

Quality Measures for Microarray Design and Experiments

마이크로어레이 설계와 실험을 위한 품질평가함수

Pan-Gyu Kim^{1*}, Hee-Jeong Jin² and Hwan-Gue Cho²

¹ Department of Computer Science, Pusan National University, Pusan, Korea

² Department of Computer Engineering, Pusan National University, Pusan, Korea

*To whom correspondence should be addressed. E-mail: pgkim@pearl.cs.pusan.ac.kr

Abstract

마이크로어레이를 이용한 발현실험이 기하급수적으로 늘어남에 따라 마이크로어레이 이미지를 자동으로 처리하는 기술에 대한 요구가 커지고 있고, 이에 대한 연구도 많아지고 있다. 마이크로어레이 이미지를 자동으로 처리하기 위해서는 각 이미지가 가지고 있는 스팟 패턴을 알아보고, 자동화 정도를 측정할 수 있는 품질평가함수가 필요하다. 우리는 본 논문에서 마이크로어레이 이미지 분석의 자동화에 대한 평가를 도와주는 스팟 패턴의 품질평가함수(quality measure)를 정의하고, 각 실험이 얼마나 잘 이루어졌는지를 예측할 수 있는 품질제어평가함수(quality control measure)를 정의한다. 또한 마이크로어레이 실험과 이미지에 대한 품질을 평가하기 위해서 이미지내의 블록들과 스팟들에 대한 통계량을 이용하고, 스팟들의 발현값에 대한 정확도를 측정하기 위한 품질평가함수들을 정의한다. 이러한 품질평가함수의 측정을 위해서 최대정규점점의 집합(maximal regular point set)과 메타 그리드를 이용한다.

Introduction

마이크로어레이를 이용한 대용량처리 (high-throughput) 실험 데이터의 분석은 현대 생명과학에서 점점 더 중요해지고 있다. 마이크로어레이는 한번에 수만 혹은 수 십만개의 유전자 발현 정보나 유전자형 (genotype)을 얻을 수 있다. 마이크로어레이 연구가 활발해지면서, 분석되어야 할 마이크로어레이 이미지의 수가 급속히 증가하고 이에 따라 자동 이미지분석에 대한 요구가 점점 더 커지고 있다. 최근 자동 마이크로

어레이 이미지 분석 시스템들이 소개되기 시작했지만[6,7], 마이크로어레이 이미지의 분석이 자동화되기 위해서는 이미지의 상태가 좋고, 일정비율이상의 유전자들이 발현되어 있어야 한다. 최근 이러한 마이크로어레이의 실험이나 결과 이미지에서의 품질을 측정하는 방법에 대한 연구[1,2,5]가 시작되었지만 이에 대한 연구는 아직 미미한 편이다. 또한 마이크로어레이 전체 이미지에서 스팟팅 패턴(spotting pattern)의 품질평가함수(quality measure)에 관련된 연구는 거의 없다.

최근 마이크로어레이를 통한 유전자 발현실험이 많이 이루어지고 있지만, 아직까지 마이크로어레이 실험은 여러 가지 실험환경이나

This work is supported by a grant of Korea Health 21 R&D Project, Ministry of Health & Welfare(02-PJ1-PG11-VN01-SV06-0029)

조건에 따라 적지 않은 오류를 가지게 된다. 이러한 오류는 마이크로어레이 실험 후 이미지 분석 시 많은 어려움을 주게 되며, 또한 조절 유전자(control gene)가 특정 영역에 집중되어 있거나, 강하게 발현되는 유전자들이 밀집되어 있으면, 서로간의 간섭으로 인하여 분석 시 어려움이 따른다. 따라서 이미지 분석 시 이러한 어려움을 줄이기 위해서 마이크로어레이 스팟팅 패턴(spottting pattern)을 고려한 디자인이 필요하다.

마이크로어레이 실험의 궁극적인 목적은 유전자 발현실험을 통하여 각 유전자들의 발현율을 알아보는 것이다. 유전자의 발현율은 이미지상의 각 유전자를 표현하는 스팟들의 색의 강도를 측정하여 계산할 수 있다. 마이크로어레이를 통한 유전자 발현 실험은 실험 환경과 조건에 따라 다양한 형태의 결과를 나타내고, 실제로 실험이 제대로 이루어지지 않아서 이미지 분석 결과가 전혀 의미가 없는 경우도 있다. 이 문제점을 해결하기 위해서 마이크로어레이 실험의 완성도를 측정하는 평가함수가 필요하다.

본 논문에서는 각 마이크로어레이 이미지의 몇 가지 통계량을 이용하여 이미지들 사이의 품질을 비교하는 문제와 자동 그리딩(auto gridding)을 위한 마이크로어레이 설계 문제를 제안한다. 본 연구에서는 마이크로어레이 스팟 패턴의 품질을 측정하기 위해서 정점 집합(point set)에서의 정규도(regularity)를 분석한다. 정규도의 분석을 위해 규칙적인 서열(regular sequence)을 구하는데, 이 때 스팟의 중심점들을 정점 집합으로 하여 규칙적인 서열을 계산하고, 구해진 규칙적인 서열들 중에서 최대정규서열(maximal regular sequence)를 구할 수 있다 [3,4].

서열이 “규칙적이다”라는 것은 서열에 포함된 모든 점들이 동일 직선상에 같은 간격으로 존재한다는 것이다. 이미지에서 공간상의 정규도를 찾는 것은 컴퓨터 비전이나, 장면 분석, 적외선 촬영 이미지에서 지

위를 찾아내는 것과 같은 분야에서 중요하게 다루어져 온 문제이며, 지금까지 많은 연구가 있어왔다[3,4]. 본 논문에서는 이를 이용하여, 마이크로어레이의 설계시의 스팟 패턴과 실험 후 이미지의 스팟 분포에 대한 품질을 측정한다. 또한 마이크로어레이 전체 실험과 한 이미지에 대한 품질을 평가하기 위해서 메타그리드를 이용하여 이미지내의 블록들과 스팟들간의 간격에 대한 통계량을 측정하고, 스팟들의 발현값에 대한 정확도를 측정하기 위한 품질평가함수들을 정의한다. Hautaniemi등[5]은 마이크로어레이 실험의 정확도에 영향을 미치는 여러 가지 요소들을 가정하고, 베이스안 네트워크를 통해서 실험의 정확도에 가장 많은 영향을 주는 요소들을 알아보는 모델을 제시하였다. 그러나, 이 모델은 각각의 스팟에 대한 품질평가를 측정해보는 것으로 이미지 전체에 대한 품질평가를 나타내기에는 미흡하다.

마이크로어레이 스팟팅 패턴 (Spotting Pattern)을 위한 품질평가함수

마이크로어레이에 놓여질 수 있는 유전자는 조절 유전자와 테스트 유전자(test gene) 두 종류가 존재한다. 조절 유전자는 마이크로어레이의 정규화와 그리드 위치를 지정해주기 위해서 마이크로어레이 설계시에 포함되는 특별한 유전자로서, 이상적인 조절 유전자는 100% 발현한다. 일반적으로, 이러한 조절 유전자들의 위치는 마이크로어레이 이미지 분석시에 미리 알고 있다.

마이크로어레이에서 그리드 위치가 (i, j) 인 유전자의 이름을 g_{ij} , 조절 유전자의 집합을 C , g_{ij} 의 발현율을 $exp(i, j)$ 라 하자. 실제 $exp(i, j)$ 는 실험 후 이미지 분석을 통해서 그 값을 알 수 있지만, 우리는 그것을 미리 알고 있다고 가정한다.

- $cpath(1, 2) = \max(1.0 \cdot 0.6 \cdot 0.2, 1.0 \cdot 0.9 \cdot 0.2) = \max(0.12, 0.18) = 0.18$
- $cpath(3, 2) = \max(1.0 \cdot 0.7 \cdot 0.2, 1.0 \cdot 0.8 \cdot 0.2) = \max(0.14, 0.16) = 0.16$
- $address(2, 2) = \max(cpath(1, 2), cpath(3, 2)) = \max(0.18, 0.16) = 0.18$

| | | | |
|-----|-----|-----|-----|
| 1.0 | 0.9 | 0.5 | 0.7 |
| 0.6 | 0.2 | 0.8 | 0.3 |
| 0.6 | 0.7 | 1.0 | 0.8 |
| 0.8 | 0.8 | 0.7 | 0.5 |

표 1. $address(i, j)$ 의 예. 4 by 4 그리드 마이크로어레이의 발현율. 조절 유전자가 (1,1)과 (3,3)에 있으므로 $cpath(1, 2)$, $cpath(3, 2)$ 를 구하고, 두 값 중 큰 값인 0.18이 (2,2) 스팟의 $address(2, 2)$ 가 된다.

| (a)A의 $exp(i, j)$ | (b)A의 $address(i, j)$ | (c)B의 $exp(i, j)$ | (d)B의 $address(i, j)$ |
|-------------------|-----------------------|-------------------|-----------------------|
| 1.0 0.9 0.5 0.7 | 1.0 0.9 0.45 0.32 | 0.6 0.9 0.5 0.7 | 0.36 0.32 0.14 0.31 |
| 0.6 0.2 0.8 0.3 | 0.6 0.18 0.8 0.24 | 0.6 0.2 0.8 0.3 | 0.6 0.14 0.45 0.13 |
| 0.6 0.7 1.0 0.8 | 0.42 0.7 1.0 0.8 | 1.0 0.7 0.8 0.8 | 1.0 0.7 0.56 0.45 |
| 0.8 0.8 0.7 0.5 | 0.45 0.56 0.7 0.4 | 0.8 1.0 0.7 0.5 | 0.8 1.0 0.7 0.3 |

표 2. 4 by 4 그리드의 예. (a)와 (c)는 g_{ij} 의 $exp(i, j)$ 를 나타낸다. (b)와 (d)는 (a),(c) 각각의 $address(i, j)$ 이다. '1.0'은 조절 유전자를 뜻한다. (b)와 (d)의 결과에서 마이크로어레이 A가 B 보다 전체 스팟 $addressability$ 가 좋다는 것을 알 수 있다.

Gene id at microarray grid position (i, j) : g_{ij}

Expression rate of g_{ij} : $exp(i, j) =$

$$\begin{cases} exp(i, j) = 1, & \text{if } g_{ij} \in C \\ 0 \leq exp(i, j) < 1, & \text{otherwise} \end{cases}$$

$nearest(i, j)$ 를 g_{ij} 에서 스팟 어드레싱(spot addressing) 확률의 값들을 지정하는 조절 유전자들 중 가장 큰 스팟 어드레싱 확률 값을 주는 조절 유전자라 하고, $cpath(i, j)$ 를 g_{ij} 로부터 $nearest(i, j)$ 까지의 패스, $|cpath(i, j)|$ 를 $cpath(i, j)$ 의 길이라고 정의하자. 일반적으로 자동 어드레싱을 할 때 어떤 유전자의 위치는 그 유전자와 가장 가까운 조절 유전자 $nearest(i, j)$ 를 이용하여 계산된다. $nearest(i, j)$ 에서 g_{ij} 의 위치를 성공적으로 찾아낼 수 있는 확률을 $address(i, j)$ 라 하면 $address(i, j)$ 는 다음과 같이 정의된다.

$$address(i, j) = \prod_{g_{ij} \in cpath(i, j)} exp(i, j)$$

표 1은 4 by 4 그리드의 마이크로어레이의 발현율을 나타낸 것이다. 색이 칠해진 (2,2) 스팟의 $address(2, 2)$ 는 마이크로어레이의 조절 유전자의 위치에 의해서 결정된다. 조절 유전자가 (1,1)과 (3,3)에 있으므로 각각의 조절 유전자로부터의 $cpath(1, 2)$, $cpath(3, 2)$ 를 구하고, 두 값 중 큰 것이 $address(2, 2)$ 가 된다. 마이크로어레이에서 스팟팅 패턴의 모든 $address(i, j)$ 합은 $\sum_{g_{ij}} address(i, j)$ 이고, 이는 스팟의 수로 나눈으로써 정규화 될 수 있다.

표 2는 4 by 4 그리드의 마이크로어레이에 대한 $address(i, j)$ 의 예이다. 칩 A와 B의 전체 $addressability$ 는 9.513, 8.021326이다. 따라서 어레이 A가 B보다 더 좋은 스팟팅 패턴을 가지고 있다고 볼 수 있다.

자동 그리딩(Gridding)을 위한 마

이크로어레이 이미지 품질평가함수

마이크로어레이의 실험(hybridization) 후 얻어지는 이미지의 자동 그리딩을 위한 품질을 평가하기 위해서 최대정규점점의 집합의 분할을 수행하며, 그 수행 방법은 다음과 같다.

- (1) 마이크로어레이 이미지에서 스팟 분할을 수행한 다음 스팟의 중심점들을 이용하여 점집합 P_0 를 만든다.
- (2) P_0 에서 ϵ -최대정규점점의 집합을 구하고, 그 중 가장 길이가 긴 ${}_eMRS_1$ 을 구한다.
- (3) P_0 에서 ${}_eMRS_1$ 의 모든 점을 제거한다.
 $P_1 = P_0 - {}_eMRS_1$.
- (4) 모든 점들이 ${}_eMRS$ 에 포함될 때까지 반복한다. ${}_eMRS$ 는 서로 교차하지 않는다.

결과적으로 모든 유전자들은 최소 하나 그리고 최대 하나의 ${}_eMRS$ 에 포함된다. 만약 조절 유전자가 포함되지 않은 ${}_eMRS_i$ 가 k 개라면, 이것은 ${}_eMRS_i$ 에 포함된 모든 점들은 한번에 위치를 결정할 수 있으므로, 적어도 k 번의 매뉴얼 작업으로 스팟 인덱스를 찾을 수 있다는 것이다. 따라서, ${}_eMRS_i$ 의 수는 실험 후 마이크로어레이 이미지 품질평가 시 중요한 척도가 된다.

마이크로어레이 이미지 품질평가함수

마이크로어레이 이미지의 품질은 실험의 정확도에 크게 의존한다. 따라서 실험자의 능숙도를 판단하고, 실험전체의 품질을 제어하기 위해서는 실험이 된 이미지의 성능을 체계적으로 평가할 필요가 있다. 우리는 이러한 성능평가를 위해서 각 마이크로어레이 이미지의 그리드구조를 저장한 메타그리

드를 이용하여, 각 이미지내의 블록과 스팟들에 대한 품질평가함수를 정의한다.

일반적으로, 한 실험을 통해서 생성되는 여러 이미지들이 포함하고 있는 그리드구조는 모두 동일하다. 그러므로 한 실험을 통해서 얻어지는 이미지들이 포함하고 있는 블록과 스팟들의 규칙성은 통계적 자료를 통해서 간단하게 알 수 있다. 또한 이러한 정보들은 똑같은 그리드 구조를 가지는 다른 칩 이미지들의 분석 시에 쉽게 그리드구조를 파악하여 분석할 수 있는 좋은 정보가 된다. 블록과 스팟들의 그리드구조를 알아보기 위해서 다음과 같은 통계량을 측정한다.

- (1) $IB_a = \text{Avg}(\text{Inter Block})$
- (2) $IB_s = \text{Std}(\text{Inter Block})$
- (3) $IS_a = \text{Avg}(\text{Inter Spot})$
- (4) $IS_s = \text{Std}(\text{Inter Spot})$
- (5) $SD_a = \text{Avg}(\text{Spot Diameter})$
- (6) $SD_s = \text{Std}(\text{Spot Diameter})$

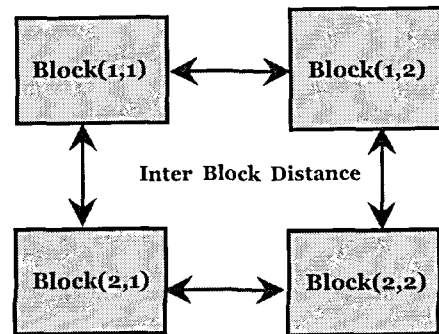


그림 1. 2 × 2 블록구조를 가질 때의 블록간 거리의 계산. 각각의 블록에 대해서 이웃한 4-연결(connected) 인접 블록과의 거리를 블록간의 거리로 나타내고 모든 블록간 거리의 평균을 IB_a , 표준편차를 IB_s 라 한다.

마이크로어레이 실험의 목적은 유전자들의 발현율을 측정하는 것이다. 이것은 이미지에서 각 스팟들의 위치에 해당하는 색의 강도를 조사하여 알 수 있다. 하지만, 마이크로어레이 실험은 실험자의 능숙도나 마

이크로어레이의 성능, 염료의 양 등에 따라 다양한 결과를 얻을 수 있고, 실제로 실험한 결과가 실제 반응을 제대로 나타내지 못할 수도 있다. 우리는 스팟의 크기나 모양, 위치 등의 정보를 고려하여 유전자의 발현 값이 얼마나 정확한지를 측정할 수 있는 품질평가함수들을 다음과 같은 5가지로 정의한다 : 신호잡음(Signal Noise), 배경잡음(Background Noise), Scale Invariant, Size Regularity, 스팟위치 에러(Spot Alignment).

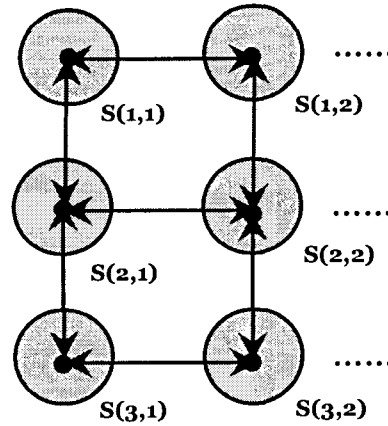


그림 2. 3 × 3 이상의 스팟구조를 가질 때의 스팟간 거리의 계산. 각각의 스팟에 대해서 이웃한 4-연결 (connected) 인접 스팟과의 거리를 스팟간의 거리로 나타내고 모든 스팟간 거리의 평균을 IS_a , 표준편차를 IS_s 라 한다.

(1) 신호 잡음(Signal Noise) :

스팟의 영역에서 블랍(blob)이나 다른 잡음과 같이 유리슬라이드에 다른 이물질이 묻게 되면, 발현값을 측정할 때 오류가 나타내게 된다. 우리는 이러한 평가함수를 전체 스팟의 영역과 그 영역에서 오염된 영역을 제외한영역의 비로 정의한다.

$$Signal\ Noise = \frac{Signal\ Area}{Signal\ Area + Ignored\ Signal\ Area}, \in [0, 1]$$

(2) 배경 잡음(Background Noise) :

마이크로어레이 이미지 분석시에 배경의 강도값도 실험결과에 많은 영향을 미친다. 그러므로 배경에 포함되어 있는 오염된 영역도 중요한 평가함수 중 하나이다. 배경이 포함하고 있는 잡음이 얼마나 되는지를 확인하여 이미지를 평가할 수 있다.

$$Background\ Noise = \frac{Background\ Area}{Background\ Area + Ignored\ Background\ Area}, \in [0, 1]$$

(3) Scale Invariant :

스팟의 모양이 원형에 가까울수록 실험의 질이 좋다고 할 수 있다. 먼저 스팟이 포함하고 있는 컴포넌트영역의 무게중심을 그 스팟의 중심이라 가정한다. 스팟의 중심이 원의 중심이고 스팟내에서 발견된 유효 컴포넌트영역을 포함하는 원들중 최소의 반지름을 가지는 원을 C_a 라 하고, 컴포넌트 영역이 포함할 수 원들중 최대 반지름을 가지는 원을 C_b 라 하면, Scale Invariant는 C_a 의 반지름과 C_b 의 반지름의 비로 정의된다(그림 3).

$$Scale\ Invariant = \frac{Radius\ of\ C_b}{Radius\ of\ C_a}, \in [0, 1]$$

(4) Size Regularity :

Scale Invariant와 비슷한 의미를 지니지만 때때로 다른 값을 가질 수 있다. 시그널 영역과 시그널 영역으로 만들어질 수 있는 이상적인 원의 넓이의 비로 나타낸다(그림 3).

$$Size\ Regularity = \frac{Signal\ Area}{Ideal\ Circle\ Area}, \in [0, 1]$$

(5) 스팟위치 에러(Spot Alignment) :

블록내 각 스팟의 위치는 모두 똑같은 거리의 그리드 형태로 놓여 있어야 한다. 각 스팟이 이상적인 그리드 형태로 놓였을 때의 위치와 실제 위치의 차이를 스팟위치 에러로 정의한다.

$$Spot\ Alignment = 1 - \frac{Avg(real\ position\ of\ spot - regular\ position\ of\ spot)}{Avg(inter\ spot\ distance/2)}, \in [0, 1]$$

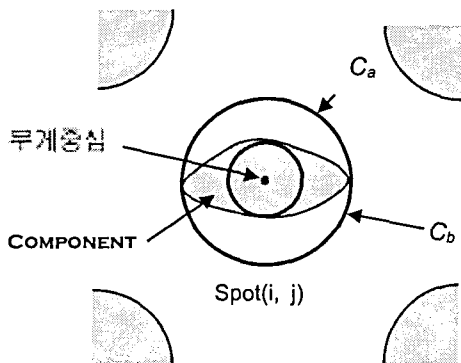


그림 3. Scale Invariant와 Size Regularity를 설명한 그림. Scale Invariant는 하나의 스팟으로 인식된 컴포넌트에서 컴포넌트를 포함하는 최소원과 컴포넌트가 포함하고 있는 최대원사이의 반지름비로 정의되고, Size Regularity는 컴포넌트를 포함하는 최소원의 넓이와 컴포넌트의 넓이의 비로 정의된다.

Experiments and Results

우리는 마이크로어레이 이미지의 품질평가함수의 테스트를 위해서 6종류의 이미지를 사용하였다. A부터 E까지는 실제 실험실에서 실험한 데이터들이고, F는 스탠포드대학에서 Yeast의 세포주기를 연구할 때 사용한 마이크로어레이 이미지이다. 먼저 각 이미지들이 포함하고 있는 칩의 특성을 알아

보기 위해서 블록과 스팟에 대한 기본적인 통계량을 표 3과 같이 알아보았다. 이러한 정보는 실험에 대한 정보를 알 수 있을 뿐만 아니라, 향후 똑같은 형태로 설계된 마이크로어레이의 발현값을 분석할 때 메타그리드에서 사용할 수 있는 유용한 정보가 될 수 있다. 또한 표 4에서 보여주는 품질평가 측정값이 실험의 1차 품질평가함수라면 표 3은 2차 품질평가함수가 될 수 있다. 각각의 이미지는 발현실험을 한 유전자들의 종류에 따라 그 발현율이 다르다. 각각의 스팟에 대하여 발현율이 높을 수록 이미지 분석시에 그리딩이 쉬우며, 나아가 자동그리딩이 가능하게 된다.

6종류의 이미지에 대하여 품질평가함수를 적용한 결과는 표 4와 같다. 모든 이미지가 공통적으로 신호잡음과 스팟위치에러에 대한 평가에서는 비교적 좋은 결과를 나타냈다. 하지만 나머지 3가지 평가함수에 있어서는 서로간의 차이를 나타내었는데, 이러한 차이를 통해서 품질의 차이를 확인할 수 있다. C와 D가 가장 좋은 결과를 나타내었으며, A, B, E도 C, D와 그리 큰 차이를 보이지 않았지만 F는 나머지 5종의 테스트 이미지 결과와는 많은 차이를 나타내었다. F와 같은 이미지는 자동 그리딩이 거의 불가능한 이미지라고 할 수 있다.

| CHIP | B | S | IMAGE RESOLUTION | BD | SD | RADIUS | C(%) |
|------|----|--------|------------------|--------|-------|--------|-------------|
| A | 16 | 16×100 | 1886×2321 | 102.41 | 35.97 | 8.16 | 141(8.8%) |
| B | 16 | 16×324 | 1900×1860 | 29.73 | 23.03 | 6.63 | 995(19.2%) |
| C | 16 | 16×324 | 1808×1800 | 31.33 | 23.02 | 7.14 | 1650(31.8%) |
| D | 16 | 16×168 | 1984×2004 | 122.79 | 24.89 | 7.49 | 1289(48.0%) |
| E | 16 | 16×342 | 2000×2000 | 22.17 | 22.93 | 8.82 | 656(12.0%) |
| F | 4 | 4×1596 | 1024×1024 | 23.17 | 10.43 | 4.13 | 715(11.2%) |

*B : NUMBER OF BLOCKS , S : NUMBER OF SPOTS, BD : BLOCK DISTANCE, SD : SPOT DISTANCE, C : NUMBER OF COMPONENT DETECTED.

표 3. Chip Specification. 각 이미지에 포함되어 있는 블록과 스팟의 그리드구조와 발현을 등을 나타낸다. 이 정보는 메타그리드의 중요한 정보로써 향후 똑같은 형태로 설계된 마이크로어레이 이미지에서는 이러한 정보를 이용하여 쉽게 분석가능하다.

| CHIP | BACKGROUND NOISE | SIGNAL NOISE | SCALE INVARIANT | SIZE REGULARITY | SPOT ALIGNMENT | 평균 |
|------|------------------|--------------|-----------------|-----------------|----------------|-------|
| A | 0.52 | 0.98 | 0.56 | 0.75 | 0.93 | 0.748 |
| B | 0.70 | 0.98 | 0.52 | 0.55 | 0.94 | 0.738 |
| C | 0.65 | 0.99 | 0.56 | 0.69 | 0.91 | 0.760 |
| D | 0.69 | 0.99 | 0.54 | 0.62 | 0.94 | 0.756 |
| E | 0.90 | 0.99 | 0.56 | 0.45 | 0.86 | 0.752 |
| F | 0.53 | 0.91 | 0.49 | 0.51 | 0.83 | 0.654 |

표 4. Quality Measure 실험결과. 6개의 데이터에 대한 품질평가표이다. 대체적으로 Signal Noise와 Spot Alignment는 좋은 결과를 보이지만 Background Noise, Scale Invariant, Size Regularity에서는 편차를 가진다.

Discussion

6종류의 이미지에 대하여 테스트한 결과 신호잡음과 스팟 위치에러는 마이크로어레이 이미지의 품질을 나타낼 때 큰 편차를 보이지 않았다. 초기 마이크로어레이 실험 시에는 이러한 부분이 많은 문제를 야기시

켰지만, 마이크로어레이어의 발달과 실험자의 주의 및 실험환경 개선 등으로 최근에는 이러한 부분이 많이 향상되었다. 이와 달리 배경잡음은 큰 편차를 나타내었는데, 이것은 유전자의 발현값을 측정할 때 정확도의 차이를 야기시킬 수 있다. 똑같은 슬

라이드에 대해서도 레이저로 얻은 이미지와 CCD카메라로 얻은 이미지는 배경잡음과 신호 잡음의 차이를 나타내는데, 특히, 미세하게 구분되어야 할 배경에서 더 큰 차이를 나타내게 되는 것이다. Scale Invariant와 Size Regularity는 실험시 환경에 많은 영향을 받는다. 특히, 슬라이드 구성시에 얼마만큼의 염료를 사용하는가에 따라 똑같이 발현되더라도 그 모양이 틀려질 수 있다. 일반적으로 넉넉한 양의 염료를 사용할수록 스팟의 모양이 원에 가까워지며 이 두가지 평가함수에서도 좋은 결과를 나타낼 것이다.

Conclusion and Future works

본 논문에서는 발현실험이후의 마이크로어레이 이미지에 대하여 자동적으로 스팟의 위치를 찾기 위한 스팟 패턴의 품질평가함수(quality measure)와 각 실험이 얼마나 잘 이루어졌는지를 예측할 수 있는 품질제어평가함수(quality control measure)를 제안하였다. 또한 마이크로어레이 전체 실험과 한 이미지에 대한 품질을 평가하기 위해서 이미지내의 블럭들과 스팟들 간의 간격에 대한 통계량을 이용하여 정의하고, 스팟들의 발현값에 대한 정확도를 측정하기 위한 품질평가함수들을 정의하였다. 이러한 품질평가함수들은 자동 그리딩을 위한 마이크로어레이 설계에서 유용하게 사용될 수 있다. 또한 한 실험을 통해서 얻어지는 동일한 디자인의 여러 마이크로어레이 이미지들 간의 품질을 평가할 수 있어서, 실험의 완성도나 실험자의 능숙도 그리고 마이크로어레이의 성능을 측정할 수 있다.

향후 개선하거나 추후 연구해야 할 과제는 다음과 같다.

- 본 논문에서 정의한 마이크로어레이 이미지 품질평가함수는 각 스팟들의 크기는 고려하지 않고, 단순하게 평가함수를 통해서 얻어지는 값만을 계산한다. 만일 어떤 스팟은 반지름이 아주 크고, 다른 스팟은 반지름이 아주 작는데, 두 스팟에

대한 평가가 비슷하게 나왔다면 반지름이 작을수록 더 잘 된 것이라 할 수 있다. 왜냐하면 스팟의 크기가 크면 클수록 더 정확한 실험결과를 얻기 때문이다. 향후 한 이미지내에 포함되어 있는 스팟의 수, 혹은 스팟의 반지름의 크기에 따라 품질평가값을 정규화하는 작업이 필요하다.

- 현재 5개의 품질평가함수 값들이 가지는 변이역은 $[0,1]$ 로 동일하다. 하지만, 구간 $[0,1]$ 에서 모든 평가함수 값들이 동일한 분포를 가지지는 않는다. 현재 전체를 통합하여 평가하는 방법은 단순하게 평균을 산출하는데, 이 방법은 모든 평가함수들이 똑같은 크기의 영향을 미치고, 유니폼 확률분포라는 가정하에 사용할 수 있는 방법이다. 앞으로 5가지 평가함수를 하나의 통합 평가함수로 구성하는 작업에 대한 연구가 이루어져야 할 것이다.
- 마이크로어레이 실험은 슬라이드에 붙이는 유전자나, 그리드구조, 반지름크기, 염료, 이미지 추출방법 등 여러 가지 조건에 따라 많은 영향을 받는다. 그러므로, 마이크로어레이 실험의 완성도를 측정하는 품질평가함수는 상황에 따라 다른 모델로 구성되어 질 수도 있다. 각 변수들 간의 상관관계를 잘 파악하여 특정 상태나 상황에 맞는 마이크로어레이 품질평가함수의 개발이 필요하다.

<http://pearl.cs.pusan.ac.kr/~arrayzer/index.html>

에서 본 연구에서 사용한 프로그램과 문서 및 설명서를 다운받거나 확인할 수 있다.

References

- [1] A. P. Alexander Kuklin and S. Shams., Quality control in microarray image analysis, *Imaging & Microscopy Research*, 1, 2001.
- [2] B. L. R. Gabriel Robins and B. S.Sethi, On detecting spatial regularity in noisy

- images, *Information Processing Letters*, 12(4), 1999.
- [3] A. B. Kahng and G. Robins, Optimal algorithms for extracting spatial regularity in images, *Pattern Recognition Letters*, 12, 1999.
 - [4] G. Robins, B. L. Robinson and Bhupinder S. Sethi., On Detecting Spatial Regularity in Noisy Images, *Information Processing Letters*, 69, 1999.
 - [5] Sampsa Hautaniemi *et al.*, A novel strategy for microarray quality control using bayesian networks, *Bioinformatics* (to appear), 2003.
 - [6] Ho-Youl Jung and Hwan-Gue Cho, An automatic block and spot indexing with k -nearest neighbors graph for microarray image analysis, *Bioinformatics*, 18: 141S-151S, 2002.
 - [7] 정호열, 황미녕, 유영중, 조환규, cDNA 마이크로어레이 이미지를 위한 그래프 모델과 분석알고리즘, *한국정보과학회논문지*, 29 (7/8): 411-421, 2002.