International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Denoising Based Clustering Algorithm for Segmentation of Microarray Image

J. Harikiran
Member IEEE and Asst. Prof., GIT, GITAM University, Visakhapatnam, India

M. l. Phanendra
Student, GIT, GITAM University, Visakhapatnam

P. Naga Srinivasu
Student, GIT, GITAM University, Visakhapatnam

Dr. P. V. Lakshmi
Professor, Department of IT, GIT, GITAM University, Visakhapatnam, India

Dr. R. Kiran Kumar
Asst. Prof., Deptt. of CS, Krishna University, Machilipatnam, India

Abstract **– A Deoxyribonucleic Acid (DNA) microarray is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip forming an array. The analysis of DNA microarray images allows the identification of gene expressions to draw biological conclusions for applications ranging from genetic profiling to diagnosis of cancer. The DNA microarray image analysis includes three tasks: gridding, segmentation and feature extraction. Clustering algorithms have been applied for segmenting the microarray image. However, noises are introduced into the images during acquisition or transmission process, affecting the segmentation results. Noise reduction is a prerequisite step prior to feature extraction attempts from microarray images. In order to overcome this drawback, this paper presents a new clustering based segmentation technique that can be used in segmenting noise microarray images. We call this method as Denoising Fuzzy Moving K-means Clustering algorithm (DFMK). The proposed algorithm is able to minimize the effect of Salt-and-Pepper noise during the segmentation process without degrading the fine details of the images. The method incorporates a noise detection stage to the clustering algorithm, producing an adaptive segmentation technique specifically for segmenting the noise microarray images. The results obtained from the proposed algorithm are more quantitative and qualitative than the conventional clustering algorithm such as K-means in segmenting the noise microarray images.**

Keywords **- Microarray, Salt-and-Pepper Noise, Image Segmentation, Image Processing.**

I. Introduction

Microarrays, are the next revolution in molecular biology, enable scientists to analyze genes, proteins and other biological molecules on a genomic scale [1]. A microarray is a collection of spots containing set of DNA probes deposited on the solid surface of glass slide. Each of the spot contains multiple copies of single DNA sequence [2].

Microarray expression technology helps in the monitoring of gene expression for tens and thousands of genes in parallel. During the biological experiment, the mRNA of two biological tissues of interest is extracted and purified. Each of the mRNA samples are reverse transcribed into complementary DNA (cDNA) copy and labeled with two different fluorescent dyes resulting in two fluorescence-tagged cDNA (red Cy5 and green Cy3). The tagged cDNA copies, called the sample probe, are hybridized with the slide's DNA spots. The hybridized glass slides are fluorescently scanned at different wavelengths (corresponding to the different dyes used), and two digital images are produced, one for each population of mRNA. Each digital image contains a number of spots of various fluorescence intensities. The intensity of each spot is proportional to the hybridization level of the cDNAs and the DNA dots, the gene expression information is obtained by analyzing the digital images [3].

The processing of the microarray images [4] usually consists of the following three steps: (i) gridding, which is the process of segmenting the microarray image into compartments, each compartment having only one spot and background (ii) Segmentation, which is the process of segmenting each compartment into one spot and its background area (iii) Feature extraction, quantifies the area of interest's intensity measure along with the features local background. This process also calculates a host of other metrics to facilitate downstream analysis [13]. The evaluation of microarray images is a difficult task as the natural fluorescence of the glass slide and non-specifically bounded DNA molecules add a substantial noise floor to the microarray image, which would affect the gridding and segmentation results [5].

At present, there is different image clustering methods [6] used for segmenting the microarray images by various researches. All the methods can be applied to noise microarray images only after performing pre-processing tasks like filtering algorithm. To tackle this problem, we propose a new clustering algorithm for segmenting noise microarray images by incorporating the noise detection stage within the clustering algorithm to introduce an adaptive clustering based segmentation technique. The adaptive behavior enables the clustering algorithm to segment the noisy image properly even in the occurrence of noise without going through any filtering stage beforehand. The inherited noise detection behavior will improve the segmentation results of microarray image by only selecting noise-free pixels for the process of segmentation. First, the gridding of noise microarray is performed using the method specified in [11], each compartment having only one spot and background and segment each compartment into one spot and its background area using the proposed method.

This paper is organized as follows. In Section II the Fuzzy Moving k-means Clustering Algorithm is presented. The proposed adaptive clustering based segmentation technique is explained in section III. Section IV analyses the results obtained from the proposed algorithm with different noise densities by using qualitative and quantitative methods.

International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

II. Fuzzy Movingk-Means Clustering Algorithm

Consider N as the number of pixels to be clustered into n_c clusters. Each pixel should have a degree of membership to clusters rather than belonging to one cluster. The membership degree of a pixel u_{ij} is a value in (0,1). The sum of all membership values of a pixel belonging to clusters should satisfy the following constraint

$$\sum_{j=1}^{c} u_{ij} = 1, \quad \forall\ i=1,2,....N \tag{1}$$

where c is the number of clusters and N is the number of pixels in microarray image.

Let v_i be the i^{th} data and c_j be the j^{th} center with predetermined initial value where i=1,2,….N and j=1,2, …..n_c. For the FMK algorithm, the objective function of clustering an image into c_j clusters is given by

$$F = \sum_{j=1}^{n_c} \sum_{i=1}^{N} u_{ij}^m d_{ij} \tag{2}$$

Where d_{ij} is the Euclidean distance from a pixel to a cluster center given as $d_{ij} = \|v_i - c_j\|^2$ and m, the fuzzy exponent, is an integer, m>1. All the data will be assigned to the nearest center based on Euclidean distance.

The fitness for each cluster and the membership value of i^{th} data to the cluster center is then calculated using:

$$f(c_j) = \sum_{i \in c_j} (\|v_i - c_j\|^2) \text{ and } u_{ij} = \frac{1}{\sum_{k=1}^{n} \left(\frac{d_{ij}}{d_{jk}}\right)^{\frac{2}{m-1}}} \tag{3}$$

After specifying the membership for each data and applying the fitness calculation process, the relationship among the centers must fulfill the following condition.

$$f(C_s) \geq \alpha_a f(C_l) \text{ and } u(C_{si}) > u(C_{li}) \tag{4}$$

Where $f(C_s)$ is the cluster that has the smallest fitness value, $f(C_l)$ is the cluster that has largest fitness value, α_a is a designed small constant with initial value equal to $0 < \alpha_a < (1/3)$, $u(C_{si})$ is the membership value of i^{th} data according to the smallest center and $u(C_{li})$ is the membership value of i^{th} data according to the largest center.

If (4) is not fulfilled, the members of C_l which are larger than C_l are assigned as members of C_s while the rest are maintained as members of C_l. Then the position of C_l and C_s are recalculated using:

$$C_s = \frac{1}{n_{c_s}} \sum_{t \in c_s} v_t \quad \text{and} \quad C_l = \frac{1}{n_{c_l}} \sum_{t \in c_l} v_t \tag{5}$$

The value of α_a is updates according to $\alpha_a = \alpha_a - \alpha_a / n_c$.

$$(6)$$

The above process is repeated until (4) is fulfilled.

III. Adaptive Clustering Based Segmentation Technique

We propose an adaptive clustering based segmentation technique for the microarray images corrupted with salt-and-pepper noise. The proposed clustering technique consists of two stages: The first stage involves the noise detection in the image and in the second stage, we perform the clustering process. The noise free pixels will give full contribution on the clustering process, whereas for the noisy pixels, the fuzzy concept is applied to determine the degree of contribution on the clustering process.

First Stage: Noise Detection

In salt and pepper type of noise, the noisy pixels takes either salt value *(gray level - 225)* or pepper value *(gray level - 0)* and it appears as black and white spots on the microarray images with certain probabilities [12]. In this step, a binary noise mask is constructed for the noise microarray image Y. When the gray level images is contaminated with salt-and-pepper noise, a noisy pixel takes either a maximum intensity value (I_{max}= 255) or a minimum intensity value (I_{min}= 0). This dynamic range [I_{max} , I_{min}] provide information about the noisy pixels in the image. The binary noise mask *N(i,j)* is constructed by assigning a binary value 1, if the intensity of the pixel located at position (i, j) in the noisy image is I_{max} or I_{min}, otherwise assign a binary value 0.

The binary noise mask is computed from the noisy image as follows:

$$N(i,j) = \begin{cases} 0, \text{ if } Y(i,j) = Imax \\ 0, \text{ if } Y(i,j) = Imin \\ 1, \text{ otherwise} \end{cases} \tag{7}$$

Where *Y(i,j)* is the pixel at location *(i,j)* with intensity *Y*, *N(i,j)=1* represents the 'noise-free' pixel to be retained in the clustering stage while *N(i,j)=0*, represents 'noise' pixels.

Second Stage: Clustering Process

In the second stage, a more versatile and powerful method of clustering-based segmentation of noise microarray image is developed. Before assigning each pixel in the microarray image to their nearest center, check whether the pixel is 'noise-free' or 'noise' by the use of binary noise mask. If the pixel *Y(i,j)* is noise, the absolute difference *G(i+k,j+l)* between the neighboring pixels and the central pixel *Y(i,j)* in 3X3 window is calculated using :

$$G(i+k,j+l) = |Y(i+k,j+l) - Y(i,j)| \text{ with } k,l \in (-1,0,1)$$

and

$$Y(i+k,j+l) \neq Y(i,j) \tag{8}$$

The maximum value of the absolute difference among the eight neighboring pixels of *Y(i,j)* in the 3x3 window will be used as the fuzzy gradient value. The fuzzy set processes the neighborhood information represented by the fuzzy gradient value to estimate a correction term which aims at cancelling the noise. Mathematically, the fuzzy set *F(i,j)* which is taken from [7][8][9] is given by

$$F(i,j) = \begin{cases} 0 & : 0 \le \max(G(i,j)) < T_1 \\ \dfrac{\max(G(i,j)) - T_1}{T_2 - T_1} & : T_1 \le \max(G(i,j)) < T_2 \\ 1 & : otherwise \end{cases} \quad (9)$$

Where $T_1 and T_2$ are the thresholds to perform partial correction, and set to 10 and 30 respectively as described in [7]. The correction term $Y^1(i,j)$ for replacing the current pixel $Y(i,j)$ is given by

$$Y1(i,j) = (1 - F(i,j)) * Y(i,j) + F(i,j) * m_{ij}, \quad (10)$$

where m_{ij} is the median of noise-free pixels in the 3x3 window. For each detected "noise pixel", the size of filtering window is initialized to 3x3. If the current filtering window does not have a minimum number of one "noise-free pixel", then the filtering window will be expanded by one pixel at each of its four sides. This procedure is repeated until minimum of one "noise–free pixel" criterion is met. The search for the noise free pixels is halted when the filtering window reached a size of 7x7. When the filtered window has reached the size of 7x7 although no "noise-free pixel" is detected, then the first four pixels in the 3x3 filtering window is used to compute m_{ij}.

$$mij = median\{ Y(i-1,j-1), Y(i,j-1), Y(i+1,j-1), Y(i-1,j)\}. \quad (11)$$

To increase the robustness of FMK clustering towards noise, these corrected values (for noise pixels) are used to replace the original pixels values during the process of assigning the data to their nearest center. The term v_i in (2) is substituted by

$$vi = \begin{cases} Y(i,j) & \text{if } N(i,j)=1 \\ Y1(i,j) & \text{if } N(i,j)=0 \end{cases} \quad (12)$$

By employing this concept into Fuzzy Moving k-means Clustering Algorithm, the new proposed algorithm is called Denoising Fuzzy moving k-means Clustering Algorithm.

IV. Experimental Results

Qualitative Analysis:

In this section, we present the experimental results of the proposed algorithm on a microarray image contaminated by different levels of salt-and-pepper noise. The proposed method is performed on a microarray image that consists of a total of 38808 pixels. The segmentation results by the proposed algorithm are shown in figure 3. Figure 1 shows the microarray image corrupted with 10 %, 20%, 30%, 40% and 50% density of salt-and-pepper noise. Figure 2 shows the segmentation results produced by K-means clustering algorithm for the noisy images in figure 1. From the results, we can visualize that the results produced in segmenting the microarray image by the clustering algorithms (K-means, Fuzzy C-means) is influenced by the noise, which indicates that the algorithms are less robust to noise mixture. In this paper, we use K-means clustering algorithm for comparison with the proposed algorithm. The noise contamination has highly affected the clustering process in the clustering algorithm which contributes to poor segmentation results. The proposed algorithm is seen to have successfully minimized the influence of noise from affecting the segmentation process and achieved satisfactory results.

Quantitative Analysis:

In this section, we have tabulated a quantitative evaluation of the clustering result by using the function [10]

$$F(I) = \sqrt{p}\, \frac{\sum_{i=1}^{p} e_i^{\,2}}{\sqrt{a_i}} \quad (13)$$

Where I is the clustered image to be evaluated, p is the number of regions found, a_i is the size of i-th region, $p(a_i)$ is the number of regions with the area a_i and e_i is defined as the sum of Euclidean distances between the features of pixels of region I and the corresponding region in the clustered image. The values of the function for the clustered microarray image using k-means and proposed method are tabulated in table 1. The findings can therefore conclude that the proposed method serve as a better approach for segmenting noise microarray image.

Table 1: Evaluation of function F(I) for two different algorithms

Noise Density	K-means (2-clusters)	Proposed Method (2-clusters)
10	74.8323	22.4939
20	76.0528	32.0731
30	81.7272	44.0232
40	88.8398	52.2130
50	92.2579	56.6151

V. Conclusion

This paper presents new adaptive clustering based segmentation technique for segmentation of noise-corrupted microarray image. The qualitative and quantitative analysis favor the proposed algorithm, producing better results as compared to the K-means clustering algorithm through its inclusion of the noise detection stage in the clustering process. This stage can reduce the effect of noise during the clustering process. In addition the proposed algorithm has also successfully preserved important features on microarray images. This method is a good technique for segmentation of noise microarray image without going through any filtering stage.

References

[1] M.Schena, D.Shalon, Ronald W.davis and Patrick O.Brown,"Quantitative Monitoring of gene expression patterns with a complementary DNA microarray", Science, 270,199,pp:467-470.

[2] Wei-Bang Chen, Chengcui Zhang and Wen-Lin Liu, "An Automated Gridding and Segmentation method for cDNA Microarray Image Analysis", 19th IEEE Symposium on Computer-Based Medical Systems.

[3] Tsung-Han Tsai Chein-Po Yang, Wei-ChiTsai, Pin-Hua Chen, "Error Reduction on Automatic Segmentation in Microarray Image", IEEE 2007.

[4] Eleni Zacharia and Dimitirs Maroulis, "Microarray Image Analysis based on an Evolutionary Approach" 2008 IEEE.

[5] Jin Hyuk Kim, Hye Youns Kim and Yong Sung Lee, " A novel method using edge detection for signal extraction from cDNA microarray image analysis", Experimental and Molecular Medicine, Vol 33, No.2, 83-88, June 2001.

[6] Volkan Uslan, Ihsan Omur Bucak, "Clustering based Spot Segmentation of cDNA Microarray Images", IEEE 2010.

[7] W.Lou, " Efficient Removal of Impulse Noise From Digital Images", IEEE Transaction on Consumer Electronics, Vol. 52, no.2, pp.523-527, 2006.

[8] K.K.V.Toh, H.Ibrahim, and M.N.Mahyuddin, " Salt and Pepper Noise Detection And Reduction Using Fuzzy switching Median Filter", IEEE Transactions on Consumer electronics, vol.54, no.4, pp.1956-1961, 2008.

[9] Kenny Kal Vin Toh, Nor Ashidi Mat Isa," Noise Adaptive Fuzzy Switching Median Filter for Salt-and-Pepper Noise Reduction", IEEE Signal Processing Letters, Vol 17, No 3, March 2010, pp 281-285.

[10] J.Liu, Y.H.Yang, " Multiresolution color image segmentation", IEEE Transaction on Pattern and Machine Inteligence, Vol. 16, no.7, pp 689-700, 1994.

[11] Eleni Zacharia, Dimitris Maroulis," A Precise and Automatic Gridding Approach to Noise-Affected and Distorted Microarray Images, Proceeding CBMS '08 Proceedings of the 2008 21st IEEE International Symposium on Computer-Based Medical Systems Pages 605-607.

[12] M.Premakumar, J.Harikiran, B.Saichandana, Dr.P.Rajesh Kumar, "Performance Evaluation of Image Fusion for Impulse Noise Reduction in Digital Images Using an Image Quality Assessment, IJCSI , International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011, pp 407-411.

[13] "Spot Edge Detection of Microarray Images Using Bi-dimensional Empirical Mode Decomposition",C3IT -2012, ELSEVEIR, Procedia Technology, Vol 4: pp227-231.

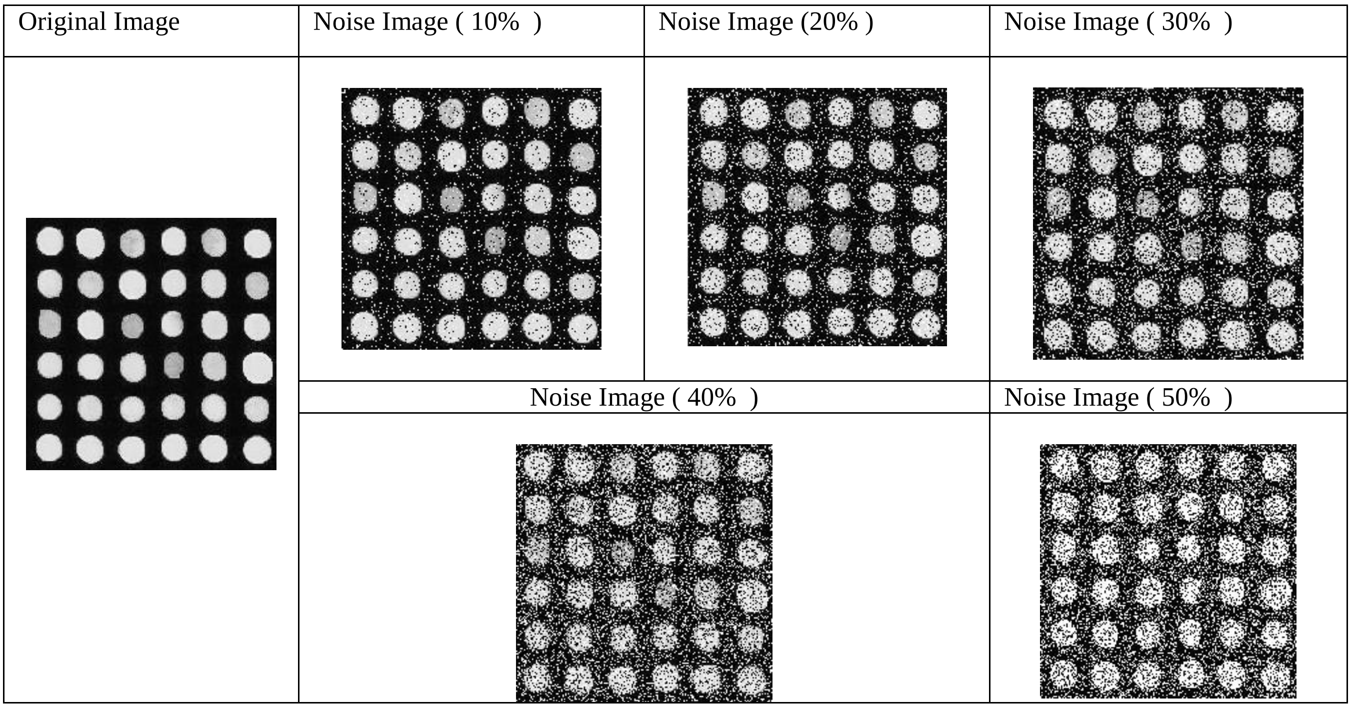


Fig.1. Microarray image corrupted by salt- and-pepper noise with different probabilities

International Journal of Electronics Communication and Computer Engineering
Volume 3, Issue 6, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

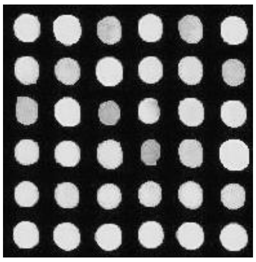
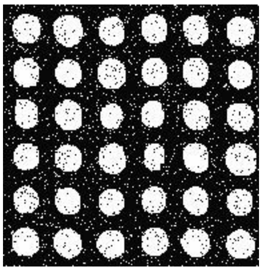
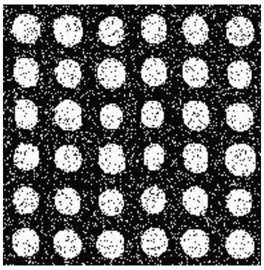
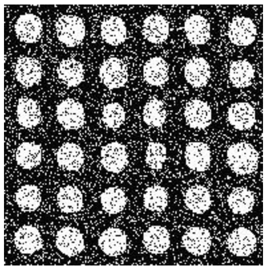
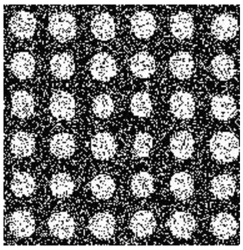
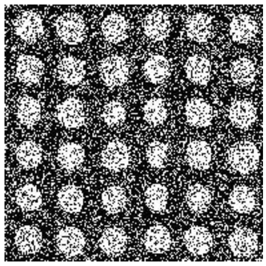
Original Image	Noise Image (10%)	Noise Image (20%)	Noise Image (30%)
			
	Noise Image (40%)		Noise Image (50%)
			

Fig.2. Results by K-means Clustering Algorithm on noise images

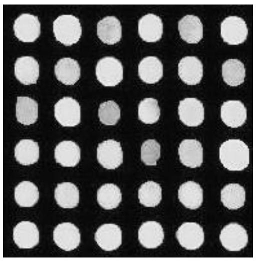
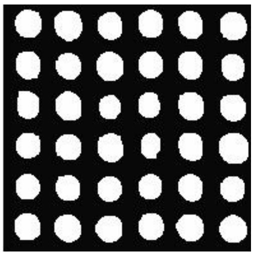
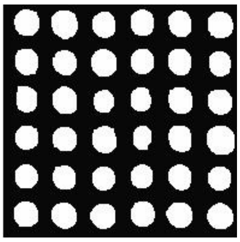
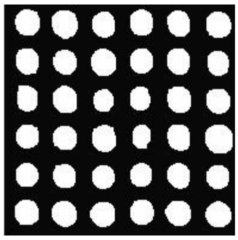
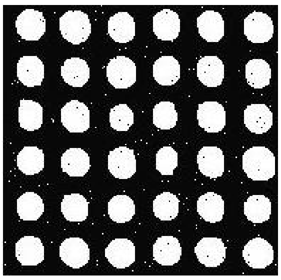
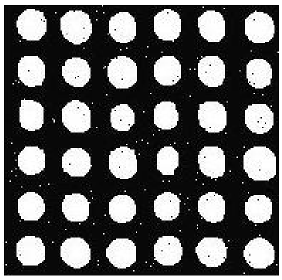
Original Image	Noise Image (10%)	Noise Image (20%)	Noise Image (30%)
			
	Noise Image (40%)		Noise Image (50%)
			

Fig.3. Results obtained by proposed method on noise images