A COST-EFFICIENT DESIGN FOR MICROARRAY IMAGE SYSTEM

Ching-Yu Huang
Department of Computer Science, Kean University, Union, NJ 07083, USA
chuang@kean.edu

Abstract - Image based Microarray processing has been widely applied to the biotechnology field for identifying SNP variations. In order to reduce the cost for manufacturers, a flexible design for handling different layouts of containers and different numbers of spots is necessary to accomplish the need for different throughputs on the same microarray instrument. The key to making the microarray technology successful is spot recognition during the image processing. There are many factors that could affect the quality and alignment of spots. Both high accuracy and automation in spot intensity acquisition is required to have a correct result. Here, we present a cost-efficient design that will accommodate for different spot layouts with minimal required information.

Keywords: Microarray, layout, spots, controls, container, SNP

1. INTRODUCTION

Single Nucleotide Polymorphism (SNP) is a one of the most common forms of genomic variation within a population in which a single nucleotide — A, T, C or G [1]. SNPs are the most common type of genetic variation among biological species. In general, a microarray is a container that has many spots on its surface, with each spot printed by oligonucleotides or other biochemistry representing a SNP [2]. Microarray systems can be adopted for SNP genotyping and DNA sequencing. Each container is printed with different numbers of SNP and can experiment with different numbers of samples. There are many microarray systems using different types of container forms such as plate, chip, slides, glasses, etc.

Most of the microarray systems use fluorescent-based biochemical assays which use a CCD camera to capture the signals of the reaction results of SNP and samples. Therefore, image processing software is required to detect the spots and retrieve the intensity values as the raw data to determine the SNP types. In order to meet different needs of markets, a manufacturer may make different models of microarray systems for different type and size of containers. It is not efficient to implement and maintain different software for different types of instruments and containers.

Predefined spot position, human interaction, and other criteria are commonly assumed for the current methods. Several spot detection methods have been developed. A high-throughput and automated SNPs genotyping method based on gold magnetic nanoparticles array and dual-color single base extension has been designed by Li[3] to fulfill the increasing need for large-scale genetic research.

A dynamic spot detection method should be only based on hardware specifications with unknown spot locations so it has the flexibility to handle different spot, well, and image layouts. The presented method could accommodate different sizes of experiments and significantly reduce the cost.

2. DESIGN AND ANALYSIS

In order to let same instrument software handle different formats of microarray containers and spot layouts, there are some special specifications required during the design process. The spots can be classified into two types: regular SNP spots and control spots. A regular SNP spot is printed with certain oligonucleotide for a specific nucleotide. The genotype of a SNP spot is determined by its signals of oligonucleotide reaction with samples comparing to other SNP spots' signals [4].

2.1 Controls

Most of the microarray has control spots assigned. They can be classified into 3 types:

- **Positive Controls:** These spots should always have reaction and light up under certain laser wave. Multiple positive controls are required at certain positions for reorganizing and aligning other spots. Each image should have at least three positive

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.47

controls at corners to verify the alignments of column and row spots. For example the black spots at each corner in Figure 1 are the positive controls.

- **Negative Controls:** These spots should have no or very low reaction that is very low intensity. The hollow spots at right bottom corner in Figure 1 is the negative control. The signal values of negative controls sometimes are considered the background intensity and they are used to normalize the regular spots or determine the quality and reliability of the experiments. Each image should have at least one negative control.
- **Biochemical Controls:** These spots are designed to activate with certain biochemistry.

These controls are very important for spot detection and alignment, the normalization of the regular SNP spots' intensity values, and to verify each other. So, their locations should not be located too close to each other, in case some sections on the container are damaged or have problems.

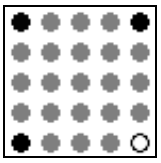


Figure 1. The black spots are the positive controls, the gray spots are the normal spots, and ring spots are the negative controls. An image has 5×5 spots.

2.2 Design Specifications

The designs of container layout, spot arrangement and the image configuration have great impact on the efficiency and cost of microarray experiments. There are 3 kinds of specifications can affect the design:

- **Instrument hardware**: camera resolution, sensor size and laser waves. The strength of the laser will greatly contribute to the spots' intensity values. Higher camera CCD resolution provides better quality of spot signals and can capture more spots at one shot. However, higher resolution and bigger sensors will also generate bigger image sizes and require more time to process the images to detect spots and extract the intensity values.
- **Container layout**: number of wells on the container, number of spots per well, spot diameter, distance between wells, and distance between spots. One well can be used by one sample or can be pooled with more than one sample. Therefore, if an experiment needs to have more than one sample/pool per run, it will be necessary to have multiple wells on a container. The more choice of different container layouts, the more flexibilities of experiments the system can provide.
- **Image layout**: the number of wells per image, and the number of image per containers, and controls' locations on the image. If a container is sliced into more than one image file, it will be necessary to have both positive and negative controls on every image because the strength of the laser could vary when CCD camera snapshot the signals of reactions each time. In order to capture all spots and wells, the container or camera might be moved internally by the instrument for every capture. The movement will cause alignment issue that the positive control spots won't be precisely at their ideal location on blueprint. That means the coordinates of the first spots will be different between images.

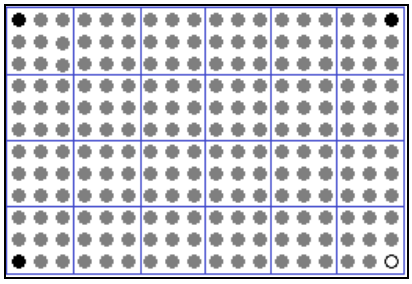
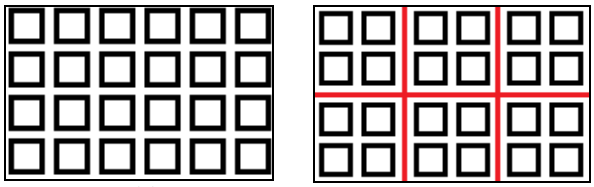


Figure 2. A container has 18×12 spots with 3 positive controls and 1 negative control. The spots are separated by the blue lines and grouped into 6×4 wells and each well will have 3×3 spots.

A container has 18×12 spots shown in Figure 2 are evenly distributed on 6×4 wells as shown in Figure 3(a). Therefore, each well has 3×3 spots. Table 1 shows if all 18×12 spots are captured on one image, only four positive and one negative control are required. If an image has 2×2 wells, the container should be equally captured into 3×2 images as shown in Figure 3(b). As discussed above, each image file should have eat least three positive and one negative controls. This design will need to have 24 controls. The 1×1 image per container will provide 20 more spots for genotyping experiments than 3×2 images per container, a 10.4% production increased. However, although same container captured into more images wastes more spots for experiments, it might provide more accurate spot detection and reliable results.

Table 1. Different layouts of a container with 18×12 spots (total 216) with different number of positive and negative controls.

# of images per container	# of positives	# of negatives	# of SNPs
1×1	3	1	212
3×2	18	6	192



(a) (b)

Figure 3. (a) A container has 6×4 wells in one image file. (b) An image has 2×2 wells and 3×3 spots per well.

3. SPOT DETECTION AND ALIGNMENT

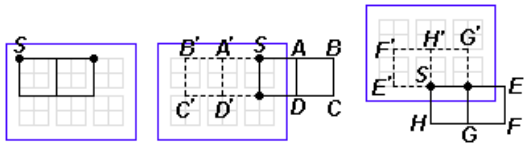
Once the spots on the container are captured and transformed into image files, the next steps will be to process each file that should include bipartition, spot detection and intensity values retrieval, and spot quality verification. In order to do further processing, the raw image needs to be bipartite to be a binary image - *background* pixel, and *object* pixel. There are many methods that have been discussed about how to do image segmentation and retrieve signals. Galinsky[5,6] proposed both Hexagonal and rectangular grids to do automatic registration of microarray images. Wang[7] described Matarray to do spot detection, segmentation, intensity acquisition, and quality determination. Angulo[8] presented an automatic no-supervised set of morphological operators for a fast and accurate spot data extraction. This paper focuses on how to detect the spots on dynamic layouts.

Since the spot locations on the containers and image files might be off more than a spot distance, without knowing where the spots are located on the images, the positive controls are required to light up to assist the automatic spot detection's accuracy.

3.1 Spot Detection

For each binary image, all the recognized spots are stored in ΣS_{ij}. A mesh with $N_{col} \times N_{row}$ virtual spots is built to validate each spot, and the distance between 2 virtual adjacent spots is D_k where N_{col} and N_{row} represent the numbers of spots in every *column* and *row*, respectively. Each ${}_{ab}S_{xy}$ location is assumed at (a, b) of mesh centered at (x, y) on the image. Its virtual spot's ${}_{ab}S_{xy}$ can be derived by: $x = a \times D_k + i$, $y = b \times D_k + j$, where $a=0$ to $N_{col}-1$ and $b=0$ to $N_{row}-1$. If x or y is over the image border, both a and b need to be adjusted as $a = l_a - a$, $b = l_b - b$, where l_a and l_b are the last non-adjusted positions. Every ${}_{ab}S_{xy}$ will be compared to against ΣS_{ij}. ${}_{ab}S_{xy}$ is considered as a true spot**,** if its ***Distance***$({}_{ab}S_{xy}, \Sigma S_{ij})$ is less than a half of the spot diameter. For a valid mesh, the number of true spots should be greater than ***Minimum***(N_{col}, N_{row}), and all its spot will be set as true. This rule will filter out the noise-like spot.

Figure 4 illustrates different situations with 3×2 wells per image. A mesh with 3×2 virtual spots will be constructed, and S indicates the first *SE*. Figure 4(a) is a simple mesh with 2 validated *SE.* In Figure 4(b), the mesh is partially out of the image and virtual spots A, B, C and D are adjusted as A', B', C' and D' respectively. Figure 4(c) shows the complication. F is at (3, 2) and the last non-adjusted position is at (2, 1). Since F is out of image, the adjusted F' will be at $(2-3+S, 1-2+S)$.



(a) (b) (c)

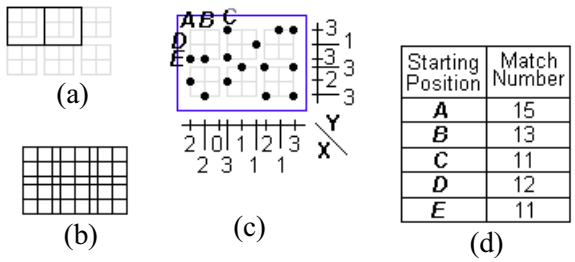
Figure 4. (a) The simple spot alignment mesh. (b) A mesh with adjusted virtual spots. (c) A complicate mesh.

Once a spot is detected, similar process will be applied again with local threshold to detect if any other spots have low signal and their shapes look not like spots on binary images with the global threshold. Figure 4(b) and (c) show that this method can ensure the spots in other 4 wells can still be automatically processed by only 2 wells with the validated spot (shown in right and bottom wells, respectively). During the process, all ${}_{ab}S_{xy}$ positions will be collected into $\Sigma {}_{ab}S_{xy}$ for both adjusted and non-adjusted virtual spots.

3.2 Positive Spot Recognition

Since the positive controls might not light up, a *Multiple Grid Mesh* (*MGM*) is built to detect the root spot – the left and upper corner spot of each well for

each image. As shown in Figure 5 (a) shows that the same size of mesh as in spot recognition will be applied on each spot to expend as a multiple mesh as shown in (b). From $\Sigma_{ab}S_{xy}$, the histogram for number of *SE* can be derived along *X* and *Y* axis as ***Hist***(X_i) and ***Hist***(Y_j) where ***Hist*** represents the amount of spots on the *i* row and *j* column. The *MGM* will then be located on each (***Hist***(X_i), ***Hist***(Y_j)). Then, a match number can be calculated between true spots and the virtual spots of *MGM.* The spot S_{ij} with the highest match number is considered as the root of this image. Therefore, the starting position of each well can be computed from the mesh. Figure 5(c) is showing as an example. The ***H***(X_i) and ***H***(Y_j) numbers are shown in bottom and right respectively. After the multiple grid mesh is applied, the match numbers for the sample spot are shown in (d). The virtual spot ***A*** is picked as the image root position.



Starting Position	Match Number
A	15
B	13
C	11
D	12
E	11

Figure 5. (a) The basic mesh with 3×2 wells. (b) The *Multiple Grid Mesh* is expended from the basic mesh for every one of the 3×3 spots inside a well. (c) Spot number histogram is calculated along both X and Y direction. (d) The spot match number on (c) with the *MGM* starts at different positions.

4. CONCLUSIONS

SNP is commonly used to detect DNA variations. With the number of samples and DNA dramatically increasing, the nanotechnology can print millions of spots on a microarray for a large amount of experiments. However, small clinic offices might only need to run low-throughput experiments compared to big labs that run high-throughput experiments, in terms of the sample sizes and the number of SNPs. It will reduce the cost a lot if an instrument can take different numbers of samples and SNP and if its software can automatically process different layouts with different numbers of spots. The spots on the images must be detected and assigned with accurate positions. To achieve this goal, a fully dynamic spot recognition method is presented in this paper for high throughput microarray image processing. Only few parameters are required from the hardware or kits. In addition, our model has the flexibility to process different kinds of spot layouts simultaneously while dynamically feeding the parameters. No human interaction is needed.

5. REFERENCES

[1] Michael C. Ellis, “Spot-On” SNP Genotyping *Genome Research*. 2000. 10: 895-897

[2] Ramesh Hariharan, "The Analysis of Microarray Data," *Pharmacogenomics*, 4(4), pp. 477-497, 2003.

[3] Song Li, Hongna Liu, Yingying Jia, Yan Deng, Liming Zhang,Zhuoxuan Lu, and Nongyue He1 “A Novel SNPs Detection Method Based on Gold Magnetic Nanoparticles Array and Single Base Extension,” *Theranostics*. 2012; 2(10): 967–975.

[4] Huang CY, Studebaker J, Yuryev A, Huang J, et al. “Auto-validation of fluorescent primer extension genotyping assay using signal clustering and neural networks.” *BMC Bioinformatics*, 2004, 5:36.

[5] Vitaly L. Galinsky, "Automatic registration of microarray images. II. Hexagonal grid," Bioinformatics, 19(14), pp. 1832-1836, 2003.

[6] Vitaly L. Galinsky, "Automatic registration of microarray images. I. Rectangular grid," Bioinformatics, 19(14), pp. 1824-1831, 2003.

[7] Xujing Wang, Soumitra Ghosh and Sun-Wei Guo, "Quantitative quality control in microarray image processing and data acquisition," *Nucleic Acids Researc*h, 29(15), 2001.

[8] Jesus Angulo and Jean Serra, "Automatic analysis of DNA microarray images using mathematical morphology," *Bioinformatics*, 19(5), pp. 553-562, 2003.