

SPR 센서로 촬영된 마이크로어레이 영상 분석을 위한 전처리 알고리즘

황지운^o 김재학 한준희
 포항공과대학교 컴퓨터공학과
 { romeo^o, inyashio^o, joonhan }@postech.ac.kr

Pre-processing Algorithm for Analysis from Microarray Image Sequence by Surface Plasmon Resonance Biosensor

Ji-Woon Hwang^o Jae-Hak Kim Joon H. Han
 Computer Vision Lab, Dept. of Comp. Sci. & Eng., POSTECH

요 약

나노 기술의 개발과 마이크로어레이(microarray)의 등장으로 생물학자들은 한번에 보다 빠르고 신속하게, 대량의 실험을 처리할 수 있게 되었다. 이와 발 맞추어 마이크로어레이를 이용한 다양한 실험 방법들이 개발되었다. 형광 염료(fluorescence dye)를 이용한 관찰 방법이 널리 이용되고 있으나, 관찰되는 형광 염료의 밝기가 실험 환경(pH, 온도)에 매우 민감하게 반응하며, 단백질을 포함한 많은 분자 물질들이 형광을 내지 않기 때문에 마이크로어레이를 이용한 분석 대상 물질들의 개수가 제한을 받는다. 본 논문에서는 직접적인 형광 염료의 사용 없이, SPR(Surface Plasmon Resonance)을 이용한 마이크로어레이 분석 시스템에서 스팟(spot)의 밝기(intensity)를 측정하기 위한 효율적인 전처리 과정을 제안하고자 한다. 전처리 과정은 크게 프로젝티브 왜곡 효과 제거, 스팟의 위치 추적, 스팟의 영역 추출, 정규화된 스팟의 밝기 측정으로 나누어진다. 특히, 이러한 과정을 거쳐서 측정된 밝기는 반응 유무의 관찰 뿐만 아니라, 실험 물질의 양적인 측정에도 이용되기 때문에 정확한 스팟의 밝기 측정에 중점을 두고자 한다.

1. 서론

마이크로어레이는 수천 혹은 수만 개 이상의 생체 물질-DNA나 단백질 등-을 일정 간격으로 금속이나 유리판에 집적시킨 칩을 의미한다. 마이크로어레이를 이용한 분석은 이러한 칩 위에 분석하고자 하는 물질들을 용매에 녹여서 뿌려줌으로써 반응 결과 및 발현 패턴을 분석하는 것이다. 이러한 발현 패턴은 영상으로 나타나며 이러한 실험 영상을 분석함으로써 결과를 얻을 수 있다.

형광 물질을 이용한 방법은 DNA 마이크로어레이 분석에 널리 이용되어 왔다. 하지만 단백질을 비롯한 많은 물질들이 형광 물질에 반응하지 않기 때문에 제한적으로 이용된다. 근래에 SPR 센서를 이용하여 마이크로어레이를 분석하는 방법이 소개되었다[1]. SPR의 이점은 형광 물질이나 다른 레이블링(labeling) 없이 높은 민감도(sensitivity)를 제공한다는 것이다.

본 논문에서는 SPR을 이용한 마이크로어레이 실험 영상을 분석하기 위해서 필요한 전처리 단계에 대한 표준화된 방법을 제시하고자 한다.

첫번째 단계는 SPR 센서의 특성상 CCD 카메라가 마이크로어레이 상에서 기울어져 촬영됨으로 인해서 생기는 프로젝티브 왜곡 효과를 제거하는 것이다. 두번째 단계는 마이크로어레이에서 각 스팟들의 위치를 자동적으로 찾고 마이크로어레이를 격자 무늬(lattice)로 나누어서 하나의 사각 영역의 중심에 하나의 스팟이 위치하도록 만드는 것이다. 세번째 단계는 하나의 사각 영역에 존재하는 하나의 스팟을 사각영역 내의 배경과 분리하여 순수한 스팟 영역만 추출하는 것이다. 마지막 단계에서는 각 사각 영역내에서 찾아진 스팟의 영역으로부터 스팟의 정규화된 밝기를 측정한다. 정규화된 밝기는 스팟 주위의 배경 밝기에 구애받지 않고 마이크로어레이의 지역적인 밝기의 편차를 배제한 밝기를 의미한다.

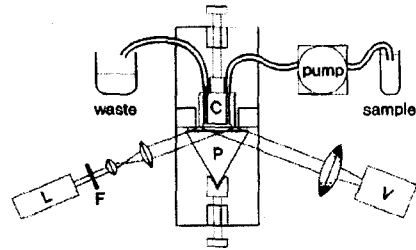


그림 1: 전형적인 Microarray 분석 시스템[2]

2. SPR 센서를 이용한 마이크로어레이 분석 시스템

그림 1은 SPR 센서를 이용한 전형적인 마이크로어레이 분석 시스템을 나타낸다. 마이크로어레이를 이용한 실험을 하기 위해서는 여러 가지 장비들이 필요하다. 우선, 실험적인 측면에서는 마이크로어레이에 분석하고자 하는 샘플(sample)을 용매에 녹여서 펌프(pump)를 통하여 플로우 셀(C)로 흘려보낸다. 이때, 용매에 녹아 있는 샘플과 마이크로어레이에 고정된 이미 알려져 있는 물질들과의 생화학적 반응을 관찰하면 된다. 여기까지는 일반적인 마이크로어레이 시스템과 같다. 주요 차이점은 실험을 관찰하는 시스템이다. 우선, 레이저나 백색광 등의 광원(L)을 마이크로어레이에 비추어 준다. 이때, 프리즘(P)를 통해 반사된 빛을 CCD 카메라(C)에서 포착하는 것이다[1].

3. 마이크로어레이 실험 영상의 전처리 단계

3.1 프로젝티브 왜곡 효과의 제거

마이크로어레이는 주로 정사각형의 금속판이나 유리판이 사용된다. 비록 마이크로어레이가 실제 정사각형일지라도 실제 비디오에 찍힌 영상은 프로젝티브 왜곡에 의해 사다리꼴 혹은 임의의 모양의 사각형이 나타난다. 이러한 프로

젝티브 왜곡현상을 없애는 방법은 다음과 같다.

1. 영상으로부터 Canny 윤곽선 검출기[3]를 이용하여 윤곽선을 검출한다.
2. 윤곽선이 검출된 영상에서 선을 찾기 위해서 정규화된 직선 방정식 $x \cos \theta + y \sin \theta = \rho$ 를 정의하고 파라미터 공간 (ρ, θ) 에 대한 허프 변환[4]을 수행한다.
3. 파라미터 공간에서 가장 값이 큰 네 개의 엔트리에 해당하는 파라미터값들이 마이크로어레이의 네 모서리에 대한 직선 방정식을 나타낸다.
4. 네 개의 직선 방정식에 대해서 각 직선이 교차하는 점을 구한다. 그림 2에서, 이 네 개의 점들 (P_1, P_2, P_3, P_4) 이 프로젝트 왜곡이 일어난 영상에서 마이크로어레이의 코너점들이 된다.
5. 프로젝트 왜곡을 제거시킨 마이크로어레이의 영상을 정의할 수 있다. 이러한 영상에 존재하는 마이크로어레이의 코너점들 (P_1, P_2, P_3, P_4) 은 마이크로어레이의 길이를 이용하여 정의할 수 있다. 이 때, 프로젝트 왜곡이 제거된 평면 Π 와 제거되지 않은 평면 Π' 사이에는 프로젝트 변환이 존재하며 이러한 변환 H 를 앞에서 구한 코너점들의 쌍을 이용하여 구할 수 있다[5].
6. 평면 Π 상에 있는 마이크로어레이의 이미지는 평면 Π' 상에 있는 마이크로어레이의 이미지로부터 H 의 역변환에 의해서 구하여진다.

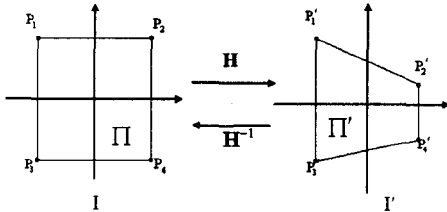


그림 2: 프로젝트 왜곡이 제거된 마이크로어레이 이미지 I와 실제 카메라로 촬영된 이미지 I'사이의 상관 관계

3.2 스왓의 위치 추적

앞에서 프로젝트 왜곡이 제거된 이미지 I를 얻었다. 이제 마이크로어레이상에서 일정한 간격으로 배열된 스왓의 위치를 추적하고 하나의 사각영역에 하나의 스왓이 존재하도록 마이크로어레이 이미지를 격자화(lattice)시켜야 한다. 마이크로어레이에서 스왓은 일정한 간격으로 놓여있으므로 특정한 축의 방향-x축 또는 y축-으로 픽셀값을 더하여보자. 이러한 1차원 신호에 대하여 1차 미분된 Gaussian 커널을 이용한 컨볼루션(convolution)[4]을 이용하여 신호에서 지역적 최대값(peak)과 지역적 최소값(valley)인 부분을 찾을 수 있다. 따라서, 두 개의 연속된 valley를 기준으로 선을 그으면 정확하게 스왓과 스왓 사이의 영역이 갈리게 된다. 만약 특정한 행이나 열의 스왓들이 제대로 빛을 발산하지 못하여, 1차원 신호에서 valley가 나타나지 않을 경우, 이미 찾은 valley와 valley 사이의 거리를 바탕으로 추론이 가능하다. x축 방향과 y축 방향으로 이러한 절차를 수행함으로써 각각의 모든 스왓이 하나의 사각 영역에 포함되도록 할 수 있다.

3.3 스왓의 영역 추출

이제는 격자화된 마이크로어레이 이미지에서 하나의 사각 영역으로부터 실제 스왓의 영역을 분리해야 한다. 실제 스왓 영역의 분리는 K-means 클러스터링[6]에 의해 수행된다. 우선 각 픽셀에 대하여 특징(feature)들을 추출한다. 추출되는 특징들은 픽셀의 RGB 색상과 사각 영역의 중심에서 픽셀까지의 유클리디언(Euclidean) 거리 d 이다. 이러한 특징들을 이용하여, 각 픽셀마다 특징 벡터(feature vector) v 를 정의한다.

$$v = \begin{bmatrix} d/(\text{Diagonal Length of Rectangle Region}) \\ R's \text{ Value}/(\text{Maximum R's Value}) \\ G's \text{ Value}/(\text{Maximum G's Value}) \\ B's \text{ Value}/(\text{Maximum B's Value}) \end{bmatrix} \quad (1)$$

그리고 각 픽셀의 RGB 색상만을 특징으로 이용하여, 각 픽셀마다 특징 벡터 u 를 정의한다.

$$u = \begin{bmatrix} R's \text{ Value}/(\text{Maximum R's Value}) \\ G's \text{ Value}/(\text{Maximum G's Value}) \\ B's \text{ Value}/(\text{Maximum B's Value}) \end{bmatrix} \quad (2)$$

그리고 다음과 같은 과정을 수행한다.

1. 각 픽셀마다 첫번째 특징 벡터 v 를 추출하여 각 픽셀들이 두개의 클러스터(스왓,배경)로 분류되도록 클러스터링을 수행한다.
 - (a) 클러스터 중심을 초기화한다. 스왓의 중심벡터는 $[0, R \text{ 값의 평균}, G \text{ 값의 평균}, B \text{ 값의 평균}]^T$ 이고 배경의 중심 벡터는 $[1, R \text{ 값의 평균}, G \text{ 값의 평균}, B \text{ 값의 평균}]^T$ 으로 설정한다.
 - (b) K-means 클러스터링 알고리즘을 적용시킨다.
 - (c) 클러스터링 결과를 저장한다.
2. 두번째 특징 벡터 u 를 이용하여 앞의 과정을 반복한다.
 - (a) 클러스터 중심을 초기화한다. 스왓의 중심벡터는 $[R \text{ 값의 평균}, G \text{ 값의 평균}, B \text{ 값의 평균}]^T$ 이고 배경의 중심 벡터는 $[R \text{ 값의 최소값}, G \text{ 값의 최소값}, B \text{ 값의 최소값}]^T$ 으로 설정한다.
 - (b) K-means 클러스터링 알고리즘을 적용시킨다.
 - (c) 클러스터링 결과를 저장한다.
3. 두 번의 클러스터링 결과에서 클러스터링된 스왓 영역의 교집합을 구한다.

여기서 사각 영역의 중심으로부터의 거리 d 가 특징으로 이용된 이유는, 이전의 과정을 통해 사각 영역의 중심에 스왓이 위치하기 때문에, 사각 영역의 중심에 있는 픽셀일수록 스왓의 영역이 될 확률이 높기 때문이다. 또한, 사각 영역의 중심으로부터의 거리 d 가 특징으로 이용되더라도 이러한 특징이 RGB 색상에 비해 클러스터링 과정에서 상대적으로 스왓의 영역의 선택에 일관된 큰 영향을 미치지 않을 것이다. 따라서 특징 d 를 포함한 클러스터링 결과와 특징 d 를 포함하지 않은 클러스터링 결과를 비교하여 특징 d 가 얼마나 큰 영향을 미치는지 알 수 있으며, 두 클러스터링 결과의 교집합을 통해서 RGB 색상 또는 d 에만 크게 영향받지 않는 영역을 구할 수 있다.

3.4 스왓의 밝기 측정

스왓의 밝기 측정은 마이크로어레이 분석에 있어 핵심적인 부분이다. 왜냐하면 마이크로어레이 특성상 스왓의 밝기 측정을 통해서, 여러가지 생물학적인 반응-DNA-DNA

, 단백질-단백질, 항원-항체-을 관찰할 수 있으며, 반응이 일어난 물질의 양을 측정하는데도 이용이 되기 때문이다.[7].

이제 스팟의 최종적인 밝기를 측정하여 보자. 마이크로어레이는 특성상 밝기의 지역적인 편차가 존재할 수 있다. 이러한 편차로 인해 배경의 밝기가 일정하지 않다. 경우에 따라 특정 영역에서 배경의 밝기가 다른 영역에 존재하는 스팟의 밝기보다 더 밝을 수도 있다. 이러한 문제를 해결하기 위해서 정규화된 밝기 측정이 필요하다. 제한된 지역에서의 배경의 밝기는 전체적인 배경의 밝기와 비교해볼 때 밝기의 편차가 크게 나지 않으며, 배경이 밝아지면 스팟도 이에 비례하여 밝아진다. 따라서 사각 영역 내에서, 스팟인 픽셀들의 밝기의 평균에 배경인 픽셀들의 밝기의 평균을 빼 값-정규화된 밝기-을 이용하면 지역적인 편차 없이 순수한 스팟의 밝기만 구할 수 있다.

4. 실험 결과 및 분석

본 실험에 이용된 영상은 단백질 마이크로어레이 실험 영상으로서 SPR 센서를 이용하여 CCD 카메라로 촬영한 영상이다. 그림 3은 촬영된 단백질 마이크로어레이 영상을 나타낸다.



그림 3: 단백질 마이크로어레이 영상

그림 4은 3.2절의 과정을 거쳐서 사각 격자로 나누어진 영상이다. 흰선은 찾을 수 있는 valley를 검은선은 추론된 valley를 나타낸다. 그림 5는 특정한 사각 영역을 3.3절의

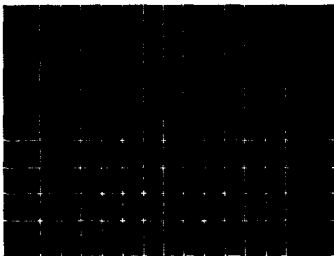


그림 4: 사각 격자로 나누어진 영상

과정을 통하여 클러스터링을 수행한 영상을 보여준다. 그림 5-(a)는 원래의 영상이고, 그림 5-(b)는 식 1에서의 특징 벡터 v 를 이용하여 클러스터링을 수행한 결과이다. 그림 5-(c)는 식 2에서의 특징 벡터 u 를 이용하여 클러스터링을 수행한 결과이다. 그림 5-(d)는 앞의 두 영상에서 스팟인 영역이 서로 겹쳐지는 부분이다. 표 1은 앞의 세가지 영상에 대한 스팟의 밝기, 배경의 밝기, 정규화된 스팟의 밝기를 나타낸다.

표 1에서 알 수 있듯이, 색상정보만 이용하여 분리한 스팟과 최종적으로 분리한 스팟간의 정규화된 밝기가 크게

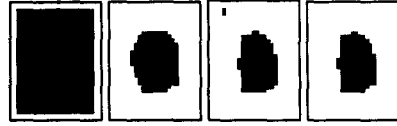


그림 5: (a)원래 영상, (b)색상정보와 거리 d 를 이용한 클러스터링 영상, (c)색상정보만 이용한 클러스터링 영상, (d)두 영상의 교집합 영상, 최종 영상

	그림5-(b)	그림5-(c)	그림5-(d)
스팟의 밝기	109.1453	114.2445	114.3984
배경의 밝기	76.3202	76.9379	77.1222
정규화된 밝기	32.8250	37.3066	37.2762

표 1: 각각의 그림에 대한 밝기 비교

차이가 나지 않음을 알 수 있다. 이러한 이유는 원래 영상이 색상 정보만으로도 충분히 영역의 분리가 되기 때문이다. 하지만, 스팟과 배경의 밝기 차이가 뚜렷하지 않은 경우에는 색상 정보만 이용하여 스팟의 영역을 추출하면 불안정하게 영역이 추출된다. 따라서, 특징 d 를 이용하여 클러스터링할 경우, 색상보다는 특징 d 에 크게 영향을 받게 되어 사각 영역의 중심 가까이 있는 모든 픽셀들이 스팟의 영역으로 추출된다. 그리고 최종적으로 분리한 스팟은 이러한 요소를 반영하여 사각 영역의 중심에 있는 픽셀들 중에서 밝기가 큰 픽셀들이 스팟인 영역으로 나타난다. 따라서, 스팟의 밝기는 어느정도 안정적인 값이 도출된다.

5. 결론 및 향후 연구 방향

기존의 방법들이 이미지 전체에서 스팟의 영역을 직접적으로 분리해내었다면, 본 논문에서는 마이크로어레이의 특성을 이용하여 격자로 나누고, 지역적인 방법으로 스팟의 영역을 분리해냈다. 또한, 기존의 방법들이 스팟 영역 분할에 중점을 둔 것에 비해, 본 논문에서는 정확한 스팟의 밝기 측정에 중점을 두었다. 본 논문에서 제시한 방법은 외부 요소-배경의 밝기 및 지역적 편차, 스팟과 배경의 밝기의 차이-의 변화에 안정적이고(stable) 융통성 있게 대처 가능하다. 향후에는 보다 정확한 밝기 측정을 위해서 노이즈-마이크로어레이의 굵힘자국, 반질, 시료가 묻은 흔적-를 제거하는 방법을 시행하고자 한다.

ACKNOWLEDGEMENTS

본 과제는 과학기술부 특정 연구 개발 사업(신기술 융합 사업)으로 지원되었습니다.

참고문헌

- [1] C. T. Campbell, "Surface plasmon resonance(spr) biosensor development," the Charles Campbell Research Group, <http://www.cpac.washington.edu/~campbell/projects/plasmon.pdf>.
- [2] C. E. B. et al, "Surface plasmon resonance multisensing," in *Anal. Chem.*, vol. 70, 1998, pp. 703-706.
- [3] J. Canny, "A computational approach to edge detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, 1986, pp. 679-698.
- [4] R. C. Gonzalez, *Digital Image Processing*. Prentice Hall, 2002.
- [5] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [6] D. G. S. Richard O. Duda, Peter E. Hart, *Pattern Classification*. Wiley-Interscience Publication, 2000.
- [7] L. S. J. et al, "Quantitative interpretation of the response of surface plasmon resonance sensors to adsorbed films," in *Langmuir*, vol. 14, 1998, pp. 5636-5648.