

## 누적 기울기를 이용한 2차원 전기영동 영상의 복잡영역 분석 알고리즘

김미애\*, 윤영우\*\*

### A Complex Region Analysis Algorithm of Two Dimensional Electrophoresis Images Using Accumulated Gradients

Mi Ae Kim \*, Young Woo Yoon \*\*

#### 요약

단백질 연구에 획기적인 기여를 한 2차원 전기영동법에 의한 결과영상을 처리하기 위하여 일반적으로 제안되었던 기존의 영상처리기법들이 가지는 문제점 중 하나인 단백질을 나타내는 여러 개의 스팟들이 겹치고 포화되어 있는 복잡영역에서 단백질 스팟을 하나로 인식하거나 검출이 불가능하여 오류를 범하게 되는 문제점에 대한 해법을 제안한다. 복잡영역내의 각 점의 누적 기울기를 산출하고 누적 기울기 영상을 워터셰드 기법으로 영역 분할하여 스팟을 분리한다. 그 결과 스팟 분리에 있어서 기존의 기법들을 적용하였을 때 보다 더 우수하고 효율적인 결과를 나타내며, 2차원 전기영동 결과영상에서 숨겨져 있었던 단백질 스팟을 더 많이 검출할 수 있고, 예측의 범위 또한 확장시킨다.

#### Abstract

A solution to the problems of recognizing as one spot or detection failures for complex regions, in which many spots representing proteins are overlapped and saturated, is suggested. The accumulated gradients of each point in complex regions are calculated, and the resulting accumulated gradient image segmented using watershed technique. The suggested solution show better and efficient result than existing method for spot separation, detects more protein spots hidden in the image of 2-dimensional electrophoresis, and expands the scope of prediction.

▶ Keyword : 이차원 전기영동(2-dimensional electrophoresis), 단백질(protein), 스팟 검출(spot detection), 복잡영역(complex region), 누적 기울기(accumulated gradient).

• 제1저자 : 김미애

• 투고일 : 2009. 04. 08, 심사일 : 2009. 04. 15, 게재확정일 : 2009. 07. 10.

\* 경북전략산업기획단 평가기획팀 연구원 \*\* 영남대 컴퓨터공학과 교수

## 1. 서론

오늘날 IT기술의 발달은 생활, 문화, 경제, 산업 등 모든 영역에 걸쳐 융합되어 급속한 성장을 보이고 있다. 그 중에서도 생물학 분야로의 융합은 생물정보학(bioinformatics)이라는 새로운 학문영역을 구축하게 되어 미국, 영국, 일본 등 과학 선진국뿐만 아니라 IT분야에 있어 세계 우위를 점하고 있는 우리나라에서도 생물정보학에 대한 연구개발이 집중적 이루어지고 있고, 이러한 생물정보학 관련 기술의 발달은 방대한 생명정보 데이터베이스의 구축, 유전체의 연구, 단백질체의 연구 등에 획기적인 발전을 보여주어 실질적인 인류 혜택을 실현할 포스트 게놈 시대에 반드시 필요한 학문분야라고 할 수 있다. 그 중 단백질체학(proteomics)은 기초생물학·의학·응용생물학 분야에서 필수적이라고 할 수 있으며, 유전자의 기능, 단백질의 기능 이상 및 구조변형 유무 등을 규명하고 질병과정을 추적하는 분석기술로서 모든 단백질의 총합인 단백질체를 다루는 학문으로 생물을 구성하고 유지하는 단백질에 관한 연구가 필수적이라고 볼 수 있다. 여기서 단백질 연구에 획기적인 기여를 한 2차원 전기영동법(2-dimensional electrophoresis: 2-DE)은 조직이나 세포 내의 다수 단백질을 소량의 시료로도 높은 분리 능력으로 분리하는 방법으로서 세포내 단백질체의 변화 양상을 탐지하고, 표시할 수 있는 기술이다[1,2]. 특히 유전정보만으로 확인이 불가능한 단백질의 발현량 및 번역 후 수식(post-translational modification) 등도 쉽게 분석할 수 있어 신호전달, 독성 및 약물작용 분야 등에 그 활용범위가 점점 넓어지고 있다[3]. 이러한 정보를 바탕으로 각종 질병의 발생 원인, 진행 상태, 생리적인 변화 등에 대해 분석 할 수 있어 활발히 연구되고 있다. 2차원 전기영동의 결과영상을 처리하기 위하여 일반적으로 제안되었던 기법으로는 워터셰드, 확산모델 매칭 등이 있으며, 이러한 영상처리기법을 이용하여 영상에서 단백질을 나타내는 스팟들을 분리해 내는 기술[3-11]이 개발되어 이용되고 있으나, 아직도 보다 완벽하고 효율적인 분석을 위하여 해결해야 할 문제점들이 많이 남아 있다. 그러한 문제점중 하나는 여러 개의 스팟이 겹치고 포화되는 복잡영역(complex region)이다. 복잡영역의 정상은 고원과 같이 평평하므로 일반적인 영상처리 기법을 적용하면 스팟 검출이 불가능하거나 단 하나의 스팟으로 인정하는 오류를 범할 수 있다. 2차원 전기영동 영상에는 이러한 복잡영역이 자주 나타난다. 본 논문에서는 스팟을 역상 영상(inverse image)의 3차원으로 표현할 때 기술로부터 고원 아래까지의 비탈의 기울기가 정상을 향하는 현

상을 이용하여, 영역내의 각 점의 누적 기울기(gradient)를 산출하여 누적 기울기 영상을 워터셰드 기법을 이용하여 영역을 분할하여 스팟을 분리하는 방법을 제시한다.

## II. 관련 연구

복잡영역에 대한 기존 연구로서 논문[12]에서는 고리연산자(ring operator)로써 피크를 검출하고, 피크를 중심으로 조정된 가우스 함수를 원영상에서 뺀 나머지 영상에서 다시 고리연산자로써 피크를 검출하고, 피크를 중심으로 가우스 함수를 조정하여 복잡영역의 스팟 검출을 제안하였다. 이 방법의 문제점은 실제의 스팟은 가우스 모델보다 피크 영역이 넓은 확산모형[13]에 가까우며, 피크를 중심으로 한 대칭형이 아니라 좌우 또는 상하로 길게 늘어드는 형태가 많다. 여러 개의 스팟이 겹친 복잡영역에서는 피크부분이 더욱 넓어져 넓은 고원(plateau)이 된다. 따라서 포화된 복잡영역에서 이 방법의 결과는 예상과 많이 달라진다. 논문[14]에서는 어두운 영역의 경계선을 구하고 [그림1]과 같이 타원방정식을 영상의 경계선에 정합시키는 방법을 제안하였다. 타원방정식을 정합시킬 때 경우의 수가 NP complete하므로, 선형프로그래밍기법으로 타원방정식을 경계선에 정합시켰다. 이 방법은 2차원 정보만 이용하므로 정확도가 떨어지고 계산 복잡도가 높은 문제점이 있다.

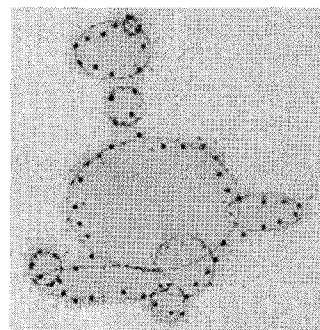


그림 1. 선형프로그래밍에 의한 타원방정식 매칭 결과(3)  
Fig. 1. The result of matching the elliptic equation through linear programming(3)

### III. 본 론

#### 1. 제안 방법

2차원 전기영동의 결과영상을 처리하는 영상처리기법들의 문제점 중의 하나인 여러 개의 스팟이 겹치고 포화되어 있는 복잡영역에서의 스팟 분리에 대한 개선 방법으로 먼저 영역내의 각 점의 누적 기울기를 산출하고, 누적 기울기 영상을 워터셰드 방법을 이용하여 영역 분할함으로써 스팟을 분리해내는 기법을 제안한다.

##### 1.1 누적 기울기

[그림 2]는 영역내의 각 점의 누적 기울기를 산출하는 그림을 나타낸다.  $p(x,y)$ 는 영상내의 화소의 위치를 나타내며,  $z$ 는 점  $p(x,y)$ 의 역상영상의 강도를 나타낸다. 영역  $R$ 내의 점  $p(x,y)$ 의 기울기 크기  $g_p$ 와 기울기 방향  $\theta_p$ 는 다음 식으로 구해진다:

$$\theta_p = \arg \left( \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y} \right).$$

$$g_p = \sqrt{\left( \frac{\partial z}{\partial x} \right)^2 + \left( \frac{\partial z}{\partial y} \right)^2}$$

여기에서  $\arg(x,y)$ 는  $x$ 축으로부터 점  $(x,y)$ 까지의 각을 의미한다.

점  $p$ 에서 시작하는 선형적  $y = \tan \theta_p \cdot x + k$  위의 점집합  $\{p_0 = p, p_1, \dots, p_i, \dots, p_n, p_{n+1}, \dots\}$ 에서  $z(p_{i-1}) \leq z(p_i) \leq z(p_{i+1}), 1 \leq i \leq n-1$  이며  $z(p_{n-1}) \leq z(p_n) \leq z(p_{n+1})$  인 점집합  $S_p = \{p, p_1, p_2, \dots, p_{i-1}, p_i, \dots, p_n\}$ 을  $p(x,y)$ 의 기울기 궤적이라 한다.

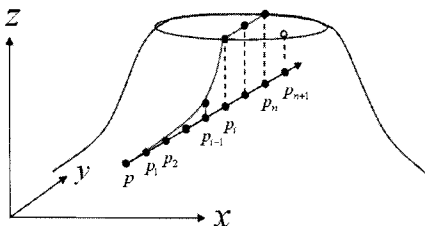


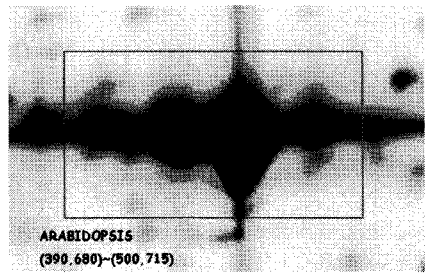
그림 2. 누적 기울기의 산출  
Fig.2. The result of accumulated gradient

영역  $R$ 내의 점  $q(x,y)$ 의 누적 기울기  $G_q$ 는 다음 식에 의해 구해진다.:

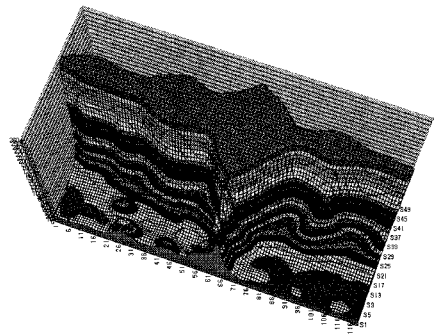
$$\forall q \in R, G_q = \sum_{p \in R} \tau_p(q) \cdot g_p.$$

$$\text{여기서 } \tau_p(q) = \begin{cases} 1, & q \in S_p \\ 0, & q \notin S_p \end{cases}$$

[그림3](a)는 여러 개의 스팟이 겹치고 포화되는 복잡영역의 예이다. [그림3](b)와 같이 복잡영역의 정상은 고원과 같이 평평하므로 워터셰드 영상처리 기법을 적용하면 [그림 4]와 같이 단 하나의 스팟으로 분할된다. [그림 5]는 이 영상의 누적 기울기를 0~255로 정규화(normalization)한 영상이다.



(a) 회색조 영상  
(a) Gray image

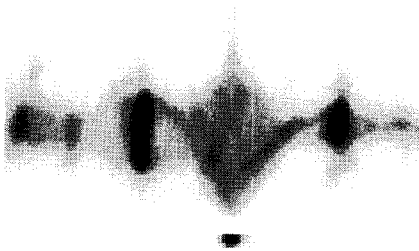


(b) 역상영상의 3차원 표현  
(b) 3-dimensional expression of inverse image

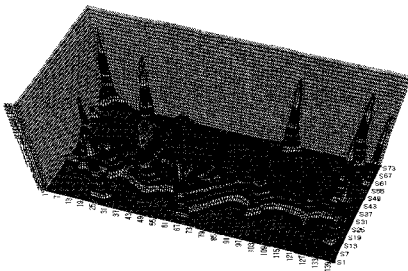
그림 3. 복잡영역의 예  
Fig.3. Example of accumulated gradient image



그림 4. 그림 3의 워터셰드 분할 결과  
Fig.4. Results of watershed separation of the figure 3



(a) 회색조 영상  
(a) Gray image



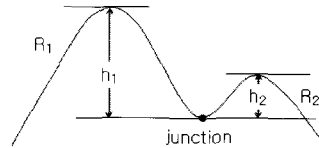
(b) 누적 기울기 역상영상의 3차원 표현  
(b) 3-dimensional expression of accumulated gradient inverse image

그림 5. 그림 3의 누적 기울기 영상  
Fig.5. accumulated gradient image of the figure 3

### 1.2 누적 기울기 영상 분할

스팟의 형태는 원형 또는 타원형이지만 여러 가지 원인으로 형태가 왜곡되며 여러 스팟이 엉겨있는 복잡영역에서는 서로간의 간섭으로 인하여 왜곡이 더욱 심해진다. 그 결과 [그림 5](b)에서 누적 기울기 영상에는 많은 피크가 생겨 과분할

(over segment)의 원인이 된다. 이에 대한 대책으로서 첫째, 워터셰드 분할 과정에서 [그림 6]과 같이 두 영역  $R_1$ 과  $R_2$ 의 접합부에서 다음 식과 같이 골짜기와 정상과의 고도차가 임계값(threshold value)  $T_h$ 보다 적은 경우에는 영역  $R_1$ 과  $R_2$ 들을 서로 병합하는 방법을 취하도록 한다.



$$h_1 \leq T_h \vee h_2 \leq T_h$$

그림 6. 영역의 병합  
Fig.6. Mergence condition of regions

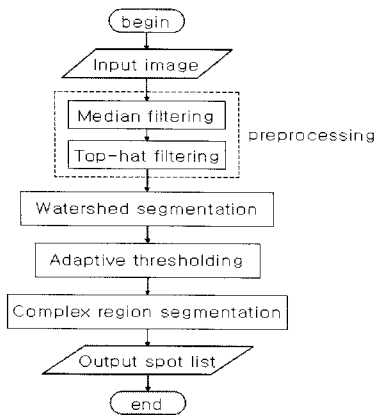
둘째, 누적 기울기 역상 영상에서 고도가 낮은 부분은 영역분할에서 제외한다. 이를 위하여 다음 식과 같이 고도가  $T_g$  이하인 영역은 고도를  $T_g$ 로 고정한다.

$$\forall p \in R,$$

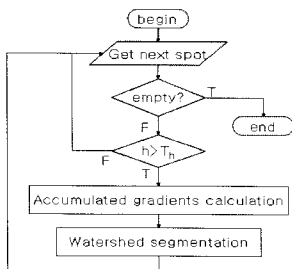
$$G_p \leq T_g \Rightarrow G_p = T_g$$

### 1.3. 시스템 흐름도

본 연구에서 제안하고 있는 스팟 검출 시스템을 살펴보면 [그림 7], 먼저 입력 영상에 중간값 필터(median filter)와 탑-햇 필터(top-hat filter)를 이용하여 전처리 단계를 거치고 난 뒤, 워터셰드 분할 알고리즘을 이용하여 영역 분할이 이루어져 스팟을 분리해내고, 그 다음 영역에 임계값을 적용하여 스팟 리스트를 검출하게 된다. 이 때 스팟들이 겹치고 포화된 복잡 영역이 나타난다면 [그림 7] (b)에서와 같이 스팟 리스트에서 임계값을 지정하여 누적 기울기를 구하고, 워터셰드 알고리즘을 이용하여 또다시 영역을 분할한 뒤 최종적으로 스팟 리스트를 결정한다.



(a) 전체 시스템 흐름  
(a) Overall system flowchart



(b) 복잡영역 분할 흐름  
(b) Complex region segmentation flowchart  
그림 7. 시스템 흐름도  
Fig. 7. System flowchart

2. 실험 결과

[그림 8]은 [그림 3] 스팟이 최종적으로 분할된 복잡영역의 스팟 영상이다. +마크는 각 스팟의 의사(pseudo) 피크를 나타낸다. 누적 기울기 영상은 왜곡이 심하므로 피크 위치를 추정하는 정보로서는 적당치 않으므로, 누적 기울기로써 분할된 영역의 원영상의 고원 부분의 중점을 의사 피크 위치로 결정하였다.

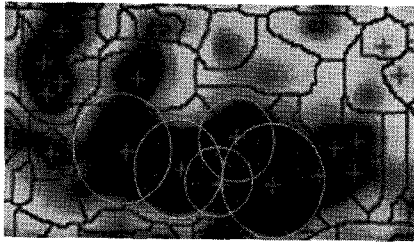


그림 8. 분할된 복잡영역  
Fig.8. Segmented complex region

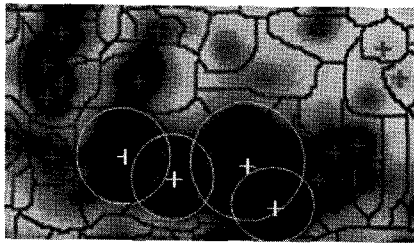
Swiss-prot 데이터베이스에서 제공하는 등전점(iso electric focusing)이 Escherichia coli의 단백질에서 각각 pH4.5-6.5, pH6-9, pH6-11인 영상과 LYMPHOCYTE\_HUMAN, NUCLEI\_LIVER\_HUMAN 영상에서 좌표값의 범위가 각각 (496,576)~(659,665), (127,604)~(225,642), (130,562)~(199,614), (210,539)~(290,611), (370,767)~(456,793)이며 임계값( $T_h$ )이 각각 9, 10, 23, 2, 2인 복잡영역을 선정하여 일반 워터셰드 방법과 누적 기울기를 이용한 방법으로 각각 실험을 실시하였으며, 그 결과를 [그림 9]에서 나타내었다. 원영상(①②③④⑤)에서 각각 일반 워터셰드 처리한 결과에서는 복잡영역에서 스팟을 구분할 수 없었으나 제안한 누적 기울기를 이용한 결과에서는 각각의 스팟을 인식하여 분리할 수 있었다.

원영상	워터셰드 결과	누적기울기 처리결과
① <i>E. coli</i> pH6-9 (127,604)~(225,642), $T_h=9$		
② <i>E. coli</i> pH6-11 (130,562)~(199,614), $T_h=10$		
③ <i>E. coli</i> pH4.5-6.5 (496,576)~(659,665), $T_h=23$		
④ LYMPHOCYTE_HUMAN (210,539)~(290,611), $T_h=2$		
⑤ NUCLEI_LIVER_HUMAN (370,767)~(456,793), $T_h=2$		

그림 9. 원 영상과 일반적인 워터셰드 처리 결과와 누적 기울기 처리 결과 영상 비교  
Fig.9. Image comparison between original image, result from general watershed process, and accumulated gradient process



(a) 결과 영상  
(a) Result image



(b) 사람의 시각에 의한 예측  
(b) The prediction through optical observation

그림 10. 누적 기울기 처리 결과와 사람의 시각에 의한 예측 결과 영상

Fig. 10. The result from accumulated gradient process and optical observation

여기서 [그림 9]의 ③ *E.coli* pH4.5-6.5 누적기울기 처리결과 영상과 사람의 시각에 의한 예측을 비교할 경우, [그림 10] (a) 결과영상에서 보이는 원내의 중심점에서 스팟의 분할영역이 고리연산자에 알맞지 않음은 물론이고 스팟의 일반적인 형태가 아님을 알 수 있다(스팟의 개수 = 5). 직관적인 예측으로도 그림 (b)와 같이 분리되는 것이 알맞은 것(스팟의 개수 = 4)이다.

#### IV. 결 론

2차원 전기영동 결과영상을 처리하기 위해 일반적으로 제안되었던 기법들이 가지는 문제점 중의 하나인 단백질을 나타내는 여러 개의 스팟들이 겹치고 포화되어 있는 복잡영역에서 단백질 스팟을 하나로 인정하거나 검출이 불가능하여 오류를 범하게 되는 문제점에 대한 개선 방법으로 복잡영역내의 각 점의 누적 기울기를 산출하고, 누적 기울기 영상을 위터체드 기법으로 영역 분할하여 스팟을 분리하는 영상처리기법을 적용하였다. 이때 영상분할에서 생길 수 있는 과분할의 해결책으로는 임계값을 적용하였고, 그 결과 스팟 분리에 있어서

기존의 기법들을 적용하였을 때 보다 더 우수하고 효율적인 결과를 나타냄을 확인하였다. 또한 2차원 전기영동 결과영상에서 숨겨져 있었던 단백질 스팟을 더 많이 검출 할 수 있고, 예측의 범위 또한 확장시켰다. 그러나 아직 본 논문에서 제안된 누적기울기를 이용하여 단백질 스팟을 검출하는 개선된 방법이 모든 단백질 스팟을 완벽하게 검출하는 것은 아니어서 향후에는 분할영역의 오류 극복에 더욱 개선된 알고리즘을 개발하여야함은 물론이고 표준영상과 실험 영상간의 최적 매칭 알고리즘을 개발하여 생물학자들에게 제공해야 할 것이다.

#### 참고문헌

- [1] Patrick, H. and O'Farrell, "High resolution two-dimensional electrophoresis of proteins", *Journal of Biological Chemistry*, 250, pp.4007-4021, 1975.
- [2] O. Gaal, G. A. Medgyesi, and L. Vereczkey, "Electrophoresis in the Separation of Biological Macromolecules", John Wiley & Sons, 1980.
- [3] Soren Naaby-Hansen, Michael D. Waterfield and Rainer Cramer, "Proteomics-post-genomic cartography to understand gene function", *Trends in Pharmacological Sciences*, Vol. 22, No. 7, July 2001.
- [4] Thitiwan Srinark and Chandra Karnbarnettu, "An image analysis suite for spot detection and spot matching in two-dimensional electrophoresis gels", *Electrophoresis*, Vol. 29, pp.706-715, 2008.
- [5] Kwan-Deok Choi, Mi-Ae Kim, and Young-Woo Yoon, "Adaptive thresholding for eliminating noises in 2-DE image", *Journal of Signal Processing and Systems*, Vol. 9, 2008.
- [6] Klaus-Peter Pleiünerl et al., "New algorithmic approaches to protein spot detection and pattern matching in two-dimensional electrophoresis gel databases", *Electrophoresis*, Vol. 20, pp. 755-765, 1999.
- [7] Akutsu, T., Kanaya, K., Ohshima, A., and Fujiyama, A., "Matching of spots in 2D electrophoresis images. Point matching under non-uniform distortions", *Lecture Notes in*

Computer Science, 1645: pp.212-222, 1999.

[8] Takahashi, K., Nakazawa, M., and Watanabe, Y., "DNA insight : an image processing system for 2-D gel electrophoresis of genomic DNA". Genome Informatics. Universal Academy Press, pp.135-146, 1997.

[9] M. Rogers, J. Graham and R. P. Tonge, "Statistical models of shape for the analysis of protein spots in two-dimensional electrophoresis gel images", Proteomics, Vol. 3, No. 6, pp.887-896, 2003.

[10] K. Kaczmarek et al., "Comparison of image-transformation methods used in matching 2d gel electrophoresis images", acta chromatographica, No. 13, 2003.

[11] Chris. Glasbey, Leila vali, and Hohn Gustafsson, "A statistical model for unwarping of 1-D electrophoresis gels", Electrophoresis, Vol. 26, pp. 4237-4242, 2005.

[12] Katsutoshi Takahashi et al., "Fully-automated spot recognition and matching algorithms for 2-D gel electrophoretogram of genomic DNA," Genome Informatics, Vol. 9, pp.161-172, 1998.

[13] E. Bettens, "Peak characterization using parameter estimation methods", PhD. thesis, University of Antwerpen, 1999.

[14] Alon Efrat et al., "Geometric Algorithms for the Analysis of 2D- Electrophoresis Gels," Journal of Computational Biology, Vol. 9, No.2, pp. 299-315, 2002.

**저자 소개**



**김미애(Miae Kim)**  
 2004년 2월 영남대 정보통신공학과 (공학사)  
 2006년 2월 영남대 컴퓨터공학과 (공학석사)  
 2009년 8월 영남대 컴퓨터공학과 (공학박사)  
 관심분야 : 생물정보학, 영상처리



**윤영우(Youngwoo Yoon)**  
 正會員  
 1972년 2월 영남대 전자공학과(공학사)  
 1983년 2월 영남대 전자공학과 (공학석사)  
 1988년 2월 영남대 전자공학과 (공학박사)  
 1988년 9월~현재 영남대 컴퓨터공학과 교수  
 관심분야 : 영상처리, 생물정보학, 컴퓨터 설계