

Machine Learning Techniques For Microarray Image Segmentation: A Review

A Sukanya

Dept. of Computer Applications
Bharathiar University
Coimbatore, India
sukan4mithul@gmail.com

R Rajeswari

Dept. of Computer Applications
Bharathiar University
Coimbatore, India
rrajeswari@rediffmail.com

Abstract— Microarray image analysis helps in the identification of new genes enables better understanding of their functions and used for the simultaneous analysis of gene expression levels under different conditions. It is used to diagnose various diseases and helps to develop more effective treatment for the targeted disease. Microarray image processing usually follows three stages (i) spot addressing or gridding (ii) segmentation (iii) intensity extraction. Among these stages segmentation is a vital stage. Four categories of methods for microarray image segmentation are (a) fixed /adaptive circle segmentation (b) Histogram based techniques (c) adaptive shape segmentation (d) Machine learning techniques. Thus the integration of machine learning in Image processing will contribute a better analysis of medical and biological data. The purpose of this paper is to discuss on various machine learning based techniques to segment the microarray images. After analyzing the features of all algorithms we conclude with several promising directions for the future research in microarray image segmentation.

Keywords— DNA Microarray processing, Supervised segmentation, unsupervised segmentation, Image segmentation.

I. INTRODUCTION

Microarray image analysis has made powerful the scientific community to understand the basic feature in the growth and development of life as well as to explore the hereditary causes of anomalies occurring in the functioning of human body [37]. Some of the applications of microarray technology are gene discovery [38], disease diagnosis and treatment [39], drug discovery and toxicological research [40]. Microarray images contain several blocks or sub grids which consist of a number of spots, placed in rows and columns as shown in fig.1. The intensity level of each spot represents the amount of sample which is hybridized with the corresponding gene.

The Microarray image analysis at gene expression level is necessary when dealing with vast amount of biological data. Microarray image processing has three stages. Among these three stages the second stage is the vital stage i.e. segmentation.

Microarray image analysis has made powerful the scientific community to understand the basic feature in the growth and development of life as well as to explore the hereditary causes of anomalies occurring in the functioning of human body [37]. Some of the applications of microarray technology are gene discovery [38], disease diagnosis and treatment [39], drug discovery and toxicological research [40]. Microarray images contain several blocks or sub grids which consist of a number of spots, placed in rows and columns as shown in figure1. The

intensity level of each spot represents the amount of sample which is hybridized with the corresponding gene.

The Microarray image analysis at gene expression level is necessary when dealing with vast amount of biological data. Microarray image processing has three stages. Among these three stages the second stage is the vital stage i.e. segmentation.

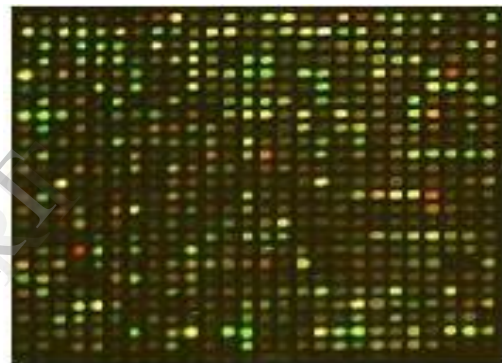


Fig. 1. The Microarray Image

There are four categories of methods for microarray image segmentation. They are (a) fixed or adaptive circle segmentation [1], [2] (b) Histogram-based techniques [41] (c) The adaptive shape segmentation [42]. Algorithms such as seed region growing, watershed transform and Markov random field have been employed. (d) The fourth category is based on machine learning techniques. There are two categories in this method. They are (a) supervised segmentation techniques and (b) unsupervised segmentation technique. More specifically, methods in unsupervised category employ clustering algorithms, such as k-means, hybrid k-means, fuzzy c-means, expectation-maximization and partitioning method for segmentation of microarray images. The Supervised method includes Bayes classifier approach. Thus the incorporation of machine learning in image processing has an enormous advantage, and provides a better analysis of medical and biological data.

The processing of microarray images [4] includes three stages. First, spots and blocks are preliminarily located from the images (gridding). Second, using the gridding information each microarray spot is segmented into background and foreground as shown in fig.2. Finally, intensity extraction calculates the foreground fluorescence background and foreground. Finally, intensity extraction calculates intensity,

which represents each gene expression level and the background intensities.

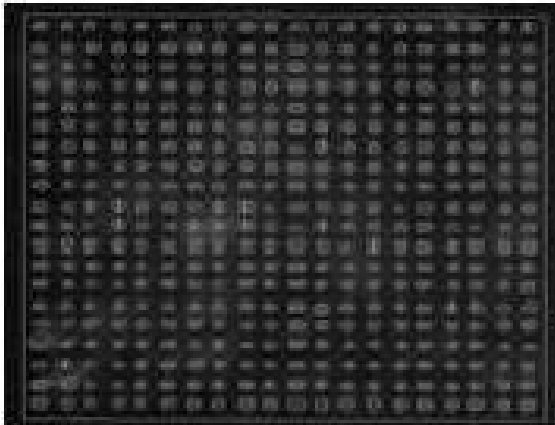


Fig. 2. The Segmented Microarray Image

The image analysis would be a rather a simple process, if all the spots had circular shape, similar size, and the background was noise and artefact free. However, a scanned microarray image has none of the above characteristics, thus microarray image analysis becomes a difficult task.

In this paper, several microarray segmentation algorithms based on machine learning techniques to segment microarray images are described as shown in figure 3. Section II describes various unsupervised techniques used for microarray image segmentation. Section III describes various segmentation techniques. Section IV gives the conclusion.

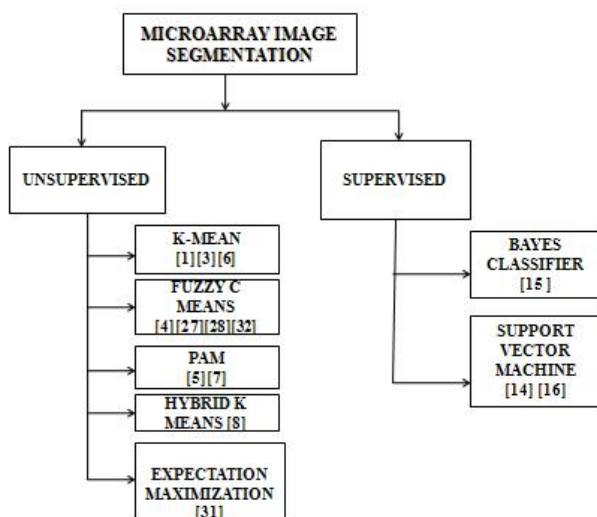


Fig. 3. The Microarray Image Segmentation

II. UNSUPERVISED SEGMENTATION TECHNIQUE

The various methods in this category employ clustering algorithms, such as k-means, hybrid k-means, fuzzy c-means, expectation-maximization and partitioning method for segmentation of microarray images.

The k-means segmentation algorithm is based on the traditional k-means clustering [3]. It employs a square-error criterion, which is calculated for each of the two clusters. K-means is commonly fed with the intensity of each pixel in the microarray image as features. However, many algorithm have already been developed, which use more intensity features of each pixel such as mean intensity of the neighbourhood of the pixel, or spatial features.

For instance, [5] employed the K-means algorithm using only the intensity of the pixel as a feature, while [8] used three intensity-based features as well as the Euclidean distance between the pixel and the center of the spot, as the fourth feature. Both the channels of the microarray image are segmented simultaneously. Thus, for each pixel the intensities from both channels are combined to one feature vector. The number of cluster centers K is set usually to two, due to the fact that the segmentation is used for characterizing the pixels of the image as foreground or background pixels.

A number of studies utilize the fuzzy c-means (FCM) [6], instead of the k-means algorithm. FCM is a data clustering technique in which a dataset is grouped into k clusters with each data point in the dataset belonging to a cluster to a specified degree. For example, a certain pixel that lies close to the centroid of a signal cluster will have a high degree of belonging or membership to that cluster and another pixel that lies far away from the centroid of a cluster will have a low degree of membership to that cluster.

A more robust method than K-means and FCM clustering technique is the Partition around Medoids (PAM) [9] clustering. PAM minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances. The algorithm first computes a number of representative objects, called medoids. A medoid can be defined as that object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal. The representative objects are called centrotypes. After finding the set of medoids, each object of the dataset is assigned to the nearest medoid.

[7] developed a segmentation method based on the PAM to extract the target intensity of the spots. The distribution of the pixel intensity in a grid containing a spot is assumed to be the superposition of the foreground and the local background. The partitioning around medoids is used to generate a binary partition of the pixel intensity distribution. The medoid of the cluster members are chosen as the cluster representatives.

Another method called the hybrid k-means algorithm is an extended version of the k-means segmentation approach [10]. The machine learning contribution includes repeated clustering in order to increase the number of foreground pixels. As long as the minimum amount of foreground pixels is not reached, the remaining background pixels are clustered into two groups and the group with pixels of higher intensity is assigned as foreground. After the clustering, the number of outlier pixels in the segmentation result is reduced with mask matching.

Model-based segmentation algorithm [11] is a two- step method for spot segmentation. The main steps of the method are model-based clustering of pixel values and spatial extraction of connected components. Model-based clustering forms the initial segmentation into at most three different clusters sharing similar intensity values, which are the background, the spot with background or artefact, and the spot

foreground. Model-based clustering relies on Gaussian mixture models, and the number of clusters is defined based on data by using Bayesian Information Criterion (BIC). Spatial connected component removal is used for excluding small disconnected clusters that are assumed to be artifacts from the spot foreground pixels. Though the algorithm actually provides spot foreground as a separate cluster,[11] used both the foreground and the spot with artefact clusters to denote the foreground. The algorithm combines both spatial and intensity information in segmentation. Similarly as in the Mann-Whitney method, a circular target mask is first used for separating all possible foreground pixels from the known back- ground. Pixels inside the target mask having larger intensity than the local background mean and standard deviation, are considered as accepted foreground pixels and are used for calculating new spot centroids. Thereafter, the foreground and background pixels for each spot are iteratively redefined.

[13] explains that DNA (Deoxyribonucleic acid) microarray image segmentation based on pattern recognition techniques performs an unsupervised classification of pixels using a clustering algorithm, and a subsequent supervised classification of the resulting regions. Additionally morphological operators are used to eliminate noise from the spots. The results obtained on various microarray images shows that this technique is quite promising for segmentation of DNA microarray images, obtaining a very high accuracy on background and noise separation.

[19] has developed a parallel strategy based on domain decomposition to determine the number of clusters and inherent properties of spectral clustering. They have tested the method on microarray images.

The segmentation of cDNA (Complementary DNA) microarray spots is required to analyze the intensities of microarray images for biological and medical investigation. [20] explains a method using kernel density estimation to segment two-channel cDNA microarray images. This method groups pixels into a foreground and a background. The segmentation performance of this model is tested and evaluated with reference to 16 microarray data. In particular, spike genes with various contents are spotted in a microarray to examine and evaluate the accuracy of the segmentation results and duplicate design is implemented to evaluate the accuracy of this model. This method can cluster pixels and estimate statistics regarding spots with high accuracy.

[22] explains the k-means, fuzzy c-means methods and thresholding methods are used and compared to segment microarray image. The intensity of each spot is calculated and the gene expression is observed. And they conclude that fuzzy c-means is more efficient than the k-means in terms of clustering the signal pixels. This is because fuzzy c-means has ensured a sensitive classification when compared with the k-means algorithm.

[28] stated that the target intensity of the spots are extracted using clustering-based segmentation. [29] demonstrated clustering based approaches such as fuzzy c-means clustering for automated spot segmentation. [21] stated the extraction of spot features from a gene microarray image, which along with the spot intensity can be used for statistical analysis of spot shape and intensity variations.

[31] demonstrated two clustering methods, fuzzy c-means and k-means algorithms and the results of both are compared. The experimental result shows that fuzzy c-means has ensured a sensitive classification of the weak spots when compared with k-means algorithm. [32] Presented a technique for removing genes noises based on the offset vector field and segmenting genes using the expectation maximization algorithm. [33] Proposed a fuzzy c-means with bi-dimensional empirical mode decomposition (FCMBEMD) for segmenting the microarray image to reduce the effect of noise. [34] Demonstrated the automatic spot detection of cDNA microarray images using mathematical morphology methods.

III. SUPERVISED SEGMENTATION TECHNIQUE

Many supervised techniques are employed to deal with microarray image segmentation. In the software Spot [14], the seeded region growing algorithm [15] was used for microarray segmentation for the first time. The algorithm segments each spot by iteratively growing separate regions with respect to a set of predefined seed points providing a starting point for the segmentation. In each iteration, the algorithm includes the most homogenous pixels from the neighborhood to the segmented regions. The algorithm aims at ensuring that the final segmented regions are as homogeneous as possible for the given connectivity constraint. Finally, the region originating from the foreground seeds is considered as the spot foreground, and the region originating from the background seeds as the background.

[16] explains that the classification technique segments the microarray image by classifying their pixels into signals, background and artefact pixels using support vector machines (SVM). This method requires training data set of spots with pixel-by-pixel information for the real images, while for the simulated images the training data are extracted directly during the production of images. This method performs direct characterization of each pixel to the designated category and it is more advantageous compared to clustering based methods.

Here we include another segmentation method called Bayes classifier approach. This method classifies the pixels of the image into two categories (foreground and background) using classification techniques. This classification-based approach directly classifies each pixel to the designated category. More specifically, the Bayes classifier [17] is employed to classify the pixels of the image into different classes. Thus, the method can classify the pixels of the image into signal, background and artifacts.

[18] explains a method based on support vector machine, classifies the foreground pixels from the background pixels. The canny method, morphological method and fixed circle method are also combined into the SVM method to improve the performance. To verify the performance of the method, images drawn from the SMD and GEO database are used for comparing with the k-means method and the GenePix.Pro. Experiment results reveal that this method has two advantages. It can segment most of the microarray spots effectively, and its segmentation result is closer to the real spot than the k-means method and the GenePix.Pro.

[23] proposed an approach based on the markov random field modeling of the microarray spot regions in which the contextual information is also considered.[24] proposed

markov random field (MRF) based approach to high level grid segmentation which is robust to common problems encountered with array images and does not require calibrations. They also developed an active contour method for single spot segmentation that describes objects in images by properties of their boundaries.

[25] presented a novel automatic approach to locate the spots without the formation of grids. [26] proposed an integration of active contour approach and Fisher criterion to capture the boundary and the region information of microarray images. [27] stated a method to reduce the error of the edge detection which is influenced by noise and tilt array during segmentation of microarray image. [36] described the segmentation of microarray images using an improved seeded region growing method.

[35] explains soft threshold provides better segmentation output. According to varying spot sizes of input image, the output segmented spot sizes also get varied. Fuzzy c-means clustering performs well when compared with adaptive threshold method of segmentation. The soft thresholding gives good segmentation and improved log intensity values when compared with all other methods. Thus this method provides accurate segmentation of spots in microarray images.

[30] proposed a new approach for segmentation of the microarray images. They have used chan-vee approximation of the mumford-shah model and the level set method for image segmentation.

IV. CONCLUSION

An overview of the already developed methods for microarray image segmentation is presented in this paper. We have categorized all the methods of machine learning techniques for microarray image segmentation. Image segmentation is an important stage in the microarray image analysis. Reliability of this stage strongly influences the results of data analysis performed on extracted gene expressions. This survey explains various machine learning techniques and their features to segment the microarray images.

ACKNOWLEDGMENT

The authors are thankful to Bharathiar University for the valuable support.

REFERENCES

- [1] M.B. Eisen, ScanAlyse, Available <http://rana.Stanford.EDU/software/>, 1999.
- [2] J.Buhler, T.Idekar, D.Haynor, D.Dapple, "Improved Techniques for finding spots on DNA microarrays", UWCSE Technical report, University of Washington, 2000.
- [3] J.B.MacQueen, "Some methods for classification and Analysis of Multivariate observations", Proceedings of fifth Berkely symposium on Mathematical statistics and probability, University of California press, Bibsonomy publication, pages 281-297,1967.
- [4] M.Schena, R.W.Davis, P.O.Brown, "Quantitative Monitoring of Gene Expression Patterns with a complementary DNA microarray", Science Journals, Volume 270, number 5235, pages 467-470, 1995.
- [5] E.Ergut, Y.Yardimci, E.Mumcuoglu, O.Konu, "Analysis of microarray images using FCM and K-means clustering algorithm", pages 116-121, In:Proceedings of International conference in signal processing, 2003.
- [6] C.James, Bezdek, Robert ehrlich, William full, "FCM:The Fuzzy C-means Clustering Algorithm", Computers and Geosciences Volume 10, Number 2-3, pages 191-203, 1984.
- [7] R.Nagarajan, "Intensity-Based Segmentation of cDNA Microarray Images", IEEE Transaction on Medical Imaging, volume 22, pages 882-889, 2003.
- [8] H.Wu, H.Yan, "Microarray Image Processing Based on Clustering and Morphological Analysis", Published In:Proceedings of the First Asia-Pacific bioinformatics conference on Bioinformatics, Volume 19, pages 111-118, 2003.
- [9] L.Kaufman, P.J.Rousseeuw, "Finding Groups in Data - An Introduction to Cluster Analysis", Wiley Interscience, New York, 1989.
- [10] D.Bozinov, Rahnenfuhrer, "Unsupervised Technique for Robust Target Separation and Analysis of DNA Microarray Spots through Adaptive Pixel Clustering, Volume 18, pages 747-756, 2002.
- [11] QLi, C.Fraley, RE.Bumgarner, KY.Yeung, A.E Raftery, Donuts, Scratches and blanks, "robust model-based segmentaion of microarray images", Bioinformatics, Volume 28, pages 823-830, 2012.
- [12] X.Wang, S.Ghosh, S.W Guo, "Quantitative quality control in microarray image processing and data acquisition", Oxford Journals, Volume 29, pages e75, 2001.
- [13] Luis Rueda, Juan Carlos Rojas, "A Pattern Classification Approach to DNA Microarray Image Segmentation", In:Proceedings of Pattern recognition of Bioinformatics, Volume 5780, pages 319-330, 2009.
- [14] YH.Yang, ML.Buckley, S.Dudoit, TP.Speed, "Comparision of methods for image analysis on cDNA microarray data", The computer Journal, Volume 41, pages 578-588, 2002.
- [15] R.Adams, L.Bischof, "Seeded region growing", IEEE Transaction on Pattern analysis and Machine intelligence, Volume 16, pages 641-647, 1994.
- [16] Nikolas Giannakeas, S.Karvelis, P.Exarchos, G.Kalatzis, "Segmentation of microarray images using pixel classification-Comparision with clustering based method", Computers in Biology and medicine, Volume 43, pages 705-716, 2013.
- [17] R.C Gonzalez, R.E.Woods, S.L Eddins, "Digital image processing using MATLAB", Journal of Optical technology, Volume 77, pages 245-252, 2004.
- [18] Guifang Shao, Tingna Wang, Hong, Zhigeng Chen, "An Improved SVM Method for cDNA Microarray image segmentation", pages 391-395, 2008.
- [19] Sandrine, Noailles, Daniel Ruiz and Ronan Guivarch, "Microarray image segmentation using parallel clustering", 6th International Con-ference on Practical Applications of Computational Biology and Bioinformatics, University of toulouse, IRIT, Volume 154, pages 1-9, 2012.
- [20] Tai-been, Henry, yun, Hsiu-jen lan, "Segmntation of cDNA microarray image using kernel density estimation", Journal of Biomedical informatics, Volume 41, pages 1021-1027, 2008.
- [21] I.Kasif, O.Hero and M.Siddiqui, "Mathematical Morphology applied to Spot Segmentation and Quantification of Gene Microarray Images", IEEE Xplore, Volume 1, pages 926-930, 2002.
- [22] M.G.Kavitha, D.S.Suresh Kumar, "Comparison of Clustering Techniques for Microarray Image Segmentation", International Journal of Scientific Engineering Research, Volume 4, Issue 9, ISSN 2229-5518, September-2013.
- [23] M.H.Asyali, M.M.Shoukri and K.S.AbuKhabar, "Segmentation of Microarray cDNA Spots using MRF-based method", IEEE Transaction on medical imaging, Volume 1-4, pages 674-677, 2003.
- [24] Mathias Katzer, Franz Kummert and Gerhard Sarger, "Methods for Automatic Microarray Image Segmentation", IEEE Transactions on NanoBioscience, Volume 2, pages 202-214, 2003.
- [25] A.Sreedevi, D.S.Jangamashetti, "Automatically Locating Spots in DNA Microarray Image using Genetic Algorithm without Gridding ", International journal on signal and image processing, Volume 4, No 6, 2009.
- [26] Jinn Ho and Wen -Liang Hwang, "Segmenting Microarray Image Spots using an Active Contour Approach", Proceedings of the International Conference on Image Processing, pages 273-276, 2007.
- [27] Tsung-Han Tsai, Chien Po Yang, Wei Chi Tsai and Pin hua Chen, "Error Reduction on Automatic Segmentation in Microarray Image", IEEE Transaction on medical imaging, pages 76-81, 2007.

- [28] Radhakrishnan Nagarajan, "Intensity-Based Segmentation of Microarray Images", IEEE Transaction on Medical Imaging, Volume 22, pages 882-889, 2003.
- [29] Wang Yu-Ping, Gunampally, Maheshwar reddy and Cai Wei-Wen, "Automated Segmentation of Microarray Spots using Fuzzy Clustering Approach", IEEE workshop on Machine learning for signal processing, pages 387-391, 2005.
- [30] A.Kaustubha Mendhurwar, Rajasekhar Kakumani and Vijay Devabhaktuni, "Microarray Image Segmentation using Chan-Vese Active Contour Model and Level Set Method", 1st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, pages 3629-3632, 2009.
- [31] Volkan Uslan and Ihsan Bucak, "Microarray Image Segmentation Using Clustering Methods", Mathematical Computational Applications, pages 240-247, 2010.
- [32] Weng Guirong and Su Jian, "Microarray Images Processing using The Offset Vector Field and Expectation Maximization Algorithm", 4th International conference on Bioinformatics and Biomedical Imaging, pages 1-3, 2010.
- [33] J.Harikiran, D.RamaKrishna, M.L.Phanendra, Dr.P.V.Lakshmi and Dr.R.Kiran Kumar, "Fuzzy C-means with Bi-dimensional Empirical Mode Decomposition for Segmentation of Microarray Image", International Journals on Computer Science Issues, Volume 11, 2012.
- [34] Chiao-ling Shih, Hung-Wen Chiu, "Automatic Spot Detection of C-DNA Microarray Images using Mathematical Morphology methods", 2003.
- [35] P.Rajkumar, Dr.Ila.Vennila, K.Nirmalakumari, "An Intelligent Segmentation Algorithm For Microarray Image Processing", International Journal on Computer Science and Engineering (IJCSE), 2013.
- [36] J.Deepa, Tessamma Thomas, "Automatic Segmentation of DNA Microarray Images using an Improved Seeded Region Growing Method", World Congress on Nature and Biologically Inspired Computing, pages 1469-1474, 2009.
- [37] J.Angulo and J. Serra, "Automatic Analysis of DNA Microarray Images Using Mathematical Morphology", Bioinformatics, Volume 19, pages 553-562, 2003.
- [38] Nicole, L.W van, L.W van, Peijnenburg, Asaph, Arjen, Esther, Jaap Keijer, "The application of DNA microarrays in gene expression analysis", Journal of Biotechnology, Volume 78, pages 271-280, 2000.
- [39] Martin Klumpp and Lorenz M Mayr, "Trends in drug discovery technologies" conference report from MipTec 8 - 11, Novartis Institutes of BioMedical Research, Volume 1, No. 4, pages 361-364, 2006.
- [40] B.Ballantyne, T.C Marrs, and T. Syversen, "Basic Elements of Toxicology. General, Applied and Systems Toxicology", Wiley Online Library, Volume 4, pages-115-212, 2009.
- [41] A.N.Jain, T.A.Tokuyasu, A.M.Snijders, R.Segraves, D.G.Albertson and D.Pinkel, "Fully automatic quantification of microarray image data", Genome Research, pages 325-332, 2002.
- [42] B.T.M. Roerdink, Arnold Meijster, "The Watershed Transform: Definitions, Algorithms and Parallelization Strategies", Fundamental Informaticae, Volume 41, pages 187-228, 2000.
- [43] A. Ash, T. Douglas, M. Charles, Matt, "Towards a novel classification of human malignancies based on gene expression patterns", The Journal of Pathology, Volume 195, pages 41-52, 2001.