ISSN: 0976-9102(ONLINE)
DOI: 10.21917/ijivp.2015.0161

AN IMPROVED FUZZY CLUSTERING ALGORITHM FOR MICROARRAY IMAGE SPOTS SEGMENTATION

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

[1]*Department of Electronics and Communication Engineering, College of Engineering Munnar, India*
E-mail: bvgpillai@gmail.com

[2]*Division of Electronics Engineering, School of Engineering, CUSAT, India*
E-mail: mythili@cusat.ac.in

Abstract

An automatic cDNA microarray image processing using an improved fuzzy clustering algorithm is presented in this paper. The spot segmentation algorithm proposed uses the gridding technique developed by the authors earlier, for finding the co-ordinates of each spot in an image. Automatic cropping of spots from microarray image is done using these co-ordinates. The present paper proposes an improved fuzzy clustering algorithm Possibility fuzzy local information c means (PFLICM) to segment the spot foreground (FG) from background (BG). The PFLICM improves fuzzy local information c means (FLICM) algorithm by incorporating typicality of a pixel along with gray level information and local spatial information. The performance of the algorithm is validated using a set of simulated cDNA microarray images added with different levels of AWGN noise. The strength of the algorithm is tested by computing the parameters such as the Segmentation matching factor (SMF), Probability of error (p_e), Discrepancy distance (D) and Normal mean square error (NMSE). SMF value obtained for PFLICM algorithm shows an improvement of 0.9 % and 0.7 % for high noise and low noise microarray images respectively compared to FLICM algorithm. The PFLICM algorithm is also applied on real microarray images and gene expression values are computed.

Keywords:

Gridding, Spot Segmentation, Local Information, Spatial Information, Typicality, Clustering, Gene Expression

1. INTRODUCTION

cDNA microarray is used to measure the gene expression levels of thousands of genes simultaneously over different time points and different experiments. This powerful tool in biotechnology has been utilized in many biomedical applications such as cancer research, infectious disease diagnosis and treatment, toxicology research, pharmacology research, and agricultural development. The enormous improvement of technology in the last decade makes it possible to simultaneously identify and quantify thousands of genes by their gene expression [1], [2], [3]. The spots on a microarray are segmented from the background to compute the gene expression. The three basic operations to compute the spot intensities are gridding, segmentation and intensity extraction. These operations are used to find the accurate location of the spot, separate spot Foreground (FG) from Background (BG) and to calculate of the mean red and green intensity ratio for gene expression. In the last decade, several software packages and algorithms have been developed for segmenting spots in microarray images. Fixed circle segmentation was the first algorithm used in ScanAlyze software where all spots are considered to be circular with a predefined fixed radius [4]. An adaptive circle segmentation technique was employed in the GenePix software [5], where the radius of each spot was not considered constant but adapts to each spot separately. Dapple software estimated the radius of the spot using the laplacian based edge detection [6].

An adaptive shape segmentation technique was used in the Spot software [7]. A histogram-based segmentation method was used in the ImaGene software [8]. Later Watershed [9] and the Seeded region growing algorithms were employed [10] in microarray image analysis. The disadvantages of the above mentioned software packages and algorithms were either the spots were considered to be circular in shape or it was a prerequisite to know the precise position of the spot's center [11]. Further segmentation algorithms based on the statistical Mann–Whitney test was also used in spot segmentation [12], which assessed the statistical significant difference between the FG and BG.

The model based segmentation technique suggested by Q. Li et al. [13] for segmenting microarray spots uses clustering techniques but removes small disconnected clusters based on some threshold assuming that they are artifacts. The spot segmentation using the Markov random field method (MRF) [14], [15] utilizes neighbor information, along with intensity information based on an MRF modeling of the compartment. Although this combination of intensity and spatial information results in a more accurate pixel classification process, it requires an initial classification of the pixels, which in turn affects the final results. Another similar approach is the segmentation method included in the Matarray toolbox of Matlab [16], which also combines both spatial and intensity information. A disadvantage of this method is that it requires input parameters in order to segment the spots.

A 3-D model for spot segmentation was introduced by Eleni Zacharia and Dimitris Maroulis [17]. In this method each real spot of the cDNA microarray image is represented in a 3D space by a 3D model. Since this optimization of 3D space model is using genetic algorithm it is time consuming. Two segmentation methods, Fuzzy gaussian mixture model (FGMM) [18] and Wavelet markov random field (WMRF) model [19] were proposed by Emmanouil I. et al. for segmenting cDNA microarray images. These algorithms were applied to each cell with the purpose of discriminating FG from BG.

Shape-independent segmentation approaches regard pixel intensities in a compartment as a single unit which includes the basic clustering and fuzzy clustering algorithm such as K-means [20], [21], Hybrid k-means [22] and Fuzzy c means (FCM) [23]. The main drawbacks of these methods are their sensitivity to noise. The improved fuzzy clustering techniques such as Possibilistic fuzzy *c* mean (PFCM) [24] and Fuzzy local

information c means (FLICM) [25] solve the noise sensitivity problem of the basic clustering algorithm to some extent, but still need further improvement. The Genetic algorithm based fuzzy c means (GAFCM) [26] method was applied to cDNA microarray images by the authors for segmenting microarray spots. This algorithm improves the FCM algorithm with the help of an optimization technique which results in better segmentation of microarray spots compared to FCM, but needs more computation time since genetic algorithm is used for optimization.

Machine learning based DNA microarray image gridding, a novel method for automatic gridding of cDNA microarray images based on the maximization of the margin between rows and columns of the spots, was introduced by D. Bariamis et al. [27]. This algorithm first estimated the microarray image rotation, followed by spot detection. Then, a set of grid lines were placed on the image in order to separate each pair of consecutive rows and columns of the selected spots. The optimal positioning of the lines was determined by maximizing the margin between these rows and columns by using a maximum margin linear classifier [27], [28].

The aim of microarray image processing is to find the gene expression from each spot. The block diagram shown in Fig.1 explains the steps involved in the calculation of gene expression from a microarray image.

In this paper, an improved fuzzy clustering algorithm which can perform better spot segmentation in the presence of noise, than the existing PFCM and FLICM algorithm is proposed. Gridding of microarray image is done using an algorithm developed by the authors [29]. The fully automatic gridding method developed enhances spot intensity in the preprocessing step as per a histogram based threshold method. The gridding step finds co-ordinates of spots from horizontal and vertical profile of the image. To correct errors due to the grid line placement, a grid line refinement technique is used [29]. Fuzzy clustering is one of the most significant techniques that is used for microarray spot segmentation. The fuzzy clustering algorithms such as FCM, PCM, PFCM and FLICM were used for spot segmentation of cDNA microarray images. The proposed method PFLICM is an improved version of fuzzy clustering algorithm. The PFCM, FLICM and PFLICM algorithm are coded in MATLAB [30]. For evaluation and testing of the algorithm both simulated and real microarray images are used. The performance of PFCM, FLICM and PFLICM algorithms are tested by evaluating the SMF, (p_e), D, and NMSE.

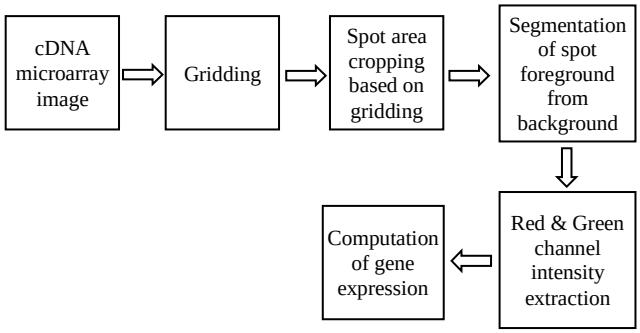


Fig.1. Block diagram of microarray image processing

1.1 MOTIVATION

In FLICM [25], [31], G_{ij} is used to measure the damping extent of the neighboring pixels with the spatial distances from the central pixel. The analysis of the fuzzy factor G_{ij} shows that the spatial information and the local gray-level information are represented by the spatial distance and the gray-level difference, respectively.

Furthermore, the local spatial relationship changes adaptively according to spatial distances from the central pixel. For the neighborhood pixels having same gray level value, the greater the spatial distance, the smaller the damping extent, and vice versa. However, the spatial distance used to measure the damping extent of the neighbors may be unreasonable in some cases. Two situations which give unreasonable result are mentioned. 1. A sub image with the central pixel corrupted by noise, whereas the other adjacent pixels are not corrupted and is homogenous. 2. A sub image with the central pixel not noisy but some adjacent pixels are corrupted by noise. In the above cases, the gray level differences between the neighboring pixels and the central pixel are different. For accurate estimation of the fuzzy factor, the damping extent of the neighboring pixels is to be treated separately. However, the damping extent of the neighboring pixels which is a function of the spatial distance, fails to analyze the impact of each neighboring pixel on the fuzzy factor.

In the proposed PFLICM, to overcome the above mentioned shortcomings of FLICM, the typicality of the pixel from PFCM algorithm is incorporated along with fuzzy factor in FLICM [31]. PFLICM solves the noise sensitivity defect and overcomes the coincident clustering problem. It incorporates typicality, local gray level and local spatial information in a fuzzy way to obtain robustness and noise insensitivity. The factor G_{ij} controls the influence of the neighborhood pixels depending on their distance from the central pixel. To make the proposed algorithm independent of the types of noise and make it a better choice for clustering pixels in the absence of a priori knowledge of the noise, the typicality of the pixel t_{ij} is included [31].

2. POSSIBILISTIC FUZZY LOCAL INFORMATION C MEANS (PFLICM)

The PFLICM is an improved fuzzy clustering algorithm which incorporates typicality of a pixel, local gray level and spatial information in grouping pixels in an image into different clusters. Let $x = x_i$ (i = 1 to N) be the pixels of a single microarray spot, where N is the total number of pixels present in the spot image. These pixels have to be clustered in two classes BG and FG. Let C_j (j =1, 2) be the prototype cluster centers of the BG and FG pixels respectively. A membership function u_{ij} represents the membership value of each pixel to be in different clusters. Based on the maximum value of the membership function each pixel is grouped. The cluster centers C_1 and C_2 are updated iteratively based on the grouped pixel. PFLICM is an iterative clustering algorithm that produces an optimal C partition by minimizing the weight within group sum of squared error objective function F^t.

$$F^t = \sum_{i=1}^{N}\sum_{j=1}^{c}\left[\left(u_{ij}^m + t_{ij}^\eta\right)d_{ij} + G_{ij}\right] + \sum_{j=1}^{c}\gamma_i\sum_{j=1}^{N}\left(1-t_{ij}\right)^\eta. \quad (1)$$

The aim of this method is to minimize the absolute value of the difference between the two consecutive objective functions F^t and $F^t + 1$ given by the Eq.(2),

$$\left\|F^{t+1} - F^t\right\| \le \varepsilon \quad (2)$$

where, ε is a small value close to zero.

In Eq.(1), d_{ij} is the Euclidean distance from a pixel to a cluster center and is given by,

$$d_{ij} = \left\|x_i - c_j\right\|^2 \quad (3)$$

where, t_{ij} is called the typicality of the pixel, each row of t_{ij} is interpreted as the possibility distribution over x. t_{ij} is calculated by Eq.(4) [24], [31]

$$t_{ij} = \frac{1}{1+\left(d_{ij}/\gamma_i\right)^{1/(m-1)}} \quad (4)$$

where, γ is a constant and is given by Eq.(5),

$$\gamma_i = K\frac{\sum_{i=1}^{N}u_{ij}^m d_{ij}}{\sum_{i=1}^{N}u_{ij}^m} K > 0 \quad (5)$$

where, K is a constant greater than zero and u_{ij} is the membership function.

In order to enhance the insensitivity to noise, a new factor G_{ij} is included in PFLICM objective function [25], [31]. This factor incorporates local gray level and local spatial information in a fuzzy way so as to obtain robustness and noise insensitivity, and also control the influence of the neighborhood pixels depending on their distance from the central pixel

$$G_{ij} = \sum_{i=N_k}\frac{1}{d_{ki}+1}\left(1-u_{ij}\right)^m\left\|x_i - c_j\right\|^2 \quad (6)$$

where, the k^{th} pixel is the center of the local window, j is the reference cluster and the i^{th} pixel belongs to the set of neighbours falling into a window around the k^{th} pixel (N_k). d_k, j is the spatial Euclidean distance between pixels k and j, u_{ij} is the degree of membership of the i^{th} pixel in the j^{th} cluster, m is the weighting exponent on each fuzzy membership, and c_j is the prototype of the centre of cluster j.

The membership function u_{ij} is given as,

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}+G_{ij}+t_{ij}}{d_{ik}+G_{ik}+t_{ik}}\right)^{1/(m-1)}}. \quad (7)$$

The prototype center c_j is updated using Eq.(8)

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

The objective function F^t has to be minimized to find the optimum cluster centers.

Finally the PFLICM algorithm is given as follows:

Step 1: Initialize the cluster centers c_j, fuzzification parameters (m and η) and the stopping condition (ε).

Step 2: Initialize randomly membership function u_{ij} and γ_i.

Step 3: Initialize randomly the typicality matrix t_{ij}.

Step 4: Set the loop count iter = 0.

Step 5: Calculate t_{ij} using Eq.(4).

Step 6: Compute fuzzy partition membership degree matrix u_{ij} using Eq.(7).

Step 7: Update cj using Eq.(8).

Step 8: If the max $\{\|F_t + 1 - F_t\|\} \le \varepsilon$ then stop otherwise iter = iter + 1 and go to step 5.

3. DATABASE USED FOR EVALUATION

The evaluation was carried out on two types of databases referred to as Type 1 and Type 2. The proposed algorithm is also applied on real microarray images downloaded from UNC microarray database.

3.1 TYPE 1

In order to objectively compare the proposed segmentation method with fuzzy clustering methods such as PFCM and FLICM, a dataset of synthetic cDNA microarray images is used [32]. This dataset contains 50 Good-quality images (GQI) and 50 Low-quality images (LQI) for which the ground truth is known. Each image having 1000 spots is digitized at 330 × 750 pixels. It was produced by the microarray simulator of Nykter, which generates synthetic cDNA microarray images with realistic characteristics [33]. The good-quality images have low variability in spot sizes and shapes, and their noise level is reasonably low. On the contrary, the low-quality images contain spots whose shape and size vary significantly. The noise levels are significantly higher in the low-quality images. This dataset is used here since it is used by different authors for comparing various established segmentation techniques [34].

3.2 TYPE 2

3.2.1 Synthetic Database:

A set of 40 microarray images, each with 225 spot, are simulated by the authors as mentioned in the literature [18], [19] for numerically evaluating and comparing the 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 area of all simulated spots are a priori determined. The intensities of each FG region are drawn from a uniform distribution using the mean FG intensities of the respective spots in the original image. The remaining BG pixels are drawn from a uniform distribution whose mean intensity is determined from the original image using the entire binary mask image. Note, all the BG intensities is drawn from a single distribution while FG intensities of each target region are drawn from an uniform

distribution whose mean is estimated separately from the original respective spot region. The spots in microarray images sometimes exhibited doughnut-like shapes. During the simulation the doughnut holes which are identified as BG during thresholding, have the same intensity distributions as the BG.

3.2.2 Real Microarray Images:

Real microarray images with category cell line & sub category drug treatment are downloaded from the UNC microarray data base [35]. Each image consists of 34 blocks or sub arrays with each block containing 625 spots. From the downloaded images we have arbitrary selected 25 microarray blocks, i.e. total 15625 spots are used for segmentation. Although the ground truth is not known, it is clear from the segmentation result that the proposed method is more efficient in segmenting the real microarray spots.

4. MEASURES USED FOR EVALUATION

To compare the performance of the proposed algorithm applied on Type 1 and Type 2 databases, with PFCM and FLICM, the following parameters are used. The Type 1 database is evaluated with two statistical parameter such as p_e and D [17], [34] and the Type 2 synthetic database with SMF, p_e, D, and NMSE.

The accuracy of the segmentation is examined with the statistical parameter probability of error p_e, which measures the missegmented pixels, and is defined as,

$$p_e = P(F)P\left(\frac{B}{F}\right) + P(B)P\left(\frac{F}{B}\right) \tag{9}$$

where, $P\left(\frac{B}{F}\right)$ is the probability of error in classifying foreground pixel as background pixels, $P\left(\frac{F}{B}\right)$ is the probability of error in classifying background pixels as foreground pixels, $P(F)$ and $P(B)$ are a priori probabilities of foreground and background pixels in the image. The minimum value of zero occurs for p_e then all of the pixels of the spots are segmented correctly. A maximum value of one for p_e indicates a situation where all of the pixels of the background are segmented as foreground and vice versa.

The parameter discrepancy distance D, which gives different weights for missegmented pixel, based on how spatially far they are located from the nearest correct segmentation result.

It is defined as,

$$D = \frac{1}{A}\sqrt{\sum_{i=1}^{N} d(i)^2} \tag{10}$$

where, N is the number of missegmented pixels, $d(i)$ is the Euclidian distance from the i^{th} missegmented pixels to the nearest pixel that actually belongs to the missegmented class. A is the total number of pixels in the image.

The Type 2 synthetic database images are corrupted by Additive white gaussian noise (AWGN) [18, 19] with the Signal-to-noise ratio (SNR) ranging from 1 to 10dB. The segmentation ability of the proposed algorithm is compared with PFCM and FLICM by finding the SMF, p_e, and NMSE for every binary spots produced by these clustering algorithms. The SMF [17]-[19], [34] for every binary spot, produced by the clustering algorithm is given by,

$$SMF = \frac{A_{seg} \cap A_{act}}{A_{seg} \cup A_{act}} * 100 \tag{11}$$

where, A_{seg} is the area of the spot, as determined by the proposed algorithm and A_{act} is the actual spot area. A perfect match in the case of *SMF* is indicated by a 100% score, any score higher than 50% indicates reasonable segmentation where as a score less than 50% indicate poor segmentation [18], [19].

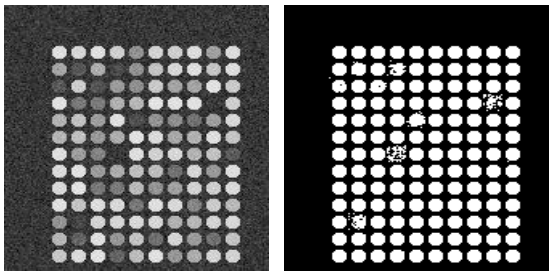
For a simulated image with known ground truth, another metric called NMSE [34] is used to measure the performance of the proposed approach which is given by,

$$NMSE = \frac{\sqrt{\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}\left(x_{ij} - \overline{x_{ij}}\right)^2}}{\frac{1}{MN}\sum_{i}^{M}\sum_{j}^{N}\left(x_{ij}\right)} \tag{12}$$

where, M and N are the dimensions of the image. x_{ij} and $\overline{x_{ij}}$ are the original and clustered image pixels respectively. NMSE is calculated for varying noise levels in the input image. A minimum value of zero is desirable for better segmentation.

5. RESULTS AND DISCUSSION

After validating the PFLICM algorithm and ensuring its correctness, it is applied on the Type 1 and Type 2 databases. The evaluation results of the proposed method on Type 1 are shown in Table.1. It includes the results of ten established segmentation techniques, as reported by Lehmussola et al. [2] and Eleni Zacharia et al. [17].

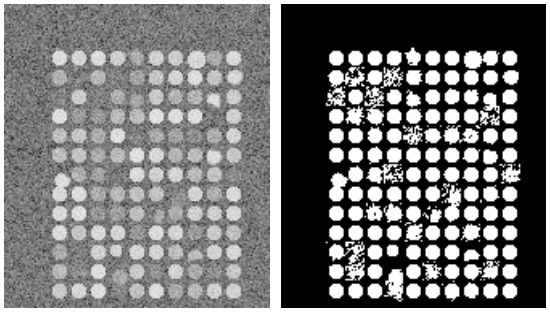


(a) (b)

Fig.2. (a). A block in a good-quality artificial microarray image (b). Spot-segmentation result using proposed method

Comparing K-means, 3D modeling, PFCM and FLICM in the case of HQI, the proposed method yields the same results. However, in the case of LQI the proposed method is better. The significant number of spots used for evaluation additionally supports these arguments. The evaluation is done on more than 50000 artificial microarray spots for which the correct segmentation result is known. The Fig.2(a) illustrates the a microarray block taken from a good-quality synthetic image, while Fig.3(a) illustrates a microarray block taken from a low-quality synthetic image. From Fig.2(b) and Fig.3(b) it is observed that the proposed approach has near optimally segmented all the microarray spots of Fig.2(a) and most of the microarray spots of Fig.3(a). It can be seen that the

proposed method has not segmented any spurious spot. The results show that the proposed segmentation algorithm can segment spots of any shape.



(a) (b)

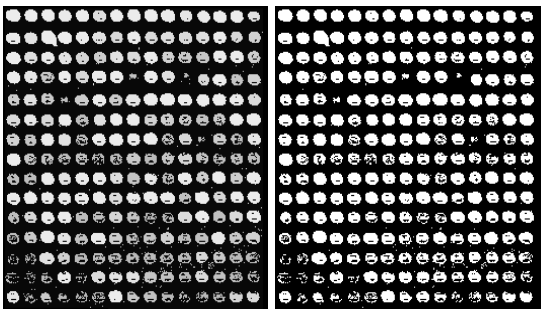
Fig.3. (a). A block in a low-quality artificial microarray image (b). Spot-segmentation result using proposed method

Table.1. The pixel based accuracy measurement of the segmentation

Algorithm	Discrepancy distance (D)		Probability of error (p_e)	
	GQI	LQI	GQI	LQI
Fixed Circle [4]	0.049	0.049	0.027	0.027
Adaptive Circle [5]	0.019	0.192	0.017	0.074
Seed region Growing [10]	0.099	0.114	0.037	0.048
Mann-Whitney [11]	0.165	0.162	0.066	0.074
Hybrid K-means [22]	0.017	0.02	0.016	0.029
Markov random field [14]	0.154	0.053	0.063	0.039
Matarray [16]	0.004	0.031	0.008	0.068
Model Based Segmentation [13]	0.094	0.101	0.052	0.067
K-means [21]	0.000	0.025	0.000	0.041
3D Spot modeling [17]	0.000	0.012	0.000	0.018
PFCM [24]	0.000	0.018	0.000	0.021
FLICM [25]	0.000	0.01	0.000	0.014
Proposed Method	0.000	0.009	0.000	0.011

The proposed algorithm is also applied on a set of 40 Type 2 synthetic microarray images, each with 225 spots which was simulated as mentioned in the literature [18, 20]. The Table.2 summarizes the average performance results obtained for segmenting spots of 40 simulated microarray images (10000 spots). The Fig.4(a) shows a simulated microarray image and Fig.4(b) shows the segmentation result obtained using the proposed algorithm.

The segmentation algorithm is applied on each image after applying the AWGN noise. The SNR value of noise is varied from 1 to 10dB. The performance measurement parameter such as SMF, p_e, and NMSE achieved for all simulated spots corresponding to different SNR levels are presented in Table.2. Regarding the SMF, the PFLICM algorithm resulted in higher spot area identification accuracy than PFCM and FLICM.



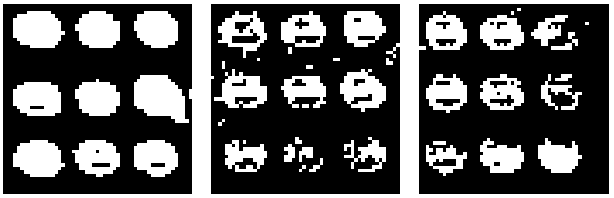
(a) (b)

Fig.4. (a). A microarray simulated image with 225 spots, (b). Segmentation result obtained for the proposed method

The ultimate goal of the segmentation process in microarray image processing is to obtain intensity measurement. Accurate segmentation of spot has a great impact on the intensity calculation. Measurements based on the pixel intensity, rather than the segmentation area such as p_e and NMSE, support the superiority of the PFLICM against PFCM and FLICM. When evaluating the results in intensity extraction perspective a lower value of p_e and NMSE are expected for the better performance of the algorithm [12]. Hence the proposed algorithm is better compared to other algorithms. The Fig.5(a)-(c) shows three subimages, from a simulated image added with AWGN noise (SNR value 5dB). The subimages includes good quality, low quality and doughnut shaped spots. The Fig.6(a)-(c) shows the segmented output of Fig.5(a)-(c) obtained using the proposed algorithm. The result shows the strength of the proposed algorithm in segmenting multi shaped and spurious spots.

(a) (b) (c)

Fig.5(a)-(c) Three subimages each with 9 spots cropped from a simulated microarray image added with AWGN noise of SNR=5db



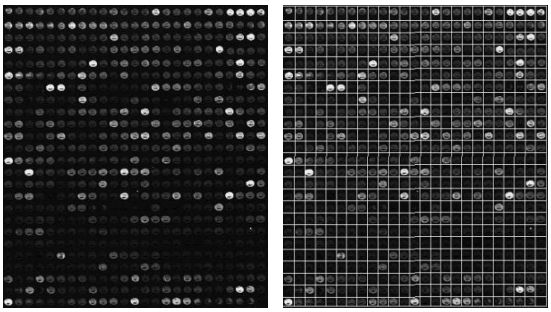
(a) (b) (c)

Fig.6(a)-(c). Segmentation output obtained for the proposed algorithm for the subimage shown in Fig.5(a-c) respectively

Table.2. The comparison of PFCM, FLICM, PFLICM algorithm based on segmentation matching factor (SMF), Probability of error (p_e) and Normalized mean square error (*NMSE*) for simulated microarray images with different levels of additive white Gaussian noise SNR(dB)

	SMF			p_e			NMSE		
SNR(dB)	**PFCM**	**FLICM**	**PFLICM**	**PFCM**	**FLICM**	**PFLICM**	**PFCM**	**FLICM**	**PFLICM**
1	66.282	81.864	82.746	0.126	0.049	0.040	0.337	0.181	0.172
2	72.513	86.230	87.136	0.096	0.030	0.028	0.275	0.138	0.128
3	78.322	90.359	91.239	0.071	0.027	0.022	0.217	0.096	0.086
4	84.153	92.641	93.546	0.038	0.020	0.016	0.158	0.074	0.064
5	89.329	94.677	95.573	0.028	0.015	0.009	0.107	0.053	0.043
6	93.799	95.854	96.654	0.017	0.009	0.007	0.062	0.041	0.031
7	96.363	96.780	97.400	0.008	0.007	0.005	0.036	0.034	0.024
8	97.032	97.332	98.197	0.006	0.006	0.004	0.03	0.027	0.016
9	98.169	98.469	99.091	0.004	0.004	0.002	0.018	0.015	0.012
10	98.552	98.652	99.410	0.003	0.003	0.001	0.014	0.013	0.010

Real microarray images downloaded from the UNC microarray data base is used for segmenting spots FG from BG. All downloaded images are having 34 sub arrays with 625 spots in each sub array. The proposed algorithm is applied on 25 such sub arrays. The Fig.7(a) and Fig.7(b) shows a real image sub array obtained from the UNC microarray data base and its gridded image respectively. The Fig.8 shows the segmented result obtained using the proposed method.

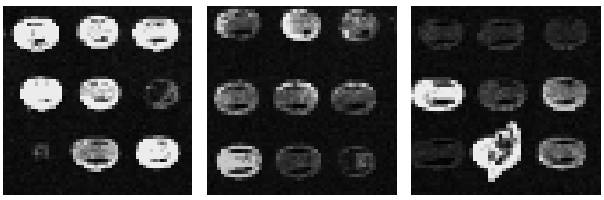


(a) (b)

Fig.7. (a) Real cDNA microarray image, (b) Gridded image of real cDNA microarray image

Fig.8. Segmentation result obtained for the proposed method

The Fig.9(a)-(c) shows three 3 × 3 subimages selected from real microarray image with high intensity, low intensity, doughnut shaped and spurious spots. The Fig.10(a)-(c) shows the segmentation results obtained using the proposed algorithm. The results show that the proposed method segments all high intensity, low intensity and doughnut shaped spots successfully even when the spots are not easily distinguishable. The method also properly segments most of the spurious spots. So the proposed algorithm is a good choice for segmenting real microarray image spots and computation of gene expression.



(a) (b) (c)

Fig.9(a)-(c). Three subimages each with 9 spots cropped from a real microarray image

(a) (b) (c)

Fig.10(a)-(c) Segmentation output obtained using the proposed algorithm for the subimage shown in Fig.9(a)-(c) respectively

6. CONCLUSION

An improved version of fuzzy clustering technique, PFLICM is proposed in this paper. The PFLICM algorithm is applied on simulated and real microarray images. The results obtained support the method strength in reliable measurement of gene expression values from real microarray images. The simulated microarray images include good quality, low quality images and synthetic images whose ground truth value is known. The number of spots used for the evaluation purpose and measures obtained support the superiority of the proposed method over other existing standard methods in microarray image processing.

For real microarray images, more than 15000 spots are segmented for calculating the gene expression values.

The combination of gridding step developed by the authors and spot segmentation step proposed in this paper will make the microarray image analysis automatic.

REFERENCES

[1] Yee Hwa Yang, Michael J. Buckley, Sandrine Dudoit and Terence P. Speed, "Comparison of Methods for Image Analysis on cDNA Microarray Data", *Journal of Computational and Graphical Statistics*, Vol. 11, No. 1, pp. 108-136, 2002.

[2] A. Lehmussola, P. Ruusuvuori and O. Yli-Harja, "Evaluating the Performance of Microarray Segmentation Algorithms", *Bioinformatics*, Vol. 22, No. 23, pp. 2910-2917, 2006.

[3] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray", *Science*, Vol. 270, No. 5235, pp. 467-470, 1995.

[4] M. B. Eisen, "ScanAlyze", http://rana.lbl.gov/Eisen Software.htm,1999.(Last accessed in March, 2010)

[5] "GenePix 4000A User's Guide", Axon Instruments Inc., Foster City, CA.

[6] Jeremy Buhler, Trey Ideker and David Haynor, "Dapple: Improved Techniques for Finding Spots on DNA Microarrays", Technical Report UWTR 2002-08-05, University of Washington, 2000.

[7] M.J. Buckley, "The spot user's guide CSIRO Mathematical and Information Science", http://www.cmis. csiro.au/IAP/Spot/spotmanual.html, 2000. (Last accessed in March, 2010)

[8] ImaGene, "ImaGene 6.1 User Manual", http://www.biodiscovery.com/index/papps-webfiles-action. (Last accessed in March, 2010)

[9] Serge Beucher and Fernand Meyer, "The Morphological Approach to Segmentation: The Watershed Transformation. Mathematical Morphology in Image Processing", *Optical Engineering*, Vol. 34, pp. 433-481, 1993.

[10] R. Adams and L. Bischof, "Seeded Region Growing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 6, pp. 641-647, 1994.

[11] D. Bozinov and J. Rahenfuhrer, "Unsupervised Technique for Robust Target Separation and Analysis of DNA Microarray Spots through Adaptive Pixel Clustering", *Bioinformatics*, Vol. 18, No. 5, pp. 747-756, 2002.

[12] Yidong Chen, Edward R. Dougherty and Michael L. Bittne, "Ratio-Based Decisions and the Quantitative Analysis of Cdna Microarray Images", *Journal of Biomedical Optics*, Vol. 2, No. 4, pp. 264-374, 1997.

[13] Q. Li, C. Fraley, R. E. Bumgarner, K.Y. Yeung and A.E. Raftery, "Donuts, Scratches and Blanks: Robust Model-based Segmentation of Microarray Images", *Bioinformatics*, Vol. 21, No. 12, pp. 2875-2882, 2005.

[14] O. Demirkaya, M.H. Asyali and M.M. Shoukri, "Segmentation of cDNA Microarray Spots using Markov Random Field Modeling", *Bioinformatics*, Vol. 21, No. 13, pp. 2994-3000, 2005.

[15] R. Lukac, K.N. Plataniotis, B. Smolka and A.N. Venetsanopoulos, "A Multichannel Order-Statistic Technique for cDNA Microarray Image Processing", *IEEE Transactions on NanoBioscience*, Vol. 3, No. 4, pp. 202-214, 2004.

[16] X. Wang, S. Ghosh and S.W. Guo, "Quantitative Quality Control in Microarray Image Processing and Data Acquisition", *Nucleic Acids Research*, Vol. 29, No. 15, e75, 2001.

[17] E. Zacharia and D. Maroulis, "3-D Spot Modeling for Automatic Segmentation of cDNA Microarray Images", *IEEE Transactions on NanoBioscience*, Vol. 9, No. 3, pp. 181-192, 2010.

[18] E.I. Athanasiadis, D.A. Cavouras, P.P. Spyridonos, D.T. Glotsos, I.K. Kalatzis and G.C. Nikiforidis, "Complementary DNA Microarray Image Processing based on the Fuzzy Gaussian Mixture Model", *IEEE Transactions on Information Technology in Biomedicine*, Vol. 13, No. 4, pp. 419-425, 2009.

[19] E.I. Athanasiadis, D.A. Cavouras, P.P. Spyridonos, D.T. Glotsos, I.K. Kalatzis and G.C. Nikiforidis, "A Wavelet-based Markov Random Field Segmentation Model in Segmenting Microarray Experiments", *Computer Methods and Programs in Biomedicine*, Vol. 104, No. 3, pp. 307-15, 2011.

[20] K. Blekas, N. Galatsanos, A. Likas and I.E. Lagaris, "Mixture Model Analysis of DNA Microarray Images", *IEEE Transactions on Medical Imaging*, Vol. 24, No. 7, pp. 901-909, 2005.

[21] Shuanhu Wu and Hong Yan, "Microarray Image Processing Based on Clustering and Morphological Analysis", *Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics*, Vol. 19, pp. 111-118, 2003.

[22] Jorg Rahneenfuhrer and Daniel Bozinov, Hybrid Clustering for Microarray Image Analysis Combining Intensity and Shape Features", *BMC Bioinformatics*, Vol. 5, pp. 47-58, 2004.

[23] V. Uslan and I.O. Bucak, "Clustering-based Spot Segmentation of cDNA Microarray Images", *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1828-1831, 2010.

[24] N.R. Pal, K. Pal, J.M. Keller and J.C. Bezdek, "A Possibilistic Fuzzy c-Means Clustering Algorithm", *IEEE Transactions on Fuzzy Systems*, Vol. 13, No. 4, pp. 517-530, 2005.

[25] Stelios Krinidis and Chatzis Vassilios, "A Robust Fuzzy Local Information C-means Clustering Algorithm", *IEEE Transactions on Image Processing*, Vol. 19, No. 5, pp. 1328-1337, 2010.

[26] V.G. Biju and P. Mythili, "A Genetic Algorithm Based Fuzzy C Mean Clustering Model for Segmenting Microarray Images", *International Journal of Computer Applications*, Vol. 52, No. 11, pp. 42-48, 2012.

[27] Dimitris Bariamis, Michalis Savelonas and Dimitris Maroulis, "*Machine Learning-Based DNA Microarray Image Gridding*", Microarray Image and Data Analysis Theory and Practice, pp. 109-127, 2014.

[28] Dimitris Bariamis, Dimitris K. Iakovidis and Dimitris Maroulis, "M^3G: Maximum Margin Microarray Gridding", *BMC Bioinformatics*, Vol. 11, No. 49, pp. 1-11, 2014.
[29] V.G. Biju and P. Mythili, "Microarray Image Gridding Using Grid Line Refinement Technique", *ICTACT Journal on Image and Video Processing*, Vol. 5, No. 4, pp. 1010-1016, 2015.
[30] The MathWorks, http://www.mathworks.com
[31] V.G. Biju and P. Mythili, "Fuzzy Clustering Algorithms for cDNA Microarray Image Spots Segmentation", *Procedia Computer Science*, Vol. 46, pp. 417-424, 2015.
[32] http://www.cs.tut.fi/sgn/csb/spotseg/
[33] Matti Nykter, Tommi Aho, Miika Ahdesmäki, Pekka Ruusuvuori, Antti Lehmussola and Olli Yli-Harja, "Simulation of Microarray Data with Realistic Characteristics", *BMC Bioinformatics*, Vol. 7, pp. 349-366, 2006.
[34] Yu-Ping Wang , Maheswar Gunampally, Jie Chen, Douglas Bittel, Merlin G. Butler and Wei-Wen Cai, "A Comparison of Fuzzy Clustering Approaches for Quantification of Microarray Gene Expression", *Journal of Signal Processing Systems*, Vol. 50, No. 3, pp. 305-320, 2008.
[35] UNC Microarray database, https://genome.unc.edu/