

An Analysis and Comparison of Quality Index Using Clustering Techniques for Spot Detection in Noisy Microarray Images

A. Sri Nagesh

Faculty, Computer Science & Engineering Department,
R.V.R. & J.C. College of Engineering,
Guntur -522019. India

asrinagesh@gmail.com

Dr. G. P. Saradhi Varma

Faculty, Information Technology Department,
S.R.K.R.Engineering. College,
Bhimavaram -534204. India

gpsvarma@yahoo.com

Dr.A.Govardhan

Faculty, Computer Science & Engineering Department,
JNTUCEH, Jagtial,
Hyderabad. India

govardhan_cse@yahoo.co.in

Dr. B. Raveendra Babu

Director, Operations,
Delta Technologies,
Hyderabad. India

rbhogapathi@yahoo.com

Abstract

In this paper, the proposed approach consists of mainly three important steps: preprocessing, gridding and segmentation of micro array images. Initially, the microarray image is preprocessed using filtering and morphological operators and it is given for gridding to fit a grid on the images using hill-climbing algorithm. Subsequently, the segmentation is carried out using the fuzzy c-means clustering. Initially the enhanced fuzzy c-means clustering algorithm (EFCMC) is implemented to effectively clustering the image whether the image may be affected by the noises or not. Then, the EFCM method was employed the real microarray images and noisy microarray images in order to investigate the efficiency of the segmentation. Finally, the segmentation efficiency of the proposed approach was compared with the various algorithms in terms of quality index and the obtained results ensures that the performance efficiency of the proposed algorithm was improved in term of quality index rather than other algorithms.

Keywords: Microarray Image, Genes, Spot Segmentation, Morphological Operator, Fuzzy K-Means, Fuzzy C-means, Enhanced fuzzy C-means Clustering (EFCM).

1. INTRODUCTION

In this research, we have proposed an efficient approach for microarray image segmentation to quantify the intensity of each spot and locate differentially articulated genes. The proposed approach contains three important steps such as, preprocessing, gridding and segmentation. The preprocessing stage contains the following process such as, top-hat filtering, binarization and morphological operations. Subsequently, the preprocessed image is given to the gridding process to accurately fit a grid into a spot. Here, we make use of hill climbing algorithm to effectively spot the grids on microarray using objective functions. Then, the image is given to the proposed clustering algorithm for segmentation. The designed clustering algorithm improves the microarray image segmentation by taking the advantage of spatial information along with the gray level pixel values. Furthermore, we have introduced the neighborhood fuzzy factor in the proposed clustering algorithm to effectively handling the spatial and intensity values in order to find the

appropriate cluster. The neighborhood fuzzy factor can be able to accurately detect the absent spots as well as the noisy spots.

The organization of the paper is as follows: Section 2 presents a brief review of some recent significant researches in Microarray image segmentation. The properties of proposed methodology for microarray image segmentation utilizing the enhanced fuzzy c-means clustering algorithm are explained in section 3. Experimental results and analysis of the proposed methodology are discussed in Section 4. Finally, concluding remarks are provided in Section 5.

2. REVIEW OF RELATED WORKS

Numerous researches based on gridding and clustering techniques have been proposed by researchers for the segmentation of microarray images. A brief review of some important contributions from the existing literature is presented in this section.

Luis Rueda and Juan Carlos Rojas [5] have proposed a pattern recognition technique based method for DNA micro array image segmentation. Using a clustering algorithm, the method has first performed an unsupervised classification of pixels, and the resulting regions have been subsequently subjected to supervised classification. Further fine tuning is achieved by discovering and merging region edges, and eliminating noise from the spots by morphological operators. The reasonable potential of the proposed technique for segmentation of DNA micro array images has been demonstrated by the very high accuracy obtained by the results on background and noise separation in various micro array images.

Volkan Uslan and Dhsan Omur Bucak [6] have performed a study in the microarray image processing to make a fine difference against the gene expressions. They have experimented and compared two methods for this. In particular, the segmentation phase of the microarray image has been analyzed. Clustering techniques have been used in addition to the segmentation methods utilized in commercial packages. They have examined the results of the application of fuzzy c-means and k-means techniques.

Maroulis D. and Zacharia E. [2] have presented a morphological modeling of spots based automatic micro array images segmenting approach. The proposed approach has been shown to be extremely effective even for noisy images and images with spots of diverse shapes and intensities by the carried out experiments.

3. PROPOSED METHODOLOGY FOR MICROARRAY SEGMENTATION USING CLUSTERING TECHNIQUES

The proposed approach consists of three important steps: preprocessing, gridding and segmentation. Initially, the microarray image is preprocessed using filtering and morphological operators and it is given for gridding to fit a grid on the images using hill-climbing algorithm. Subsequently, the segmentation is carried out using the proposed clustering algorithm, which is developed utilizing the fuzzy c-means clustering. The enhanced fuzzy c-means clustering algorithm (EFCMC) proposed in this paper makes use of the neighborhood pixel information along with the gray level information to effectively clustering the image whether the image may be affected by the noises or not. Then, the proposed method was employed the real microarray images and noisy microarray images in order to investigate the efficiency of the segmentation.

3. (A) PROPERTIES: ABSENT SPOT DETECTION AND NOISE TOLERANCE

Property 1: Absent Spot Detection

In general, the main challenge behind the microarray image segmentation is to accurately detect the absent spots (case 1) and in addition to accurately segment the high intensity spots (case 2). By looking into these challenges, case 2 can be easily achieved by the conventional clustering algorithms. But, the absent spots can be very difficult to find by the traditional clustering algorithms so in order to detect absent spots accurately, we make use of the neighborhood

dependent fuzzy factor to balance the image details whenever the membership values of the pixel values is calculated. The factor proposed in the clustering algorithm is adaptively changed in all iteration by considering the intensity values of neighborhood pixels and thus preserving the insensitiveness to boundary values by converging it to the central pixel's value.

Property 2: Noise Tolerance

The proposed algorithm can efficiently tackled the following two challenges even if the microarray image is corrupted by the noises.

Case 1: if the central pixel is not affected by the noise and some pixels within its neighbors may be corrupted by noise.

Case 2: if the central pixel is corrupted by noise and the other pixels within its neighbors may not be corrupted by noise. These two cases are easily dealt with the proposed clustering algorithm due to the introduction of the neighborhood dependent fuzzy factor which is easily ignoring the added noises. On the other hand, it can be adaptively adjusted their membership values according to its neighborhood pixel so that the segmentation accuracy of the proposed clustering algorithm can be improved even if the image is corrupted by the noises.

3. (B) QUALITY ASSESSMENT ANALYSIS

The input image taken for microarray image segmentation is given to the proposed algorithm to obtain the segmented results. Then, the quality index is computed based on the definition given below to assess the quality of the proposed approach. The quality index given in [7, 8] is used to evaluate the performance of the proposed approach in microarray image segmentation. The quality index is defined as follows,

$$q_{index}(IDSpot) = \frac{q_{com2R}(IDSpot) + q_{com2G}(IDSpot)}{2} \tag{1}$$

$$q_{com} = (q_{size} * q_{sig-noise} * q_{bkg1} * q_{bkg2})^{1/4} * q_{sat} \tag{2}$$

$$q_{size} = \exp(-|F_{pixel} - F_{pixel0}| / F_{pixel0}) \tag{3}$$

$$q_{sig-noise} = F_{mean} / (F_{mean} + B_{mean}) \tag{4}$$

$$q_{bkg1} = f1 / (BSD / B_{mean}), f1 = 1 / [\max(BSD / B_{mean})] \tag{5}$$

$$q_{bkg2} = f2 * (bkg0 / (bkg0 + B_{mean})), f2 = 1 / [\max(bkg0 / (bkg0 + B_{mean}))] \tag{6}$$

$$q_{sat} = \begin{cases} 1 & ; \% sat \leq 10 \\ 0 & ; else \end{cases} \tag{7}$$

Where, F_{pixel} → number of pixel per spot

F_{pixel0} → Average number of pixel per spot

F_{mean} → Mean of foreground pixel intensities per spot

- B_{mean} → Mean of local background pixel intensities
- BSD → Standard deviation of local background per spot
- $bkg0$ → Global average of background per array
- $\%sat$ → Percentage of saturated pixel per spot

Here, q_{size} assesses the irregularities of spot size, $q_{sig-noise}$ is a measure for the signal-to-noise ratio, q_{bkg1} quantifies the variability in local background and q_{bkg2} scores the level of local background.

Here, the quality index is computed for each spots presented in the microarray image after applying the clustering algorithms such as, k-means clustering, FCM and the proposed clustering (EFCMC). Then, the quality index obtained is plotted as graph shown in figure 4 and figure 5 for both channels. Fig 8.a and 9.a shows the input microarray image from different channels and Fig 4.b and 5.b illustrates the comparative quality index graph of the three algorithms corresponding to the input image. By analyzing these graphs, the proposed algorithm exactly detects the absent spots, which has zero quality index compared with other algorithms and at the same time, the intensity spots are accurately segmented since their quality index is greater than the other algorithms.

4.1 Experimental Dataset

The performance of the proposed approach is carried out in a set of real microarray images obtained from the publically available database [9]. The image taken from the database contains 24 blocks and each block contains 196 spots, i.e. 14×14 rows and columns of spots. Here, we have taken one block containing 196 spots from the real images and the experimentation is carried out on the extracted block.

4.2 Segmentation Results

This section describes the segmentation results of the proposed approach, which is then compared with the results obtained by the k-means clustering and fuzzy c-means clustering algorithms described in [48, 1]. The overall segmentation results of the proposed approach are given in figure 2.

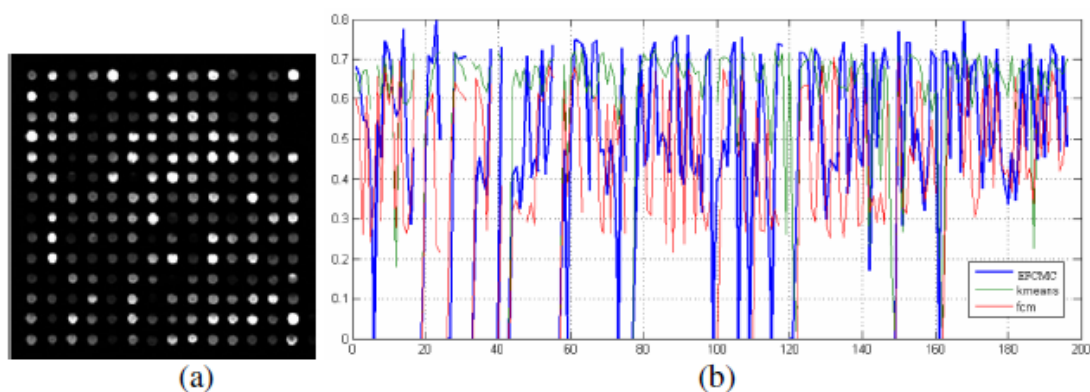


Fig. 1: (a) input microarray image-red channel (b) Comparative Quality index graph

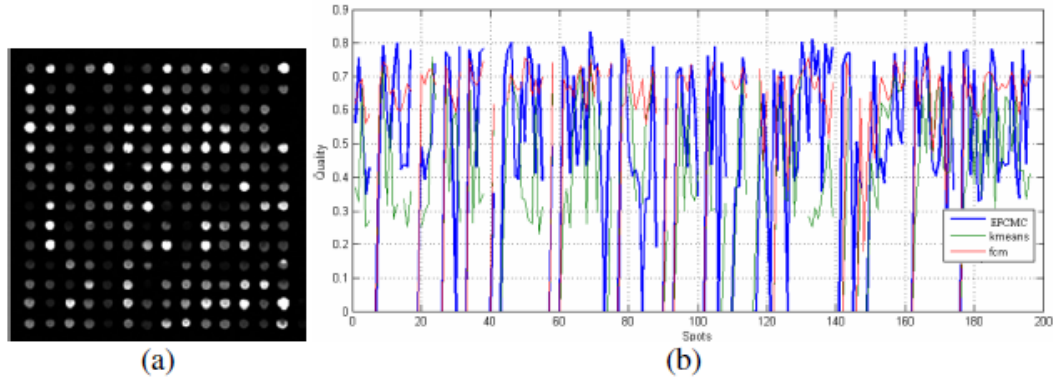


Fig. 2: (a) input microarray image-Green channel (b) Comparative Quality index graph

4.3 Analysis: Absent Spot Detection and Noise Tolerance

Here, we have analyzed the property of the proposed algorithm in identifying the low intensity spots and the detecting of absent spots. For analysis, we have taken typical spots, low intensity spots, absent spots and spots with various noises and then, different algorithms are applied on those spots to identify the efficiency of the algorithms. The obtained results are tabulated in the following figure 10. For a typical spot, three algorithms provide the identical results and for low intensity spots, FCM and EFCMC achieved better results compared with k-means clustering. The results obtained by the proposed algorithm for the absent spot is better compared with the k-means and FCM and those algorithms failed to identify the absent spots as per figure shown in below. For Gaussian and salt and pepper noise, the proposed approach accurately segments the spots and correctly removes the noisy pixels.

	Raw Image	K-means clustering	Fuzzy c-mean clustering	Proposed clustering
Typical spot				
Low intensity spot				
Absent spot				
Spots with Gaussian noise				

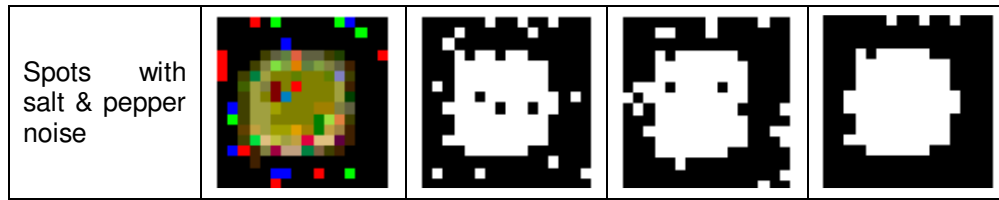


Fig. 3: Segmentation results in a typical spot, low intensity spot, absent spot and noisy spots using K-means, FCM and EFCMC

4.4 Quality Assessment Analysis for the Noisy Images

The noise tolerance property of the proposed approach is analyzed by adding the salt & pepper noise in the microarray image. The input image is added with the salt & pepper noise and it is given to the proposed approach for segmentation. The results obtained by the proposed approach are used to compute the quality index so that the noise tolerance property is analyzed. The noisy input shown in fig.5.a and 7.a is given to the different algorithms, which provides the segmented results shown in figure 4 and 6. As per segmentation results of the proposed approach, the noisy pixels are exactly removed but in case of k-means and FCM, the noisy pixels still presented in the results. When we looking into the quality index graph shown in fig 5 and 7, the proposed approach provide the zero quality index for absent spots but other algorithms cant able to provide the accurate results for absent spot and at the same time, it falsely identify the absent spots.

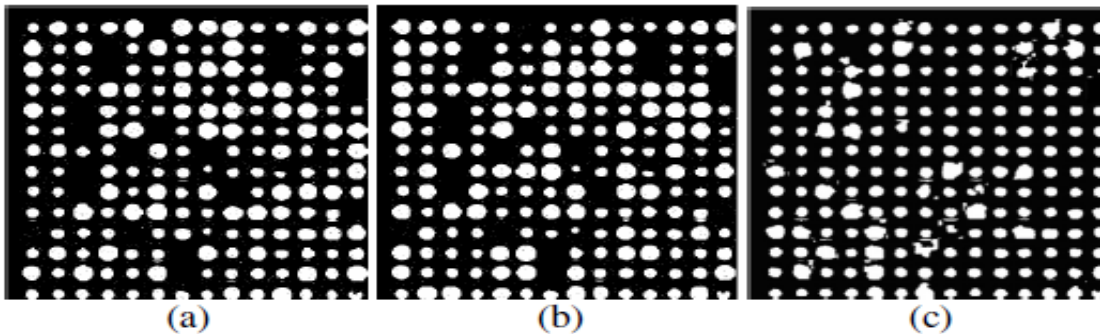


Fig. 4: Segmentation results of salt & pepper -Green channel (a) k-means clustering (b) FCM clustering (c) EFCMC

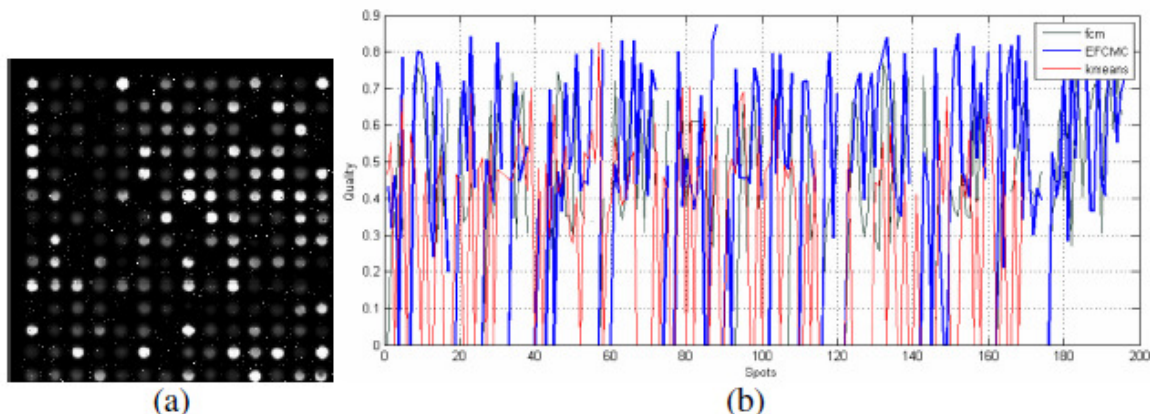


Fig. 5: (a) Noisy input (salt & pepper)-Green channel (b) Quality index graph-Green channel

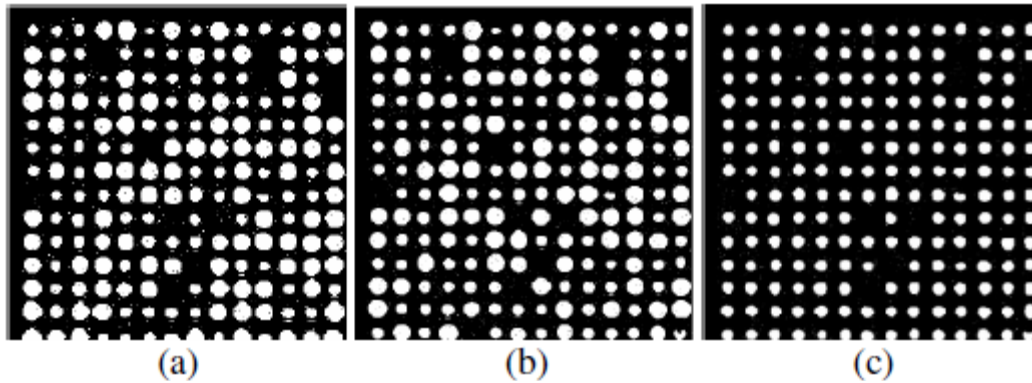


Fig. 6: Segmentation results of salt & pepper- Red channel (a) k-means clustering (b) FCM clustering (c) EFCMC

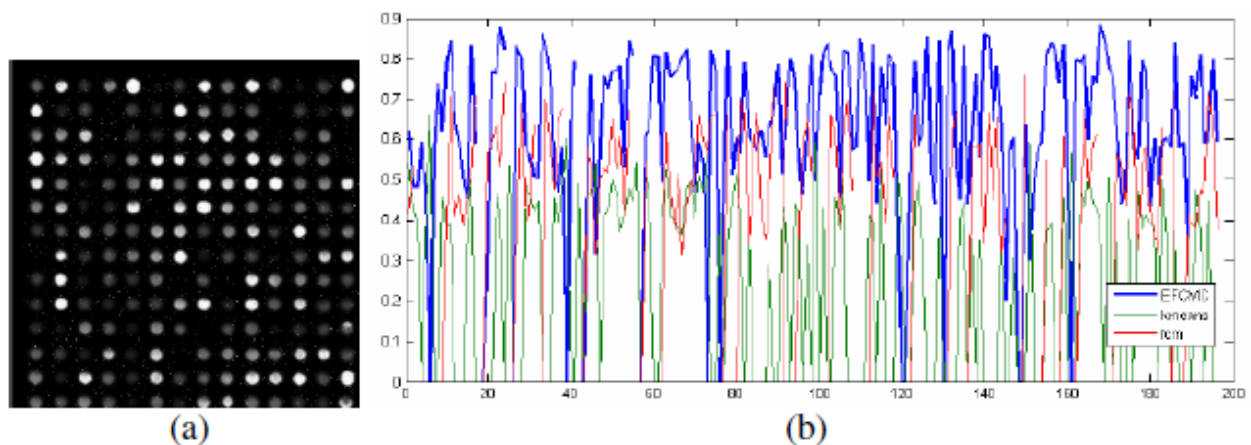


Fig. 7: (a) Noisy input (salt & pepper)-Red channel (b) Quality index graph-Red channel

5. CONCLUSION

In this paper, we developed and implemented and utilized an enhanced Fuzzy C-means clustering algorithm (EFCM) which was compared with the various algorithms in terms of quality index to investigate the performance efficiency in segmenting the microarray spot images. The comparative analysis proved that the proposed EFCM algorithm improved the quality index when compared with other algorithms.

ACKNOWLEDGEMENTS

The Authors would like to thank all the authors and contributors for this outcome of the paper.

6. REFERENCES

- [1] Wu, H., Yan, H., "Microarray Image Processing Based on Clustering and Morphological Analysis", In First Asia Pacific Bioinformatics Conference, 111-118, 2003.
- [2] Maroulis D., Zacharia, E., "Microarray image segmentation using spot morphological model", in proceedings of the 9th International Conference on information Technology and Applications in Biomedicine, Larnaca, pp: 1-4, 2009.
- [3] A.Sri Nagesh, Dr.A.Govardhan,Dr G.P.S.Varma, Dr G.S.Prasad, "An Automated Histogram Equalized Fuzzy Clustering based Approach for the Segmentation of Microarray images."ANU Journal of Engineering and Technology, pp 42-48 volume 2, Issue 2 December 2010, ISSN: 0976-3414.

- [4] "Microarray Images", from <http://lmpp.nih.gov/lymphoma/data/rawdata/>
- [5] Luis Rueda, Juan Carlos Rojas, "A Pattern Classification Approach to DNA Microarray Image Segmentation", in Proceedings of the 4th IAPR International Conference on Pattern Recognition in Bioinformatics, 2009.
- [6] Kaushik Suresh, Debarati Kundu, Sayan Ghosh, Swagatam Das, Ajith Abraham and Sang Yong Han, "Multi-Objective Differential Evolution for Automatic Clustering with Application to Micro-Array Data Analysis", Sensors, Vol. 9, pp. 3981-4004, 2009.
- [7] Sebastiano Battiato, Gianpiero Di Blasi, Giovanni Maria Farinella, Giovanni Gallo and Giuseppe Claudio Guarnera, "Adaptive techniques for microarray image analysis with related quality assessment", vo. 16, no.4, 2007.
- [8] U. Sauer, C. Preininger, and S. R. Hany, "Quick & simple: quality control of microarray data", Bioinformatics, Advance Access, 2004.
- [9] Laurie Heyer, "MicroArray Genome Imaging & Clustering (MAGIC) Tool", Davidson College, Available: <http://www.bio.davidson.edu/projects/magic/magic.html>
- [10] Ergüt E, Yardimci Y, Mumcuoglu E, Konu O. "Analysis of microarray images using FCM and K-means clustering algorithm", In: Proceedings of International Conference on Signal Processing, p. 116–21, 2003.
- [11] Volkan Uslan and Đhsan Ömür Bucak, "Microarray Image Segmentation Using Clustering Methods", Mathematical and Computational Applications, Vol. 15, No. 2, pp. 240-247, 2010.
- [12] A.Sri Nagesh, Dr G.P.S.Varma, Dr.A.Govardhan "An Improved Iterative Watershed and Morphological Transformation Techniques for Segmentation of Microarray Images" IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT, 2010.