Microarray Image Analysis Using Genetic Algorithm

B. Sivalakshmi[1], N. Naga Malleswara Rao[2]
[1]Acharya Nagarjuna University, Guntur, A.P, India
[2]RVR and JC College of Engineering, Guntur, A.P, India
*Corresponding author, e-mail: jkr@btp.or.id

Abstract

Microarray technology allows the simultaneous monitoring of thousands of genes. Based on the gene expression measurements, microarray technology have proven powerful in gene expression profiling for discovering new types of diseases and for predicting the type of a disease. Gridding, segmentation and intensity extraction are the three important steps in microarray image analysis. This paper presents microarray image analysis using Genetic Algorithm. A new algorithm for microarray image contrast enhancement is presented using Genetic Algorithm. Contrast enhancement is crucial step in extracting edge information in image and finally this edge information is used in gridding of microarray image. Mostly segmentation of microarray image is carried out using clustering algorithms. Clustering algorithms have an advantage that they are not restricted to a particular shape and size for the spots. In this paper, segmentation using Genetic Algorithm by optimizing K-means index and Jm measure is presented. The qualitative analysis shows that the proposed method achieves better segmentation results than K-means and FCM algorithms.

1. Introduction

The most powerful tool in molecular genetics for biomedical research is Microarray, which allows parallel analysis of the expression level of thousands of genes. The most important aspect in microarray experiment is image analysis. The output of image analysis is a matrix consisting of a measure of intensity of each spot in the image. This measure denotes gene expression ratio (transcription abundance) between the test and control samples for the corresponding gene. The positive expression indicates the over-expression, while negative expression indicates under-expression between the control and treatment genes. The main components in microarray image analysis are localization, segmentation and spot quantification [1]. This paper mainly focuses on enhancement and segmentation of microarray image using Genetic Algorithm (GA). The segmentation of microarray using GA is done by optimizing two objective functions, KMI (K-means index) and Jm measure, while the enhancement is done by using Entropy value as optimization function. The main applications of microarray technology are Gene discovery, Drug discovery, Disease diagnosis, Toxicological research etc. The microarray image analysis is shown in figure 1.
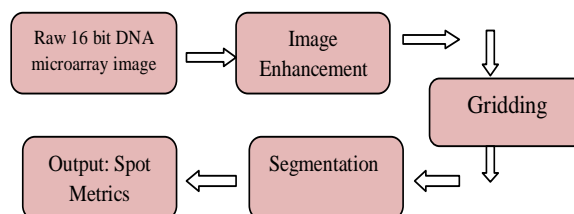

Figure 1. Microarray Image Analysis

Currently clustering algorithms are used for segmentation of microarray image. Clustering algorithms has the advantages over other methods such as non-dependency of

shape of spot, does not require any prior knowledge of the image objects and sensitivity to noisy pixels [2]. Till now Clustering algorithms such as K-means and FCM are used for segmentation of microarray image. In this paper, GA based segmentation of microarray is presented. The quantitative values shows that GA based segmentation achieved better segmentation than K-means and FCM.

The paper is organized as follows: Section 2 presents Microarray image enhancement using GA, Section 3 presents Gridding of Microarray Image, Section 4 presents Segmentation of Microarray image using GA, Section 5 presents Experimental results and Section 6 reports conclusions.

2. Microarray Image Enhancement Using Ga:

Bi-dimensional Empirical Mode Decomposition: Empirical mode decomposition [3] is a signal processing method that nondestructively fragments any non-linear and non-stationary signal into oscillatory functions by means of a mechanism called shifting process. These oscillatory functions are called Intrinsic Mode Functions (IMF), and each IMF satisfies two properties, (a) the number of zero crossings and extrema points should be equal or differ by one. (b) Symmetric envelopes (zero mean) interpret by local maxima and minima. The signal after decomposition using EMD is non-destructive means that the original signal can be obtained by adding the IMFs and residue. The first IMF is a high frequency component and the subsequent IMFs contain from next high frequency to the low frequency components. The shifting process [4] used to obtain IMFs on a 2-D signal (image) is summarized as follows:
Let I(x,y) be a Image used for EMD decomposition. Find all local maxima and local minima points in I(x,y).

1) Upper envelope Up(x,y) is created by interpolating the maxima points and lower envelope Lw(x,y) is created by interpolating minima points. This interpolation is carried out using cubic spline interpolation method.
2) Compute the mean of lower and upper envelopes denoted by Mean(x,y).

$$Mean(x, y) = \frac{(Up(x, y) + Lw(x, y))}{2}$$

(1)

3) This mean signal is subtracted from the input signal.

$$Sub(x, y) = I(x, y) - Mean(x, y)$$

(2)

4) If Sub(x,y) satisfies the IMF properties, then an IMF is obtained .

$$IMF_i(x, y) = Sub(x, y)$$

(3)

5) Subtract the extracted IMF from the input signal. Now the value of I(x,y) is

$$I(x, y) = I(x, y) - IMF_i(x, y)$$

(4)

6) Repeat the above steps (b) to (f) for the generation of next IMFs.

7) This process is repeated until I(x,y) does not have maxima or minima points to create envelopes. Original Image can be reconstructed by inverse EMD given by

$$I(x, y) = \sum_{i=1}^{n} IMF_i(x, y) + res(x, y)$$

(5)

Image Contrast Enhancement: If the contrast of the microarray image is low, the quality of the edges extracted from the image will be poor. This edge information is primary source for automatic gridding of microarray image [5]. The quality of the spot edges can be improved by applying this GA based contrast enhancement algorithm to the original image prior to the computation of Gridding.

1) First divide the microarray image into Intrinsic Mode Functions (IMF) by using Bi-dimensional Empirical Mode Decomposition. $BEMD(F) = [imf_1, imf_2, ..., imf_n]$.\

2) The enhanced image is obtained by the summation of IMFs, with each IMF multiplied by weight w_i. The weights for IMFs are optimized using GA.

$$EMI = \sum_{i=1}^{n} w_i * imf_i \tag{6}$$

Where w_i denotes the weight of ith imf and EMI represents the enhanced microarray image.

3) The Genetic Algorithm is applied as follows.

a) Randomly generate chromosomes representing initial population with values indicating the weights of IMFs. The length of chromosome is equal to the number of IMFs generated from BEMD, and each element represents the corresponding IMF weight. The values of weights differ between 0 and 1 and the summation of weights is equal to 1.

b) Calculate the fitness value of each chromosome using information entropy of image as objective function. The information entropy is defined as

$$\text{Entropy} = -\text{sum}(p.*\log2(p)) \tag{7}$$

c) Based on Entropy values, the chromosomes are selected using roulette wheel technique.

d) To generate new chromosomes from the current population, two operators such as crossover and mutation are used with probabilities 0.7 and 0.07.

e) Maximum number of iterations is used as stopping criteria. The values in the chromosome are used as optimum weights in constructing the enhanced image.

3. Microarray Image Gridding

Gridding is the process of dividing the microarray image into blocks (sub-gridding) and each block again divided into sub-blocks (spot-detection). The final sub-block contains a single spot and having only two regions spot and background. Existing algorithms for gridding are semi-automatic in nature requiring several parameters such as size of spot, number of rows of spots, number of columns of spot etc. In this paper, a fully automatic gridding algorithm designed in [6] is used for sub-gridding and spot-detection.

4. Microarray Image Segmentation Using Ga

It is the process of grouping the pixels of image in sub-block into two non-overlapping regions, spot region and back-ground region. Clustering algorithms are mostly used for segmentation of microarray image. In microarray image, the number of clusters used to segment the image is two, foreground and background regions. If the number of clusters is three, the third cluster denoted edge pixels.

Genetic Algorithms [7] belong to the class of evolutionary algorithms that are based on principles of natural selection and genetics. It is a search technique used in computing true solutions to optimization problems that is driven by natural evolution process. GA performs parallel search of the solution space rather than point by point search. Genetic Algorithm consists of three operators namely, Selection, Crossover and Mutation.

The Genetic Algorithm mechanism can be abstracted as follows [8].

1) The initial population of solutions is randomly generated across the search space.
2) Using an objective function, the fitness of each individual solution in the population is evaluated.
3) Using this fitness values, the solutions in the population are selected.
4) New population is created from selected solutions using the crossover and mutation operators.
5) The new population is replaced instead of old population.
6) Repeat iteratively from (2) to (5) until a stop criterion is satisfied. Each iteration of this GA process is called generation.

GA is a method of parallel search of the solution space based on two assumptions inspired by evolutionary biology. 1) The measure of problem solving ability by an individual in the population is determined by its fitness value. 2) New individuals which are obtained by combining different individuals in the population have more problem solving ability.
Image segmentation using GA:

The Genetic Algorithm is applied as follows.

a) Assume P chromosomes in the population where P is the size of the population. Each chromosome is encoded with K cluster centers that are randomly selected from the image.
b) Using an objective function, the fitness value of each chromosome is evaluated. Three Different indices, such as K-means index (KMI) and Jm measure are used as objective functions individually. For computing the measures, the centers z_1, z_2, ..., z_k encoded in a chromosome are first extracted. The membership values u_{ik}, i=1,2,....K and k=1,2,....n are computed as follows:

$$u_{ik} = \frac{1}{\sum_{j=1}^{K} \left(\frac{D(z_i, x_k)}{D(z_j, x_k)} \right)^{\frac{2}{m-1}}}$$

(8)

Where $D(z_i, x_k)$ is the Euclidean distance between two points x_k and cluster center z_i. The centers encoded in a chromosome are updated using the following equation:

$$z_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}, \qquad 1 \leq i \leq K$$

(9)

The J_m measure [9][14] which is to be minimized is defined as

$$J_m = \sum_{j=1}^{n} \sum_{k=1}^{K} u_{kj}^m D^2(x_j, z_k), \qquad 1 \leq m \leq \infty$$

(10)

Where m is the fuzzy exponent, D denotes the Euclidean distance between two points x_j and z_k and u_{kj} denotes the membership values.
The k-means index [10] which is used as the objective function in this GA process is defined as follows:

$$KMI = \frac{1}{\sum_{k=1}^{K} \sum_{i=1}^{N} \|x_i - z_k\|^2}$$

(11)

Where K number of clusters and z_k is the cluster centers

c) The selection of chromosomes is done based on the fitness value using roulette wheel technique.
d) By applying crossover and mutation operators with rate 0.8 and 0.07, a new population is produced from the parents. This new population replaces the old population.Maximum number of iterations is used as stopping criteria.

After the execution stops, the highest fitness value chromosome is selected and the values in this chromosome represent the solution to the segmentation of image.

5. Experimental Results

The proposed enhancement and segmentation algorithm using GA is performed on two microarray images drawn from the standard microarray database corresponds to breast category aCGH tumor tissue. Image 1 consists of a total of 38808 pixels and Image 2 consists of 64880 pixels. Image enhancement using GA is performed prior to gridding in-order to improve the contrast of image. Gridding is performed on the input images by the method proposed in [6], to segment the image into compartments, where each compartment is having only one spot region and background. The gridding output is shown in figure 2. Segmentation algorithm is applied to each compartment for segmenting the foreground and background region. The output of the GA based segmentation method with Jm measure as objective function on a compartment from image 1 and image 2 is shown in figure 3.
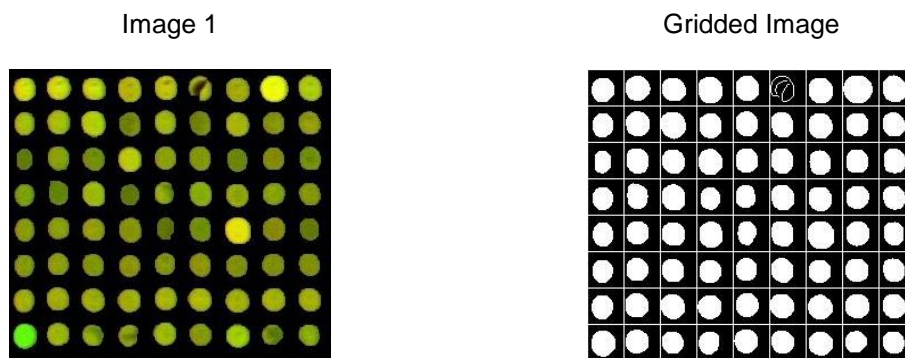
Quantitative Analysis: Quantitative analysis is a numerically oriented procedure to figure out the performance of algorithms without any human error. The Mean Square Error (MSE) [11] is significant metric to validate the quality of image. It measures the square error between pixels of the original and the resultant images. The MSE is mathematically defined as

$$MSE = \frac{1}{N} \sum_{j=1}^{k} \sum_{i \in c_j} ||v_i - c_j||^2 \tag{12}$$

Where N is the total number of pixels in an image and xi is the pixel which belongs to the jth cluster. The lower difference between the resultant and the original image reflects that all the data in the region are located near to its centre. Table 1 shows the quantitative evaluations of segmentation algorithms. The results confirm that GA with Jm measure as objective produces the lowest MSE value for segmenting the microarray image than k-means [12] and FCM [13].

Table 1. MSE values

Method	MSE Values (Compartment No 1) In image 1	MSE Values (Compartment No 8) In image 2
K-means	282.781	296.47
Fuzzy c-means	216.392	228.69
GA with K-means Index	221.411	232.74
GA with Jm measure	198.327	206.276

Image 1   Gridded Image

Image 2   Gridded Image

Figure 2. Spot –detection

(Compartment No 1)   (Compartment No 8)
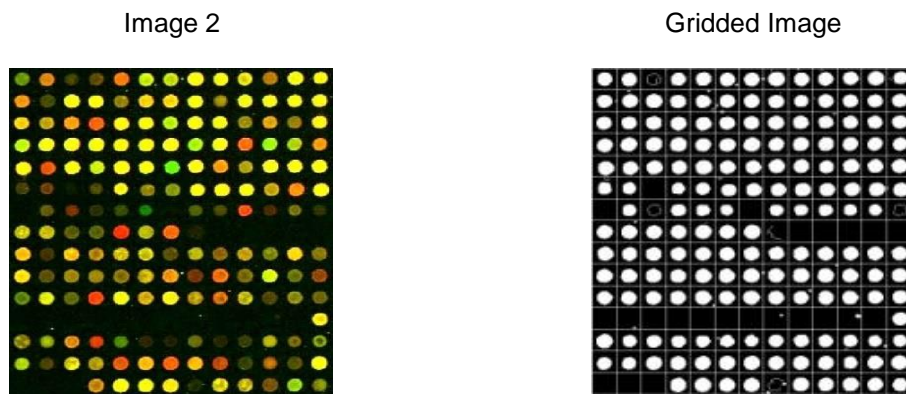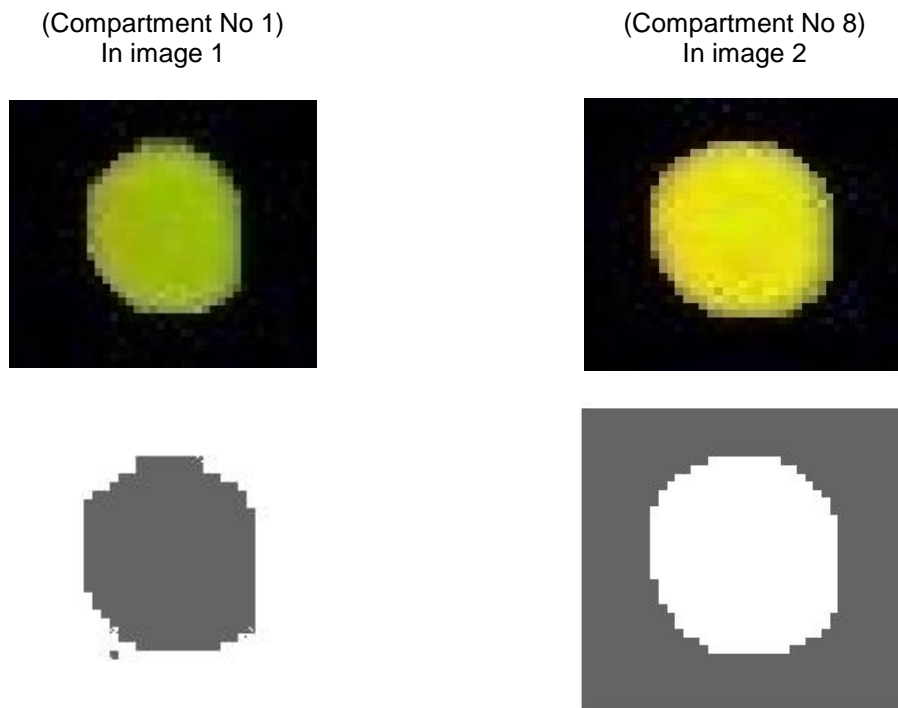In image 1   In image 2

Figure 3. Segmentation using GA with Jm measure

6. Conclusions

Microarray technology is used for parallel analysis of gene expression ratio of different genes in a single experiment. The analysis of microarray image is done with gridding, segmentation and information extraction. The expression ratio of each and every gene spot denotes the transcription abundance between two genes under experiment. This paper presents analysis of microarray image using genetic algorithm. The segmentation using GA optimizes two objective functions. Optimizing cluster validity indices will produce better results than K-means and FCM algorithm. Spot information includes the calculation of Expression Ratio in the region of every gene spot on the microarray image. The expression-ratio measures the transcription abundance between the two sample genes

References

[1] J Harikiran, et.al. Vector Filtering Techniques for Impulse Noise Reduction with Application to Microarray images. *International Journal of Applied Engineering Research*". 2015; 10(3): 7181-7193.

[2] J Harikiran, Dr PV Lakshmi, Dr R Kiran Kumar. "Fast Clustering Algorithms for Segmentation of Microarray Images*", International Journal of Scientific & Engineering Research*. 2014; 5(10): 569-574,

[3] J Harikiran, et.al. Spot Edge detection in Microarray Images using Bi-dimensional Empirical Mode Decomposition. C3IT-2012, Elsevier Procedia Technology, (Science Direct). 2012; (4): 19-25,

[4] B Saichandana, J Harikiran, Dr K Srinivas, Dr R Kiran Kumar. Application of BEMD and Hierarchical Image Fusion in Hyper spectral Image Classification, *International Journal of Computer Science and Information Security*. 2016; 14(5): 437-445.

[5] J Harikiran, et.al. A New Method of Gridding for Spot Detection in Microarray Images. *Computer Engineering and Intelligent Systems*. 2014; 5(3): 25-33,

[6] J Harikiran, Dr PV Lakshmi, Dr R Kirankumar. Automatic Gridding Method for microarray images. *Journal of Theoretical and Applied Information Technology*. 2014: 65(1): 235-241.

[7] J Holland (1975)," Adaptation in Natural and Artificial Systems". Ann Arbor, MI: Univ. of Michigan Press.

[8] B Saichandana, et.al. "Image Fusion in Hyperspectral Image Classification using Genetic Algorithm", *Indonesian Journal of Electrical Engineering and Computer Science*. 2016; 2(3): 703-711.

[9] B Sanghamitra, et.al. Multiobjective Genetic Clustering for Pixel Classification in Remote Sensing Imagery, *IEEE Transactions on Geoscience and Remote Sensing*. 2007; 45(5): 1506-1512.

[10] Yeh Fen, Yang, et.al. *"Genetic Algorithms for Multispectral Image Classification",* http://www.isprs.org/proceedings/XXXVI/part3/singlepapers/P_16.pdf.

[11] [11] J Harikiran, Dr PV Lakshmi, Dr R Kiran Kumar. Multiple Feature Fuzzy C-means clustering algorithm for segmentation of microarray image", *International Journal of Electrical and Computer Engineering*. 2015; 5(5): 1045-1053,

[12] B.Saichandana et.al. Clustering Algorithm combined with Hill climbing for Classification of Remote Sensing Image. *International Journal of Electrical and Computer Engineering*. 2004; 4(6): 923-930.

[13] J Harikiran et.al. "Fuzzy C-means with Bi-dimensional empirical Mode decomposition for segmentation of Microarray Imge", *International Journal of Computer Science issues*. 2012; 9(5); 273-279.

[14] R Kiran Kumar, B Saichandana, K Srinivas. "Dimensionality Reduction and Classification of Hyperspectral Images using Genetic Algorithm". *Indonesian Journal of Electrical Engineering and Computer Science*. 2016; 3(3).