

## Gridding and Segmentation Method for DNA Microarray Images

B. Sivalakshmi<sup>1\*</sup>, N.Naga Malleswara Rao<sup>2</sup>

<sup>1</sup> Research Scholar, Department of CSE, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

<sup>2</sup> Professor, Department of IT, RVR and JC College of Engineering, Guntur, India

[bslakshmi85@gmail.com](mailto:bslakshmi85@gmail.com)<sup>1</sup>

**Article History:** Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

**Abstract:** This article mainly explores meshing and segmentation techniques for microarray image analysis. The term "grid" refers to dividing an image into subgrids of dots and then dividing them into point detection. Most of the existing methods depend on input parameters such as the number of rows / columns, the number of points in each row / column, the size of the subarrays, etc. This article proposes a fully automatic mesh generation algorithm. This can remove any initialized parameter without any manual intervention. In the segmentation step, clustering algorithms are used because they do not consider the size and shape of the spots, do not depend on the initial state of the pixels, and do not require post-processing. In this article, a new method is proposed to estimate the initial parameters (centroid and number of clusters) required by any clustering algorithm. Qualitative and quantitative analysis shows that the algorithm can perform grid processing on microarray images well, and improves the performance of the clustering algorithm.

**Keywords:** Microarray image, Image segmentation, Mathematical morphology, clustering algorithms.

### 1. INTRODUCTION

Microarray technology is used to analyze gene expression values of thousands of points in parallel [1]. The output of the microarray experiment is the image extracted from the hybrid microarray slide using a sensor with two different wavelengths, Cy3 and Cy5. The analysis of the microarray image is carried out in three stages, namely gridding, segmentation and quantification. The gridding process is divided into two stages.

First, divide the image into sub-arrays, called sub-grids, and then divide these sub-arrays into gene point regions, called point detection [1]. Segmentation is the process of grouping the spot pixels and background pixels in each spot area. Quantification is the process of finding the value of gene expression, which is equal to the logarithmic ratio of the green and red intensity values of a point [2].

The gene expression value depends on the intensity values of foreground pixels (spot areas) and background pixels. Most of the existing algorithms for microarray image analysis are semi-automatic, which means that manual intervention is required to initialize the parameters to execute the gridding algorithm [3]. This article offers a fully automatic mesh generation algorithm. Microarray technology is widely used in genetics, disease diagnosis, drug development, and pharmacology [4]. The microarray image analysis mechanism is shown in Figure 1.

In this article, in the second part, a grid generation algorithm is presented, in the third part, an algorithm for estimating the necessary parameters of the clustering algorithm is presented, in the fourth part, a clustering algorithm is presented, and in the fifth. part, experimental results and conclusions are presented.

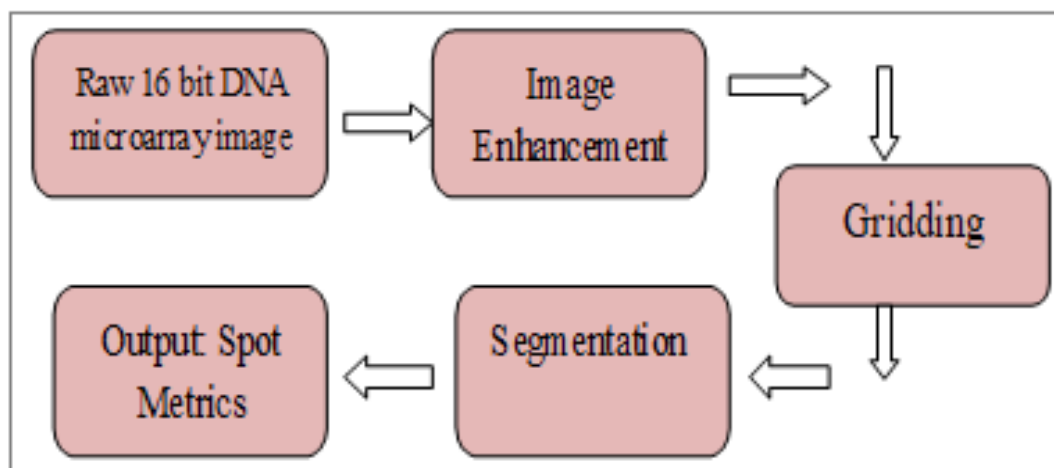


Figure 1: Architecture for Microarray Image Analysis.

## 2. MICROARRAY GRIDDING:

The grid is the most important step in microarray image analysis. For sequential analysis of microarray images, a fine mesh can improve the efficiency of the segmentation and quantification steps. To capture each spot of a gene from a microarray image, the image is first divided into sub arrays, and then these sub arrays are divided into spot regions. The first step is called global meshing (sub-meshing), and the second step is local meshing (point detection). The end result of the mesh is an area with one spot and a background. The existing meshing algorithms are divided into manual and semi-automatic. When building a mesh manually, the user must specify all the parameters necessary for building the mesh, such as the number of sub arrays, the number of points and the size of the points. In a semi-automatic program, the user part provides parameters such as the number of rows/columns and the number of points in each row/column. This paper proposes an algorithm to automatically divide the grid using the horizontal and vertical contours of the microchip image. Figures 2 and 3 show the algorithm used to detect sub-grids and points.

## 3. MICROARRAY SEGMENTATION

Image segmentation is the process of dividing an image into regions, and for microarray images, it is the process of dividing a subnet into foreground and background regions. This foreground area corresponds to the spot area. Using this spot of the segmented image, gene expression levels are estimated. This segmentation is a difficult task, since the intensities in this area are heterogeneous, and the sizes and shapes of the points also differ from each other. Several statistical segmentation methods are proposed for segmentation of microarray images. Shape-based circular segmentation [6], in which a circular mask with a specific radius is used to identify a point. But all spots are not of equal shape. Region growing method [5] in this the spot area and background area are separated by selecting seed pixels in both areas and regions are extracted by using some predefined criteria. Selection of seed pixel is a difficult task. Thresholding method [6], by using the histogram of the image, a suitable threshold is estimated, and the image is divided into two regions. This estimation is done using Mann Whitney test. Morphology based segmentation [7], uses the morphological operations such as hit or miss transforms to segment the image. Here the process depends on the selection of mask used for morphological operations. Supervised learning-based segmentation [8] uses support vector machine algorithm to segment the image. In this paper, clustering algorithms are used to segment the image. These algorithms are efficient than other existing segmentation methods, as they do not depend on the shape and size of the spot. This algorithm does not require any mask and seed pixels. Every clustering algorithm depends on the number of clusters and initial values of centroids. If these values are estimated, the number of iterations required for any algorithm to segment an image can be minimized. The required number of centroids and clusters for microarray image segmentation are two, as every algorithm divides into two regions background and spot area. Existing algorithm used minimum value in the background area and maximum value in the foreground area for segmenting the image into two regions. This article presents an algorithm for estimating centroids and the number of clusters using empirical mode decomposition. A block diagram for estimating the number of clusters is shown in Figure 4. This article uses clustering algorithms such as k-mean, moving k-mean, and FCM.

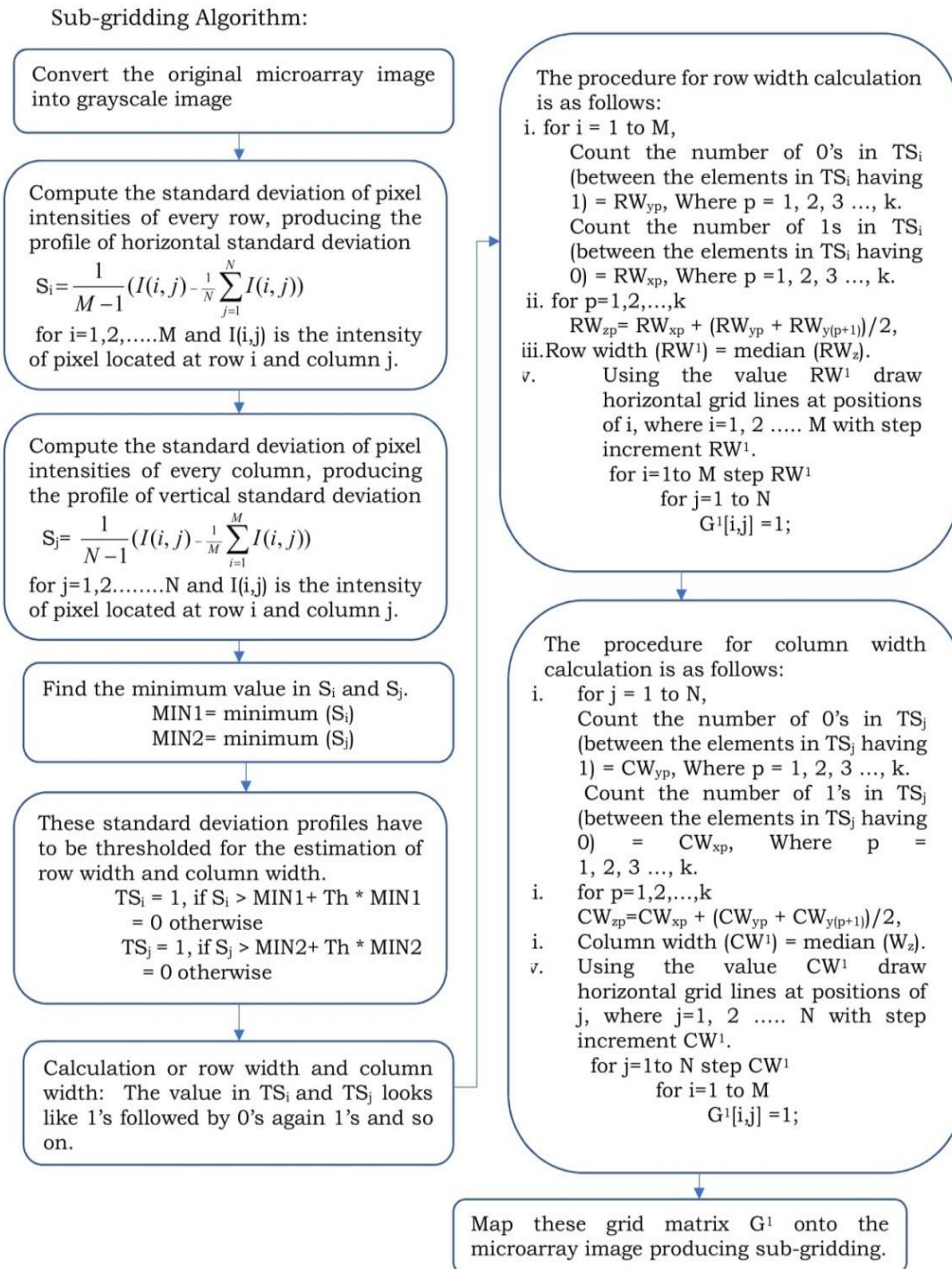


Figure 2: Sub-gridding Algorithm

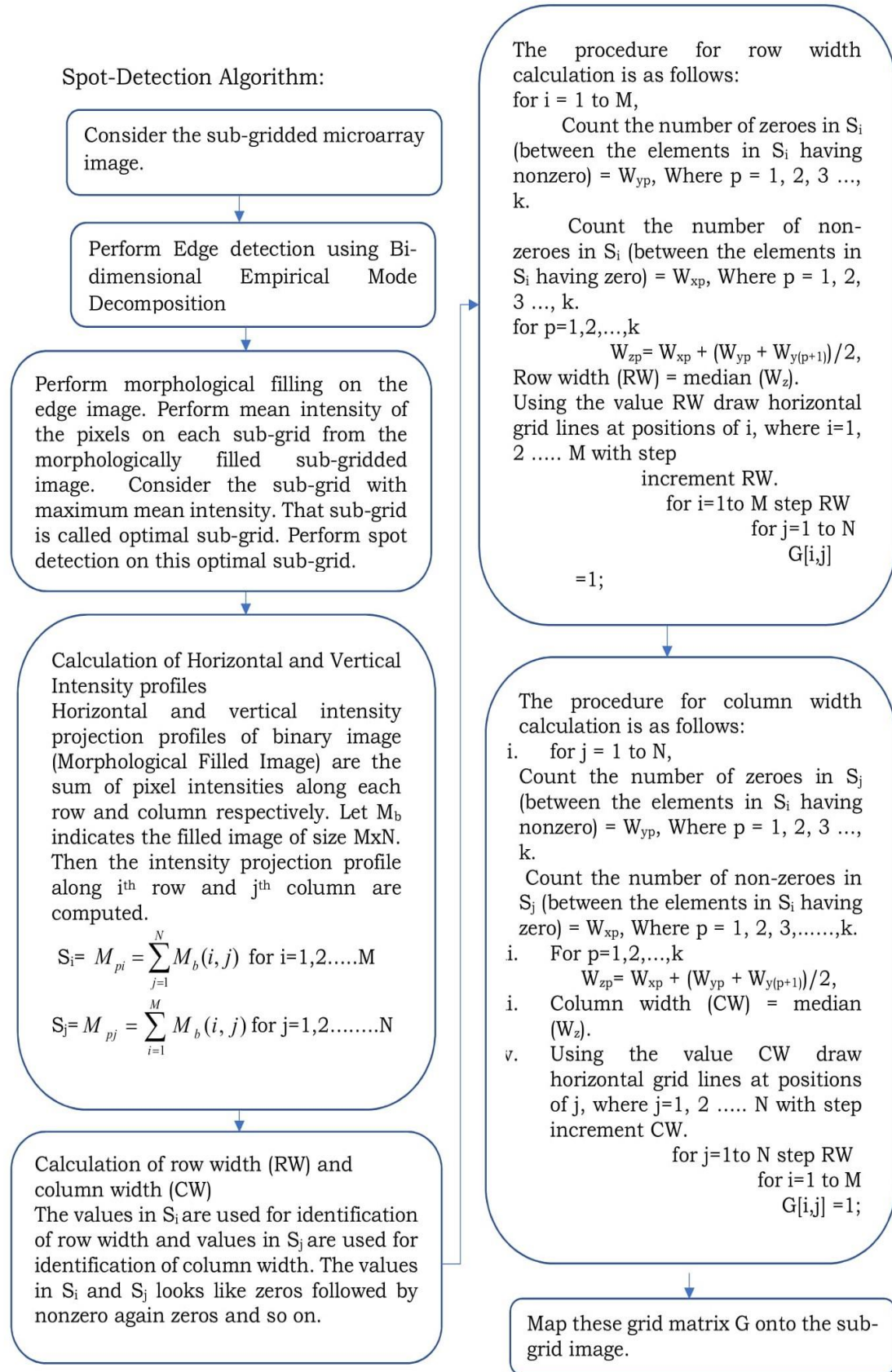


Figure 3: Spot Detection Algorithm

Algorithm for Estimation of Centroids and Number of Clusters using EMD:

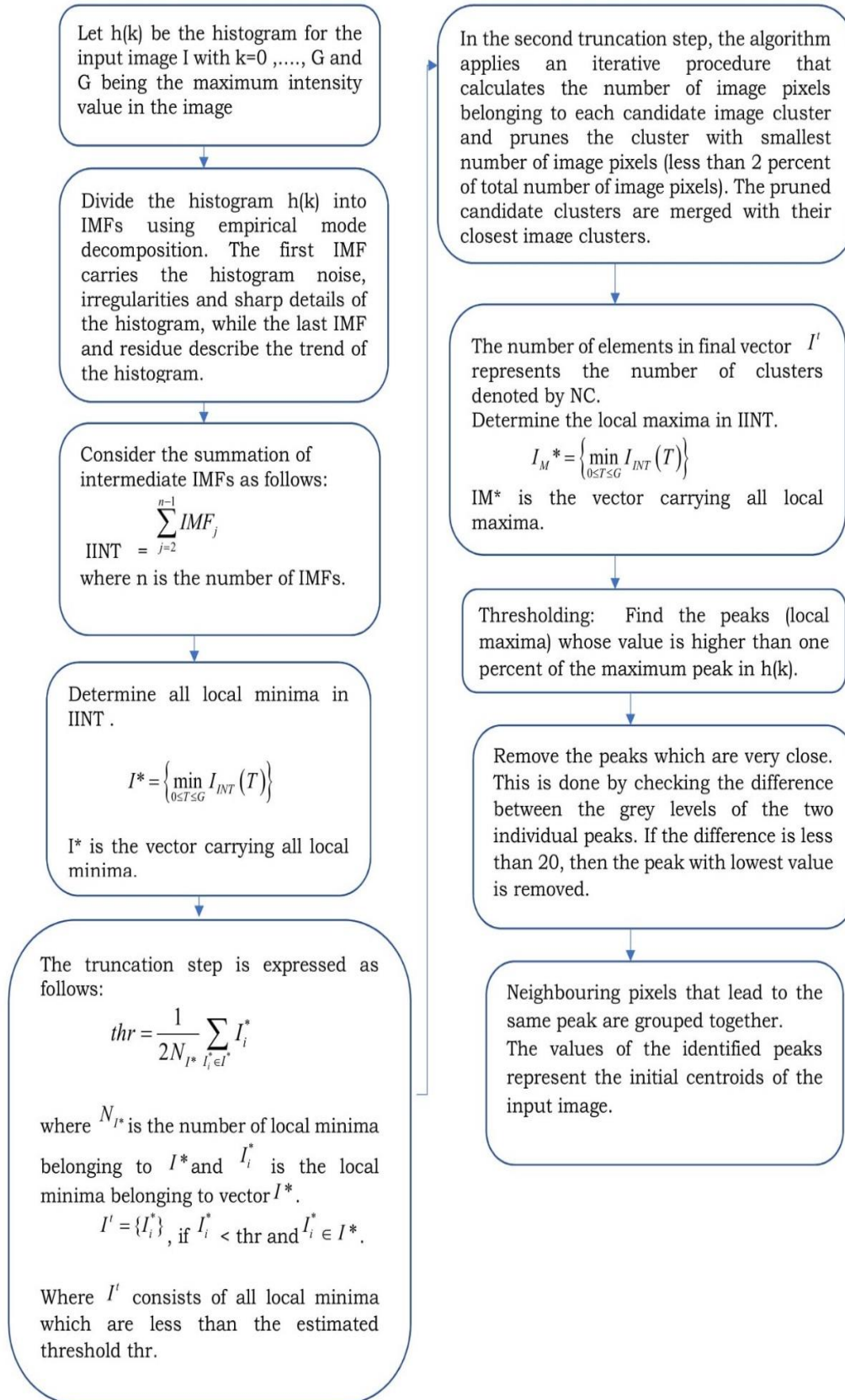


Figure 4: Algorithm for estimating clustering parameters

#### 4. CLUSTERING ALGORITHMS

In this article, k-mean [9], moving k-mean [10], and FCM algorithm [11] are used to segment microarray images. The two parameters required by these algorithms are estimated using the algorithm described in Section 3.

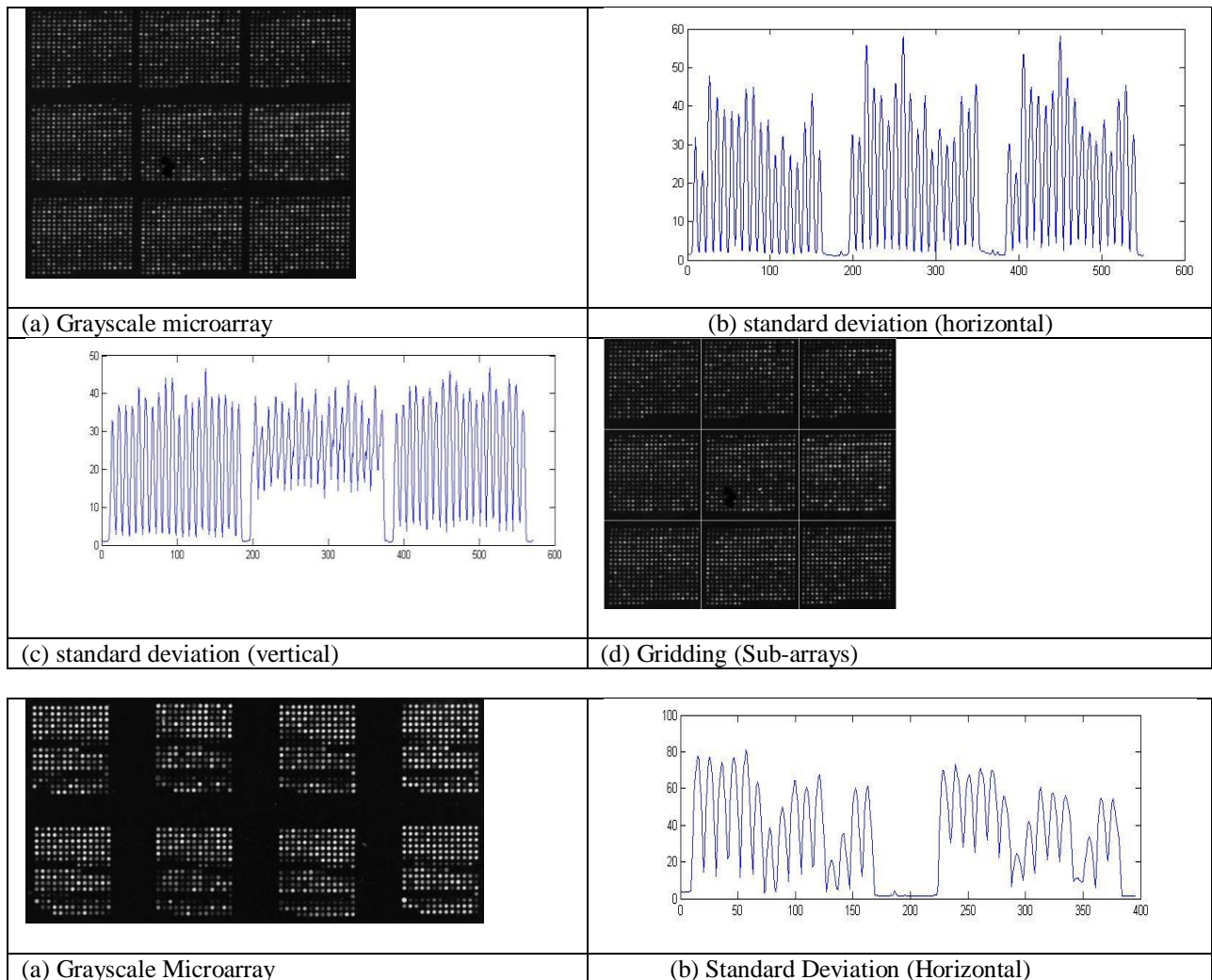
#### 5. EXPERIMENTAL RESULTS

The qualitative and quantitative analysis of proposed sub-gridding and spot detection algorithms are performed on two microarray images of breast cancer a CHG tumor tissue. The qualitative analysis of proposed sub-algorithm is shown in figure 5. The qualitative analysis of proposed spot detection algorithm is shown in figure 6. Table 1 shows a quantitative analysis of the proposed mesh generation algorithms in comparison with existing methods.

The precision of the mesh generation algorithm is estimated by the formula

$$\text{Percentage accuracy} = \frac{\text{Number of spots perfectly gridded}}{\text{Total number of spots}} * 100$$

After sub-gridding and spot detection algorithms, individual spots are extracted. These individual spots are segmented using clustering algorithms with initial values are estimated using proposed estimation of number of clusters and centriods algorithm. A sample spot is extracted from the image and segmented using clustering algorithm. Three different clustering algorithms are used in the presented work. Table 2 shows the number of iterations used by different clustering algorithms in segmenting the spot image with and without estimation of initial parameters. Table 3 shows the MSE values [11] of different segmentation algorithms. The segmented step on a single spot is shown in figure 7.



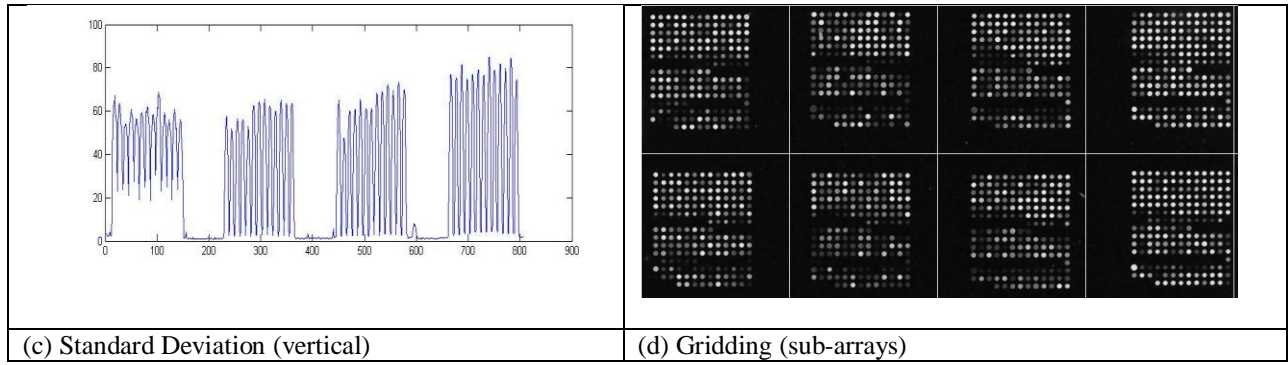
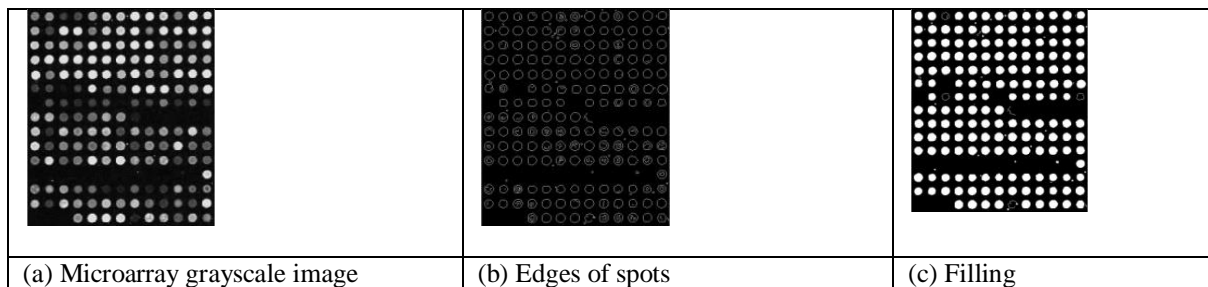
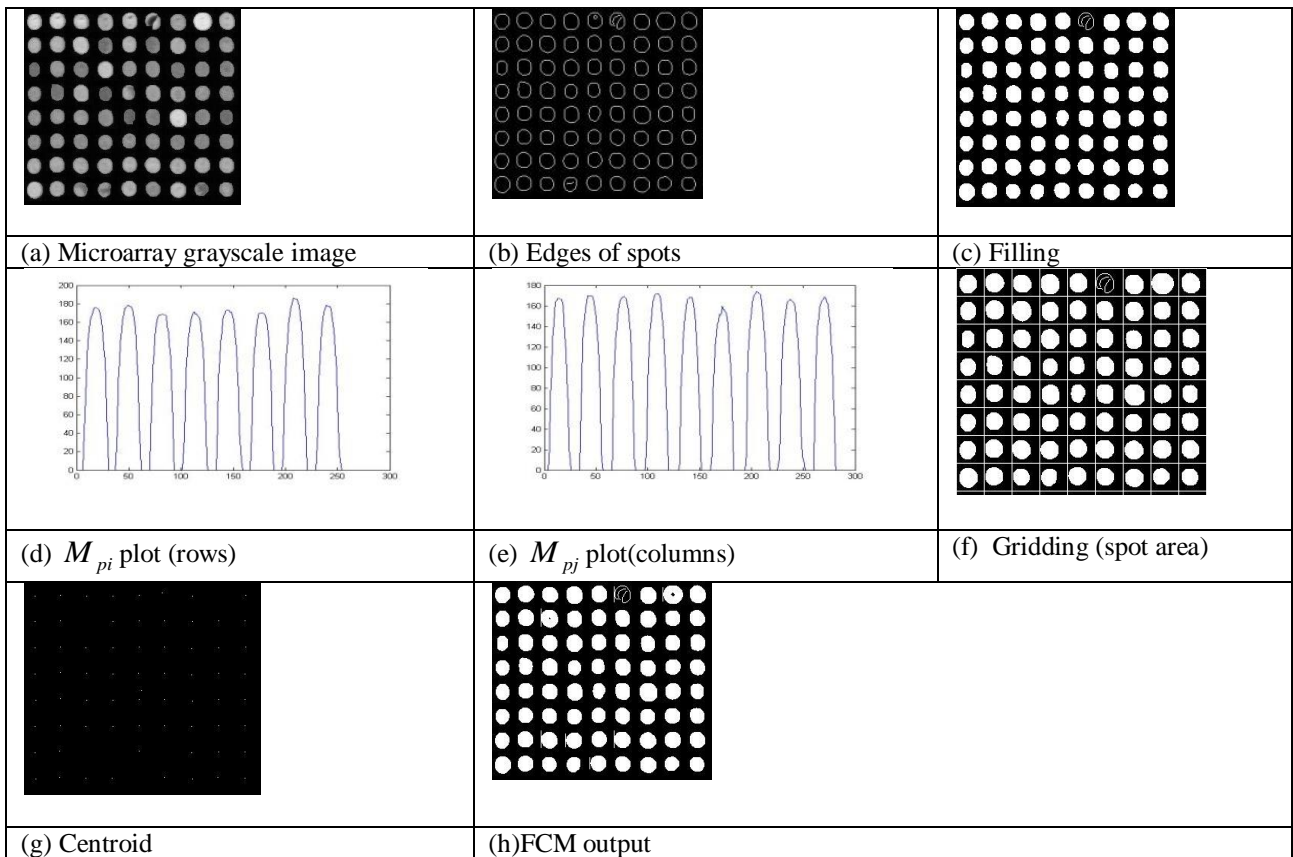


Figure 5: Sub-gridding



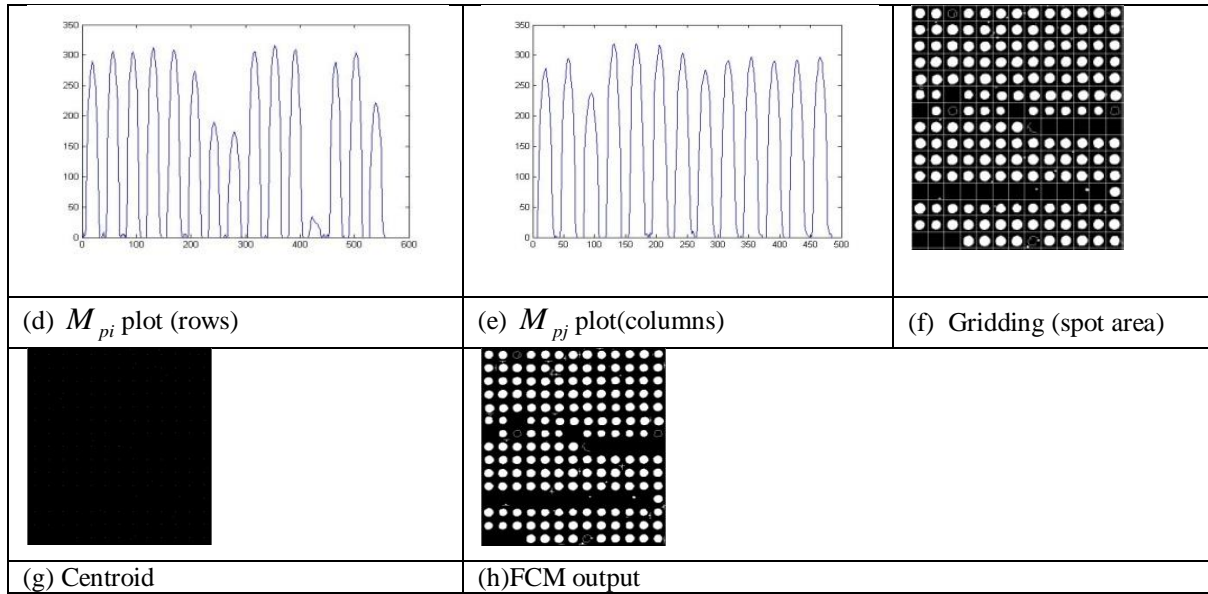


Figure 6: Spot-Detection

Table 1: Percentage accuracy of gridding

Method	Percentage Accuracy Image 1	Percentage Accuracy Image 2
Hirata [Microarray Gridding by mathematical Morphology]	87	79
Jain [Fully Automatic Quantification of microarray image data]	89	81
Wang [Precise Gridding of microarray images by detecting and correcting rotations in subarrays]	91	82
Shuqing Zhao [Microarray Image Processing Based on Mathematical Morphology]	90	84
Deepa .J [A New Gridding Technique for High Density Microarray Images Using Intensity Projection Profile of Best Sub Image]	92	88
Proposed	96	91

Table 3: MSE values

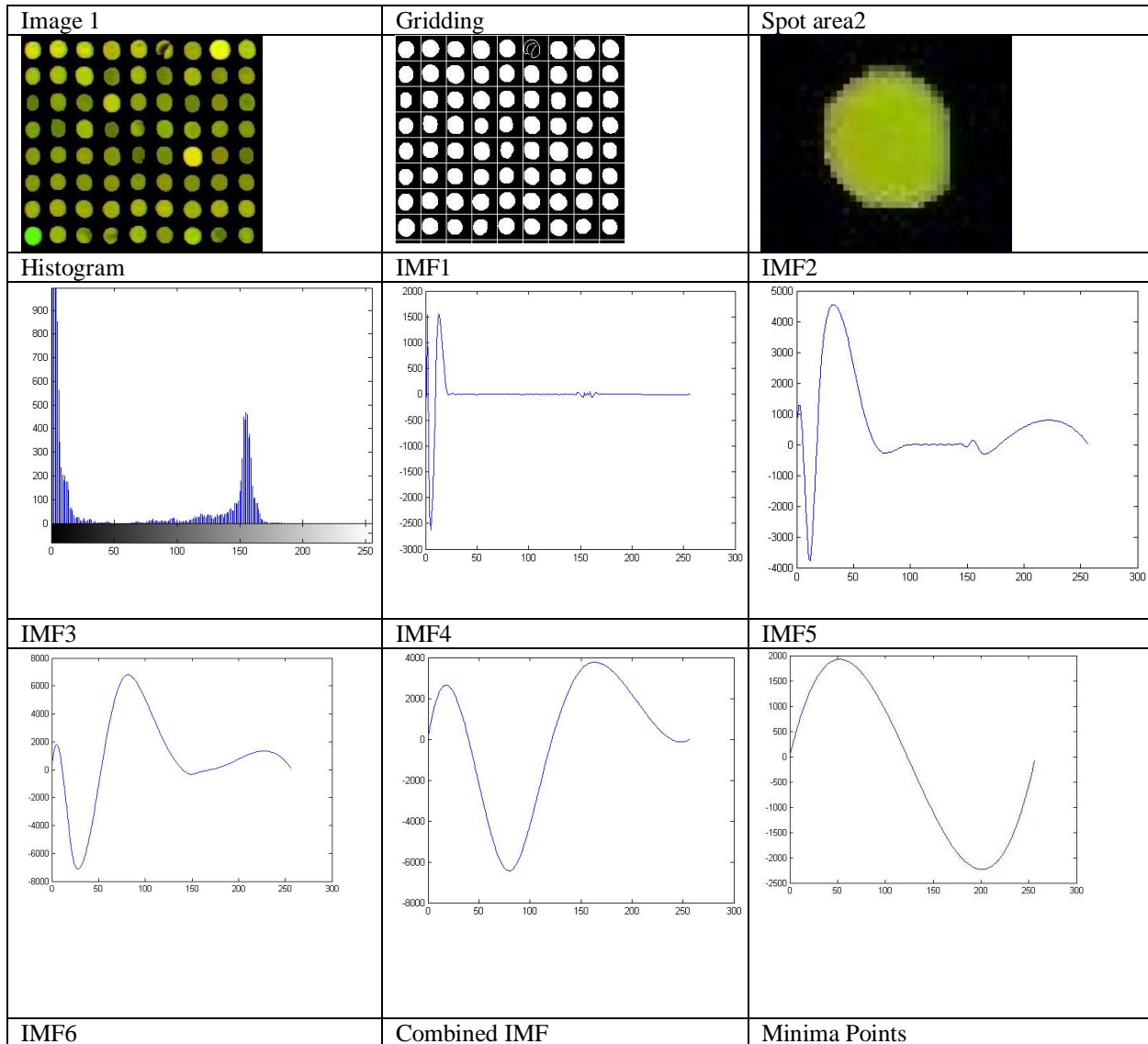
Method	MSE Values (Compartment No 1) In image 1	MSE Values (Compartment No 8) In image 2
K-means	282.781	346.47
Moving K-means	226.92	238.79
Fuzzy C-means	208.327	196.276

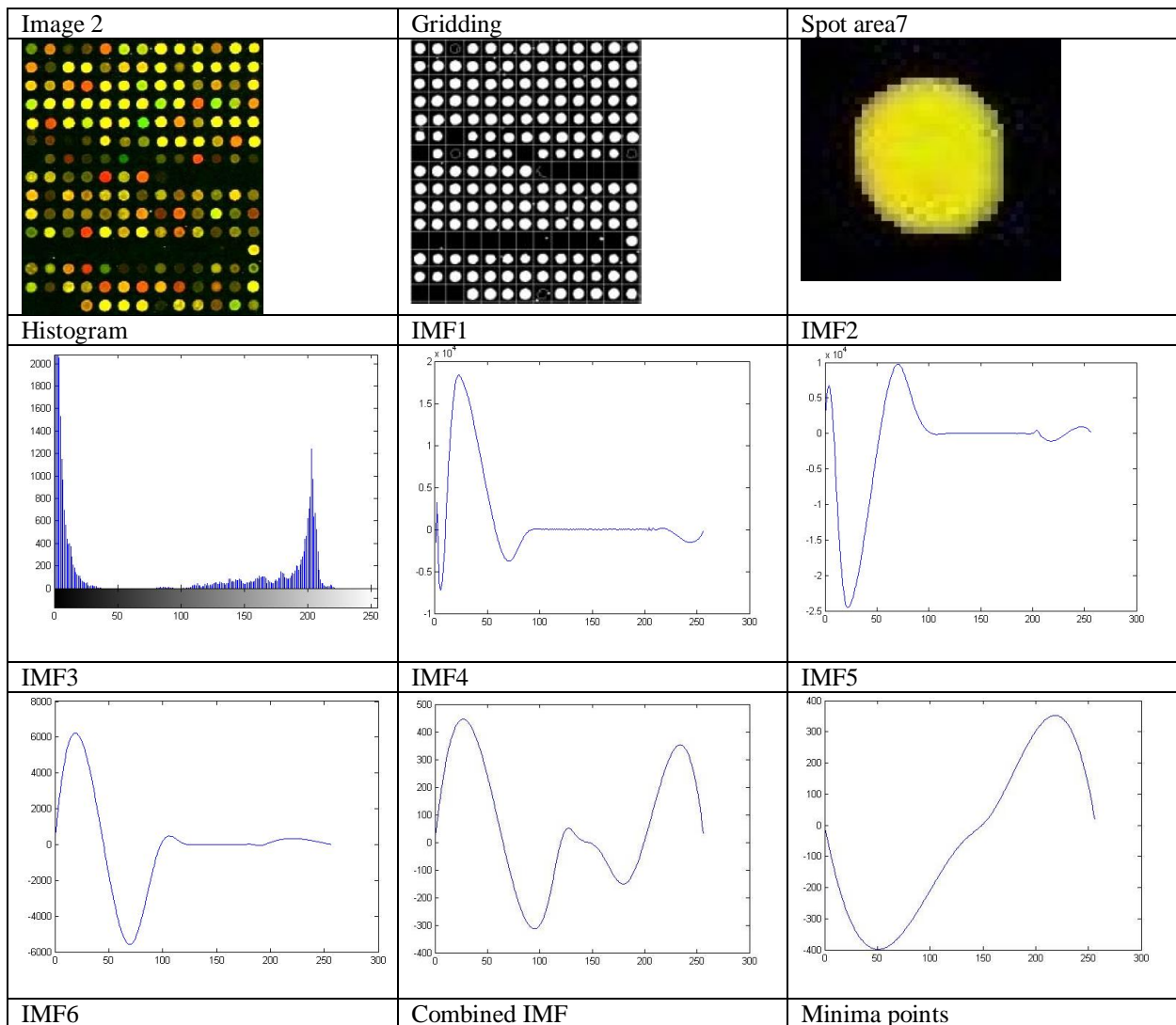
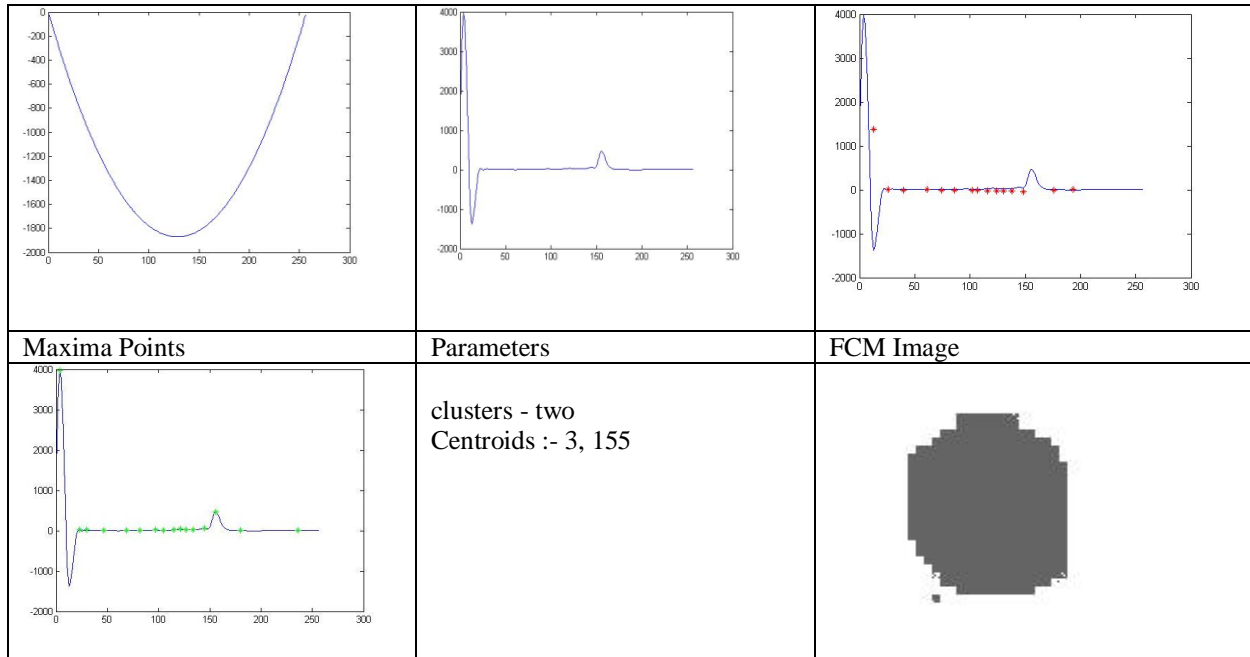


Table 2: Comparison of iterative step numbers

	Clustering algorithm	Iterative steps (without ECNC)	Iterative steps (with ECNC)
(Compartment No 1) In image 1	K-means	10	4
	Moving k-mens	14	6
	Fuzzy C-means	17	9

	Clustering algorithm	Iterative steps (without ECNC)	Iterative steps (with ECNC)
(Compartment No 8) In image 2	K-means	11	6
	Moving k-mens	16	12
	Fuzzy C-means	19	11





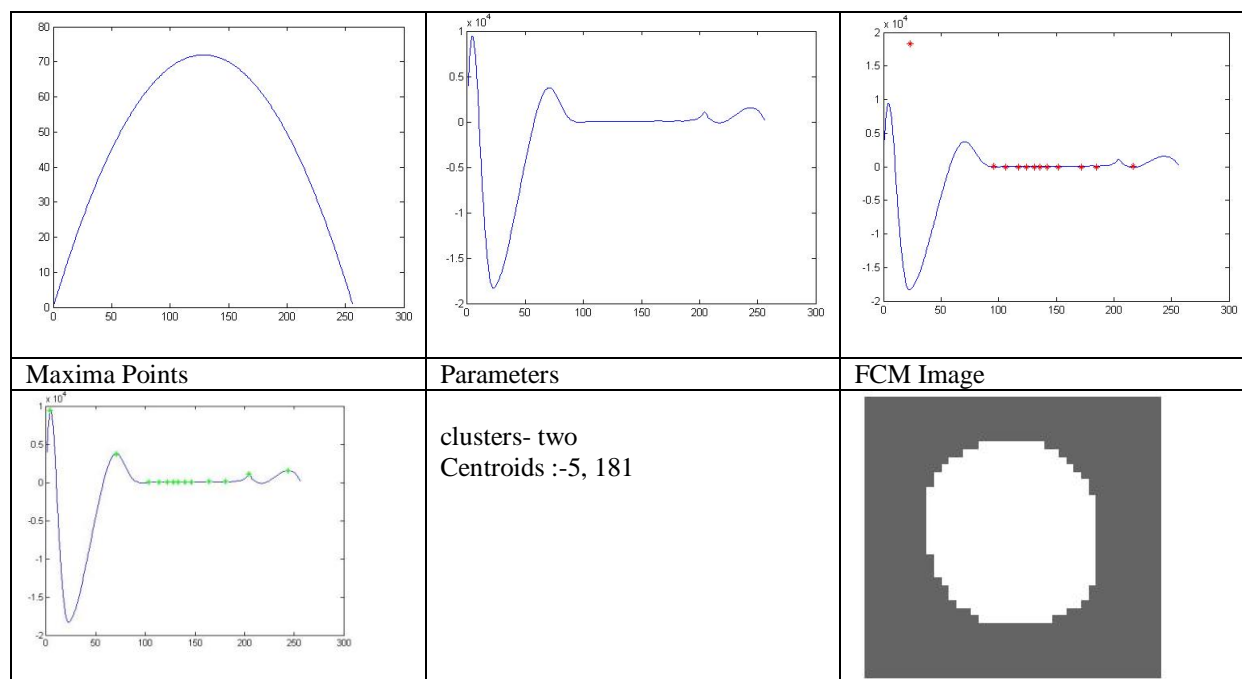


Figure 7: Segmented Result

## 6. CONCLUSIONS

The microarray image analysis is a sequential process with gridding, segmentation and quantification process. Any error in the gridding and segmentation stages, the gene expression value will be affected. This paper presents an automatic gridding algorithm and estimates the parameters required for clustering in segmentation process. The experimental results show the proposed algorithm grids the image with 96% accuracy and decreases the number of iterations by estimating the required parameters for segmenting the image by clustering algorithm.

## REFERENCES

1. J.Harikiran et. al., "Fast Clustering Algorithms for Segmentation of Microarray Images", IJSER, Volume 5, Issue 10, pp 569-574, October 2014.
2. J.Harikiran et. al., "Noise Removal in Microarray Images using Variational Mode Decomposition Technique", TELKOMNICA Telecommunication Computing Electronics and Control, Vol. 15, No. 4, pp. 1750 – 1757, December 2017.
3. J.Harikiran et. al., "Spot Edge detection in Microarray Images using Bi-dimensional Empirical Mode Decomposition", Proceedings of C3IT-2012, Vol 4, pp 19-25.
4. M.Katzer at.al.," Methods for automatic microarray image segmentation" IEEE transactions on NanoBioscience, vol 2, No 4, pp. 202-214, 2003.
5. L.Reuda et.al., "A pattern classification approach to DNA image segmentation", Pattern Recognition in Bioinformatics, lecture Notes in Computer Science, Volume 5780, pp, 319-330, 2009.
6. J.Rahnenfuhreret.al.,"Hybrid Clustering for microarray image analysis combining intensity and shape features", BMC bioinformatics, volume 5, No 47, pp.1-11, 2004.
7. N.Karimiet.al.,"Segmentation of DNA microarray images using an adptive graph based method", IET Image Processing Vol 4, No1 pp.19-27, 2010.
8. E.Zacharia et.al.,"3D spot modelling for automatic segmentation of cDNA microarray images", IEEE Transactions on NanoBioscience, Volume 9, pp. 181-192, November 2010.
9. B.Sivalakshmi et.al.,"Microarray Image Analysis using k-means Clustering algorithm", IJRAT, Vol 6, No 12, December 2018.
10. B.Sivalakshmi et.al, "Microarray Image Analysis using Adaptive Data Clustering Algorithms", IJAER, Vol 14, No 7, pp.1638-1643, 2019.
11. B.Sivalakshmi,"Microarray Image Analysis using Multiple Feature Data Clustering Algorithms", Springer SIST, Vol 105, 2019.