

Development of a New Software Package for Processing and Analyzing DNA Microarray Images

Jinho Choi

Taejon Christian International School, Daejeon, Republic of Korea
jchoi0223@gmail.com

Heejun Choi

Department of Electronics and Computer Science, Air Force Academy, Republic of Korea
hjchoi01@yahoo.com

Received 23 July 2010; Revised 8 September 2010; Accepted 20 September 2010

Microarray technology is an interdisciplinary technique that promises a revolutionary progress toward better health and improved quality of life. The paper focuses on the development of an efficient software package, equipped with already well-known methods; also some new methods are proposed that will allow the processing and analysis of thousands of genes on microarray images. The microarray analysis software package (called SmartArray), newly proposed in this paper verifies, through microarray analysis, dramatic changes in the mRNA, protein, and activity level in the rat retina during light deprivation, which have been demonstrated in previous biological experiments. The analysis results demonstrate that SmartArray can successfully find many changes in gene expression levels in each subarray and classify them according to their significance.

Categories and Subject Descriptors: Software & Applications [**Bioinformatics**]

General Terms: DNA microarray, gene expression, microarray analysis

Additional Key Words and Phrases: snake algorithms, template alignment, segmentation

1. INTRODUCTION

Recently, people have begun to turn their eyes to a new stage, which is expected to promise a better future with regard to our health [Gerhold et al. 1999]. This advanced technology is called bioinformatics. Bioinformatics incorporates the development of databases to store and search data, and statistical tools and algorithms to analyze and determine relationships between biological data sets, such as macromolecular sequences, structures, expression files and biochemical pathways [Westhead et al. 2002].

Copyright(c)2010 by The Korean Institute of Information Scientists and Engineers (KIISE). Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Permission to post author-prepared versions of the work on author's personal web pages or on the noncommercial servers of their employer is granted without fee provided that the KIISE citation and notice of the copyright are included. Copyrights for components of this work owned by authors other than KIISE must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires an explicit prior permission and/or a fee. Request permission to republish from: JCSE Editorial Office, KIISE. FAX +82 2 521 1352 or email office@kiise.org. The Office must receive a signed hard copy of the Copyright form.

Microarray technology is a good example of the incorporation of these technologies and needs the support of various interdisciplinary fields such as biology, chemistry, mathematics, engineering, etc. Microarray technology has brought many advantages in biological experimentations. For example, in the past, traditional experimental methods of biological assays such as gels and Northern blots required a large amount of experiment time to obtain a small volume of data, allowing for the study of only one or a handful of genes per experiment [Dozmorov et al. 2002; Schena 2003]. Also recent advances in DNA microarray technology have revolutionized the way scientists study genes, allowing for the study of thousands of genes, or even whole genomes, simultaneously.

Until now, many types of microarrays have been developed, and companies that developed these microarrays have also produced most of the software packages that analyze their own microarrays. The three most widely used microarray analysis packages are ImaGene of BioDiscovery, Inc. and GenePix Pro of MDS, Inc., which process color-based fluorescent microarray images; and AtlasImage of Clontech, Inc., which processes gray-scale-based nylon membrane microarrays [BioDiscovery 2006; Clontech 2007; MDS 2008]. Microarray analysis software packages are generally optimized to process certain types of microarrays, such as fluorescent microarrays or nylon membrane microarrays, and their performance is continuously improved, based on the result of biological experiments [White et al. 2005; Novikov et al. 2007].

AtlasImage software package processes nylon membrane microarrays. However, it is not appropriate for dealing with a variety of customized arrays with different array formats [Clontech 2007]. The microarray analysis software package (called SmartArray), proposed in this paper demonstrates an excellent performance in analyzing our customized nylon membrane microarrays, which fit the purpose of our biological experiments, as well as the existing nylon membrane arrays. Also, it is expected to accommodate upcoming customized arrays that will be used in future experiments, improving the performance of the gene data analysis algorithms.

This paper concentrates on studying methods and implementing a system for processing and analyzing DNA microarray images, which are very suitable for analyzing gene expression data [Duggan et al. 1999; Li et al. 2005]. The study of comparative gene expression by DNA microarray consists of the following steps [Hall 2001]:

- Step 1: Target preparation and probe preparation
- Step 2: Printing target onto a microarray
- Step 3: Labeling of probe samples
- Step 4: Hybridization of probe samples to a microarray
- Step 5: Scanning the hybridized array
- Step 6: Analyzing the scanned image

The microarray experimentation from step 1 to step 4 is based on a background of molecular biology. The major objective of this paper is to implement a system for processing and analyzing DNA microarray images, which is achieved in steps 5 and 6.

2. BACKGROUND

2.1 An Overview of Microarrays

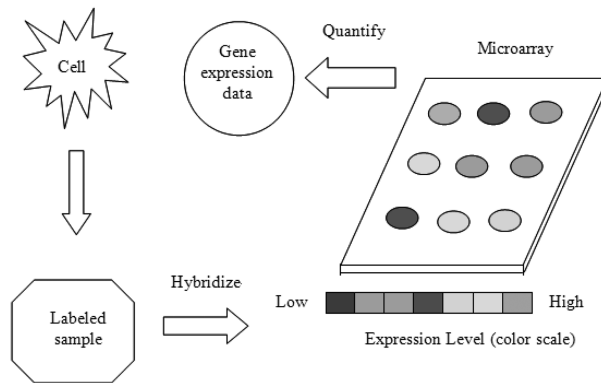


Figure 1. Overview of microarray experiment.

A microarray is a device that enables a rapid and quantitative analysis of gene expression patterns, patient genotypes, drug mechanisms, and disease onset and progression on a genomic scale. It makes it easier to determine exactly what genes are active in particular cells. For DNA microarrays, the strategy is to isolate the mRNAs made in the cells, and use these molecules as templates for making a cDNA mixture for base pairing with DNA from different genes.

A microarray is a small glass chip containing tens of thousands of DNA and protein sequences, and is used to investigate probe samples made by labeling mRNAs from cells, tissues, and other biological sources. The cognate sequences on the chip cause each spot to glow with intensity proportional to the activity of the expressed gene.

Microarray is a new scientific word originating from the Greek word *mikro*, meaning small [Schna 2003]. It is an ordered array of microscopic elements on a planar substrate, which allows the specific binding of genes or gene products. It is composed of many spots arranged in rows and columns. There are four criteria for qualifying a microarray as follows:

• **Criteria 1: Ordered**

The spots on a microarray need to be arranged in uniform size and spacing in rows and columns for automated and consistent processing.

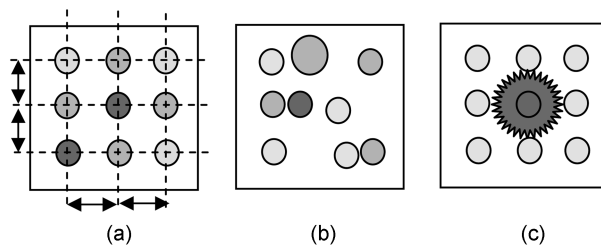


Figure 2. Good or bad example of an ordered array: (a) Uniform size and spacing, (b) Different size and irregular placing, (c) Signal blooming

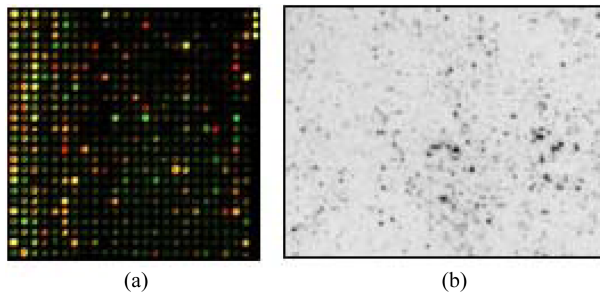


Figure 3. Microarray images obtained by various detection methods: (a) Fluorescent detection, (b) Radioactive detection.

- **Criteria 2: Microscopic**

Microscopic can be defined as any object that cannot be seen clearly without the use of a microscope. It is smaller than anything nominal, about 1 mm (1000 μm). Microarrays made by photolithography and the other semiconductor-based strategies typically produce elements in the range 15–350 μm .

- **Criteria 3: Planar**

A planar substrate such as glass, plastic, silicon or nylon is needed for parallel and unbending support. It also supports the automated production of microarrays and provides accurate scanning and imaging. Currently, glass is the most widely used material.

- **Criteria 4: Specific binding**

Specific binding refers to unique biochemical interactions between probe molecules in a solution and their cognate target molecules on the microarray. Binding specificity allows a gene or gene product to be analyzed quantitatively with a single microarray target element. Quantitative analysis is the process of measuring the number, amount, or concentration of the molecules present in a sample. Therefore, specific binding is a critical aspect of a microarray analysis.

Microarray images captured by different detection methods can be stored in the graphic file format and processed by a variety of methods. Figure 3 shows microarray images produced by different detection methods.

After that, template alignment is performed on a microarray image to find gene spots. Then, quantification and data analysis processes are subsequently followed.

2.2 Gene Expression Data Analysis

Microarray data are commonly composed of an $n \times m$ matrix, M where there are n rows for genes and m columns for samples. M_{ij} is the expression of gene i in sample j . Figure 4 shows a microarray data matrix of 5000 genes on six samples [Jagota, M. 2001].

At present, several types of microarrays from different companies or institutes are in use. Among them, the three most widely used microarrays are the cDNA microarray,

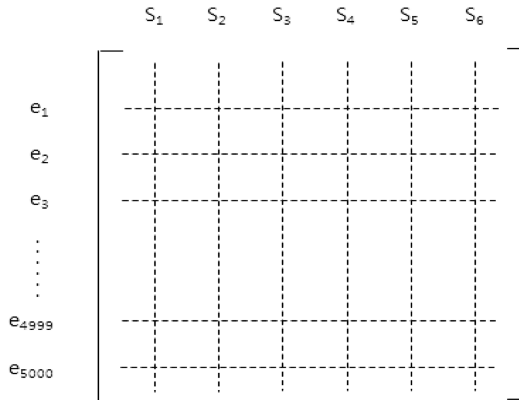


Figure 4. A microarray data matrix of 5000 genes on six samples.

nylon membranes, and oligonucleotide silicon chips. Accordingly, there exist a variety of data forms fitted to their own microarray types.

Before visualizing or analyzing microarray data, it is sometimes useful to process them as log-transformed or spotted array data. When comparing data from different gene microarrays it is necessary to perform microarray-to-microarray scaling to account for variation in the overall microarray intensity [Jagota, M. 2001].

$$e_{ij} = \log \frac{r_{ij}}{g_i} \quad (1)$$

where r_{ij} is the actual expression level and g_i is the expression level in a control.

The method suppresses an outlier in which $r_{i,j} \gg 0$ and $g_i \rightarrow 0$. On the other hand, it introduces a new outlier, when $r_{i,j} \rightarrow 0$ and $g_i \gg 0$.

Other types of processing include subtracting the mean from each component of a gene expression pattern. This is the most popular method due to its simplicity. The expression, e_{ij} is as follows:

$$e_{ij} = \frac{r_{ij}}{g_i} \quad (2)$$

where r_{ij} is the actual expression level and g_i is the expression level in a control.

In this expression, a problem may occur as it produces an outlier when $r_{i,j} \gg 0$ and $g_i \rightarrow 0$.

Nylon membranes and plastic arrays of Clontech, Inc. use the raw intensity or the adjusted value obtained by subtracting the background after measuring a raw intensity and a background value for each gene. The raw intensity or the adjusted value can be chosen by users [Clontech 2007].

3. DEVELOPMENT OF A MICROARRAY ANALYSIS SOFTWARE PACKAGE

The purpose of this paper is to develop a user-friendly and efficient system for the analysis of microarray images. The system includes a number of modules: a template alignment module, a quantification module and analysis module. It implements these

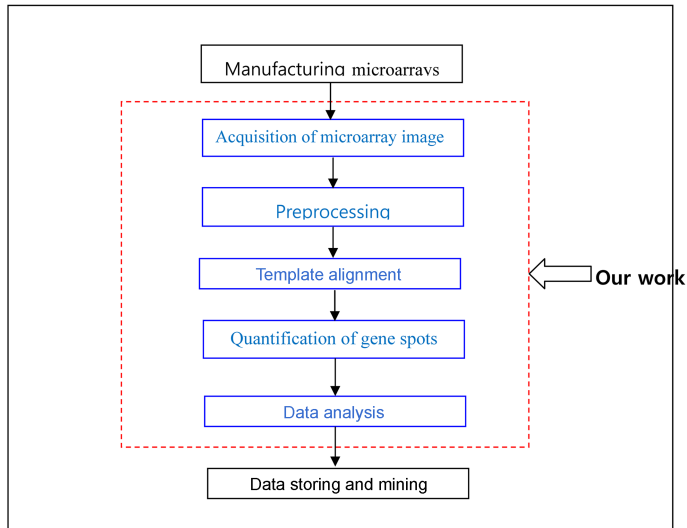


Figure 5. Overview of a microarray analysis system.

modules using existing methods, such as image rotation, threshold, etc., and newly proposed methods, such as DAC algorithm and new snake algorithm.

3.1 Preprocessing

At the first stage, microarray images captured by many different devices need to be preprocessed for successful quantification. This system includes many filters such as pixel-based operations and geometric transformation. The main factor that causes imprecise template alignment is when the microarray images are tilted. The first step of this stage is to correct tilted pixel images.

To get the deviation angle, α by obtaining two referred points such as (x_1, y_1) and (x_2, y_2) , the following geometric transformation equation is applied to all pixels in a microarray image. To correct tilted images, our system uses the following algorithm.

- Get the deviation angle, α by obtaining two referred points.

$$R = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

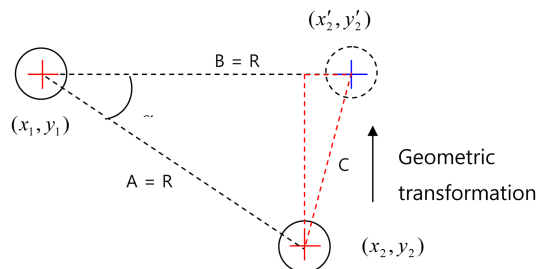


Figure 6. Derivation of a corrected point.

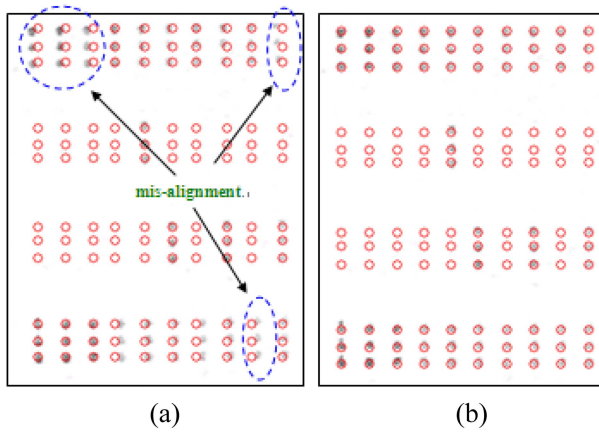


Figure 7. Example showing errors due to misalignment: (a) Tilted image, (b) Corrected image.

$$x'_2 = x_1 + R \tag{4}$$

$$C = \sqrt{(x_2' - x_2)^2 + (y_2 - y_2')^2} \tag{5}$$

$$C^2 = A^2 + B^2 - 2AB \cdot \cos \alpha \text{ by cosine rule} \tag{6}$$

$$\alpha = \alpha \cos((A^2 + B^2 - C^2)/(2AB)) \tag{7}$$

- Apply geometric transformation algorithm to all pixels in a microarray image [Foley et al. 1992].

$$\begin{bmatrix} x_2' \\ y_2' \end{bmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \cdot \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \tag{8}$$

Figure 7 shows a misaligned image due to tilting and a corrected image using the above algorithm.

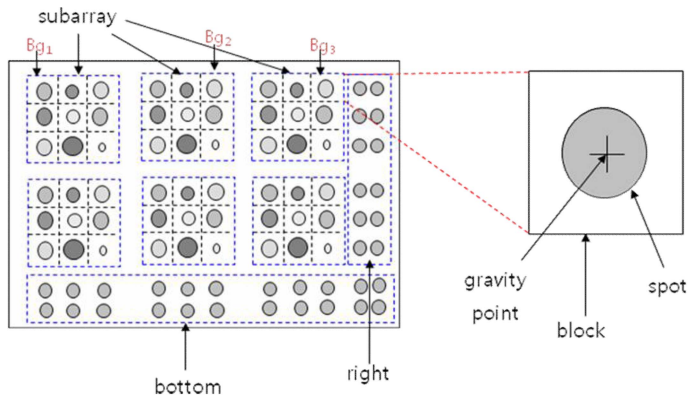


Figure 8. A microarray format with 2 × 3 subarray, 3 × 3 spots, and reference sections.

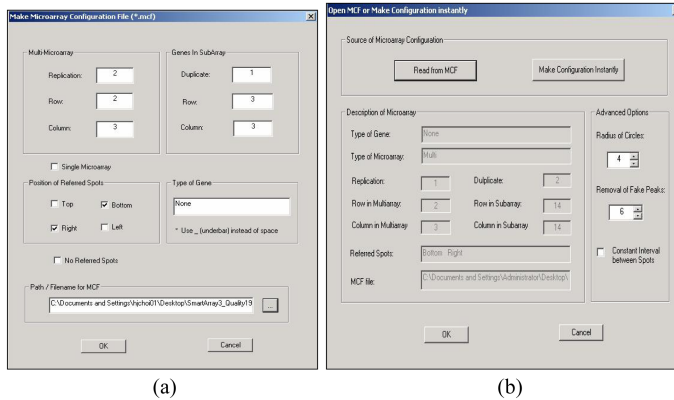


Figure 9. Dialog boxes for user-defined template: (a) Making a new template, (b) Opening an existing template.

3.2 Template Alignment

At the second stage, a new algorithm, called DAC (divide-and-conquer) algorithm, is applied to make templates for various types of microarray images by finding the gravity points at the centers of gene spots. At first, the DAC algorithm calculates averages for the row and the column directions, and puts them into arrays. Then, it decides signs which indicate many local maximums represented by upward slopes, downward slopes, or potential peaks. At last, it finds real peaks, called gravity points, and removes fake peaks simultaneously. The algorithm for automatic template alignment proceeds as follows:

- Make a user-defined template if a template does not exist.
- Calculate averages for the row and the column directions.

$$Average_i = \frac{\sum_{j=0}^{NumberOfPixel-1} X_j}{NumberOfPixel} \tag{9}$$

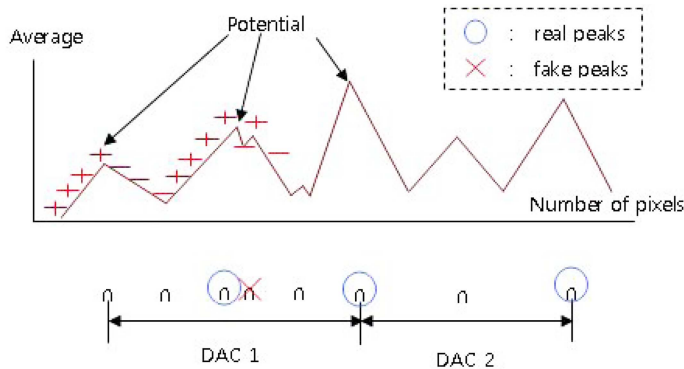
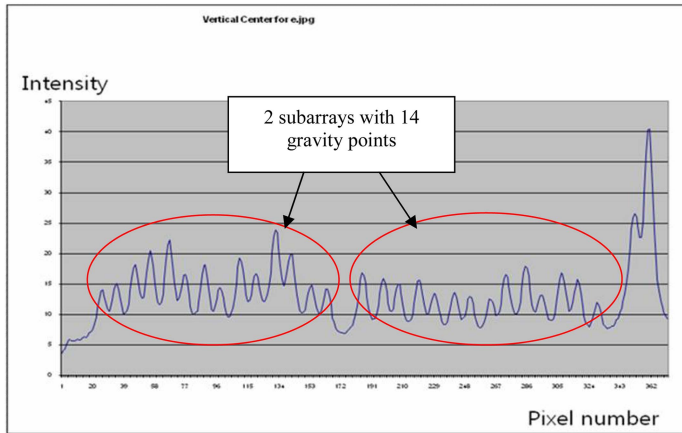
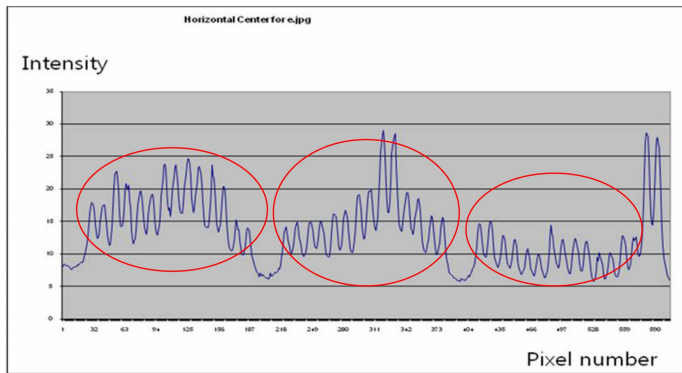


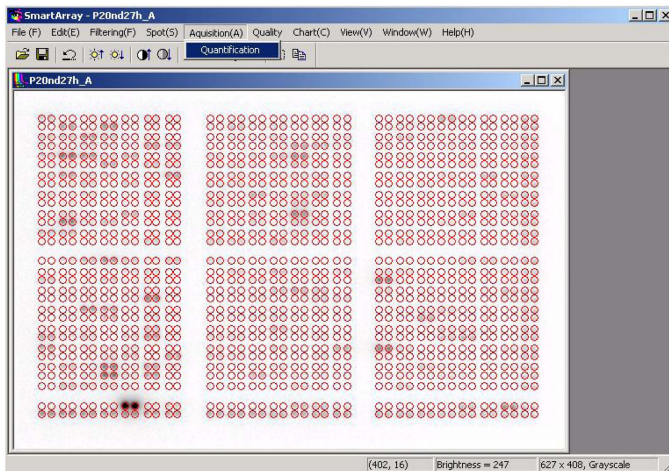
Figure 10. Finding real peaks among potential peaks and removing fake peaks.



(a)



(b)



(c)

Figure 11. Automatic template alignment: (a) Sketching averages in row, (b) Sketching average in column, (c) Finding gravity points with a circled form of template

where *NumberOfPixel* is the number of pixels in row or column scanned, *i* is the index of a row or column, *j* is the number of genes in a subarray, and X_j is the intensity of the pixel in a row or column.

- Set signs to the values in the average arrays and mark potential peaks.
 - If the present average is greater than the previous average, set a positive sign (+1).
 - Else, set a negative sign (-1).
 - If a sign changes from positive to negative, mark the potential peak.
- Find real peaks by applying DAC function to potential peaks with zero value and then remove the fake peaks simultaneously.
 - DAC (*L*, *R*)
 - DAC (*L*, *real_peak-1*)
 - DAC (*real_peak*, *R*)
 - where *L* is first index value and *R* is final index value in the average array.
 - DAC (*L*, *R*)
 - {– Find the local maximum between *L* and *R*.
 - Remove fake peaks by setting potential peaks near real peak to non-zero}
- Divide into various types of blocks using gravity points, called real peaks, as seen in Figure 10.
- If the gravity points are not correctly matched on a microarray image, repeat the process from the first step to the final step again after changing the parameter of “Removed Fake Peaks”.

As a result of repeated experiments, the template turned out to work well for other microarrays, as well as the Clontech’s nylon membrane microarray.

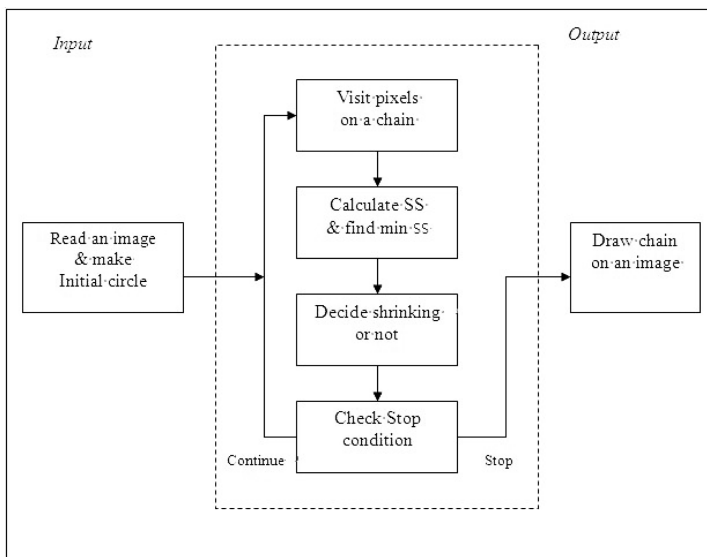


Figure 12. An overview of a new snake algorithm using the minimum SS condition.

3.3 Quantification of Gene Spots

At the third stage, quantification is needed to find the gene expression data for each gene. The gene expression data obtained by this process is used for the comparison of a control and a test microarray. In general, segmentation techniques such as threshold and snake algorithm are used for precise quantification. In our system, the new snake algorithm suitable for microarray images is used.

There are various approaches to implement shrinking operations. All of them use some conditions such as minimum SS (sum of squares), minimum energy (greedy algorithm), etc., in order to detect the boundary of the object [Sonka et al. 1999; Li et al. 2005]. In the following experiment, the new snake approach using the minimum SS is used. While each pixel on a chain is visited permanently, the SS of each pixel is calculated and each pixel moves toward the direction of the minimum SS. The stop check for a whole system occurs at the first pixel of a chain. If all pixels on a chain do not move anymore, the snake system stops and draws a chain as output on an original image.

Figure 13 shows the results of the new snake algorithm applied to different objects. This approach used in our system uses minimum SS and chain-shrinking operations.

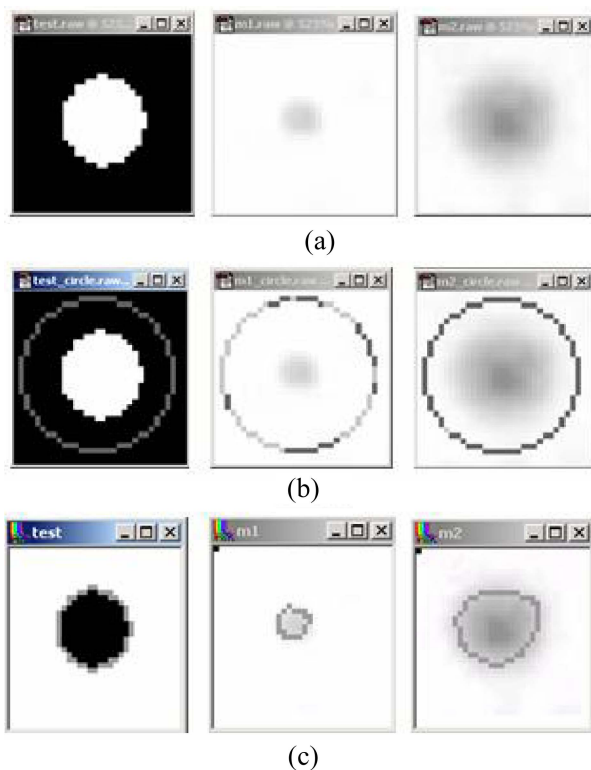


Figure 13. Segmentation of different types of spots using a new snake algorithm: (a) Spots before segmentation, (b) Initial circles surrounding the spots, (c) Implementation for different objects

3.4 Analysis of Gene Expression Data

At the fourth stage, a control microarray and a test microarray need to be compared to analyze gene expression data which are found through several previous processes. The control microarray contains non-disease genes while the test microarray includes some disease genes extracted from a living creature. The gene analysis between a control microarray and a test microarray is performed as follows:

If $Gene_Ex_{Test_Array} \geq Gene_Ex_{Control_Array}$,

$$Gene_Analysis_value_{i,j} = -1.0 \times \left(\frac{\sum_{i=1,j=1}^{n,m} (Gene_Ex_{i,j})_{Control_Array}}{\sum_{i=1,j=1}^{n,m} (Gene_Ex_{i,j})_{Test_Array} \times Norm_Coeff} \right) \quad (10)$$

Otherwise,

$$Gene_Analysis_value_{i,j} = -1.0 \times \left(\frac{\sum_{i=1,j=1}^{n,m} (Gene_Ex_{i,j})_{Control_Array}}{\sum_{i=1,j=1}^{n,m} (Gene_Ex_{i,j})_{Test_Array} \times Norm_Coeff} \right) \quad (11)$$

where i is the index for a subarray, j is a gene number in a subarray, $Gene_Ex_{i,j}$ is the value for each gene spot, and $Norm_Coeff$ is the normalization parameter.

The normalization parameter is used in case the brightness level of two compared microarray images is different. The normalization coefficient is multiplied to the test microarray image:

$$Norm_coeff = \frac{\sum_{i=1,j=1}^{n,m} (Gene_Ex_{i,j} - Bg_j)_{Test_Array}}{\sum_{i=1,j=1}^{n,m} (Gene_Ex_{i,j} - Bg_j)_{Control_Array}} \quad (12)$$

where i is a index for a subarray, j is a gene number in a subarray, and $Gene_Ex_{i,j}$ is the value for each gene spot.

On the above equation, when a gene expression of a test microarray is less than that of a control microarray, -1 is multiplied to the ratio of two gene expression data, in order to represent that the gene was downregulated. When the gene analysis value is -2 , a test gene is downregulated two-fold relative to its control gene; and when it is 2 , a test gene is upregulated two-fold relative to its control gene.

Once the analysis tool in the software package processes gene expression data, a result chart and a gene analysis report are created. Figure 14 shows a GAR file containing a heading and the results of a gene analysis.

In Figure 14, the first subarray (Row 1, Col 1) contains five changed genes that were downregulated. The second subarray (Row 1, Col 2) contains five upregulated

```

Case4 - Notepad
File Edit Format Help
# Microarray_No P20nd27h
# Type_of_Gene Rat
# Types_of_Gene_Ex Mean
# Top_Percentage 50
# RowInMultiarray 2
# ColInMultiarray 3
# RowInSubarray 14
# ColInSubarray 14
# Dup 2
# Avg_Dup_Genes 1
# Raw_Data 0
# Distinguish_spots 0.000000
# Discard_Dup_genes 0.000000

# Row 1 Col 1
Gene_No row col Ratio
55 4 13 -1.958333
13 1 13 -1.916667
11 1 11 -1.833333
97 7 13 -1.783784
83 6 13 -1.722222

# Row 1 Col 2
Gene_No row col Ratio
149 11 9 2.482759
143 11 3 2.416667
165 12 11 2.142857
135 10 9 2.100000
179 13 11 1.833333

# Row 1 Col 3
Gene_No row col Ratio
59 5 5 1.823529
47 4 5 -1.714286
11 1 11 1.636364
87 7 3 1.571429
33 3 5 -1.444444
    
```

Figure14. A GAR (gene analysis report) file.

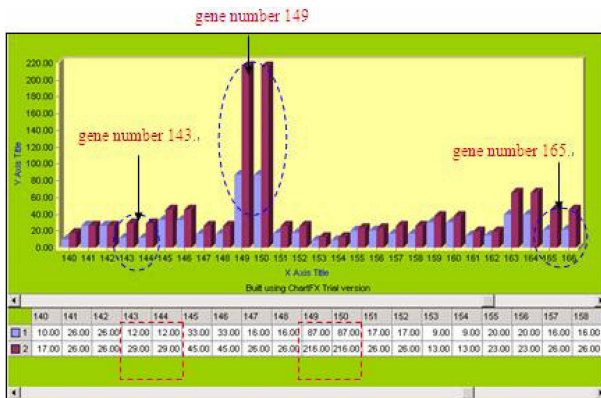


Figure 15. The drawing chart for subarray 2 (Row 1, Col 2).

genes that were significantly changed, especially gene numbers 149, 143, and 165. The third subarray (Row = 1, Col = 3) contains four upregulated genes and one down-regulated gene.

Using the draw chart function of our system, a variety of chart formats may be used to compare the gene expression data of subarrays. Figure 15 depicts how different each gene is and which genes have changed significantly. The consecutive genes were duplicated and averaged for comparison of the gene expression between two microarray images.

4. COMPARISON OF MICROARRAY ANALYSIS SOFTWARE PACKAGES

Microarray analysis software packages function in different methods, from detection to gene data analysis, depending on the type of microarray. Currently, MDS, Inc.'s

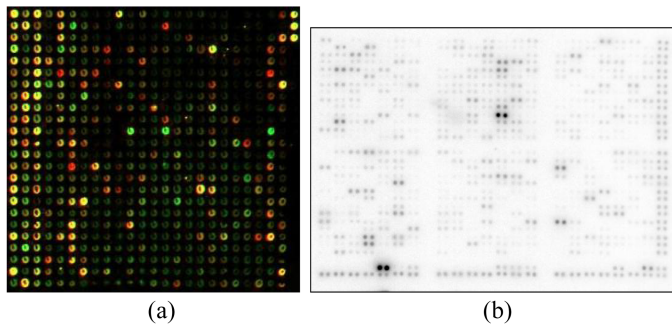


Figure 16. Different types of microarrays: (a) A color-based fluorescent microarray, (b) A nylon membrane microarray.¹

color-based fluorescent and Clontech's nylon membrane arrays are the most widely used for biological experiments.

Color-based fluorescent microarrays are scanned by a color laser scanner with two wavelengths, such as green 532 nm (Cy3, excited by green laser light) and red 635 nm (Cy5, excited by red laser light), and saved as 8-bit or 16-bit color images. Of the two color images, one is a control (otherwise called reference) array and another is a test array. Clontech's nylon membrane array is scanned and saved as two 8-bit grayscale images, one image for the control array and one image for the test array.

Currently, GenePix Pro and ImaGene are widely used to analyze color-based fluorescent microarrays, while AtlasImage is widely used to analyze Clontech's nylon membrane microarrays. Table I shows a comparison of the functionality of the 3

Table I. Comparison of the functionality of microarray analysis software packages.

	GenePix	ImaGene	AtlasImage	SmartArray
Company	MDS, Inc.	BioDiscovery, Inc.	Clontech, Inc.	Newly proposed
Type of microarray	Fluorescent	Fluorescent	Nylon membrane	Nylon membrane
Scanned image	8 or 16 bit color	8 or 16 bit color	8 bit grayscale	8 bit grayscale
Creation of template	Automatic	Automatic	Automatic	Automatic
Segmentation of spots	Threshold or morphological method	Threshold	Threshold	Threshold Or new snake
Quantification of gene data	Mean, median	Mean, median	Mean, median	Mean, median
Data analysis (normalization)	Present	Present	Present	Present

¹We appreciate Dr. Nigel G.F. Cooper for providing microarray images.

software packages mentioned above and the new software package (called SmartArray), proposed in this paper.

All the software packages in Table I have the necessary functions for analyzing microarray images, and there is no difference in functionality, except for the scanned image's format. As mentioned previously, processing and analysis methods (normalization, gene expression range, gene analysis value, etc.) vary in accordance with the type of microarray. Accordingly, GenePix and ImaGene were not suitable for our experiment; we used Clontech's nylon membrane microarray.

Also, AtlasImage is limited in dealing with customized nylon membrane microarrays as it lacks template flexibility. In particular, since AtlasImage 2.7 was introduced in 2007, Clontech, Inc. has not upgraded the software or provided much customer support. This means there is no guarantee that all software packages are suitable for all arrays [Novikov 2007].

Our SmartArray was developed to support the biological experiments we were performing, and it includes all functions that can process and analyze customized nylon membrane microarrays. The functions were implemented using the existing algorithms, such as image rotation, threshold segmentation, etc., and newly proposed algorithms, such as DAC algorithm and new snake segmentation algorithm. Also, the functions of SmartArray software package can be extended, with the support of biologists, to process and analyze color-based fluorescent microarrays.

5. CASE STUDY USING EXPERIMENTAL RESULTS

5.1 Background about Treatments

In previous biological experiments, it has been demonstrated that CaMKII and NMDA receptors undergo dramatic changes in their mRNA, protein and activity level in the rat retina during light deprivation. These changes are a result of developmental regulation. In order to better understand this process and find the possible pathway, a group of rats at postnatal 13 days were raised in complete darkness for 1 week and sacrificed at postnatal 20 days (P20dark), while another group of rats were raised in a normal diurnal light/dark cycle and sacrificed at p20 simultaneously as a control (P20normal).

Total RNAs were extracted from P20dark and P20normal respectively, and [³²P] dATP-labeled cDNA probes were synthesized and hybridized to Atlas arrays (nylon filter arrays of Clontech Inc.) containing 588 genes. Then, the autoradiograms obtained were analyzed. Next, a semiquantitative PCR was conducted to confirm the upregulation or downregulation of a number of genes. Accordingly, the purpose of this experiment was to find the genes that were significantly changed between two different environments: the presence and absence of light.

5.2 Experimental Results and Analysis

The microarrays are analyzed using gene expression data for three different methods: mean, median and max. The analysis tool has three user-defined parameters: "Normalization", "List from biggest changed spots", and "Selection of number of spots". In this case study, the normalization was activated by default. The parameter

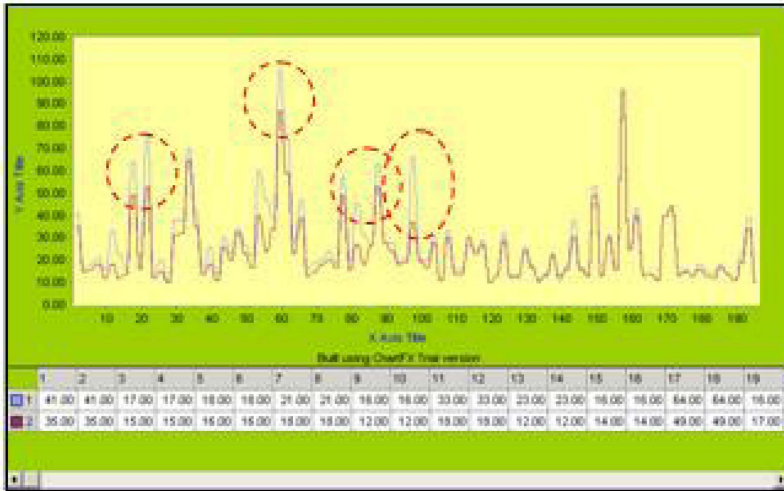


Figure 17. The drawing chart showing differences of two subarrays (for subarray 1).

Table II. Analysis results for subarrays 1, 2 and 3.

	Subarray 1			Subarray 2			Subarray 3		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
1	13	11	55	149	149	149	47	47	47
2	55	55	11	143	165	135	59	17	15
3	11	97	13	165	135	165	33	15	59
4	97	81	83	135	143	143	67	11	11
5	83	13	181	179	139	147	15	59	17
6	81	83	97	35	7	163	11	33	3
7	53	21	81	167	157	133	107	107	87
8	39	51	21	57	163	7	153	93	153
9	21	67	53	193	179	167	87	79	23
10	9	53	69	163	81	111	17	117	145

of “List from biggest changed spots” was selected for listing the most changed genes.

Tables II and III show the ten significantly changed genes for mean, median and max of each subarray. In the tables, bold numbers indicate genes were upregulated, and normal numbers indicate genes were downregulated.

For the mean of subarray 1, all genes were downregulated and gene numbers 55, 13, and 11 were approximately downregulated two fold. In addition, the analysis results for subarray 1 correspond to the circled areas of Figure 17. For median and max of subarray 1, all genes were downregulated.

For the mean of subarray 2, all genes except 57 were upregulated, and the first four genes were the most changed genes. Genes 149 (Row 11, Col 9) and 143 (Row 11, Col

Table III. Analysis results for subarrays 4, 5 and 6.

	Subarray 4			Subarray 5			Subarray 6		
	Mean	Median	Max	Mean	Median	Max	Mean	Median	Max
1	179	115	115	151	151	11	11	11	67
2	115	179	179	127	127	99	81	41	77
3	103	13	13	25	11	83	35	77	149
4	13	95	63	75	51	13	3	21	33
5	117	113	113	11	191	171	57	35	11
6	3	69	95	185	183	131	77	51	159
7	49	97	139	29	57	51	33	151	57
8	95	35	117	59	185	191	91	85	125
9	139	3	61	171	67	179	167	13	167
10	113	117	169	183	99	9	67	3	13

3) were 2.471264 and 2.333333, respectively, and the most changed genes in the analysis of the entire microarray. For the median and max of subarray 2, the first four genes were the most changed genes in the analysis of the whole microarray and their numbers were identical to the numbers of the most changed genes for the mean. In addition, up- or downregulation of genes were consistent regardless of the types of gene expression such as mean, median and max. For example, gene number 77 of subarray 6 was upregulated for mean, median and max as seen in Table III.

The case study analysis shows that the results of mean, median and max were consistent though the orders were different, as seen in the gene numbers of the below groups. The bold numbers mean they appear in the three groups and the underscored numbers appear in the two groups. In addition, the consistency of up- or downregulation was maintained in the analysis of whole the microarray regardless of types of gene expression such as mean, median and max. Accordingly, three types of gene expression can be used by users, and as discussed in the next chapter, the study of which gene expression is best for which case will be done comprehensively by computer scientists and biologists.

6. CONCLUSIONS

The microarray technology is an advanced one and is an interdisciplinary pursuit combining the fields of engineering and biology. The goal of this pursuit includes combining technology and natural sciences to promote better health and quality of life.

In this paper, we concentrated on developing a software package for analyzing microarray images, based on many methods of image processing and computer graphics. The results of our experiment show that our software package is suitable for analyzing microarray images, even though it leaves much room for improvement for more refined operations. In the near future, an advanced version of our software package is expected to be developed through our continuous work with the biologists.

REFERENCES

- BioDiscovery, Inc. 2006. *ImaGene User Manual*. <http://www.biodiscovery.com>
- Clontech, Inc. 2007. *AtlasImage 2.7 User Manual*. <http://www.clontech.com>.
- DUGGAN, D. J., BITTNER, M., CHEN, Y., MELTZER, P., AND TRENT, J. M. 1999. Expression profiling using cDNA microarrays, *Nature genetics supplement*, 21, 10–14.
- SCHEMA, M. 2003. *Microarray*, John Wiley & Sons.
- DEBOUCK, C. AND GOODFELLOW, P. N. 1999. DNA microarrays in drug discovery and development, *Nature genetics supplement*, 21, 48–50.
- DOZMOROV, L. AND CENTOLA, M. 2002. *Bioinformatics*. An associative analysis of gene expression array data.
- FOLEY, J., DAM, A. V., FEINER, S., AND HUGHES, J. 1992. *Computer graphics*, Addison-Wesley Publishing Company, Inc.
- GERHOLD, D., RUSHMORE, T., AND CASKEY, C. T. 1999. DNA chips: promising toys have become powerful tools, *Trends Biochemistry Science* 24, 5, 168–173.
- HALL, W. C. 2001. Online Microarray Analysis, *Master Thesis*, Univ. of Louisville.
- JAGOTA, M. 2001. Microarray analysis and visualization. *Bioinformatics*, BAY press.
- LI, Q., FRALEY, C., BUMGARNER, R. E., YEUNG, K. Y., AND RAFTERY, A. E. 2005. Donuts Scratches and Blanks: Robust Model-Based Segmentation of Microarray Images. *Bioinformatics*, 21, 2875–2882.
- Mathworks, Inc. 2010. *MATLAB Bioinformatics Toolbox 3 User Manual*. <http://www.mathworks.com>.
- MDS, Inc. 2008. *Axon GenePix Pro 7 Software User Guide*. <http://www.moleculardevices.com>
- NIH. 2010. *The MSCV Analyst's Toolbox User Guide*. <http://affylims.cit.nih.gov>.
- NOVIKOV, E. AND BARILLOT, E. 2007. Software package for automatic microarray image analysis (MAIA). *Bioinformatics* 23, 639–640.
- SONKA, M., HLAVAC, C., AND BOYLE, R. 1999. *Image processing, analysis, and machine vision*, PWS Publishing Co.
- WESTHEAD, D. R., PARISH, J. H., AND TWYMAN, R. M. 2002. *Bioinformatics*, BIOS Scientific Publishers.
- WHITE, A. M., DALY, D. S., WILLSE, A. R., PROTIC, M., AND CHANDLER, D. P. 2005. *Bioinformatics*, Automated Microarray Image Analysis Toolbox for MATLAB.



Jinho Choi He is a senior attending Taejon Christian International School in Daejeon, Korea. He completed the SAST (Science in Applied Science and Technology) program and RABS (Research Apprenticeship in Biological Sciences) internship at the University of Pennsylvania and Cornell University, respectively. Also, through an internship with the Korea Air Force Academy Department of Electronics and Computer Science, he joined the DNA Microarray Project at the Air Force Academy and performed biological data analysis. His research interests include DNA microarray analysis and immunology.



Heejun Choi He received his B.S. degree in Computer Science from the Korea Air Force Academy in Cheongwon, Korea, and his M.S. degree in Communication and Computer Engineering from POSTECH in Pohang, Korea. He received his Ph.D. degree in Computer Science and Engineering from the University of Louisville, Kentucky, USA. He has been a faculty member at the Department of Electronics and Computer Science of the Korea Air Force Academy in Cheongwon, Korea. His research interests include bioinformatics, image processing, and computer vision.