Journal of Physics: Conference Series, Volume 2071, 012032
International Conference on Biomedical Engineering (ICoBE 2021)

DNA Microarray Image Segmentation Using Markov Random Field Algorithm

Khairul Anuar Mat Said [a], Asral Bahari Jambek [b]

Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Malaysia

[a] anuarsaid91@gmail.com, [b] asral@unimap.edu.my

Abstract. A deoxyribonucleic acid (DNA) microarray image requires a three-stage process to enhance and preserve the image's important information. These are gridding, segmentation, and intensity extraction. Of these three processes, segmentation is considered the most difficult, as its function is to differentiate between features in the foreground and background. The elements in the foreground form the object or the vital information of the image, while the background features less critical information for DNA microarray image analysis. This paper presents a study that utilises the Markov random field (MRF) segmentation algorithm on a DNA microarray image. The MRF algorithm evaluates the current pixel depends on its neighboring pixels. The experimental results show that the MRF algorithm works effectively in the segmentation process for a DNA microarray image.

1. Introduction

Scientists can investigate thousands of gene expressions simultaneously using a DNA microarray image [1]. Initially, these gene expressions are kept on a glass slide containing thousands of probes [2]. The glass slide is then used to perform hybridisation between two samples. The two cDNA (complementary DNA) samples are stained with different fluorescent dyes; Cy3 dye is used for the normal sample, and Cy5 dye is used for the malignant sample [3]. When the hybridisation step is completed, the DNA microarray image is created, and the intensity of the spots on the image is calculated. The intensity of the dots shows their state, and the aggregate results allow scientists to evaluate and study gene expression [4], [5]. A high-quality DNA microarray image is required to generate this information.

The DNA microarray image may become polluted during the scanning procedure, compromising gene expression analyses [6]. One way for improving and optimising the microarray image is image processing [7]. The processing of the microarray image is divided into three parts. First, gridding (addressing) is employed to determine each spot's location. Second, segmentation to detect the features in the image's foreground (object). Finally, intensity extraction is employed to calculate the intensity of each spot [8].

An MRF segmentation for a DNA microarray image is presented in this research. Based on the experimental results, the performance of this algorithm is then evaluated. Section II introduces and explores various ways to image segmentation using MRF algorithms. Section III describes the approach employed in this investigation. Section IV discusses and analyses the experimental data, and Section V concludes this study.

2. Literature reviews

The authors surveyed computer vision and imaging modelling of MRF in the paper [9]. Over the last decade, various MRF models have been developed for a variety of vision applications. According to the results of their survey, higher-order MRFs are commonly used. MRFs have also been utilised to improve

Authors' Final Manuscript.
Published paper link: https://iopscience.iop.org/article/10.1088/1742-6596/2071/1/012032/pdf

machine learning approaches in image processing, parameter learning, and MRF model structure learning. Due to the promising outcomes, MRFs attract lots of scientific interest currently.

The researchers evaluated the evidentiary Markov random field (EMRF) to conventional MRF, fuzzy MRF (FMRF), and traditional evidential approaches for image segmentation in their research [10]. The label field was defined using a credal partition based on evidence theory, and the EMRF model was optimised using the iterated conditional modes (ICM) algorithm. The experimental results suggest that the proposed algorithm outperforms the traditional MRF, FMRF, and evidential techniques in terms of segmentation.

In their publication [11], the authors suggested an automatic segmentation method for image data from complicated organotypic three-dimensional (3D) cell culture models based on MRFs. An organotypic culture is a three-dimensional cell culture model of epithelial tumours such as prostate cancer. The suggested method accurately captures aspects of multicellular tumour and cancer-associated fibroblast (CAFs) structures in 3D stack picture data; these features are extraordinarily distinct but equally important in biological terms. As a result, this approach was able to identify crucial traits that could not previously be determined. Moreover, the experimental results demonstrate that the proposed approach works well.

The authors of the paper [12] attempted to solve the energy minimisation challenge in robotic perception using a proposed solution involving a novel superpixel-based segmentation. First, geometrical constraints are encoded using the MRF inference. Then, the geometrical information of the superpixels is described in the energy function, and a global minimum for the energy function can be determined using the rapid primal-dual (FastPD) approach. The NYU v1 dataset was used to test this strategy, and the results were examined and analysed. Compared to other contemporary and commonly used approaches, the results showed that the suggested method had higher values for correct detection and lower values for "missed instances" and "noise instances."

By merging a collection of top-down and bottom-up segmentation maps, the authors in [13] developed a domain-specific MRF (DS-MRF) segmentation to increase segmentation performance. The MRF framework governs both segmentation maps. Bottom-up segmentation (BUS) maps are created by altering the parameters of an unsupervised segmentation algorithm, whereas top-down segmentation maps are created by combining domain-specific maps. The proposed method was used to recognise airports in remote sensing images. The experimental findings show that the suggested method outperforms efficient graph-based segmentation (EG), multiscale normalised cut, mean shift, and ROI-SEG methods.

Table 1. The comparison between various applications using MRF algorithm-based segmentation.

Paper	Application	Type of image	Threshold	Accuracy	Complexity	Special features
[9]	Survey	Colour	N/A	N/A	Normal	N/A
[10]	Case study	Various images	No	N/A	Normal	EMRF
[11]	Medical	Multi-channel 3D image	No	No	Normal	N/A
[12]	Machine vision	Colour	Yes	Yes	Normal	Superpixel-based segmentation
[13]	Remote sensing	Colour	Yes	Yes	Normal	DS-MRF

Authors' Final Manuscript.
Published paper link: https://iopscience.iop.org/article/10.1088/1742-6596/2071/1/012032/pdf

Table 1 summarizes various applications that use MRF segmentation, as explained in this section. The comparison demonstrates that MRF segmentation is utilised in various applications, and the experimental results show that the algorithm is effective in image segmentation. In this study, a conventional MRF segmentation is performed on a DNA microarray image, and the experimental results will be examined.

2.1. Markov Random Field

The MRF algorithm evaluates the current pixel value, taking into consideration the neighbouring pixels [14]. Figure 1 shows an example of a neighbourhood system, where N0 denotes the site of interest and N1, N2, N3, N4, and N5 its neighbours. This neighbourhood system and its groupings, known as cliques, can be understood as follows. Pixels labelled N_1 indicate the sites of the first-order neighbourhood system, as shown in figure 1 (a). Pixels labelled N_1, and N_2 indicate the sites of the second-order neighbourhood system, as shown in figure 1 (b). Figure 1 (c) shows the *n*th-order neighbourhood system, in which n = 5. The neighbouring sites can be viewed as a single element enclosing the site (N_0).

Figure 2 shows some examples of several types of a clique; the cliques for the first-order neighbourhood system are shown in (a). The cliques in (a), (b), (c), and (d) constitute the second-order neighbourhood system. The figure shows that the number of types of cliques increases as the order of the neighbourhood increases.

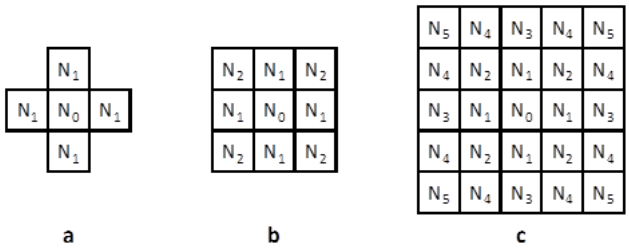


Figure 1. Neighbourhood systems [14]: (a) first-order; (b) second-order; (c) fifth-order.

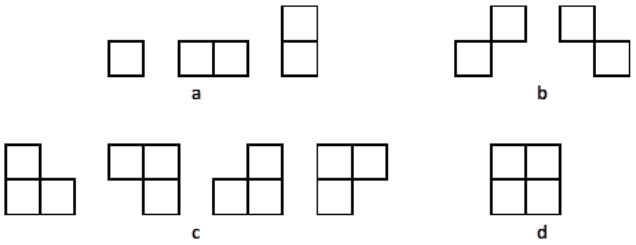


Figure 2. Types of clique [14]: (a) single-site and horizontal and vertical pair-site; (b) diagonal pair-site; (c) triple-site; (d) quadruple-site.

The probability in equation (1) below is defined as the Gibbs distribution [14]. The parameter Z is the normalising constant, β is a positive constant, and $U(x)$ is the energy function also knows as Gibbs energy [15].

$$P(x) = Z^{-1} \times exp\,[-\beta U(x)] \tag{1}$$

where

$$Z = \sum_{x \in L} exp[\beta U(x)] \tag{2}$$

$$U(x) = \sum_{c \in C} V_c(x) \tag{3}$$

Authors' Final Manuscript.
Published paper link: https://iopscience.iop.org/article/10.1088/1742-6596/2071/1/012032/pdf

where $V_c(x)$ denoted the sum over all clique potentials of the given neighbourhood systems. The *s* and *n* are the neighbours of each other, which the only *n* is the elements of neighbourhood systems [15].

$$V_C(x) = \begin{cases} 1 & x_s \neq x_n \\ -1 & x_s = x_n \end{cases} \quad (4)$$

Equation (5) defined the Bayesian theorem, represented by equation (6). The $P(Y)$ have a total probability that equal to one, and thus considered to be a constant. Therefore, the posterior probability $P(X|Y)$ is proportional to the prior probability $P(X)$ and likelihood probability $P(Y|X)$, as expressed in equation (7) [15].

$$Posterior = \frac{Likelihood \times Prior}{Evidence} \quad (5)$$

$$P(X|Y) = P(Y|X)\, P(X) \,/\, P(Y) \quad (6)$$

$$P(X|Y) \propto P(X)\, P(Y|X) \quad (7)$$

Equation (8) defined the conditional probability $P(Y|X)$ follow a Gaussian distribution by considering the image intensity representing either the foreground or background [15].

$$P(Y|X) = (2\pi\sigma_s^2)^{-0.5} exp\{-[(y - \mu_s)/\sigma_s{}^2]\} \quad (8)$$

where *y* is the observed image, and μ_s and σ_s are the parameters of the distribution of the x_s.

Equation (9) is the maximum a posteriori (MAP) estimation of posterior given by equation (7). Then equation (10) is produced by substitute equation (1) and equation (8) into equation (9). Then the equation (10) is optimising further by taking the negative and generates the minimisation of the equation as stated in equation (11) [15].

$$\hat{x} = arg \max_x \{log\, p\,(y|x) + log\, p\,(x)\} \quad (9)$$

$$\hat{x} = arg \max_x \{-[(y - \mu_s)/\,\sigma_s] - 0.5\, log(2\pi\sigma_s^2) - \beta U(x)\} \quad (10)$$

$$\hat{x} = arg \min_{x_s \in L} \{[(y - \mu_s)/\,\sigma_s] + 0.5\, log(2\pi\sigma_s^2) + \beta U(x)\} \quad (11)$$

3. Methodology

This section discusses the segmentation of a DNA microarray image [16] (2200×7300 pixels) using the MRF algorithm [15]. Figure 3 shows a fragment of an image of size 446×431 pixels from the DNA microarray image, which is used as the input image for this work. The yellow box shown in figure 3 is the worst case for this segmentation process.

Figure 4 presents a flow chart of MRF segmentation. Firstly, the input image is converted into a greyscale image. Then, an initial labelled image, based on the input image, is generated. Next, MRF segmentation is applied to the initial labelled image to generate a new labelled image. This process continues until the maximum number of iterations is reached. Finally, the segmented image is produced after the iteration is completed. In this study, a second-order neighbourhood system was chosen, and the iteration was set to 5. All experiments were performed using MATLAB R2019a software on a Windows 7 operating system with a 2.50GHz Intel Core i5 CPU and 8GB of RAM.

Authors' Final Manuscript.
Published paper link: https://iopscience.iop.org/article/10.1088/1742-6596/2071/1/012032/pdf

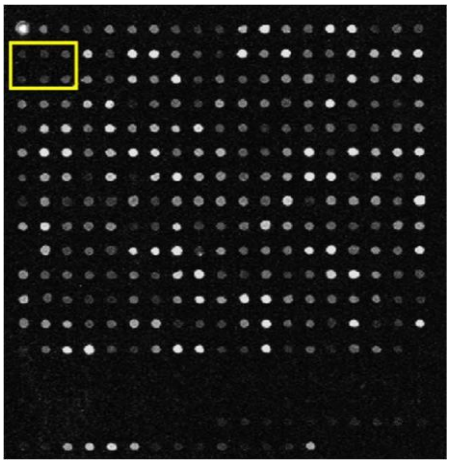


Figure 3. The 446×431 pixel region of a DNA microarray image of 2200×7300 pixels used as an input image [16].

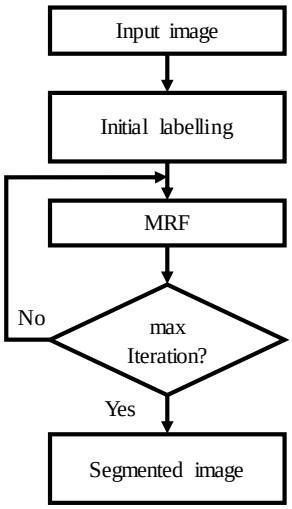
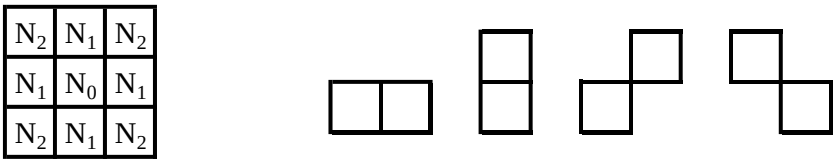


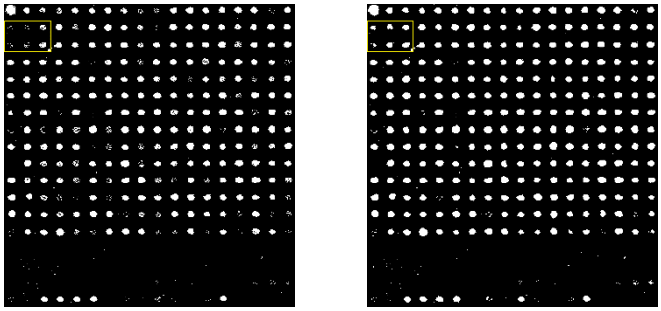
Figure 4. Flowchart of the MRF segmentation algorithm of microarray image processing [15].



(a) (b)

Figure 5. (a) The second-order neighbourhood system, and (b) the horizontal, vertical, and diagonal pair-site cliques.

The first labelling was generated as part of the initial segmentation of the input image, with the initial foreground labelled as two, and the initial background labelled as one. In this investigation, the second-order neighbourhood system is selected to guide the Gibbs energy, as stated in equation (3). This system allows the comparing method only included the neighbourhood in labelled N_1 and N_2 for each N_0, as shown in figure 5 (a). Following the system, the comparing method will be using horizontal, vertical, and diagonal pair-site to calculate the Gibbs energy, as shown in Figure 5 (b).



(a) (b)

Figure 6. (a) Initial labelled image and (b) the MRF segmentation results after five iterations.

(a) (b) (c) (d) (e)

Figure 7. The iteration effects concerning MRF segmentation: (a) first iteration, (b) second iteration, (c) third iteration, (d) fourth iteration and (e) fifth iteration.

The total energy is the summation between the Gibbs energy and the log-likelihood, which can be calculated as stated in equation (8) of each labelled possibility. The possibility refers to the site's label (N_0), label '1' and '2'. After both total energies are computed, the label with minimum total energy is selected as the new site (N_0) label. Then, the total energy of the following site is calculated till the new labelled image is generated. This new labelled image results from the MRF segmentation of the initial labelled image, and the process is repeated until the maximum iteration is reached. After five iterations, the final segmented image is generated. The label '2' of the segmented image representing the foreground while label '1' representing the background.

4. Results and Discussion

The previous section described the MRF algorithm and the steps used in the segmentation in this experiment. The segmentation experiment was performed using MATLAB simulation tools on a Windows operating system. The experimental results of all steps are presented here. Firstly, the microarray image is cropped to the size described above. Next, this cropped image is converted from true colour (RGB) to a greyscale image, as shown in figure 3.

Figure 6 presents the initial labelled image and the MRF segmentation result after five iterations. Firstly, the initial labelled image is generated based on the input image, as shown in figure 6 (a). The labels are determined by the value of the foreground and background means. A value of 2 is triggered if the intensity is close to the foreground mean; otherwise, one is triggered. Next, based on the initial labelled image, the likelihood energy and the Gibbs energy are computed. A new labelling image is produced based on the total energy computed. Finally, the process is repeated until the maximum iteration is reached, with the MRF segmentation results are shown in figure 6 (b). The value of 2 representing the foreground while the value of 1 representing the background.

Figure 7 presents the MRF segmentation results for the yellow box. The results are taken from the first to the fifth iterations, as, beyond that, the results show no changes. The results show that each iteration of the MRF segmentation improves the labelling of the image.

5. Conclusion

In this paper, a segmentation process based on the MRF algorithm is used, demonstrating that this approach is suitable for performing segmentation on a DNA microarray image. An evaluation of each

Authors' Final Manuscript.
Published paper link: https://iopscience.iop.org/article/10.1088/1742-6596/2071/1/012032/pdf

pixel that considers its neighbourhood pixels offers improvements in classifying pixels into the foreground and background features. The results show that the MRF algorithm performs well in the segmentation of a DNA microarray image.

Acknowledgment
This research was supported by the Science Fund (03-01-15-SF0229) from the Ministry of Science, Technology and Innovation (MOSTI), Malaysia.

References

[1] G. S. C. Kumar, R. K. Kumar, G. A. Naidu, and J. Harikiran, "Noise removal in microarray images using variational mode decomposition technique," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 4, pp. 1750–1756, 2017.
[2] S. M. Joseph and P. S. Sathidevi, "CDNA Microarray Image Enhancement for Effective Gridding of Spots," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2019-Octob, pp. 326–331, 2019.
[3] M. V. Klimushina, N. G. Gumanova, and V. A. Metelskaya, "Direct labeling of serum proteins by fluorescent dye for antibody microarray," *Biochem. Biophys. Res. Commun.*, vol. 486, no. 3, pp. 824–826, 2017.
[4] L. Roszkowiak and C. Lopez, "PATMA: parser of archival tissue microarray," *PeerJ*, vol. 4, p. e2741, 2016.
[5] H. Saberkari, M. Shamsi, and H. B. Ghavifekr, "A shape-independent algorithm for fully-automated gridding of cDNA microarray images," *Comput. Electr. Eng.*, vol. 62, pp. 135–150, 2017.
[6] G. Shao, D. Li, J. Zhang, J. Yang, and Y. Shangguan, "Automatic microarray image segmentation with clustering-based algorithms," *PLoS One*, vol. 14, no. 1, pp. 1–22, 2019.
[7] A. Mohandas, S. M. Joseph, and P. S. Sathidevi, "An Autoencoder based Technique for DNA Microarray Image Denoising," *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 1366–1371, 2020.
[8] M. M. Saeid, Z. B. Nossair, and M. A. Saleh, "A Fully Automated Spot Detection Approach for cDNA Microarray Images using Adaptive Thresholds and Multi-resolution Analysis," *IEEE Access*, vol. 7, pp. 1–1, 2019.
[9] C. Wang, N. Komodakis, and N. Paragios, "Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey," *Comput. Vis. Image Underst.*, vol. 117, no. 11, pp. 1610–1627, 2013.
[10] Z. Zhang, D. Han, Y. Yang, Zhe Zhang, Deqiang Han, and Yi Yang, "Image segmentation based on evidential Markov random field model," *2015 Int. Conf. Control. Autom. Inf. Sci.*, pp. 239–244, 2015.
[11] S. Robinson, L. Guyon, J. Nevalainen, M. Toriseva, M. Nees, and M. Åkerfelt, "Segmentation of image data from complex organotypic 3D models of cancer tissues with markov random fields," *PLoS One*, vol. 10, no. 12, pp. 1–26, 2015.
[12] T. Hamedani, R. Zarei, and A. Harati, "Superpixel based RGB-D image segmentation using Markov random field," in *2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP)*, 2015, pp. 89–94.
[13] O. Oztimur Karadag and F. T. Yarman Vural, "Image segmentation by fusion of low level and domain specific information via Markov Random Fields," *Pattern Recognit. Lett.*, vol. 46, pp. 75–82, 2014.
[14] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, vol. 26, no. 4. London: Springer London, 2009.
[15] O. Demirkaya, M. H. Asyali, and P. K. Sahoo, *Image Processing with MATLAB: Applications*

in Medicine and Biology. CRC Press, 2008.
[16] Microarrays Inc., "Expert Laboratory Services," 2016. [Online]. Available: http://www.microarrays.com/services.php. [Accessed: 19-Feb-2016].

Authors' Final Manuscript.
Published paper link: https://iopscience.iop.org/article/10.1088/1742-6596/2071/1/012032/pdf