

# Scientific Research and Essays

Volume 11 Number 12 30 June 2016  
ISSN 1992-2248



## ABOUT SRE

**The Scientific Research and Essays (SRE)** is published twice monthly (one volume per year) by Academic Journals.

**Scientific Research and Essays (SRE)** is an open access journal with the objective of publishing quality research articles in science, medicine, agriculture and engineering such as Nanotechnology, Climate Change and Global Warming, Air Pollution Management and Electronics etc. All papers published by SRE are blind peer reviewed.

### Contact Us

Editorial Office: [sre@academicjournals.org](mailto:sre@academicjournals.org)

Help Desk: [helpdesk@academicjournals.org](mailto:helpdesk@academicjournals.org)

Website: <http://www.academicjournals.org/journal/SRE>

Submit manuscript online <http://ms.academicjournals.me/>.

## Editors

### **Dr. NJ Tonukari**

*Editor-in-Chief*

*Scientific Research and Essays*

*Academic Journals*

*E-mail:sre.research.journal@gmail.com*

### **Dr. M. Sivakumar Ph.D. (Tech).**

*Associate Professor*

*School of Chemical & Environmental Engineering*

*Faculty of Engineering*

*University of Nottingham*

*JalanBroga, 43500 Semenyih*

*SelangorDarul Ehsan Malaysia.*

### **Prof. N. Mohamed ElSawi Mahmoud**

*Department of Biochemistry, Faculty of science, King*

*Abdul Aziz university,*

*Saudi Arabia.*

### **Prof. Ali Delice**

*Science and Mathematics Education Department, Atatürk*

*Faculty of Education,*

*Marmara University, Turkey.*

### **Prof. Mira Grdisa**

*RudjerBoskovicInstitute, Bijenickacesta54,*

*Croatia.*

### **Prof. Emmanuel HalaKwon-Ndung**

*Nasarawa State University Keffi Nigeria*

*PMB1022 Keffi,*

*Nasarawa State.*

*Nigeria.*

### **Dr. Cyrus Azimi**

*Department of Genetics, Cancer Research Center,*

*CancerInstitute, Tehran University of Medical Sciences,*

*Keshavarz Blvd.,*

*Tehran, Iran.*

### **Dr. Gomez, Nidia Noemi**

*National University of San Luis,*

*Faculty of Chemistry, Biochemistry and Pharmacy,*

*Laboratory of Molecular Biochemistry*

*EjercitodelosAndes950-5700 SanLuis*

*Argentina.*

### **Prof.M.Nageeb Rashed**

*Chemistry Department-Faculty of Science,*

*Aswan South Valley University,*

*Egypt.*

### **Dr. John W. Gichuki**

*KenyaMarine& FisheriesResearchInstitute,*

*Kenya.*

### **Dr. Wong Leong Sing**

*Department of Civil Engineering,*

*College of Engineering,*

*Universiti Tenaga Nasional,*

*Km7, JalanKajang-Puchong,*

*43009Kajang, SelangorDarulEhsan, Malaysia.*

### **Prof. Xianyi Li**

*College of Mathematics and Computational Science*

*Shenzhen University Guangdong,518060*

*P.R.China.*

### **Prof. Mevlut Dogan**

*Kocatepe University, Science Faculty, Physics Dept.*

*Afyon/Turkey.*

*Turkey.*

### **Prof. Kwai-Lin Thong**

*Microbiology Division, Institute of Biological Science,*

*Faculty of Science, University of Malaya, 50603,*

*KualaLumpur,*

*Malaysia.*

### **Prof. Xiacong He**

*Faculty of Mechanical and Electrical Engineering, Kunming*

*University of Science and Technology, 253 XueFu Road,*

*Kunming,*

*P.R.China.*

### **Prof. Sanjay Misra**

*Department of Computer Engineering*

*School of Information and Communication Technology*

*Federal University of Technology, Minna,*

*Nigeria.*

### **Prof. Burtram C. Fielding Pr. Sci. Nat.**

*Department of Medical BioSciences*

*University of the Western Cape Private Bag X17*

*Modderdam Road*

*Bellville,7535,SouthAfrica.*

### **Prof.Naqib Ullah Khan**

*Department of Plant Breeding and Genetics*

*NWFP Agricultural University Peshawar 25130,*

*Pakistan*

## Editorial Board

**Prof. Ahmed M. Soliman**

*20 Mansour Mohamed St., Apt 51, Zamalek,  
Cairo,  
Egypt.*

**Prof. Juan José Kasper Zubillaga**

*Av. Universidad 1953 Ed. 13 depto 304,  
México D.F. 04340,  
México.*

**Prof. Chau Kwok-wing**

*University of Queensland Instituto  
Mexicanodel Petroleo, Eje Central  
Lazaro Cardenas Mexico D.F.,  
Mexico.*

**Prof. Raj Senani**

*Netaji Subhas Institute of Technology,  
Azad Hind Fauj Marg, Sector 3,  
Dwarka, New Delhi 110075, India.*

**Prof. Robin J Law**

*Cefas Burnham Laboratory,  
Remembrance Avenue Burnhamon Crouch, Essex  
CM08HA,  
UK.*

**Prof. V. Sundarapandian**

*Indian Institute of Information Technology and  
Management-Kerala  
Park Centre,  
Technopark Campus, Kariavattom P.O.,  
Thiruvananthapuram-695581, Kerala, India.*

**Prof. Tzung-Pei Hong**

*Department of Electrical Engineering,  
And at the Department of Computer Science and  
Information Engineering  
National University of Kaohsiung.*

**Prof. Zulfiqar Ahmed**

*Department of Earth Sciences, box 5070,  
Kfupm, dhahran-31261, Saudi Arabia.*

**Prof. Khalifa Saif Al-Jabri**

*Department of Civil and Architectural Engineering  
College of Engineering, Sultan  
Qaboos University  
P.O. Box 33, Al-Khod 123, Muscat.*

**Prof. V. Sundarapandian**

*Indian Institute of Information Technology &  
Management-Kerala  
Park Centre,  
Technopark, Kariavattom P.O.  
Thiruvananthapuram-  
695581, Kerala India.*

**Prof. Thangavelu Perianan**

*Department of Mathematics,  
Aditanar College,  
Tiruchendur-628216 India.*

**Prof. Yan-ze Peng**

*Department of Mathematics,  
Huazhong University of Science and  
Technology, Wuhan 430074, P.R.  
China.*

**Prof. Konstantinos D. Karamanos**

*Université Libre de Bruxelles,  
CP231 Centre of Nonlinear  
Phenomena And Complex Systems,  
CENOLIB Boulevard de Triomphe  
B-1050,  
Brussels,  
Belgium.*

**Prof. Xianyi Li**

*School of Mathematics and Physics,  
Nanhua University, Hengyang City,  
Hunan Province,  
P.R. China.*

**Dr. K.W. Chau**

*Hong Kong Polytechnic University  
Department of Civil & Structural  
Engineering, Hong Kong Polytechnic  
University, Hung Hom, Kowloon,  
Hong Kong,  
China.*

**Dr. Amadou Gaye**

*LPAO-SF/ESPPo Box 5085 Dakar-Fann SENEGAL  
University Cheikh Anta Diop  
Dakar SENEGAL.*

**Prof. Masno Ginting**

*P2F-LIPI, Puspiptek-Serpong,  
15310 Indonesian Institute of Sciences,  
Banten-Indonesia.*

**Dr. Ezekiel Olukayode Idowu**

*Department of Agricultural Economics,  
Obafemi Awolowo University, Ife-Ife,  
Nigeria.*

**ARTICLE**

<b>New methodology for microarray spot segmentation and gene expression analysis</b>	<b>126</b>
Amjad A. Hudaib, Hussam N. Fakhouri and Rawan Ghnemat	

*Full Length Research Paper*

## New methodology for microarray spot segmentation and gene expression analysis

Amjad A. Hudaib<sup>1</sup>, Hussam N. Fakhouri<sup>2\*</sup> and Rawan Ghnemat<sup>3</sup>

<sup>1</sup>Computer Information Systems Department, King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan.

<sup>2</sup>Computer Science Department, King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan.

<sup>3</sup>Computer Science Department, King Hussein Faculty of Computing Sciences, Princess Sumaya University For Technology, Jordan.

Received 30 December, 2015; Accepted 17 June, 2016

**DNA microarray analysis is the main core in genome mapping. Each microarray image contains millions of information about genes. Microarray analysis is considered one of the most recent and important technologies in exploring the genome. One of the key steps in microarray analysis is to extract gene information from the gene spots, these information represent gene expression levels in the microarray. This paper proposes a new methodology to improve microarray spot analysis based on spot extracted segments. It concentrates on each spot segment area independently rather than analyzing all the spots area together of the microarray image. This paper provides a formal model to enhance the intensity values obtained from gene expression levels of the microarray at any intensity expressed level. It also this paper presents the adaptive threshold techniques to be used for microarray segmentation. The experimental results show that the mean of the gene expression intensity value was 87.77.**

**Key words:** Microarray images, microarray analysis, image segmentation, gridding, microarray addressing, spot localization, spot extraction.

### INTRODUCTION

Recent technology combines various biological, medical and engineering analyses to the field of genome mapping (Lukac and Plataniotis, 2006). Microarray analysis is powerful and efficient technology for measuring gene expression level of thousands of genes simultaneously; microarray technology allows the comprehensive measurement of the expression level genes (Uslan and

Bucak, 2010; Schena et al., 1999). Typical applications of microarrays include the quantification of RNA expression profiles for a system under different experimental conditions (Angulo and Serra, 2003). Several researches were conducted, numerous algorithms have been proposed and software tools have been developed but they were not fully automated and require human

\*Corresponding author. E-mail: [h.fakhouri@ju.edu.jo](mailto:h.fakhouri@ju.edu.jo).

interaction. Almost all the software systems tested require human intervention (Moffitt et al., 2011). It requires the user to specify the geometry of the array, such as the number of grids, number of rows and columns, etc. For example: SPOT from UCSF (Emmanouil et al., 2009), IMAGEGENE from Bio Discover (Rueda and Vidyadharan, 2006) and DAPPLE from University of Washington (Buhler et al., 2000).

The first step in microarray image analysis is addressing the spots locations the second step is spot segmentation (Tiwari, 2005; Wang et al., 2007; Efron et al., 2001) thirdly estimate the intensities of the gene spot (Newton et al., 2001; Yang et al., 2002).

There are five main issues that should be considered while analyzing microarray images that include gridding the microarray, determining the location of spots in the microarray image, removing noise and unwanted particles, removing background and finally image enhancement.

The main aim for microarray image analysis is to determine the percent of gene in every spot of the microarray and this will be done by determining the percent of red or green color's in a specified spot. The process of analyzing each spot basically depends on the segmentation method used to extract the spot. The current processes of extracting the spot and perform further analysis depend on analyzing and enhancing the microarray image global values for all the spots.

This paper proposes a new methodology to improve microarray spot analysis based on spot extracted segments area taking into account the segment background and noise. It concentrates on each spot segment area independently rather than analyzing all the spots area together of the microarray image. The methodology involves noise removal, background subtraction, percent of color calculation and gene profiling. This paper provides a formal model to enhance the intensity values obtained from gene expression levels of the microarray at any intensity expressed level.

The remainder of the paper is organized as follows. First is a discussion of the related work, followed by data presentation. Furthermore, the proposed algorithm and its application to the microarray images was presented and finally, conclusions are summarized.

## LITERATURE REVIEW

Segmentation of microarray analysis is classified into four types: fixed circle segmentation, adaptive circle segmentation, histogram segmentation and adaptive shape segmentation (Zacharia and Maroulis, 2008; Antonio and Ceccarelli, 2004; Giannakeas and Fotiadis, 2008; Bariamis et al., 2010). The fixed circle segmentation algorithm assumes that the spot has a perfect circle shape and all spots have the same size (Zacharia and Maroulis, 2008). The adaptive circle segmentation

algorithm assumes that the spot has a circular shape and permit adjusting the size of each spot (Antonio and Ceccarelli, 2004; Stefano and Luo, 2004). It provides more accurate results than the fixed circle algorithm (Zacharia and Maroulis, 2008; Wu and Yan, 2003).

The histogram segmentation (Giannakeas and Fotiadis, 2008) uses a clustering algorithm to partition the pixels based on their intensity values.

The adaptive shape segmentation segments a spots by its shape either by the seeded region growing (SRG) (Bariamis et al., 2010) or the globally optimal geodesic active contours (GOGAC) (Alhadidi et al., 2006). The spot is adaptive in size and can be of irregular shape. (Jain et al., 2003; Lee et al., 2000).

Alhadidi et al., 2006 developed a new algorithm for determining grid in the microarray images based on spots position in the microarray image Alhadidi et al., 2006.

Clustering is also used in microarray image segmentation (Saal et al., 2002). Clustering have shown some advantages when applied to microarray image segmentation such as reducing computational time, and producing complete segmentation despite the false edges of the spots (Saal et al., 2002; Mangalam 2001; Kamberova and Shah, 2002). However, clustering produces irregular shaped spots which makes clustering based algorithms produce noisy pixels in the foreground regions and incorrect quantization measures for spots intensity (Bariamis et al., 2010; Meher, et al., 2011).

The segmentation algorithms extract gene spots based on either the whole background of the microarray image or the on the locally segmentation of microarray images in each sub-image. However, microarray images consist mostly of low-intensity features that are not well distinguished from the background, which leads to errors that propagate to all stages of the microarray analysis (Zacharia and Maroulis, 2008; Adwan et al., 2013; Fuyong, 2013). In this paper, we propose an algorithm that divides the microarray image into small images and extract gene spot using adaptive threshold based on the background of each small image.

## PROPOSED METHODOLOGY

The proposed methodology segments the gene spots from the spot area using adaptive threshold. The methodology removes the noise from the spot area and calculates the spot intensity for each spot individually in order to reduce the error that result from processing the whole microarray background.

Figure 1 shows the steps of the proposed methodology where the input for this algorithm is the microarray image, and the target output is the gene intensity expression level. The methodology steps are subsequently expressed.

### Crop microarray image

The microarray image represents the whole genome mapping. It consists of thousands of sub-microarrays. The microarray image is

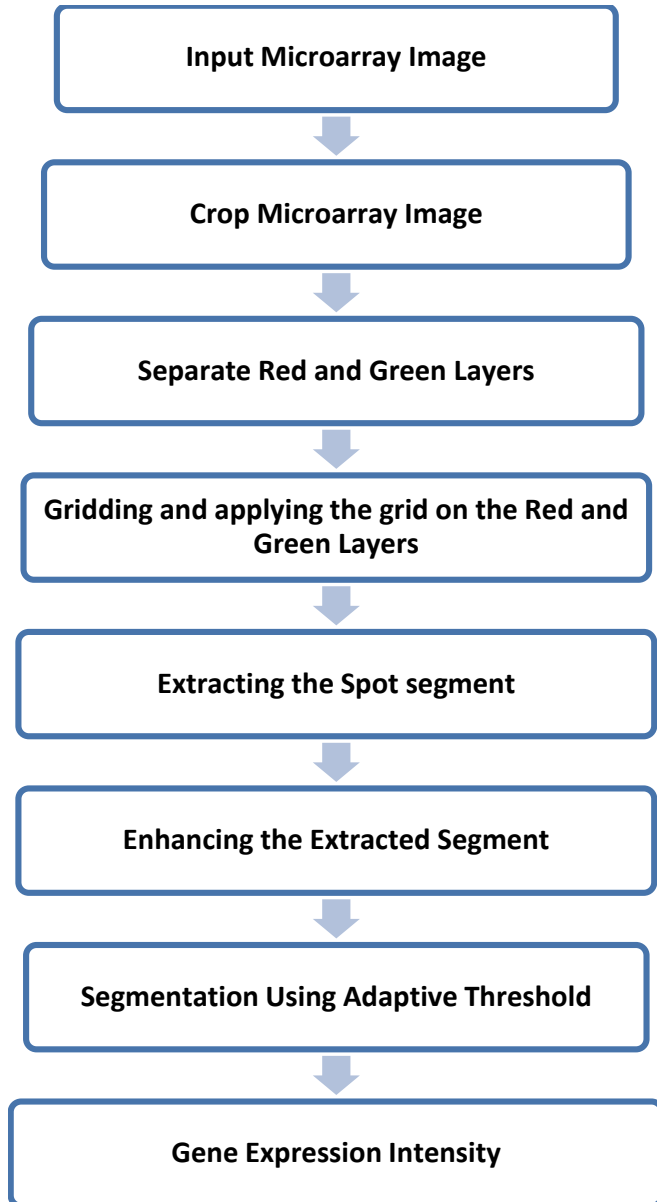


Figure 1. Proposed methodology steps.

cropped into sub-microarray images. Each sub-microarray image will be used to extract each spot area separately.

The microarray is cropped using crop image tool in MATLAB software by determining the positing of the spot this step is done manually. Figure 2 shows microarray image and cropped sub-microarray image.

#### Separate red and green layers

The microarray is stained with two dyes: Cy5 (red) and Cy3 (green) (Newton et al., 2001), usually Cy3 for the control and Cy5 for the experimental channel (Newton et al., 2001; Alhadidi et al., 2006). The microarray are scanned at 540 nm (green) for the control (Cy3) and 630 nm (red) for the experimental channel (Cy5). The microarray image is produced by scanning the microarray

monochromatic images, which are registered into two channels red and green. The red and green layers are produced by Cy5 and Cy3 dyes respectively with a zero blue component (Antoniol and Ceccarelli, 2004; Kim et al., 2001) The output of this process is sub-microarray image with two colors red and green.

The separation of the red and green layers from the sub-microarray image is done using the Matlab red Map and green Map tools. Figure 3 shows the red and green layers of a sub-microarray image.

#### Gridding and applying the grid on red and green layers

This step determine the spot position in the sub-microarray image using grid lines. The grid lines are the lines that are drawn at the center point between any two adjacent spot rows or columns. The intersection of these lines results a square that surrounds each spot.

To apply gridding, first we calculate the intensity of the spot row and spot columns, where the intensity represents the value of red and green colors in each pixel. Then the autocorrelation function is applied to enhance the self-similarity of the vertical and horizontal intensity averages of rows and columns. After that, we calculate the center of peak for each spot row and spot column. The lines of the grid will be drawn at the center of each peak. Figure 4 shows grid lines on red and green layers.

#### Extracting the spot segment

This step extracts spot segments according to the spot location in the grid. The result of gridlines (row and columns) intersection is used to determine the spot segments coordinates. Every segment has four coordinates:  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$ ,  $(x_4, y_4)$ , where  $(x, y)$  values are relevant to gridline intersection points (Alhadidi et al., 2006; Fuyong, 2013). Each extracted segment contains only one spot. Every spot intensity value represents the gene expression level for the gene that the spot represent (Figure 5).

#### Enhancing the extracted segment

Enhancing the extracted segment is a key step before calculating the gene expression level of the microarray image. This enhancement involves three steps: (i) applying logarithmic transformation, (ii) enhancing the contrast, and (iii) image sharpening.

#### Applying logarithmic transformation (LG)

Since microarray spots are different in shape and have large variation in brightness, the logarithmic transformation is applied using Equation 1 (Jiao and Sun, 2010). This step is aims to equalize large variations in spots color magnitude, and leads to increase the visibility of low brightness spots.

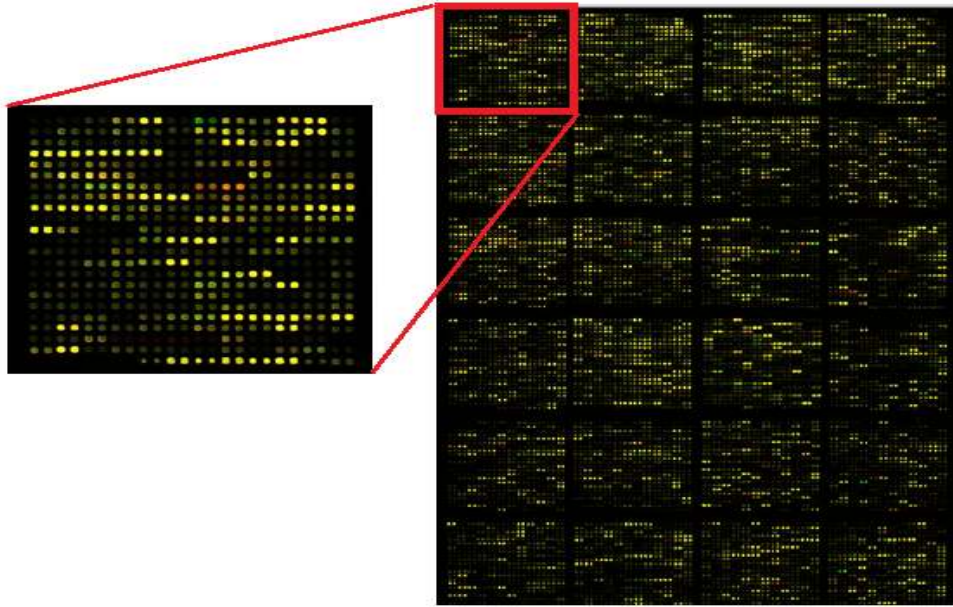
$$LG = c * \log(1+f) \dots\dots (1)$$

Where  $c$  is a constant value,  $f$  is the value of the image pixels, the greater the value of  $c$  the better the intensity values appearance.

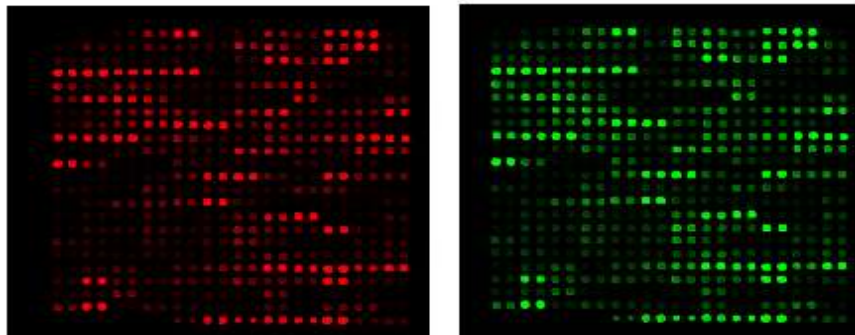
#### Enhancing the contrast

The poor absorption by some microarray genes causes slow spot intensity of red and green levels (Antoniol and Ceccarelli, 2004;

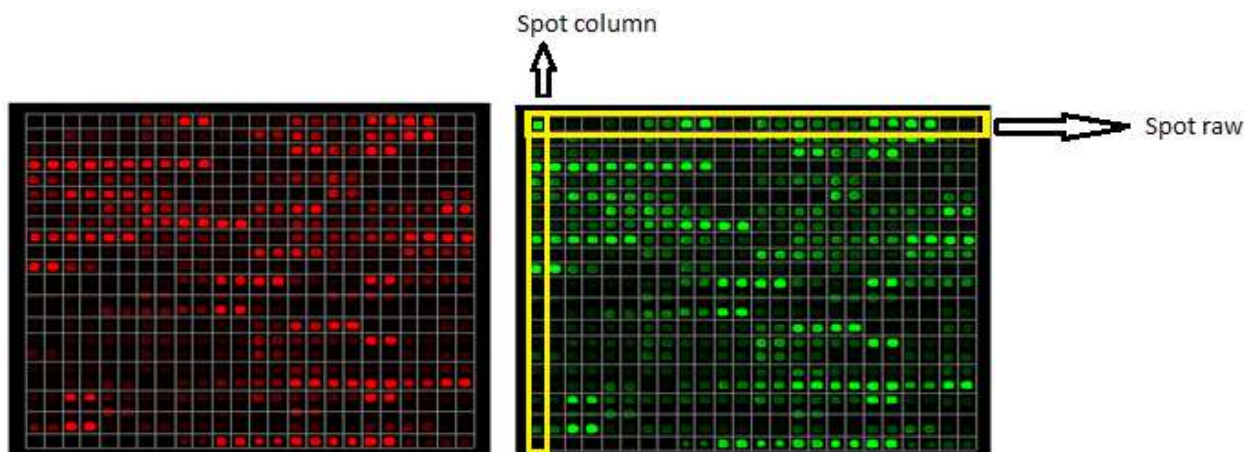




**Figure 2.** Sub-microarray image cropped from microarray image.



**Figure 3.** Red and green layers of a sub-microarray image.



**Figure 4.** Grid lines on red and green layers.

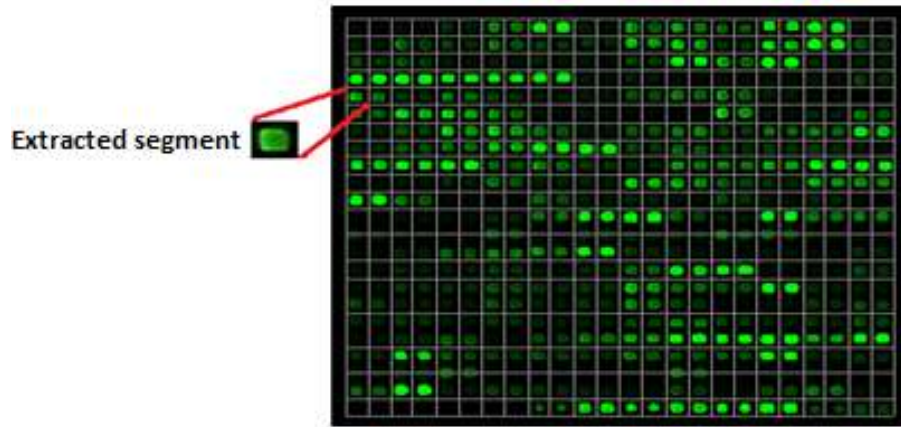


Figure 5. Extracting the spot segment based on spot location in the grid.

Giannakeas and Fotiadis, 2008). Therefore, we need to enhance the contrast of these spots to resolve the low spot intensity. The nonlinear contrast stretching as shown in Equation 2 (Kim et al., 2001) is used to solve this problem.

$$p_{in}(r_k) = \frac{n_k}{n} \quad 0 \leq r_k \leq 1 \quad 0 \leq k \leq L-1 \quad (2)$$

Where  $L$  is the total number of gray levels in the microarray image.  $N_k$ : Number of pixels with gray value  $r_k$ ,  $n$ : Total number of pixels in the spot image.  $P_{in}$ : denote the probability density of the Gray values in the spot image.

**Image sharpening**

Microarray images are taken either by digital scanner or digital camera which produces soft images without sharpen edges. The sharpen edges are needed for the segmentation step. The shock filter is applied to sharpen these edges (Jiao and Sun, 2010). The shock filter enhances the image and produces a sharp discontinuity called shock at the borderline between the spots and the background.

**Segmentation using adaptive threshold**

To calculate the gene expression level for spots, first we need to segment the spots from extracted spot segments. To achieve this segmentation the constant mean adaptive threshold of the extracted spot segment intensity is applied by calculating the average intensity ( $M$ ) of the foreground and the background using Equation 3.

$$M = (IF + IB) / 2 \quad (3)$$

Where,  $IF$  is the intensity values of the foreground, and  $IB$  is the intensity values of the background.

Due to the variation in the spot intensities, the average intensity ( $M$ ) is not suitable to be considered as a threshold value at some spot segments. Therefore, the threshold ( $T$ ) is calculated using Equation 4 (Kamberova and Shah, 2002).

$$T = C \times W \times M \quad (4)$$

Where  $C$  is the constant,  $W$ : is the spot segment size.

The constant value ( $C$ ) is calculated using the mean of the experimental results that resulted from testing the set of spot images. In our experimental results, we calculate the constant value ( $C$ ) using 300 extracted spot segments. All of the extracted spot set images have shown the same behavior as the value of the constant fall in the range of 0.5 to 0.01.

The spot segment size ( $W$ ) is the window size, which represents the dimension of the spot segments that extracted from the microarray image  $w$  is calculated using Equation 5.

$$W = w * h \quad (5)$$

Where  $w$  is the distance between any two horizontal grid lines,  $h$  is the distance between any two vertical grid lines. Because the grid lines are parallel the perpendicular distance between them is constant. The distance between grid lines  $L1$ , and grid line  $L2$ , is the distance between the two intercepts of these lines (Fuyong, 2013).

**Gene expression intensity**

The extracted segment usually includes the spot, the background, noise, and other objects such as piece of the surrounded spots or other artifacts. To calculate the gene expression intensity, a spot measurement model is used as Shawn in Equation 6 [R].

$$\text{Spot measured value} = \gamma A/B + \delta + e - bi \quad (6)$$

Where,  $\gamma A/B$  is the log-transformed true fold change of gene of condition A with respect to condition B (log (intensity of red color / intensity of green color)),  $\delta$ - dye effect,  $e$  – measurement error with  $E[e] = 0$  and  $\text{Var}(e) = \sigma^2$ , and  $bi$  is the background intensity.

However, this model does not taking the other objects and the other artifacts (OOA) that may exist in the extracted segment into consideration. Therefore, in order to enhance the spot measure value (Equation 6), we propose equation 7 to obtain better for quality gene expression intensity (GEI) value that takes all factors (OOA) into consideration.

$$GEI = \gamma A/B + \delta + e - (bi + OOA) \quad (7)$$

Where,  $\gamma A/B$  is the log-transformed true fold change of gene of condition A with respect to condition B (log (intensity of red color / intensity of green color)),  $\delta$ - dye effect,  $e$  – measurement error with

**Table 1.** The gene expression intensity.

Sub microarray No.	GEI
1	84.32
2	85.63
3	87.32
4	89.21
5	82.41
6	90.87
7	87.89
8	93.41
9	91.2
10	80.31
11	84.57
12	85.39
13	86.49
14	91.32
15	95.34
16	93.84
17	82.36
18	91.84
19	83.91
20	86.34

**Table 2.** Comparison between the proposed method and GMM.

Sub microarray No.	Proposed methodology	GMM
1	84.32	63.73
2	85.63	65.26
3	87.32	68.59
4	89.21	71.66
5	82.41	74.2
6	90.87	77.25
7	87.89	79.39
8	93.41	81.01
9	91.2	82.34
10	80.31	83.27
11	84.57	83.64
12	85.39	83.64
13	86.49	83.86
14	91.32	83.92
15	95.34	83.98
16	93.84	83.8
17	82.36	82.12
18	91.84	82.68
19	83.91	83.27
20	86.34	82.59

$E[e] = 0$  and  $\text{Var}(e) = \sigma^2$ ,  $b_i$  is the background intensity and OOA is the objects and the other artifacts intensity values.

We subtracted the background intensity of the extracted segment for each spot rather than taking the whole image background. And this is due to that the whole image background intensity gave false and inaccurate result for some spots area's with low intensity values.

The experimental results for tested sets of spots show a negative values when we used the whole image background. But, using the extracted segment background area with all its objects and other artifact, it gave positive results even with the low intensity values.

## EXPERIMENTAL RESULTS

The experimental results and the comparison of the methodology were carried out in a set of real microarray images, obtained from genome database (Image Dataset, 2016). The experimented genome database for whole yeast Genome (ISB Version - provided by Laura Hoopes and Allen Kuo) (Figure 2). The yeast genome contain 20 sub microarray images each one contains 480 extracted segments.

The experimental results show that the proposed methodology is able to calculate the intensity for all of the studied sub microarray images.

Table 1 shows the calculated gene expression intensity (GEI) for the 20 sub microarray. The average of the calculated GEI is equal to 87.77.

To evaluate the proposed methodology, we compare its results with the results of the Gaussian Mixture Models

(GMM) by (Emmanouil et al., 2009). Table 2 shows the experimental results for the proposed methodology and the results of GMM. The comparison of the proposed methodology and GMM algorithms was carried out in a set of real microarray images, available at (Image Dataset, 2016). The experimental results are presented in Table 2. The average of the calculated GEI using the proposed methodology is equal to 87.77, and for the GMM is equal to 79.01. The experimental results show that the proposed methodology has ability to segment the spot and calculate the gene expression intensity, and the experimental result was comparable. Figure 6 represents the comparison between the two methodologies.

Some spot in the extracted segments were hazy with low resolution as shown in Figure 7. Even though, the methodology was able to calculate the intensity for such images and this is due to the enhancement steps that were able to deal with such images and treat them to give correct results.

Some spots were not fully expressed in the microarray images and this is due to incorrect staining of the microarray or incorrect capturing to some parts of the microarray images; this lead to low intensity value of the extracted segments. The proposed methodology was able to detect and identify the low intensity of the spot because the methodology relay on applying adaptive threshold on extracted segment images rather than using the microarray full image.

Some methods of segmentation that relay on spot shape such as fixed circle segmentation or adaptive

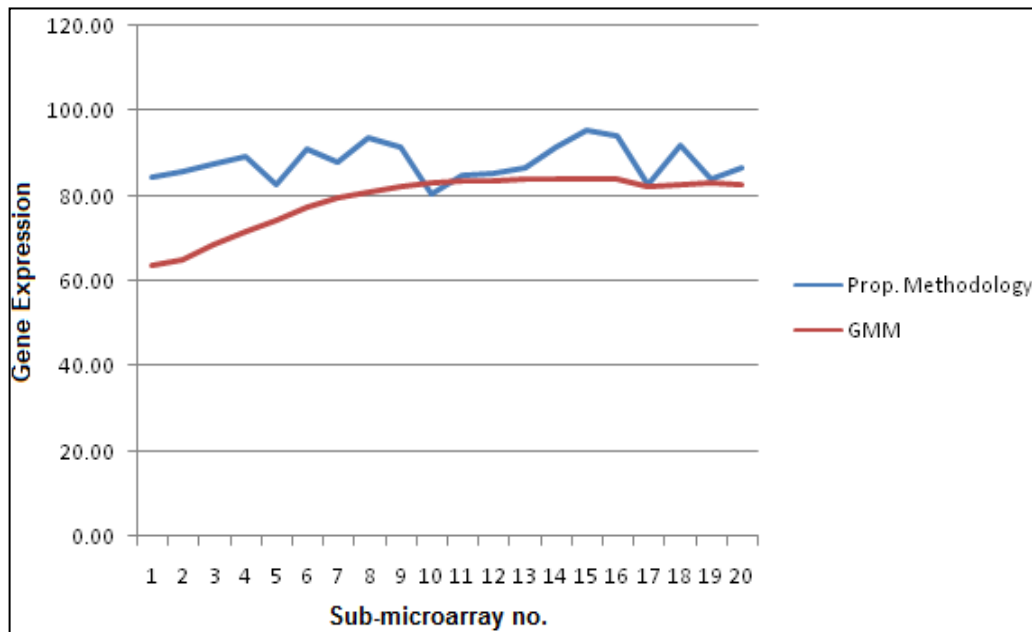


Figure 6. Low resolution captured segments.



Figure 7. Low resolution captured segments.

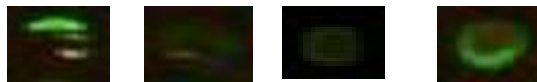


Figure 8. Spot segments with small portion of spot gene expression.

circle segmentation (Zacharia and Maroulis, 2008; Antoniol and Ceccarelli, 2004) are not able to detect spot with unclear shape because they rely on the shape of the spot and not the extracted segment of the spot. Some cases are presented in Figure 8.

However, the proposed methodology is able to calculate the intensity of the spots with unclear shapes because it depends on the segmentation using adaptive threshold rather than segmentation using the spot shape and calculating the intensity of the every extracted segments rather than extracting the spot based on its shape.

A comparison between the features of the proposed methodology with other methodologies and techniques was summarized in Table 3, which shows a comparison according to spot extraction, enhancement methods, noise removal, background, spot feature analysis and spot size used in microarray analysis.

To test the ability of the proposed methodology to get gene expression intensity, we tested the accuracy of the 20 sub microarray yeast genome with the noises, that is, Gaussian white noise, salt and pepper noise, multiplicative noise. In fact, the aim of this test is to show the ability of the proposed methodology to extract the spot and measure its intensity despite the circumstances of preparation of the microarray and the resulted noise. Since the proposed methodology rely on extracting each spot segment separately and removing the noise with the proposed steps then it was able to calculate the intensity with these different added noises. The results of intensity accuracy calculation are shown in Table 4.

Microarray image modified by adding Gaussian white noise of mean  $m$  and variance  $v$  to the image, where  $m$  is zero mean noise with 0.01 variance then, we calculate the GEI of the sub microarray images after adding the

**Table 3.** Comparison between proposed methodology and proposed techniques.

Name	Fixed circle segmentation (Zacharia and Maroulis, 2008)	Adaptive circle segmentation (Antoniol and Ceccarelli, 2004)	Histogram segmentation (Giannakeas and Fotiadis, 2008)	Adaptive shape segmentation (Bariamis et al., 2010)	Proposed methodology
Spot extraction	Fixed circle segmentation	Adaptive circle segmentation	Histogram segmentation	Adaptive shape segmentation	Adaptive threshold
Enhancement	No	Yes	No	No	Yes
Noise removal	No	No	No	No	Yes
Background	Whole microarray image background	Whole microarray image background	Whole microarray image background	Whole microarray image background	Spot segment extracted background
Spot feature analysis	Red intensity, green intensity for spots	Red intensity, green intensity for spots	Red intensity, green intensity for spots	Red intensity, green intensity for spots	Red intensity, green intensity for spots.
Spot size	assumes that all spots have the same size	assumes that the spot has a circular shape and permit adjusting the size of spots	based on spots intensity values rather than spot size	The spot is adaptive in size and irregular in shape	based on the intensity values of the spot rather than spot size

**Table 4.** GEI after adding noise: G-GEI, SP-GEI, M-GEI for the 20 sub microarray images.

Sub-microarray No.	GEI (Without adding noise)	G-GEI	SP-GEI	M-GEI
1	84.32	84.07	83.31	82.31
2	85.63	85.37	84.60	83.59
3	87.32	87.06	86.27	85.24
4	89.21	88.94	88.23	87.17
5	82.41	82.16	81.50	80.53
6	90.87	90.60	89.87	88.79
7	87.89	87.63	86.92	85.88
8	93.41	93.13	92.38	91.27
9	91.2	90.74	90.01	88.93
10	80.31	79.91	79.27	78.31
11	84.57	84.15	83.56	82.55
12	85.39	84.96	84.37	83.44
13	86.49	86.06	85.45	84.51
14	91.32	90.86	90.22	89.23
15	95.34	94.86	94.20	93.16
16	93.84	93.37	92.71	91.69
17	82.36	81.95	81.37	80.48
18	91.84	91.38	90.74	89.74
19	83.91	83.49	82.90	81.99
20	86.34	85.91	84.96	84.02

Gaussian noise (we call it G-GEI).

Microarray image is modified by adding salt and pepper noise to the image with noise density value.05. Then we calculate the GEI of the sub Microarray image after salt and pepper noise (we call it SP-GEI). Microarray image is also modified by adding multiplicative noise to the image, using the Equation 9 (Goodman, 1976),

$$J = I + n * I \quad (9)$$

Where  $n$  is uniformly distributed random noise with mean

0 and variance  $v$  is equal to 0.04.

Then we calculate the GEI of the sub Microarray image after adding multiplicative noise (we call it M-GEI). Table 4 expresses the accuracy results of 20 sub Microarray images after adding Gaussian, Salt and Pepper, and Multiplicative noises.

The results show that the proposed methodology was able to calculate the gene expression intensity values for the different spot segments after adding noise and this is due to using the local spot segment enhancement methods that is applied to each spot segment separately

and then calculating the intensity of each spot which represent the gene expression level. The mean of the accuracy for the whole 20 yeast genome sub-microarray was 87.77. After the addition of Gaussian white noise, salt and pepper noise, and multiplicative noise the mean of the GEI values were 87.40 for G-GEI, 86.73 for SP-GEI and 85.72 for M-GEI.

## Conclusion

This paper proposes a new methodology for spot gene expression intensity in microarray images. The proposed methodology analyzes spot in the extracted segments independently in the microarray image and uses the adaptive threshold for each segment separately rather than analyzing all the spots together in correspondence to the global background and noise of the microarray image. It also improves the microarray measurement model to calculate the gene expression intensity values. The experimental results showed that the proposed methodology was able to calculate the intensity values of spots in all extracted segments with accuracy of 87.77.

## Conflict of Interests

The authors have not declared any conflict of interest.

## REFERENCES

- Adwan O, Hudaib A, Fakhouri H, Alnemer L (2013) Microarray Gene Expression Extraction and Archiving Based on Local Spot Locations. *IJCSI Int. J. Comput. Sci.* Vol. 10(6/2):203-208.
- Alhadidi B, Fakhouri H, Al Mousa O (2006). cDNA Microarray Genome Image Processing Using Fixed Spot Position, *Am. J Appl. Sci.* 3(2):1730-1734.
- Angulo J, Serra J (2003). Automatic Analysis of DNA Microarray Images Using Mathematical Morphology. *Bioinformatics* 19(5):553-562.
- Antonio G, Ceccarelli M (2004). A Markov Radom Field Approach to Microarray Image Gridding, Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th Int. Conf. 3:550-553.
- Bariamis D, Iakovidis D, Maroulis D (2010). M3G: Maximum margin microarray gridding. *BMC Bioinformatics* 49:11-49
- Buhler J, Ideker T, Haynor D (2000). Dapple Improved Techniques for Finding Spots on DNA Microarrays. UW CSE Technical Report UWTR 2000-08-05.
- Efron B, Tibshirani R, Storey J, Tusher V (2001). Empirical Bayes Analysis of A Microarray Experiment. *J. Am. Stat. Assoc.* 96:456.
- Emmanouil I, Dionisis A, Panagiota P, Dimitris G (2009). Complementary DNA Microarray Image Processing Based on the Fuzzy Gaussian Mixture Model. *IEEE Trans. Information Technol. Biomed.* 13(4):419-425.
- Fuyong C (2013). Application of Operator's Solution in Linear Partial Differential Equations, Fifth International Conference on Computational and Information Sciences (ICIS), pp. 839-841
- Giannakeas N, Fotiadis DL (2008). Image Processing and Machine Learning Techniques for the Segmentation of cDNA Microarray Images, Handbook of research on advanced techniques in diagnostic images and biomedical application.
- Goodman J (1976). Some Fundamental Properties of Speckle. *J. Opt. Society Am.* 66(11):1145-1150.
- Image Dataset (2016). Retrieved from <http://www.bio.davidson.edu/projects/magic/magic.html>
- Jain A, Tokuyasu T, Snijders A, Segraves R, Albertson D, Pinkel D (2003). Fully Automatic Quantification of Microarray Image Data, *Genome Res.* 12(2):325-332.
- Jiao L, Sun Z (2010). Improvement of Image Contrast with Local Adaptation, *MultiMedia and Information Technology, Int. Conf.* 2:51-54.
- Kamberova G, Shah S (2002). DNA Array Image Analysis - Nuts and Bolts. *Data Analysis Tools for DNA Microarrays*, edition 2, DNA Press.
- Kim J, Kim L, Hwang S (2001), An Advanced Contrast Enhancement Using Partially Overlapped Sub-Block Histogram Equalization, *IEEE Trans. Circuits and Systems Video Technol.* 11(11):475-484.
- Lee M, Kuo F, Whitmore GA, Sklar J (2000). Importance Of Replication In Microarray Gene Expression Studies: Statistical Methods and Evidence from Repetitive cDNA Hybridizations. *Appl. Math.* 97(18):9834-9839.
- Lukac R, Plataniotis K (2006). "cDNA Microarray Image Segmentation Using Root Signals" *Wiley Periodicals, Inc.* 16:51-64.
- Mangalam H, Stewart J, Zhou J, Schlauch K, Waugh M, Chen G, Weller JW (2001). Gene X: An Open Source gene expression database and integrated tool set, *IBM Syst. J.* 40(2):552-569.
- Meher JK, Meher, PK, Dash GN (2011). Preprocessing of Microarray by Integrated OSR and SDF Approach for Effective DE noising and Quantification, International Conference on Information and Network Technology IACSIT Press, Singapore 4 p.
- Moffitt R, Yin-Goen Q, Stokes T, Torrance J, Phan J, Young A, Wang M (2011). "caCORRECT2: Improving The Accuracy And Reliability Of Microarray Data In The Presence Of Artifacts" *BMC Bioinformatics* 12:1
- Newton M, Kendzierski C, Richmond C, Blattner F, Tsui K (2001). On Differential Variability of Expression Ratios: Improving Statistical Inference about Gene Expression Changes From Microarray Data. *J. Comput. Biol.* 8(1):37-52.
- Rueda L, Vidyadharan R (2006). A Hill-climbing Approach for Automatic Gridding of cDNA Microarray Images. *IEEE Trans. Comput. Biol. Bioinformatics* 3:72-83.
- Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C (2002). BioArray Software Environment: A Platform for Comprehensive Management and Analysis of Microarray Data. *Biomed* 3:8.
- Schena M, Shalon D, Davis R, Brown P (1999). "Quantitative Monitoring of gene expression patterns with a complementary DNA microarray", *Science* pp. 467-470.
- Stefano L, Luo Y (2004). Gridding and compression of microarray images, Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conf. pp. 122-130
- Tiwari R (2005). Comparison of Microarray Image Analysis Software, ACM-SE 46 Proceedings of the 46th Annual Southeast Reg. Conf. 77-81:1151-1160.
- Uslan V, Bucak I (2010). "Microarray Image Segmentation Using Clustering Methods," *Math. Comput. Appl.* 15(2):240-247.
- Wang Y, Mark Q, Zhang K, Shih F (2007). A Hierarchical Refinement Algorithm for Fully Automatic Gridding in Spotted DNA Microarray Image Processing. *Information Sci.* 177(4):1123-1135.
- Wu S, Yan H (2003), Microarray Image Processing Based On Clustering And Morphological Analysis, Proceedings of the First Asia-Pacific Bioinformatics Conference on Bioinformatics 19:111-118.
- Yang YH, Buckley MJ, Dudoit S, Speed TP (2002). Comparison of Algorithms for Image Analysis On cDNA Microarray Data. *J. Comput. Graph. Stat.* 11(1):108-136.
- Zacharia E, Maroulis D (2008). Micoarray Image Gridding Via an Evolutionary Algorithm. *IEEE Int. Conf. on Image Process.* pp. 1444-1447.



# Scientific Research and Essays

## Related Journals Published by Academic Journals

- African Journal of Mathematics and Computer Science Research
- International Journal of Physical Sciences
- Journal of Oceanography and Marine Science
- International Journal of Peace and Development Studies
- International NGO Journal

**academicJournals**