Fuzzy C-means with Bi-dimensional Empirical Mode Decomposition for Segmentation of Microarray Image

J.Harikiran[1], D.RamaKrishna[2], M.L.Phanendra[3], Dr.P.V.Lakshmi[4], Dr.R.Kiran Kumar[5]

[1]Member IEEE and Assistant Professor
GIT, GITAM University, Visakhapatnam, India

[2,3,4]Gitam Institute of Technology
GITAM University, Visakhapatnam, India

[5]Assistant Professor
Department of Computer Science
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 intensity extraction. Clustering algorithms have been applied for segmenting the microarray image. Among these approaches, Fuzzy C-means (FCM) method is simple and efficient one. However, microarray image contains noise and the noise would affect the segmentation results. In this paper, we propose to combine the FCM method with Bi-dimensional Empirical Mode Decomposition for segmenting the microarray image in order to reduce the effect of noise. We call this method as Fuzzy C-means with Bi-dimensional Empirical Mode decomposition (FCMBEMD). We use an adaptive local weighted averaging filter in the BEMD method for removing the noise in the image and finally K-means algorithm is used for segmentation of image. Using the FCMBEMD method on microarray image, we obtain better results than those using FCM only. Using the FCMBEMD method to analyze microarray image can save time and obtain more reasonable results.

Keywords: Microarray Image; Bi-Dimensional Empirical Mode Decomposition; Image Segmentation, Image Processing;

1.Introduction

Microarrays, widely recognized as 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 DNA 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) Intensity extraction, which calculates red and green foreground intensity pairs and background intensities.

However, 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 segmentation results [7]. In this paper we propose to combine Fuzzy C-means method with bi-dimensional Empirical Mode Decomposition (BEMD) for segmentation of microarray image. The Empirical Mode Decomposition (EMD) method was first proposed by Huang [5] and then Lin et al. [8] proposed an iterative filtering as an alternative algorithm for EMD. Usually, EMD is used to analyze the intrinsic components of a signal. These components are called intrinsic mode functions (IMF). Most noisy IMFs are considered as noise in the signal. If we remove noisy IMFs from the raw data, the trend component can be obtained. Then we can use the trend as the denoised data to perform clustering analysis. In this paper we denoise the microarray image using BEMD and finally segment the image with Fuzzy C-means clustering algorithm.

The paper is organized as follows: Section II presents Bi-dimensional Empirical Mode Decomposition, Section III presents, Fuzzy C-means Clustering Algorithm, Section IV presents Experimental results, and finally Section V reports conclusion.

2. Bi-Dimensional Empirical Mode Decomposition

The EMD proposed by Dr. Norden Huang [5], was a technique for analyzing nonlinear and non-stationary signals. It serves as an alternative to methods such as wavelet analysis and short-time Fourier transform. It decomposes any complicated signal into a finite and often small number of Bi-dimensional Intrinsic Mode Functions (IMF). The IMF is symmetric with respect to local zero mean and satisfies the following two conditions.

1. The number of extrema and the number of zero crossings must either be equal or differ by one.
2. At any point, the mean value of the envelope defined by local maxima and local minima is zero, indicating the function is locally symmetric.

The decomposition method in EMD is called Shifting Process. The shifting process of the two-dimensional signal such as image can be adapted from the one-dimensional signal as follows.

1. Let $I_{original}$ is defined as an image to be decomposed. Let j=1 (index number of IMF), I= $I_{original}$ (the residue).
2. Identify the local maxima and local minima points in I.
3. By using interpolation, create the upper envelope E_{up} of local maxima and the lower envelope E_{lw} of local minima.
4. Compute the mean of the upper envelope and lower envelope.
   $E_{mean}= [E_{up} + E_{lw}]/2$
5. $I_{imf} = I- E_{mean}$.
6. Repeat steps 2-5 until I_{imf} can be considered as an IMF.
7. IMF(j)= I_{imf}, j=j+1, I = I- I_{imf} ,
8. Repeat steps 2-7 until all the standard deviation of two consecutive IMFs is less than a predefined threshold or the number of extrema in I is less than two.

The first few IMFs obtained from BEMD contain the high frequency components which correspond to salient features in original image and the residue represents low frequency component in the image. The original image can be recovered by inverse BEMD as follows:

$$I = RES+ \sum_{j} IMF(j)$$

(1)

The BEMD process is shown in figure 1.

Lin et al. [8] proposed iterative filtering as an alternative algorithm for EMD. Instead of using the envelopes generated by spline, in the new algorithm Lin use a "moving average" to replace the mean of the envelope. The essence of the shifting algorithm remains the same. We use a low pass filter to generate a moving average replacing the mean of the envelope. The simplest choice of moving average is an adaptive local weighted average. We use the filter

given by $L(x,y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} w(i,j) I(x+i, y+j)$

where $w(i,j)$ is called the mask or filter coefficients at (i,j). We select the mask $w(i,j) = \frac{(a+b) - |(i+j)| + 2}{a+b+2}$, i=-a,…..,a and j=-b,…..,b.

The shifting process (iterative filtering) of the two-dimensional signal such as image can be adapted from the one-dimensional signal as follows.

1. Let $I_{original}$ is defined as an image to be denoised. Let j=1 (index number of IMF), I= $I_{original}$ (the residue).
2. Let L be a low pass filter such that L(I) represents the moving average of I.
3. Now define $I_{imf} = I - L(I)$.
4. Repeat steps 2-3 until I_{imf} can be considered as an IMF.
5. IMF(j)= I_{imf}, j=j+1, I = I- I_{imf} ,
6. Repeat steps 2-5 until IMF has atmost one local maximum or local minimum.

The original image can be recovered by inverse as follows:

$$I = RES + \sum_{j} IMF(j)$$
(2)

3.   Fuzzy C-Means clustering Algorithm

In order to segment the image, the location of each spot must be identified through gridding process. Hirata [9] presented an automatic gridding method by using the horizontal and vertical profile signal of the image to perform the image gridding. The algorithm can satisfy the requirements of microarray image segmentation.

Algorithm Fuzzy C-Means(x,n,c,m)
Input:
N=number of pixels to be clustered;
x = {x_1, x_2,..., x_N}: pixels of microarray image;
c=2: foreground and background clusters;

m=2: the fuzziness parameter;
Output:
u: membership values of pixels
Begin
Step_1: Initialize the membership matrix u_{ij} is a value in (0,1) and the fuzziness parameter m. The sum of all membership values of a pixel belonging to clusters should satisfy the constraint expressed in the following.

$$\sum_{j=1}^{c} u_{ij} = 1 \qquad (3)$$

for all i= 1,2,…….N, where c is the number of clusters and N is the number of pixels in microarray image.

Step_2: Compute the centroid values for each cluster c_j. Each pixel should have a degree of membership to those designated clusters. So the goal is to find the membership values of pixels belonging to each cluster. The algorithm is an iterative optimization that minimizes the cost function defined as follows:

$$F = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^{m} \, || x_j - c_i ||^2 \qquad (4)$$

where u_{ij} represents the membership of pixel x_j in the ith cluster and m is the fuzziness parameter.

Step_3: Compute the updated membership values uij belonging to clusters for each pixel and cluster centroids according to the given formula.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{||x_j - v_i||}{||x_j - v_k||} \right)^{2/(m-1)}},$$

and

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}}.$$

(5)

Step_4: Repeat steps 2-3 until the cost function is minimized.
End.

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
ISSN (Online): 1694-0814
www.IJCSI.org

319

4. Experimental Results

The proposed FCMBEMD method of the microarray image processing are performed on a sample microarray slide that has 48 blocks, each block consisting of 110 spots. A sample block has been chosen and 36 spots of the block have been cropped for simplicity. The sample image is a 198*196 pixel (gray scale) image that consists of a total of 38808 pixels. In-order to check the efficiency of the proposed method, Gaussian noise is added into the microarray image. The noisy microarray image is filtered by an iterative filtering algorithm using EMD. After denoising, in the 9th IMF, we obtain the smooth image, which we used as a denoised microarray image for analysis purpose. The FCM clustering algorithm is used for segmentation of filtered and noisy microarray image. The experimental results shows that noise in the microarray image affects the segmentation results and that segmentation structures still can be found in the microarray image without any biological significance. Because we have removed noise in microarray image, now there is no segment structure in the microarray image without any biological significance. The experimental results of the proposed method are shown in figure 2. The number of pixels clustered as spots and background for original image segmented using fuzzy c-means and segmented image by using the proposed method have been presented in table 1.

Table 1: The number of pixels clustered as spots and background

	Spots	Background
Original Image segmented by Fuzzy C-means	14563	24245
Noisy Image segmented by proposed method	13674	25134

5.Conclusion

In this paper, we proposed to combine Fuzzy C-means with empirical mode decomposition for segmenting the microarray image. The noisy microarray image is first filtered by an iterative filtering algorithm using BEMD. Then, the filtered microarray image is segmented using Fuzzy C-means clustering algorithm. Using the proposed method, there is no segment structure in the microarray image without any biological significance. In the future work, the same BEMD concept can also be applied to any clustering algorithms for segmenting the noisy microarray image.

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. N.E.Huang, Z.Shen, S.R.Long, "The empirical mode decomposition and the Hilbert Spectrum for non-linear and non-stationary time series analysis". Proc. Roy. Soc, London.A, Vol.454, pp.903-995, 1998.
6. Volkan Uslan, Omur Bucak, " clustering based spot segmentation of microarray cDNA Microarray Images ", International Conference of the IEE EMBS , 2010.
7. 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.
8. L. Lin, Y.Wang and H.Zhou, Iterative filtering as an alternative for empirical mode decomposition. Advances in Adaptive Data Analysis. 1(4): 543-560, 2009.
9. R.Hirata, J.Barrera, R.F.Hashinoto and D.o.Dantas, " Microarray gridding by mathematical morphology", in Proceedings of 14th Brazilian Symposium on Computer Graphics and Image Processing, 2001, pp: 112-119
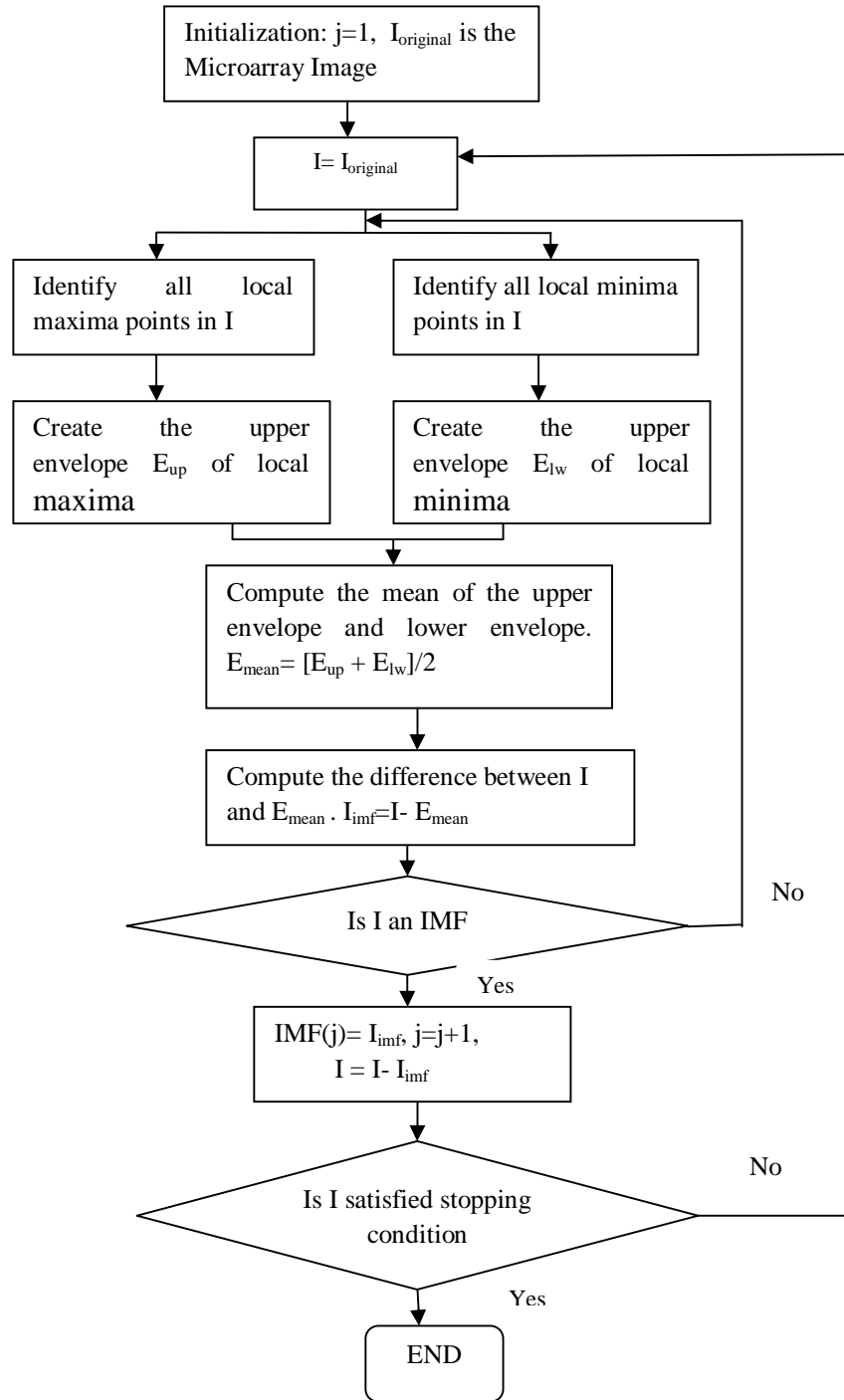
IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
ISSN (Online): 1694-0814
www.IJCSI.org

320

Initialization: j=1, $I_{original}$ is the Microarray Image

$I = I_{original}$

Identify all local maxima points in I

Identify all local minima points in I

Create the upper envelope E_{up} of local maxima

Create the upper envelope E_{lw} of local minima

Compute the mean of the upper envelope and lower envelope. $E_{mean} = [E_{up} + E_{lw}]/2$

Compute the difference between I and E_{mean} . $I_{imf} = I - E_{mean}$

Is I an IMF

No

Yes

$IMF(j) = I_{imf}$, $j = j+1$, $I = I - I_{imf}$

Is I satisfied stopping condition

No

Yes

END

Fig 1 : BEMD Process

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012
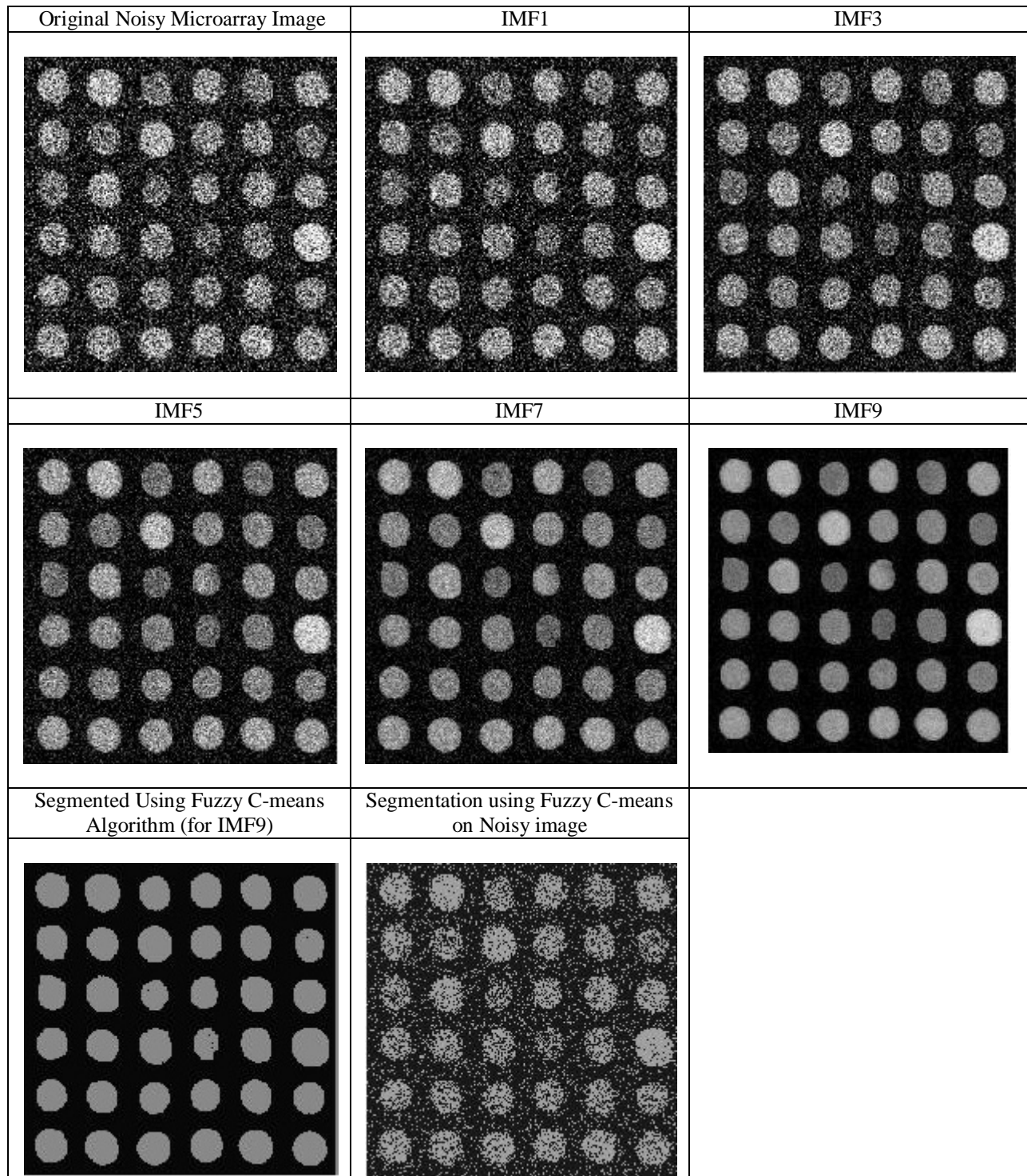ISSN (Online): 1694-0814
www.IJCSI.org

321

Fig 3: Experimental results on a microarray image