Egyptian Informatics Journal 20 (2019) 173–178

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

Microarray spot segmentation algorithm based on integro-differential operator

R.M. Farouk [a], M.A. SayedElahl [b,*]

[a] Department of Mathematics, Faculty of Science, Zagazig University, P. O. Box 44519, Zagazig, Egypt
[b] Alrayan-Collages, P.O.Box 41411-167, Al Madinah Al Munawwarah, Kingdom of Saudi Arabia

ARTICLE INFO

Article history:
Received 27 March 2018
Revised 2 December 2018
Accepted 2 April 2019
Available online 6 April 2019

Keywords:
Integro-differential operator
cDNA microarray spot image analysis
cDNA microarray spot image segmentation

ABSTRACT

Microarray technology has increased exponentially in the past few years. Such technology encompasses various applications related to drug design, tumor detection, clinical diagnosis and treatment, and environmental health research. However, the segmentation of the spots in microarray image is still a problem due to variations of spots qualities such as spots shapes and sizes for example. Hence, we have introduced a new microarray spot segmentation algorithm based on Integro-Differential Operator (IDO). Because it includes a spot detection step, this IDO can be used to find any spot regardless of its size and shape within the microarray image. The IDO algorithm is applied on the cDNA microarray images to improve the accuracy and the efficiency of the spots segmentation process. As the IDO is able to segment each spot based on the fact that the illumination difference between the inside and the outside of the pixels in spot edge circle is maximum. We have also tested our algorithm on the Stanford Microarray Database (SMD). The numerical results show that, the IDO is a powerful technique to improve the overall spots segmentation on the microarray image and to also give better results in spot segmentation and analysis process.

1. Introduction

DNA segmentation refers to the process of automatically detecting the spots (genes) in a given microarray image. This process helps in extracting features from the discriminative texture of the cDNA, while excluding the surrounding regions. A microarray image showing genes (spots) is seen in Fig. 1.

Spot segmentation plays a key role in the performance of the gene recognition system. This is because improper segmentation can lead to incorrect feature extraction from less discriminative regions, thereby reducing the recognition performance [1].

cDNA microarray image analysis pass through three major steps:

Step one: cDNA image noise filtering

This is the process of removing noise from cDNA microarray image without affecting the genes data. A fast noise detection method without resorting to pairwise distance computations between each pixel is discussed in [2].

Step two: Spot addressing and localization

This is a separation process of the spots into a distinct cells [3], which is based on calculating the vertical and horizontal image intensity profile, as discussed in [4–6].

Step three: Microarray image segmentation process

The process splits the image pixels into foreground and background pixels [7]. Microarray image segmentation techniques can be categorized into four categories;

1.1. Fixed circle segmentation

This method assumes that all the spots have perfect circle shape and same size [8]. Some cDNA microarrays segmentation software have been developed based on the fixed circle segmentation method, like ScanAlyze developed by Eisen in 1999 [9], and GenePix developed by Axon Instruments Inc. in 1999 [10].

* Corresponding author at: Alrayan-Collages, P.O.Box 41411-167, Al Madinah Al Munawwarah, Kingdom of Saudi Arabia. Tel.: +966556442768.
E-mail address: ma0777@yahoo.com (M.A. SayedElahl).
Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.eij.2019.04.001
1110-8665/

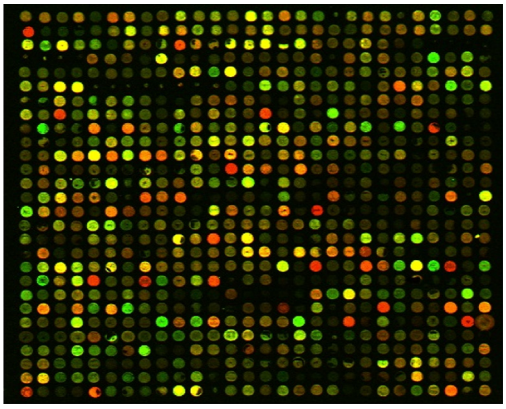

Fig. 1. cDNA microarray image.

1.2. Adaptive shape segmentation

This method assumes that the spots have a circular shape but also allows for adjusting the size of each spot [11]. Adaptive shape segmentation methods, like watershed [12] and seed region growing [13], is based on selecting start points and enlarging the detected spot area step by step until some criterion is reached. Seed region growing has been implemented in the spot software [14].

1.3. Histogram segmentation

This segmentation is used by computing a threshold value using -Whitney test, where Mann-Whitney use a circular target mask to cover all the foreground pixels, and compute a threshold [15]. The pixels with intensity lower than the threshold value is considered as background and the pixels with intensity value higher than the threshold value is considered as foreground [16].

1.4. Clustering

The main idea behind the clustering algorithm is to group the pixels into clusters [17]. Clustering methods have been employed for segmenting the microarray images such as K-means [18] and Fuzzy C-means [19].

In this paper we have used IDO to separate the cDNA microarray spots from its background. Segmentation by IDO is fast, reliable and extensively employed in image processing and analysis [21]. IDO is included in adaptive shape segmentation and it is a powerful technique where it combines the frontier approach and the region approach, IDO makes a fast detection of both edges and regions [22]. The paper organized as follows. In the second section, the proposed techniques are explained in detail. Experimental results obtained after applying the proposed method on real images from the Stanford Microarray Database (SMD) [23] are discussed in the third section. Finally, the conclusion is analyzed in the fourth section.

2. The proposed techniques

The previous HCT spot segmentation approaches assume that the boundary of gene is a circle. However, according to our observation, circle cannot model this boundary accurately. To improve the quality of segmentation, a novel IDO is proposed to detect the irregular boundary of spot. The method can successfully detect all the spot boundaries in the SMD database [23] and increase the recognition accuracy. This new algorithm based on applying Integro-Differential Operator on cDNA microarray spots images for spots segmentation. The input cDNA microarray image comes across several steps before the proposed technique is performed as shown in the flowchart at Fig. 2.

It typically involves the following steps:

1. Read the cDNA microarray image and crop a spot region of 500×500 pixels.
2. Spot addressing and localization based on the calculation of vertical and horizontal image intensity profile [6].
3. Each spot is extracted as individual image.
4. Integro-differential operator [20] is applied on each spot image for spot segmentation.
5. Compare the segmentation result of Integro-Differential Operator with Hough circle transformation (HCT) method [11] and measure the gene expression level.

2.1. Dataset

More than seven hundred of real microarray images from the Stanford Microarray Database (SMD) [23] are employed to illustrate the analytic power of our proposed technique. These images are cropped to 500×500 pixels spot region as shown in Fig. 3.

2.2. Spot addressing and extraction

Spot addressing and extraction is one of the most challenge process in gene analysis process. This is a separation process of the spots into a distinct cells as shown in Fig. 6 [6].

2.2.1. Spot addressing algorithm

Consider an image $A = \{a_{ij} | i = 1, 2, \ldots, n | j = 1, 2, \ldots, m\}$

The horizontal and vertical intensity profile calculated as follows:

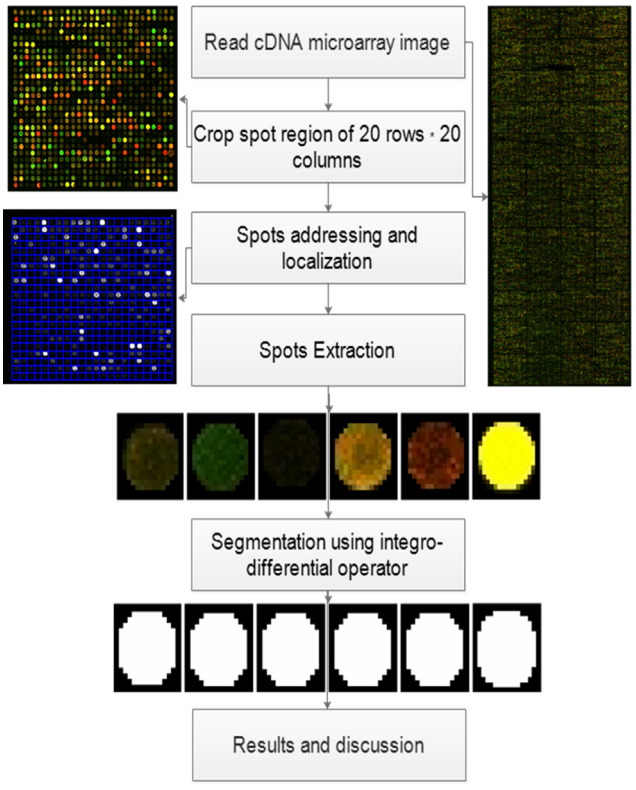

Fig. 2. The flowchart of the proposed technique.

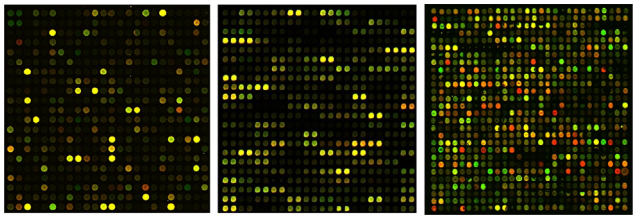

Fig. 3. Examples of the cropped cDNA microarray images.

$$H_i = \sum_{j=1}^{m} a_{ij} \quad i = 1, 2, \ldots, n \tag{1}$$

$$H_j = \sum_{i=1}^{n} a_{ij} \quad j = 1, 2, \ldots, m \tag{2}$$

Use equation 1 and 2 to get the horizontal and the vertical grid as shown in Fig. 4.

Then calculate the columns width CW and Rows width RW by counting the number of zeros in H_i and H_j respectively.

Then draw the gridding map, which separate the spots into distinct cells as shown in Fig. 5.

Finally, extract each spot as a separate image as shown in Fig. 6.

2.3. Integro-Differential Operator

Integro-Differential Operator (IDO) is included in adaptive shape segmentation and it is a powerful technique where it combines the frontier approach and the region approach, it makes a fast detection of both edges and regions. IDO is based on the fact that the illumination difference between inside and outside of pixels in spot edge is maximum. In other words, if you calculate the difference in values of pixels gray level in spot, this value is higher than any other object in the image. This fact turns to color of spot and color of background. Fig. 8 present a result of segmented images using IDO.

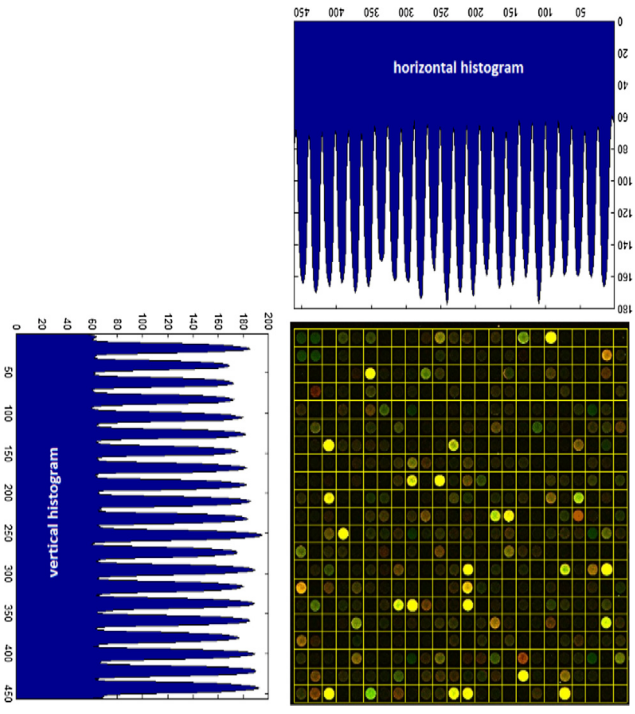

Fig. 4. Steps of gridding.

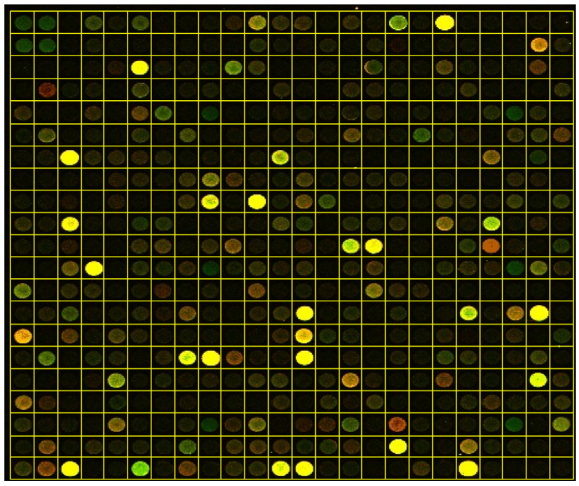

Fig. 5. The result of automatic gridding.

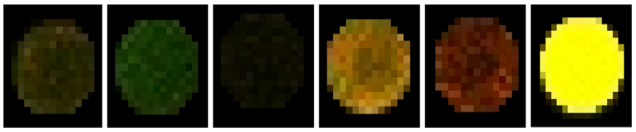

Fig. 6. Examples of the extracted spots as a result of the automated grading process.

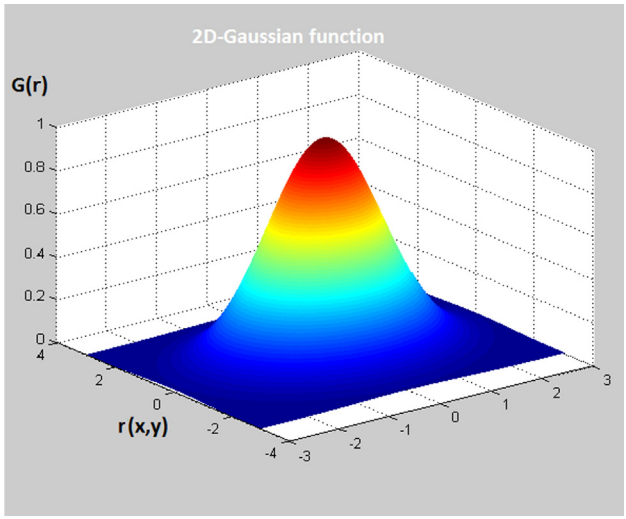

Fig. 7. 2D Gaussian smoothing function.

2.3.1. IDO curve fitting algorithm

First: Let $I(x, y)$ containing a spot image like images in Fig. 6.

Then:

- Search over the image domain (x, y) for the maximum in the blurred partial derivative with respect to increasing radius r of the normalized contour integral of $I(x, y)$ along a circular arc ds of radius r, and center coordinates (x_0, y_0), where the spot radius range from 7 to 11 pixel.
- For locating the inner and the outer boundaries in a spot image use the following optimization:

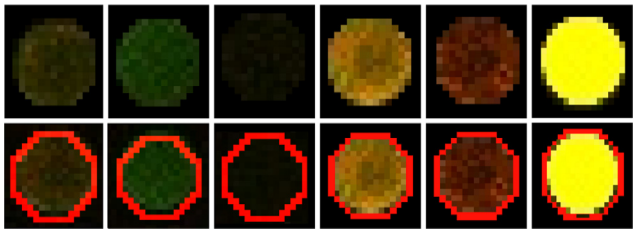

Fig. 8. Example of segmented spots using IDO.

$$\max(r,x_o,y_o)\left|G_\sigma(r)*\frac{\partial}{\partial r}\int_{r,x_0,y_0}\frac{I(x,y)}{2\pi r}ds\right| \tag{3}$$

where the symbol $*$ denotes convolution and $G_\sigma(r)$ is Gaussian smoothing function with sigma $\sigma = 0.5$.

$$G_\sigma(r) = \frac{1}{2\pi\sigma^2}e^{-\frac{r^2}{2\sigma^2}} \tag{4}$$

2.4. Hough circle transformation (HCT)

Hough transform is recognized as a robust curve detection technique [11]. Hough transformation aims to find the circular patterns within an image, even noise existence, as shown in Fig. 9.

2.4.1. HCT algorithm

Consider a point (x_i, y_i) in the image. The general equation of a circle is:

$$(x_i-a)^2+(y_i-b)^2-r^2=0 \tag{5}$$

where a and b are the coordinates of the center and r is the radius of the circle.

The parametric equations for a circle in polar coordinates are:

$$x_i = a + r\cos\theta \tag{6}$$

$$y_i = b + r\sin\theta \tag{7}$$

Where θ is the gradient angle.

Solving for the parameters of the circle we obtain the equations:

$$a = x_i - r\cos\theta \tag{8}$$

$$b = y_i - r\sin\theta \tag{9}$$

We can eliminate the radius from the pair of equations above to yield

$$b = a\tan\theta - x_i\tan\theta + y_i \tag{10}$$

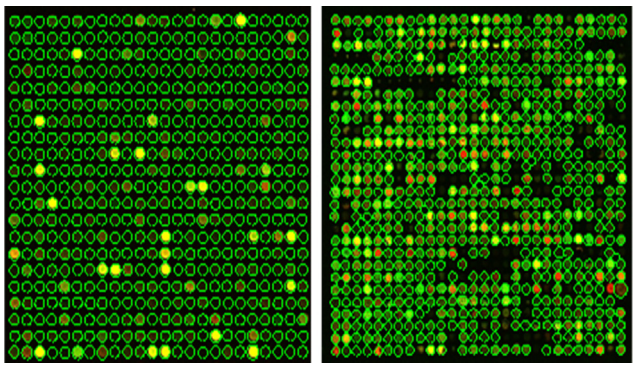

Fig. 9. The result of applying CHT on the cDNA microarray image.

For the edge point in the image, increment the parameters (a, b) along the equation 10 then, get the local maxima of the results which represent the center of the circle.

Fig. 9 shows a random picked microarray image after applying CHT algorithm, in which each spot has a green circle around it, after that, each spot can be extracted as individual image as shown in Fig. 10.

3. Experimental results

Image Processing Toolbox and Matlab 2012 [24] is used to evaluate our new approach. Real twelve cDNA microarray images from SMD are employed to evaluate the efficiency and accuracy of the proposed technique. This real image is cropped to the 200∗200 pixel region.

Fig. 3 shows examples of the cropped spot regions, and then the automated gridding algorithm [6] is applied to extract each spot in the cropped image as individual image, as in Fig. 6. This algorithm is based on calculating the horizontal and vertical intensity profile using Eqs (1) and (2) respectively to get the rows and columns width and draw the grid map as shown in Fig. 5. After that IDO algorithm is applied in order to get a circle around the spot in the spot image, as in Fig. 8. The pixels inside the circle are clustered as a signal and the ones outside the circle as background as in Fig. 11. In IDO algorithm we use the spot radius range from 7 to 11 pixel to get the radius r and the spot center point (x_0, y_0), then apply the Eq. (3) for locating the inner and the outer boundaries and use 2D Gaussian function, presented in Eq. (4) and Fig. 7 with scale sigma $\sigma = 0.5$ as a smoothing function. Finally compare the segmentation result of the proposed technique with other techniques like GenePix® [10] and HCT [11] using quality index value Q-index and gene expression level GL.

4. The quality index (Q-index)

The Quality index (Q-index) [25] is used to evaluate the quality of the segmentation results of the proposed method (IDO) and HCT via the results of GenePix® as presented in the Figs. 12A–12C.

Let $o = \{o_i | i = 1, 2, 3, \ldots, N\}$ and $t = \{t_i | i = 1, 2, 3, \ldots, N\}$ be the original and the test image, respectively.

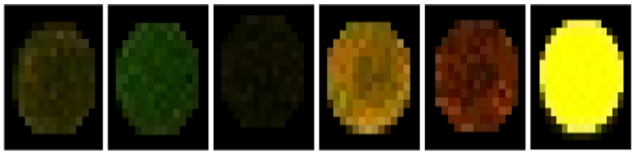

Fig. 10. Example of the extracted spots.

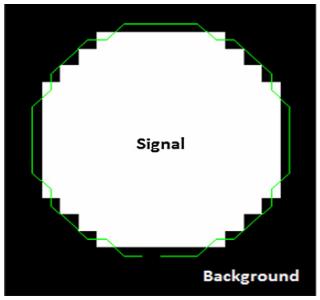

Fig. 11. Clustered spot image.

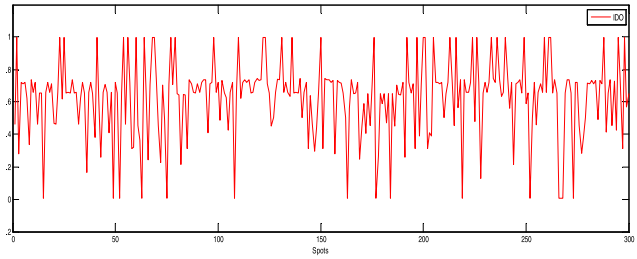

Fig. 12A. Q-index using IDO via GenePix®.

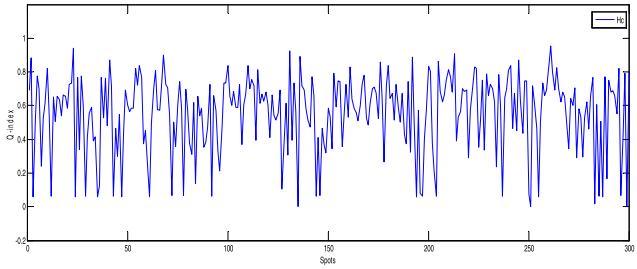

Fig. 12B. Q-index using HCT via GenePix®.

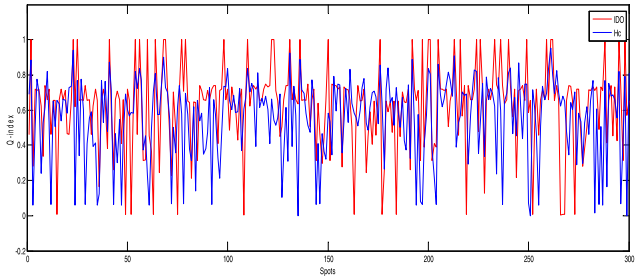

Fig. 12C. Q-index using IDO and HCT via GenePix®.

$$Q-index=\frac{4\sigma_{ot}\bar{O}\bar{T}}{(\sigma_o^2+\sigma_t^2)(\bar{O}^2+\bar{T}^2)} \tag{11}$$

where

$$\bar{O}=\frac{1}{N}\sum_{i=1}^{N}O_i \tag{12}$$

$$\bar{T}=\frac{1}{N}\sum_{i=1}^{N}T_i \tag{13}$$

$$\sigma_o^2=\frac{1}{N-1}\sum_{i=1}^{N}(O-O_i)^2 \tag{14}$$

$$\sigma_t^2=\frac{1}{N-1}\sum_{i=1}^{N}(T-T_i)^2 \tag{15}$$

$$\sigma_{ot}=\frac{1}{N-1}\sum_{i=1}^{N}(O-O_i)(T-T_i) \tag{16}$$

From the numerical results of Q-index for IDO and HCT via GenePix® in Table 1, we found the Q-index of IDO variant from 0.0049 to 1 with a mean value of 0.6292. This is better than the values of the Q-index of HCT which variant from 2.9055e to 04 to 0.9545 with a mean value of 0.5616.

Table 1
The numerical results of Q-index for IDO and HCT via GenePix®.

	Min(Q-index)	Max(Q-index)	Mean(Q-index)
IDO	0.0049	1	0.6292
HCT	2.9055e−04	0.9545	0.5616

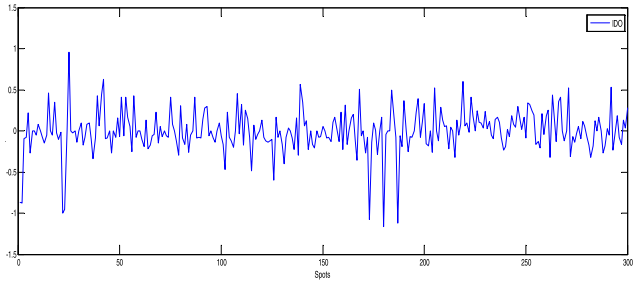

Fig. 13A. Gene expression level using IDO.

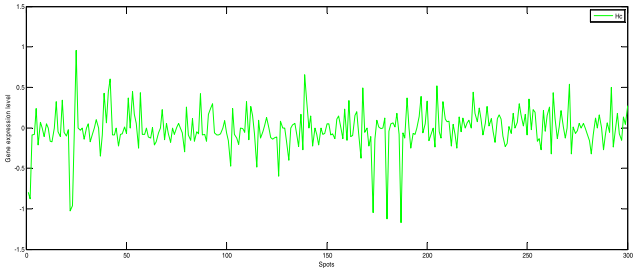

Fig. 13B. Gene expression level using HCT.

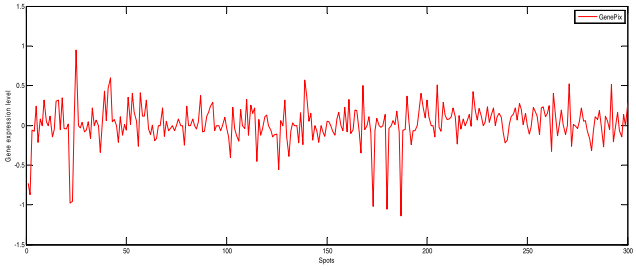

Fig. 13C. Gene expression level using GenePix®.

Table 2
Numerical results of GL for IDO, HCT and GenePix®.

	Min(GL)	Max(GL)	Mean(GL)
IDO	−1.1476	0.9531	0.0141
HCT	−1.1676	0.9531	0.0122
GenePix®	−1.1421	0.9531	0.0145

Also the visual results presented in the Figs. 12A–12C show that the IDO has the best result comparing with the HCT method.

4.1. Gene expression level Gl

Gene expression level *Gl* is measured for each spot (gene) as in Figs. 13A–13C [25].

$$Gl=\log\frac{\text{Red layer intensity}}{Green\,\text{layer intensity}} \tag{17}$$

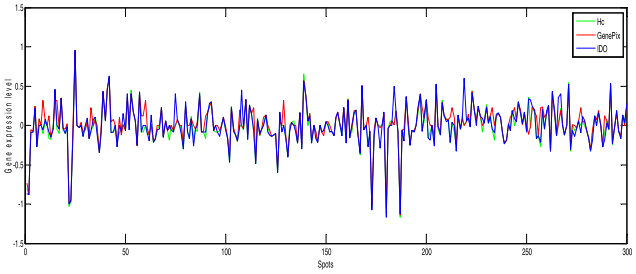

Fig. 13D. Gene expression level using IDO, HCT and GenePix®.

The numerical results of gene expression level in Table 2 and visual results in Figs. 13A–13D show that the IDO and GenePix® have a convergent result.

5. Conclusion

This paper presents an enhanced spot segmentation approach using Integro- Differential Operator, which includes a spot detection step. IDO used to find spots regardless to its size and shape within the microarray image. It's able to segment each spot based on the fact that the illumination difference between inside and outside of pixels in spot edge circle is maximum. The IDO algorithm is applied on the cDNA microarray images to improve the accuracy and the efficiency of the spots segmentation process. We have tested our new algorithm on the Stanford Microarray Database (SMD). This new method was compared to other methods like HCT and GenePix®. The numerical and visual results show that, the proposed technique is an effective in spot addressing, localization and segmentation.

References

[1] Farouk RM, Badr EM, Sayed Elahl MA. Recognition of cdna microarray image using feedforward artificial neural network. Int J Artificial Intell Appl (IJAIA) 2014;5(5).
[2] Emre Celebi M, Kingravi1 Hassan A, Uddin Bakhtiyar, Alp Aslandogan Y. A fast switching filter for impulsive noise removal from color images. J Imag Sci Technol 2007;51(2):155–65.
[3] Giannakeas N, Kalatzis F, Tsipouras MG, Fotiadis DI. Spot addresses for microarray images structured in hexagonal grids. Comput Methods Program Biomed 2012;106(1):1–13.
[4] Deepa J, Thomas T. Automatic gridding of DNA microarray images using optimum subimage. Int J Recent Trends Eng1 (4), 2009..
[5] Rueda L, Rezaeian I. A fully automatic gridding method for cDNA microarray images. BMC Bioinformatics 2011;12–113.
[6] Giannakeas N, Fotiadis DI. An automated method for gridding and clustering-based segmentation of cDNA microarray images. Comput Med Imaging Graph 2009;33:40–9.
[7] Karim R, Mahmud S. A review of image analysis techniques for gene spot identification in cDNA microarray images, international conference of next generation. Inform Technol; 2011.
[8] Li Qunhua, Fraley Chris, Bumgarner Roger E, Yeung Ka Yee, Adrian E. Raftery, robust model-based segmentation of microarray images. Bioinformatics 2005;21:2875–82.
[9] Eisen MB, ScanAlyze, http://rana.lbl.gov/EisenSoftware.htm, QuantArray Analysis Software, Operator's Manual, 2013
[10] Axon Instruments Inc., GenePix 4000, A User's Guide, 1999
[11] Farouk RM, SayedElahl MA. Robust cDNA microarray image segmentation and analysis technique based on Hough circle transform. HFSP J 2015;9(7). ISSN: 1955-2068, , part 1.
[12] Park C, Lee K, Lee S, Automatic Microarray Image Segmentation Based on Watershed Transformation. In 17th Inter-national Conference on Pattern Recognition (ICPR'04), Cambridge, UK. 786–789, 2004.
[13] Deepa J. Automatic segmentation of DNA microarray images using an improved seeded region growing method. Nat Biol Inspired Comput 2009. NaBIC 2009. World Congress on, 2009.
[14] Buckley. Spot User's Guide. Sydney, Australia: CSIRO Mathematical and Information Sciences; 2000.
[15] Chen Y, Dougherty ER, Bittner ML. Ratio based decisions and the quantitative analysis of cDNA microarray images. J Biomed Opt 1997;2(4):364–74.
[16] Chen TB, Lu, HHS, Lan YJ, Segmentation of cDNA microarray images by kernel density estimation, 41 (6), 1021–1027, (2008
[17] Shuanhu W, Yan H, Microarray Image Processing Based on Clustering and Morphological Analysis, The First Asia-Pacific bioinformatics conference on Bioinformatics -Australia, (19), 111–118, 2003
[18] Maguluri LP, Rajapanthula K, Srinivasu PN. A comparative analysis of clustering based segmentation algorithms in microarray images. Int J Emerg Sci Eng (IJESE) 2013;1(5):27–32.
[19] Biju VG, Mythili P, Fuzzy Clustering Algorithms for cDNA Microarray Image Spots Segmentation, International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014.
[20] Tadmor E, Athavale P. Multiscale image representation using integro differential equations. Inverse Prob Imaging 2009;4:693–710.
[21] Radman A, Jumari K, Zainal N. Fast and reliable iris segmentation algorithm. IET Image Process 2013;7(1).
[22] Athavale P, Tadmor E. Integrodifferential equations based on (BV, L1) minimization. SIAM J Imag Sci 2011;4(1):300–12.
[23] Gollub J, Ball CA, Binkley G, Demeter K, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaplper M, Matese JC, Schroeder M, Brown PO, Botstein D, Sherlock G. The Stanford Microarray Database (SMD): data access and quality assessment tools. Nucleic Acids Res 2003;31:94–6.
[24] MathWorks, Image Processing Toolbox: User's Guide, 2012.
[25] Tarca AL, Romero R, Draghici S. Analysis of microarray experiments of gene expression profiling. Am J Obstet Gynecol 2006;195(2):373–88.