

Complementary DNA microarray image processing based on the Fuzzy Gaussian Mixture Model

Emmanouil I. Athanasiadis, Dionisis A. Cavouras, Panagiota P. Spyridonos, Dimitris Th. Glotsos, Ioannis K. Kalatzis, and George C. Nikiforidis

Abstract—The objective of this work was to investigate the segmentation ability of the Fuzzy Gaussian Mixture Model (FGMM) clustering algorithm, applied on complementary DNA (cDNA) images. Following a standard established procedure, a simulated microarray image of 1600 cells, each containing one spot, was produced. For further evaluation of the algorithm, three real microarray images were also used, each containing 6400 spots. For the task of locating spot borders and surrounding background in each cell, an automatic gridding process was developed and applied on microarray images. The FGMM and the Gaussian Mixture Model (GMM) algorithms were applied to each cell, with the purpose of discriminating foreground from background. The segmentation abilities of both algorithms were evaluated by means of the segmentation matching factor, coefficient of determination and concordance correlation, in respect to the actual classes (foreground-background pixels) of the simulated spots. Pairwise correlation and mean absolute error of the real images among replicates were also calculated. The FGMM was found to perform better and with equal processing time, as compared to the GMM, rendering the FGMM algorithm an efficient alternative for segmenting cDNA microarray images.

Index Terms—cDNA Microarrays, Fuzzy Gaussian Mixture Models, Segmentation.

I. INTRODUCTION

Microarray imaging is used for the concurrent identification of thousands of genes in the field of bioinformatics [1]. By finding the location of the spots in a complementary DNA (cDNA) microarray experiment, calculations of the mean fluorescence intensity value are obtained, that are closely related to the expression of a specific gene. Thus, a precise localization and outlining of a spot are essential to obtain a more accurate intensity

measurement, leading to a more precise expression measurement of a gene.

For the task of measuring spot intensity values, three major steps were followed [1]-[2]: 1/ the gridding step, for the precise localization of the cells, 2/ the segmentation step, for distinguishing each cell's foreground from background and 3/ the intensity extraction step, for calculating the mean fluorescence value of each spot.

In the past, several techniques and software packets have been developed for the task of processing microarray images [3]-[7]. In the ScanAlyze [3] software, a fixed circle segmentation method is used, where all spots are considered to be circular with a fixed predefined radius. In the GenePix [4] software, an adaptive circle segmentation technique is employed. According to that method, the radius of each spot is not constant but adapts to each spot separately. In the Spot [6] software, an adaptive shape segmentation technique is performed. In the latter technique, the most representative algorithms employed are the watershed [8] and the seeded region growing [9]. Nevertheless, in all those techniques, the disadvantages are either that spots are considered to be circular in shape or a-priori knowledge of the precise position of the spots' centers is a pre-requisite [10]. In the ImaGene [7] software, a histogram based segmentation method is applied, in which the values between the 80th and the 95th histogram percentile contribute to the calculation of the mean intensity value.

Segmentation algorithms based on the statistical Mann-Whitney test was also used [26], assessing the statistical significance difference between foreground and background. Markov random field (MRF) modeling for microarray image segmentation was also adopted [19]. However, initial segmentation of the foreground and background is essential and affects the final result [25]. Active contour techniques have been also used [27]. Nevertheless, snakes have no good performance in noisy images. Gaussian Mixture Model (GMM) algorithm has been also adapted to microarray images [15], providing promising results in segmenting spots.

The Fuzzy Gaussian Mixture Model (FGMM) clustering algorithm is an effective clustering technique that has found application in many areas of pattern recognition, such as in

Manuscript received March 28, 2007, revised July 12, 2007. E. I. Athanasiadis thanks the Greek State Scholarships Foundation (I.K.Y.), under Grand 4690/2005-2006, for funded the above work.

E. I. Athanasiadis, P. P. Spyridonos, D. Th. Glotsos 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, 26500 Rion - Patras, Greece. (phone: 0030-2610-997745 e-mail: mathan@upatras.gr).

D. A. Cavouras and I. K. Kalatzis are with 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: 0030-210-5385375; fax: 0030-210-5910975; e-mail: cavouras@teiath.gr.).

voice recognition [11]-[12], but it has not been applied so far in microarray images.

In the present study, the FGMM clustering algorithm was suitably adapted for processing microarray images and it was implemented with purpose to investigate the segmentation ability of the algorithm. Additionally, a comparison of the FGMM algorithm with the GMM algorithm, which has been shown to be effective in microarrays segmentation [13], was performed. Both methods, FGMM and GMM, were developed in MATLAB® [23]. Evaluation was performed by calculating the segmentation matching factors [14] of each algorithm in respect to the actual classes (foreground-background pixels) of the simulated spots. Additionally, coefficient of determination, as well as concordance correlation was also calculated with regard to the actual mean intensities values of the simulated spots. The performance of the two algorithms was also tested in three actual microarray images, each one consisting of 6400 spots, by calculating correlation and mean absolute error among the replicates.

II. METHODS AND MATERIAL

A. Automatic Gridding Process

A cDNA image consists of an arrayer, several sub-arrays, and thousands of spots corresponding to specific genes (see Fig. 1). Gridding is the procedure of segmenting each sub-array into numerous cells, each cell containing one spot and its background. Gridding renders the procedure of spot finding easier, since segmentation may be applied within each individual cell automatically. The gridding algorithm adopted in the present study consisted of the following steps [13]:

1/Determination of sub-array regions. First, average intensities were calculated along image rows and columns for both Red ('R') and Green ('G') channels (R for Cyanine Cy3 and G for Cyanine Cy5), thus, forming two signals for each channel. Second, noise suppression was performed by means of a low pass filtering mask [13]. Third, sub-array regions were determined by finding the local minima of either the R or the G signals in the horizontal and vertical axes [15]-[16]; multiple experimentation showed that by choosing either R, or G channels, no significant differences were observed in sub-array boundary localization.

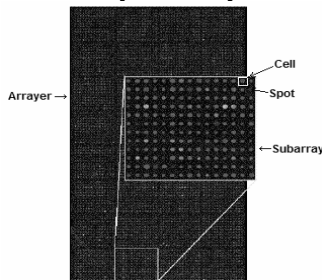
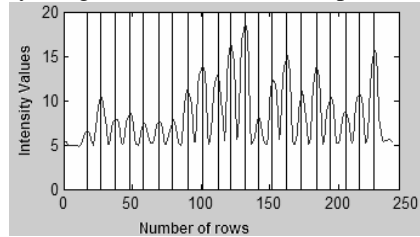
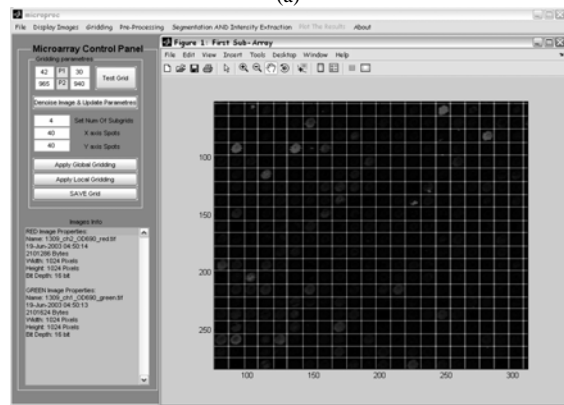


Fig. 1: An Arrayer consists of 8x4 sub-arrays, each sub-array of 19x21 cells and spots respectively.

2/Determination of cells. A similar procedure to sub-array determination was followed for identifying automatically the centers of the spots (local maxima) and the boundaries of the cells (local minima), employing an algorithm based on regional connectivity properties of pixels. Figure 2 demonstrates the automatic localization of the maximum signal values of an R-signal, as well as the result of the gridding step, applied to a sub-array region of a microarray image that contains 19x21 spots.



(a)



(b)

Fig. 2: (a) automatic localization of spot centers (local maxima using Matlab function 'imregionalmax') and (b) cell determination (local minima using Matlab function 'imregionalmin'), applied on a real cDNA 19x21 sub-array microarray image collected from the Davidson College [21].

B. Gaussian Mixture Models

Let $X = \{x_1, x_2, \dots, x_T\}$ be a sequence of T vectors with x_i intensity values, and $\theta = \{p_i, \mu_i, C_i\}$ for $i = 1, \dots, c$, be a set of parameters to be maximized, using the Expectation Maximization technique [13], where parameters p_i , μ_i and C_i correspond to mixture weights, mean vectors, and covariance matrixes respectively of a mixture of c Gaussian Distributions.

The major task of the Gaussian Mixture Model [13]-[15] algorithm is to maximize the likelihood function that is described by (1):

$$p(X | \theta) = \prod_{t=1}^T p(x_t | \theta) \quad (1)$$

where the likelihood function of each vector $p(x_t | \theta)$ is computed by (2) and (3) as a probability density function of multiple Gaussians:

$$p(x_t | \theta) = \sum_{i=1}^c p(x_t, i | \theta) = \sum_{i=1}^c p_i p(x_t | i, \theta) \quad (2)$$

$$p(x_t | i, \theta) = \frac{\exp\left\{-\frac{1}{2}(x_t - \mu_i)' C_i^{-1} (x_t - \mu_i)\right\}}{(2\pi)^{d/2} |C_i|^{1/2}} \quad (3)$$

where $(x_t - \mu_i)'$ is the transpose matrix of $(x_t - \mu_i)$, C_i^{-1} is the inverse matrix and $|C_i|$ is the determinant of the covariance matrix C , for each class i , and d is the number of features used.

For the maximization of $p(X_t | \theta)$, an auxiliary function Q , named expected log-likelihood function [22], may be estimated employing (4):

$$Q(\theta, \bar{\theta}) = \sum_{t=1}^T p(i | x_t, \theta) \log[\bar{p}_i p(x_t | i, \bar{\theta})] \quad (4)$$

where $\bar{\theta} = \{\bar{p}_i, \bar{\mu}_i, \bar{C}_i\}$ are computed by using (5)-(7) (Maximization Step).

$$\bar{p}_i = \frac{1}{T} \sum_{t=1}^T p(i | x_t, \theta) \quad (5)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T p(i | x_t, \theta) x_t}{\sum_{t=1}^T p(i | x_t, \theta)} \quad (6)$$

$$\bar{C}_i = \frac{\sum_{t=1}^T p(i | x_t, \theta) (x_t - \mu_i)(x_t - \mu_i)'}{\sum_{t=1}^T p(i | x_t, \theta)} \quad (7)$$

The posterior probabilities $p(i | x_t, \theta)$ are then recomputed by using (8) (Estimation Step).

$$p(i | x_t, \theta) = \frac{\bar{p}_i p(x_t | i, \bar{\theta})}{\sum_{k=1}^c \bar{p}_k p(x_t | k, \bar{\theta})} \quad (8)$$

Equation (4) is then calculated, and the process is repeated until there is no significant change in Q .

C. Fuzzy Gaussian Mixture Models

In the Fuzzy Gaussian Mixture Models FGMM [11]-[12] algorithm, a modification of the Fuzzy C Means FCM clustering technique [24] is performed for the estimation of parameters θ of GMM. In the FCM algorithm, the objective is to minimize (9).

$$J(U, \theta) = \sum_{t=1}^T \sum_{i=1}^c u_{it}^m d_{it}^2 \quad (9)$$

where $U = \{u_{it}\}$ is a fuzzy c -partition of the initial vector X , u_{it} is the probability of vector x_i of belonging to class i , m is the fuzzy membership of u_{it} , called the degree of fuzziness, and d is a dissimilarity measurement [17], defined by (10).

$$d_{it}^2 = -\log p(x_t, i | \bar{\theta}) = -\log \bar{p}_i p(x_t | i, \bar{\theta}) \quad (10)$$

Substituting (10) to (9), (11) is derived:

$$J(U, \bar{\theta}) = -\sum_{t=1}^T \sum_{i=1}^c u_{it}^m \log \bar{p}_i p(x_t | i, \bar{\theta}) \quad (11)$$

Minimization of (11) may be accomplished by using Lagrange multiplier methods [18]. Calculation of the new fuzzy parameters is achieved by using (12) – (14) (Minimization Step).

$$\bar{p}_i = \frac{\sum_{t=1}^T u_{it}^m}{\sum_{i=1}^c \sum_{t=1}^T u_{it}^m} \quad (12)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T u_{it}^m x_t}{\sum_{t=1}^T u_{it}^m} \quad (13)$$

$$\bar{C}_i = \frac{\sum_{t=1}^T u_{it}^m (x_t - \mu_i)(x_t - \mu_i)'}{\sum_{t=1}^T u_{it}^m} \quad (14)$$

After the calculation of the new parameters, matrix U is recomputed according to (15) (Estimation Step).

$$u_{it} = \left[\sum_{k=1}^c \left(\frac{d_{it}}{d_{kt}} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (15)$$

Equation (11) is then calculated, and the process is repeated until there is no significant change in J .

D. Segmentation Process

For every cell determined by the gridding process, the following procedure was performed:

1/ The $N \times M$ cell image, considering R and G channels separately, was converted into a vector X , with dimensions $l \times NM$.

2/ The clustering algorithms (GMM and FGMM) were separately applied to the vector X , in order to discriminate the data into two categories or classes ($c=2$), the foreground (FG) and background (BG) class, denoted by zeros and ones, respectively. Next, the binary vector was transformed into a binary cell, following the inverse procedure.

E. Intensity Extraction

Representative spot intensity was obtained by subtracting the mean of the FG from the BG, according to (16).

$$I = \mu_{FG} - \mu_{BG} \quad (16)$$

Where μ_{FG} and μ_{BG} are the mean foreground and mean background respectively, both calculated from the corresponding labeled cell pixels.

F. Material

For the numerical evaluation of the clustering ability of the two segmentation algorithms, a simulated cDNA image was produced as described in literature [19]-[20]. In order to generate spots with realistic characteristics, the following procedure was adopted. A true cDNA image, consisting of 1600 spots, was used as a template, and its binary version was produced by employing a thresholding technique [19] (see Fig. 5). In the simulated image, the location as well as the area of each spot was a-priory known. The mean intensity value of each spot was pre-defined, ranging between 0 and $2^{16}-1$ for both the R and G channels [19]. Spot intensities were produced using an exponential distribution with mean value the pre-defined mean intensity value (using Matlab function ‘expnd’ and ‘expfit’). Background intensities were drawn from a single exponential distribution, with mean value determined from the true cDNA image’s mean intensity background [19].

To investigate the algorithms’ performance in presence of noise, the simulated image was corrupted with additive white Gaussian noise [15], ranging the Signal to Noise Ratio (SNR) from 1dB to 20 dB.

The accuracy of segmentation was numerically calculated using the segmentation matching factor (SMF) (17) [14] for every binary cell, produced by the clustering algorithm.

$$SMF = \frac{A_{cal} \cap A_{real}}{A_{cal} \cup A_{real}} \quad (17)$$

where A_{cal} 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 whereas, a score less than 50% indicates poor segmentation [14].

Additionally, measurements based on regression analysis, coefficient of determination r^2 and concordance correlation p_c (18) were calculated [19].

$$p_c(A, B) = \frac{2S_A S_B r}{S_A^2 + S_B^2 + (\bar{A} - \bar{B})^2} \quad (18)$$

where A and B are two samples, \bar{A} and \bar{B} are the mean values, S_A and S_B are the standard deviation of the samples. The higher the p_c value, the better the performance of the algorithm.

Mean Square Error (MSE) [15] for each calculated intensity value, estimated by each of the two methods in respect to the actual intensity value, for all 1600 simulated spots was performed.

Additionally, to investigate the performance of the two algorithms, as well as to confirm the results obtained by using the simulated images, three real microarray images [21] were also used. In these images, the global pattern of gene expression was remarkably stable [25][28]. Logarithmic intensity ratios were calculated using a custom made Graphic User Interface (GUI) program, developed in


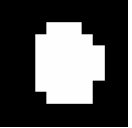



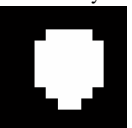


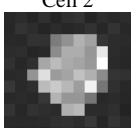

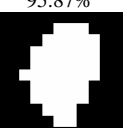
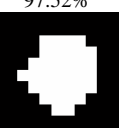
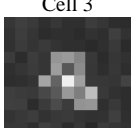
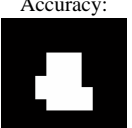
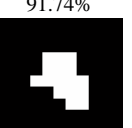

MATLAB for the purposes of the present study. Logarithmic ratios were normalized with LOWLESS method [1].

III. EXPERIMENTAL RESULTS AND DISCUSSION

Following the gridding procedure, the GMM and FGMM algorithms were applied identifying the two classes, the foreground and background pixels.

The degree of fuzziness m of FGMM algorithm is a free parameter and it was experimentally adjusted by measuring the SMF. Ranging the fuzzy parameter from 1.2 to 5, SMF measurements were obtained. An example of 3 randomly chosen cells is illustrated in Fig. 3. It becomes evident, that setting the degree of fuzziness greater than 3 does not affect significantly the algorithm behavior. Subsequent validation results were obtained using m equal to 3.

TABLE I
SEGMENTATION RESULTS FOR 4 DIFFERENT CELLS

Original Cells	Actual Boundaries	GMM Result	FGMM Result
			
Cell 1	Accuracy:	87.60%	100.00%
			
Cell 2	Accuracy:	95.87%	97.52%
			
Cell 3	Accuracy:	91.74%	100.00%
			
Cell 4	Accuracy:	95.87%	96.69%

Comparative results for 4 different cells obtained from the G channel of the simulated microarray. The first column indicates the simulated spot with the surrounding area, the second column indicates the actual boundaries of the spot and the third and the forth columns present the segmentation results of the GMM and FGMM algorithms as well as the corresponding matching factors ($c=2, m=3$).

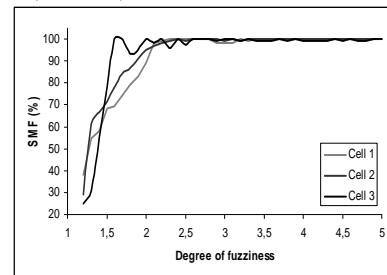


Fig. 3: SMF Results achieved by FGMM at three different cells, by adjusting the degree of fuzziness from 1.2 to 5.

TABLE II

MEAN SMF , CORRELATION COEFFICIENT, CONCORDANCE CORRELATION AND MEAN SQUARE ERROR CALCULATIONS IN SIMULATED DATA

SNR (dB)	SMF GMM (%)	SMF FGMM (%)	GMM r^2	FGMM r^2	GMM p_c	FGMM p_c	MSE GMM	MSE FGMM
1	63.73	66.07	0.89	0.95	0.83	0.86	4.8076	4.6879
2	65.26	69.00	0.90	0.95	0.83	0.86	4.7883	4.6720
3	68.59	73.50	0.90	0.96	0.83	0.87	4.8121	4.6678
4	71.66	76.07	0.90	0.97	0.83	0.88	4.8103	4.6264
5	74.20	78.84	0.91	0.98	0.84	0.90	4.7929	4.5446
6	77.25	81.92	0.91	0.98	0.85	0.91	4.7782	4.4839
7	79.39	84.48	0.91	0.98	0.85	0.93	4.8018	4.4297
8	81.01	87.05	0.91	0.99	0.87	0.94	4.7254	4.328
9	82.34	89.62	0.91	0.99	0.87	0.95	4.7825	4.2465
10	83.27	91.37	0.91	0.99	0.87	0.97	4.7115	4.1023
11	83.58	92.94	0.91	0.99	0.89	0.97	4.7272	3.9999
12	83.64	93.92	0.91	0.99	0.89	0.98	4.7010	3.8437
13	83.86	94.90	0.91	0.99	0.91	0.99	4.6951	3.7375
14	83.92	95.07	0.91	0.99	0.91	0.99	4.6480	3.5864
15	83.98	95.97	0.91	0.99	0.92	0.99	4.7051	3.4497
16	83.80	97.29	0.91	0.99	0.91	0.99	4.6622	3.2798
17	82.12	97.34	0.91	0.99	0.93	0.99	4.7155	3.5993
18	82.68	97.17	0.91	0.99	0.93	0.99	4.6900	3.3896
19	83.27	97.93	0.91	0.99	0.93	0.99	4.7397	3.7335
20	82.59	98.03	0.91	0.99	0.93	0.99	4.7670	3.2370

The mean SMF, the coefficient of determination, the concordance coefficient and MSE achieved by each of the two different techniques in R channel for all 1600 simulated spots, in respect to different SNR levels.

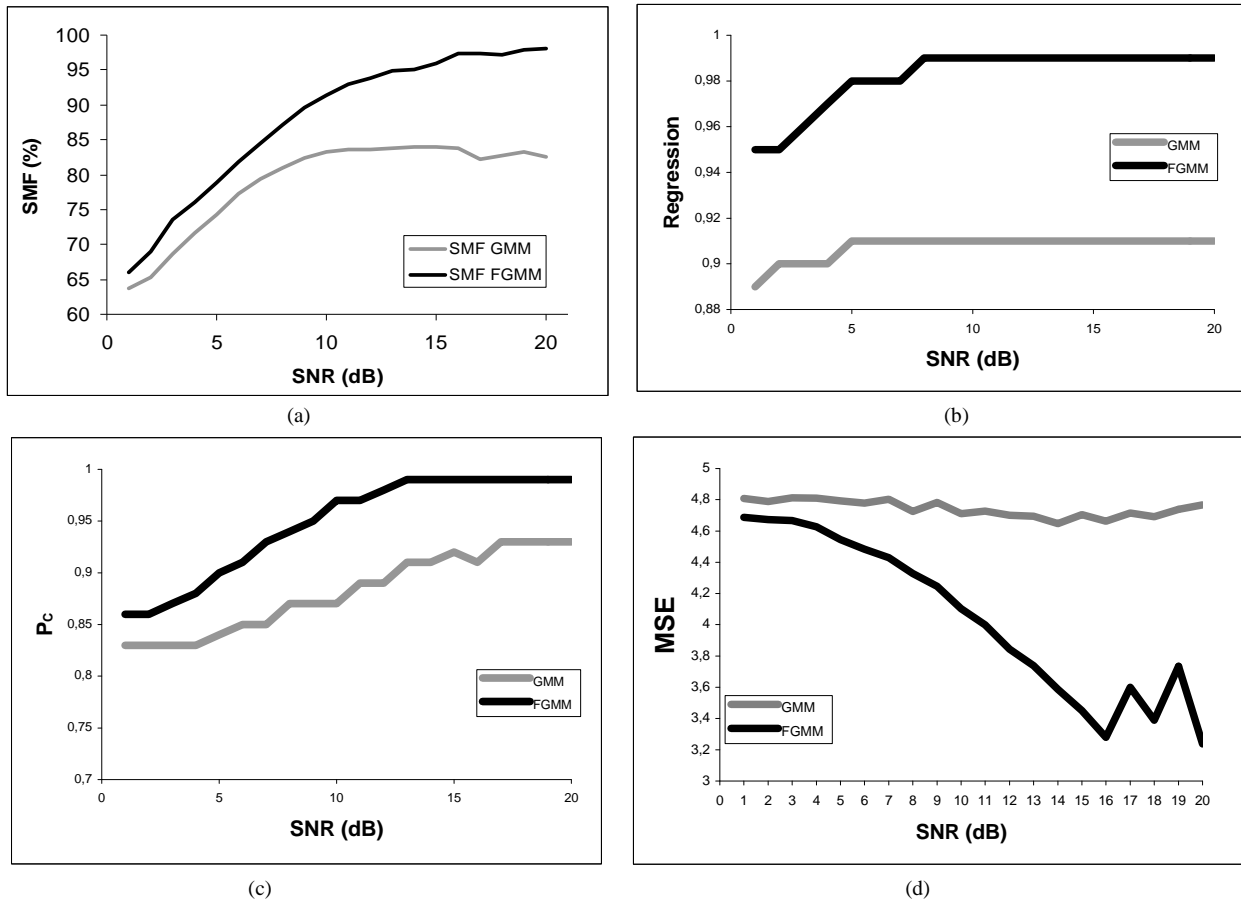


Fig. 4: (a) SMF, (b) regression coefficient, (c) concordance correlation and (d) MSE calculated by using GMM and FGMM algorithms in respect to additive white Gaussian noise with different SNR.

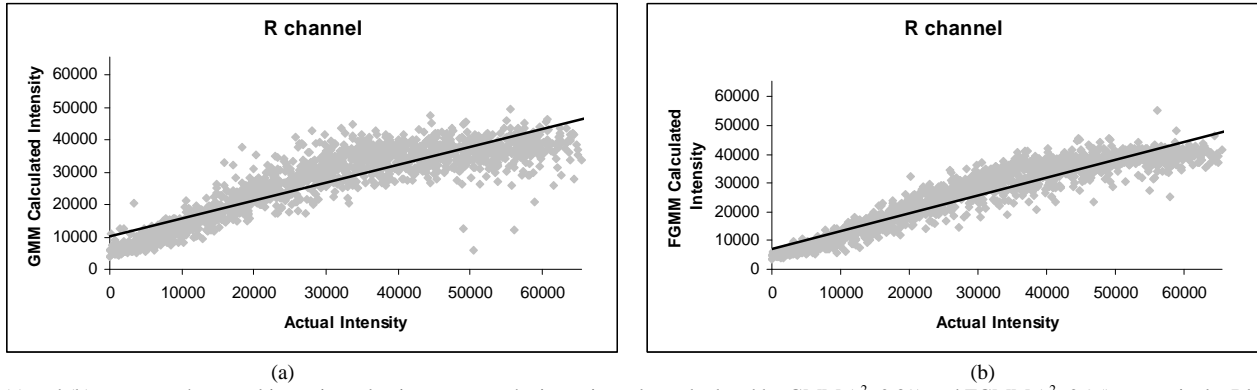


Fig. 5 (a) and (b) represent the actual intensity value in respect to the intensity value calculated by GMM ($r^2=0.89$) and FGMM ($r^2=0.95$) respectively. Black line is the linear regression line. The SNR of the additive white Gaussian noise was set to 1dB in the present example.

Table II summarizes the segmentation performance results for the simulated data for both algorithms. The mean SMF, the coefficient of determination, the concordance coefficient and MSE achieved for all simulated spots, are presented in respect to different SNR levels (Fig. 4(a)-(d)).

Regarding the SMF, the FGMM algorithm resulted in higher spot area identification accuracy than the conventional GMM (Table II, Fig. 4(a)). An illustrative example of spot area detection is given in Table I, for 4 different randomly chosen cells.

The ultimate goal of the segmentation process in microarray imaging is to obtain an intensity measurement. Accurate delineation of spot has a great impact on the intensity calculation. To this end, metrics based on the intensity, rather than the segmentation area were also extracted. Coefficient of determination, concordance correlation and MSE measurements, support the superiority of FGMM against GMM, when evaluating the results in intensity extraction perspective (Table II).

The higher the coefficient of determination and the concordance correlation values, the better the performance of the algorithm is [19].

Coefficient of determination for the R channel with SNR equal to 1 dB is illustrated in Fig. 5(a) and 5(b) using the GMM and FGMM algorithm respectively. In this particular example, the obtained r^2 was equal to 0.89, using the GMM algorithm whereas using the FGMM approach the resulted r^2 was equal to 0.95 (Table II). It is clear that the fuzzy

approach of the GMM is more tolerant to noise, providing better results compared to the conventional GMM approach.

Furthermore, the comparison of the algorithms was carried out in a set of three real microarray images, concerning *Saccharomyces cerevisiae*, obtained from publicly available database [21]. The software used for the creation of microarray images was the ArrayMaker Version 1.8.5. Additionally, a detailed description about the scanner, as well as instructions for the software used for the creation of the microarray images, is also available at: <http://cmgm.stanford.edu/pbrown/scanner.html>. Each spot should have the same gene expression ratio throughout replication experiments, thus correlation between each experiment should be maximal [25]. Moreover, pairwise mean absolute error (MAE) was calculated among the replicates. Box plots for correlation and MAE are illustrated in fig. 6(a) and 6(b) respectively. It should be noted that the higher the correlation and the lower the MAE values, the better the performance of the segmentation algorithm [25].

The FGMM algorithm achieved a correlation of 0.97 and MAE close to 2×10^4 . The GMM algorithm resulted in a lower correlation value that of 0.90 and a higher MAE of the order of 5×10^4 .

The results obtained by the real microarray images verify the simulated experiments and support the performance superiority of the FGMM algorithm in segmenting DNA images.

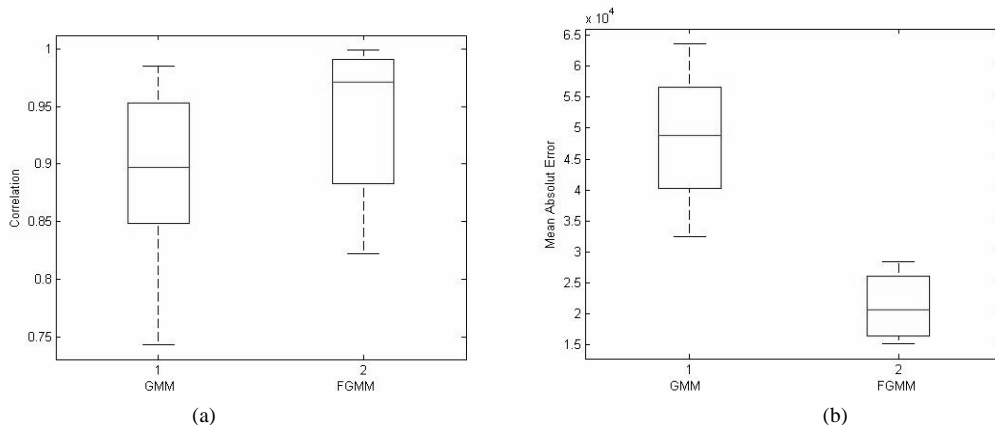


Fig 6: Box plots that illustrate (a) the correlation and (b) the MAE among the replicates.

I. CONCLUSIONS

In the present study, the FGMM clustering technique is proposed to model complicated cell microarray image distributions. The FGMM approach proved more accurate in spot delineation and intensity computation than the conventional GMM algorithm, thus, providing more reliable means for estimating gene expression on cDNA microarray images.

Errors during segmentation are being propagated to the intensity extraction step, leading to inaccurate intensity ratio calculations degrading thus the gene expression results. The use of an accurate and robust segmentation algorithm, will significantly contribute in more reliable measurements, improving the interpretation of gene expression results.

It should be noted that the processing time of the GMM was identical to the FGMM algorithm, providing the FGMM algorithm an efficient alternative for segmenting cDNA microarray images.

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] S.Beucher, F. Meyer, "The morphological approach to segmentation: The watershed transformation", *Optical Engineering*, Vol. 34, pp. 433-481, 1993.
- [9] R. Adams and L. Bischof, "Seeded Region Growing", *IEEE Trans. Pattern Anal. Machine Intell.*, vol 16, pp 641-647, 1994.
- [10] D. Bozinov and J. Rahenfuhrer, "Unsupervised technique for robust target separation and analysis of DNA Microarray spots through adaptive pixel clustering", *Journal of Bioinformatics*, vol. 18, pp 747-756, 2002.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] Betal D, Roberts N, Whitehouse GH., "Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology". *Br. J. Radiol.*:70(837):903-17, 1997.
- [15] 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.
- [16] S.Lonardi and Y. Luo, "Gridding and Compression of Microarray Images". *IEEE Computational Systems Bioinformatics Conference CSB*, 2004.
- [17] James C. Bezdek , "Pattern Recognition with fuzzy objective function algorithms", Plenum Press, New York and London, 1987.
- [18] X.D. Huang, Y. Ariki, and M.A. Jack, "Hidden Markov models for speech recognition", Edinburgh University Press, 1990.
- [19] O. Demirkaya, M. H. Asyali and M.M. Shoukri, "Segmentation of cDNA Microarray Spots Using Markov Radom Field Modeling", *Bioinformatics*, vol. 21 no. 13, pp. 2994-3000, 2005.
- [20] Y. Balagurunathan, E. R. Dougherty, Y. Chen, M.L. Bittner and J.M. Trent, "Simulation of cDNA Microarray via a parameterized random signal model", *Journal of Biomedical Optics*, vol. 7, pp. 507-523, 2002.
- [21] Laurie Heyer, MicroArray Genome Imaging & Clustering (MAGIC) Tool, Davidson College.
Available: <http://www.bio.davidson.edu/projects/magic/magic.html>
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm" *J. Roy. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.
- [23] The MathWorks, Inc. Software, MATLAB®.
- [24] James C. Bezdek, "Pattern Recognition with fuzzy objective function algorithms", Plenum Press, New York and London, 1987.
- [25] Antti Lehmussola, Pekka Ruusuuvuori and Olli Yli-Harja, "Evaluating the performance of microarray segmentation algorithms", *Bioinformatics*, Vol 22, 2910-2917, 2006.
- [26] Y. Chen, et al, "Ratio-based decisions abd the quantitative analysis of cDNA microarray images", *J. Biomed. Opt.*, 2, 264-374, 1997.
- [27] T Srinark, C Kambhamettu , "a microarray image analysis system based on multiple snakes", *Journal of Biological Systems Special Issue*, 2004
- [28] 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, pg 680, 1997.
- [29] Stanford University, School of Medicine
Available: <http://cmgm.stanford.edu/pbrown/scanner.html>