Analysis of Microarray Image Spots Intensity: A Comparative Study

M.Anandhavalli, Chandan Mishra, M.K.Ghose

***Abstract*—Microarray technology gives rise to a challenge to interpret the meaning of the immense amount of biological information formatted in numerical matrices from the expression levels of thousands of genes, possibly all genes in an organism. To meet this challenge various methods have been developed using both traditional and innovative techniques to extract, analyze and visualize gene expression data generated from DNA Chip. The large amount of information provided by microarray images requires automatic techniques to develop accurate and efficient processing. Each spot in the microarray contains the hybridization level of a single gene. In this paper, the comparative analysis of measurement of intensity of microarray spots using the intensity transformation methods such as gray level, logarithmic, gamma and contrast stretching after cropping a particular section of microarray image is performed, and the numeric data sheet is prepared for these transformations.**

***Index Terms*—DNA Microarray chip, Contrast Stretching, Logarithmic Stretching, Gamma Stretching.**

I. Introduction

The DNA microarray technique has empowered the scientific community to understand the fundamental aspects underlining the growth and development of life as well as to explore the genetic causes of anomalies occurring in the functioning of the human body. Gene is a segment of DNA, which contains the formula for the chemical composition of one particular protein. Genes serve as the blueprints for proteins and some additional products, and mRNA is the first intermediate during the production of any genetically encoded molecule. The concentration of a specific mRNA molecule is usually called the expression level of the respective gene, and it serves as an indicator of the amount of end product that is currently being produced [1]. DNA Microarray technique first appeared in E.M. Southern's publication of 1970 has now become an efficient technology platform for identifying valid biomarkers by hybridizing transcriptomic specimens against almost genome-wide genes simultaneously (Brazma et al., 2001). Spotted DNA microarrays have become a very useful technology which arised at the middle of the last decade, when the first DNA microarray was developed by Schena et al. [8]. Microarrays allow researchers to compare and analyze thousands of genes simultaneously, and to study their interactions and relations. The reader is referred to [2],[4],[6],[9] for a detailed explanation on how spotted microarray experiments are carried out.

Manuscript received July 23, 2009.
M.Anandhavalli, Chandan Mishra, M.K.Ghose, Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, East Sikkim, India - 737136. e-mail: anandhigautham@gmail.com.

II. DNA MICROARRAY

Microarrays (or biochips) allow for the simultaneous study of all the genes in an organism in a single experiment. This is made possible by spotting (placing) thousands of short DNA sequences on a surface. For microarrays manufactured using in situ synthesis (such as the ones studied in this work) each spot is fairly circular and the microarray image itself is very structured. Microarray technology relies on hybridization between the genes (messenger RNA or cDNA) and the DNA probes spotted on the array. Two gene pools are used; a test one and a control one. The genes bind to the probes on the array and become immobilized. The gene sequences are labeled with fluorescent dyes Cy3 and Cy5 (control and test) respectively [7]. Thus, the level of gene expression, corresponding to the amount of gene sequences immobilized on a specific spot on the array, is proportional to its intensity.

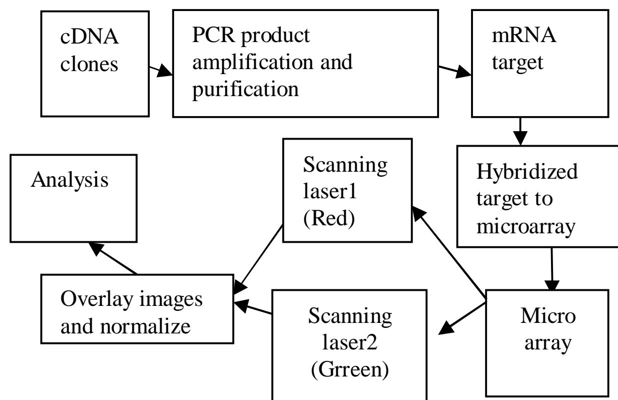


Fig. 1 Steps in DNA Microarray Experiment

III. GENE EXPRESSION DATA

A typical DNA microarray experiment provides the expression profiles of several tens of samples (say Ns % 100), over several thousand (Ng) genes. These results are summarized in an Ng×Ns expression table or gene expression matrix; each row corresponds to one particular gene and each column to a sample. Entry E_{gs} of such an expression table stands for the expression level of gene g in sample s. The original gene expression matrix obtained from a scanning process contains noise, missing values, and systematic variations arising from the experimental

IACSIT International Association of Computer Science and Information Technology WWW.IACSIT.ORG

procedure [1]. Data pre-processing is indispensable before any analysis can be performed.

IV. BACKGROUND MATERIALS

The following backgrounds were reviewed for Intensity Transformation, Logarithmic transformation Contrast Stretching and Gamma correction methods.

A. Intensity or gray level Transformation

In gray level transformation, the pixel value of the transformed image g at (x, y) depends only on the intensity of the original image f at that point (x,y) and the transformation function T is related by the expression g(x,y) = T[f(x,y)]. For detailed idea on Intensity transformation, we can refer [7].

B. Logarithmic transformation

It is implemented by using the expression g=c*log(1+double(f)), where c is the constant. One of the principle uses of the log transformation is to compress the dynamic range. When log transformation is applied, it is often desirable to bring the resulting compressed value back to full range of display. For detailed idea on logarithmic transformation, we can refer [7].

C. Contrast Stretching

The function representing contrast stretching is given as $s= T(r) = 1+(1+(m/r)^E)$, where r represents the intensities of input image, s corresponds to the intensity values in the output image, and E controls the slop of the function. For further detailed idea, we can refer [7].

D. Gamma Correction

Gamma Correction is implemented using the imadjust function g = imadjust(f, [l_{in},h_{in}], [l_{out},h_{out}], gamma) in MATLAB. This function maps the intensity value in image f to new values in g, such that the values between l_{in} and h_{in} map to values between lout and hout . The input image can be of class uint8, uint16, or double and the output image has same class as the input. All the inputs to the function imadjust, other than f are specified as values between 0 and 1. Using empty matrix ([ ]) for [l_{in}, h_{in}] or for [$_{lout}$,h_{ou}t] results in default value [0 1]. If h_{out} is less than l_{out} then output intensity is reversed. The parameter gamma specifies the shape of the curve that maps the intensity value in f to create g. If gamma is less than 1, the output image will be brighter. If gamma is greater than 1, the output image will be darker. For further detailed idea, we can refer [7].

V. METHODS

Model involves the essential pre-analysis issues of microarray image data including Cropping, Separation into RGB, Grayscale conversion, gridding pre-processing, global thresholding, local thresholding, applying different transformations and preparing the datasheet, which is given in Fig 2. The module is implemented in MATLAB and Windows XP environment.

A. Image Reading

The microarray image is stored as a J-PEG file. The image is much larger than the screen size, and hence scaled by the imshow( ) function in MATLAB.

B. Cropping

Image in MATLAB is cropped using imcrop( ) function which provides us the facility to study a particular region of interest.

C. RGB separation

This image is stored in RGB format. We are only interested in the red and green planes. In order to extract the red plane, simply index layer 1 and for the green plane, layer 2. It is Important to note that the spot shape in not same in both cases.

D. Grayscale Conversion

Before calculating the intensity transformation, it is necessary to find the locations of spots. So, it is important to perform grayscale conversion so that spot locations can be easily focused.

i. Histogram Equalization

Histogram equalization is implemented with the help of the following functions: J=histeq(I,hgram) and hgram=ones(1,n)* prod(size(A))/n. histeq() enhances the contrast of images by transforming the values in an intensity image, so that the histogram of the output image approximately matches a specified histogram. When we supply a desired histogram hgram, histeq chooses the grayscale transformation T to minimize where c0 is the cumulative histogram of A; c1 is the cumulative sum of hgram for all intensities k. This minimization is subject to the constraints that T must be monotonic and c1(T(a)) cannot overshoot c0(a) by more than half the distance between the histogram counts at a. histeq uses this transformation to map the gray levels in X (or the colormap) to their new values.

ii. Adding noise to image

It is implemented with the help of imnoise() function J=imnoise(I, 'gaussian', m, v). It adds Gaussian white noise of mean m and variance v to the image I. The default value is zero mean noise with 0.01 variance.

iii. Low pass filtering

It is implemented using function J = wiener2 (I, [m n], noise). This function filters the image I using pixel wise adaptive Wiener filtering, using neighborhoods of size m-by-n to estimate the local image mean and standard deviation. The default value of m and n is 3. The additive noise (Gaussian white noise) power is assumed to be noise.

E. Gridding

The goal of gridding or addressing microarray objects is to reliably compute intensity of corresponding spots in accord with microarray layout design [3]. The object intensity of respective microarray spot shows the expression level of arrayed probe sequence from related gene. Given top left corner as the reference of an intended microarray image, the

gene expression of spot objects are computed by image segmentation algorithms along with the layout design of designated spot size and spacing in horizontal and vertical which is shown in Fig.3.

F. Global Thresholding

It has been used to segment spots from background. Applying a single threshold level to the whole image so all spots are detected equally is generally a good idea. However, it will not work due to large differences in spot brightness.

G. Local Thresholding

Alternatively, the gridding can be used to determine local threshold values for each spot. Unfortunately, the results are mixed. Weak spots showed up well but spots with bright perimeters were as bad as the original global threshold before log space transformation. Since both techniques have their merits, let's combine the best of both approaches. This can be done using logical operation on the binary masks. These spot segmentation results are indeed much better.

H. Transformations

The transformations which are discussed in Section IV are used to equalize large variation in magnitude of the image spots.

I. Data Sheet

It is basically Microsoft Excel file which contains the microarray spot data which include the intensity of red and green planes respectively and position of spot measurements.

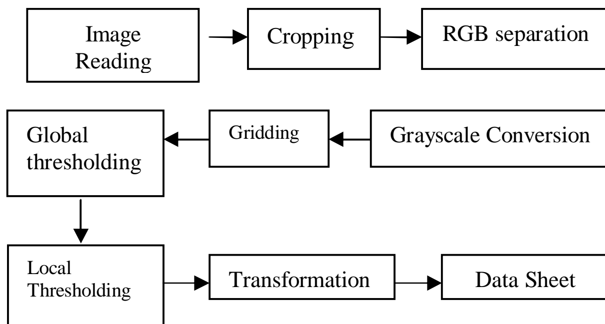


Fig. 2 Steps involved in pre-analysis of microarray image data

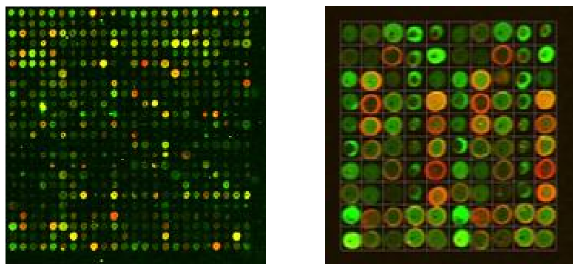
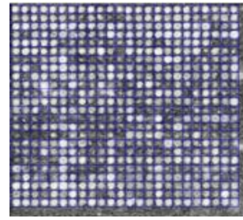
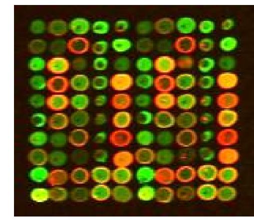


Fig.3 (a) The original image (b) Image after crapping



(c) Image after gridding



(d) Image after transformation

VI. EXPERIMENTAL RESULTS

The performance of the various transformations are evaluated on real microarray images from a public database of the Lymphoma/Leukemia Molecular Profiling Project and the images are available for free download from http://llmpp.nih.gov/lymphoma/data/rawdata.

The performance of the different intensity transformations are discussed in the table 1.

Table 1: Comparison of various intensity transformations

Transformation	Advantages	Shortcomings
1.Gamma	Gives better result for parameter gamma less than 1	Result improves only when used with histogram equalization where parameter gamma is greater than 1
2.Logarithmic	Compress the dynamic range	Needs to bring the resulting compressed value back to full range of display
3.Contrast Stretching	Gives better result when value of E increases.	Limited by thresholding function for binary images.

The graph in Fig.4 is plotted between the number of spots and the Intensity of red spots which is obtained applying logarithmic transformation and contrast stretching. Both the transformation uses gamma correction also. The numbers of spots are taken on X-axis and the Intensity value is taken on Y-axis.

The graph in Fig.5 is plotted between the number of spots and the Intensity of green spots which is obtained applying logarithmic transformation and contrast stretching. Both the transformation uses gamma correction also. The numbers of spots are taken on X-axis and the Intensity value is taken on Y-axis.

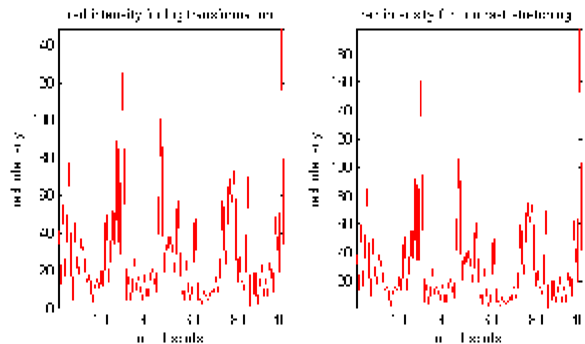


Fig. 4 Intensity of Red Spots.

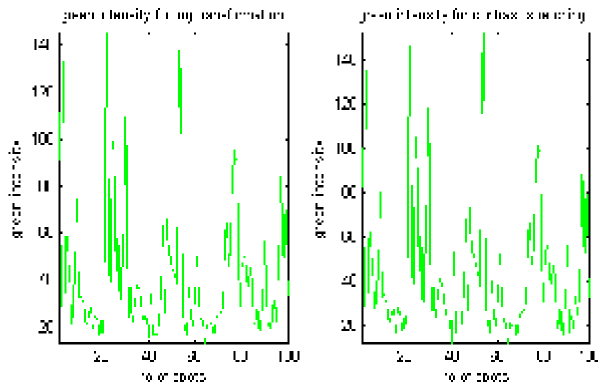


Fig. 5 Intensity of Green Spots

The data sheet in given Fig. 6 shows the logarithmic transformation applied with gamma correction in the microarray image. The first four columns represents the position of spot and the next two columns represents the red and green spot intensity and the last column represents the gene expression value which is a simple logarithmic value of red intensity divided by green intensity respectively.

The Data sheet in given Fig.7 shows the contrast stretching applied with gamma correction. The first four columns represents the position of spot and the next two columns represents the red and green spot intensities and the last column represents the gene expression value which is a simple logarithmic value of red intensity divided by green intensity respectively.

1	5	5	17	18	39	103	-0.97117
2	5	23	17	17	12	28	-0.8473
3	5	40	17	18	60	140	-0.8473
4	5	58	17	17	17	34	-0.69315
5	5	75	17	17	41	60	-0.38077
6	5	92	17	18	75	42	0.579818
7	5	110	17	18	7	20	-1.04982
8	5	128	17	18	43	44	-0.02299
9	5	146	17	17	22	74	-1.21302
10	5	163	17	18	18	33	-0.60614
11	5	181	17	18	30	29	0.033902
12	5	199	17	18	30	31	-0.03279
13	5	217	17	17	16	20	-0.22314
14	5	234	17	18	14	26	-0.61904
15	5	252	17	18	19	28	-0.38777
16	5	270	17	17	3	19	-1.84583
17	5	287	17	18	12	26	-0.77319
18	5	305	17	17	16	26	-0.48551
19	5	322	17	18	12	14	-0.15415
20	5	340	17	18	24	26	-0.08004
21	5	358	17	18	8	16	-0.69315
22	5	376	17	18	57	143	-0.91979
23	5	394	17	17	14	21	-0.40547
24	5	411	17	18	28	91	-1.17865
25	22	5	17	18	54	93	-0.54362

Fig. 6 Data sheet for Spot Intensity Values

From Fig.3 and Fig.4 which represents the red and green spots intensities respectively for logarithmic transformation and contrast stretching with gamma correction, it is clear that for red spots with contrast stretching gives better result than logarithmic transformation but for green spots both the results are almost same..

1	5	5	17	18	39	100	-0.94161
2	5	23	17	17	13	29	-0.80235
3	5	40	17	18	59	135	-0.82774
4	5	58	17	17	17	34	-0.69315
5	5	75	17	17	42	62	-0.38946
6	5	92	17	18	85	46	0.61401
7	5	110	17	17	9	24	-0.98083
8	5	127	17	18	49	46	0.063179
9	5	145	17	18	25	80	-1.16315
10	5	163	17	18	18	33	-0.60614
11	5	181	17	18	39	34	0.137201
12	5	199	17	18	29	30	-0.0339
13	5	217	17	17	16	20	-0.22314
14	5	234	17	18	13	24	-0.6131
15	5	252	17	18	18	27	-0.40547
16	5	270	17	17	3	17	-1.7346
17	5	287	17	18	13	28	-0.76726
18	5	305	17	17	16	24	-0.40547
19	5	322	17	18	13	15	-0.1431
20	5	340	17	18	25	26	-0.03922
21	5	358	17	18	9	18	-0.69315
22	5	376	17	18	58	146	-0.92316
23	5	394	17	17	14	21	-0.40547
24	5	411	17	18	28	91	-1.17865
25	22	5	17	18	65	110	-0.52609

Fig. 7 Data sheet for Spot Intensity Values

VII. CONCLUSION

We have implemented and compared the different transformations such as logarithmic transformation, contrast stretching and gamma correction methods, to compute the spot intensity in microarray images, and found that each one of the transformation has its own set back and advantages. From this analysis, we conclude that it is better to apply logarithmic transformation with gamma correction for parameter gamma greater than 1 or to use contrast stretching for suitable value E for controlling the slop of the function which will give a better result or equivalent result as compared to traditional logarithmic transformation.

REFERENCES

[1] Angulo J. and Serra J. Automatic analysis of DNA microarray images using mathematical morphology. Bioinformatics, 19(5):553–562, 2003.

[2] Attwood T. K. and Parry-Smith D. J. Introduction to bioinformatics. Addison Wesley Longman Limited, Harlow, England, 1999.

[3] Bajcsy P. An overview of DNA microarray grid alignment and foreground separation approaches. EURASIP J. on Applied Signal Proc., Article ID 80163:1–13, 2006.

[4] Berrar D. P., Dubitzky W., and Granzow M., editors. A practical approach to microarray data analysis. Kluwer Academic Publishers, Dordrecht, 2003.

[5] Blekas K., Galatsanos N. P., et al. Mixture model analysis of DNA microarray images. IEEE Trans. on Med. Imag., 24(7):901–909, 2005.

[6] Katzer M., Kummert F., and Sagerer G. Methods for automatic microarray image segmentation. IEEE Trans. on Nano-Bioscience, 2(4):202–214, 2003.

[7] Rafael C.Gonzalez and Richard E.Woods, Digital Image processing, Second Edition.

[8] Schena M., Shalon D., et al. Quantitative monitoring of gene expression patterns with a complementary cDNA microarray. Science, 270:467–470, 1995.

[9] Yang Y., Buckley M., et al. Comparison of methods for image analysis on cDNA microarray data. Technical Report #584, Dep. of Statistics, UCB, Nov. 2000.