips are used for the qualitative and quantitative identification of thousands of genes at the same time [1]. Biological and technological specifications concerning *cDNA* microarray technology can be found in [1]. When using *cDNA* microarray images, the objective is to calculate the fluorescence intensity value for each spot on the chip. This

intensity value is closely related with the expression abundance of the specific gene. However, accurate assessment of the spot's intensity value would require precise segmentation of the spot-pixels from the spot's background. Although, it has been advocated [1] that an inaccurate segmentation algorithm with a background correction method has no impact on the calculated spot intensity value, nevertheless, it has been recently proved [20,31] that segmentation methods can significantly influence microarray data precision, leading to inaccurate intensity extraction results.

Thus, the exact location and accurate identification of the boundary of each spot are crucial for more precise gene expression measurements.

In order to measure spot intensities, three steps are followed [1]-[2]: 1/ the addressing or gridding step, for the individual localization of cells (spots and their background), 2/ the segmentation step, for the discrimination of the spot from its background in each cell, and finally, 3/ the intensity extraction step, where calculation of the intensity of each spot is performed. The segmentation step is the most crucial due to the fact that errors during segmentation propagate to the intensity extraction step, leading to inaccurate intensity calculations.

In previous studies, several segmentation methods have been proposed in order to overcome limitations of currently used segmentation methods [3]-[7]. More specifically, in the publicly available *SCANALYZE* [3] and *GenePix* [4], fixed and adaptive circle segmentation methods have been adopted respectively. According to those two methods, all spots are considered to be circular with a fixed predefined radius; radius is constant in the former and adaptive in the latter technique. This assumption is the main drawback of those techniques due to the fact that spots are not perfect circles but are irregular in shape. Seeded Region Growing (*SRG*) [8] is one representative method of adaptive shape segmentation and it has been employed by the *SPOT*[®] [6] software. However, the limitation of *SRG* is that all initial seeds must be accurately defined [9]. In the *ImaGene* [7] software, a histogram based segmentation method is used, where values between the 80th and the 95th histogram percentile contribute to the calculation of the intensity value of the spot. Segmentation algorithms, based on the statistical Mann-Whitney test, that apply a thresholding filter on the cell, have also been introduced [10]. Active contour

Manuscript received January 29, 2008, revised April 27, 2009. E. I. Athanasiadis thanks the Greek State Scholarships Foundation (I.K.Y.), under Grand 4690/2007-2008, for funded the above work.

E. I. Athanasiadis, P. Georgiadis, and G. C. Nikiforidis are with the Medical Image Processing and Analysis (*M.I.P.A.*) Group, Laboratory of Medical Physics, School of Medical Science, University of Patras, 26 500 Rion - Patras, Greece. (phone: +30-2610-997745 e-mail: mathan@upatras.gr).

D. A. Cavouras, D. Th. Glotsos, and I. K. Kalatzis are with the Medical Image and Signal Processing (*MED.I.S.P.*) Laboratory, Department of Medical Instruments Technology, Technological Educational Institute of Athens, Ag. Spyridonos Street, Aigaleo, 122 10, Athens, Greece. (phone: +30-210-5385375; fax: +30-210-5385302; e-mail: cavouras@teiath.gr.).

techniques have also been applied [11], where an initial contour is located on the spot-image and a contour deformation is formed for the task of delineating the spot's boundaries. However, deformation that makes use of spatial information is apt to noise. Mixture Model [12]-[13] algorithms have been adopted for the task of segmenting microarray images using the Bayes' rule. However, these methods are only based on the intensity and histogram of the image and do not take into account dependences among adjacent pixels (contextual).

In the last decade, Markov Random Field (*MRF*) [15] models, that use contextual and textural information, have been proposed for the task of segmenting [19] microarray images. In [27], multiresolution approaches have been proposed with the use of the Discrete Wavelet Transform (*DWT*) for the task of modeling the labels at various scales; label is a binary variable that indicates whether a pixel belongs to the foreground (spot) "1" or the background "0". That modeling scheme has taken into account interscale and intrascale statistical dependencies. However, during the *DWT* decomposition process, image size is reduced by half and no further information has been used. Additionally, *DWT*-wavelet domain images have been used to model only the Energy of labels in the *MRF* model.

In the present study, we directly deal with the segmentation problem and we assume that the gridding stage has already been done and the location of each cell is known, as explicitly described in a previous study by our group [14]. Here, a new segmentation approach based on the *MRF* and the Stationary Wavelet Transform (*SWT*) [16] is proposed. The 1-level *SWT* of the image was used to form two images, a denoised image, using hard thresholding filter [24], and a magnitude image, from the amplitudes of the *SWT*'s horizontal and vertical components. The *SWT* was employed, since it retains the size of the decomposed components and information contained within the wavelet domain is not compressed as in the case of the *DWT*, and, thus, elements from the denoised and magnitude images could be suitably combined in actual scale to form the proposed wavelet-based *MRF* (*WMRF*) model for segmenting spots from their background. In addition, with the use of the stationary wavelet transform, additional textual (i.e. the intensity in the wavelet domain) and contextual information (i.e. the details in the wavelet domain), which are "hidden" in the wavelet domain, were also used to model both the labels and the feature's Energy.

II. MATERIALS AND METHODS

A. Material

Simulated microarray images were produced, as described in [19], for the task of numerically evaluating the segmentation methods. More precisely, by applying a simple thresholding segmentation technique [19] to an actual *cDNA* microarray image with 1600 spots, a binary

image was produced, which was in turn used as a template for the creation of a simulated image. Thus, the location, boundary, and area of all simulated spots were *a priori* determined. Moreover, the intensities of each spot (foreground) were formed from intensity values randomly drawn from an exponential distribution, which had a predefined mean intensity value that was randomly picked between 0 and $2^{16}-1$, which is the range of a typical actual microarray image. Additionally, background intensities were similarly produced but from a single exponential distribution with mean intensity value equal to 4000, which is close to the mean background intensity of actual microarray images. Finally, the so formed simulated microarray image was corrupted with additive Gaussian noise [13] with Signal to Noise Ratio (*SNR*) 1, 3, 5, 7 and 9 *dB*, thus, producing five different microarray images. The latter were employed for assessing the robustness of the simple Fuzzy C Means (*FCM*) [18], the conventional *MRF*, and the proposed Wavelet based *MRF* (*WMRF*) segmentation algorithms, in determining the spot's boundaries.

In addition, the performances of the segmentation algorithms were estimated on five actual microarray images, which were replicates at five different distinct time intervals of the same experiment for both the Red and Green channels [1], [25]. Those microarray images were obtained from a free public available database [25], well established and widely used in the literature concerning *Saccharomyces cerevisiae* [25]. The software used for the creation of the microarray images was the ArrayMaker (Version 1.8.5). Detailed descriptions may be found in [26]. In the common reference Green channel, normalized spot-image intensities (see section *D*) are expected to remain constant with time, and this feature is usually employed in assessing the performance (reproducibility) of the segmentation techniques [20]: the steadier the segmentation result, in terms of mean spot intensity value at different time-intervals, the better the segmentation algorithm.

B. Markov Random Field (*MRF*)

MRFs were first introduced in computer vision by Geman [22]. The main benefit of the *MRF* model is that it can manipulate textural and contextual information of the image.

Suppose that a feature vector F has been extracted from a random image X and that Y is the result of segmentation (labels "0" for Background and "1" for Foreground or spot) based on the observed features F . According to the Bayesian theory, the segmentation problem can be expressed by using relation (1).

$$P(Y|F) = \frac{p(F|Y)P(Y)}{p(F)} \quad (1)$$

where, $P(Y|F)$ is the *a posteriori* probability of Y given the observed features F , $p(F|Y)$ is the conditional probability of F given Y , $P(Y)$ is the *a priori* probability of Y that is used to describe the label distribution, and $p(F)$ is the probability distribution of F .

The objective is to find the maximum *a posteriori* probability $P(Y | F)$ in (1); $p(F)$ in (1) was omitted since it does not vary with respect to any solution [23]. If we assume the energy form of $P(Y)$ to be E_L and of $p(F | Y)$ to be E_F^c , where c refers to the class (foreground, background), then the Total Energy can be derived from the sum of the two energies [23].

$$E^c = E_L + aE_F^c \quad (2)$$

where a is a weighted parameter used to determine the individual energy percentage contribution in respect to the total energy.

The *a posteriori* probability may be expressed, using Gibbs distribution [22] and [23], as in (3).

$$P(Y | F) = \frac{1}{Z} e^{-\frac{E^c}{T}} \quad (3)$$

where, T is a constant, E^c is the energy function of class c and Z is a normalized constant given in (4) for all possible configurations of Y .

$$Z = \sum_{c=1}^2 e^{-\frac{E^c}{T}} \quad (4)$$

In order to model the labels distribution (E_L), the Multilevel Logistic Model [15] has been adopted with 1st order neighbor system [23] as in (5)

$$E_L(y) = \sum_{t \in \text{Neighbor}} \delta(y_t, y_{center}) \quad (5)$$

where y is the value of the label ("1" for the foreground or spot and "0" for the background) t are the neighboring pixels of the y_{center} center, $\delta(y_t, y_{center}) = -1$ if $y_t = y_{center}$ else $\delta(y_t, y_{center}) = 1$.

In addition, if we assume that features follow a Gaussian distribution [23], we can model the feature distribution (E_F) as in (6).

$$E_F(f) = \sum_{k=1}^K \left[\frac{(f^k - \mu_c^k)^2}{2(\sigma_c^k)^2} + \log(\sqrt{2\pi}\sigma_c^k) \right] \quad (6)$$

where, f is the feature vector, μ and σ are the mean value and the standard deviation of each feature k of class c , and K is the number of features. In our case, K was set to unity due to the fact that only the intensity of the pixel was used.

Finally, for the maximization of $P(Y | F)$ in (1), the Expectation Maximization algorithm was applied according to the following scheme [23]:

- 1) For each cell, a k-means clustering algorithm [18] was applied in order to make an initial rough estimation of the 2 classes, the foreground (spot) and background classes.
- 2) For each class, μ and σ were calculated.
- 3) E_L in (5) and two E_F^c in (6) energies for each class c were calculated by means of a 3x3 moving window within each cell. Next, two total energies in (2) were computed. Finally, based on (4), two *a posteriori* probabilities were found and classification was performed to the highest *a posteriori* probability.
- 4) Label values Y were redefined.

- 5) The procedure was repeated from step 2 with the updated labels, until $E_{new}^c - E_{old}^c < \text{threshold}$.

C. Proposed Wavelet MRF (WMRF) Method

The SWT transform was first applied onto the microarray image, using one-level decomposition as an optimal solution to our microarray data; next the hard thresholding technique, as shown in (7), was employed to the detail images.

$$W_{out} = \begin{cases} W_{in} + Th \cdot (G-1) & \text{if } W_{in} > Th \\ W_{in} - Th \cdot (G-1) & \text{if } W_{in} < -Th \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where, W_{out} denotes the output and W_{in} the input coefficient values of the details (Horizontal (Ho), Vertical (Ve) and Diagonal (Di)). Th and G are threshold and gain values respectively. Threshold was automatically calculated by using the principle of Stein's Unbiased Risk Estimate [30]. In each detail image, an individual threshold value was selected with the use of "thselect" *MATLAB* command.

Finally, the denoised (D) and magnitude (M) images were generated (see Fig.1). Images D and M were used in the calculation of the new energies WE_L and WE_F^c (see relations (9) and (11)) of the labels and features respectively.

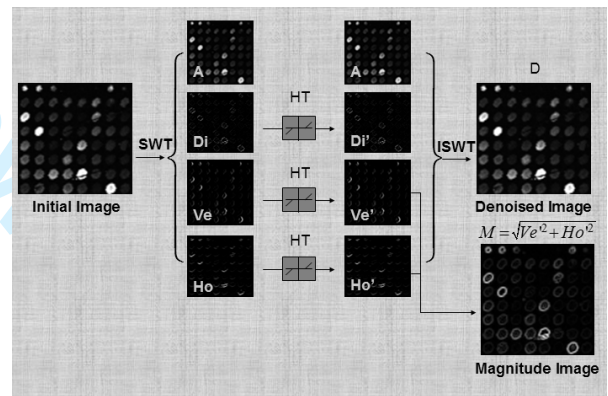


Fig 1: SWT-decomposition of the initial image, filtering by hard thresholding (HT), and formation of the filtered (D) and magnitude (M) images.

The magnitude image [24] was formed as in (8)

$$M = \sqrt{Ho'^2 + Ve'^2} \quad (8)$$

where Ho' and Vi' are the Horizontal and the Vertical wavelet decomposed detail images respectively at the 1st level of decomposition.

Following a trial and error procedure, the biorthogonal 1.5 mother wavelet [29] was chosen as optimum in delineating spot boundaries. For the calculation of E_L in (5), the k-means clustering algorithm was applied to both denoised and magnitude images and, as a result, two binary images were produced, a rough estimation of spot and background pixels. WE_L energy was calculated employing (9)

$$WE_L(y) = w_1 \sum_{t \in \text{Neighbor}} \delta(y_t, y_{center})_M + (1 - w_1) \sum_{t \in \text{Neighbor}} \delta(y_t, y_{center})_D \quad (9)$$

where $w_1 \in [0, 1]$ is a weight constant.

For the calculation of WE_F^c a smoothing kernel (10) was first applied to the magnitude image M

$$L_{mask} = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (10)$$

and WE_F^c was calculated as follows

$$WE_F^c = w_2 \left[\frac{(I_M - \mu_M^c)^2}{2(\sigma_M^c)^2} + \log(\sqrt{2\pi}\sigma_M^c) \right] + (1 - w_2) \left[\frac{(I_D - \mu_D^c)^2}{2(\sigma_D^c)^2} + \log(\sqrt{2\pi}\sigma_D^c) \right] \quad (11)$$

where I_M is the smoothed magnitude image, I_D is the denoised image, μ_M^c and μ_D^c are the mean values of class c of smoothed magnitude image M and denoised image D , respectively. Finally, σ_M^c and σ_D^c are the standard deviations of M and D images of class c , respectively. $w_2 \in [0, 1]$ is a weight parameter.

The two new SWT -total energies were then calculated by (12).

$$WE^c = WE_L + aWE_F^c \quad (12)$$

D. Intensity Extraction and Normalization

The outcome of each segmentation algorithm was a binary image, in which white pixels (255) corresponded to spots and black pixels (0) to background. In each cell-image (spot with its background) the mean intensities of both spot (foreground) and background were calculated by reference to the corresponding intensities on the original image. All spot mean intensities, on both simulated and actual microarray images, were background corrected by subtracting from each spot's mean the background's mean [1]. Regarding actual microarray images, spot intensities were calculated for both Red R (control) and Green G (common reference) channels.

Following intensity extraction, a normalization process was essential for removing systematic variations from the measured spot intensities in $cDNA$ microarray experiments [28] and, consequently, for assessing spot intensities' variation at different time intervals. In the present study, an intensity dependent normalization method (lowess) [28] was implemented for normalizing spot intensities. More precisely, MA plots were firstly constructed for one of the five replicate images, where $M = \log_2(I_R/I_G)$, $A = (1/2)\log_2(I_R I_G)$, and I_R and I_G are the green and red channels spot mean intensities (see Fig. 2).

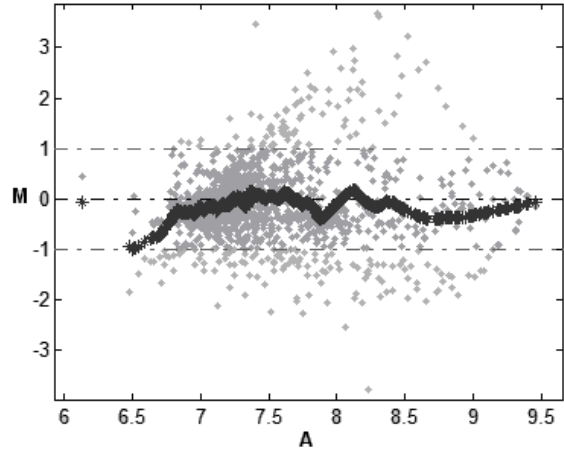


Fig 2: MA plot and lowess normalization method.

MA plots have been previously found useful in identifying artifacts and intensity depended patterns [28]. Each MA plot was next fitted by a polynomial regression function (\hat{M} : predicted using lowess technique), which was used to normalize M , by subtracting from each value of M the corresponding predicted \hat{M} , $\dot{M} = M - \hat{M}$. Regarding the green channel, which is the common reference channel, individual I_G intensity values were normalized according to (13) [...]

$$\log_2 I_G = A - \frac{\dot{M}}{2} \quad (13)$$

Normalized green channel mean spot intensity values were calculated for the spots of the five individual images (replicates) and were used for further processing. Regarding red channel spots, mean spot intensities were left intact, since they were not employed in the evaluation of the precision of the algorithm.

E. Evaluation

For the simulated microarray images the following metrics were calculated:

1) Segmentation matching factor (SMF) [14].

$$SMF = \frac{B_{segment} \cap B_{actual}}{B_{segment} \cup B_{actual}} \quad (14)$$

where $B_{segment}$ and B_{actual} are the binary versions of the segmented and simulated actual cell-images respectively.

2) Coefficient of determination r^2 [19]

$$r^2 = \frac{\sum_{i=1}^{\text{All spots}} (I_{segment(i)} - \bar{I}_{actual})^2}{\sum_{i=1}^{\text{All spots}} (I_{actual(i)} - \bar{I}_{actual})^2} \quad (15)$$

where $I_{segment}$ and I_{actual} are the mean intensity values of the calculated and simulated actual spots respectively, i refers to individual cell images ($i=1 \dots 1600$), and \bar{I}_{actual} is the overall

mean of the spot intensity values of the simulated actual image.

Regarding the real microarray images, the reproducibility of the segmentation techniques was quantified by means of:

1) Mean Absolute Error (*MAE*) [20].

$$MAE_{spot} = \frac{1}{n} \sum_{i=1}^n |I_i - \bar{I}| \quad (16)$$

where n is the number of replicates ($n = 5$), I_i is the normalized mean spot intensity value and \bar{I} is the spot's overall mean, calculated from the means of the corresponding spots in the n replicates.

2) Coefficient of Variation (*CV*) [21].

$$CV_{spot} = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (I_i - \bar{I})^2}}{\frac{1}{n} \sum_{i=1}^n I_i} \quad (17)$$

The lower the *MAE* and the *CV* values, the better is the reproducibility of the method being evaluated.

Finally, all methods and metrics described in the above sections were performed by means of custom made programs developed in MATLAB® [17].

III. EXPERIMENTAL RESULTS AND DISCUSSION

The most essential task of the segmentation process is the discrimination of each spot's foreground from its background. Since it is unfeasible to know each spot's exact location, extend, and outline on real microarray images, simulated data were generated for validating the segmentation algorithms [20]. Five simulated microarray images were formed for quantifying the segmentation performances of the proposed wavelet based *MRF* technique, the conventional *MRF* technique as well as the *FCM* clustering algorithm. The segmentation matching factor (*SMF*) and the coefficient of determination (r^2) were both used for evaluation purposes, since the former concerns segmentation accuracy and the latter intensity extraction precision, and the results are illustrated in Tables I and II respectively. In addition, a graphic representation of the results is shown in Fig. 3 and 4.

TABLE I
SMF RESULTS ON SIMULATED MICROARRAY IMAGES

<i>SNR</i> (dB)	<i>FCM</i>	<i>MRF</i>	<i>WMRF</i>
1	89.22	92.15	92.66
3	89.94	92.96	93.53
5	91.25	94.24	94.56
7	93.81	96.30	96.69
9	94.69	97.20	97.47

Comparative *SMF* results for the 5 simulated images with different *SNR* levels. The 2nd, 3rd, and 4th columns indicate the *SMF* by using *FCM*, *MRF* and the proposed *WMRF* segmentation techniques respectively.

TABLE II

r^2 RESULTS ON SIMULATED MICROARRAY IMAGES

<i>SNR</i> (dB)	<i>FCM</i>	<i>MRF</i>	<i>WMRF</i>
1	0.84	0.90	0.92
3	0.86	0.93	0.94
5	0.92	0.95	0.96
7	0.96	0.97	0.98
9	0.97	0.98	0.99

Comparative r^2 results for the 5 simulated images with different *SNR* levels. The 2nd, 3rd, and 4th columns indicate r^2 by using *FCM*, *MRF* and the proposed *WMRF* segmentation techniques respectively.

According to our findings on the simulated data, the proposed *WMRF* algorithm achieved the highest *SMF* and r^2 scores as compared to the conventional *MRF* and *FCM* techniques at all *SNR* levels. More precisely, at the 1dB *SNR* Gaussian noise level (see Table I), the *FCM* scored 89.22 whereas the *MRF* and *WMRF* scored 92.15 and 92.66 respectively. This *SMF* difference, of about 3%, between the *FCM* and *WMRF* was sustained for higher *SNR* levels. In contrast, in the case of the r^2 , differences among *FCM*, *MRF*, and *WMRF* decreased for higher *SNR* levels (see Table II).

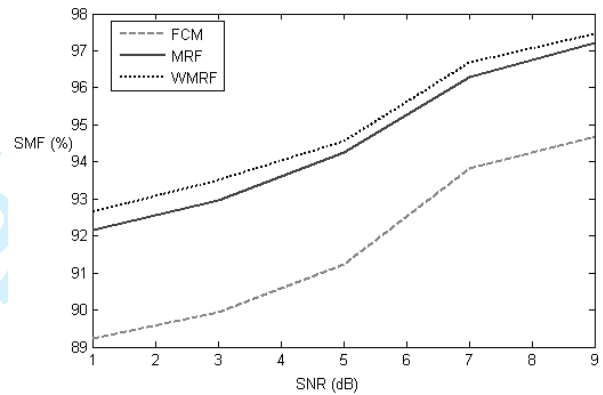


Fig 3: *SMF* results on simulated data using *FCM*, *MRF* and *WMRF* for different *SNR* levels.

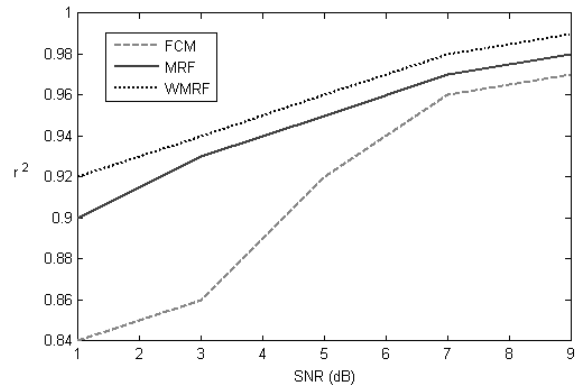


Fig 4: r^2 results on simulated data using *FCM*, *MRF* and *WMRF* for different *SNR* levels.

Furthermore, the proposed *WMRF* and the conventional *MRF* techniques were both used to segment five actual microarray images. The algorithms' performances were assessed by the reproducibility of the results on the green

channel and they were quantified employing two metrics, *MAE* (sameness) and *CV* (variation). Moreover, for comparison reasons, the publicly available software programs *SCANALYZE*[®] [3] and *SPOT*[®] [6] were employed for segmenting the same microarray images; the *SCANALYZE* software makes use of the Fixed Circle (*FC*) segmentation technique and the *SPOT* of the Seeded Region Growing (*SRG*). Results concerning *MAE* and *CV* for all 6400 spots are illustrated in Table III and Fig 5.

TABLE III
MAE AND CV RESULTS ON REAL MICROARRAY IMAGES

	<i>MAE</i>	<i>CV</i>
<i>MRF</i>	1215	1.15
<i>WMRF</i>	497	0.88
<i>SCANALYZE (FC)</i>	503	0.90
<i>SPOT (SRG)</i>	1180	0.93

Results for the four segmentation techniques by means of *MAE* and *CV*, applied on five real microarray images.

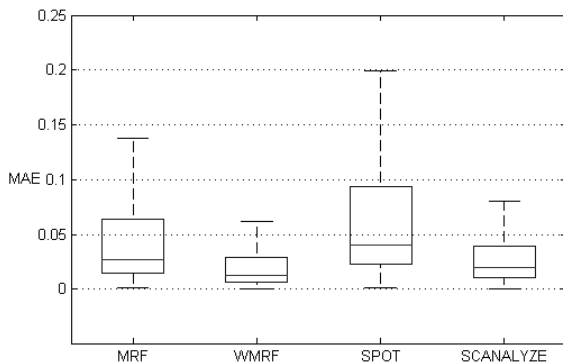


Fig 5: Normalized boxplots for the four segmentation techniques, using *MRF*, *WMRF*, *SPOT* and *SCANALYZE* applied on five real microarray images.

The lowest *MAE* and *CV* measurements were obtained by the proposed *WMRF* method. More precisely, *WMRF* scored 497 mean *MAE* for the 6400 spots, which is close to the score attained by the *FC* technique. Additionally, the *MAE* of the proposed *WMRF* technique was half that of the conventional *MRF*, which is indicative of the improvement on *MRF* that wavelet based information may bring about. Additionally, the proposed method achieved a *MAE* (497) score lower, but close, to *SCANALYZE*'s *FC* technique (503), but definitely lower than that of the conventional *MRF* technique (1215) and of the *SPOT*'s *SRG* technique (1180) (see Table III). Additionally, the *CV* values of *WMRF*, *FC*, *SRG*, and *MRF* (0.88, 0.90, 0.93, 1.15) were close to unity with the *WMRF* having the lowest value.

Regarding segmentation processing time, for an actual microarray image of 6400 spots it took approximately 23 minutes for the *WMRF* against about 20 minutes for conventional *MRF* on a Pentium IV 3GHz desktop PC with 1GB RAM. Nevertheless, it should be noted that the algorithms have not been optimized yet, for optimal time-processing results, as compared with *SCANALYZE*'s fast response times, using the fixed-circle technique. Although our method, as compared to *Scanalyze (FC)*, shows a slight

improvement in the precision of reproducibility, its value is on the delineation accuracy of the spots' boundaries and, hence, of the spot's expression level, that can significantly influence microarray intensity extraction results [20, 31].

CONCLUSIONS

A new method (*WMRF*), based on the *SWT* and the *MRF* was proposed, for improving the segmentation of microarray images. The *WMRF* method was tested on both simulated and actual *cDNA* microarray images and was compared against the *FCM* and the ordinary *MRF* methods and against two publicly available software packets, *SPOT* and *SCANALYZE*. Results revealed the potentiality of the proposed method, which may be attributed a)to the employment of the *SWT*, for retaining the actual scale on the wavelet transformed images, and b)to the formation of novel labels and features, that utilize additional hidden textual and contextual information, to design the proposed *WMRF* scheme.

REFERENCES

- [1] Y.H. Yang, M. J. Buckley, S. Duboit and T.P.Speed , "Comparison of methods for Image Analysis on cDNA Microarray Data", *Journal of Computational and Graphical Statistics*, vol. 11, pp 108-136, 2002.
- [2] M. Schena, D. Shalon, R.W. Davis and P. O. Brown, "Quantitative monitoring of gene expression patterns with a complementary DNA microarray", *Science* 270, pp. 467-470, 1995.
- [3] M.B. Eisen, *ScanAlyze* (1999). Available: <http://rana.lbl.gov/EisenSoftware.htm>
- [4] Axon Instruments, Inc. (1999): *GenPix 4000A User's guide*
- [5] GeneSifter data center, Available: <http://www.genesifter.net/web/dataCenter.html>
- [6] M.J. Buckley (2000), *The Spot user's guide*. CSIRO Mathematical and Information Science. Available: <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] R. Adams and L. Bischof, "Seeded Region Growing", *IEEE Trans. Pattern Anal. Machine Intell.*, vol 16, pp 641-647, 1994.
- [9] Dat Tran, Michael Wagner, Yee W. Lau and Mitsuo Gen, "Fuzzy Methods for Voice-Based Person Authentication", *IEEJ (Institute of Electrical Engineers of Japan) Transactions on Electronics, Information and Systems*, vol. 124, no. 10, pp. 1958-1963, 2004.
- [10] Joseph L. DeRisi, Vishwanath R. Iyer, Patrick O. Brown, "Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale", *SCIENCE*, vol 278, pp 680, 1997.
- [11] T Srinark, C Kambhamettu, "a microarray image analysis system based on multiple snakes", *Journal of Biological Systems Special Issue*, vol 12, pp 127-157, 2004
- [12] K. Blekas, N.P. Galatsanos and I. Georgiou, "An unsupervised Artifact Correction Approach for the Analysis of DNA Microarray Images", *Proc. IEEE International Conf. on Image Processing (ICIP)*, vol 2, pp 165-168, 2003.
- [13] K. Blekas, N.Galatsanos, A. Likas, and I.E. Lagaris, "Mixture Model Analysis of DNA Microarray Images", *IEEE Transactions on Medical Imaging*, vol 24, pp. 901-907, 2005.
- [14] Emmanouil Athanasiadis, Dionisis Cavouras, Panagiota Spyridonos, Ioannis Kalatzis and George Nikiforidis, "An Automatic Microarray Image Gridding Technique Based on Continuous Wavelet Transform", *Lecture Notes in Computer Science*, vol. 4673, pp 854-870, 2007.
- [15] S.Z. Li, *Markov Random Field Modeling in Computer vision*. New York: Springer- Verlag, 2001

- [16] James E. Fowler, "The Redundant Discrete Wavelet Transform and Additive Noise", *IEEE Signal Processing Letters*, Vol. 12, NO. 9, 2005.
- [17] The MathWorks, Inc. Software, MATLAB®.
- [18] Dat Tran and Michael Wagner, "Fuzzy C-Means Clustering-Based Speaker Verification", *Lecture Notes in Computer Science: Advances in Soft Computing - AFSS 2002*, N.R. Pal, M. Sugeno (Eds.), pp. 318-324, Springer-Verlag, 2002.
- [19] O. Demirkaya, M. H. Asyali and M.M. Shoukri, "Segmentation of cDNA Microarray Spots Using Markov Random Field Modeling", *Bioinformatics*, vol. 21, pg. 2994-3000, 2005.
- [20] Lehmussola, A. et al., "Evaluating the performance of microarray segmentation algorithms", *Bioinformatics*, vol. 22, pp. 2910-2917, 2006.
- [21] George F. Reed, Freyja Lynn, and Bruce D. Meade, "Use of Coefficient of Variation in Assessing Variability of Quantitative Assays", *Clinical and Diagnostic Laboratory Immunology*, vol. 9, No. 6, pp. 1235-1239, 2002.
- [22] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 6, pp 721 - 741, 1984.
- [23] Huawu Deng, and David A. Clausi, "Unsupervised Segmentation of Synthetic Aperture Radar Sea Ice Imagery Using a Novel Markov Random Field Model", *IEEE Trans. On Geoscience and Remote Sensing*, vol. 43, No. 3, 2005.
- [24] P Sakellaropoulos, L Costaridou and G Panayiotakis, "A wavelet-based spatially adaptive method for mammographic contrast enhancement", *Phys. Med. Biol.*, Vol. 48, pp. 787-803, 2003
- [25] Joseph L. DeRisi, Vishwanath R. Iyer, Patrick O. Brown, "Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale", *SCIENCE*, Vol 278, pp 680, 1997.
- [26] Stanford University, School of Medicine
Available: <http://cmgm.stanford.edu/pbrown/scanner.html>
- [27] Hideki Noba, Mahdad N. Shirazi, Eiji Kawaguchi, "MRF-based texture segmentation using wavelet decomposed images", *Pattern Recognition*, vol 35, pp. 771-782, 2002
- [28] Yee Hwa Yang, Sandrine Dudoit, Percy Luu, David M. Lin, Vivian Peng, John Ngai, and Terence P. Speed, "Normalization for cDNA microarray data : a robust composite method addressing single and multiple slide systematic variation", *Nucleic Acid Research*, vol. 30, No. 4 e15, 2002
- [29] Burrus CS, Gopinath RA, Guo H, Introduction to Wavelets and Wavelet Transforms. Prentice Hall, Englewood Cliffs, NJ. 1998.
- [30] Donoho, D.L., and Johnstone, I.M. "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, vol 81, pp. 425-455, 1994.
- [31] Ahmed, A.A. et al. "Microarray segmentation methods significantly influence data precision". *Nucleic Acids Res.*, vol. 50 pp. 32, 2004.



Emmanouil I. Athanasiadis was born in Nafplio, Greece, in 1981. He received the B.Sc. degree in Medical Instruments Technology from the Technological Institution of Athens, Athens, Greece, in 2004, and his M.Sc. in Medical Physics from the Medical School of the University of Patras in 2006, funded by the Greek State Scholarships (I.K.Y.) Foundation. Since 2006, he is a PhD candidate in Medical Physics at the University of Patras, Patras, Greece, and he is funded by the Greek State Scholarships Foundation (I.K.Y.). His research interests include medical image processing and analysis with focus in the field of bioinformatics.



Dionisis A. Cavouras (IEEE Member) was born in Kalamata, Greece, in 1951. He received his B.Sc. ('74) in Electronic Engineering, and the M.Sc. ('76) and Ph.D. ('81) in Systems Engineering from the City University, London, England. He was a research assistant at the Dept. of Nuclear Medicine, Guy's Hospital, London, England between 1976-1981, he worked as a Research Fellow at the Dept. of Computed Tomography, Hellenic Air-force Hospital, Athens, Greece, between 1994-1991 and since then he is a professor of medical imaging processing at the

Dept. of Medical Instrumentation Technology, TEI-Athens, and director of the laboratory of Medical Image and Signal Processing. He is also an adjunct lecturer on medical image processing at the Dept. of Informatics, University of Athens, Greece, and the Dept. of Medical Physics, University of Patras, Greece, and on Network Computing at the School of Engineering and Design, Brunel University, UK. His research interests include medical image processing, image analysis, pattern recognition, medical statistics, medical physics, and network computing. He has published numerous technical and medical papers as well as conference proceedings.



2007.

Dimitris Th. Glotsos received the B.Sc degree in Medical Instruments Technology from the Technological Educational Institute of Athens in 2000, Greece and the M.Sc and PhD degrees in Medical Physics from the University of Patras in 2002 and 2006 respectively, Greece. He has been working as an adjunct laboratory instructor at the Technological Educational Institute of Athens, Greece, since October



Pantelis V. Georgiadis (IEEE Member) was born in Thessaloniki, Greece, in 1980. He received the B.Sc. degree in medical instruments technology from the Technological Institution of Athens, Athens, Greece, in 2004, and the M.Sc. degree in data communication systems from the Brunel University, West London, UK, in 2006. Since May 2006, he is a Ph.D. candidate in medical physics at the University of Patras, Patras,

Greece. His research interests include medical image processing, image analysis, telemedical applications, and electromagnetic field variations prior to earthquakes.



Ioannis K. Kalatzis completed his BSc in Physics from the University of Athens, Greece in 1987 and his PhD in Medical Physics from the Medical School of the University of Athens in 2000. During his PhD thesis he was involved in tomographic image processing in the Nuclear Medicine Department of the Hippokrateio General Hospital of Athens, as well in research projects concerning image processing and pattern recognition applications in medicine. He is currently research fellow at the Laboratory of Medical Image and Signal Processing in the Department of Medical Instruments Technology of the Technological Educational Institution of Athens. Dr Kalatzis's research interests include medical image and signal processing and analysis, with focus in the field of pattern recognition.



George C. Nikiforidis received his Laurea on Physics and his M.Sc. in Atomic and Nuclear Physics both from the University of Milan, Italy in 1973 and 1980 respectively and his Ph.D. in Medical Physics from the University of Patras, Greece in 1981. He is currently a Professor of Medical Physics and the director of the Department of Medical Physics, University of Patras, Greece. He is also the Dean of the School of Health Sciences in the same institution. He is the director of the post-graduate course on Medical Physics, from the same institution. He has been the principal investigator or been involved in a variety of national or European research and development projects.