

SLIC Superpixel Based Self Organizing Maps Algorithm for Segmentation of Microarray Images

Durga Prasad Kondisetty¹, Mohammed Ali Hussain²,

¹Dept. of Computer Science, Bharathiar University, Tamilnadu, India

²Dept. of Computer Science & Engineering, KL University, Vijayawada, AP, India

Article Info

Article history:

Received Jun 9, 2017

Revised Jan 13, 2018

Accepted Feb 16, 2018

Keyword:

Empirical mode decomposition

Image Segmentation

Microarray Images

Self organizing maps

Simple linear iterative

clustering

ABSTRACT

We can find the simultaneous monitoring of thousands of genes in parallel Microarray technology. As per these 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, Intensity extraction, Enhancement and Segmentation are important steps in microarray image analysis. This paper gives simple linear iterative clustering (SLIC) based self organizing maps (SOM) algorithm for segmentation of microarray image. The clusters of pixels which share similar features are called Superpixels, thus they can be used as mid-level units to decrease the computational cost in many vision applications. The proposed algorithm utilizes superpixels as clustering objects instead of pixels. The qualitative and quantitative analysis shows that the proposed method produces better segmentation quality than k-means, fuzzy c-means and self organizing maps clustering methods.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Durga Prasad Kondisetty,

Research Scholar,

Dept. of Computer Science, Bharathiar University, Tamilnadu, India.

Email: dpkondisetty@gmail.com.

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 analysis of output of image is a matrix consisting of intensity measure of each spot in the image. This is denotes gene expression ratio (transcription abundance) between control samples for the corresponding gene and the gene test. The negative expression indicates under-expression while positive expression indicates the over-expression between the control and treatment genes. The main components in microarray image analysis are localization, segmentation and spot quantification [1]. The main applications of microarray technology are Gene discovery, Drug discovery, Disease diagnosis, Toxicological research etc [2]. The microarray image analysis is shown in figure 1.

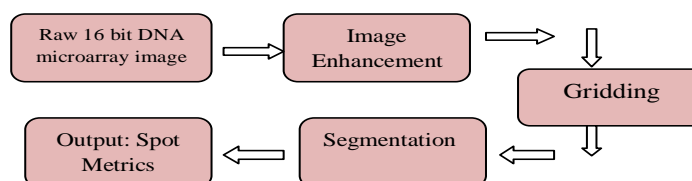


Figure 1: Analysis of Microarray Image

The microarray images is a difficult task as the fluorescence of the glass slide adds noise floor to the microarray image [3] [18]. The processing of the microarray image requires noise suppression with minimal reduction of spot edge information that derives the segmentation process. This paper describes reduce the noise in microarray images using Empirical Mode Decomposition [EMD] method. The BEMD method [5] decomposes the image into several Intrinsic Mode Functions [IMF], in which the first function is the high frequency component, second function next high frequency component and so on; the last function denotes the low frequency component. The mean filter is applied only to the few first high frequency components leaving the low frequency components, as the high frequency components contain noise. The image is reconstructed by combining the filtered high frequency components and low frequency components. After noise removal, segmentation, Expression ratio and gridding calculations are the important tasks in analysis of microarray image. Any noise in the microarray image will affect the subsequent analysis [6].

In the proposed literature of many microarray image segmentation approaches have Fixed circle segmentation [7], Adaptive circle Segmentation Technique [8], Seeded region growing methods [9] and clustering algorithms [10] are the methods that deal with microarray image segmentation problem. This paper mainly focuses on clustering algorithms. These algorithms have the advantages that they are not restricted to a particular spot size and shape, does not require an initial state of pixels and no need of post processing. These algorithms have been developed based on the information about the intensities of the pixels only (one feature). In this paper, SLIC super pixel based self organizing maps clustering algorithm is proposed. The qualitative and quantitative results show that proposed method has segmented the image better than k-means, fuzzy c-means and self organizing maps clustering algorithms.

2. BI-DIMENSIONAL EMPIRICAL MODE DECOMPOSITION-DWT THRESHOLDING METHOD

Empirical mode decomposition [11] 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 functions are called Intrinsic Mode Functions (IMF), and it satisfies two properties, (i) the number of zero crossings and extrema points should be equal or differ by one. (ii) Symmetric envelopes (zero mean) interpret by local maxima and minima [12]. 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 used to obtain IMFs on a 2-D signal (image) is summarized as follows:

- a) Let $I(x,y)$ be a Microarray image used for EMD decomposition. Find all local maxima and local minima points in $I(x,y)$.
- b) Upper envelope $Up(x,y)$ is created by interpolating the maxima points and lower envelope $Lw(x,y)$ is created by interpolating minima points. The cubic spline interpolation method for interpolation is carried out as:
- c) Compute the mean of lower and upper envelopes denoted by $Mean(x,y)$.

$$Mean(x, y) = \frac{(Up(x, y) + Lw(x, y))}{2} \quad (1)$$

- d) This mean signal is subtracted from the input signal.

$$Sub(x, y) = I(x, y) - Mean(x, y) \quad (2)$$

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

$$IMF_i(x, y) = Sub(x, y) \quad (3)$$

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

$$I(x, y) = I(x, y) - IMF_i(x, y) \quad (4)$$

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

- g) 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

$$I(x, y) = \sum_{i=1}^n IMF_i(x, y) + res(x, y) \quad (5)$$

The mechanism of de-noising using BEMD-DWT is summarized as follows

- Apply 2-D EMD for noisy microarray to obtain IMF_i (i=1, 2, ...k). The kth IMF is called residue.
- The first intrinsic mode function (IMF1) contains high frequency components and it is suitable for denoising. This IMF1 is denoised with mean filter. This de-noised IMF1 is represented with DNIMF1.
- The denoised image is reconstructed by the summation of DNIMF1 and remaining IMFs given by

$$RI = DNIMF1 + \sum_{i=2}^k IMF_i \quad (6)$$

where RI is the reconstructed band and the flow diagram of BEMD-DWT filtering is shown in figure

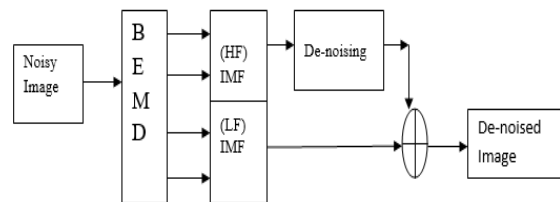


Figure 2: Flow Diagram of BEMD-mean filtering method

3. MICROARRAY IMAGE GRIDDING

The process of dividing the microarray image into blocks (sub-gridding) and each block again divided into sub-blocks (spot-detection) is called Gridding. 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 [13] is used for sub-gridding and spot-detection..

4. SLIC SUPERPIXELS

Simple linear iterative clustering (SLIC) is an adaption of k-means for Superpixel generation, with two important distinctions: i) the number of distance calculations in the optimization is dramatically reduced by limiting the search space to a region proportional to the Superpixel size. This reduces the complexity to be linear in the number of pixels N and independent of the number of superpixels k. ii) A weighted distance measure combines color and spatial proximity, while simultaneously providing control over the size and compactness of the superpixels.

The algorithm of SLIC superpixels generation is given below [14].

- Initialize p initial cluster centers in $C = [k, x, y, r, s]^T$ by sampling pixels at regular grid steps S.
- For generation of equal sized super pixels the grid interval S is given by $S = \sqrt{\frac{N}{p}}$
- Set label $k(j)=-1$ for each pixel j.
- Set distance $d(j) = \infty$ for each pixel j.
- For each cluster center C do
- For each pixel j in a $2S \times 2S$ region around C do
- Compute the distance D between C and j.
- The distance D depends on pixel's color (color proximity) and pixel position (spatial proximity), whose values is known. The value of D is given by

$$d_c = \sqrt{(l_j - l_i)^2 + (x_j - x_i)^2 + (y_j - y_i)^2} d_s = \sqrt{(r_j - r_i)^2 + (s_j - s_i)^2}$$

$$D = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \quad (7)$$

The maximum spatial distance expected within a given cluster should correspond to the sampling interval, $NS = S$. Determining the maximum color distance N_c is not so straightforward, as color distances can vary significantly from cluster to cluster and image to image. The value of N_c in the range from [1, 40].

9. if $D < d(i)$ then set $d(i)=D$ and $k(i)=p$ go to 6.

10. Goto 5, the same process for each cluster

11. Compute new cluster centers.

12. The clustering and updating processes are repeated until a predefined number of iteration is achieved. The SLIC algorithm can generate compact and nearly uniform superpixels with a low computational overhead.

5. FUZZY C-MEANS CLUSTERING ALGORITHM

The FCM algorithm for segmentation of microarray image is described below [15]:

1. Take randomly K initial clusters from the $m*n$ image pixels.

2. Initialize membership matrix u_{ij} with value in range 0 to 1 and value of $m=2$.

Assign each pixel to the cluster C_j $\{j=1,2,\dots,K\}$ if it satisfies the following condition $[D(\cdot, \cdot)]$ is the Euclidean distance measure between two values.

$$u_{ij}^m D(I_i, C_j) < u_{iq}^m D(I_i, C_q), q = 1, 2, \dots, K$$

$$j \neq q \quad (8)$$

The new membership and cluster centroid values as calculated as

$$u_{ik} = \frac{1}{\sum_{j=1}^K \left(\frac{D(C_i, I_k)}{D(C_j, I_k)}\right)^{\frac{1}{m-1}}}, \text{ for } 1 \leq i \leq K \quad (9)$$

$$C_j^{\wedge} = \frac{\sum_{j=1}^n u_{ij}^m I_j}{\sum_{j=1}^n u_{ij}^m}$$

3. Continue 2-3 until each pixel is assigned to the maximum membership cluster [16].

6. SLIC SUPERPIXEL BASED SOM CLUSTERING ALGORITHM

The SLIC algorithm generates superpixels which are used in our clustering algorithm. The superpixels are generated based on the color similarity and proximity in the image plane. The algorithm depends on two values N_s and N_c , the higher value of N_s corresponds to more regular and grid-like Superpixel structure and lower value of N_c captures more image details. The SLIC Superpixel based SOM clustering algorithm is given below:

1. Collect necessary information of superpixels by generate the superpixels representation of original image.

2. Initialize cluster centroids $v_i, i=1, \dots, C$.

3. The objective function F is given by

$$F = \sum_{i=1}^C \sum_{j=1}^Q \gamma_j u_{ij}^m \| \xi_j - v_i \|^2 + \frac{\alpha}{N_R} \sum_{i=1}^C \sum_{j=1}^Q u_{ij}^m \left(\sum_{s_r \in N_j} \gamma_r \| \xi_r - v_i \|^2 \right)$$

$$+ \sum_{j=1}^Q \lambda_j \left(1 - \sum_{i=1}^C u_{ij} \right) \quad (10)$$

4. The membership values u_{ij} is updated given by

$$u_{ij} = \left(\sum_{k=1}^C \left(\frac{\gamma_j \|\xi_j - v_i\|^2 + \frac{\alpha}{N_R} \sum_{s_r \in N_j} \gamma_r \|\xi_r - v_i\|^2}{\gamma_j \|\xi_j - v_k\|^2 + \frac{\alpha}{N_R} \sum_{s_r \in N_j} \gamma_r \|\xi_r - v_k\|^2} \right)^{1/(m-1)} \right)^{-1} \quad (11)$$

Where γ_j is the number of pixels in superpixel s_j ,

u_{ij} denotes the membership of superpixel s_j to the i th cluster.

Q is the number of superpixels in images and

ξ_j is the average color value of superpixel s_j ,

N_j stands for the set of neighboring superpixels that are adjacent to s_j and

N_R is the cardinality of N_j .

$\|\cdot\|$ is a norm metric, denoting Euclidean distance between pixels and clustering centroids.

The parameter m is a weighting exponent on each SOM membership and determines the amount of self mapping of the resulting classification.

5. The cluster centroids v_i is updated given by

$$v_i = \left(\sum_{j=1}^Q u_{ij}^m \left(\gamma_j \xi_j + \frac{\alpha}{N_R} \sum_{s_r \in N_j} \gamma_r \xi_r \right) \right) \left(\sum_{j=1}^Q u_{ij}^m \left(\gamma_j + \frac{\alpha}{N_R} \sum_{s_r \in N_j} \gamma_r \right) \right)^{-1} \quad (12)$$

6. Repeats Steps 3 to 4, until $\|V_{\text{new}} - V_{\text{old}}\| < \varepsilon$.

7. EXPERIMENTAL RESULTS

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) [17] is significant metric to validate the quality of image. It measures the square error between pixels of the original and the resultant images.

Qualitative Analysis: The proposed clustering algorithm is performed on two microarray images drawn from the standard microarray database corresponds to breast category a CGH tumor tissue [18]. Image 1 consists of a total of 38808 pixels and Image 2 consists of 64880 pixels. Gridding is performed on the input images by the method proposed in [13], to segment the image into compartments, where each compartment is having only one spot region and background. The gridding output is shown in figure 3. After gridding the image into compartments, such that each compartment is having single spot and background, compartment no 1 from image 1 and compartment no 12 from image 2 are extracted. Superpixels are generated for these two compartments using SLIC and segmented using SLIC based SOM algorithm. The Superpixel generation and segmentation is shown in figure 3.

The MSE is [18] mathematically defined as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \|v_i - c_j\|^2 \quad (13)$$

Where N is the total number of pixels in an image and x_i is the pixel which belongs to the j th 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 clustering algorithms. The results confirm that SLIC based SOM algorithm produces the lowest MSE value for segmenting the microarray image.

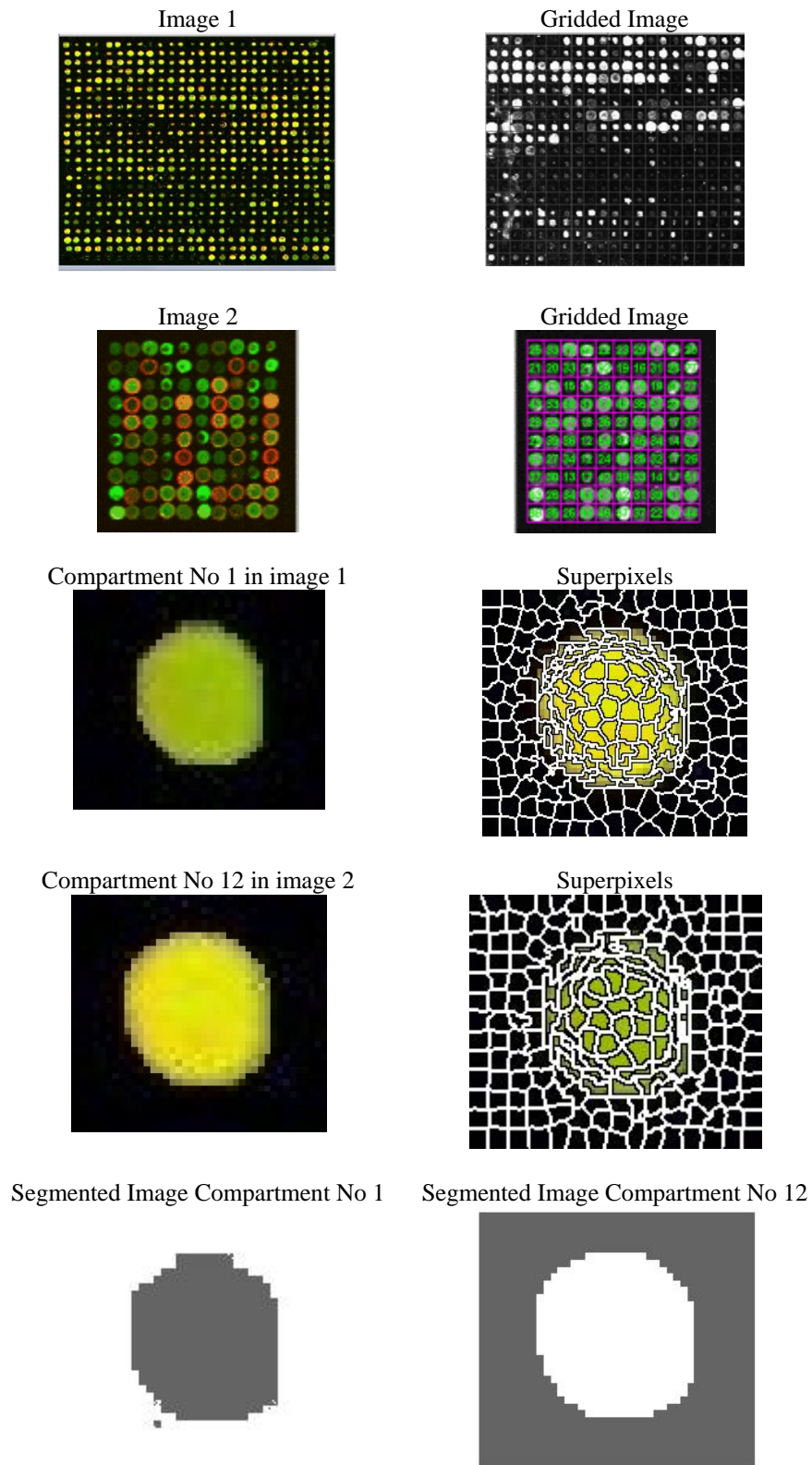


Figure 3: Super pixel based SOM segmentation

Table 1: MSE Values

Method	Compartment	
	No 1	No 12
K-means	96.4	93.6
Fuzzy c-means	93.1	89.4
Self organizing maps	84.7	80.9
SLIC SOM	82.8	77.4

8. 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 segmentation, information extraction and gridding. The transcription abundance between two genes under experiment is the expression ratio of each and every gene spot. Clustering algorithms have been used for microarray image segmentation with an advantage that they are not restricted to a particular size and shape for the spots. This paper describes SLIC based self organizing maps clustering algorithm for segmentation of microarray image. 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. The proposed method performs better noise suppression and produces better segmentation results.

REFERENCES

- [1] J.Harikiran, A.Raghu, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar, "Edge Detection using Mathematical Morphology for Gridding of Microarray Image", *International Journal of Advanced Research in Computer Science*, Volume 3, No 2, pp.172-176, April 2012.
- [2] J.Harikiran, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar, "Fast Clustering Algorithms for Segmentation of Microarray Images", *International Journal of Scientific & Engineering Research*, Volume 5, Issue 10, pp 569-574, 2014.
- [3] Bogdan Smolka, et.al. "Ultrafast Technique of Impulsive Noise Removal with Application to Microarray Image De-noising", *ICIAR 2005*, LNCS 3656, pp. 990–997, 2005. Springer-Verlag Berlin Heidelberg 2005.
- [4] Hara Stefanou, et.al. "Microarray Image Denoising Using a Two-Stage Multiresolution Technique", *2007 IEEE International Conference on Bioinformatics and Biomedicine*.
- [5] Lakshmana Phaneendra Maguluri, et.al, "BEMD with Clustering Algorithm for Segmentation of Microarray Image", *International Journal of Electronics Communication and Computer Engineering* Volume 4, Issue 2, ISSN: 2278-4209 2013.
- [6] J.Harikiran, D.Ramakrishna, B.Avinash, Dr.P.V.Lakshmi, Dr.R.Kiran Kumar, "A New Method of Gridding for Spot Detection in Microarray Images", *Computer Engineering and Intelligent Systems*, Vol 5, No 3, pp.25-33, 2014
- [7] M.Eisen, ScanAnalyze User's manual, 1999,
- [8] J.Buhler, T.Ideker and D.Haynor, "Dapple:Improved Techniques for Finding spots on DMA Microarray Images", Tech. Rep. UWTR 2000-08-05, University of Washington, 2000.
- [9] Frank Y. Shih and Shouxian Cheng, "Automatic seeded region growing for color image segmentation", 0262-8856/2005.
- [10] Aliaa Saad El-Gawady, et.al. "Segmentation of Complementary DNA Microarray Images using Marker-Controlled Watershed Technique", *International Journal of Computer Applications (0975 – 8887)* Volume 110 – No. 12, January 2015.
- [11] J.Harikiran, et.al. "Multiple Feature Fuzzy C-means Clustering Algorithm for Segmentation of Microarray image", *IAES International Journal of Electrical and Computer Engineering*, Vol. 5, No. 5, pp. 1045-1053, 2015.
- [12] J.Harikiran et.al. "Fuzzy C-means with Bi-dimensional empirical Mode decomposition for segmentation of Microarray Image", *International Journal of Computer Science Issues*, volume 9, Issue 5, Number 3, pp.273-279, 2012.
- [13] J.Harikiran, B.Avinash, Dr.P.V.Lakshmi, Dr.R.Kirankumar, "Automatic Gridding Method for microarray images", *Journal of Theoretical and Applied Information Technology*, volume 65, Number 1, pp.235-241, 2014.
- [14] Radhakrishna Achanta et.al."SLIC Superpixels Compared to state-of-the-art Superpixel Methods", *Journal of LaTeX Class Files*, Vol 6, No. 1, December 2011.
- [15] Hamidreza Saberhari, et.al. " Fully Automated Complementary DNA Microarray Segmentation using a Novel Fuzzy-based Algorithm", *IEEE Transactions on Information Technology in Biomedicine*, Vol 5, Issue 3, Jul-Sep 2015.
- [16] Luis Rueda and Vidya Vidyadharan, "A Hill-climbing Approach for Automatic Gridding of cDNA Microarray Images", *IAES International Journal of Electrical and Computer Engineering*, volume 4, No 6, December 2014, pp.923-930.
- [17] B.Saichandana et.al. "Hyperspectral Image Classification using Genetic Algorithm after Visualization using Image Fusion", *International Journal of Computer Science And Technology*, Volume 7, No. 2, June 2016, pp.2229-4333.
- [18] Durga Prasad Kondisetty, Dr. Mohammed Ali Hussain. "A Review on Microarray Image Segmentation Methods", *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 12, December 2016.

-
- [19] Komang Ariana, et.al., “Color Image Segmentation using Kohonen SOM”, *International Journal of Engineering and Technology (IJET)*, Vol. 6, No. 2, December 2014.
 - [20] Kristof Van Laerhoven. et.al, “Combining the Self-Organizing Map and K-Means Clustering for On-line Classification of Sensor Data”, *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 12, December 2001.
 - [21] M.N.M. Sap, et.al., “Hybrid Self Organizing Map for Overlapping Clusters”, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 14, No. 12, December 2016.
 - [22] Janne Nikkil, et.al., “Analysis and visualization of gene expression data using SOM”, *Neural Networks 15 (2002)* 953–966 Vol. 14, No. 12, December 2002.