

## 다중 피크의 영역 성장 기법에 의한 전기영동 젤의 영상 분석

(Image Analysis of Electrophoresis Gels by using Region Growing with Multiple Peaks)

김 영 원 <sup>†</sup> 전 병 환 <sup>\*\*</sup>

(Young Won Kim) (Byung Hwan Jun)

**요 약** 최근 생명공학(BT)에 대한 관심이 집중되면서, 새로운 생리활성 물질을 찾거나 유전자 정보를 분석하기 위한 목적으로 전기영동 젤의 영상 분석 기술에 대한 요구가 급증하고 있다. 이를 위해서는 젤 영상의 레인에서 각 밴드의 위치와 양을 정확히 측정해야 한다. 기존 연구에서는 주로 레인의 프로파일에 피크를 탐색하는 접근방법을 사용하는데, 이 피크의 위치는 밴드에 있는 최대 밝기 화소의 위치도 아니고 더욱이 밴드 무게중심의 위치도 아니기 때문에 밴드의 대표 위치로 인정하기 어렵다. 또한, 피크 추출을 쉽게 하기 위해 다양한 영상 향상 처리를 적용하기 때문에 밴드의 양을 측정하기에는 부적절한 경우가 많다. 본 논문에서는 영상의 상대적인 밝기를 변화시키지 않으면서 먼저 밴드의 영역을 추출한 후, 밴드 영역의 밝기 합으로 양을 구하고 이의 무게중심을 밴드 위치로 정하는 방식을 채택한다. 실제로, 먼저 젤 영상 히스토그램에 엔트로피기반 임계치를 설정하여 레인을 추출한 후, 밴드 영역 추출을 위해 서로 다른 세 가지 방법을 시도한다. 첫째, 추출된 레인을 이등분하는 중심선을 탐색하여 피크와 밸리를 찾고, 피크의 상하 밸리를 각 밴드의 최소 포함 박스영역으로 지정하는 방법(MER), 둘째, 앞의 방법에서와 같이 구한 피크를 영역 성장의 시드로 사용하여 이웃하는 밴드와의 중첩을 해결하면서 밴드 영역을 추출하는 방법(RG-1), 셋째, 이와 달리 레인을 삼등분하는 두 탐색선에서 피크를 찾고 동일한 밴드에 속하는 피크 쌍을 결정한 후 영역을 성장하는 방법(RG-2)을 제안한다. 이상의 세 방법을 비교하기 위해 밴드의 위치 및 양을 측정한 결과, 밴드 위치의 평균 오차는 레인의 길이를 단위 크기로 정규화 할 때, MER 방법이 6%, RG-1 방법이 3%, RG-2 방법이 1%로 나타났다. 또한, 밴드 양의 평균 오차는 레인 내 밴드들의 양의 합을 단위 크기로 정규화 할 때, MER 방법이 8%, RG-1 방법이 5%, RG-2 방법이 2%로 나타났다. 결과적으로, RG-2 방법이 밴드의 위치 및 양 추출에 있어서 정확도가 가장 높은 것으로 판명되었다.

**키워드** : 전기영동, 레인 및 밴드 추출, 영역 성장, 유전자 및 단백질 분석, 젤 영상 분석

**Abstract** Recently, a great interest of bio-technology(BT) is concentrated and the image analysis technique for electrophoresis gels is highly requested to analyze genetic information or to look for some new bio-activation materials. For this purpose, the location and quantity of each band in a lane should be measured. In most of existing techniques, the approach of peak searching in a profile of a lane is used. But this peak is improper as the representative of a band, because its location does not correspond to that of the brightest pixel or the center of gravity. Also, it is improper to measure band quantity in most of these approaches because various enhancement processes are commonly applied to original images to extract peaks easily. In this paper, we adopt an approach to measure accumulated brightness as a band quantity in each band region, which is extracted by not using any process of changing relative brightness, and the gravity center of the region is calculated as a band location. Actually, we first extract lanes with an entropy-based threshold calculated on a gel-image histogram. And then, three other methods are proposed and applied to extract bands. In the MER method, peaks

· 본 연구는 한국과학재단 지정 공주대학교 자원재활용 신소재 연구센터의 지원에 의한 것입니다.  
<sup>†</sup> 비 회 원 : 공주대학교 대학원 컴퓨터공학과 forever@kongju.ac.kr  
<sup>\*\*</sup> 중 신 회 원 : 공주대학교 공과대학 정보통신공학부 교수 bhjun@kongju.ac.kr  
 논문접수 : 2001년 10월 25일  
 심사완료 : 2003년 2월 13일

and valleys are searched on a vertical search line by which each lane is bisected. And the minimum enclosing rectangle of each band is set between successive two valleys. On the other hand, in the RG-1 method, each band is extracted by using region growing with a peak as a seed, separating overlapped neighbor bands. In the RG-2 method, peaks and valleys are searched on two vertical lines by which each lane is trisected, and the left and right peaks may be paired up if they seem to belong to the same band, and then each band region is grown up with a peak or both peaks if exist.

To compare above three methods, we have measured the location and amount of bands. As a result, the average errors in band location of MER, RG-1, and RG-2 were 6%, 3%, and 1%, respectively, when the lane length is normalized to a unit value. And the average errors in band amount were 8%, 5%, and 2%, respectively, when the sum of band amount is normalized to a unit value. In conclusion, RG-2 was shown to be more reliable in the accuracy of measuring the location and amount of bands.

**Key words:** electrophoresis, lane and band extraction, region growing, DNA/RNA and protein analysis, gel image analysis

1. 서론

최근 유전자 정보를 분석하거나 새로운 활성 물질을 찾기 위한 생명공학(BT)에 대한 관심이 고조되고 있다. 전기영동(electrophoresis)이란 단백질이나 핵산(RNA/DNA) 조각과 같은 분자들이 전류의 영향 하에서 아가로스 젤(agarose gel)이나 폴리아크릴아미드 젤(polyacrylamide gel)과 같은 특정 매질을 통과하여 이동하게 하는 전기화학적 분리과정이다[1]. 이는 분자의 크기에 따른 전하량의 차이를 이용한 기술이며, DNA 순서나 단백질의 종류 혹은 양을 조사하기 위한 응용에 필수적으로 사용되어 왔다.

젤 영상은 여러 레인(lane)으로 구성되고 각 레인은 하나 이상의 수평 막대 모양의 밴드(band)를 갖는데, 각 밴드의 위치와 양을 정확하게 측정하는 것이 중요하다. 초기에는 젤 영상을 사람이 직접 분석하였는데 이는 매우 방대하고 지루한 작업이다. 최근에는 이를 자동화하고 보다 신속하고 정확하게 정성적 혹은 정량적 정보를 추출하기 위해 젤 영상 분석(gel image analysis)에 대한 연구가 활발하게 진행되고 있다. 대표적인 연구기관으로는 National Human Genome Research Institute(NHGR)[2], National Institute of Mental Health(NIMH)[3], Biomathematics and Statistics Scotland(BioSS)[4] 등이 있다. 상업용 시스템 중에서 Media Cybernetics Inc.[5]의 Gel-Pro, UVitec Inc.[6]의 UVIdoc, Nonlinear Dynamics[7]의 Phoretix는 레인에 지정된 수직 탐색선의 프로파일이나 평균 밝기 프로파일에서 피크 위치를 찾아 이를 밴드의 위치로 간주하는 방법을 사용한다. 레인 프로파일에서의 피크는 밴드에 있는 최대 밝기 화소와 일치하지 않으며 더욱이 밴드 전체 영역의 무게중심도 아니기 때문에, 밴드의 대표 위치로 적절하지 않다. 또한 Eastman Kodak Inc.[8]의

EDAS 290, Alpha Innotech Inc.[9]의 Alpha Imager, AABI Inc.[10]의 Biotech Software, Vilber Lourmat [11]의 Bio-Profil, ECO Solutions Inc.[12]의 EGA System, Curvex Inc.[13]의 Gelyzer 등은 레인 추출 단계에서부터 잡음 제거나 향상을 위한 다양한 영상처리 기법을 적용하고 있다. 이로 인해 밴드의 밝기가 변화될 수 있으며 사용자의 개입에 많이 의존한다.

본 논문에서는 밴드의 영역을 먼저 추출한 후, 밴드 영역의 밝기로부터 양을 구하고 이의 무게중심을 대표 위치로 정하는 방식을 고려한다. 이를 위해, 먼저 젤 영상의 히스토그램에 엔트로피기반 임계치 설정 기법[14]을 적용하여 이진화를 하고 수직 프로젝션을 수행하여 레인들을 추출한다. 그런 다음, 각 레인에서 밴드의 위치를 검출하기 위한 탐색선의 수와 밴드 영역을 분할하는 방법에 따라, [그림 1]과 같이 세 가지 밴드 영역 추

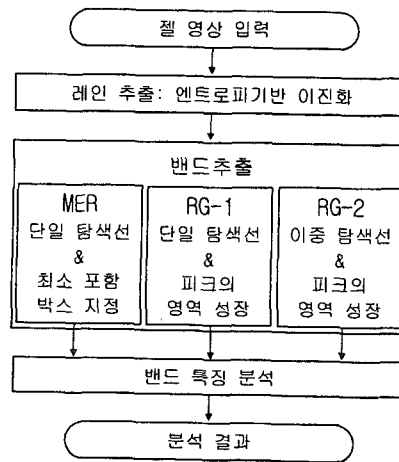


그림 1 제안하는 세 가지 젤 영상 분석 방법

출 방법을 고려한다. 즉, 레인을 이동분하는 하나의 탐색선에서 밸리와 피크를 찾아 각 밴드의 최소 포함 박스로 설정하는 방법(MER), 앞서 언급한 하나의 중심탐색선에서 피크를 찾고 이를 시드(seed)로 사용하여 밴드의 영역을 성장시키는 방법(RG-1), 그리고 레인을 삼등분하는 두 탐색선에서 피크 쌍을 결정한 후 이들을 모두 시드로 사용하여 밴드의 영역을 성장시키는 방법(RG-2)을 제안하고 비교한다.

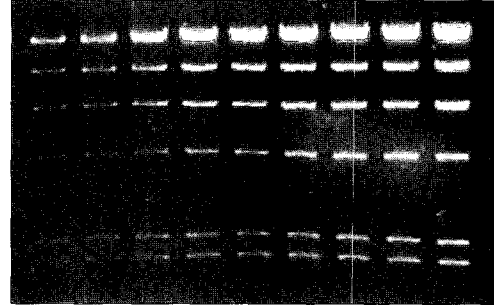
2. 레인 추출

젤 영상에 있는 레인들은 일반적으로 폭과 간격이 일정하고 밴드와 배경간의 밝기에 어느 정도 차이가 있는 등 비교적 추출하기 쉬운 특성을 가지고 있다. 예로써, Machado 등[15]은 영상을 최소 필터로 전처리 한 후, 수직으로 밝기의 합이나 최대로 프로젝션한 결과를 최소나 중간 값으로 필터링하여 잡음을 제거한 다음, 1차 모멘트를 이용한 반복적인 기법으로 레인의 중심축을 추출한다. 그러나 이 방법에서는 레인 사이의 배경 구간을 분리하지 못하고 레인의 중심축만을 추출하며, 다양한 영상처리에 의해 밴드 밝기에 왜곡이 생길 가능성이 크다. Jedra 등[16]은 수직라인별 평균 밝기 프로파일을 스무딩(smoothing)한 후 반복적인 임계치 설정 기법으로 레인을 추출하는데, 이 방법에서는 영상의 배경 밝기 성분이 제거되지 않는다. Ye 등[17]은 원형의 구축 요소에 의한 모폴로지의 폐쇄(closing)를 적용하여 배경을 산정하고 이를 원 영상에서 빼는 'Top-Hat Transform'을 수행하여 점진적으로 변화하는 배경을 제거한 후, 프로파일의 에지를 분석하여 레인을 추출하는데, 이 방법에서는 구축 요소의 모양과 크기에 따라 결과 영상에서의 밝기 변화가 크고, 역시 밴드의 밝기에 왜곡이 생길 수 있다.

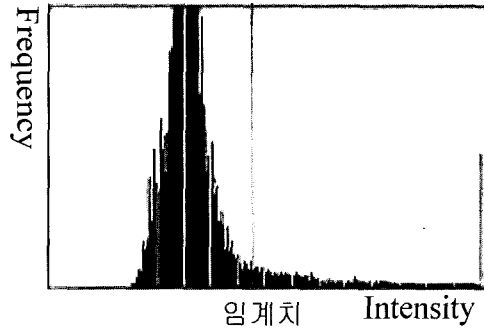
본 논문에서는 밴드의 밝기를 왜곡시킬 수 있는 잡음 제거나 영상 향상을 지양하고, 영상 히스토그램에 엔트로피 기반 임계치 설정 기법[14]을 적용하여 이진화 한 후, 이를 프로젝션하여 레인을 추출하는 방법을 제안한다.

[그림 2]는 제안하는 레인 추출 과정이다. 먼저, 원 영상 (a)에서 명암도  $i$ 의 확률을  $p_i$ 라고 할 때, 엔트로피  $H(i) = H_w(i) + H_B(i)$ 를 최대로 하는 임계치  $t$ 를 구하면 (b)와 같다.

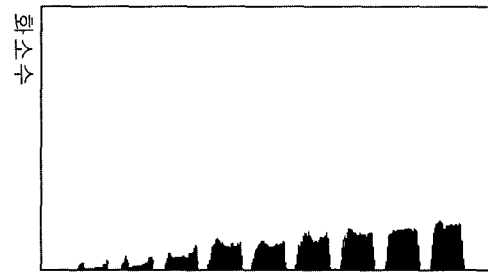
$$\text{이때, } P_i = \sum_{j=0}^i p_j \text{ 이고, 물체 확률분포 } W: \frac{p_0}{P_i}, \frac{p_1}{P_i}, \dots, \frac{p_i}{P_i} \text{의 엔트로피는 } H_w(i) = - \sum_{j=0}^i \frac{p_j}{P_i} \log \left( \frac{p_j}{P_i} \right) \text{이고,}$$



(a) 원 영상

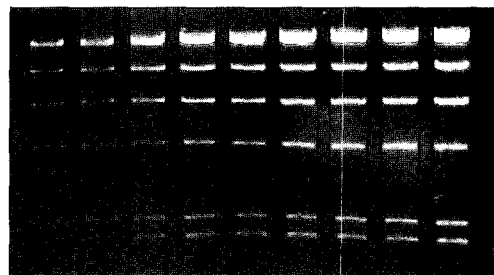


(b) 히스토그램



영상 X좌표

(c) 수직 프로젝션 결과



(d) 레인 추출 영상  
그림 2 레인 추출 과정

배경 확률분포  $B: \frac{p_{t+1}}{1-p_t}, \frac{p_{t+2}}{1-p_t}, \dots, \frac{p_{255}}{1-p_t}$  의 엔트로피는  $H_B(t) = \log\left(\frac{p_t}{1-p_t}\right) - \sum_{i=t+1}^{255} \frac{p_i}{1-p_t}$  이다.

이 임계치로 영상을 이진화하고 수직으로 프로젝션하면 (c)와 같이 된다. 추출된 라인 후보들의 평균 폭을 구하고, 이 폭의 일정 비율(70%) 미만인 라인 후보들을 잡음으로 간주하여 제거하면 최종적으로 (d)와 같이 레인을 추출할 수 있다.

### 3. 밴드 추출 및 분석

일반적으로 젤 영상에서 밴드의 이동 거리는 해당 분자량에 반비례하고 전기영동 시간에 비례하며, 각 밴드의 밝기는 해당 물질 성분(분자)의 양에 비례한다. 또한, 전기영동 과정에서 전류가 균일하지 못할 경우 밴드의 모양이 왜곡되거나 혹은 일부분이 흐려지거나 심지어 가운데가 끊어진 형태로도 나타날 수 있다. 이러한 밴드를 추출하기 위한 방법의 예로써, 황덕인 등[18]은 Hessian 행렬에 의한 DNA 밴드의 지형적 특성과 곡률을 이용하여 밴드의 위치를 추출하는 방법을 제안하는데, 이 방법에서 추출하는 밴드 위치는 중심이 아니라 한쪽 경계 부분에 해당하므로 밴드 두께가 다양하게 나타나는 단백질 분석에는 적용하기 어려우며 양의 추출은 고려하지 않고 있다. Jedra 등[16]은 앞서 언급한 라인 추출에서와 유사하게 각 레인의 평균 밝기 프로파일을 구하고 이를 스무딩하여 TDNN(Temporal-Delay Neural Network)이나 TOM(Temporal Organization Map)을 사용하여 참조 패턴과 정합을 시도한다. 따라서 이미 구축된 참조패턴 집합과의 정합에 의한 종류 판별만이 가능하고 구성 성분들의 양을 측정하지는 못한다. Ye 등[17]은 각 라인 영상의 평균 밝기 프로파일에서 피크를 탐지하고 주변의 두 경계를 찾아 그 사이의 면적을 해당 밴드의 양으로 간주한다. 그러나 앞서 언급한 라인 추출에서 항상 처리된 영상을 사용하기 때문에 양의 측정이 정확하지 않고, 특히 밴드간의 중첩이 있는 경우 정교한 영역 분할이 어렵다. Bajla 등[19]은 각 라인에서 DNA 밴드를 추출하기 위해, 먼저 GDD(Geometry-Driven Diffusion) 필터로 정제한 후, 누적 행 차이의 지역 극대를 밴드의 경계로 추출하고 계단상수 함수로 극대 쌍을 결정하여 밴드의 박스를 지정한다음, 최대 미분에 의해 밴드 경계를 추출한 후 무게중심축을 구하여 밴드의 위치를 파악한다. 이 방법은 프로파일 접근방법을 탈피하고 있고 무게중심 축을 찾는 등 비교적 정교하게 위치를 추출하고 있으나, 영상 정제 과

정에서의 밴드 밝기 변화로 밴드 양의 추출에는 부적합한 측면이 있고 사용자의 개입에 상당히 의존하고 있다.

본 논문에서는 밴드의 영역을 정교하게 추출한 후 그 위치와 양을 측정하기 위해, [그림 3]에서 나타낸 것과 같이 세 가지 방법을 고려한다. 먼저, MER방법은 하나의 탐색선과 최소 포함 박스를 결합한 방법으로 탐색선에서 찾아진 피크와 상하 밸리를 이용하여 밴드의 최소 포함 박스를 지정하고 밴드를 추출하는 방법으로 ①→②→③→④→⑦→⑧→⑪의 과정으로 수행된다. RG-1방법은 찾아진 피크를 시드로 이웃하는 화소와의 중첩을 해결하면서 영역 성장(region growing)[20]을 하는 방법으로 ①→②→③→④→⑦→⑨→⑩→⑪의 과정으로 수행된다. RG-2방법은 레인을 삼등분하는 두 탐색선에서 피크를 찾고 동일한 밴드에 속하는 피크 쌍을 결정한 후 영역을 성장하는 방법으로 ①→②→③→⑤→⑥→⑦→⑨→⑩→⑪의 과정으로 수행된다.

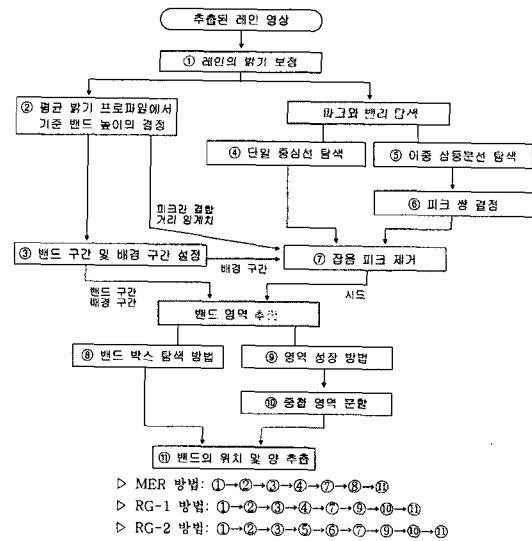


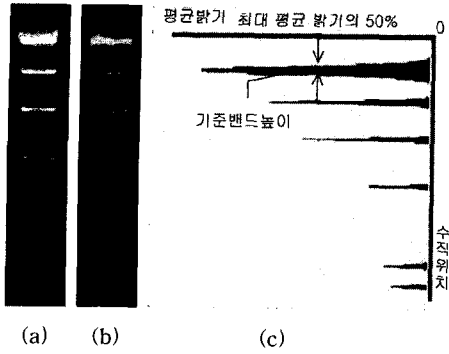
그림 3 세 가지 밴드 추출 방법의 흐름도

#### 3.1 레인의 밝기 보정

[그림 4]의 (a)와 같이 추출된 라인 영상의 각 행마다 좌우 배경의 평균 밝기를 구하여 이를 뺀으로써 배경 성분을 제거한 보정 영상은 (b)와 같이 된다. 이와 같이 하면, 밝기가 점진적으로 변하는 젤 영상에서도 적응적으로 배경 성분을 제거할 수 있다.

#### 3.2 라인 프로파일에서의 기준 밴드 높이의 설정

[그림 4]의 (b)의 보정된 라인 영상의 평균 밝기 프로파일은 (c)와 같다. 이때, 최대 평균 밝기의 50% 지점



(a)레인영상, (b)배경 성분제거, (c)평균 밝기 프로파일  
 그림 4 레인 영상에서의 기준 밴드 높이를 구하는 과정

에서 형성되는 최대 밴드 높이를 기준 밴드 높이로 사용한다.

**3.3 밴드 및 배경 구간 설정**

밴드 및 배경 구간은 [그림 5]와 같이 설정한다. 먼저, 레인의 수직 프로파일 상에서 최대 평균 밝기의 10%를 임계치로 설정하고, 프로파일의 피크에서 상하로 탐색하면서 임계치 미만이 되기 직전까지를 밴드 구간으로 설정하고, 이웃한 밴드 구간 사이를 배경 구간으로 설정한다. 이와 같이 구한 밴드 및 배경 구간에서 상하 각각 5% 마진을 제외한 것을 최종 구간으로 설정한다.

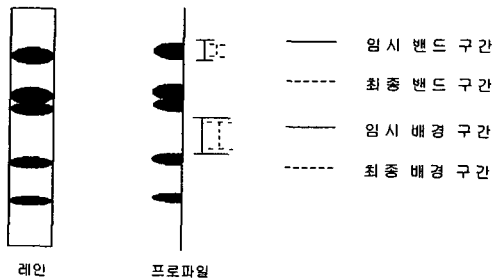


그림 5 밴드 및 배경 구간 설정

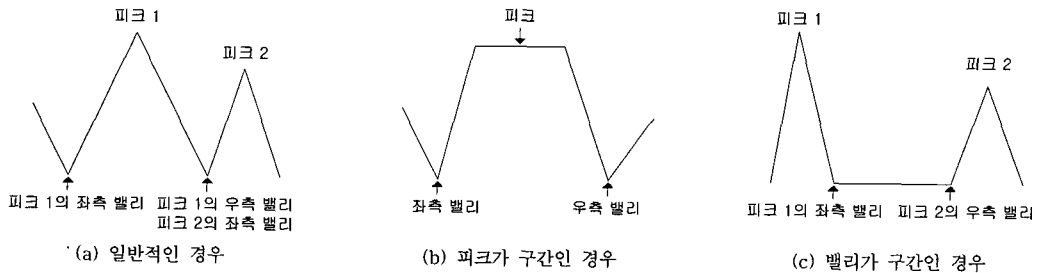


그림 6 피크가 밸리의 추출

**3.4 피크와 밸리 추출**

본 논문에서는 각 레인에 대해서 레인을 이동분하는 단일 탐색선 또는 레인을 삼등분하는 이중 탐색선에서 현재 화소( $I_{cur}$ )와 이전 화소( $I_{pre}$ )의 밝기를 비교하면서 피크와 밸리를 추출한다. 먼저 탐색선상에서 최초로 배경 밝기보다 큰 화소를 좌측 밸리로 설정하여 다음과 같은 피크 탐색 및 밸리 탐색 과정을 반복 수행한다.

▷ <피크 탐색 과정>

- 1)  $I_{cur} > I_{pre}$  : 다음 화소 탐색
- 2)  $I_{cur} = I_{pre}$  : 구간 정보 표시
- 3)  $I_{cur} < I_{pre}$  : ① (구간이 아닌 경우) 이전 화소를 피크로 결정  
 ② (구간인 경우) 구간의 중앙 위치를 피크로 인정  
 → <밸리 탐색 과정> 호출

▷ <밸리 탐색 과정>

- 1)  $I_{cur} < I_{pre}$  : 다음 화소 탐색
- 2)  $I_{cur} = I_{pre}$  : 구간 정보 표시
- 3)  $I_{cur} > I_{pre}$  :  
 ① (구간이 아닌 경우) 이전 화소를 우측 밸리로 인정하고 다음 피크의 좌측 밸리로도 인정  
 ② (구간인 경우) 구간의 시작점은 현재 피크의 우측 밸리로 인정 구간의 끝점은 다음 피크의 좌측 밸리로 인정 → <피크 탐색 과정> 호출

[그림 6]의 (a)는 피크나 밸리가 하나의 화소로 나타나는 일반적인 경우이고, (b)는 피크가 구간을 이루는 경우로 구간의 중심을 피크로 선택하는 방법을 나타내고, (c)는 밸리가 구간을 이루는 경우로 구간의 시작 화소는 좌측 피크의 우측 밸리로, 구간의 끝 화소는 우측 피크의 좌측 밸리로 선택하는 방법을 나타낸다.

[그림 7]의 (a)는 레인을 삼등분하는 두 탐색선상에서 피크를 추출한 것이고, (b)는 이동분하는 단일 탐색선상에서의 피크 추출을 나타낸 것이다. (a)의 경우에는 가



(a) 이중 탐색선



(b) 단일 탐색선

그림 7 초기 피크 추출의 예



(a) 이중탐색선



(b) 단일 탐색선

그림 8 잡음 피크가 제거된 예

운대가 흐린 아령 모양이나 심지어 끊어진 밴드의 경우에도 피크를 놓치지 않고 추출할 수 있다.

### 3.5 피크 쌍 결정

본 논문에서는 두 탐색선에서 추출된 각 피크의 수직 거리를 이용하여 형제인지를 판별하는 방법을 제안한다. 먼저, 왼쪽 수직 탐색선에서 추출된 각 피크에 대해 오른쪽 수직 탐색선에서 수직 거리가 가장 가까운 피크를 찾아 대응시키고, 마찬가지로 오른쪽 탐색선에서 추출된 피크에 대해서도 대응하는 좌측 피크를 찾는다. 그런 다음, 서로 지정한 피크 쌍이 동일한 경우에 형제 피크로 인정하고 두 개 중에서 보다 밝은 피크를 대표 피크로 간주한다. 그렇지 않고, 서로 지정한 피크쌍이 다를 경우에는 각각 단독의 대표 피크로 인정한다.

### 3.6 잡음 피크 제거

이상의 과정에서 불필요하게 추출된 잡음 피크를 제거하기 위해, 세 가지 방법을 사용한다. 즉, 이웃 피크간의 거리에 따른 제거, 피크와 밸리와 밝기 차이에 따른 제거, 배경 구간에 위치하는 피크의 제거 등이다. 이때, 이웃 피크간의 비교를 위해서는 단독 피크나 쌍을 이룬 두 피크 중 대표 피크를 사용한다.

먼저, 이웃 피크간의 거리에 따른 제거는 이웃하는 피크간 수직 거리가 가장 짧은 이웃을 찾아 그 거리가 임계치 이내이면 화소 밝기가 작은 피크를 제거하는 과정을 반복하는 방법이다. 이때, 거리 임계치는 앞서 언급한 기준 밴드 높이의 80%로 한다. 피크와 밸리의 밝기 차이에 따른 제거는 각 밸리마다 좌우 피크와의 밝기

차이를 구하고, 이 중 작은 차이가 큰 차이의 30% 이하이면 현재 밸리와 작은 밝기 차이의 피크를 제거한다. 또한, 피크의 위치가 배경 구간에 속할 경우 잡음

피크로 간주하여 제거한다.

최종적으로 살아남은 피크들은 밴드 영역 성장의 시드로 사용된다. [그림 8]은 [그림 7]에서 추출된 피크들 중 잡음 피크가 제거된 결과이다.

### 3.7 밴드 영역 추출

밴드 영역 추출을 위한 밝기 임계치는 다음과 같이 각 밴드마다 적응적으로 결정한다. 각 밴드에 대해 위 배경의 평균 밝기와 아래 배경의 평균 밝기 중에서 큰 값을 기준 배경 밝기라 하자. 밴드 구간에서의 최대 평균 밝기와 기준 배경 밝기의 차이의 15%를 기준 배경 밝기에 더하여 이를 임계치로 사용한다.

밴드 영역 추출을 위해서는, 앞서 언급한 바와 같이, 레인의 폭 구간과 밴드의 상하 밸리를 경계로 하는 밴드의 최소 포함 박스를 설정하고 이 박스 내에서 임계치 이상인 화소를 밴드 영역으로 추출하는 방법(MER 방법), 단일 탐색선에서 추출된 피크를 시드로 임계치 이상인 이웃 화소를 포함시키면서 영역을 성장시키는 방법(RG-1 방법), 이중 탐색선에서 추출된 형제 피크 중 먼저 보다 밝은 피크를 시드로 영역을 성장시키고 이 영역 내에 나머지 형제 피크가 포함되지 않은 경우 그 피크를 시드로 밴드 영역을 추가 성장하는 방법(RG-2 방법)을 사용한다. 이때, 추출된 영역이 매우 작아 그 폭이 밴드 폭의 30% 미만인 경우에는 인정하지 않음으로써 잡음 영역을 제거할 수 있다.

### 3.8 밴드간의 중첩 영역 분할

영역 성장 방법의 경우 이웃한 밴드의 분자량이 유사하면 밴드간의 중첩이 발생할 수 있다. 중첩 영역을 분할하기 위해서는, 먼저 단독인 경우에는 피크의 y축 좌표를, 그리고 형제인 경우에는 평균 y축 좌표를 밴드의

기준 수평축으로 결정한다. 그런 다음, 두 수평축 사이의 중첩된 영역을 각 열마다 수직으로 탐색하면서 최소 밝기 화소를 경계로 중첩 영역을 분할한다.

[그림 9]는 중첩된 상하 밴드간의 각 열을 수직으로 탐색할 때 수직 탐색선 ①과 같이 중첩 구간이 있는 경우와 ②와 같이 중첩에 배경 구간이 있어 중첩이 없는 경우의 예를 나타낸다.

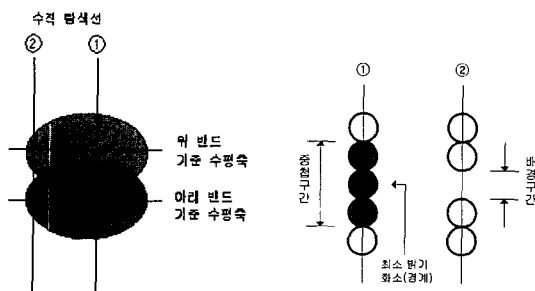


그림 9 밴드간의 중첩 영역 분할

#### 4. 실험 및 결과 분석

##### 4.1 실험 환경

실험을 위해서 단백질 시료는 SDS-6H와 SDS-7을, DNA 시료는 Lambda DNA Hind III Digest와 100bp Ladder를 전기영동 하였다. 이때 정확한 밴드의 위치 및 양을 측정하기 위해, 레인 내에 투여하는 물질의 양을 일정하게 증가시킨 후, 전기영동 시간 및 투여하는 물질의 양에 따라 다양한 영상을 산출하고, 영상 취득 장치를 이용하여 약 320×240 내외의 크기를 갖는 영상 총 26장을 취득하였다.

##### 4.2 평가 기준

앞서 제안한 MER, RG-1, RG-2 방법이 밴드의 위치 및 양을 얼마나 정확하게 측정하는지를 평가하기 위하여 단순 평가 및 정밀 평가를 수행한다.

단순 평가 기준은 [표 1]과 같다. 먼저, 밴드 수의 오차는 총 밴드 수  $N_T$ 에 대한 레인 내 밴드 수 오차의 합  $N_p$ 의 비율  $R_p$ 로 계산하고, 밴드 양 증가의 오차는 총 밴드 양 증가 횟수  $N_I$ 에 대한 관찰된 밴드 양 증가의 총 오류 횟수  $N_q$ 의 비율  $R_q$ 로 계산한다.

한편, 정밀 평가 기준은 [표 2]와 같다. 먼저, 밴드 위치의 정밀 평가 기준은 첫 번째 밴드의 중심에서 마지막 밴드의 중심까지의 수직 거리를 단위 길이로 정규화할 때, 레인들 사이에 대응하는 각 밴드  $j$ 에 대한 오차  $e_p(j)$ 의 평균  $E_p$ 를 사용한다. 밴드 양의 정밀 평가 기

표 1 단순 평가 기준

밴드 수의 오차	밴드 양 증가의 오차
$R_p = \frac{N_p}{N_T}$	$R_q = \frac{N_q}{N_I}$
$N_T = N_B \times N_L$	$N_I = N_B \times (N_L - 1)$
$N_p = \sum_{i=1}^{N_L}  N_B - n_B(i) $	$N_q = \sum_{i=1}^{N_L-1} \sum_{j=1}^{N_B} f(i, j)$
$n_B(i)$ : 레인 $i$ 에서 추출된 밴드 수	$f(i, j) = \begin{cases} 0 & q_{ij} < q_{i+1, j} \\ 1 & q_{ij} \geq q_{i+1, j} \end{cases}$
$N_L$ : 실제 레인 수	$q_{ij}$ : 레인 $i$ 에서 밴드 $j$ 의 정규화된 양
$N_B$ : 각 레인내의 실제 밴드 수	( $0.0 \leq q_{ij} \leq 1.0$ )

표 2 정밀 평가 방법

밴드 위치	밴드 양
$E_p = \frac{\sum_{j=1}^{N_B} e_p(j)}{N_B}$	$E_q = \frac{\sum_{j=1}^{N_B} e_q(j)}{N_B}$
$e_p(j) = \frac{\sum_{i=1}^{N_L}  y_i - y_{ij} }{N_L}$	$e_q(j) = \frac{\sum_{i=1}^{N_L}  q_i - q_{ij} }{N_L}$
$\bar{y}_j = \frac{\sum_{i=1}^{N_L} y_{ij}}{N_L}$	$\bar{q}_j = \frac{\sum_{i=1}^{N_L} q_{ij}}{N_L}$
$y_{ij}$ : 레인 $i$ 에서 밴드 $j$ 의 정규화된 위치	$q_{ij}$ : 레인 $i$ 에서 밴드 $j$ 의 정규화된 양
( $0.0 \leq y_{ij} \leq 1.0$ )	( $0.0 \leq q_{ij} \leq 1.0$ )
$N_B$ : 각 레인내의 실제 밴드 수	$N_L$ : 실제 레인 수

준은 각 레인 내 모든 밴드 양의 합을 단위 크기로 정규화 할 때, 레인 간에 대응하는 각 밴드 양에 대한 오차  $e_q(j)$ 의 평균  $E_q$ 를 사용한다.

##### 4.3 평가 결과

제안한 세 가지 방법으로 추출된 밴드에 대해 단순 평가와 정밀 평가를 수행한 결과는 [표 3]과 같다. 이때, 밴드 수가 정확하게 추출된 레인에 대해서만 정밀 평가를 수행한다.

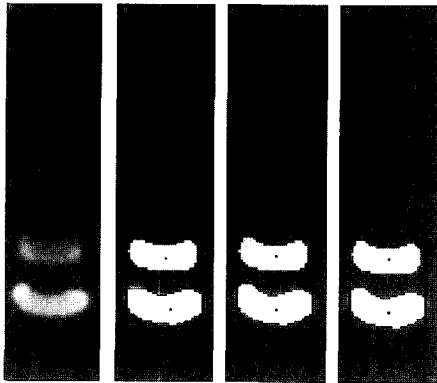
표 3 밴드 추출 방법의 평가 결과

평가 기준		분석 방법		
		MER	RG-1	RG-2
단순 평가	밴드 수의 오류율 $R_p$	13%	10%	7%
	밴드 양 증가의 오류율 $R_q$	15%	11%	8%
정밀 평가	밴드 위치의 평균 오차 $E_p$	6%	3%	1%
	밴드 양의 평균 오차 $E_q$	8%	5%	2%

제안하는 세 가지 방법으로 밴드 영역을 추출한 결과는 다음과 같다. 이때, 이웃한 밴드를 구분할 수 있도록 회색과 흰색을 번갈아가면서 사용하고 있고, 중앙의 검은 점들은 각 밴드의 위치인 무게 중심을 나타낸다.

[그림 10]의 (a)는 휘어진 밴드가 있는 원 영상이다.

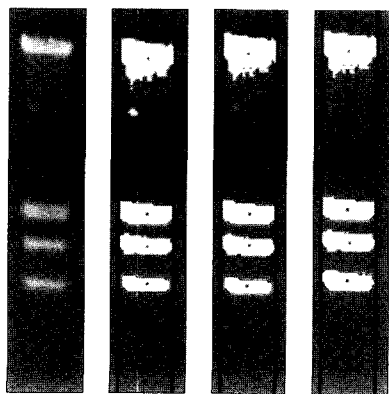
『MER』 방법으로 추출한 결과 (b)에서는 최소 포함 박스로 제한되어 두 번째 밴드의 윗부분이 잘려있는 반면, 영역 성장을 적용한 (c)와 (d)에서는 밴드의 형태가 제대로 추출되어 있다.



(a) 원영상 (b) MER (c) RG-1 (d) RG-2

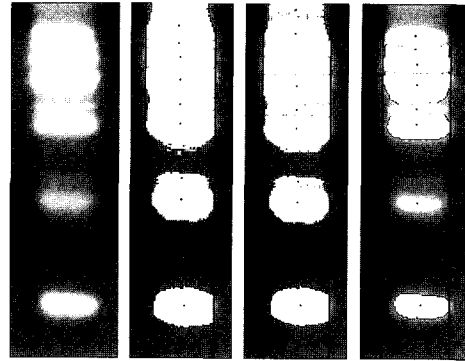
그림 10 휘어진 밴드 추출의 예

[그림 11]의 (a)는 잡음이 포함된 원 영상이다. MER 방법으로 추출한 결과 (b)에서는 최소 포함 박스 내의 잡음을 밴드 영역에 포함시켜 추출하는 반면, 영역 성장을 적용한 (c)와 (d)에서는 떨어져 있는 잡음이 포함되지 않는다.



(a) 원 영상 (b) MER (c) RG-1 (d) RG-2

[그림 11] 잡음이 포함된 밴드 추출의 예

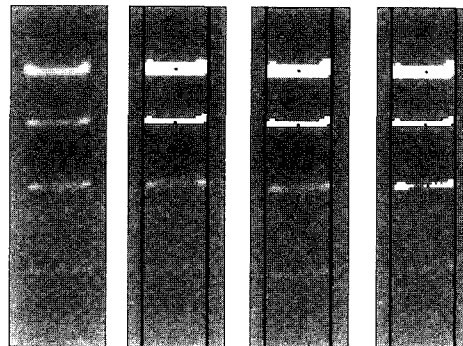


(a) 원 영상 (b) MER (c) RG-1 (d) RG-2

그림 12 중첩이 심함 밴드 추출의 예

[그림 12]의 (a)는 전기영동 시간의 부족으로 밴드간의 중첩이 발생한 원 영상이다. 밴드의 영역이 최소 포함 박스내로 제한되는 (b)의 경우에는, (c) 혹은 (d)와 달리 밴드의 상하 경계가 평탄하게 추출되고 있다. 이때, 중앙 탐색선 상에서 피크를 탐색하는 (a)와 (b)의 경우에 맨 윗부분에 있는 잡음을 시드로 인정하여 잘못된 밴드를 추출하고 있다.

[그림 13]의 (a)는 밴드의 중간 부분이 매우 흐린 원 영상이다. 중앙 탐색선에서 피크를 탐색하는 (b)와 (c)의 경우에는 세 번째 밴드를 검출하지 못하는 반면, 이 중 탐색선에서 피크를 찾는 (d)에서는 끊어진 경우에도 하나의 밴드로 추출하고 있다. 전기영동 과정에서 밴드의 가운데가 가장자리보다 흐려지는 현상이 빈번히 발생하므로 RG-2 방법을 사용하면 이러한 문제에 대처할 수 있다. 한편, 매우 흐리게 나타난 마지막 밴드는 세 방법 모두 추출하지 못하고 있다.



(a) 원 영상 (b) MER (c) RG-1 (d) RG-2

그림 13 중간이 끊어진 밴드 추출의 예



[그림 14]는 기존 시스템에서 주로 사용하는 최대 또는 평균 프로파일 방법과 제안하는 RG-2 방법을 사용하여 추출된 밴드의 위치를 나타낸다. (a)는 레인의 각 행에서 최대 밝기값을 취하는 최대 프로파일을 사용한 것으로, 잡음에 민감하고 밴드의 위치가 중앙이 아닌 가장자리에서 흔히 나타날 수 있다. (b)는 평균 프로파일을 사용한 것으로, 평균 밝기의 최대 위치를 레인의 중앙에 표시하였다. 일반적으로 평균 밝기의 최대 위치는 밴드에 있는 최대 밝기 화소와 일치하지 않으며, 특히 밴드가 휘거나 기울어진 경우에는 밴드 영역이 아닌 배경을 많이 포함시켜 평균하기 때문에 밴드의 대표 위치로 사용하기에는 부적절한 경우가 많다. (c)는 제안하는 RG-2 방법을 적용하여 밴드 영역의 무게중심을 밴드의 위치로 추출한 결과이다. 이렇게 추출된 밴드 위치는 주로 밴드의 중앙에서 나타남을 관찰할 수 있다.

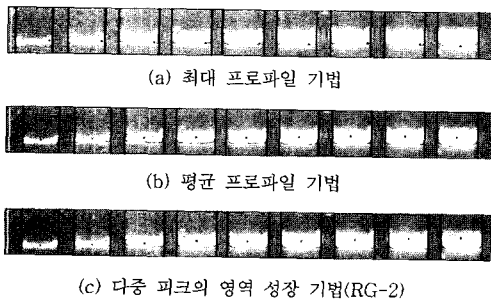


그림 14 프로파일 기법과 다중 피크의 영역 성장 기법의 밴드 위치 추출 예

[그림 15]는 생체 물질 분석을 위해 산출된 실제 젤 영상 상에 제안하는 RG-2 방법을 적용하여 밴드를 추출한 결과이다. (a)는 밴드들의 크기가 현저하게 다르더라도 밴드를 제대로 추출하는 것을 나타내고, (b)는 잡음이 많은 영상에서도 밴드를 정확하게 추출하는 것을 보여준다. (c)에서는 심하게 휘어진 경우에도 밴드를 원래 형태로 추출하는 것을 볼 수 있다. (d)에서는 맨 좌측 레인과 맨 우측 레인의 첫 밴드들과 같이 밴드가 번진 경우에도 적절한 경계를 검출하고 있다. 그러나 첫 레인과 둘째 레인의 마지막 밴드들과 같이 주변의 배경 밝기와 유사하게 심하게 흐린 경우에는 검출하지 못하고 있다. 이를 해결하기 위해서는 적응적인 임계치 설정 과정에서 보완이 필요하다. 한편, 본 연구에서는 밴드의 밝기가 물질의 농도에 비례한다고 가정하고 무게 중심을 구하고 있다. 보다 정교한 무게 중심을 찾기 위해서는 밴드 밝기와 농도와의 관계를 규명하는 연구가 뒤따라야 한다.

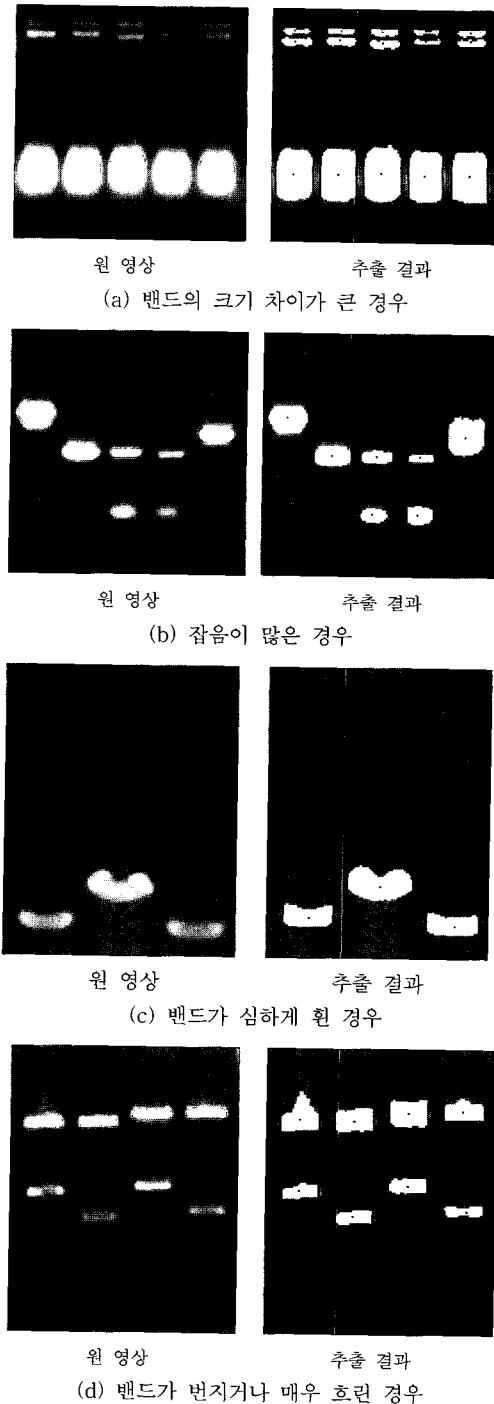


그림 15 RG-2에 의한 생체 물질 분석용 실제 젤 영상들의 추출 결과

**5. 결론**

본 논문에서는 프로파일기반 접근방법을 사용하는 대부분의 기존 연구들이 갖는 밴드의 위치와 양 추출의 부정확성을 보완하기 위해, 먼저 밴드의 양을 왜곡시킬 수 있는 영상처리를 가급적 지양하면서도 엔트로피와 프로젝션을 기반으로 추출된 레인에서 적응적으로 배경 성분을 제거한다. 그런 다음, 각 레인에서 밴드 위치와 양을 측정하기 위하여, 밴드의 위치를 포착하기 위한 탐색선의 수와 밴드 영역을 결정하는 방법에 따라 MER, RG-1, RG-2의 세 가지 분석 방법을 제안한다.

제안한 방법들을 평가한 결과, 다중 피크의 영역 성장으로 밴드 영역을 추출한 후 그 영역의 밝기를 누적하여 양을 측정하고 무게 중심을 대표 위치로 정하는 RG-2 방법이 밴드의 위치 및 양 추출에 있어서 가장 높은 정확도를 나타냈다. 향후에는, 매우 흐린 밴드도 검출할 수 있도록 적응적인 임계치 설정 과정에서의 보완이 필요하고, 전기영동 과정에서 균일하지 않은 전류 세기에 의해 가장자리보다 가운데에 위치한 레인에 있는 밴드들이 빠르게 진행하면서 생기는 스마일 현상을 보정하기 위한 연구도 필요하다.

**감사의 글**

본 논문의 실험을 위한 젤 영상 취득에 도움을 주신 공주대학교 화학과 이기평교수님, 생명과학과 전기화교수님, 벤처기업 DNA정보의 황정희연구원께 감사드립니다.

**참고 문헌**

[1] O. Gaal, G. A. Medgyesi, and L. Vereczkey, Electrophoresis in the Separation of Biological Macromolecules, John Wiley & Sons, 1980.  
 [2] <http://www.nhgri.nih.gov> : National Human Genome Research Institute(NHGRI)  
 [3] <http://www.nimh.nih.gov> : National Institute of Mental Health(NIMH)  
 [4] <http://www.bioss.sari.ac.uk> : Biomathematics and Statistics Scotland(BioSS)  
 [5] <http://www.mediacy.com> : Media Cybernetics Inc.  
 [6] <http://www.uvitec.demon.co.uk> : UVItect Inc.  
 [7] <http://www.phoretix.com> : Nonlinear Dynamics  
 [8] <http://www.kodak.com> : Eastman Kodak Inc.  
 [9] <http://www.alphainnotech.com> : Alpha Inno-

tech Inc.  
 [10] <http://www.aabi.com> : Advanced American Biotechnology & Imaging(AABI) Inc.  
 [11] <http://www.vilber.com> : Vilber Lourmat  
 [12] <http://www.ecosol.co.kr> : ECO Solutions Inc.  
 [13] <http://www.curvex.co.kr> : Curvex Inc.  
 [14] J. R. Parker, Algorithm for Image Processing and Computer Vision, John Wiley & Sons, 1997.  
 [15] A. M. C. Machado, M. F. M. Campos, A. M. Siqueira, and O. S. F. Carvalho, "An Iterative Algorithm for Segmenting Lanes in Gel Electrophoresis Images", Proc. of the XI SIBGRAPI, pp.140-146, Oct. 1997.  
 [16] M. Jedra, N. E. Khattabi, M. Limouri, and A. Essaid, "Recognition of Seed Varieties Using Neural Networks Analysis of Electrophoretic Images", Proc. of Int'l Joint Conf. on Neural Networks, Vol. 5, pp.5521-5528, Jul. 2000.  
 [17] X. Ye, C. Y. Suen, M. Cheriet, and E. Wang, "A Recent Development in Image Analysis of Electrophoresis Gels", Proc. of Vision Interface'99, pp.432-438, May 1999.  
 [18] 황덕인, 공성근, 조성원, 조동섭, 이승환, "영상의 지형적 특징에 의한 유전밴드 인식", 정보과학회논문지, B편, 제26권, 제11호, pp.1350-1358, 1999.  
 [19] I. Bajla, I. Hollander, and K. Burg, "Improvement of Electrophoretic Gel Image Analysis", Measurement Science Review, Vol. 1, No.1, pp.5-10, Sep. 2001.  
 [20] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Addison Wesley, 1993.



**김 영 원**  
 1997년 한국방송통신대학 전자계산학과(이학사). 2001년 공주대학교 대학원 전자계산학과(이학석사) 2001년~현재 공주대학교 대학원 컴퓨터공학과 박사과정 관심분야: 얼굴 분석 및 모델링, 젤 영상 분석, 가상현실 등

**전 병 환**  
 정보과학회논문지: 소프트웨어 및 응용 제 30 권 제 2 호 참조