Original Article

www.jmss.mui.ac.ir

Fully Automated Complementary DNA Microarray Segmentation using a Novel Fuzzy-based Algorithm

Hamidreza Saberkari, Sheyda Bahrami, Mousa Shamsi, Mohammad Javad Amoshahy, Habib Badri Ghavifekr, Mohammad Hossein Sedaaghi

Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran

Submission: 15-04-2015 Accepted: 13-06-2015

ABSTRACT

DNA microarray is a powerful approach to study simultaneously, the expression of 1000 of genes in a single experiment. The average value of the fluorescent intensity could be calculated in a microarray experiment. The calculated intensity values are very close in amount to the levels of expression of a particular gene. However, determining the appropriate position of every spot in microarray images is a main challenge, which leads to the accurate classification of normal and abnormal (cancer) cells. In this paper, first a preprocessing approach is performed to eliminate the noise and artifacts available in microarray cells using the nonlinear anisotropic diffusion filtering method. Then, the coordinate center of each spot is positioned utilizing the mathematical morphology operations. Finally, the position of each spot is exactly determined through applying a novel hybrid model based on the principle component analysis and the spatial fuzzy c-means clustering (SFCM) algorithm. Using a Gaussian kernel in SFCM algorithm will lead to improving the quality in complementary DNA microarray segmentation. The performance of the proposed algorithm has been evaluated on the real microarray images, which is available in Stanford Microarray Databases. Results illustrate that the accuracy of microarray cells segmentation in the proposed algorithm reaches to 100% and 98% for noiseless/noisy cells, respectively.

Key words: *Algorithms, array sequence analysis, breast cancer, fuzzy clustering, gene expression, noise*

INTRODUCTION

The discovery of microarray technology in 1995 made it possible for researchers to analyze 1000 of gene expression levels simultaneously.[1] This can be very useful in identifying genetic diseases in different molecular levels. This technology has been exploited in many clinical applications in recent decades, some of these applications are:[2,3]

- Cancer: To determine the differences between normal and abnormal cells, classification of tumors and identifying the risk factors
- Pharmaceutical: To determine the relation between gene expression profiles and their response to various drugs
- Toxicology: To determine the relation between response to various toxins and deviations made in different tissues of genetic profiles in facing different toxins.

Complementary DNA (cDNA) microarray contains thousands of individual DNA strands, which is printed by robotic arrayer on the high-density array (often on a glass slide). In general, DNA experiment consists of these steps as follows:[4]

- Probe preparation: The first step in DNA microarray was to prepare the DNA glass slide for every patient. To do this, a DNA strand corresponding to a specific gene has been assigned to each spot. A sample of glass slide with 17×25 spots is shown in Figure 1
- Target and reference sample preparation: In this step, messenger RNA samples related to the test and the reference strands are isolated. These two samples are reversely transcripted, and the cDNA is extracted. Then, it is amplified using the polymerase chain reaction, and finally, it is labeled. Labeling often is carried out by two fluorescent dyes as red Cy5 and green Cy3
- Hybridization: In this step, two labeled samples are mixed together on the substrate and are placed on the single strand DNA, which are related to the specified DNA strands
- Washing the glass slides
- Detection: In this step, the intensities of red and green dyes are calculated using a scanner, which leads to create two images of microarray.

The raw data of cDNA microarray are often stored as a 16-bit image in the tagged image format for each dye. Different

Address for correspondence:
Hamidreza Saberkari, Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. E-mail: h_saberkari@sut.ac.ir

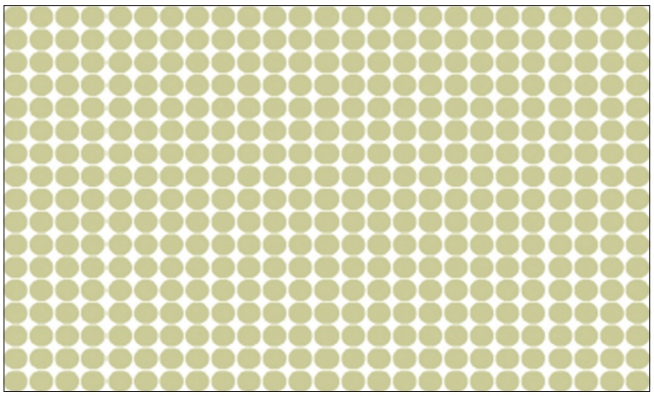

Figure 1: A sample glass of DNA microarray

dyes absorb and emit lights in a different wavelength, that is, Cy5 and Cy3 emit lights in wavelengths of 510–550 nm and 630–660 nm, respectively.

Microarray images contain multiple blocks (which are referred to sub-networks) and these sub-networks consist of many spots, which are situated in rows and columns. In general, there are four main steps to analyze the microarray images:[5]

- Preprocessing: This step is used to eliminate the background noise and artifacts
- Gridding: Determining the position of each spot and also the center of the coordinates is carried out in this step
- Segmentation: Pixels in the microarray images are classified into foreground (spot) and background
- Determining the gene expression levels: In this step, assigned pixels to each spot are used to determine the gene expression levels.

Among the aforementioned steps, segmentation plays a key role in analyzing of these images. In recent decades, the different software was introduced in the literature for segmentation. Some of these packages are Scanalyze: a system for aligning and merging range data was the primary tool used in the Digital Michelangelo Project (1997-2004),[6] Dapple: DAPPLE was developed by Liz Rossin in the lab of Mark J Daly,[7] ImaGene: ImaGene® 9.0 from BioDiscovery is the leading-edge microarray analysis software that offers optimal performance with the broadest and most refined features available today,[8] and SpotFinder.[9] The main problem of this software was; all the parameters should be set manually, as well as the center of spot should be situated by human intervention, which has a negative influence on the analysis of gene expression levels. In order to overcome the aforementioned drawbacks, different methods are proposed, which can be divided into three categories:[5]

- Shape-based segmentation methods: Shape-based methods are based on the specific shape of spots. There are two conventional methods for this approach; the fixed circle segmentation algorithm implemented by scanAlyze[6] software and adaptive circle segmentation algorithm presented by Buhler.[7] In these two methods, a circular template is placed on each spot. Sarder *et al*.[10] proposed the parametric circular technique with the elliptic centers for segmentation of spots in noisy microarrays. The shape-based methods suffer from not being able to segment the noncircular spots
- Shape-independent segmentation methods: These methods relieve the main problem of shape-based methods. Seeded growing regions (SRG) algorithm used in spot software for segmentation of irregular spots is a common approach proposed.[11] In this algorithm, a collection of seeds is considered to each cell for the first step. In the iterative procedure, similar neighborhood pixels are considered as spots. The main restriction of SRG algorithm is its high sensitivity on the selection of suitable seed. In,[12-15] two methods have been proposed in order to assign the pixels belonging to the spots and backgrounds named K-means and the hybrid k-means method, respectively. These methods will show poor performance if their spots have low contrast. The other method, which is based on clustering the pixel values, utilizes elimination method in order to eliminate the nonconnected clusters. In this approach, small clusters are considered as artifacts. However, the dimension of each cluster should be manually adjusted, which is the main problem of this method. In,[16-19] active contour and multiple snake methods are proposed. However, both of them suffer from their inappropriate performance in noisy images
- Hybrid methods: In these methods, a set of information of the intensity and neighborhood are combined together. In,[20] Markov random field (MRF) approach has been proposed. In[21] Gottardo presented a method based on the MRF, in which the intensity of the background and foreground are shown by *t*-distribution. Nagarajan proposed another method, which the segmentation procedure of each spot is conducted based on the correlation of statistical information of spots.[22] This method does have a good performance in diagnosing of low intensity spots. However, its operation will be restricted when the microarray image has a low quality. In other method proposed by Zacharia, a novel algorithm based on the combination of the three-dimensional (3D) modeling of spots and genetic algorithm is utilized for the segmentation task.[23] In this method, each spot in cDNA microarray is considered as 3D model through the solving an optimization problem using the genetic algorithm. This method is efficient for noisy images. However, the main problem is the high computational complexity because of using the genetic algorithm.

In this paper, a novel fuzzy-based algorithm for spot segmentation in microarray cells has been proposed. In recent years, fuzzy clustering algorithms have been exploited for segmentation task in the microarray images.

These algorithms have some advantages: (1) They are able to distinguish spots, which have low intensity (because of using the neighborhood information for each spot), (2) these algorithms are almost independent from noises, and (3) spot diagnosis is done more accurately. The main problem of fuzzy clustering algorithms is their high computational complexity due to update the membership functions in each iteration. To resolve this drawback, first the main informative components of each microarray cell are extracted using the principle component analysis (PCA) algorithm. The rest of the paper is organized as follows:

In Section II, the dataset used in this paper is introduced. The block diagram of the proposed algorithm is expressed in Section III, and different steps are explained in more details. Simulation results are interpreted in Section IV. Finally, Section V includes the conclusion of the paper.

DATA SET USED IN THIS PAPER

In this paper, the Stanford Microarray Database[24] was utilized to evaluate the performance of proposed algorithm. The sub-networks of the microarray contain 576 spots, which create an image with 24 × 24 rows and columns (the total pixels are 112896). In addition, annotation images are extracted by means of a constant radius circle and are used as a ground truth images. The Binary versions are produced for all of the images. Inner/outer pixels of the binary images have been indicated the signal pixels (spot)/ the background, respectively.

PROPOSED ALGORITHM

An efficient algorithm based on the spatial fuzzy clustering approach has been proposed for microarray images segmentation. Figure 2 shows the block diagram of our proposed algorithm. First, a preprocessing step is applied in order to eliminate of noise and artifacts, which leads to improve the image quality. Then, mathematical morphology operations are applied in order to grid the image into sub-networks. Finally, spatial fuzzy clustering algorithm for each sub-grid region is performed.

Preprocessing

In this paper, a method based on the nonlinear diffusion filtering has been proposed to background noise suppression in the both red and green channels of a microarray image. This method has a physical inspiration, originated from mass and heat transfer rules. It is used to create a balance in concentration deviation between two environments. We can model the spots and also the noise pixels in microarray image as a concentration and a little inhomogeneity in density, respectively. Noise inhomogeneity is smoothed by utilizing the diffusion law,

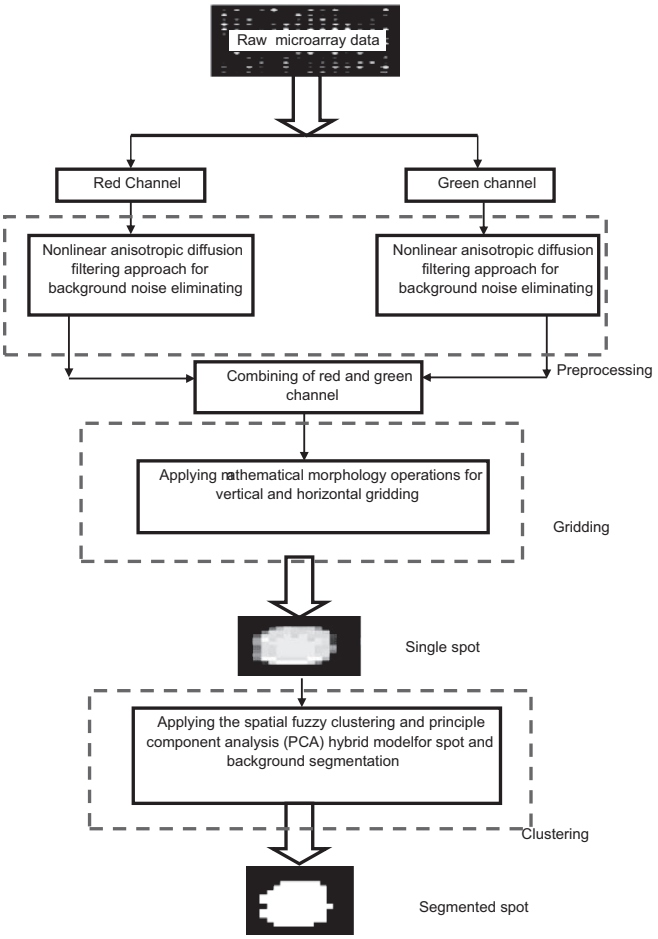

Figure 2: Block diagram of the proposed algorithm

and this phenomenon leads to reducing of noise in the microarray image.[25]

Gridding

After denoising, we should determine the coordinate of each spot and create a gridded image. A simple method for gridding proposed in this paper is the projection of microarray image in the length of rows and columns based on the morphology reconstruction operations. Assume, one sub-network of cDNA microarray is shown as $f = \{a_{xy}\}$ in which $x \in [1,h]$ and $y \in [1,w]$. Gridding procedures based on the morphology operation are as follows:[26]

- Calculation of vertical and horizontal projection signals:

$$H(y)=\sum_{x=0}^{w-1} f(x,y)$$
$$V(x)=\sum_{y=0}^{h-1} f(x,y) \quad (1)$$

- Filtering of horizontal projection signal using of morphological reconstruction operations:

$$H^{rec}(i)=\gamma^{rec}\left(H(i),H_{\eta}(i)\right) \quad (2)$$

where,

$$H_\eta(i)=H(i)-\bar{H}$$
$$\bar{H}=\frac{1}{w}\sum_{i=1}^{w}H(i) \quad (3)$$

and γ shows opening morphology operation defined as:

$$\left.\begin{aligned}&Opening:\gamma_B(f(x))=\delta_B[\varepsilon_B(f)]\\&Erosion:\varepsilon_B(f(x))=\inf_{y\in B}\{f(x-y)\}\\&Dilation:\delta_B(f(x))=\sup_{y\in B}\{f(x-y)\}\end{aligned}\right\}\ f:D_f\to T \quad (4)$$

In Eq. (4), f is the amounts of gray level of image in point (x, y), D_f is a subset of Z^2 and T is a set of gray levels:

- Determining the quantity of residual signal:

$$H_r(i)=H(i)-H^{rec}(i) \quad (5)$$

- Estimating the optimal value of threshold (t_H) which is defined as below:

$$t_H=\frac{1}{2}\cdot\frac{1}{w}\sum_{i=1}^{w}H_r(i) \quad (6)$$

- Obtaining the square value of the binary signal by t_H and finding the border lines on the right and left sides of each interval
- Calculating the middle of each interval of binary signals and drawing the straight lines.

Vertical gridding is exactly the same way as the method explained above for the horizontal gridding.

In Figure 3a, a sample of microarray image has been shown, which is related to a patient suffered from breast cancer. Figure 3b and e are the red and green channels extracted from the original image, respectively. The results of horizontal projection by applying mathematical morphology operations and also the gridded network of the microarray image have been shown in Figure 3d-f, respectively. As can be seen in Figure 3g, the coordinate of each spot in every sub-network is calculated very well.

Segmentation

In this paper, the hybrid spatial fuzzy clustering algorithm and PCA have been proposed for the segmentation of the each cell. The application of Fuzzy C-means clustering together with PCA gives more specific segmentation results for microarray images. The fuzzy clustering algorithm assigned pixels to the classes by means of membership functions. Assuming $X = (x_1, x_2,...x_N)$ is an image by N pixels that belong to the c class (cluster). The fuzzy clustering algorithm is segmented an image by minimizing the cost function. The cost function is defined as:[27]

$$J_m=2\sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m\|x_k-v_i\|^2 \quad (7)$$

where, $V = [v_1, v_2,...v_c]$ is a vector indicates the center of clusters and N is the number of inputs. μ_{ij} indicates the membership of x_j pixel in i^{th} cluster. In this paper, we used the Gaussian kernel function. The Gaussian kernel has two adjustable parameters named the center of each cluster and the Gaussian kernel parameter. We chose the Gaussian kernel parameter equal to 150. Then, for each cluster center, we calculated the Euclidean distance between each data point to the cluster center. Furthermore, in Eq. 7, m is a constant that control the degree of fuzziness of partitions. In our proposed algorithm, the amount of m varies

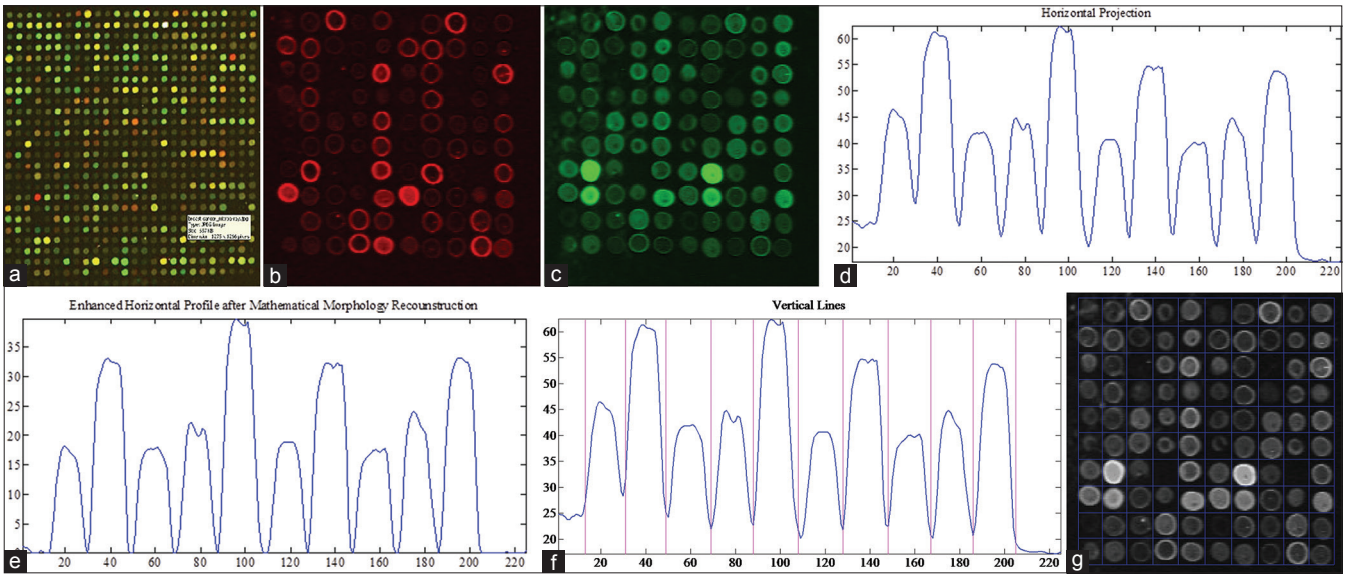

Figure 3: Results of the mathematical morphology in the gridding of breast cancer microarray image. (a) Microarray image includes the red and green channels. (b and c) Extraction of the red and green channels in the microarray image. (d) Extraction of the one-dimensional vertical projection signal. (e) Reconstruction of the vertical projection of signal using the mathematical morphology operations. (f) Drawing of the horizontal lines. (g) Final gridded image

from 1.2 to 4. The norm of Euclidean distance ($\|.\|$) which is defined as the distance between of pixels and the mean of clusters is calculated as below:

$$D_{ikA}^2 = \|x_k - v_i\|_{A_D}^2 = (x_k - v_i)^T A_D (x_k - v_i) \quad (8)$$

In which A_D is a diagonal matrix:

$$A_D = \begin{bmatrix} \left(1/\sigma^2\right) & 0 & \cdots & 0 \\ 0 & \left(1/\sigma^2\right) & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \left(1/\sigma^2\right) \end{bmatrix} \quad (9)$$

The deviation of x_k from v_i can be seen from Eq. 9. Minimizing the cost function is a nonlinear problem. There are many approaches proposed in the literatures for minimizing this function such as a genetic algorithm and grouped coordinates. However, the most common method is the Picard iteration loop by primary condition. Using the Lagrange multipliers, we can rewrite the cost function as follows:

$$J(X;U,V) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ij})^m D_{ikA}^2 + \sum_{k=1}^{N} \lambda_k \left(\sum_{i=1}^{c} \mu_{ik} - 1 \right) \quad (10)$$

By considering Eq. 10 and $D_{ikA}^2 > 0, \forall i,k$, $m > 1$ the cost function will be minimized. Then membership function and the center of the clusters are updated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{D_{ikA}}{D_{jkA}} \right)^{\frac{2}{m-1}}} \qquad 1 \le i \le c,\ 1 \le k \le N$$
$$v_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m x_k}{\sum_{k=1}^{N} (\mu_{ik})^m} \qquad 1 \le i \le c \quad (11)$$

Eq. 11 indicates the v_i as a weighted average of the points of the data belonging to clusters. The weights can be considered as the membership matrix.

One of the most important characteristics of the microarray images is the strongly correlation of one pixel by the neighbor pixels. In other words, these neighbor pixels have a similar property, so the probability of assigning a special pixel to the neighbor pixels is very high. In conclusion, the necessity of a spatial parameter in the segmentation algorithms based on the intensity becomes apparent. We utilized a spatial parameter h_{ij} in the proposed clustering algorithm and named. The spatial parameter in the spatial fuzzy clustering algorithm can be defined as:

$$h_{ij} = \sum_{k \in N_{x_j}} \mu_{ik} \quad (12)$$

where N_{x_j} indicates a square window centered of x_j in spatial area. In this paper, we used a window size of 5 × 5. The spatial function h_{ij} will be maximized for the centered pixel if the most of the neighbors of one pixel belongs to the same class. The execution procedure of the spatial fuzzy clustering algorithm is the same as fuzzy clustering algorithm. The membership function varies as follows:

$$\mu'_{ij} = \frac{\mu_{ij}^p h_{ij}^q}{\sum_{k=1}^{c} \mu_{kj}^p h_{kj}^q} \quad (13)$$

where, P and q are the control parameter of each two function. In the first step (similar to fuzzy clustering algorithm), the membership function is calculated. In the second step, the information of membership function is mapped to the spatial area, and then the spatial function is calculated. It should be noticed that in the homogeneous region, the spatial functions are similar to the main membership function. As a result, there is no important variation in the clustering results. However, for the noisy pixels, the above relation reduces the weight of the noisy clusters by labeling the neighbor pixels. Consequently, the wrong classified pixels were easily modified from noisy areas. The block diagram of the spatial fuzzy clustering algorithm is shown in Figure 4.

The main problem of the fuzzy clustering algorithm is its high computational complexity due to updating its membership function in each iteration. To solve this problem, we use PCA method in order to project of n dimension data into the q, in which q is smaller than n ($q < n$). In PCA algorithm, $2d$ dimensional mean (μ) vector and also $d \times d$ dimensional covariance matrix (Σ) are calculated for all of the data, then eigenvectors and eigenvalues are calculated, and are sorted based on the descending orders. The mean vector is proportional to the maximum eigenvalue as a first main component. Moreover, the covariance matrix can be defined as:

$$F = \frac{1}{N} (x_k - \mu)(x_k - \mu)^T \quad (14)$$

Where, N is the number of objects in dataset and μ is the mean of data. PCA algorithm can be summarized as shown in Figure 5.

In the proposed PCA + SFCM algorithm, the main n and q dimensions are extracted while noisy and irrelevant dimensional features that could reduce the performance of clustering algorithm are eliminated. The PCA is exploited

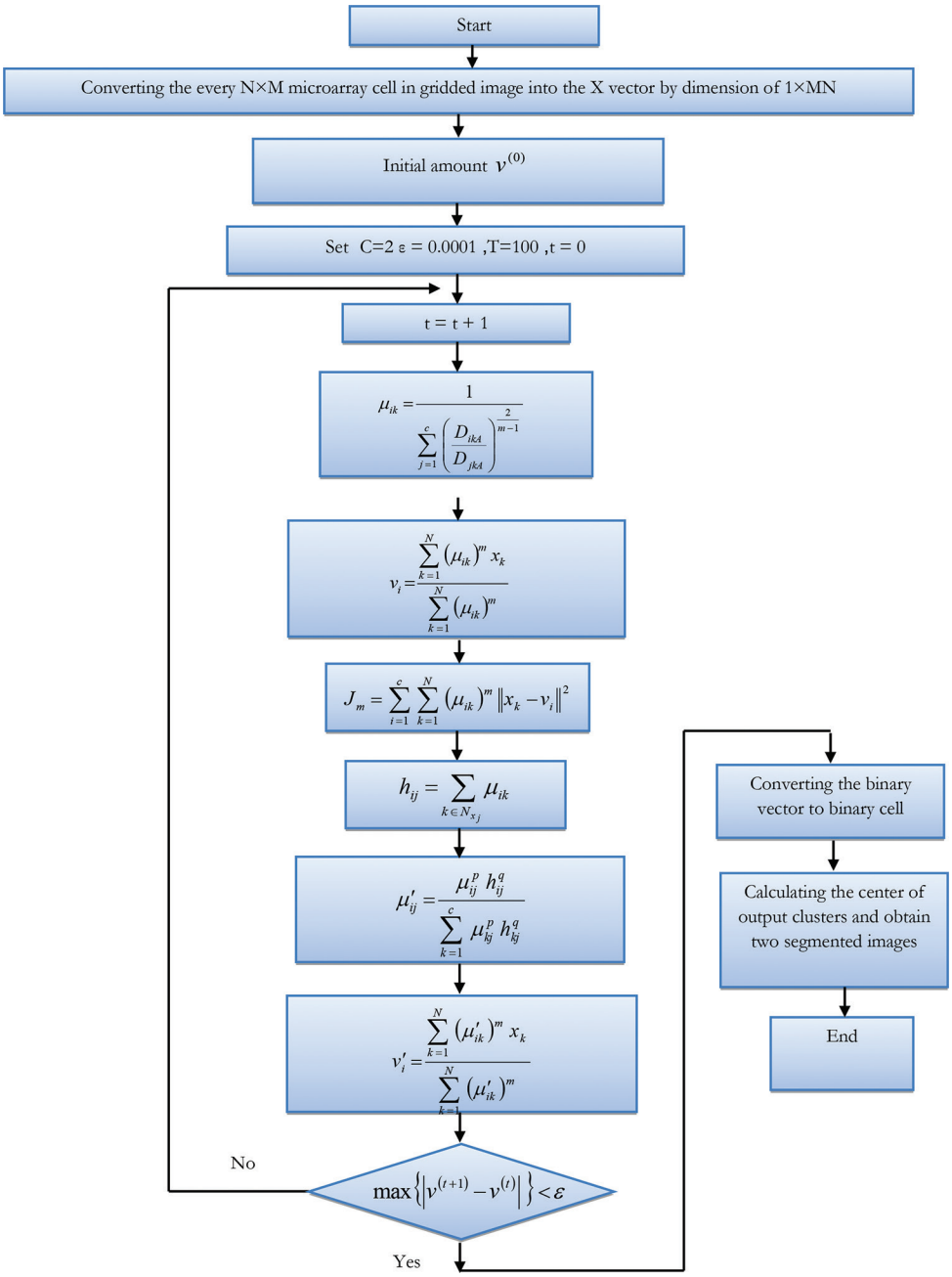

Figure 4: Proposed algorithm of the spatial fuzzy clustering for microarray cell segmentation

as a preprocess that generates dimension reduction data with noisy decreased. Such refine data are later employed in spatial fuzzy c-means clustering (SFCM) learning process.

RESULTS AND DISCUSSION

In Figure 6, the results of segmentation of the proposed algorithm have been shown. In this paper, three approaches, FCM, Otsu, and Gaussian mixture model (GMM) have been implemented in order to compare the performance the proposed algorithm. Furthermore, to evaluate the quantitative performance of our proposed algorithm, we use the evaluation criteria as below:[28,29]

- Segmentation matching factor (SMF). This parameter measures the pixels that are misclassified and defined as:

$$SMF = \frac{A_{segment} \cap A_{actual}}{A_{segment} \cup A_{actual}} \quad (15)$$

Where, $A_{segment}$ and A_{actual} are the binary versions of the real and segmentation images, respectively. For SMF parameter we have:

- If SMF = 100%, we have perfect matching of images
- If SMF >50%, the result of segmentation is acceptable
- If SMF <50%, the result of segmentation is weak.

- Coefficient of Determination (r^2). This parameter denotes the strength of the linear association between simulated and calculated images and can be defined as below:

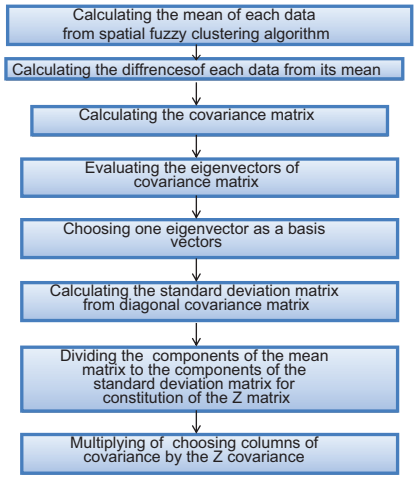

Figure 5: Principle component analysis algorithm

$$r^2 = \frac{\sum_{i=1}^{All\ spots} \left(I_{segment\,(i)} - \bar{I}_{actual}\right)^2}{\sum_{i=1}^{All\ spots} \left(I_{actual\,(i)} - \bar{I}_{actual}\right)^2} \quad (16)$$

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 … 1600), and $\bar{I}_{actual}$ is the overall spot intensity values of the simulated actual image

- Concordance correlation (P_C). This parameter measures the agreement between simulated and calculated data and is used to evaluate the reproducibility of the proposed segmentation algorithms. P_C is obtained as below:

$$P_c(A,B) = \frac{2 S_A S_B r}{S_A^2 + S_B^2 + \left(\bar{A} - \bar{B}\right)^2} \quad (17)$$

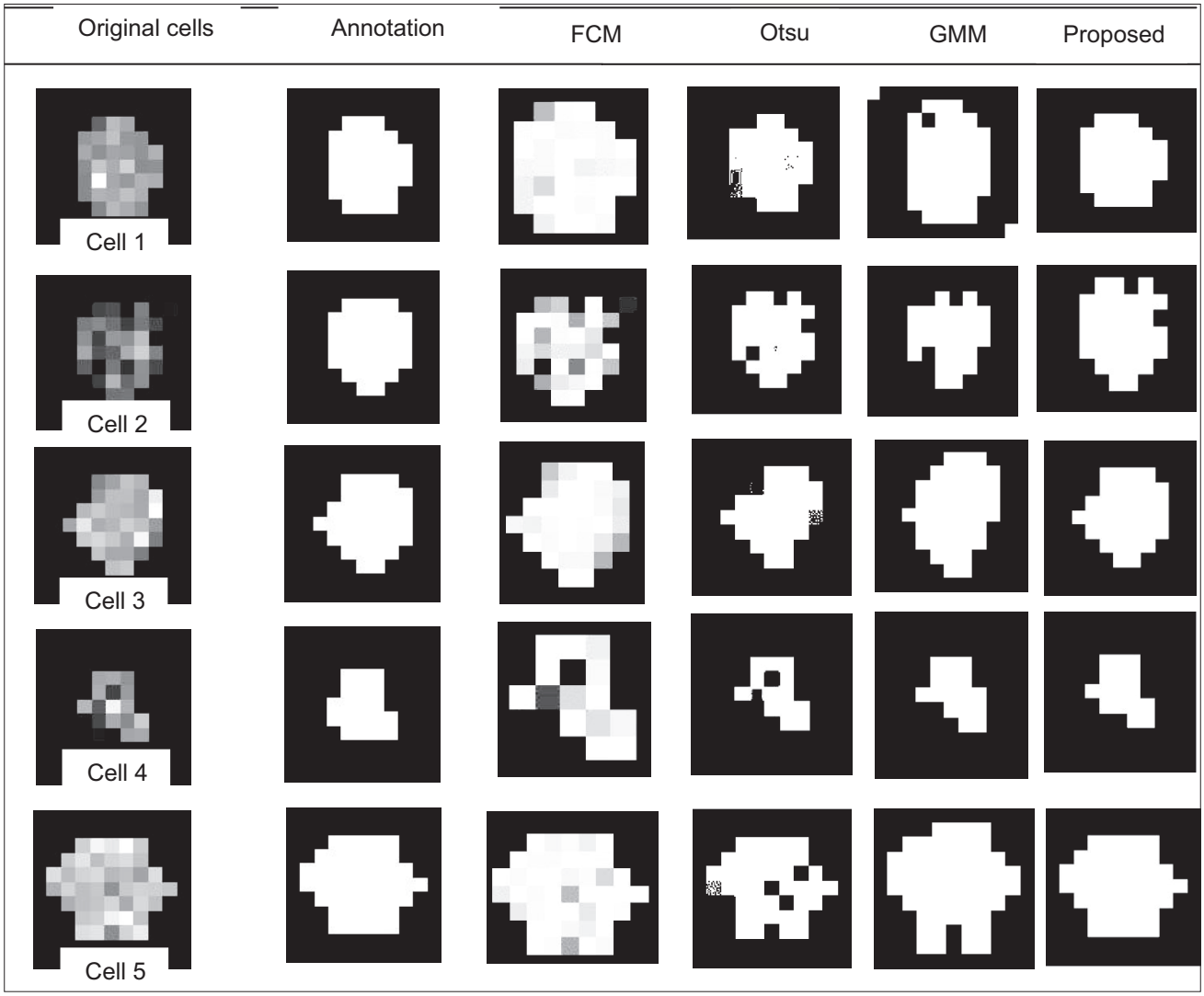

Figure 6: Comparison of the proposed algorithm and other approaches in microarray segmentation

Where, A and B are two samples, $\overline{A}$ and $\overline{B}$ are the mean values, S_A and S_B are the standard deviation of the samples. The higher value of P_c leads to the better performance of the segmentation algorithm.

To see the effect of fuzziness degree (m in Eq. 7) in the performance of our algorithm, we have changed the amount of this free parameter from 1.2 to 4, and in each situation, the quantity of SMF is calculated. Table 1 indicates the results. It can be seen that setting the degree of fuzziness >3 does have no significant effect on the behavior of our algorithm. Hence, the amount of fuzziness degree is equal to three.

For evaluating the amount of stability of the proposed algorithm in facing to the noise, the microarray images are corrupted with additive white Gaussian noise by the signal to noise ratio (SNR) of 1, 3, 5, 7, and 9 (dB). Tables 2-4 show the quantities of SMF, r^2 and P_c for the proposed algorithm and other segmentation methods, respectively. As can be seen from these tables, the proposed PCA-SFCM algorithm achieved the highest SMF, r^2 and P_c values as compared to FCM, Otsu and GMM algorithms at all five different SNR levels. When the SNR is reached to 1 dB, the amount of SMF in FCM is 89.2, whereas the Otsu and the GMM algorithms scored 85.4 and 90.7, respectively. This amount in our proposed algorithm is 92.5. This SMF differences are sustained for higher SNR levels and are shown in Tables 2-4. In contrast, in the case of the r^2, differences among FCM, Otsu, GMM, and our proposed algorithms decreased for higher SNR levels. Furthermore, the proposed algorithm has a good stability against the noise.

CONCLUSION

In this paper, segmentation of noisy cells is performed by extracting the amount of pixels intensity in each spot of microarray image. First, the artifacts in microarray images were reduced using nonlinear anisotropic diffusion method. Then, the center and also the position of each cell in microarray image were localized by applying the mathematical morphological reconstruction. Finally, the spot locations and the intensities of each spot were accurately calculated using a novel hybrid PCA and SFCM algorithm. The performance of the proposed algorithm was compared with the other methods based on some criteria measures, and results have been proved the superiority of our algorithm.

As mentioned above, the method exploited for gridding of microarray image was the mathematical morphology reconstructions. This method has two major problems: (1) The noise of the florescent has a bad effect on this approach and (2) the threshold is considered as a mean value of filtered signal in which this amount of threshold is not optimized. Our aims in the future chores will be choosing an adaptive threshold value using intelligent algorithms for gridding and also combine the Gaussian kernels besides

Table 1: Effect of degree of fuzziness (m) upon the performance of the proposed algorithm

Degree of fuzziness	Cell 1 (SMF)	Cell 3 (SMF)	Cell 5 (SMF)
1.2	38.2154	28.9265	24.2548
1.3	55.3688	61.3594	30.8965
1.4	58.3455	66.2314	54.9625
1.5	68.3259	71.6531	78.3654
1.6	69.3647	77.6215	100
1.7	74.15487	84.3514	100
1.8	79.6987	86.2165	93.2364
1.9	82.9254	90.3654	95.6958
2	89.3148	95.3124	100
2.1	98.6145	96.3654	98.6354
2.2	99.3014	98.6547	100
2.3	100	99.2514	95.0325
2.4	100	100	100
2.5	100	99.3654	97.2145
2.6	100	100	100
2.7	100	100	100
2.8	100	100	100
2.9	98.5987	100	99.1457
3	98.6478	99.2514	100
3.1	98.3694	100	100
3.2	100	100	99.1457
3.3	99.1865	100	100
3.4	100	100	99.1457
3.5	100	100	99.1457
3.6	100	100	99.1457
3.7	100	100	100
3.8	100	100	99.1457
3.9	100	100	100
4	100	100	99.1457

SMF – Segmentation matching factor

Table 2: Comparison of the quantities amount of SMF in the proposed algorithm and other methods by changing the SNR ratio

SNR	FCM	Otsu	GMM	Proposed
1	89.2	85.4	90.7	92.5
3	89.9	87.3	91.5	93.6
5	91.2	89.5	92.2	95.0
7	93.1	90.6	93.3	97.9
9	94.6	91.4	95.3	98.2
	SD=2.2394	SD=2.4643	SD=1.7861	SD=2.5442

SMF – Segmentation matching factor; SNR – Signal to noise; FCM – Fuzzy c-means; GMM – Gaussian mixture model; SD – Standard deviation

Table 3: Comparison of the quantities amount of r^2 in the proposed algorithm and other methods by changing the SNR ratio

SNR	FCM	Otsu	GMM	Proposed
1	0.84	0.80	0.89	0.932
3	0.87	0.85	0.91	0.951
5	0.93	0.89	0.93	0.959
7	0.95	0.93	0.94	0.981
9	0.97	0.96	0.98	0.993
	SD=0.0550	SD=0.0635	SD=0.0339	SD=0.0242

SNR – Signal to noise; FCM – Fuzzy c-means; GMM – Gaussian mixture model; SD – Standard deviation

Table 4: Comparison of the quantities amount of P_c in the proposed algorithm and other methods by changing the SNR ratio

SNR	FCM	Otsu	GMM	Proposed
1	0.80	0.75	0.83	0.88
3	0.80	0.76	0.83	0.89
5	0.81	0.78	0.84	0.92
7	0.82	0.80	0.85	0.95
9	0.84	0.82	0.87	0.97
	SD=0.0167	SD=0.0826	SD=0.0167	SD=0.0383

SNR – Signal to noise; FCM – Fuzzy c-means; GMM – Gaussian mixture model; SD – Standard deviation

the spatial fuzzy clustering algorithm to reach to more improvement in the segmentation process.

REFERENCES

1. Saberkari H, Shamsi M, Joroughi M, Golabi F, Sedaaghi MH. Cancer classification in microarray data using a hybrid selective independent component analysis and υ-Support Vector Machine Algorithm. J Med Signals Sens 2014;4:291-8.
2. Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. Trends Genet 2003;19:649-59.
3. Qiu P, Plevritis SK. Reconstructing directed signed gene regulatory network from microarray data. IEEE Trans Biomed Eng 2011;58:3518-21.
4. Campbell M, Heyer LJ. Discovering Genomics, Proteomics and Bioinformatics. 2nd ed. ISBN-13: 978-0805382198: Pearson Benjamin Cummings; 2007. p. 233-8.
5. Yang YH, Buckley MJ, Dudoit S, Speed T. Comparison of methods for image analysis on cDNA microarray data. J Comput Graph Stat 2002;11:108-36.
6. Eisen MB. ScanAlyze; 1999. ScanAlyze User Manual, Version 2.32; Stanford University: Stanford, CA. [Last accessed on 2015 Apr 08].
7. Buhler J, Ideker T, Haynor D. Dapple: Improved Techniques for Finding Spots on DNA Microarrays. UW CSE Technical Report UWTR, August 05, 2000; 2000. p. 1-12.
8. Biodiscovery Inc. ImaGene; 2005. Available from: http://www.biodiscovery.com. [Last accessed on 2015 Apr 08].
9. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, *et al.* A concise guide to cDNA microarray analysis. Biotechniques 2000;29:548-50, 552.
10. Sarder P, Nehorai A, Davis PH, Stanley SL. Estimating gene signals from noisy microarray images. IEEE Trans Nanobioscience 2008;7:142-53.
11. Beare R, Buckley MJ. The Spot User's Guide, CSIRO Mathematical and Information Sciences; 2000. Available from: http://www.cmis.csiro.au. [Last accessed on 2015 Apr 08].
12. Bozinov D, Rahnenführer J. Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering. Bioinformatics 2002;18:747-56.
13. Rahnenführer J, Bozinov D. Hybrid clustering for microarray image analysis combining intensity and shape features. BMC Bioinformatics 2004;5:47.
14. Nagarajan R. Intensity-based segmentation of microarray images. IEEE Trans Med Imaging 2003;22:882-9.
15. Nagarajan R, Peterson CA. Identifying spots in microarray images. IEEE Trans Nanobioscience 2002;1:78-84.
16. Li Q, Fraley C, Bumgarner RE, Yeung KY, Raftery AE. Donuts, scratches and blanks: Robust model-based segmentation of microarray images. Bioinformatics 2005;21:2875-82.
17. Ho J, Hwang WL. Automatic microarray spot segmentation using a Snake-Fisher model. IEEE Trans Med Imaging 2008;27:847-57.
18. Srinark T, Kambhamettu C, Kambhamettu R. A framework for Multiple Snakes. In Proc. Computer Vision and Pattern Recognition; 2001. p. 202-9.
19. Katzer M, Kummert F, Sagerer G. Methods for automatic microarray image segmentation. IEEE Trans Nanobioscience 2003;2:202-14.
20. Demirkaya O, Asyali MH, Shoukri MM. Segmentation of cDNA microarray spots using markov random field modeling. Bioinformatics 2005;21:2994-3000.
21. Lukac R, Plataniotis KN, Smolka B, Venetsanopoulos AN. A multichannel order-statistic technique for cDNA microarray image processing. IEEE Trans Nanobioscience 2004;3:272-85.
22. Nagarajan R, Upreti M. Correlation statistics for cDNA microarray image analysis. IEEE Trans Nanobioscience 2006;3:232-8.
23. Zacharia E, Maroulis D. 3-D spot modeling for automatic segmentation of cDNA microarray images. IEEE Trans Nanobioscience 2010;9:181-92.
24. Stanford Microarray Database. Available from: http://smd.princeton.edu//. [Last accessed on 2015 Apr 08].
25. Weickert J. Anisotropic diffusion in image processing, in ECMI Series. Stuttgart, Germany: Teubner; 1998.
26. Angulo J, Serra J. Automatic analysis of DNA microarray images using mathematical morphology. Bioinformatics 2003;19:553-62.
27. Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern 1974;3:32-57.
28. Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans Med Imaging 2002;21:193-9.
29. Lehmussola A, Ruusuvuori P, Yli-Harja O. Evaluating the performance of microarray segmentation algorithms. Bioinformatics 2006;22:2910-7.

How to cite this article: Saberkari H, Bahrami S, Shamsi M, Amoshahy MJ, Ghavifekr HB, Sedaaghi MH. Fully Automated Complementary DNA Microarray Segmentation using a Novel Fuzzy-based Algorithm. J Med Sign Sence 2015;5:182-91.

Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHIES

Hamidreza Saberkari was born in Rasht, Iran. He received the B.Sc. degree in Electrical Engineering from Guilan University, Rasht, IRAN, in 2011. In 2013, he received his M.Sc. degree in Communication Engineering from Sahand University of Technology, Tabriz, IRAN. His research interests include optimization algorithms, genomic signal processing, Bioinformatics, pattern recognition, Bio-MEMS.

E-mail: h_saberkari@sut.ac.ir

Sheyda Bahrami was born in Naqadeh, Iran in 1991. She received the B.Sc. degree in Biomedical Engineering from Sahand University of Technology in 2013. Now, she is M.Sc. student in Biomedical Engineering at this University. Her research interests are Biomedical Image Processing and Signal Processing.

E-mail: sh_bahrami@sut.ac.ir

Mousa Shamsi received his B.Sc. degree in Electrical Engineering (major: Electronics) from Tabriz University, Tabriz, IRAN, in 1995. In 1996, he joined the University of Tehran, Tehran, IRAN. He received his M.Sc. degree in Electrical Engineering (major: Biomedical Engineering) from this university in 1999. From 1999 to 2002, he taught as a lecturer at Sahand University of Technology, Tabriz, Iran. From 2002 to 2008, he was a PhD student at the University of Tehran in Bioelectrical Engineering. In 2006, he was granted with the Iranian government scholarship as a visiting researcher at the Ryukyus University, Okinawa, Japan. From December 2006 to May 2008, he was a visiting researcher at this University. He received his PhD degree in Electrical Engineering (major: Biomedical Engineering) from University of Tehran in December 2008. From December 2008 to April 2013, he was an assistant professor at Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. From April 2013, he is an associate professor at Faculty of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. His research interests include medical image and signal processing, genomic signal processing, pattern recognition, adaptive networks, and facial surgical planning.

E-mail: shamsi@sut.ac.ir

Mohammad Javad Amoshahy received the B.S. degree in applied mathematics and the M.S degree in coding and cryptography both from Isfahan University of Technology, Isfahan, Iran in 2005 and 2008, respectively. Now, he is a PhD student in Biomedical Engineering at Sahand University of Technology, Tabriz, Iran. His research interests include artificial intelligence; especially swarm intelligence, evolutionary computation, image processing, pattern recognition, and their applications to brain MRI segmentation.

E-mail: m_amoshahy@sut.ac.ir

Habib Badri Ghavifekr received the B.S. degree from Tabriz University, Iran, and continued his study in Germany and received the M.S. degree (Diploma Engineer) from Technical University of Berlin, in 1995, both in electrical engineering. Immediately after that, he joined the Institute for Microperipheric at the Technical University of Berlin (nowadays BeCAP: Berliner Center for Advanced Packaging) as a scientific assistant and in 1998 the Fraunhofer Institute for Reliability and Microintegration (FhGIZM) as a research assistant. In 2003, he received his PhD in electrical engineering from Technical University of Berlin. Since 2005, he is an assistant professor at Sahand University of Technology, Tabriz, Iran. His research interests are microsystem technologies, microelectronic packaging, MEMS, and electronic measurement system for industrial applications.

E-mail: badri@sut.ac.ir

Mohammad Hossein Sedaaghi was born in Tehran, Iran. He received the B.Sc. and M.Sc. degrees from the Sharif University of Technology, Tehran, IRAN, in 1986 and 1987, respectively. In 1998, he received the Ph.D. degree from Liverpool University. He is now a professor at Sahand University of Technology, Tabriz. His research interests include signal/image processing, pattern recognition, machine learning and biometrics.

E-mail: sedaaghi@sut.ac.ir