A New Method of Gridding for Spot Detection in Microarray Images

J.Harikiran[1], D.RamaKrishna[2], B.Avinash[3], Dr.P.V.Lakshmi[4], Dr.R.KiranKumar[5]
1. Member IEEE and Assistant professor, Department of IT, GITAM University, Visakhapatnam
2. Assistant professor, Department of CSE, GIT, GITAM University
3. Project Fellow, UGC Major Research Project, GIT, GITAM University, Visakhapatnam
4. Professor and Head, Department of IT, GITAM University, Visakhapatnam.
5. Assistant professor, Department of Computer science, Krishna University, Machilipatnam.

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. The gridding process is usually divided into two main steps: sub-gridding and spot detection. In this paper, a fully automatic approach to detect the location of spots is proposed. Each spot is associated with a gene and contains the pixels that indicate the level of expression of that particular gene. After gridding, the image is segmented using fuzzy c-means clustering algorithm for separation of spots from the background pixels. The result of the experiment shows that the method presented in this paper is accurate and automatic without human intervention and parameter presetting.

Keywords: Microarray Image, Mathematical Morphology, 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 [3]. The processing of the microarray images [5] 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.

Many approaches have been proposed for spot detection in microarray images. Hirata [6] presented an automatic sub-array and spot gridding method using the horizontal and vertical profile signal of the image. User assistance was required in this method to fix image rotation and check if the segmentation is correct. This method is valid only if the sub-array sizes are equal. Jain [7] proposed a gridding algorithm based on axis projection of image intensity along the rows and columns of the microarray image. The algorithm requires large number of spots and is not robust to misalignment of different grids. Y.Wang [8] demonstrated a fully automatic gridding methodology using intensity projection profile of microarray image. The method is sensitive to contaminations and large number of missing spots. Shuqing Zhao [13] proposed microarray image processing using mathematical morphology. An improved gridding method based on mathematical morphology is proposed, which is characterized by filtering out the block noise and filtering projection plots. Several parameters about the sub-array and spots are required during the gridding and spotting procedure which can be preset in advance or acquired from database. Deepa.J [14] proposed automatic gridding of DNA microarray images using optimum subimage. The approach is based on the selection of optimum subimage and the parameters for gridding are calculated using the intensity projection profile of the sub-image.

In this paper, a fully automatic gridding algorithm for spot detection is presented. After gridding, fuzzy C-means clustering algorithm is used for segmentation of microarray image into spots and image background. The algorithm is automatic and accurate for misalignment of spots in microarray image. Furthermore, when we apply this algorithm on different microarray images, human intervention and parameter presetting is unnecessary. The paper is organized as follows: section II presents edge detection using adaptive multi-structure morphological algorithm, Section III presents Morphological filling, Section IV presents gridding algorithm, Section V presents the Fuzzy c-means clustering algorithm, Section VI presents the qualitative and quantitative results, and finally Section VI repots conclusions.

2. Edge Detection Using Adaptive Multistructure Morphological Algorithm

Because of the unicity and fixity of structural element (SE) in traditional Edge detection using mathematical morphology, there are two main deficiencies: on the one hand, a single SE can only detect the edge of the same direction with the SE, but is not sensitive to different directions; on the other hand large-scale SE has strong ability to restrain noise, but the detected edge image is rough; small-scale SE is good at checking the details of the edge, but weak at noise suppression. In order to effectively restrain noise and preserve image edge information, we use adaptive multi-structure morphological algorithm to get the edge images [12]. We calculate the gray scale distance of original image to adaptively define the weights of SEs. The eight structuring elements of different directions with the size of 5X5 are shown in figure 1.

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

SE with 0^0

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

SE with 22.5^0

$$\begin{matrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{matrix}$$

SE with 45^0

$$\begin{matrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{matrix}$$

SE with 67.5^0

$$\begin{matrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{matrix}$$

SE with 90^0

$$\begin{matrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{matrix}$$

SE with 112.5^0

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{matrix}$$

SE with 135^0

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

SE with 157.5^0

Figure 1. SEs with different directions.

$$\begin{matrix} a_9 & a_8 & a_7 & a_6 & a_5 \\ a_{10} & a_{23} & a_{24} & a_{25} & a_4 \\ a_{11} & a_{22} & a_1 & a_2 & a_3 \\ a_{12} & a_{21} & a_{20} & a_{19} & a_{18} \\ a_{13} & a_{14} & a_{15} & a_{16} & a_{17} \end{matrix}$$

Figure 2. Image Sub-block

Figure 2 shows the sub-block with the size 5X5, in which a_1 is the gray-scale value of the center pixel, and $a_2,a_3,a_4.....,a_{25}$ stand for its neighborhood gray scale value, then the gray-scale distance of a_1 and its neighborhood can be performed as:

$d_k = |a_1 - a_k|,\ k = 2,3,4,....,25.$ (1)

The larger the gray-scale distance, the higher extent of salutation, and the bigger possibility that the pixel is an edge point in the image. The Edge Gray-Scale distances of a_1 can be defined as follows.

$G_1(x,y) = d_4 + d_5 + d_6 + d_7 + d_8 + d_9 + d_{10} + d_{12} + d_{13} + d_{14} + d_{15} + d_{16} + d_{17} + d_{18} + d_{19} + d_{20} + d_{21} + d_{23} + d_{24} + d_{25};$ --for direction 0^0

$G_2(x,y) = d_4 + d_5 + d_6 + d_7 + d_8 + d_9 + d_{10} + d_{12} + d_{13} + d_{14} + d_{15} + d_{16} + d_{17} + d_{18} + d_{19} + d_{20} + d_{21} + d_{22} + d_{23} + d_{24} + d_{25};$ --for direction 22.5^0

$G_3(x,y) = d_2 + d_3 + d_4 + d_6 + d_7 + d_8 + d_9 + d_{10} + d_{11} + d_{12} + d_{14} + d_{15} + d_{16} + d_{17} + d_{18} + d_{19} + d_{20} + d_{22} + d_{23} + d_{24};$ ----for direction 45^0

$G_4(x,y) = d_2 + d_3 + d_4 + d_5 + d_7 + d_8 + d_9 + d_{10} + d_{11} + d_{12} + d_{13} + d_{14} + d_{15} + d_{16} + d_{17} + d_{18} + d_{19} + d_{21} + d_{22} + d_{23} + d_{24} + d_{25};$
---for direction 67.5^0

$G_5(x,y) = d_2 + d_3 + d_4 + d_5 + d_6 + d_8 + d_9 + d_{10} + d_{11} + d_{12} + d_{13} + d_{14} + d_{16} + d_{17} + d_{18} + d_{19} + d_{20} + d_{21} + d_{23} + d_{25};$ ----for direction 90^0

$G_6(x,y) = d_2 + d_3 + d_4 + d_5 + d_6 + d_7 + d_9 + d_{10} + d_{11} + d_{12} + d_{13} + d_{14} + d_{15} + d_{17} + d_{18} + d_{19} + d_{20} + d_{21} + d_{22} + d_{23} + d_{24} + d_{25};$
---for direction 112.5^0

$G_7(x,y) = d_2 + d_3 + d_4 + d_5 + d_6 + d_8 + d_9 + d_{10} + d_{11} + d_{12} + d_{13} + d_{14} + d_{16} + d_{17} + d_{18} + d_{19} + d_{20} + d_{21} + d_{23} + d_{25};$ ----for direction 135^0

$G_8(x,y) = d_2 + d_3 + d_4 + d_5 + d_6 + d_7 + d_8 + d_{10} + d_{11} + d_{12} + d_{13} + d_{14} + d_{15} + d_{16} + d_{18} + d_{19} + d_{20} + d_{21} + d_{24} + d_{25};$
---for direction 157.5^0

As for the whole image, the gray-scale distances of each edge and adaptive weights of SEs can be calculated as below:

$$ED_k = \sum_{x=2}^{M-1} \sum_{y=2}^{N-1} G_k(x,y),\ k=1,2,....,8 \qquad (1)$$

$$w_k = ED_k / \left(\sum_{k=1}^{8} ED_k \right),\ k=1,2,....,8 \qquad (2)$$

The edge E extracted adaptively by multi-structure morphology is given by

$$E = \sum_{k=1}^{8} w_k [(I \circ b_k) \oplus b_k - (I \bullet b_k) \Theta\ b_k)] \quad (3)$$

3. Morphological Filling

A hole may be defined as a background region surrounded by a connected border of foreground pixels. The filling holes in an image are based on set dilation, complementation and intersection [11]. The following procedure fill all the holes with 1's until $X_k = X_{k-1}$.

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k=1, 2, 3.... \qquad (4)$$

Where B is a symmetric structuring element and A is the binary image (output of Edge Detection). X_0 is an array containing 0's with the same size of A except at the locations corresponding to the point in each hole is 1. The union of X_k with the image A gives the filled region.

4. Automatic Spot Detection Algorithm

A microarray image contains a number of sub-grids and each sub-grid representing a two dimensional array of spots. Gridding refers to accurately locating each spot within a microarray image namely sub-gridding and spot detection. The sub-grid detection is done by the method in [10]. After sub-grid detection, the aim is to separate the sub-grid into spot regions by means of a grid. The output of gridding is to obtain a 2D matrix G of same size of sub-grid. Initially the values of G[i,j]= 0, for i=1,….,M and j=1,….,N.

The steps of the automatic spot detection algorithm are as follows:

Step 1: Convert the RGB Microarray image into grayscale image

Step 2: Perform Edge detection using method in section 1 on the grayscale microarray image.

Step 3: Perform morphological filling on the edge image obtained from step 2.

Step 4: Calculation of Horizontal and Vertical Intensity profiles

Horizontal and vertical intensity projection profiles of binary image (Morphological Filled Image) are the sum of pixel intensities along each row and column respectively. Let M_b indicates the filled image of size MxN . Then the intensity projection profile along i^{th} row and j^{th} column are computed using (3) and (4).

$$S_i = M_{pi} = \sum_{j=1}^{N} M_b(i,j) \qquad i=1\ldots..M \qquad (5)$$

$$S_j = M_{pj} = \sum_{i=1}^{M} M_b(i,j) \qquad j=1\ldots\ldots..N \qquad (6)$$

Step 5: Calculation of row width (RW) and column width (CW)

The values in S_i are used for identification of row width and values in S_j are used for identification of column width. The values in S_i and S_j looks like zeros followed by nonzero again zeros and so on.

The procedure for row width calculation is as follows:

i. For i = 1 to M,

Count the number of zeroes in S_i (between the elements in S_i having nonzero) = W_{yp}, Where p = 1, 2, 3 …, k.

Count the number of non-zeroes in S_i (between the elements in S_i having zero) = W_{xp}, Where p = 1, 2, 3 …, k.

ii. For p=1,2,…,k $W_{zp}= W_{xp} + (W_{yp} + W_{y(p+1)})/2$,

iii. Row width (RW) = median (W_z).

iv. Using the value RW draw horizontal grid lines at positions of i, where i=1, 2 ….. M with step increment RW.

```
for i=1to M step RW
      for j=1 to N
            G[i,j] =1;
```

The procedure for column width calculation is as follows:

i. For j = 1 to N,

Count the number of zeroes in S_j (between the elements in S_i having nonzero) = W_{yp}, Where p = 1, 2, 3 …, k.

Count the number of non-zeroes in S_j (between the elements in S_i having zero) = W_{xp}, Where p = 1, 2, 3 …, k.

ii. For p=1,2,…,k $W_{zp}= W_{xp} + (W_{yp} + W_{y(p+1)})/2$,

iii. Column width (CW) = median (W_z).

iv. Using the value CW draw horizontal grid lines at positions of j, where j=1, 2 ….. N with step increment CW.

```
for j=1to N step RW
      for i=1 to M
            G[i,j] =1;
```

Step 6: Map these grid matrix G onto the grayscale image.

Step 7: Compute all the connected components in the gridded optimum image using 8-adjacency. The number of connected components is equal to the number of spots. Assign a label for pixels in each connected component generating a label matrix.
Step 8: The pixels with one particular label represents a region of the single spot. Using region properties, compute the centroid and area for each region (spot). From the centroids of the spots, we can find the distance between the spots.

4. Segmentation Using Fuzzy C-Means Clustering Algorithm

The Fuzzy C-means [4][9] is an unsupervised clustering algorithm. The main idea of introducing fuzzy concept in the Fuzzy C-means algorithm is that an object can belong simultaneously to more than one class and does so by varying degrees called memberships. It distributes the membership values in a normalized fashion. It does not require prior knowledge about the data to be segmented. It can be used with any number of features and number of classes. The fuzzy K-means is an iterative method which tries to separate the set of data into a number of compact clusters. The segmented microarray image using fuzzy c-means is shown in figure 7.
The Fuzzy K-means algorithm is summarized as follows:
Algorithm Fuzzy K-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 and segmented Image
Begin
Step_1: Initialize the membership matrix u_{ij} is a value in (0,1) and the fuzziness parameter m (m=2). The sum of all membership values of a pixel belonging to clusters should satisfy the constraint expressed in the following.

$$\sum_{j=1}^{c} u_{ij} = 1 \qquad (7)$$

for all i= 1,2,…….N, where c (=2) 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:

$$F = \sum_{j=1}^{N} \sum_{i=1}^{c} u_{ij}^{m} \, \| x_j - c_i \|^2 \qquad (8)$$

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 u_{ij} 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

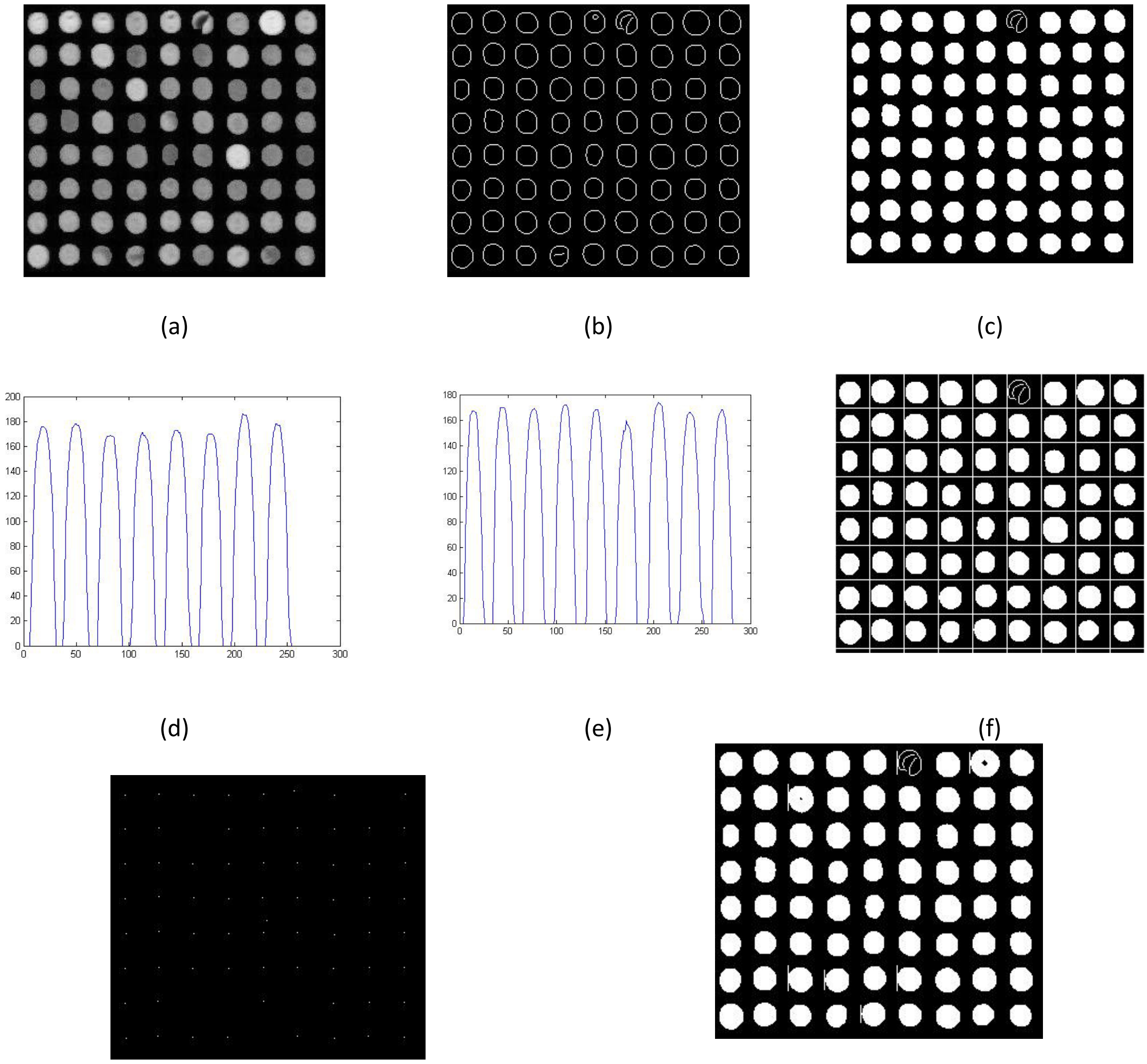
$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^{m} x_j}{\sum_{j=1}^{N} u_{ij}^{m}}.$$

(9)

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

5. Qualitative and Quantitative Results

The proposed spot detection algorithm is performed on a two different microarray slides drawn from the Stanford microarray Database corresponds to breast category aCGH tumor tissue. The first sub-grid slide is a 261*289 pixel image (Figure 3) that consists of a total of 75429 pixels. The second sub-grid slide is a 559*489 pixel image (Figure 4) that consists of total 273351 pixels. The output of the proposed automatic spot detection algorithm on two microarray sub-grids is shown in figure3 and figure 4.

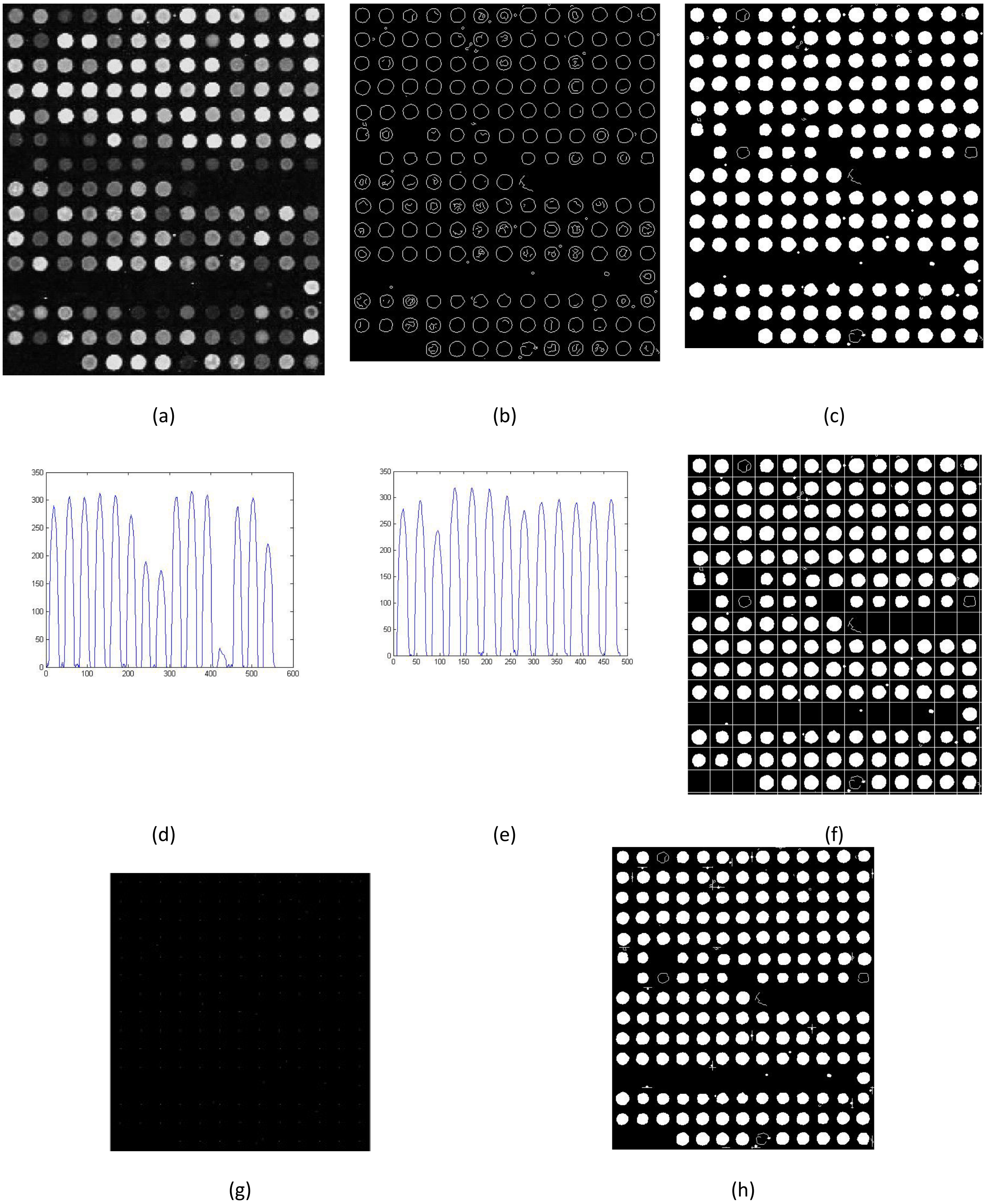


(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 3: a) Grey scale Image, b) Edge Detection, c) Morphological Filling d) Plot of M_{pi} (for all rows) (S_i) e) Plot of M_{pj} (for all columns)(S_j) f) gridded image g) centroids h) segmented image using fuzzy c-means



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 4:, a) Grey scale Image, b) Edge Detection, c) Morphological Filling d) Plot of M_{pi} (for all rows) (S_i) e) Plot of M_{pj} (for all columns)(S_j) f) gridded image g) centroids h) segmented image using fuzzy c-means

The accuracy of the gridding algorithm was calculated as

$$\text{Percentage accuracy} = \frac{\text{Number of spots perfectly gridded}}{\text{Total number of spots}} \text{ X100} \quad (10)$$

The performance of the proposed spot gridding algorithm was evaluated by comparing the results with the methods in [6], [7], [8] , [13] ,[14] are shown in table 1 .

Table 1: Percentage accuracy of gridding

Method	Percentage Accuracy figure 1	Percentage Accuracy figure 2
Hirata [6]	87	79
Jain [7]	89	81
Wang [8]	91	82
Shuqing Zhao[13]	90	84
Deepa .J [14]	92	88
Proposed	96	91

After gridding, the segmentation of spots from the background pixels is done by using fuzzy c-means clustering algorithm. The method is implemented in such a way that the intensity value of each pixel and the pixels of the image has been grouped in two clusters. The number of pixels clustered as spot and background for two different microarray images has been presented in Table 2.

Table 2: The number of pixels clustered as spots and background

Method	Total Number of Pixels	Spots	Background
Image 1	75429	40535	34714
Image 2	273351	111669	161682

6. Conclusion

In this paper, a fully automatic gridding method for separating spot centers in microarray sub-grids has been proposed. The proposed automatically locates the individual spots without any input parameters and human intervention. It can be proved that percentage accuracy of gridding is high with the methods that use projection profile of the entire image. After gridding, the image is segmented using fuzzy c-means clustering algorithm. The proposed method is accurate and automatic, which takes a microarray sub-grid as input image and makes no assumptions about the size of the spots, rows and columns in the grid.

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] Volkan Uslan, Ihsan Omur Bucak, “Clustering based Spot Segmentation of cDNA Microarray Images “, 32nd Annual International Conference of IEEE EMBS , IEEE 2010.
[5] J.Harikiran, 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, vol 3, No 2, pp- 172-175. March 2012.

[6] Roberto Hirata Jr, Junior Barrera, R.F. Hashimoto, O.D.Daniel," Microarray Gridding by mathematical Morphology", 2001 IEEE, PP. 112-119.
[7] A.N.Jain, Tokuyasu, Snijders," Fully Automatic Quantification of microarray image data", Genome research, 12(2), 32-332.
[8] Y.Wang, F.Y.Shih, M.Ma, "Precise Gridding of microarray images by detecting and correcting rotations in subarrays", in proceedings of sixth International conference on computer vision, Patternrecognition and Image Processing, 2010.
[9] J.Harikiran," Fuzzy C-means with Bi-dimensional Empirical Mode decomposition for segmentation of Microarray image", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012.
[10] L.Reuda, "Sub-grid Detection in DNA Microarray Images", Proceedings of the IEEE Pacific-RIM Symposium on image and video technology, pp 248-259, 2007.
[11] R.Gongalez, R.E.Woods," Digital Image Processing", 4th Edition, Printice Hall India, 2010.
[12] J.Harikiran, 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- 172-175. March 2012.
[13] Shuqing Zhao, Qiaoning Yang, "Microarray Image Processing Based on Mathematical Morphology", Proceedings of ICSP 2006, IEEE.
[14] Deepa J, Tessamma Thomas," A New Gridding Technique for High Density Microarray Images Using Intensity Projection Profile of Best Sub Image", Computer Engineering and Intelligent systems, Vol 4 No 1 : pp 7-18.