Possibilistic reformed fuzzy local information clustering technique for noisy microarray image spots segmentation

V. G. Biju[1,2,*] and P. Mythili[2]

[1]Department of Electronics and Communication Engineering, College of Engineering Munnar, Munnar 685 612, India
[2]Division of Electronics, School of Engineering, Cochin University of Science and Technology, Cochin 682 022, India

The cDNA microarray image provides useful information about thousands of gene expressions simultaneously. This information is used by bioinformatics researchers for diagnosis of different diseases and drug designs. Microarray image spot segmentation using an improved fuzzy clustering algorithm is proposed in this article. The proposed Possibilistic Reformed Fuzzy Local Information C Means (PRFLICM) algorithm is a variant of Possibilistic Fuzzy Local Information C Means (PFLICM) algorithm. The parameters used for testing the proposed algorithm include segmentation matching factor (SMF), probability of error (p_e), discrepancy distance (D), normalized mean square error and sum of square distance (SSD). The performance of the algorithm is validated with a set of simulated cDNA microarray images with known gene expression values. From the results of SMF, the proposed PRFLICM shows an improvement of 0.4% and 0.1% for high noise and low noise microarray images respectively when compared to PFLICM algorithm. The proposed algorithm is applied to yeast microarray database (YMD) and is used to find the yeast cell life cycle generated genes. The results show that the proposed algorithm has identified 101 cell life cycle regulated genes out of 104 such genes published in the YMD database.

Keywords: Fuzzy clustering, gene expression, image processing, microarray.

cDNA microarray technology helps molecular biologists to measure simultaneously the activity of thousands of biomolecules in the cell under different experimental conditions[1–3]. This powerful tool in biotechnology has been utilized in many biomedical applications such as cancer research, infectious disease diagnosis, toxicology research, pharmacology research and agricultural development. Spots foreground (FG) from microarray image are segmented from the background (BG) to compute gene expression (GE). The stages involved in microarray image processing are gridding, segmentation, information extraction and GE computation.

*For correspondence. (e-mail: bvgpillai@gmail.com)

These operations are used to find the accurate location of each spot, separate each spot FG from BG and compute GE value. The log to the base 2 value of the ratio of mean or median red and green plane intensities of each spot FG is the GE value.

Zacharia and Maroulis[4] proposed a 3D model for microarray spot segmentation where a 3D model was used to represent spot in a 3D space. The 3D space was optimized using genetic algorithm.

Athanasiadis *et al.*[5,6] proposed two algorithms, fuzzy gaussian mixture model (FGMM) and wavelet markov random field (WMRF) model for segmenting microarray spots. The methods were applied on both simulated and real microarray images.

Uslan *et al.*[7] used two clustering methods, Fuzzy C-Means (FCM) and K-means algorithms for segmentation of microarray image spots. Results showed that FCM could segment spots more accurately than K-means algorithm, but the segmentation accuracy of FCM was poor in medium and high noise spots.

The Genetic Algorithm based Fuzzy C Means (GAFCM)[8] method was applied to cDNA microarray images for segmenting microarray spots. The method improved FCM algorithm by optimizing the cluster centre c_j using genetic algorithm (GA) which resulted in better spot segmentation compared to FCM. The algorithm effectively segmented low and medium noise spots, but failed to segment high noise spots effectively. For high noise microarray images the GAFCM algorithm and other existing fuzzy clustering algorithms resulted in poor segmentation. So an improved fuzzy clustering algorithm, PFLICM, was put forward to enhance spot segmentation accuracy of high noise microarray images. PFLICM algorithm incorporated the properties of PFCM[9] and FLICM[11,12] algorithms to improve spot segmentation accuracy of high noise microarray images. Even though PFLICM algorithm improved the high noise microarray image spot segmentation considerably compared to GAFCM algorithm, it failed to select some isolated noisy edge pixels[12]. Hence in this article, to improve the spot segmentation accuracy of noisy spots, PRFLICM algorithm has been proposed.

Background

The aim of microarray image processing is to compute GE from each spot. In this article gridding of the microarray image is done on the basis of a grid line refinement method present[13]. It determines effectively the line segments constituting the borders between adjacent blocks or spots and does not require any input parameters or human intervention. Gridding helps to find each spot co-ordinate in a microarray image and is used to crop spot sub-image from microarray image. In this article, to improve the spot segmentation accuracy of noisy spots, PRFLICM was proposed. It is an improved version of PFLICM algorithm. PFLICM is an hybrid of PFCM[9] and FLICM[10,11] algorithm. The hybrid PFLICM uses the typicality of the pixel and local spatial information to improve clustering. The local spatial information taken from FLICM algorithm[10] measures the damping extent of the neighbouring pixels with the help of spatial distance from the central pixel. The local spatial relationship changes adaptively according to spatial distances from the central pixel. For the neighbourhood pixels having the same grey level value, the greater the spatial distance the smaller the damping extent and vice versa. FLICM was applied to microarray spot segmentation to identify its drawback[11]. In case of neighbourhood pixels which do not have same grey level values, the FLICM algorithm failed to assign these pixels to the right cluster.

This happens under two conditions. (1) The central pixel is affected by noise, whereas the other adjacent pixels are not and are homogenous. (2) The central pixel is not noisy, but some adjacent pixels are affected by noise. For case 1, the grey level difference between neighbouring pixels and central pixel is different and for case 2, the grey level difference between noise pixel and central pixel is different (Figure 1). In such cases the damping extent of the neighbouring pixels which is a function of the spatial distance, fails to analyse the impact of each neighbouring pixel on the fuzzy factor.

Because the local spatial information function (G_{ij}) used in PFLICM fails to detect certain isolated noise pixels, an improved version of PFLICM named PRFLICM is proposed in this article. For evaluation and testing of the algorithm, both simulated and real microarray images are used. The performance of existing and proposed PRFLICM algorithms is tested by evaluating the SMF, P_e, NMSE and SSD.

PRFLICM

The proposed PRFLICM algorithm is a modified version of PFLICM. The fuzzy clustering algorithms are modified for improving segmentation which is an important factor in extracting gene expression values accurately. PRFLICM is an improved version of PFLICM which includes the advantages of RFLICM[14]. In PRFLICM, G_{ij} in PFLICM has been replaced by G'_{ij} from Reformed Fuzzy Local Information C Means (RFLICM)[14].

The replaced reformulated fuzzy factor G'_{ij} is a function of local coefficient of variation C_u. Let $x = x_i$ ($i = 1$ to N) be the pixels of a microarray spot sub image. These pixels have to be clustered into two classes, BG and FG. Let c_1 and c_2 be cluster centres of BG and FG pixels respectively. Based on the maximum value of the membership function u_{ij}, each pixel is grouped into BG and FG. The cluster centres are updated iteratively based on the grouped pixel. PRFLICM aims at iteratively improving u_{ij} or c_j by minimizing the absolute value of the difference between two consecutive values of the objective functions F_t.

The objective function of the proposed PRFLICM is given by eq. (1)

$$F^t = \sum_{i=1}^{N}\sum_{j=1}^{c}[(u_{ij}^m + t_{ij}^\eta)d_{ij} + G'_{ij}\ ] + \sum_{j=1}^{c}\gamma_i \sum_{i=1}^{N}(1-t_{ij})^\eta, \qquad (1)$$

where m and η are fuzziness parameters, d_{ij} is the Euclidean distance from a pixel to a cluster centre and is given by eq. (2)

$$d_{ij} = \| x_i - c_j \|^2, \qquad (2)$$

t_{ij} is the typicality of the pixel and is calculated using eq. (3)

$$t_{ij} = \frac{1}{1+(d_{ij}/\gamma_i)^{1/(m-1)}}, \qquad (3)$$

where γ_i is a constant and is given by eq. (4) and K is a constant greater than zero

$$\gamma_i = \mathrm{K}\frac{\sum_{i=1}^{N} u_{ij}^m d_{ij}}{\sum_{i=1}^{N} u_{ij}^m} \quad K > 0. \qquad (4)$$

The reformed fuzzy factor G'_{ij} is defined using eq. (5)

$$G'_{ij} = \begin{cases} \sum_{i=N_k} \dfrac{1}{2+\min\left(\left(\dfrac{C_u^j}{C_u}\right)^2, \left(\dfrac{C_u}{C_u^j}\right)^2\right)} \\ \qquad \times (1-u_{ij})^m \| x_i - c_j \|^2 \ \text{ if } c_u^j \geq \bar{C}_u \\ \sum_{i=N_k} \dfrac{1}{2-\min\left(\left(\dfrac{C_u^j}{C_u}\right)^2, \left(\dfrac{C_u}{C_u^j}\right)^2\right)} \\ \qquad \times\ (1-u_{ij})^m \| x_i - c_j \| \ \text{ if } c_u^j < \bar{C}_u\,. \end{cases} \qquad (5)$$

C_u value in G_{ij} reflects the degree of grey-value homogeneity in a local window. It exhibits high values at edges (the areas affected by noise) and produces a low value in the homogeneous regions. C_u computes the damping extent of the neighbours on the basis of the area where the neighbouring pixels are located. For example, if the neighbouring pixel and the central pixel are located in the same region, such as the homogeneous region or the area affected by noise, the value of C_u obtained will be very close and vice versa. In general, compared to the spatial distance, the discrepancy of C_u between neighbouring pixels and the central pixel is relatively in accordance with the grey-level difference between them. In addition, it helps to exploit more local information since the local coefficient of each pixel is computed in a local window

$$C_u = \mathrm{var}(x)/\bar{x}, \tag{6}$$

where var(x) and $\bar{x}$ are the intensity variance and the mean in a local window of the image respectively. In eq. (5) C_u^j represents the local coefficient of variation of neighbouring pixels, and $\bar{C}_u$ is the mean value that is located in a local window. The reformulated factor G'_{ij} balances the membership value of the central pixel taking into account the local coefficient of variation, as well as the grey level of the neighbouring pixels[14]. Any distinct difference between the results of local coefficient of variation obtained by the neighbouring pixel and central pixel, the weightings added of the neighbouring pixel in G'_{ij} will increase to suppress the influence of outliers.

In PRFLICM, the membership function u_{ij} is computed using eq. (7). The coefficient of variation of each pixel is computed as

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\dfrac{d_{ij} + G'_{ij} + t_{ij}}{d_{ik} + G'_{ik} + t_{ik}}\right)^{1/(m-1)}}. \tag{7}$$

The cluster centre C_j represented by eq. (8) is updated by using the modified eq. (7) of u_{ij}.

$$c_j = \frac{\sum_{i=1}^{N}(u_{ij}^m)x_i}{\sum_{i=1}^{N}(u_{ij}^m)}. \tag{8}$$

The PRFLICM algorithm is given as follows.

Step 1: Initialize c_j, fuzzification parameters m and the number of iterations (itermax), iter = 0 and stopping condition or error (ε). Step 2: Find u_{ij} from equation of FCM. Step 3: Initialize randomly the typicality matrix t_{ij}. Step 4: Set the loop count iter = 0. Step 5: Calculate d_{ij} using eq. (2). Step 6: Calculate t_{ij} using eq. (3). Step 7: Calculate G_{ij} using eq. (5). Step 8: Compute the membership degree function u_{ij} using eq. (7). Step 9: Update C_j using eq. (8). Step 10: If the $\max\|F_{t+1} + F_t\| \le \varepsilon$ or iter = itermax then stop, otherwise iter = iter + 1 and go to step 5.

Database used for evaluation

To quantify the effectiveness of the proposed approach in microarray image processing, simulated as well as real microarray image database is used.

Simulated database

Database are simulated and are used for validation purposes. The advantage of using a simulated database is that the exact values of spot parameters such as spot area, mean or median FG intensity, mean or median BG intensity, GE values, etc. are known a priori. A set of 40 microarray images, each with 225 spot were simulated[5,6] for numerically evaluating and comparing various segmentation methods. In order to generate spots with realistic characteristics, the following procedure is adopted. A true cDNA image is used as a template and its binary version is produced by employing a thresholding technique. Thus the location, boundary and the area of all simulated spots are a priori determined. Intensities of each FG region are drawn from a uniform distribution using mean FG intensities of respective spots in the original image. BG pixels intensities are the same as that of the mean BG intensity of the original image.

Yeast microarray database

The algorithms applied to yeast microarray database (YMD) experiments are used to find cell life cycle regulated genes in yeast[15]. From the database the images and GE values of three experiments such as alpha factor, Cdc15 and elutriation are taken for analysis. Alpha factor experiment (pheromone experiment) contains 18 images taken at a time interval of 7 min, starting at zero time and ending in 119 min. Cdc15 experiment contains 24 images taken at different time intervals, starting after 10 min and ending in 290 min. Elutriation experiments consist of 14 images taken at a time interval of 30 min, starting at zero time and ending in 390 min.

Measures used for evaluation

The accuracy of segmentation technique is evaluated using parameters such as SMF, p_e, NMSE and SSD.

SMF

SMF is used to measure the accuracy of any segmentation algorithm. SMF[16–18] for every binary spot produced by the clustering algorithm is given by

$$\mathrm{SMF} = \frac{(A_{\mathrm{seg}} \cap A_{\mathrm{act}})}{(A_{\mathrm{seg}} \cup A_{\mathrm{act}})} \times 100, \tag{9}$$

where A_{seg} is the area of spot as determined by the proposed algorithm and A_{act} 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.

p_e

Pixel level accuracy of segmentation is examined with statistical parameter p_e, which measures improperly segmented pixels and is defined as

$$p_e = P(F) \cdot P(B/F) + P(B) \cdot P(F/B), \tag{10}$$

where $P(B/F)$ is probability of error in classifying FG pixel as BG pixels, $P(F/B)$ is probability of error in classifying BG pixels as FG pixels, $P(F)$ and $P(B)$ are a priori probabilities of FG and BG pixels in the image. When all the pixels of a spot are correctly segmented, p_e takes the minimum value of zero. p_e takes the maximum value of one when all the pixels are incorrectly segmented[2].

NMSE

NMSE is used to measure the performance of the proposed approach which is given by

$$\frac{\left[\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}(x_{ij} - \overline{x_{ij}})^2\right]^{1/2}}{\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}(x_{ij})}, \tag{11}$$

where M and N are dimensions of the image. x_{ij} and $\overline{x_{ij}}$ are the original and clustered image pixels respectively. A minimum value of zero is desirable for better segmentation[19].

SSD

$$\mathrm{SSD} = \sum_{i=1}^{N}\sum_{r=1}^{R}(i_{ir} - \overline{i_{ir}})^2, \tag{12}$$

where N is the total number of spots in the microarray, and R is the total number of replicates. i_{ir} is the log ratio of the ith spot on the rth replicate and $\overline{i_{ir}}$ is the mean of the log ratio across all replicates for the ith spot. SSD calculates the variation in the log ratio estimate. A smaller value of SSD or less variation shows stability of the method[19]. SSD is used to find the stability of the estimated gene expression levels obtained using the proposed algorithms.

Sensitivity and specificity

The parameters such as sensitivity and specificity are used to test the correctly identified cell life cycle regulated genes in YMD.

True positive (TP): Cell life cycle regulated genes that are correctly identified as cell life cycle regulated genes. False negative (FN): Cell life cycle regulated genes that are incorrectly identified as not cell life cycle regulated genes. True negative (TN): Not cell life cycle regulated genes that are correctly identified as not cell life cycle regulated genes. False positive (FP): Not cell life cycle regulated genes that are incorrectly identified as cell life cycle regulated genes.

Sensitivity measures the proportion of positives that are correctly identified.

$$\mathrm{Sensitivity} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}). \tag{13}$$

Specificity measures the proportion of negatives that are correctly identified.

$$\mathrm{Specificity} = \mathrm{TN}/(\mathrm{TN} + \mathrm{FP}). \tag{14}$$

The sensitivity and specificity values range between 0 and 1. A maximum value of 1 is expected for the parameters.

Results of simulated microarray database

The proposed algorithm is applied on simulated microarray database explained earlier. Forty microarray images, each with 225 spots, are used for analysis. To analyse the spot segmentation efficiency of the proposed algorithm under noise, the simulated images are added with AWGN noise with SNR varying from 1 to 10 dB. The spot segmentation accuracy is computed using the parameters SMF, p_e, NMSE and SSD. The performance measurement parameter such as SMF, p_e and NMSE achieved for simulated spots (average result of 40 simulated microarray image spots) corresponding to different SNR levels is presented in Tables 1–3 respectively. Regarding SMF, the proposed PRFLICM algorithm resulted in higher SMF than PFLICM algorithm. The p_e and NMSE should have a value zero for ideal spot segmentation.

Table 1. Comparison of PFCM, GAFCM, FLICM, RFLICM, PFLICM and PRFLICM algorithm based on SMF for simulated microarray images with different levels of AWGN

SNR (dB)	PFCM	GAFCM	FLICM	RFLICM	PFLICM	PRFLICM
1	66.282	68.012	81.864	82.016	82.846	83.722
2	72.513	73.914	86.230	86.312	87.236	87.635
3	78.322	79.924	90.359	91.301	91.359	91.677
4	84.153	84.915	92.641	92.711	93.646	93.866
5	89.329	90.113	94.677	94.689	95.673	95.804
6	93.799	94.614	95.854	95.862	96.864	96.976
7	96.363	96.375	96.780	96.791	97.600	97.734
8	97.032	97.116	97.352	97.357	98.397	98.517
9	98.169	98.217	98.469	98.472	99.191	99.302
10	98.552	98.618	99.652	99.657	99.810	99.818
Without noise	98.986	99.152	99.739	99.745	99.902	99.913

Table 2. Comparison of PFCM, GAFCM, FLICM, RFLICM, PFLICM and PRFLICM algorithm based on p_e for simulated microarray images with different levels of AWGN

SNR(dB)	PFCM	GAFCM	FLICM	RFLICM	PFLICM	PRFLICM
1	0.169	0.160	0.091	0.091	0.086	0.081
2	0.137	0.130	0.069	0.068	0.064	0.062
3	0.108	0.100	0.043	0.043	0.043	0.042
4	0.079	0.075	0.037	0.036	0.032	0.031
5	0.053	0.049	0.027	0.027	0.022	0.021
6	0.031	0.027	0.021	0.021	0.016	0.015
7	0.018	0.018	0.016	0.016	0.012	0.011
8	0.015	0.015	0.013	0.013	0.008	0.007
9	0.009	0.009	0.008	0.008	0.004	0.003
10	0.007	0.007	0.002	0.002	0.001	0.001
Without noise	0.001	0.001	0.000	0.000	0.000	0.000

Table 3. Comparison of PFCM, GAFCM, FLICM, PFLICM and PRFLICM algorithm based on NMSE for simulated microarray images with different levels of AWGN

SNR(dB)	PFCM	GAFCM	FLICM	RFLICM	PFLICM	PRFLICM
1	0.270	0.256	0.145	0.145	0.137	0.130
2	0.220	0.209	0.110	0.110	0.102	0.099
3	0.173	0.161	0.069	0.069	0.069	0.067
4	0.127	0.121	0.059	0.058	0.051	0.049
5	0.085	0.079	0.043	0.042	0.035	0.034
6	0.050	0.043	0.033	0.033	0.025	0.024
7	0.029	0.029	0.026	0.026	0.019	0.018
8	0.024	0.023	0.021	0.021	0.013	0.012
9	0.015	0.014	0.012	0.012	0.006	0.006
10	0.012	0.011	0.003	0.003	0.002	0.001
Without noise	0.002	0.002	0.000	0.000	0.000	0.000

Tables 2 and 3 show that the proposed PRFLICM has value near zero when compared to other existing methods. The SSD shows the stability of the proposed algorithm in segmenting microarray spot and GE calculation. In the eq. (12), N is taken as 225 and R is taken as 10. SSD calculates the variation in the log ratio estimate. A small value for SSD shows the stability of the method. Table 4 shows average result obtained for a simulated microarray image with 10 replicates obtained by adding the noise with the same SNR value[19].

The noise levels were varied from 1 dB to 10 dB. The results show minimum variation for the proposed PRFLICM algorithms when compared to other algorithms for all levels of noise. The SMF, p_e, NMSE and SSD results show the efficiency and robustness of the proposed PRFLICM algorithm over other existing algorithms. The ultimate goal of the segmentation process is to obtain intensity measurement. Accurate segmentation of spot has a great impact on the intensity calculation. Measurements such as SMF, p_e and NMSE support the superiority

Table 4. Comparison of PFCM, GAFCM, FLICM, PFLICM and PRFLICM algorithm based on SSD for simulated microarray images with different levels of AWGN

SNR (dB)	PFCM	GAFCM	FLICM	RFLICM	PFLICM	PRFLICM
1	0.387	0.367	0.218	0.208	0.202	0.189
2	0.232	0.219	0.144	0.139	0.134	0.121
3	0.235	0.225	0.082	0.080	0.074	0.069
4	0.079	0.071	0.059	0.060	0.058	0.057
5	0.091	0.087	0.049	0.048	0.048	0.048
6	0.064	0.054	0.036	0.036	0.035	0.035
7	0.050	0.045	0.033	0.034	0.033	0.033
8	0.032	0.031	0.017	0.017	0.016	0.016
9	0.018	0.018	0.017	0.015	0.016	0.014
10	0.017	0.017	0.016	0.015	0.015	0.014

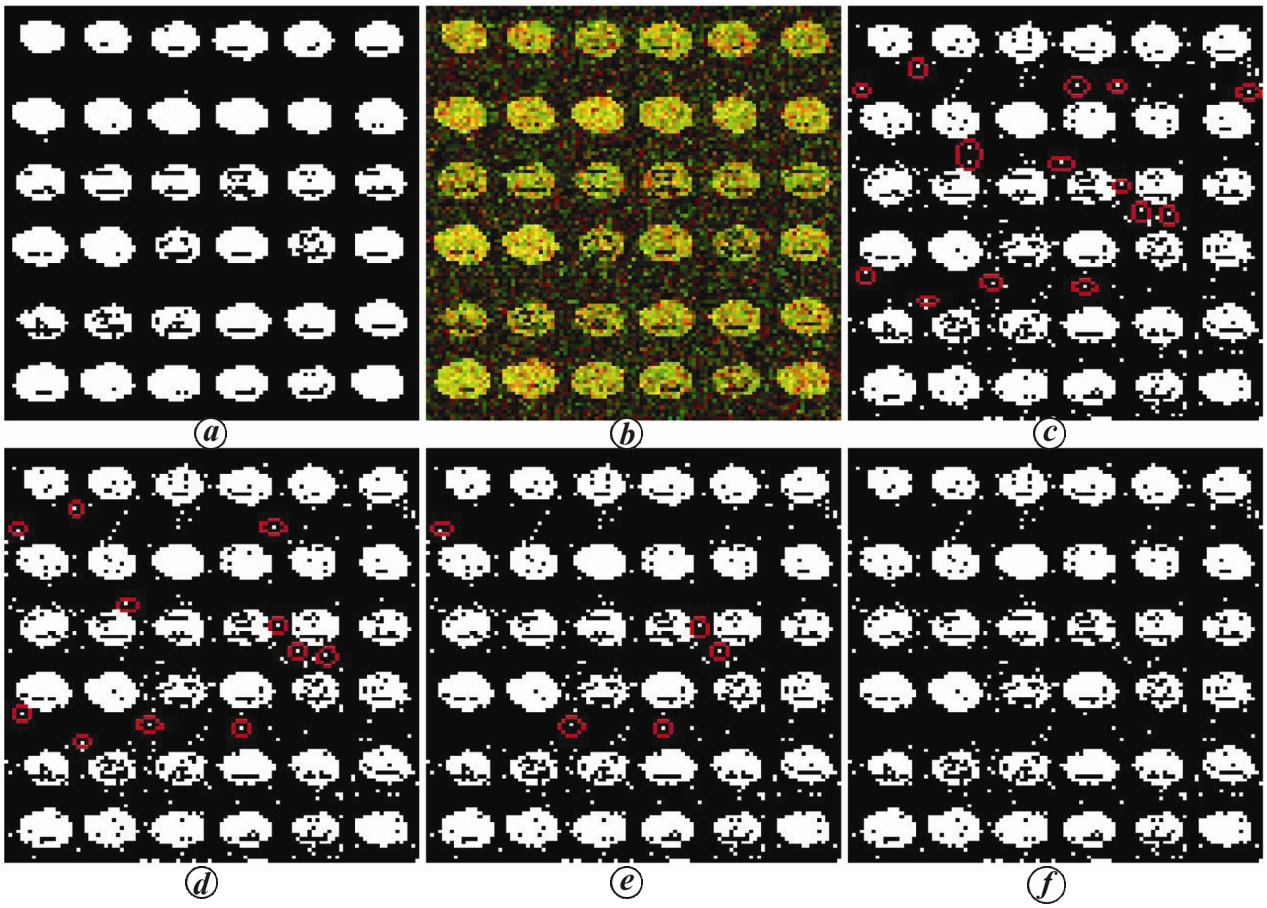

Figure 1. ***a***, The subimage of reference mask; ***b***, A simulated subimage added with AWGN noise of 4 dB; ***c–f***, Segmentation result obtained using FLICM, RFLICM, PFLICM and the proposed PRFLICM algorithms respectively.

of the proposed PRFLICM against other existing algorithms.

Figure 1 *a* shows a reference mask used to simulate the microarray simulated image. Figure 1 *b* shows the simulated image added with AWGN of SNR 4 dB. The reference image is obtained from a real microarray image. Figure 1 *c–f* shows the segmented output for Figure 1 *b* using FLICM, RFLICM, PFLICM and the proposed PRFLICM algorithms respectively. The spots marked using circles (Figure 1 *c–e*) show the difference in selecting noisy pixels using FLICM, RFLICM and PFLICM algorithms with respect to the proposed PRFLICM algorithm. From the figures it is clear that the segmentation accuracy of the proposed PRFLICM algorithm is far better for noisy spots compared to other existing algorithms.

Results of real microarray

YMD is used for finding cell life cycle regulated genes in yeast[15]. From the database, images and GE values of alpha factor, cdc15 and elutriation time course experiments are used in this article for evaluation and comparison of results. Figure 2 shows a yeast sub-array gridded image after applying the grid line refinement method introduced earlier[13]. The spot sub-images are cropped using the co-ordinates obtained from the grid line refinement method and the algorithms such as FLICM, RFLICM, PFLICM and the proposed PRFLICM algorithm are applied on YMD to separate each spot FG from BG.

The GE values are computed from each spot FG and are analysed using hierarchical clustering and PCA. PCA computes variance and covariance matrix of GE values and captures the variability of each gene and the extent to which it co-varies with every other gene. The PCA first and second principal component obtained from the actual database (alpha factor experiment) verses FLICM, RFLICM, PFLICM and PRFLICM obtained values are depicted in Figures 3–6 respectively. From Figure 6, it is seen that most data points of the proposed PRFLICM algorithm overlap with the actual data base value compared to the other existing algorithms. This shows the proposed PRFLICM algorithm's efficiency in GE computation.

The sensitivity and specificity of the parameter are used to measure the correctly identified cell life cycle regulated genes in YMD. Its value ranges between 0 and 1. A maximum value of 1 is expected for both the

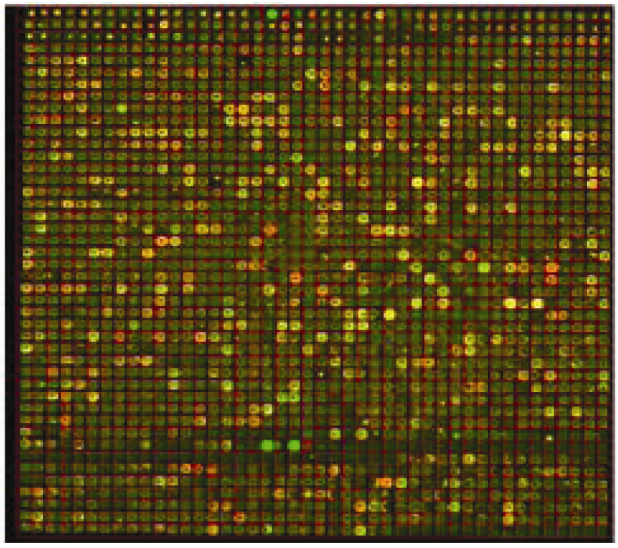

Figure 2. Yeast microarray sub-array gridded image.

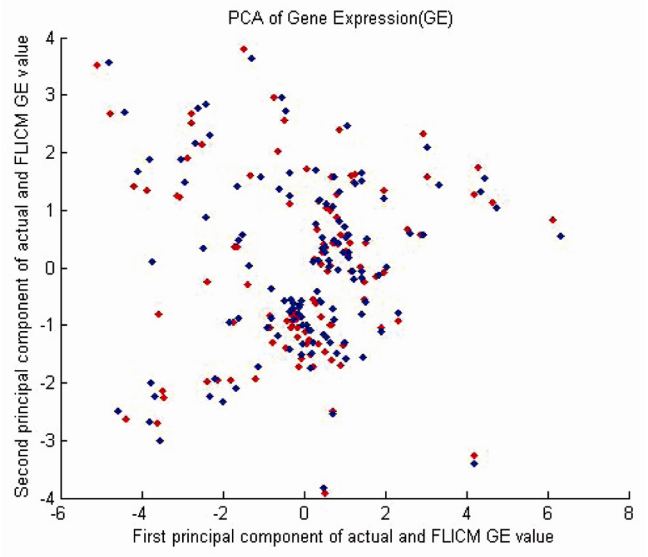

Figure 3. PCA result obtained for FLICM GE value.

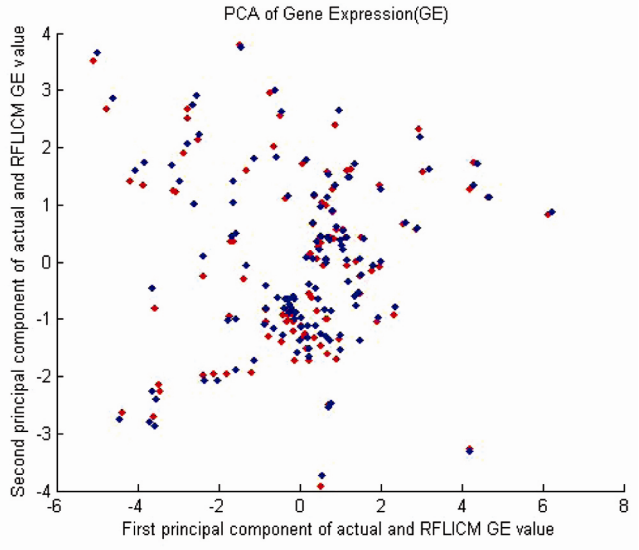

Figure 4. PCA result obtained for RFLICM GE value.

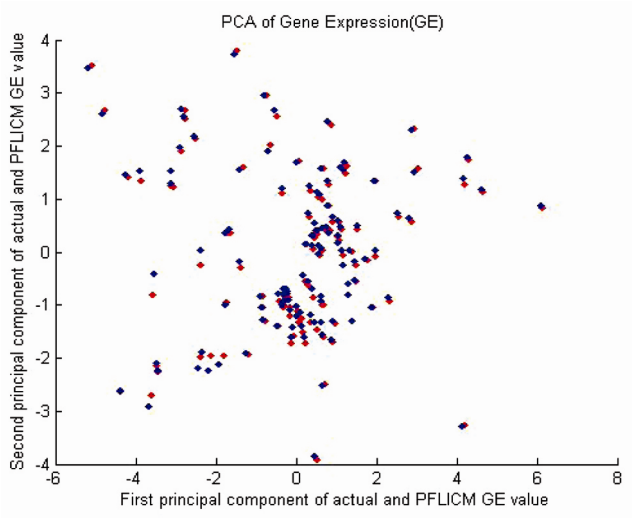

Figure 5. PCA result obtained for PFLICM GE value.

Table 5. Cell life cycle regulated genes identified from YMD

Cell life cycle regulated genes	FLICM	RFLICM	PFLICM	PRFLICM
Total identified	125	123	117	116
Correctly Identified	96	96	101	101
Sensitivity	0.9231	0.9231	0.9712	0.9712
Specificity	0.9965	0.9969	0.9979	0.998

Total number of genes in YMD is 6178 and the total number of cell life cycle regulated genes are 104.

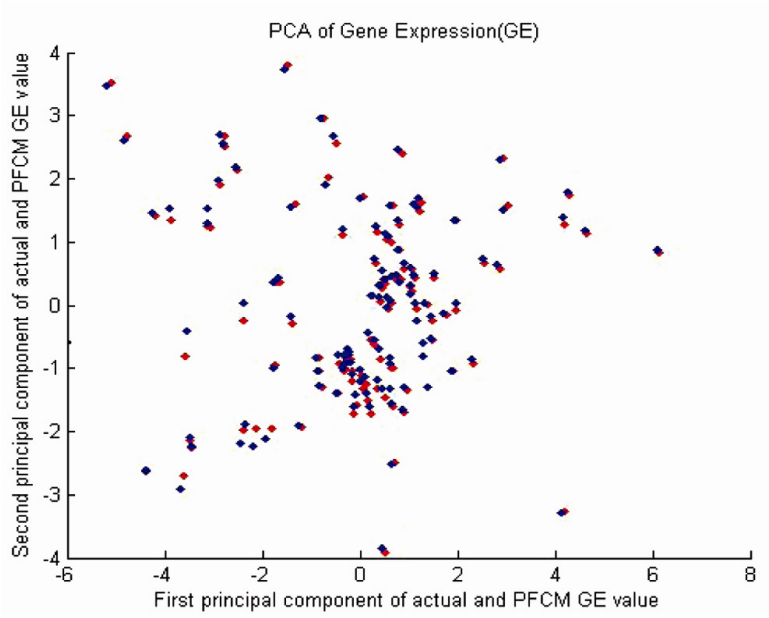

Figure 6. PCA result obtained for PRFLICM GE value.

parameters. From Table 5, the proposed method has identified 101 cell life cycle regulated genes out of 104 in the YMD[15]. The results show that the existing PFLICM and the proposed PRFLICM algorithms have values close to 1. Though the sensitivity of PFLICM and PRFLICM is the same, the specificity of PRFLICM is better than PFLICM. This makes the performance of PRFLICM better when compared to PFLICM and other existing algorithms.

To study the presence of a particular disease in any living being, a bioinformatics researcher has to separate the mRNA from all the genes collected from both diseased as well as normal cell samples. These means are used to produce microarray images and are further processed by using image processing algorithms to obtain the GE values. These GE values are used to further identify the differentially expressed genes. These differentially expressed genes help the bioinfomatics researcher to identify the nature and severity of the disease.

Microarray image processing, GE analysis and the precise identification of the differentially expressed genes are major challenges in microarray experiments. In bioinformatics labs, analysis of these differentially expressed genes alone is made to carry out the diagnosis and prognosis of diseased cell or tissues. Hence identification of the exact differentially expressed genes will narrow down the job of the researcher in the lab which remains a challenge.

Conclusion

The proposed algorithm is applied on simulated microarray images. These images are used to validate the proposed algorithm and to analyse the performance of the proposed algorithm on noisy spots. From the results, for SMF, PRFLICM shows an improvement of 0.4% and 0.1% for high noise and low noise microarray images respectively compared to PFLICM algorithm. The algorithm performs comparatively better than the other existing algorithms. The number of spots used for evaluation and comparison of various parameters obtained support the superiority of the proposed method over other existing standard methods in microarray image processing. The proposed algorithm is applied on YMD images to find cell life cycle regulated genes in yeast. The GE values are computed from the microarray images using the proposed PRFLICM algorithm. The GE values are analysed using hierarchical clustering and PCA. The analysis

identified 101 from 104 cell life cycle regulated genes of yeast.

The proposed method can be applied on any real microarray images used to find differently expressed genes in diseased cell or organs of any organism. This information is used by the bioinformatics researchers for disease diagnosis and drug discovery.

1. Yang, Y. H., Buckley, M. J., Dudoit, S. and Speed, T. P., Comparison of methods for image analysis on cDNA microarray data. *J. Comput. Graph. Stat.*, 2002, **11**(1), 108–136.
2. Lehmussola, A., Ruusuvuori, P. and Yli-Harja, O., Evaluating the performance of microarray segmentation algorithms. *Bioinformatics*, 2006, **22**(23), 2910–2917.
3. Schena, M., Shalon, D., Davis, R. W. and Brown, P. O., Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science-New York Then Washington, 1995, pp. 467–467.
4. Zacharia, E. and Maroulis, D., 3-D spot modeling for automatic segmentation of cDNA microarray images. *IEEE Trans. Nanobiosci.*, 2010, **9**(3), 181–192.
5. Athanasiadis, E., Cavouras, D., Spyridonos, P., Glotsos, D., Kalatzis, I. and Nikoforidis, G., Segmentation of microarray images using gradient vector flow active contours boosted by Gaussian mixture models. In *Second International Conference on Experiments/Process/System Modeling/Simulation/Optimization (2nd IC-EpsMsO)*, Athens, Greece, 2007.
6. Athanasiadis, E., Cavouras, D., Kostopoulos, S., Glotsos, D., Kalatzis, I. and Nikiforidis, G., A wavelet-based Markov random field segmentation model in segmenting microarray experiments. *Comput. Methods Programs Biomed.*, 2011, **104**(3), 307–315.
7. Uslan, V. and Bucak, I. Ö., Microarray image segmentation using clustering methods. *Math. Comput. Appl.*, 2010, **15**(2), 240–247.
8. Biju, V. G. and Mythili, P., A genetic algorithm based fuzzy C mean clustering model for segmenting microarray images. *Int. J. Comput. Appl.*, 2012, **52**(11), 42–48.
9. Pal, N. R., Pal, K., Keller, J. M. and Bezdek, J. C., A possibilistic fuzzy c-means clustering algorithm. *IEEE Trans. Fuzzy Syst.*, 2005, **13**(4), 517–530.
10. Krinidis, S. and Chatzis, V., A robust fuzzy local information C-means clustering algorithm. *IEEE Trans. Image Process.*, 2010, **19**(5), 1328–1337.
11. Biju, V. G. and Mythili, P., Fuzzy clustering algorithms for cDNA microarray image spots segmentation. *Procedia Comput. Sci.*, 2015, **46**, 417–424.
12. Biju, V. G. and Mythili, P., An improved fuzzy clustering algorithm for microarray spots segmentation. *ICTACT J. Image Video Process.*, 2015, **6**(2), 1107–1114.
13. Biju, V. G. and Mythili, P., Microarray image gridding using grid line refinement technique. *ICTACT J. Image Video Process.*, 2015, **5**(4), 1010–1016.
14. Gong, M., Zhou, Z. and Ma, J., Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering. *IEEE Trans. Image Process.*, 2012, **21**(4), 2141–2151.
15. Spellman, P. T. *et al*., Comprehensive identification of cell cycle–regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, 1998, **9**(12), 3273–3297.
16. Tran, D. and Wagner, M., Noise clustering-based speaker verification. *Lecture Notes in Computer Science*, 2002, pp. 325–331.
17. Betal, D., Roberts, N. and Whitehouse, G. H., Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology. *Br. J. Radiol.*, 1997, **70**(837), 903–917.
18. Athanasiadis, E. I., Cavouras, D. A., Spyridonos, P. P., Glotsos, D. T., Kalatzis, I. K. and Nikiforidis, G. C., Complementary DNA microarray image processing based on the fuzzy Gaussian mixture model. *IEEE Trans. Inform. Technol. Biomed.*, 2009, **13**(4), 419–425.
19. Wang, Y. P., Gunampally, M., Chen, J., Bittel, D., Butler, M. G. and Cai, W. W., A comparison of fuzzy clustering approaches for quantification of microarray gene expression. *J. Signal Process. Syst.*, 2008, **50**(3), 305–320.

Received 22 September 2016; revised accepted 13 April 2017

doi: 10.18520/cs/v113/i06/1072-1080