

균일 격자 구조 탐색을 이용한 마이크로어레이 반점 주소 결정 알고리즘

(An Algorithm for Spot Addressing in Microarray using Regular Grid Structure Searching)

진 희 정 * 조 환 규 **

(Hee-Jeong Jin) (Hwan-Gue Cho)

요 약 최근 마이크로어레이 실험기술의 개발로 인해서 생물학자들은 한꺼번에 수천 혹은 수만 개의 유전자 발현실험이 가능하게 되었다. 마이크로어레이를 이용한 유전자 발현 패턴 분석에 필요한 이미지의 분석 작업은 사용자의 많은 수작업이 필요하며, 올바른 결과를 얻기 위해서 많은 주의가 필요하다. 그러므로 사용자의 수작업을 최소화하고 정확한 발현결과를 얻기 위해서 마이크로어레이 이미지의 자동 분석 방법이 필요하다. 일반적으로 마이크로어레이 데이터는 반점(spot) 위치의 변동이나 모양, 크기가 고르지 않는 것과 같은 다양한 문제로 인하여 자동 분석이 어렵다. 특히 블록과 반점의 주소를 결정하는 것은 마이크로어레이 분석 중 어려운 단계이며, 대부분 상용 프로그램에서는 수작업을 통해서 해결하거나, 수작업이 필요한 반자동시스템을 이용하고 있다. 본 논문에서는 균일 격자(regular grid) 구조 탐색을 이용하여 새로운 블록과 반점의 주소를 결정하는 알고리즘을 소개한다. 본 알고리즘에서는 입력된 반점들의 중심점을 이용하여, 균등 일직선 서열(equally spaced and collinear sequence)을 생성하고 이를 통하여 이미지의 기울기와 단위길이를 계산한다. 계산되어진 기울기와 단위길이를 이용하여 가상점을 허용한 균등 일직선 서열을 다시 생성하고, 이를 이용하여 마이크로어레이의 주소를 결정한다. 실험 결과 다양한 실험 데이터에 대하여 매우 안정적이며, 신뢰성이 높은 결과를 얻을 수 있었다. 본 알고리즘에 대한 자세한 정보는 <http://jade.cs.pusan.ac.kr/~autogrid>에 정리되어 있다.

키워드 : 생물정보학, 이미지 처리, DNA 마이크로어레이, 균일 격자 구조

Abstract Microarray is a new technique for gene expression experiment, which has gained biologist's attention for recent years. This technology enables us to obtain hundreds and thousands of expression of gene or genotype at once using microarray. Since it requires manual work to analyze patterns of gene expression, we want to develop an effective and automated tools to analyze microarray image. However it is difficult to analyze DNA chip images automatically due to several problems such as the variation of spot position, the irregularity of spot shape and size, and sample contamination. Especially, one of the most difficult problems in microarray analysis is the block and spot addressing, which is performed by manual or semi automated work in all the commercial tools. In this paper we propose a new algorithm to address the position of spot and block using a new concept of regular structure grid searching. In our algorithm, first we construct maximal ϵ -regular sequences from the set of input points. Secondly we calculate the rotational angle and unit distance. Finally, we construct ϵ -regularity graph by allowing pseudo points and then we compute the spot/block address using this graph. Experiment results showed that our algorithm is highly robust and reliable. Supplement information is available on <http://jade.cs.pusan.ac.kr/~autogrid>.

Key words : bioinformatics, Image processing, DNA microarray, regular grid structure

1. 마이크로어레이 소개

분자생물학적인 실험기술(High-throughput technology)과 로봇 기술의 발달로 인해 등장하게 된 DNA 마이크로어레이(microarray 또는 microchip)라는 새로운 개념의 기술 도입으로 하나의 칩(chip)상에서 유전체

* 학생회원 : 부산대학교 전자계산학과

hjjin@pearl.cs.pusan.ac.kr

** 정 희 원 : 부산대학교 전기전자정보컴퓨터공학부 교수

hgcho@pusan.ac.kr

논문접수 : 2004년 2월 13일

심사완료 : 2004년 6월 22일

(genome)의 발현 양상을 탐색하거나, 동시에 수천 개의 유전자들 간의 상호작용을 관찰할 수 있게 되었다. 마이크로어레이 기술은 짧은 시간에 엄청난 양의 데이터를 생성할 수 있는 가능성을 지니고 있어 생명과학 분야의 새로운 연구 도구로 각광을 받게 되었다. 마이크로어레이는 기존의 대표적인 유전공학 방법인 southern과 northern blot, SNP(Single Nucleotide Polymorphism) 탐색, 그리고 DNA sequencing 등 유전 공학의 여러 실험들의 기능을 할 수 있다. 기존의 대표적인 유전공학 방법과 마이크로어레이와의 대표적인 차이점은 한 번에 수백, 수천 개의 유전자를 짧은 시간 내에 검색할 수 있다는 것이다. 이러한 마이크로어레이에 대한 분석은 현재 생물정보학 분야에서 가장 많이 연구되고 있는 분야 중 하나이다.

1.1 마이크로어레이의 원리

1.1.1 유전물질의 종류에 따른 분류 및 활용

마이크로어레이에는 슬라이드에 유전물질을 붙여서 사용하는데, 붙이는 유전물질의 종류에 따라 Oligonucleotide 칩과 cDNA 마이크로어레이 칩으로 나눌 수 있다. Oligonucleotide chip은 20~50개의 nucleotide로 이루어진 DNA probe를 집적한 것으로, 4개의 nucleotide(A,G,C,T)의 조합으로 이루어진 probe와 시료 DNA를 합성(hybridization)시키면 둘의 염기 서열 일치 정도에 따라 시그널 크기가 변화된다. 시그널의 변화를 통하여, DNA 염기 서열과 돌연 변이된 염기 서열을 분석할 수 있다. cDNA 칩은 이미 밝혀진 ORF(open reading frame) 또는 EST를 칩에 집적하는 것으로, 수백 base pair 이상의 DNA 서열을 집적한다. 생체 내 mRNA를 역전사(reverse transcription) 효소로 cDNA

를 합성하여 위 ORF와의 합성시키면 그 발현 양에 따라 시그널의 크기가 변화된다. 이를 통하여, 특정 유전자의 발현 정도 분석할 수 있다. 그림 1과 같이 두 개의 다른 환경에서 얻은 세포들로부터 mRNA를 추출하여 이들을 역전사 시킬 때 각각 다른 색깔의 형광 물질을 염기를 집어넣어 빨간색(Cy5)이나, 녹색(Cy3)을 띠는 cDNA를 합성한다. 이와 같이 합성된 두 개의 cDNA를 똑같은 양으로 섞어서 하나의 cDNA 마이크로어레이 칩에 결합시킨다[1-3]. 마이크로어레이 실험 결과 이미지의 형광 정도는 그 유전자의 발현 정도를 알려주는 것으로 이들 정보를 분석하는 것이 DNA 마이크로어레이 칩 분석 시스템이다.

1.1.2 제작 방법에 따른 분류 및 활용

마이크로어레이는 제작방법에 따라 핀 마이크로어레이(pin microarray), 잉크젯(inkjet), photolithography, electronic array로 나뉘어 진다. 핀 마이크로어레이는 1995년 Stanford 대학 생화학과에서 처음 개발되었으며, 약 2~3천개의 유전자를 1cm²안에 붙일 수 있다. 처음에는 유전자의 발현 측정의 목적으로 cDNA를 붙여놓은 칩을 만들었기 때문에 cDNA 마이크로어레이라고 불렀지만, 지금은 돌연변이를 측정할 수 있도록 oligonucleotide를 붙인 칩도 제작되고 있다. 잉크젯 마이크로어레이는 잉크젯의 원리를 이용하여 칩을 제작하는 것으로, 핀 마이크로어레이와 거의 비슷하다. 다만 핀 대신에 컴퓨터 잉크젯 프린터에서 사용하는 것과 같은 원리의 카트리지(cartridge)를 사용한다. Photolithography 칩은 미국의 실리콘밸리에 있는 Affymetrix 회사에서 개발한 기술로서, 컴퓨터 칩을 만들기 위해서 사용하는 photolithography라는 기술을 사용하여, 수만

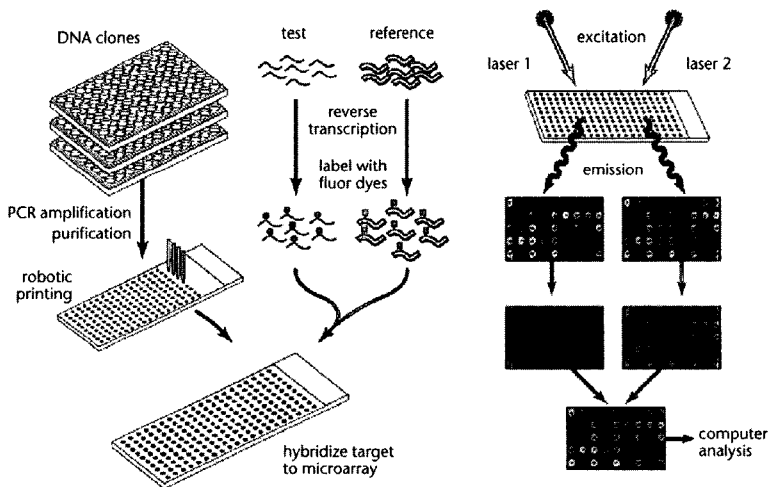


그림 1 cDNA 칩의 제작 과정과 실험과정[1]

표 1 마이크로어레이 제작 기술 분류 : 마이크로어레이는 제작 기술에 따라, 핀 마이크로어레이(pin microarray), 잉크젯(inkjet), photolithography, electronic array로 나뉘어 진다.

마이크로어레이 제작 기술	칩 종류	특징	제작사
pin microarray	cDNA oligonucleotide	<ul style="list-style-type: none"> • pin을 이용한 micro dotting • 1995년 stanford 대학에서 개발 	<ul style="list-style-type: none"> • Hyseq • Incyte
inkjet	cDNA oligonucleotide	<ul style="list-style-type: none"> • inkjet 원리를 이용한 micro dotting 	<ul style="list-style-type: none"> • Incyte
photolithography	oligonucleotide	<ul style="list-style-type: none"> • Photolithography를 이용한 oligonucleotide 직접 합성 	<ul style="list-style-type: none"> • Affymetrix
electronic array	oligonucleotide	<ul style="list-style-type: none"> • 전기를 이용한 oligonucleotide addressing 	<ul style="list-style-type: none"> • Clinical Micro Sensors • Nanogen

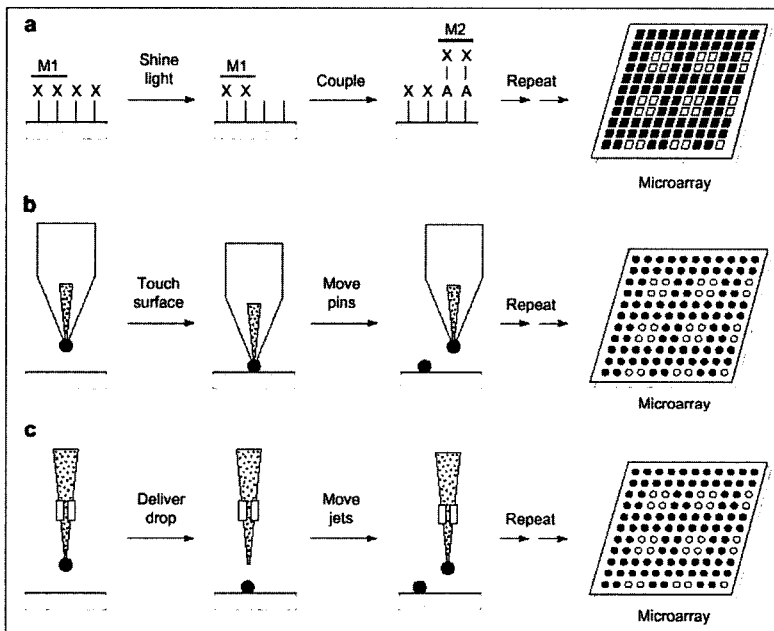


그림 2 마이크로어레이 제작 방법에 따른 분류 : (a) photolithography, (b) pin microarray, (c) inkjet 방법

개의 다른 염기들을 하나의 유리위에서 직접 합성하는 것이다. 초기에는 $1.28cm^2$ 내에 65,000개의 다른 oligonucleotide를 가진 칩을 만들었고, 지금은 400,000개의 다른 oligonucleotide를 가진 칩을 만들고 있다. 마지막으로 electronic array는 DNA가 (-)전하를 띠는 성질을 이용하여 칩 표면에 있는 특정 위치에 (+)전기를 넣어서 그 위치에 원하는 유전자를 붙게 만드는 방법이다. 이 원리를 이용한 칩이 미국의 Nanogen에서 개발되었고, 지금은 10,000개의 DNA를 이러한 방법으로 붙일 수 있는 칩이 개발되었다. 표 1과 그림 2는 제작 방법에 따른 마이크로어레이의 종류를 정리한 것이다.

1.2 마이크로어레이의 구조

마이크로어레이 실험을 통해서 각 형광물질에 따라 다른 이미지를 생성하게 되는데, 일반적으로 16bit 회색

조 TIFF 이미지를 생성한다. 보통 마이크로어레이 실험 시에는 Cy3, Cy5 두 개의 형광물질을 사용하기 때문에, 녹색과 적색 두 채널에 해당하는 두 개의 16bit 회색조 TIFF 이미지를 생성한다. 이 두 개의 회색조 TIFF 이미지를 각각 분석할 수도 있고, 두 개의 이미지를 RGB (B=0)에 해당하는 색으로 매핑해서 하나의 이미지로 변환하여 분석할 수도 있다. 그림 3은 두 개의 회색조 이미지를 하나로 합한 모습이다.

이들 이미지의 가장 큰 특징은 격자 모양으로 구분되어 있다는 것이다. 마이크로어레이 이미지는 격자 모양으로 구분되어 있어, 격자 내부에 존재하는 데이터를 각각 구분하여 분석하기가 용이하다. 일반적으로 격자 구조를 가진 이미지 데이터를 분석하는 경우에는 격자 구조를 결정하는 것을 먼저 수행하여야한다. 이런 과정을

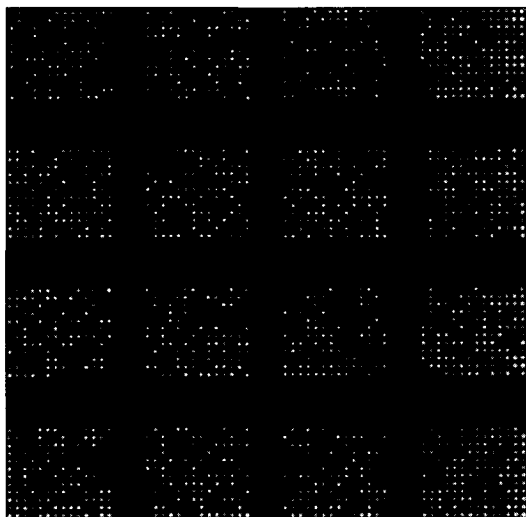


그림 3 마이크로어레이 실험을 통해서 생성된 이미지의 예 : 두 개의 녹색과 적색의 회색조 이미지를 RGB에 매핑해서 하나의 이미지로 변환한 것이다. 본 이미지는 4×4의 블록들과 각 블록 내에 14×12개의 반점으로 구성되어 있다.

일반적으로 그리(griding), 또는 주소 할당(addressing)이라 부르고, 이미지 데이터 분석의 기초가 되므로 아주 중요한 단계라고 할 수 있다. 한번 잘못 계산된 격자의 위치는 결과적으로 잘못된 분석 결과를 도출하기 때문에 정확한 격자 위치 계산이 필요하다. 하지만, 일반적으로 격자 구조를 가진 이미지 데이터를 실제 분석해보면, 격자 구조라는 것이 이미지 상에서는 하나의 그리드 셀과 다른 그리드 셀 간의 단위거리가 모두 똑같지 않고 종축과 횡축에 대해서 각각 평행하지 않고 약간씩 기울어져 있는 경우가 대부분이다. 그림 4는 마이크로어레이의 격자구조를 나타낸다.

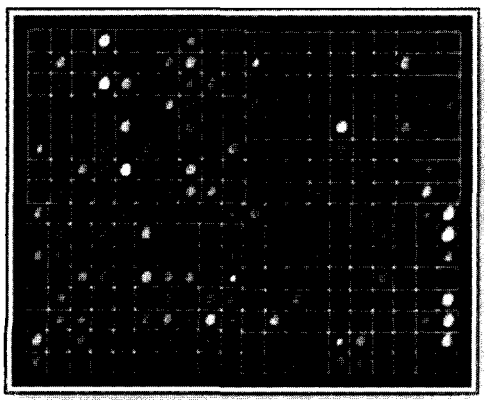


그림 4 마이크로어레이의 격자 구조

이런 마이크로어레이 이미지가 이상적인 경우일 경우 아래와 같은 다섯 가지 특징을 지니게 된다.

- (i) 모든 블록의 크기가 일정하다.
- (ii) 블록 간의 간격도 일정하다.
- (iii) 블록 안의 반점의 위치는 그 반점이 속한 행과 열의 직선의 교점위에 있어야 한다.
- (vi) 반점의 모양은 완벽한 원형이며, 모든 반점들의 크기는 동일하다.
- (v) 격자의 위치는 실험에 사용한 모든 슬라이드들에 대한 이미지들 안에서는 동일하다.

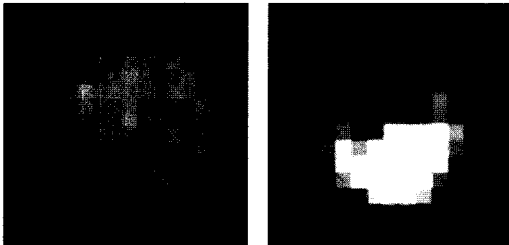
하지만 실제 실험 결과로 나오는 마이크로어레이의 이미지는 이러한 조건들을 만족하지 않기 때문에, 마이크로어레이 이미지 분석을 완전 자동화하기가 매우 어렵다. 최근 자동 마이크로어레이 이미지 분석 시스템들이 소개되기 시작했지만[4], 마이크로어레이 이미지의 분석이 자동화되기 위해서는 이미지의 상태가 좋고, 일정 비율이상의 유전자들이 발견되어 있어야 한다. 마이크로어레이 실험 시에 발생할 수 있는 이러한 문제점을 알아보고 이미지 분석시의 자동화 기능과 그 결과의 정확성을 알아보기 위해서 마이크로어레이의 품질을 측정하는 방법에 대한 연구가 진행되고 있다[5]. Kim 등의 연구[5]에서는 각 마이크로어레이 이미지의 몇 가지 통계량을 이용하여 이미지들 사이의 품질을 비교하고, 분석된 마이크로어레이 이미지 결과에 대하여 그 품질을 측정하는 5가지 품질평가함수를 정의하여 각각의 마이크로어레이 실험에 대한 품질을 정량화하여 나타내었다. 그리고 전체 실험과 한 이미지에 대한 품질을 평가하기 위해서 메타그리드를 이용하여 이미지내의 블록들과 반점들 간의 간격에 대한 통계량을 측정하고, 반점들의 발현 값에 대한 정확도를 측정하기 위한 품질평가함수들을 정의했다. Kuklin[6] 등과 Hautaniemi[7] 등의 연구 결과를 적용하여 잡음비율, 반점의 roundness를 하나의 품질 측정요소로 결정하고, 반점의 자동 격자 주소에 대한 가능성을 측정하기 위해서 반점배열의 균일화 규칙을 추가하였다. 그림 5는 마이크로어레이 이미지에서의 여러 오류를 보여준다. (a)와 (b)는 시그널 잡음에 대한 것으로, (a)는 이상적인 반점으로 시그널 잡음이 거의 없는 것을 볼 수 있다. (b)는 시그널 잡음을 포함하고 있는 반점의 예이다. (c)는 마이크로어레이 배정에 잡음이 포함된 경우로, 이러한 잡음은 반점의 발현강도를 측정할 때 많은 오류를 나타낼 가능성이 높다. (d)와 (e)는 반점의 roundness의 문제를 보여준다. 칩 실험 시에 정확한 spotting작업이 이루어지고 발현실험이 정확하게 이루어질 경우 반점의 모양은 이상적인 원형의 모습을 나타낸다. 하지만, spotting 시에 문제가 발생하거나 실험 때 염료의 양 등에 문제가 발생하면 그림 5의 (e)와



(a) 반점의 잡음이 적은 경우 (b) 반점의 잡음이 많은 경우



(c) 배경 잡음을 많이 포함하고 있을 경우 : 배경 잡음은 실험과정에서 조작 실수나 이물질의 의한 오염으로 발생한다.



(d) (e)

그림 5 마이크로어레이 이미지에서의 여러 오류들 -
 (a) Signal noise가 거의 없는 이상적인 형태의 반점. (b) 많은 signal 잡음을 포함하고 있는 반점. (c) 배경 잡음을 많이 포함하고 있을 경우. (d) 거의 정확한 원형의 모양을 가진 반점. (e) 정확한 원의 모양을 가지지 않는 경우의 반점

같이 이상적인 원형을 나타내지 못하게 된다.

본 논문에서는 마이크로어레이 이미지를 분석하는 방법에 있어서 각 반점들의 격자 상의 위치를 결정하여 주는 알고리즘을 제시한다. 이를 위해서 본 논문에서는 마이크로어레이 이미지 분석에 맞게 균일 격자 구조를 사용하여 전산학적인 측면에서 자동화된 격자 주소 결정 방법을 설명하려고 한다.

2. 마이크로어레이 이미지의 격자 주소 결정 문제

마이크로어레이 이미지 분석의 최종 목표는 각 반점에 위치하는 유전자들의 발현 정도를 알아내는 것이다. 이미지상에 위치한 유전자들의 정확한 격자 위치를 알아내야한다. 유전자의 정확한 격자 위치는 마이크로어레이 이미지내에서의 블록의 위치와 그 블록 내에 위치한 유전자의 위치 두 가지 모두를 알아내는 것이다. 마이크로어레이 이미지가 이상적인 특징을 보일 경우에는 격자 주소 결정이 쉽게 해결되지만, 실제 마이크로어레이 이미지는 여러 오류를 포함하고 있으므로, 자동적으로 격자 주소를 결정하는 것은 아주 어려운 일이다. 이를 해결하기 위해 지금까지 Buhler[8], Brown[9], Steinfath[10], Jung[4,11] 등 많은 연구가 이루어졌다. Buhler 등은 이미지상의 반점 위치를 찾으면서 이들 반점의 모양과 위치에 대한 품질 제어 및 측정에 대한 방법론과 격자 주소 결정을 위한 방법으로는 여러 실험 이미지에 대해서 평균적인 반점의 위치를 미리 측정하고 마이크로어레이의 물리적인 정보를 이용하여 각 반점들의 격자 주소를 결정하는 방법을 제시하였지만, 한 종류의 마이크로어레이 실험에서만 가능하다는 단점을 가지고 있다[8]. Brown 등은 마이크로어레이 이미지 데이터를 분석하는 데 있어서 고려되어야 할 품질 제어를 측정하기 위해 통계적인 측정 방법을 사용하였고[9], Steinfath 등은 마이크로어레이에 의해서 생성되는 이미지에 대한 분석과 이미지 분석의 결과인 유전자 발현 프로파일을 클러스터링 한 결과를 보여주는 마이크로어레이 분석 도구를 소개하였다[10]. Jung 등은 ϵ -graph를 이용하여, 사용자의 개입 없이 격자 주소를 결정하는 알고리즘을 소개하였으나, 격자 구조가 10° 이내의 기울기를 가지는 경우에 안전하게 수행 가능하였다. 이처럼, 현재의 마이크로어레이 실험은 동시에 수천 개의 유전자간의 상호작용도 알 수 있게 되므로 방대한 양의 데이터가 생성되며, 또한 반점의 위치 변동이나 모양, 크기의 다양, 샘플의 오염과 같은 다양한 문제로 인하여 자동적으로 분석하기 어려우며, 이를 위한 효율적인 알고리즘이 필요하다.

3. 점들의 균일 정도

이미지에서 공간상의 균일(regularity) 정도 값을 찾는 것은 컴퓨터 비전이나, 장면 분석, 적외선 촬영 이미지에서 지뢰를 찾아내는 것과 같은 분야에서 중요하게 다루어져 온 문제이다.

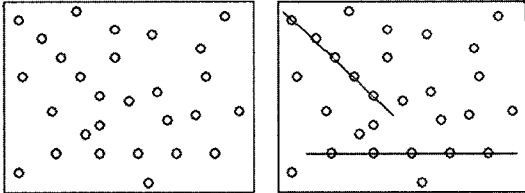
3.1 균등 일직선 서열(균일 서열)

서열(sequence)이 “균등 일직선(regular 또는 ESC : Equally-Spaced and Collinear Sequence)하다”라는 것

은 서열에 포함된 점들의 간격이 일정하며, 각 점들이 일직선상에 위치하고 있다는 것이다. 이를 정의하면 다음과 같다.

- ① Collinear : $P \subset E^2$ 이고, $\overline{P} \subseteq \overline{E^2}$ 일 때 ($|\overline{P}| \geq 2$), 만약 \overline{P} 의 모든 점이 하나의 직선상에 존재하면, \overline{P} 는 collinear하다.
- ② Equally spaced : 만약 $\overline{P} = \{\overline{p}_1, \overline{p}_2, \dots, \overline{p}_n\}$ 이 equally spaced이면, $|\overline{P}| > 3$ 이고, $\overline{p}_i - \overline{p}_{i-1} = \overline{p}_{i+1} - \overline{p}_i$ 이다.

그림 6은 주어진 입력 데이터에서 최대 균일 서열(균등 일직선 서열, regular sequence)을 생성한 것이다. 최대 균일 서열이라는 것은 입력으로 주어진 n개의 점으로 이루어진 전체 집합 P에서 가장 긴 균일 서열들을 말한다. 그림 6의 (b)에서 생성된 두 직선이 균일함을 알 수 있다. 균일화에 속하는 모든 원소들은 인접한 두 점 사이의 거리가 일정하며, 모든 원소들은 일직선상에 존재한다.



(a) 입력 포인트 집합 (b) 생성된 regularity

그림 6 입력된 포인트 집합에서의 균일화 : 균일 서열은 그 서열에 포함된 모든 점들이 동일 직선상에 위치하며, 인접한 점들 사이의 간격이 일정한 것을 말한다.

$dist(a, b)$ 가 a, b 두 점사이의 거리이고, 어떤 점집합 R이 균일 서열이라면, 다음과 같은 조건이 성립한다.

$$R = (p_1, p_2, \dots, p_n),$$

$$dist(p_i, p_{i+1}) = dist(p_{i+1}, p_{i+2}) \text{ and}$$

$$collinear(p_i, p_{i+1}, p_{i+2}), \text{ for } 1 \leq i < n-1$$

Kahng 등은 입력으로 주어진 n개의 점으로 이루어진 전체 집합 P에서 가장 긴 균일 서열을 찾기 위해서 입력 포인트 집합의 점들을 정렬하여 3개의 포인트로 이루어진 균일 서열들을 모두 찾아서 그것을 모두 그래프로 만든 다음 연결되는 그래프가 있으면 그것들을 연결해나가는 방법을 사용하였다[12]. 이는 1차원과 다차원에서의 문제로 나누어 생각할 수 있다.

1차원에서는 입력 포인트 집합 P를 모두 정렬한다. 정렬된 집합에서 P_A 를 i번째 점, P_B, P_C 를 $i+1, i+2$ 번

째 점이라 하면, N개의 모든 점 P_A, P_B, P_C 세 점의 거리를 비교한다.

$$dist_1 = P_B - P_A, \quad dist_2 = P_C - P_B$$

만약 $dist_1 > dist_2$ 이면, $P_c = P_{c+1}$ 를 수행하여, 다시 $dist_1$ 과 $dist_2$ 의 거리를 비교하고, 반대이면 $P_B = P_{B+1}$ 를 수행하고, 다시 $dist_1$ 과 $dist_2$ 의 거리를 비교한다. 비교를 하다 $dist_1 = dist_2$ 인 점 P_B, P_C 가 나타나면, P_A, P_B, P_C 는 3점으로 이루어진 균일 서열이 된다. 3점으로 이루어진 모든 균일화의 집합이 생성되면, 가장 긴 균일 서열을 구하기 위해서 각 세 점으로 이루어진 균일 서열을 그래프로 재구성한다. 생성된 그래프들 사이에 연결된 곳이 있으면, 이들을 하나로 연결하여 하나의 그래프로 바꾸게 된다. 연결되는 모든 그래프들을 연결하여 새로운 그래프로 재구성하고 나면, 남아 있는 결과 그래프들의 집합이 우리가 구하고자 하는 가장 긴 균일화가 된다. 2차원 이상에서 가장 긴 균일 서열을 구하기 위해서는 입력으로 들어온 점들을 각 좌표축으로 투영(projection)한다. 투영된 새로운 점들에서 regularity들을 구하게 되는데, 이것은 투영된 좌표축에서는 균일 서열이 성립하더라도 실제로는 그렇지 않을 수 있다. 따라서 투영한 축에서 균일 서열이 성립하면, 이를 실제 포인트들에서도 성립하는지를 체크해 주어야 한다. 체크하는 것은 균일 서열을 구축할 때 실행하면 되므로 1차원에서의 실행시간과 동일한 시간이 소요된다.

3.2 ε 간격을 고려한 균등 일직선 서열

균일 서열에 기반 하여, ε 간격을 고려한 균일 서열(ε-regular sequence)은 서열에 포함된 점들이 균일 서열이 있는 축에서 최대 ε만큼 떨어져있을 때를 말한다. 이를 식으로 나타내면 다음과 같다.

- ① ε-regular sequence : $P \subset E^2$ 이고, $\overline{P} \subseteq P$ 일 때, \overline{P} 가 ε-regular sequence이면, 모든 $1 \leq x_i \leq n$ 인 $|x_i - \overline{x}_i| \leq \epsilon, |y_i - \overline{y}_i| \leq \epsilon$ 인 $\overline{P} = \{\overline{p}_1, \overline{p}_2, \dots, \overline{p}_n\}$ 인 균일 서열이 존재한다.

실제, 본 연구에서도 ε을 고려한 균일 서열을 사용한다. 마이크로어레이의 반점들의 위치나 크기 등이 이상적이지 않으므로, 이를 고려하기 위한 것이다. 그림 7은 ε을 고려한 균일 서열의 예를 보여준다.

Robins 등은 regularity에 포함되는 점들이 균일 서열에서 ε 만큼의 간격 안에 위치하는 것을 허용하여 균일 서열을 찾아내는 방법론을 제시하였다[13]. 이 논문에서의 방법은 $O(n^2 \log n)$ 만큼의 시간이 사용된다. Robins 등의 알고리즘에서 설명한 모든 최대 ε을 고려한 균일 서열을 구하는 방법은 원래의 입력 이미지를 $8 * \epsilon$ 크기의 조직으로 나누어 각각을 cell이라 하고, 임

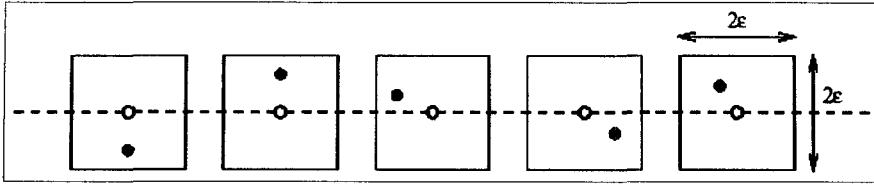


그림 7 ε을 고려한 균일 서열의 예[13] : 각 포인트들이 균일 서열에서 최대 ε 만큼 떨어져 있다.

력 포인트 집합 P를 cell 좌표에 의해서 정렬한다. 정렬된 각 점들을 한 쌍씩 만들어, 각 쌍에서 한 방향씩 쌍에 포함되지 않은 점을 추가하면서 ε을 고려한 균일 서열이 될 수 있는지를 검사하고, 만족하면 확장해나간다. 모든 가능한 쌍에서 이와 같은 방법을 수행하면, 최대 ε을 고려한 균일 서열을 구할 수 있게 된다.

본 논문에서는 오랫동안 전산학 분야에서 연구되어 온 균일 서열을 마이크로어레이 이미지 분석에 맞게 변형하여 마이크로어레이 이미지를 자동 분석하는 알고리즘을 제시한다. 마이크로어레이 이미지의 반점의 수는 수천 개에서 수만 개가 되므로 Kahng나 Robin에서 제시되었던 방법론을 그대로 적용할 수 없다. 따라서 본 문서에서는 마이크로어레이 이미지 데이터에 맞게 변형한 균일 서열 방법론에 대해서 소개한다.

4. 격자 주소 결정 알고리즘

본 논문에서는 격자 주소 결정의 자동화를 위하여 마이크로어레이 이미지에서 발현된 반점들의 중심을 입력으로 하여, 가상 점을 허용한 ε을 고려한 균일 서열을 생성하여, 블록과 반점의 격자 주소 알고리즘에 사용한다. 그림 8은 본 논문에서 제시한 알고리즘을 구현한 시스템이다.

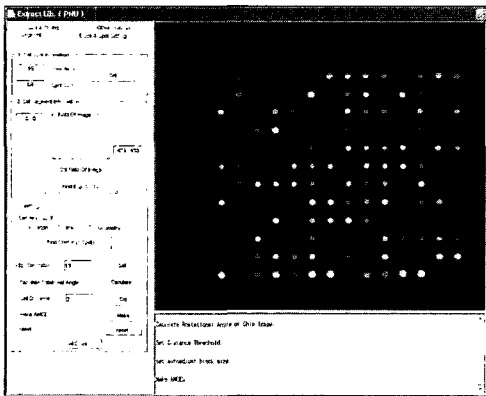


그림 8 본 논문에서 구현한 시스템 화면 : 14×12개의 반점으로 이루어진 하나의 블록을 가진 마이크로어레이 이미지의 블록 주소를 결정한 것이다.

4.1 전처리 과정

본 논문에서 제안한 격자 주소 알고리즘을 위해서 마이크로어레이 이미지에서 실제 주소 알고리즘을 적용하기 전에 전처리 과정을 거치게 된다. 이는 격자 알고리즘에서 사용할 입력 데이터를 생성하는 단계이다. 입력 데이터는 마이크로어레이 실험에서 발현된 반점들의 중심점의 집합이다. 반점의 중심점을 구하기 위해서는 먼저 고정된 일정한 임계값 T 이상의 강도를 가지는 픽셀의 세그먼트를 찾아서, 중심을 계산한다. 각 세그먼트의 중심점을 s_p 라 표기한다. 본 논문에서는 MBR, MASS, Geometry의 세 가지 반점 중심을 고려한다[4,11].

- 최소 경계 사각형(MBR, Minimum Bounding Rectangle) 중심

각 세그먼트를 둘러싼 MBR을 찾아서, MBR의 중심을 반점의 중심이라 생각한다.

- 무게(MASS) 중심

각 세그먼트의 픽셀들의 평균 위치를 반점의 중심으로 고려한다.

- 기하학적(Geometry) 중심

각 세그먼트 내에서 외부로부터 차례로 각 픽셀의 레이어를 결정하여 가장 높은 레이어를 가지는 픽셀을 반점의 중심으로 한다.

사용자의 입력에 따라 위 세 가지 방법에 의하여 각 반점들의 세그먼트 중심점의 집합 S_p 를 생성한다. S_p 는 블록과 반점들의 격자 주소를 계산하기 위해 사용된다. 그림 9는 입력이미지에서 일정 T값 이상 발현된 반점들의 기하학적 중심점을 구한 것이다. 입력된 이미지의 각 블록의 왼쪽 아래 반점((a)의 붉은색 사각형 부분)들이 왼쪽으로 기울어져 있어, 그 중심점들이 왼쪽으로 기울어져 있는 것((b)의 붉은색 사각형 부분)을 볼 수 있다.

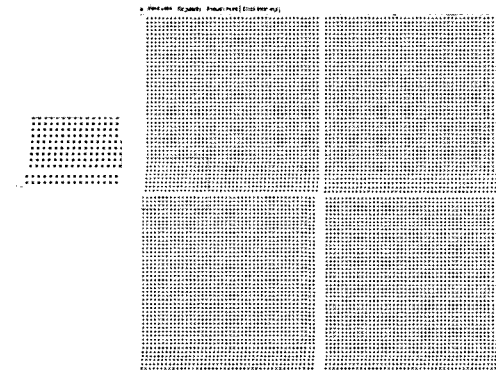
4.2 블록 주소 알고리즘

블록 주소 결정 알고리즘의 주요 성질을 간단히 설명하면 다음과 같다.

- ① 기울어진 각과 단위 길이(unit distance)를 구한다.
- ② 기울어진 각에 따라 집합 P의 중심점들을 가장 긴 ε을 고려한 균일 서열을 생성한다. 이 때, 하나의 가상 점을 허용한다.



(a) 입력 이미지



(b) 기하학적 중심점

그림 9 세그먼트의 중심점 : 입력된 이미지의 고정된 임계값 T 이상의 발현 값을 가진 세그먼트들에서 기하학적 중심점을 생성한다.

③ 생성된 regularity들 중 공통 노드를 가지는 균일 서열들을 하나로 합한 후, 각 균일 서열의 MBR을 계산한다. 이 때, 각 MBR은 각 블록이 된다.

다음 장부터는 각 단계에 대해서 자세히 살펴보겠다.

4.2.1 기울어진 각과 단위 길이 계산

본 논문에서 구현한 시스템에서는 블록 주소를 결정하기 위해서 마이크로어레이 이미지의 기울어진 정도와 단위 길이를 구하게 된다. 그림 10은 본 시스템에서 기울어진 각과 단위 길이를 계산하는 것을 나타낸 것이다.

기울어진 정도와 단위 길이를 구하기 위해서, 본 시스템에서는 최대의 ϵ 을 고려한 균일 서열을 생성한다. 최대의 ϵ 을 고려한 균일 서열을 구하기 위한 입력 데이터는 $2*\epsilon \times 2*\epsilon$ 로 나눈 셀의 중심을 사용한다. 반점들의 중심점은 반점의 위치 변동이나 모양, 크기의 다양, 샘플의 오염과 같은 다양한 문제로 인하여 동일 간격의 같은 일직선상에 위치하지 않으므로, 이를 그대로 regularity를 구하는데 사용할 수 없다. 따라서 본 시스

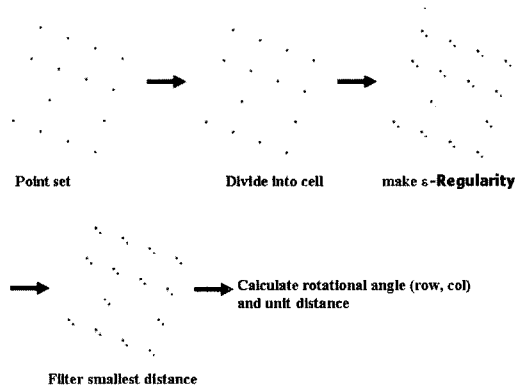


그림 10 단위길이와 기울어진 각을 계산하는 도식도

템에서는 셀의 중심점을 사용한다. 셀 중심점은 ϵ 만큼의 크기($2*\epsilon \times 2*\epsilon$)의 셀(cell)로 전체 이미지를 분할하고, 적어도 하나의 반점을 포함하고 있는 셀을 c_v (valid cell)라고 정의한다. c_v 의 중심점들로 이루어진 점집합 C_p (set of points of valid cell)을 구성하고, 이를 최대의 ϵ 을 고려한 균일 서열을 계산할 때 사용한다. 그림 11은 s_p , c_v , c_p 의 관계를 나타낸다. 그림 11은 하나의 블록을 표현한 것으로, 블록 내에 4×4 개의 반점들이 존재하며, 이중 발현된 반점들의 중심점을 구한 것이 s_p 이다. 전체 이미지를 $2*\epsilon \times 2*\epsilon$ 로 나눈 셀들 중에 s_p 가 포함되어 있는 셀이 c_v 이며, c_v 의 중심점이 c_p 이다. 그림에서 하늘색 점은 s_p 를 나타내며, 초록색 점은 c_p 를 나타낸다.

그림 12는 C_p 를 입력으로 본 시스템에서 생성한 최대의 ϵ 을 고려한 균일 서열이다. 파란색으로 표현된 점은 s_p 이며, 주황색으로 표현된 점은 c_p , 초록색으로 그려진 것이 균일 서열이다. 이전의 연구에서 고려되었던 최대의 ϵ 을 고려한 균일 서열은 그림 8에서 생성한 최

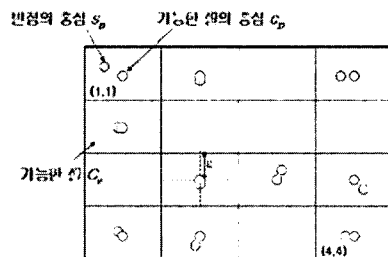


그림 11 s_p , c_v , c_p 의 관계 : 4×4 반점을 가진 하나의 블록을 표현한 것이다. s_p 가 포함되어 있는 c_v 의 중심점이 c_p 가 된다. c_p 들의 집합인 C_p 를 이용하여, maximal ϵ -regularity를 구하게 된다.

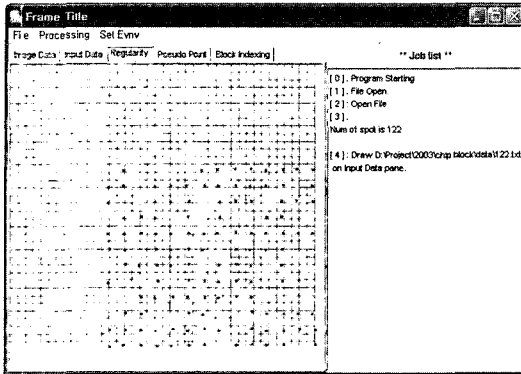


그림 12 본 시스템에서의 maximal ϵ -regularity 생성 : 마이크로어레이 이미지 분석에 맞추어 현재 생성된 maximal ϵ -regularity들 보다 단위 길이가 긴 것들의 생성을 제약한다.

대의 ϵ 을 고려한 균일 서열보다 길이가 긴 것들이 더 있을 수 있다. 하지만, 본 시스템은 마이크로어레이 이미지 분석에 초점을 둔 것이므로, 생성되는 균일 서열의 생성 시 제약 조건을 준 것이다. 가장 작은 거리를 가지는 균일 서열을 reg_{min} , reg_{min} 집합을 R_{min} 이라고 하자. 전체 이미지에서 블록별로 기울어진 정도가 약간씩 다르거나, 블록의 특정 부분이 한쪽으로 기울어져 있을 수 있으므로, reg_{min} 하나만을 사용하여 θ 를 구하면 전체 이미지를 왜곡 시킬 수 있다. 따라서 θ 를 구하기 위해서 R_{min} 의 모든 reg_{min} 를 고려한다. [알고리즘 1]은 θ 를 계산하는 알고리즘이다. 계산된 모든 R_{min} 의 각을 0보다 큰 각 $angle1$, 0보다 작은 각 $angle2$, 0인 각의 수를 각각 n_1 , n_2 , n_3 를 계산하여 이 중 가장 많은 수를 가진 각을 선택하고, 그것의 평균을 블록의 기울기 각 θ 로 계산한다.

위 길이 d_n 는 R_{min} 의 거리가 된다.

4.2.2 기울어진 각과 하나의 가상 점을 허용한 ϵ 를 고려한 균일 서열 생성

기울어진 각이 계산되어지면, 모든 C_b 를 $-\theta$ 만큼 회전시켜 변환한다.

$$c_b(x_i, y_i) \rightarrow c_b'(x_i', y_i')$$

$$x_i' = x_i \cdot \cos(-\theta) - y_i \cdot \sin(-\theta)$$

$$y_i' = x_i \cdot \sin(-\theta) + y_i \cdot \cos(-\theta)$$

이렇게 변환된 C_b' 는 마이크로어레이 이미지가 기울어지지 않았을 때의 가능한 셀들의 중점이 된다. C_b' 를 이용하여, 가로와 세로만을 고려하여 maximal ϵ -regularity를 생성한다. 각 regularity를 생성할 때에는 하나의 가상 점(pseudo point)을 허용한다. 마이크로어레이

알고리즘 1 기울어진 각 θ 를 계산하는 알고리즘

```

Algorithm: Computation of Rotational Angle
Input:  $R_{min}$ ; /*set of  $reg_{min}$ , which is regularity with minimum distance.*/
Output:  $\theta$  /*microarray rotational angle*/

·  $angle1 = 0$ ; /* sum of angles which is larger than 0 */
·  $angle2 = 0$ ; /* sum of angles which is smaller than 0 */
·  $n_1 = 0$ ; /* the set of  $n_1$  which is number of  $angle1$  */
·  $n_2 = 0$ ; /* the set of  $n_2$  which is number of  $angle2$  */
·  $n_3 = 0$ ; /* the set of  $n_3$  which is number of other(angle is 0) */

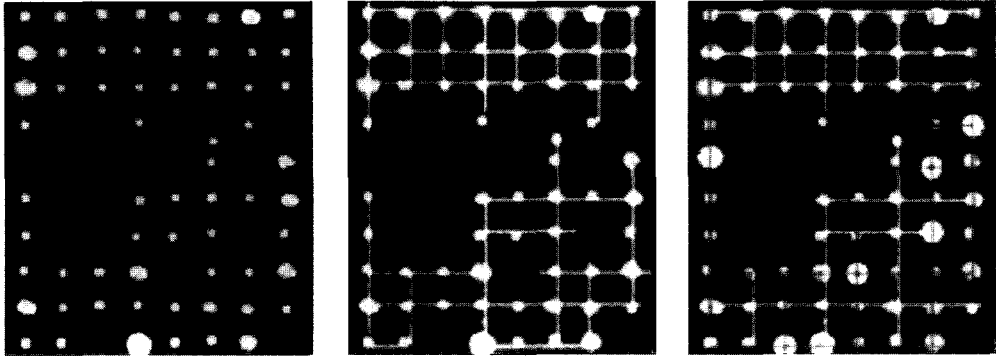
for each element  $R_{min}$  in set  $R_{min}$ 
    Point p1, p2 = first and second point in  $R_{min}$ ;
     $\theta = \text{atan}(p1.p2)$ ;
    if  $\theta > 0$  then  $angle1 += \theta$ ;  $n_1++$ ;          endif
    else if  $\theta < 0$  then  $angle2 += \theta$ ;  $n_2++$ ;    endelsif
    else then  $n_3++$ ;                             endels
endfor

if  $\max(n_1, n_2, n_3) = n_1$  then  $\theta = angle1 / n_1$ ; endif
else if  $\theta > 0$  then  $\theta = angle2 / n_2$ ;          endelsif
else then  $\theta = 0$ ;                               endels
    
```

이의 반점들은 발현이 되는 것이 있고, 그렇지 않은 것이 존재한다. Maximal ϵ -regularity는 그 정의에서도 나타나듯이 regularity에 속하는 인접한 원소들 간의 거리가 일정하여야한다. 하지만, 마이크로어레이 데이터에서는 발현이 되는 유전자와 그렇지 않은 유전자가 존재한다. 일정 수준 이상 발현한 세그먼트들의 중심점만을 입력으로 regularity를 생성하면, 발현되지 않은 반점들로 인하여, 하나로 생성될 수 있는 regularity들이 조각으로 나누어지거나, 생성되지 않는 경우가 있게 된다. 이를 보완하기 위하여, 본 방법론에서는 하나의 가상 점을 허용하여 regularity를 구성한다. 하나의 가상 점을 추가하여 regularity를 구성한다는 것은 regularity를 생성할 때, 추가되는 점이 일직선상에는 존재하나 길이가 생성되고 있는 regularity의 단위길이의 두 배일 경우, regularity와 추가되는 점 사이에 가상 점을 하나 허용하여 regularity를 생성해나가는 것이다. 그림 13은 하나의 블록 (a)에서의 regularity를 생성한 경우 (b)와, 하나의 가상 점을 허용하여 생성한 regularity의 예이다. (c)에서 6개의 가상점이 추가되었다.

알고리즘 2는 하나의 가상 점을 허용하여, 균일 서열을 구성하는 알고리즘이다. 하나의 가상 점(pseudo point)을 허용한 균일 서열을 p -regularity라고 둔다.

그림 14는 그림 5의 왜곡된 이미지에 대한 균일 서열을 구성한 것이다. 각 블록의 아랫부분에 부분적으로 왜곡된 부분의 반점들도 모두 균일 서열로 연결되어 있는 것을 볼 수 있다.



(a) 입력 이미지 (b) 생성된 균일 서열 (c) 하나의 가상 점들을 허용한 균일 서열

그림 13 하나의 가상 점을 허용한 균일 서열 : 마이크로어레이 이미지에서는 발견되지 않는 반점들로 인하여, 하나로 생성될 수 있는 regularity가 나뉘어져서 나오거나 생성되지 않을 수 있다. 이러한 점을 보완하기 위하여 하나의 가상 점을 허용한 regularity를 생성한다.

알고리즘 2 하나의 가상 점을 허용하여 p -regularity를 구성하는 알고리즘

```

Algorithm: Construction of regularity using a pseudo point
Input:  $C_p$  /* set of transformational  $C_p$  */
Output:  $R$  /* set of  $p$ -regularity */

MakeRegularity(x coordinate):
MakeRegularity(y coordinate):

Function MakeRegularity(coordinate coord)
    sort  $C_p$  by coord position;
    while all element  $c_p$  in set  $C_p$ 
        for each set of points which are located in same coord coordinate
            if  $dis(p_i, p_{i+1}) > d_n$ 
                if  $dis(p_i, p') = dis(p', p_{i+1})$  and  $collinear(p_i, p', p_{i+1})$ 
                    add  $p_{i+1}$  in regularity  $reg_{current}$ 
                endif
            else
                add  $reg_{current}$  in a  $R$ 
            endif
        else
            add  $p_{i+1}$  in regularity.
        endif
    endfor
endwhile
    
```

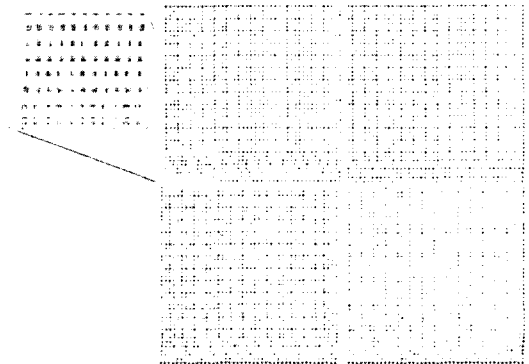


그림 14 하나의 가상 점을 허용한 regularity 생성. 왜곡된 이미지에 대해서도 모두 연결되는 regularity를 생성할 수 있다.

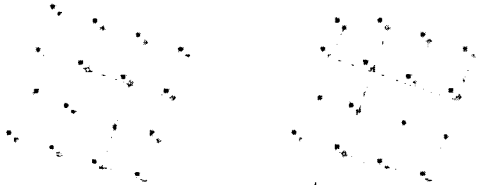
그림 15에서와 같이, 하나의 가상 점을 허용하여 생성한 균일 서열들은 마이크로어레이의 행과 열을 고려하여 생성한 것이므로, 공통된 노드들을 가지게 된다. 본 알고리즘에서는 균일 서열들을 이용하여 블록을 구하기 위해서 연결된 균일 서열들을 포함할 수 있는 MBR을 만들고, 그것을 하나의 블록으로 생각한다. 이를 위해서 균일 서열들에서 서로 공통된 노드들을 지나는 것들을 하나의 그래프로 연결한다. 그림 15의 (b)에서는 두 개의 그래프가 생성되고, 이들의 MBR을 구한 것이다. 생성된 MBR은 셀의 중심점을 이용하여 만들어진 것이므로, MBR의 네 꼭지점이 속하는 셀에 반점이 포함되어 있다면, 그 꼭지점을 반점의 위치로 변환한다.

4.2.3 블록 생성

하나의 가상 점을 허용한 regularity들을 생성하면, 이를 입력으로 하여, 공통 노드를 가지는 regularity들을 하나의 그래프로 연결한다. 모든 공통 노드들을 가진 균일 서열들을 그래프로 생성하면, 각 그래프들의 MBR을 계산한다. 그림 15는 생성된 균일 서열들에서 MBR을 구하는 과정을 나타낸 것이다.

4. 반점 주소 결정

블록이 결정되면, 각 반점 주소는 일차적으로 블록 주



(a) 공통 노드를 가지는 regularity들을 하나의 그래프로 연결



(b) 각 그래프들의 MBR을 구한다.

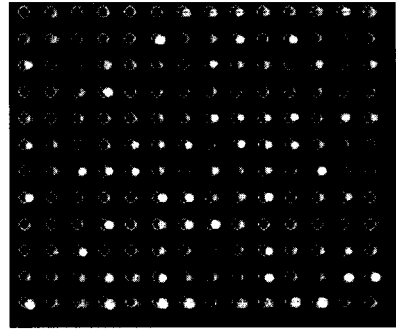


(c) MBR을 이용하여, 블록을 결정한다.

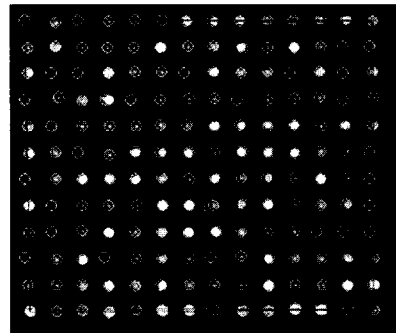
그림 15 균일 서열들을 이용하여, 블록을 생성하는 과정 : (a) 하나의 가상 점을 허용하여 생성한 regularity들에서 서로 공통된 노드를 가지는 균일 서열들을 하나의 그래프로 연결한다. (b) 생성된 그래프들의 MBR을 계산한다. 각 생성된 균일 서열과 그래프들은 θ 만큼 회전시켜 변환 시킨 것이므로, 사용자에게 보여줄 때에는 다시 θ 만큼 회전시켜야한다. (c) 생성된 MBR의 네 꼭지점이 있는 셀에 반점이 존재하면, 꼭지점을 반점의 위치로 변환한다.

소에서 구해진 단위 길이를 이용하여, 균일하게 블록을 분할하여, 그 중심점을 반점의 주소로 할당한다. 반점의 주소는 블록의 제일 왼쪽 위에 위치하는 반점에서부터 시작하여, 행과 열에서 단위 길이만큼 떨어질 때마다 행과 열에 대해서 1씩 더하면서 주소를 부여한다. 그런 다음, 현재 설정되어 있는 반점 주소에서 반점의 반지름 내에 떨어져 있는 가장 가까운 곳에 위치한 발현된 반점의 중심이 있는 지를 찾아보고, 있다면 현재 주소를 발현된 반점의 중심점으로 변경한다.

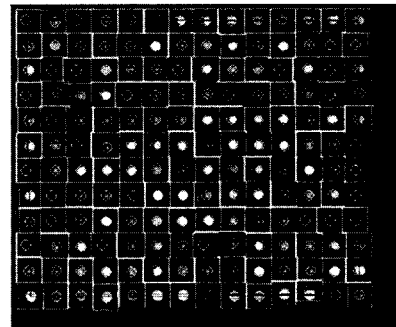
그림 16은 본 논문에서 구현한 시스템에서 반점 주소를 결정한 화면이다. (a)는 생성된 블록에서 일정하게



(a) 블록에서 단위 길이만큼 일정하게 나는 반점의 주소



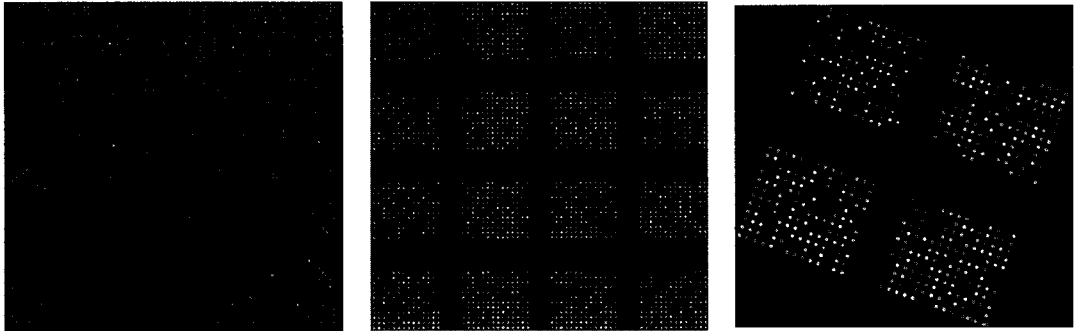
(b) 발현된 반점을 고려하여, 변환한 반점의 주소



(c) 반점의 영역을 할당

그림 16 반점의 주소 할당 : (a) 블록의 위치에서 단위 길이만큼 일정하게 나누어서 각 반점들의 주소를 할당한 화면. 블록의 왼쪽 위의 반점이 (1,1)위치가 된다. (b) 발현된 반점들을 고려하여, 반점 주소를 변환한 화면. (c) 주소가 결정된 반점들의 영역을 표현한 화면

주소를 부여한 것이고, (b)는 발현된 반점의 중심점을 고려하여, 반점 주소를 변환한 것이다. (c)는 주소가 결정된 반점들의 영역을 표현한 것이다. 반점들의 위치가 정확한 일직선상에 놓여있지 않기 때문에, (c)에서의 반점들의 영역의 위치가 일정하지 않는 것이다.



(a) 2×2 블록의 블록과 반점의 주소 결정의 결과 (b) 4×4 블록의 블록과 반점 주소 결정의 결과 (c) 시계 방향으로 20° 회전한 이미지의 블록과 반점 주소 결정의 결과

그림 17 (a) $e=1.5$ 화소이고, 이미지의 해상도는 1024×1024, (b) $\epsilon = 4$ 화소이고, 이미지의 해상도는 1705×1654, (c) $e=2$ 화소이고, 이미지의 해상도는 1022×996이다.

5. 실험 결과

본 논문에서는 제시한 블록 주소 결정 알고리즘을 테스트하기 위해서 가상의 이미지 데이터뿐만 아니라 실제 이미지 데이터에 대해서도 수행하였다. 사용한 테스트 데이터는 다음의 3가지로 구분된다.

1. Stanford 마이크로어레이 이미지 데이터 : 블록 수는 2×2이며, 하나의 블록 내의 반점 수는 44×44이다. 반점의 중심점은 MBR 방법을 이용하여 생성하였다.
2. 배추에 관한 실제 마이크로어레이 데이터 : 블록 수는 4×4이며, 하나의 블록 내의 반점 수는 12×14이다. 반점의 중심점은 Geometry 방법을 이용하여 생성하였다.
3. 가상 데이터 : 블록수가 2×2인 배추에 관한 실제 마이크로어레이 데이터를 시계 방향 시계 방향(cw, clockwise)으로 20° 회전한 데이터이다. 반점의 중심점은 MBR 방법을 이용하여 생성하였다.

그림 17은 본 논문에서 구현한 시스템에서의 블록 주소 결정의 결과이다. (a)의 이미지는 배추에 관한 실제 마이크로어레이 데이터이며, (b)는 Stanford 마이크로어레이 이미지 데이터, (c)는 배추에 관한 실제 마이크로어레이 데이터를 시계 방향(cw, clockwise)으로 20° 회전한 가상 데이터이다. (a)는 그림 9에서 설명한 각 블록의 아래쪽 반점들이 한쪽으로 치우쳐져 있는 이미지의 블록과 반점의 주소 결정 결과이며, (b)의 이미지는 발현 정도나 간격이 비교적 일정한 이미지에서의 주소 결정 결과이다. (c)는 블록수가 2×2인 이미지를 cw 방향으로 20° 회전한 가상 데이터의 실험 결과이다. 그림 17의 (a) 결과를 통하여, 본 논문에서 제안하는 블록 주소 결정 알고리즘으로 블록 내 부분적인 왜곡이 있는 이미지에서도 블록 주소 결정이 잘 수행됨을 알 수 있

으며, (c)와 같이 회전한 이미지에 대해서도 정확하게 블록과 반점의 위치를 결정할 수 있음을 알 수 있다.

6. 결론과 향후 과제

현재 개발되어진 대부분의 DNA 칩 분석 시스템들은 정확한 이미지 처리를 위해서는 사용자의 많은 개입이 필요하고, 이로 인하여 한 연구자가 처리할 수 있는 이미지의 양이 줄어들며, 여러 사람이 이미지를 처리하게 되므로 이미지 분석에서 일관성이 떨어진다는 것이다. 본 논문에서는 이를 해결하기 위해서 마이크로어레이 이미지 분석에 있어서 완전 자동화된 블록과 반점들의 주소 결정 알고리즘을 제시하였다. 본 논문에서 사용한 알고리즘은 이미지 분석에서 많이 연구되어왔던 점들의 균일(regularity 또는 Equally Spaced and Collinear sequence)에 대한 연구를 마이크로어레이 이미지에 맞게 하나의 가상 점을 허용한 ϵ 을 고려한 균일 서열 문제로 변환하여 사용하였다. 본 논문에서 제안한 방법은 반점의 수가 n 이고, 적어도 하나의 반점을 포함한 셀의 수가 m 일 때, 최악의 조건($n=m$ 이고, 가장 작은 거리를 가장 나중에 찾을 경우)에서 $O(m^2)$ 만큼의 시간이 소요된다. 본 논문에서 제시한 알고리즘의 주요 특징은 다음과 같다.

- 하나의 가상 점을 허용한 ϵ 을 고려한 균일 서열 : 기존 이미지 분석 분야에서 연구되어 왔던 균일 서열을 마이크로어레이 이미지 분석에 맞도록 변형하여 사용하였다.
- 사용자의 적은 입력 : 기존의 매뉴얼 방식으로 주소 결정을 수행하면, 사용자가 각각의 블록의 위치를 모두 지정해 주어야하므로, 많은 사용자의 입력이 필요하다. 하지만, 본 논문에서 제시한 알고리즘은 사용자

의 입력을 최소화한 완전 자동 주소 결정론이므로, 사용자가 쉽게 사용할 수 있다.

- 본 논문에서 제시한 알고리즘은 마이크로어레이 이미지만이 아니라, 격자 구조를 가진 다른 이미지 데이터에도 사용할 수 있다.

향후 연구과제는 다음과 같다. 먼저 ϵ 값의 자동 계산이 필요하다. 본 논문에서 제시한 방법론에서 사용자의 개입은 ϵ 값의 입력이다. 따라서 현재 발현된 반점들의 간격이나, 기울기를 고려하여 ϵ 값을 자동으로 계산해 줌으로써, 사용자의 개입을 줄여줄 수 있다. 그다음 정확한 반점의 발현 값 분석이 필요하다. 본 논문에서 제시된 이미지 처리는 마이크로어레이의 블록과 반점들의 주소를 결정해주는 것이다. 이러한 분석이 끝나고 나면, 반점들의 정확한 발현 정도의 수치를 계산하여, 실험에서 유의한 유전자들을 찾게 된다. 본 논문에서 제시한 블록과 반점의 주소 결정 후, 반점의 정확한 발현 수치를 계산함으로써 본 시스템에서 보다 많은 분석을 시행할 수 있을 것이다. 본 알고리즘에 대한 자세한 정보는 <http://jade.cs.pusan.ac.kr/~autogrid>에 정리되어 있다.

참 고 문 헌

- [1] David J. Duggan, Michael Bitter, Yidong Chen, Paul Meltzer, and Jeffery M. Trent. "Expression profiling using cDNA microarrays," *Nature Genetics*, 12 : 10-14, 1999.
- [2] M.Schena and R.W.Davis. "Parallel Analysis with Biological Chips: PCR Applications," *Academic Press*, 445-446, 1999.
- [3] M.Schena. "Microarray Biochip Technology," *Eaton Publishing*, 2000.
- [4] Ho-Youl Jung and Hwan-Gue Cho. "An Automatic Block and Spot Indexing with k-Nearest Neighbors Graph for Microarray Image Analysis," *Bioinformatics(Supple.)*, 18 : 141-151, 2002.
- [5] 김판규, 진희정, 조환규, "Quality Measures for Microarray Design and Experiment," *KSBI*, 2 : 155p, 2003.
- [6] Kuklin APA and Shams S, "Quality control in microarray image analysis," *Image & Microscopy Research*, 1, 2001.
- [7] Hautaniemi S, Edgren H, Vesanen P, Wolf M, Jarvinen AK, Yli-Harja O, Astola J, Kallioniemi O and Monni O, "A novel strategy for microarray quality control using Bayesian networks," *Bioinformatics*, 19: 2031-2038, 2003.
- [8] J. Buhler, T. Ideker, and D. Haynor, "Dapple: Improved Techniques for Finding Spots on DNA Microarrays," *University of Washinton CSE Technical Report UWTR*, 2000-08-05, 2000.
- [9] C.S. Brown, P. C. Goodwin, and P. K. Sorger,

"Image metrics in the statistical analysis of DNA microarray data," *Proceedings of National Academy of Sciences of the United States of America*, 98 : 8944-8949, 2001.

- [10] M. Steinfath, W. Wruck, H. Seidel, H. Lehrach, U. Radelof, and J. O'Brien, "Automated image analysis for array hybridization experiments," *Bioinformatics*, 17 : 634-641, 2001.
- [11] 정호열, 황미녕, 유영중, 조환규, "cDNA 마이크로어레이 이미지를 위한 그래프 모델과 분석 알고리즘", *한국정보과학회논문지*, 29 (7/8) : 411-421. 2002.
- [12] Andrew B.Kahng and Gabriel Robins, "Optimal algorithms for extracting spatial regularity in images," *Pattern Recognition Letters*, 12, 1991.
- [13] Gabriel Robins, Brian L.Robinson and Bhupinder S.Sethi. "On Detecting Spatial Regularity in Noisy Images," *Information Processing Letters*, 69. 1999.



진희정

2000년 부산대학교 전자계산학과 졸업(학사). 2002년 부산대학교 전자계산학과 졸업(이학석사). 2002년~2003년 국립보건원 유전체연구부 역학정보실 선임연구원. 2003년~현재 부산대학교 전자계산학과 박사과정. 관심분야는 생물정보학, 알

고리즘 이론



조환규

1984년 서울대학교 계산통계학과 졸업(학사). 1986년 한국과학기술원 전자계산학과 졸업(공학석사). 1990년 한국과학기술원 전자계산학과 졸업(공학박사). 1990년~현재 부산대학교 전기전자정보컴퓨터공학부 교수. 관심분야는 그래프이론,

생물정보학, 그래픽스