RECOGNITION OF CDNA MICROARRAY IMAGE USING FEEDFORWARD ARTIFICIAL NEURAL NETWORK

R. M. Farouk[1], E. M. Badr[2] and M. A. SayedElahl[2]

[1]Department of Mathematics, Faculty of Science, Zagazig University,
P. O. Box 44519, Zagazig, Egypt
[2] Faculty of information Science, Banha University, Egypt.

Abstract

The complementary DNA (cDNA) sequence considered the magic biometric technique for personal identification. Microarray image processing used for the concurrent genes identification. In this paper, we present a new method for cDNA recognition based on the artificial neural network (ANN). We have segmented the location of the spots in a cDNA microarray. Thus, a precise localization and segmenting of a spot are essential to obtain a more exact intensity measurement, leading to a more accurate gene expression measurement. The segmented cDNA microarray image resized and used as an input for the proposed artificial neural network. For matching and recognition, we have trained the artificial neural network. Recognition results are given for the galleries of cDNA sequences . The numerical results show that, the proposed matching technique is an effective in the cDNA sequences process. The experimental results of our matching approach using different databases shows that, the proposed technique is an effective matching performance.

Keywords: *DNA recognition, cDNA microarray image segmentation, artificial neural network, DNA sequence analysis.*

1.Introduction

Biometrics refers to a distinctive, measurable characteristics automated system that can identify an individual by measuring their characteristics or traits or patterns, and comparing it to those on record. The biometric identifiers includes two categories, the physiological and the behavioral characteristics. A biometric would identify ones by iris, fingerprint, voice, cDNA, hand print and signature or behavior [1]. DNA biometrics is one of the most exact forms of identifying any given person. Each individual has its own individual map for all cell made, this map exist in everybody cell. This map (DNA) is the frame that explains who we are intellectually and physically, even an individual is an identical twin, this means there is no two persons have the same exact DNA [2]. Bioinformatics and Molecular biologists use microarray technology for identifying genes in a biological sequence and predicting the function of the identified gene within a larger system. The technology of Microarray based on creating DNA microarrays, which is composed of thousands of DNA sequences (probes), fixed to a silicon or glass substrate [3,4]. Rapid progress has been made in improving the recognition rates. This development has greatly improved the scope of using cDNA for identification purposes. Due to the increasing of the crime rates everyday, there is an urgent need for a system which is safe and fast. The cDNA matching systems have a popular use in criminal trials, especially in the proving rape cases. The general cDNA recognition system based on feature extraction and classification of the image using the features extracted. The

current works focus on enhancing the feature extraction and matching and improving the accuracy of the results. The main problems surrounding cDNA biometrics is that it is not a quick process to identify someone by their cDNA [5]. Applications of DNA microarrays are becoming a common tool in many areas of microbial research, including pathogenesis, microbial physiology, ecology, epidemiology, phylogeny, fermentation optimization and pathway engineering [6]. Different DNA based molecular techniques viz. PCR, AFLP, RAPD, RFLP, SSCP, real time PCR, DNA microarray has been used for identification and to assess the genetic diversity among the parasite population [7]. Microarrays constructed with dozens to millions of probes on their surface to allow high throughput analysis of many biological processes, which performed simultaneously on the same sample. It enables systematic surveys of variations in DNA sequence and gene expression. Now Microarrays are widely used for deoxyribonucleic acid resequencing, comparative genomic hybridization, single-nucleotide polymorphism genotyping and gene expression analysis [8]. Automatic grid alignment jointly with the FPGA based hardware architectures are an effective solution for automated and fast microarray image processing, ride the insufficiency of presenting software platforms in case of needing a large number of microarray analysis [9].

1.1.Outline

The rest of this paper is organized as follows: in section 2 we have discussed cDNA microarray images preprocessing; enhancing, gridding, segmentation, in section 3, we have used the segmented microarray images as an input of ANN. This method includes the training of the entire pattern at once against the other earlier works of training and in section 4, we have discussed our results. The comprehensive algorithm is shown in Fig 1.
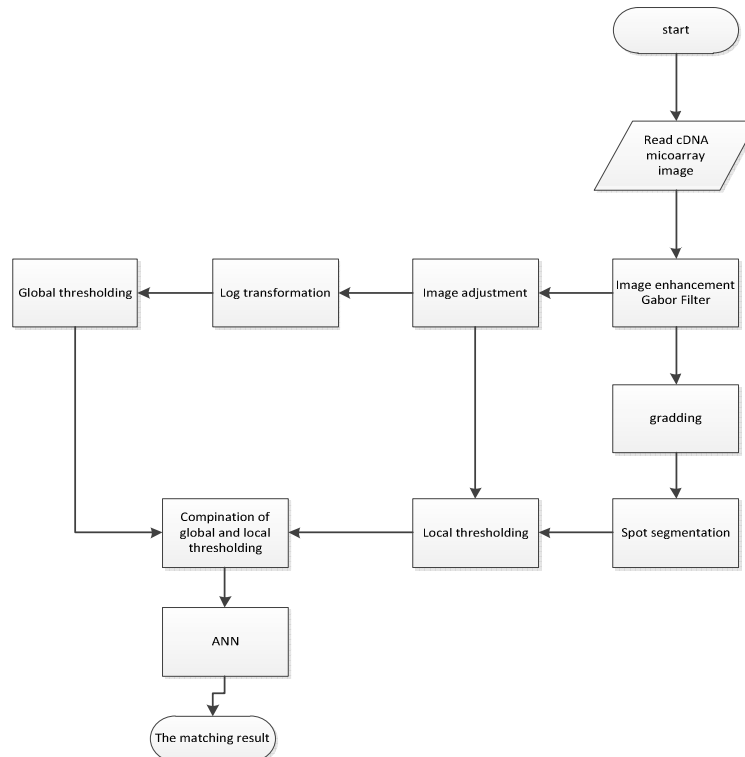

Fig. 1 Flowchart of the proposed technique

2.cDNA image enhancement

Microarray imaging considered an essential tool for massive scale analysis of gene expression. Microarrays are arrays of glass microscope slides, in which thousands of discrete DNA sequences are printed by a robotic array, thus, forming circular spots of known diameter. It consists of thousands of spots representing chains of cDNA sequences, arranged in a two-dimensional array [10].

The precision of the gene expression rely on the further image processing and experiment itself. There are several studies and approach to reduce and enhance the noise on cDNA microarray images, introduced during the experiment. cDNA microarray image analysis is a new and exciting computing paradigm with biological background. It includes the concepts ranging from molecular biology, statistics to computation while include theories and practices of computer science. Study in DNA microarray includes the gene expression, genetic data management, repository system and the algorithms of image analysis [11]. Microarray image enhancement by denoising using stationary wavelet transform provides a better performance in denoising than traditional wavelet transform method [12]. Removing noise impairments while preserving structural information in cDNA microarray images using fuzzy vector filtering framework, Noise removal is performed by tuning a membership function which utilizes distance criteria applied to cDNA vectorial inputs at each image location [13]. The task in microarray image enhancement involves removing the image background, noise from scanning and illustrating the guide spots of the cDNA [14]. It typically involves the following steps:

(i) Filter cDNA Image and get the guide spots (using Gabor filter).
(ii) Identify the location of all spots on the microarray image.
(iii) Generates the grid within each block which subdivides the block into n × m sub regions, each containing at most one spot.
(iv) Segment the spot, if any, in each sub region.

2.1 Gabor filtering

The Gabor filter method has been widely used both in multi-resolution image processing and feature extraction for object identification [15]. Given an image with size M x N, a 2D Gabor function is a Gaussian modulated by a sinusoid. It is a non-orthogonal wavelet and it can be particularly by the frequency of the sinusoid and the standard deviations of Gaussian x and y [16]:

$$G\ k = \frac{k^2}{\sigma^2} \left\{ \exp\left(-\left(\frac{x^2 + y^2}{2\sigma^2}\right)\right) \left\|k^2\right\| \left(\exp^{i\vec{k}\vec{x}} - \exp\frac{\sigma^2}{2}\right) \right.$$

(1)

$$\vec{x} = (x, y)\ \ pixel\ coordinates$$

$$\vec{k} = k\cos\varphi + k\sin\varphi$$

$$\vec{x}.\vec{k} = k\cos\varphi x + k\sin\varphi y$$

Euler form Substitute

$$\exp(ikx) = \underbrace{\left|\cos\left(x*k\cos\varphi + y*k\sin\varphi\right)\right.}_{Real\ part} + i\underbrace{\left|\sin\left(x*k\cos\varphi + y*k\sin\varphi\right)\right|}_{Imaginary\ part}$$

$$Gkreal\ part = \frac{k^2}{\sigma^2}\left\{\exp^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)\|k^2\|}\right.\left\lfloor\cos\left(x*k\cos\varphi+y*k\sin\varphi\right)\right.$$

$$Gk\ img\ part = \frac{k^2}{\sigma^2}\left\{\exp^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)\|k^2\|}\right.\boxed{\sin\left(x*k\cos\varphi+y*k\sin\varphi\right)}$$

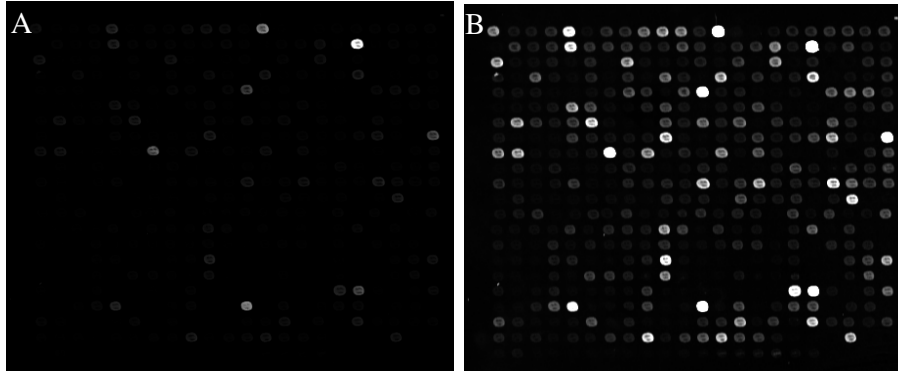The Gabor filter used to enhance the image and Fig. 2.

Fig. 2 The effect of Gabor filter on (A) original image, (B) Gabor filtered image

2.2 Gridding

Images from microarray experiments are highly structured since they are composed of high intensity spots located on a regular grid. The format of the spots is almost circular, although differences are possible. The ideal microarray image has the following properties [17,18]

(i)   All the sub grids are about the same size.
(ii) The spacing between sub grids is regular.
(iii) The spots lie on the center of the intersections of the lines of the sub grid.
(iv) The shape and size is the same for all the spots and quite circular.
(v)   Have a fixed grid location for a given type of slides.
(vi) No dust or contamination is on the slide.
(vii) There is uniform and minimum background intensity across the image.

 It goes without saying that almost all real microarray images violate at least one of these conditions. In fact, we frequently observed variations on the spot position, irregularities on the spot shape and size, contamination, and global problems that affect multiple spots. Generally, the size and the shape of the spots might fluctuate significantly across the array. The microarray

 image gridding is the separation process of the spots into distinct cells, to generate the correct grid for the microarray image, consider the location of good quality spots (which act as guide spots). A good-quality spot have to have a circular shape, appropriate size, and with intensity consistently higher than the background. Furthermore, its position have to agree the overall spot geometry as instructed by the printing process. After the guide spots are found, the correct grid

can be generated based on their geometry [19,20], Fig. 4. To find the grid lines in the microarray images is based on a series of steps:

(i) Create a horizontal profile to get the mean intensity for each column of the image.
(ii) Autocorrelation to enhance the self-similarity of the profile. The smooth result promotes peak finding and estimation of spot spacing Fig. 3.
(iii) Locate centers of the spots.
(iv) Determine divisions between spot, The midpoints between adjacent peaks provides grid point locations.
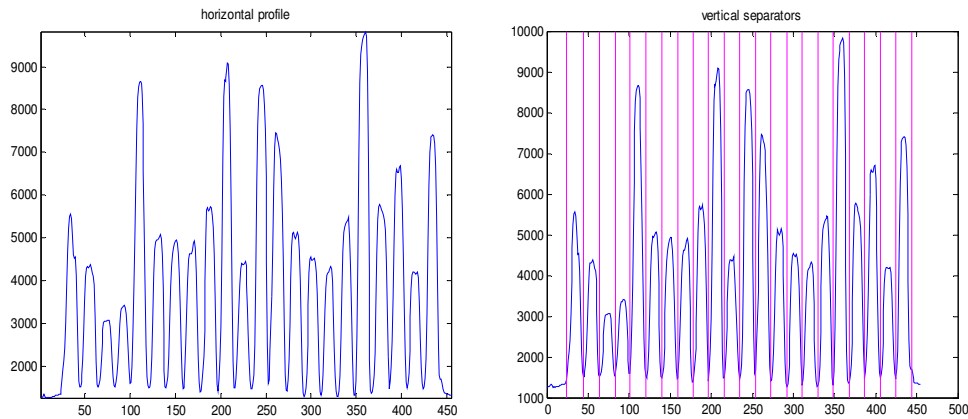(v) Transpose the image and repeat the previous steps to get the vertical grid.

Fig. 3 Steps of gridding
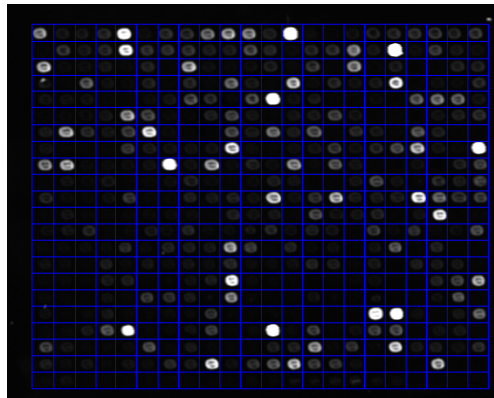

Fig. 4 Result of automatic gridding

2.3 Segmentation

Image segmentation is the process of distinguishing objects from the background and partitioning the image into several regions, each having its own properties. It is usually the first step in vision systems, and is the basis for further processing such as description or recognition. In microarray image processing, segmentation refers to the classification of pixels as either the signal or the surrounding area. Most images segmentation approaches can be placed in one of five categories [21]: clustering or threshold-based methods, boundary detection methods, region growing methods, shape-based methods, and hybrid methods. In order to get the best result of cDNA microarray image we follow these steps:

(i) Image adjustment.

(ii) Segment spots from the background by thresholding (apply logarithmic transformations, then threshold intensities)

(iii) Local thresholding

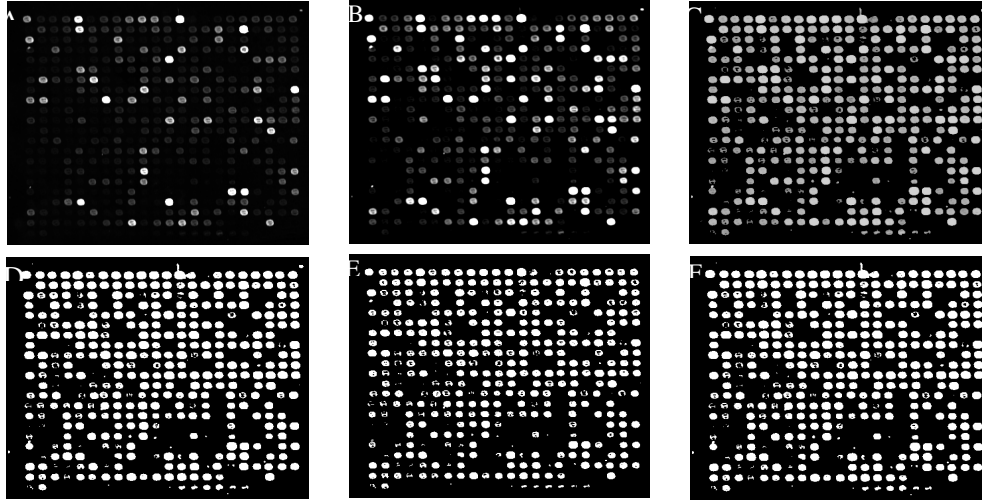(iv) Combine the result of the local threshold with the global threshold result.

Fig. 5 The steps of segmentation A, Gabor filtered image, B, image adjustment, C, Logarithmic transformation, D, Global thresholding, E, Local thresholding, F, the combined image

3.DNA Matching by ANN

DNA sequence recognition includes knowing if, whether or not a new input has been presented before and stored in. The neural network classifier learning is an easy process because it is dependent only on the locally available information, but since the information is mixed together at the storage elements, unambiguous retrieval of stored information is not simple. The gradient descent is generalized to a non-linear, multi-layer feed forward network called back-propagation. This is a supervised learning algorithm that learns by first computing an error signal and then propagates the error backward through the network by assuming the network weights are the same in both the backward and forward directions. Back-propagation is a very popular and widely used network learning algorithm [22].

In this paper, we explain our work on training the cDNAs on the MDANN by using the back propagation algorithm and then matching the cDNA for recognition from the database. The ANN used here consists of 21 neurons (the input layer is of eight neurons, the hidden layer is of 12 neurons, and the output layer is of one neuron). Fig. 7 shows the structure of the ANN but the connections between the input and the hidden layer are not shown completely for the purpose of clarity and the complete structure is obtained by extending the connections to the entire image.
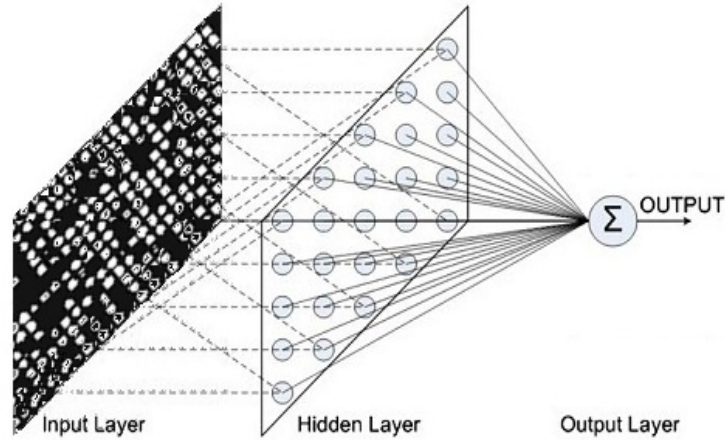
Fig.6 the structure of ANN

3.1 Algorithm for the ANN

Let M_1 and M_2 are structures containing m layers and k elements in each field and comprise of internal weights $W_{dl}, W_{dr}, W_{dl1}, W_{dr1}$, are matrices denoting the entire strengths of the image to neuron connection differingonents are single rodifferentenoting the weights of the connection of neurons with the image pixel points, their differential values, the bais values for the neurons, their differential values.

Step 1: calculation of the derivatives:

Now, let i refers to the number cDNA we are trying to train on the network, j refers to the number of regresses which is given by one in excess of the number of neurons, k refers to the dimension of the input matrix, and n refers to the number of neurons.

$$\delta = actual\ output - required\ output \tag{2}$$

For the weights eq. 3 and eq. 4

$$A_d(j) = A_d(j) + \delta, \qquad if\ \ j = 1 \tag{3}$$

$$A_d(j) = A_d(j) + \delta\left(\frac{2}{1 + \exp(-2(M_{j1}x\ I(i)xM_{j2})} - 1\right), \qquad if\ \ j > 1 \tag{4}$$

Where the symbols M_{j1} and M_{j2} refer to elements in the jth layer elements of the structures M_1 and M_2 respectively and $I(i)$ refers to the ith matrix of cDNA input image to the network. The values are calculated for all values of j. Now f refers to the activation function and is given by the eq. 5.

$$f = \left(\frac{-2}{1 + \exp(-2(M_{j1} x \, I(i) x \, M_{j2})^2} \right) \tag{5}$$

For the internal weights eq. 5 and eq. 7.

$$W_{dl1}(j,k) = W_{dl1}(j,k) + (A_w(j) \times \delta \times f \times (M'_{j2} \times I(i,k,:)')) \tag{6}$$

$$W_{dr1}(j,k) = W_{dr1}(j,k) + (A_w(j) \times \delta \times f \times (M'_{j1} \times I(i,:,k))) \tag{7}$$

Where $W_{dl1}(j,k), W_{dr1}(j,k)$ refer to elements in the jth row and k^{th} column of the matrices W_{dl1} and W_{dr1} respectively. $I(i,k,:)$ refers to the kth row elements of the ith input in the structure and $I(i,:,k)$ refers to the kth column elements of the ith input in the structure. These values are calculated for all the coordinates (i,k).

For the bias values eq. 8.

$$A_{d1}(j) = A_{d1}(j) + (\delta \times f \times A_w(j)) \tag{8}$$

Step 2 : updating the weights and bias values and setting their derivatives back to zero:

For all values of j

For the weights eq. 9, eq. 10 and eq. 11.

$$A_w(j) = A_w(j) + (\alpha \times A_{dw}(j) - (\eta \times A_d(j)) \tag{9}$$

$$A_{dw}(j) = (\alpha \times A_{dw}(j) - (\eta \times A_d(j)) \tag{10}$$

$$A_d(j) = 0 \tag{11}$$

For the bias values eq. 12, eq. 13 and eq 14.

$$A_b(j) = A_b(j) + (\alpha \times A_{db}(j) - (\eta \times A_{d1}(j)) \tag{12}$$

$$A_{db}(j) = A_{db}(j) - (\eta \times A_{d1}(j)) \tag{13}$$

$$A_{dl}(j) = 0 \tag{14}$$

Where η and α refer to the initial learning rate and the momentum factor respectively.

Step 3: updating the internal weights and setting the derivatives back to zero:

For all the coordinates

$$M_{jk1} = M_{jk1} + (\alpha \times W_{dl}(j,k)) - (\eta \times W_{dl1}(j,k)) \tag{15}$$

$$W_{dl}(j,k) = (\alpha \times W_{dl}(j,k)) - (\eta \times W_{dl1}(j,k)) \tag{16}$$

$$W_{dl1}(j,k) = 0 \tag{17}$$

$$M_{jk2} = M_{jk2} + (\alpha \times W_{dr}(j,k)) - (\eta \times W_{dr1}(j,k)) \tag{18}$$

$$W_{dr}(j,k) = (\alpha \times W_{dr}(j,k)) - (\eta \times W_{dr1}(j,k)) \qquad (19)$$

$$W_{dr1}(j,k) = 0 \qquad (20)$$

Where M_{jk1} and M_{jk2} refer to the kth element in the jth layer of the structure M_1 and M_2 respectively.

Step 4: calculating the output values after the training.

$$o = A_w(j) \ for \ j = 1$$

$$o = A_w(1) + \sum_{j=2}^{n} A_w(j) \times \left(\frac{2}{1 + \exp((M_{j1} \times I(j) \times M_{j2}) + A_b(j))} \right) for \ j > 1 \qquad (21)$$

The iteration goes on until the value calculated in Step 4 is the target value that was assigned to the image. This value is unique for every distinct microarray image and when a random microarray image is tested on the network, this value concludes as to whether the microarray image is already present in the database or not and if present the exact cDna to which it corresponds to. The ANN is trained by using 110 microarray image from [23] . During the training stage, each microarray image is assigned to a distinct arbitrary value. Now the trained ANN is subjected to a test to evaluate if the network is capable of recognizing a microarray image that have already trained on it and also to evaluate if it can reject a microarray image that have not trained on it. We will use some trained microarray images to test the ability of the ANN to identify the trained cDNA.
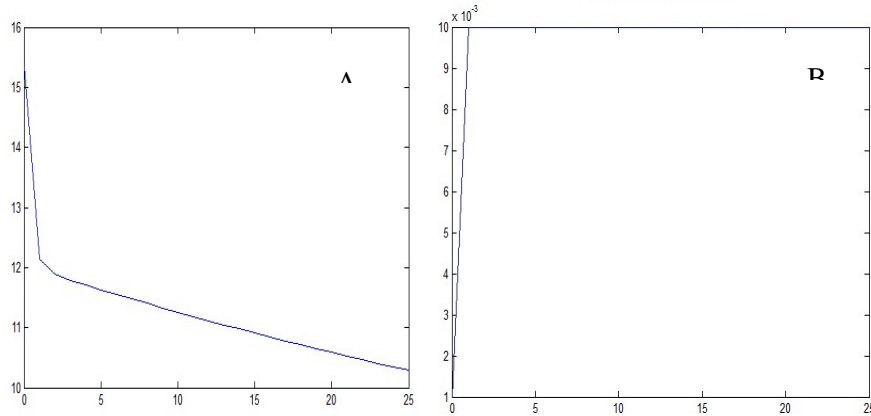

Fig.7 (A) the relation between the performance and the number of iterations,(B) the relation between the learning rate and the number of iterations

4.Result

The microarray image is augmented in order to enable the proper extraction of gene spots. The segmented image which separates the background from foreground area is shown in Fig. 5 (a-c). Different enhancement operators are applied to microarray image like global and local thresholding as shown in Fig. 5(D-E). The combined microarray image Fig. 5 (F) is resized to 20x20. The resized image is trained on the ANN by using the gradient descent training procedure. The performance curve has been plotted against the number of iterations. It is observed that the

performance is steadily improving and leading to a value around zero Fig. 6 (A). The rate of the learning curve has been shown Fig 6 (B), and this refers to the rate at which the network is learning. This rate is altered continuously in the algorithm every time the network is subject to iteration.

5.Conclusion

In this paper, we present a new method for cDNA recognition based on the artificial neural network (ANN). Microarray imaging is applied for the synchronous identification of thousands of genes. We have segmented the location of the spots in a cDNA microarray images. The segmented cDNA microarray image is resized and it is used as an input for the proposed artificial neural network. For matching and recognition, we have trained the artificial neural network. Recognition results are given for the galleries of cDNA sequences. The numerical results show that, the proposed matching technique is an effective in the cDNA sequences process. The experimental results of our matching approach using different databases shows that, the proposed technique has an effective matching performance.

Acknowledgement

References

[1] Rattani A., Adaptive Biometric System based on Template Update Procedures, PhD thesis, University of Cagliari, Italy, 2010

[2] Abdul Ahad H. Biometrics-The Human Password, JITPS, 1 (1), 29–42, 2010.

[3] Chee M., Yang R. and Hubbell E., Accessing genetic information with high-density DNA arrays, Science, 610–614, 1996.

[4] Fenstermacher D. Introduction to Bioinformatics, Journal of the American Society for Information Science and Technology, 56 (5), 440-446, 2005.

[5] Edward Connors E.,Lundregan T.,Miller N.and McEwen T. Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial, National Institute of Justice 2006.

[6] Ye R., Wang T., Bedzyk L. , Croker K., Applications of DNA microarrays in microbial systems, Journal of Microbiological Methods, 47, 257–272, 2001.

[7] Ahmed M., Singh M. N., Bera A. K. , Bandyopadhyay S. and Bhattacharya D.,Molecular basis for identification of species/isolates of gastrointestinal nematode parasites, Asian Pacific Journal of Tropical Medicine,589-593, 2011.

[8] Ward K., M. D, Microarray technology in obstetrics and gynecology: A guide for clinicians, American Journal of Obstetrics and Gynecology, 195: 364–72, 2006.

[9] Belean B., Borda M., Le Gal B. and Terebes R., FPGA based system for automatic cDNA microarray image processing, Computerized Medical Imaging and Graphics, in press, 2012.

[10] Schena, M. Microarray biochip technology, Eaton Publishing Company, ISBN: 1881299376, 2000.

[11] Karim R., Mahmud S., A review of image analysis techniques for gene spot identification in cDNA Microarray images, International Conference of Next Generation Information Technology, 2011.

[12] Wang XH, Istepanian RS, Song YH.,Microarray Image Enhancement by Denoising Using Stationary Wavelet Transform, IEEE Trans Nanobioscience, 2(4), 184-9, 2003.

[13] Lukac R., Plataniotis K., Smolka B., Venetsanopoulos A., cDNA microarray image processing using fuzzy vector filtering framework, Fuzzy Sets and Systems ,152, 17–35, 2005.

[14] Liew A., Yan H. and Yang M., Robust adaptive spot segmentation of DNA microarray images, Pattern Recognition, 36 (5), 1251–1254, 2003.

[15]   Havukkala I., Pang S., Jain V., and Kasabov N., Classifying MicroRNAs by Gabor Filter Features from 2D Structure Bitmap Images on a Case Study of Human MicroRNAs, Journal of Computational and Theoretical Nanoscience ,2(4): 499–505, 2005.

[16]   Ng C., Lu G., and Zhang D., Performance Study of Gabor Filters and Rotation Invariant Gabor Filters, Proceedings of the 11th International Multimedia Modelling Conference ,158-162, 2005 .

[17]   Wang Y., Shih F. and Ma M., Precise Gridding of Microarray Images by Detecting and Correcting Rotations in Subarrays, the 8th Information Sciences Conference, 2005.

[18]   Kuklin A. Laboratory automation in microarray image processing, American Laboratory ,64–67, 2000.

[19]   J. D. and Thomas T.,Automatic Gridding of DNA Microarray Images using Optimum Subimage, International Journal of Recent Trends in Engineering ,1 (4), 2009.

[20]   Rueda L. and Rezaeian I. A fully automatic gridding method for cDNA microarray images, BMC Bioinformatics, 12-113, 2011.

[21]   Li Q., Luis R., A and Alioune N., Spot Detection and Image Segmentation in DNA Microarray Data, Applied Bioinformatics, 4 (1): 1-11,2005

[22]   Anderson, James A. An introduction to neural networks, Prentice Hall Publications 1995.

[23]   http://malaria.ucsf.edu/comparison/comp_SupplementalData.php, university of California, San Francisco.