

2차원 전기영동 영상의 단백질 반점 정합을 위한 비대칭 확산 모형

최 관 덕[†] · 윤 영 우^{††}

요 약

2차원 전기영동 영상 분석 프로그램의 반점 검출 단계는 영상 분할 알고리즘을 사용해서 겔 영상을 반점 영역으로 분할하고 각 반점 영역을 반점 형태 모형에 정합하여 다음 단계에 필요한 반점 정보를 정량화한다. 현재 영상 분할 알고리즘으로는 분수령 기법이 일반적으로 사용되며, 대표적인 반점 형태 모형으로는 가우스 모형, 확산 모형이 있다. 확산 모형이 가우스 모형보다 실제의 반점 형태에 좀 더 가깝기는 하지만, 반점 형태는 매우 다양하며 특히 x -축과 y -축에 대해서 비대칭적인 형태를 보인다. 반점이 비대칭적 형태인 이유는 2-DE 처리가 통상 이상적인 환경 하에서 이루어질 수 없기 때문에 단백질이 완전히 확산되지 못하기 때문으로 알려져 있다. 따라서 본 논문에서는 비대칭 확산 모형을 제안한다. 비대칭 확산 모형은 초기에는 단백질이 하나의 원으로부터 확산되지만, 시간이 흐름에 따라 x -축과 y -축에 대해서 비대칭적으로 확산된다고 가정하는 모형이다. 실험으로서 19개의 겔 영상에 대해서 세 모형별로 반점 정합을 수행하고 세 모형의 비교를 위해서 SNR의 평균을 구하였다. 실험결과인 SNR의 평균은 가우스 모형이 14.22dB, 확산 모형이 20.72dB, 비대칭 확산 모형이 22.85dB이었다. 실험결과로써 비대칭 확산 모형이 가우스 모형과 확산 모형에 비해서 반점 정합에 보다 더 효율적이며 적합한 모형을 확인하였다.

키워드 : 2차원 전기영동, 반점 검출, 반점 형태 모형, 가우스 모형, 확산 모형, 비대칭확산모형

Asymmetric Diffusion Model for Protein Spot Matching in 2-DE Image

Kwan-Deok Choi[†] · Young-Woo Yoon^{††}

ABSTRACT

The spot detection phase of the 2-DE image analysis program segments a gel image into spot regions by an image segmentation algorithm and fits the spot regions to a spot shape model and quantifies the spot informations for the next phases. Currently the watershed algorithm is generally used as the segmentation algorithm and there are the Gaussian model and the diffusion model for the shape model. The diffusion model is closer to real spot shapes than the Gaussian model however spots have very various shapes and especially an asymmetric formation in x -coordinate and y -coordinate. The reason for asymmetric formation of spots is known that a protein could not be diffused completely because the 2-DE could not be processed under the ideal environment usually. Accordingly we propose an asymmetric diffusion model in this paper. The asymmetric diffusion model assumes that a protein spot is diffused from a disc at initial time of diffusing process, but is diffused asymmetrically for x -axis and y -axis respectively as time goes on. In experiments we processed spot matching for 19 gel images by using three models respectively and evaluated averages of SNR for comparing three models. As averages of SNR we got 14.22dB for the Gaussian model, 20.72dB for the diffusion model and 22.85dB for the asymmetric diffusion model. By experimental results we could confirm the asymmetric diffusion model is more efficient and more adequate for spot matching than the Gaussian model and the diffusion model.

Keywords : 2-DE; 2-Dimensional Electrophoresis, Spot Detection, Spot Shape Model, Gaussian Model, Diffusion Model, Asymmetric Diffusion Model

1. 서 론

2차원 전기영동(2-Dimensional Electrophoresis : 2-DE)

은 분자 생물학 분야의 중요한 도구로서 생물학적 표본 내의 단백질을 종류별로 분리하는 기법이다[1]. 2-DE는 1975년 O'Farrell[1]에 의해 개발된 이후 현재까지 단백질 분리 기법으로서 독보적인 위치를 고수하고 있으며, 독물학, 임상화학, 암 연구, 알츠하이머 연구 등 많은 분야에서 사용된다. 특히 포스트게놈 시대의 대표적 연구 분야인 단백체학(proteomics)을 뒷받침하는 기술로서 중요한 역할을 한다[2,3].

2-DE는 두 단계의 생화학적 분리를 차례로 수행한다. 첫

* 본 논문은 교신저자의 2006학년도 영남대학교 연구년제 수행에 의한 것입니다.

† 정 회 원 : 대구과학대학 의무기록정보과 조교수

†† 종신회원 : 영남대학교 전자정보공학부 교수(교신저자)

논문접수 : 2008년 6월 3일

수 정 일 : 1차 2008년 7월 21일

심사완료 : 2008년 8월 21일

번째 단계는 등전점(Iso-Electric Focusing : IEF) 전기영동이라 하는데, 표본을 고정화 페하 구배(Immobilized pH Gradient : IPG) 겔(gel) 막대 위에 놓고 전기장을 가하면 표본내의 단백질들은 고유의 등전점(iso-electric point : pI)에 따라 이동하여 분리된다. 두 번째 단계는 SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis)라 하는데, 첫 번째 단계의 결과에 전기장을 가하면 단백질들은 각각의 분자량(molecular weight : Mw)에 따라 이동하여 분리된다. 첫 번째 단계를 1차원 전기영동이라 하고, 두 번째 단계를 2차원 전기영동이라 하며, 두 번째 단계 처리 이후에 겔은 염색 과정을 거친다. 2-DE 처리 결과는 (그림 1)과 같이 2차원 반점(spot) 패턴으로 나타나며, 패턴 내의 각각의 반점은 개별적인 단백질을 의미한다. 따라서 2-DE 결과를 응용하기 위해서는 반점을 단백질로 동정(identify; 同定)하는 과정이 필요하다[1-3].

반점을 동정하는 방법으로는 반점을 로봇으로 추출한 후 질량분광기를 이용하는 방법과 이미 분석된 2-DE 결과와 새로운 결과를 서로 시각적으로 교차비교하여 동정하는 방법이 있다. 전자의 경우, 정밀분석이 가능하나 한 번에 하나

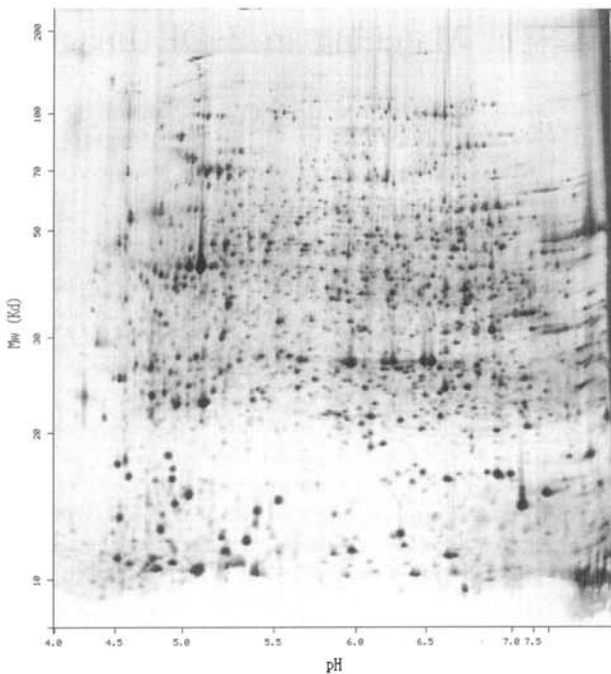
의 반점밖에 동정 못하므로 작업시간이 많이 걸린다. 후자의 경우는 단백질 반점 패턴의 복잡성 때문에 육안 분석이 매우 어려워서 컴퓨터로써 자동적으로 동정하기 위한 연구가 일찍부터 수행되어 왔다. 컴퓨터로써 자동적으로 동정하기 위해서는 겔을 고해상도의 스캐너나 레이저 광도계 등을 사용해서 디지털 영상화하고, 디지털화된 겔 영상을 겔 영상 분석 프로그램으로 분석해야한다[1-3].

겔 영상 분석 프로그램은 일반적으로 (그림 2)와 같이 반점 검출 단계, 겔 정합 단계, 정량적 비교 분석 단계로 구성된다. 반점 검출 단계는 소스 겔 영상을 반점 영역별로 분할한 다음, 형태 모형에 반점을 정합하여, 이 후의 단계에 필요한 반점 정보(즉, 반점의 위치, 고도, 체적, 형태 모형 매개변수 등)를 정량화한다[4-9]. 겔 정합 단계는 전 단계에서 검출된 반점들을 이미 분석된 겔 영상 즉 마스터 겔 영상의 반점들과 교차비교하여 실제의 단백질들로 동정하는 단계이다[6-11]. 정량적 비교 분석 단계는 소스 겔의 동정 반점들을 마스터 겔의 동정 반점들과 비교하여 반점의 유무(단백질 발현 유무), 반점 체적의 차(단백질 발현량) 등의 결과를 내는 단계이다[2,3].

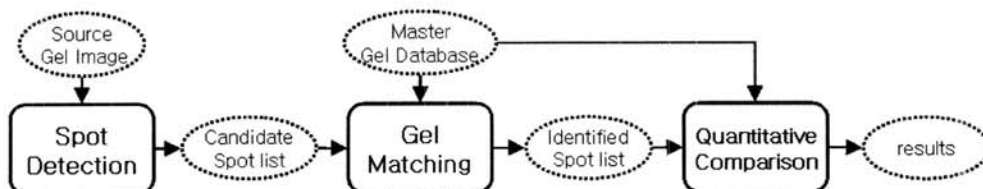
반점 검출 단계의 구현은 적합한 영상분할 기법과 반점 형태 모형을 정하는 것이 핵심이며, 겔 영상 내의 잡영(noise)의 처리, 잡영으로 보이기 쉬운 미약한 반점의 검출, 중첩된 반점(overlapping spot)의 분리를 위한 기법이 필요하다[2,3,12].

현재 영상분할 기법으로 잡영에 대한 견고성 때문에 분수령(watershed) 기법[13]이 일반적으로 사용되지만, 겔 영상의 배경이 불균질한 이유로 영역들이 과분할되는 문제점이 있다. 대표적인 반점 형태 모형으로는 가우스 모형(Gaussian model)[14], 확산 모형(diffusion model)[15]이 있다. 2-DE에서 단백질 반점은 확산 과정으로 형성된다. 가우스 모형은 단백질이 한 점(point)에서 확산된다고 가정한 모형이며, 확산 모형은 단백질이 한 원(disc)에서 확산된다고 가정한 모형이다.

확산 모형이 가우스 모형에 비해서 좀 더 반점의 형태에 가깝기는 하지만, 실제 반점의 형태는 매우 다양하며 특히 정점(peak)을 기준으로 x -축과 y -축 각각에 대해서 비대칭적으로 확산된 형태를 보인다. 반점 형태가 대칭적인 확산 형태가 되지 않는 이유는 2-DE 처리가 통상 이상적인 환경 하에서 이루어질 수 없기 때문에 단백질이 완전히 확산되지 못하기 때문으로 알려져 있다[16]. 본 논문에서는 이러한 비대칭적인 확산 형태에 잘 들어맞는 비대칭 확산 모형(asymmetric diffusion model)을 제안한다.



(그림 1) 겔 영상 예 : 겔 영상 DICTYSLUG (설명을 위해 좌표축에 이름이 붙여짐. 좌표축을 제외한 겔 영상의 크기: 1004×1109 화소)



(그림 2) 겔 영상 분석 프로그램

본 논문의 구성은 다음과 같다. 1장의 서론에 이어, 2장에서 반점 형태 모형에 관한 기존 연구를 기술한다. 3장에서는 제안 모형인 비대칭 확산 모형에 대해서 기술하며, 4장에서 비대칭 확산 모형을 실험으로써 검증한다. 5장에서 결론을 맺는다.

2. 기존 연구

반점 형태를 정의하는 기법에는 비매개변수(non-parametric) 기법과 매개변수(parametric) 기법이 있다. 비매개변수 기법은 반점 형태를 규정하지 않고 2차원적인 윤곽으로 반점을 정의하는 기법이며, 매개변수 기법은 반점 형태를 규정하는 반점 모형(spot model)을 정의하고, 반점을 모형에 정합(matching; fitting)시켜 구한 모형의 매개변수로서 반점을 정의하는 기법이다. 비매개변수 기법은 모형 정합에 필요한 시간적 경비가 필요하지 않은 이점이 있다. 반면에 매개변수 기법은 정합이 완료되면 반점을 매개변수만으로 표현할 수 있으므로 공간적 경비가 적게 들며, 이 후의 단계에 필요한 알고리즘 고안에 보다 효과적이다[15,16]. 본 논문에서는 매개변수 기법에 대해서만 다룬다.

매개변수 기법에 사용되는 대표적인 반점 모형으로는 가우스 모형[14]과 확산 모형[15]이 있다.

2.1 가우스 모형

2차원 전기영동에서 단백질 반점은 확산 과정으로 형성된다. 가우스 모형은 단백질이 한 점에서 확산된다고 가정된 모형으로서 수식(1)[14]과 같이 정의된다.

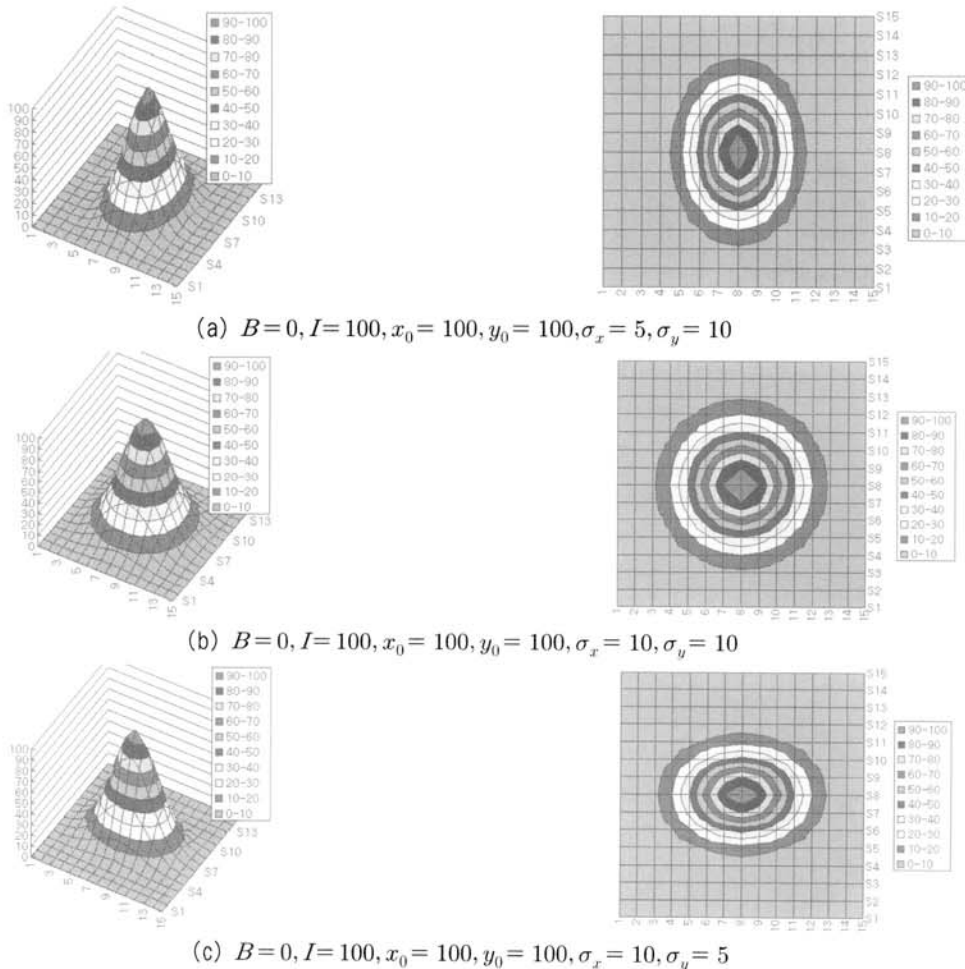
$$S(x,y) = B + I \exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2}\right) \exp\left(-\frac{(y-y_0)^2}{2\sigma_y^2}\right) \quad (1)$$

수식(1)에서 B 는 배경의 강도, I 는 정점의 강도(즉, 고도), x_0 와 y_0 는 정점 좌표, σ_x 와 σ_y 는 x -축 방향, y -축 방향에 대한 확산계수를 의미한다.

가우스 모형을 매개변수벡터로 표현하면 수식(2)와 같다.

$$G_s = (B, I, x_0, y_0, \sigma_x, \sigma_y) \quad (2)$$

(그림 3)에 가우스 모형의 예를 그림으로 보인다. 이 그림에서는 가우스 모형의 형태를 비교하여 보이기 위해서, (a),



(그림 3) 가우스 모형

(b), (c) 예 전체에 대한 매개변수 B 는 0, 매개변수 x_0 는 100, 매개변수 y_0 는 100으로 고정시키고, 매개변수 σ_x 와 매개변수 σ_y 의 값은 다르게 정하여 가우스 모형으로 정의한 반점 형태의 예를 보인다.

2.2 확산 모형

가우스 모형은 실제의 단백질 반점의 형태와 꼭 들어맞지 않으며, 특히 상부가 포화되어 평평한 형태(flat-top)에 잘 들어맞지 않는다. Bettens[15]는 이것을 해결하기 위해서 단백질이 중심 부근에서 포화되는 현상을 설명하는 반점 형성 과정의 물리적 현상을 적용한 모형인 확산 모형을 제안하였다.

확산 모형은 단백질이 초기에 농도가 C_0 이고 반지름이 a 인 원으로부터 확산된다고 가정한 모형으로 수식(3)[15]과 같이 정의된다.

$$S(x,y) = B + \frac{C_0}{2} \left[\operatorname{erf}\left(\frac{a'+r'}{2}\right) + \operatorname{erf}\left(\frac{a'-r'}{2}\right) \right] + \frac{C_0}{r'\sqrt{\pi}} \left[\exp\left(-\left(\frac{a'+r'}{2}\right)^2\right) - \exp\left(-\left(\frac{a'-r'}{2}\right)^2\right) \right] \quad (3)$$

with

$$r' = \sqrt{\frac{(x-x_0)^2}{D_x} + \frac{(y-y_0)^2}{D_y}}$$

수식(3)에서 $S(x,y)$ 는 반점 S 내의 화소 (x,y) 의 농도(즉, 강도), B 는 배경의 강도, C_0 는 정점 (x_0,y_0) 의 초기농도(즉, 고도), erf 는 가우스 오차함수(즉, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)dt$),

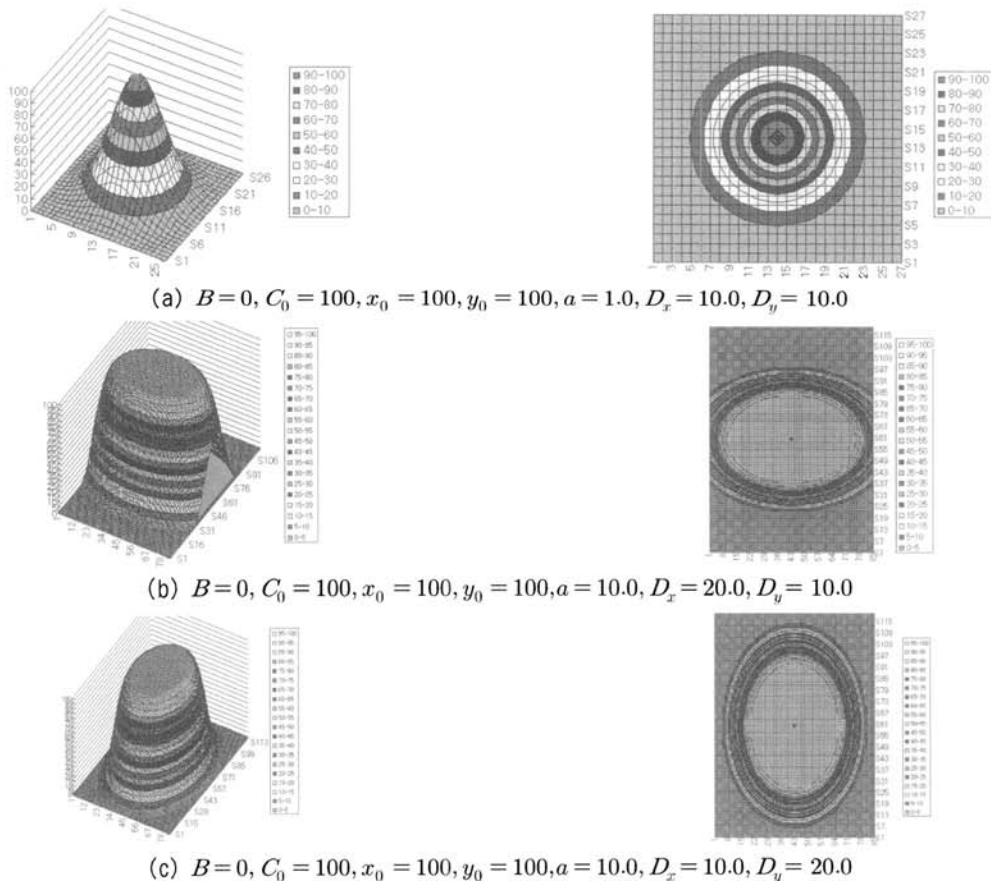
$a' = \sqrt{\frac{D}{t}} a$ (즉, a' 은 시간 t 에서의 확산 진행 비율로 계산한 초기 원의 부분 원의 반지름), $D'_x = D_x t$, $D'_y = D_y t$, a 는 초기 원의 반지름, D 는 초기 원의 확산계수, D_x 는 x -축 방향 확산계수, D_y 는 y -축 방향 확산계수, r' 은 시간 t 에서의 화소 (x,y) 를 포함하는 원의 반지름을 의미한다.

확산 모형은 반지름 a 가 0에 가까울 때 가우스 모형과 동일하다. 즉 확산 모형은 가우스 모형을 포함한다.

확산 모형을 매개변수벡터로 표현하면 수식(4)와 같다. 즉, 확산 모형은 배경의 강도가 B 이며, 중심점이 (x_0,y_0) 이고 반지름이 a 인 원이 농도가 C_0 부터 0이 될 때까지 확산계수 D_x , D_y 에 따라 확산되어 반점을 형성하는 것으로 요약하여 표현할 수 있다.

$$D_s = (B, C_0, x_0, y_0, a, D_x, D_y) \quad (4)$$

(그림 4)에 확산 모형의 예를 그림으로 보인다. (그림 4)의 각각의 예는 반지름이 a 인 원으로부터 확산 계수 D_x 와



(그림 4) 확산 모형

D_y 에 따라 확산이 완료된 반점 형태를 보이며, 그림에 표기된 눈금의 단위는 화소이다. 이 그림에서는 확산 모형의 형태를 비교하여 보이기 위해서, (a), (b), (c)에 전체에 대한 매개변수 B 는 0, 매개변수 x_0 는 100, 매개변수 y_0 는 100, 매개변수 C_0 는 100으로 고정시키고, 매개변수 a 와 매개변수 D_x 와 매개변수 D_y 의 값들은 다르게 정하여 확산 모형의 예를 보인다.

3. 비대칭 확산 모형

확산 모형이 가우스 모형보다 반점의 형태에 보다 가깝지만, 실제 겔 영상 내의 반점의 형태는 매우 다양하며 불규칙하다. 반점 형태가 불규칙한 이유는 2-DE 기법의 특성상 처리가 통상 이상적인 환경 하에서 이루어지는 것이 아니어서, 단백질이 완전하게 확산되지 않기 때문으로 알려져 있다.

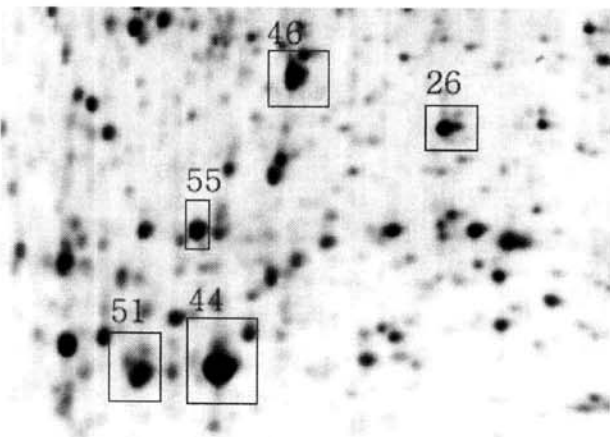
(그림 5)는 (그림 1)의 겔 영상 DICTYSLUG의 일부분을 확대한 것으로서 반점의 형태들을 살펴보면 55번 영역과 같이 확산 모형의 형태를 가진 것도 있지만 대부분의 반점들이 그렇지 않음을 볼 수 있다.

(그림 6)은 (그림 5)의 사각형으로 표시된 영역들을 2차원과 3차원 그래프로 보인다.

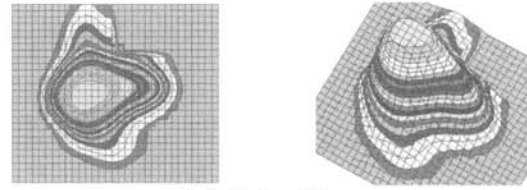
(그림 6)에서 반점의 형태는 정점을 기준으로 x -축과 y -축에 대칭을 이루는 형태라기보다는 정점을 기준으로 x -축과 y -축 각각에 대해서 비대칭인 형태에 가깝다. 비대칭 확산 모형은 이러한 비대칭 형태에 착안한 모형이다.

비대칭 확산 모형은 단백질이 초기에 농도가 C_0 이고 반지름이 a 인 원으로부터 확산되지만, 시간이 흐름에 따라 x -축과 y -축에 대해서 비대칭적으로 확산된다고 가정한 모형이다.

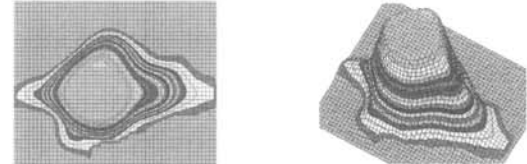
비대칭 확산 모형은 수식(3)의 확산 모형 정의식의 r' 을 네 개의 확산 방향 즉, 음의 x -축 방향, 양의 x -축 방향, 양의 y -축 방향, 음의 y -축 방향에 대한 경우로 수정하여 수식(5)와 같이 정의한다.



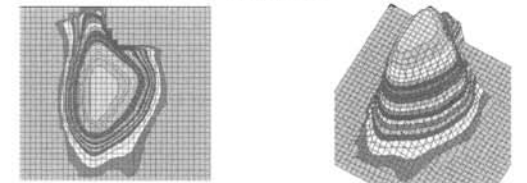
(그림 5) 다양한 반점 형태



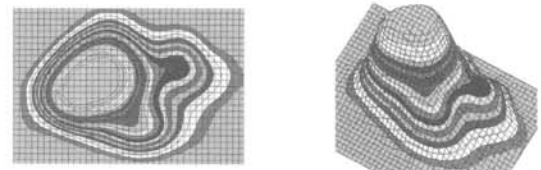
(a) Region 26



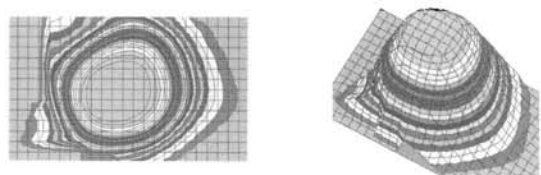
(b) Region 44



(c) Region 46



(d) Region 51



(e) Region 55

(그림 6) (그림 5)의 사각형 영역

$$S(x,y) = B + \frac{C_0}{2} \left[\operatorname{erf}\left(\frac{a'+r'}{2}\right) + \operatorname{erf}\left(\frac{a'-r'}{2}\right) \right] + \frac{C_0}{r' \sqrt{\pi}} \left[\exp\left(-\left(\frac{a'+r'}{2}\right)^2\right) - \exp\left(-\left(\frac{a'-r'}{2}\right)^2\right) \right] \quad (5)$$

with

$$r' = \sqrt{\frac{(x-x_0)^2}{D_{xp}} + \frac{(y-y_0)^2}{D_{yp}}}, \quad x > 0 \text{ and } y > 0$$

$$r' = \sqrt{\frac{(x-x_0)^2}{D_{xm}} + \frac{(y-y_0)^2}{D_{yp}}}, \quad x < 0 \text{ and } y > 0$$

$$r' = \sqrt{\frac{(x-x_0)^2}{D_{xm}} + \frac{(y-y_0)^2}{D_{ym}}}, \quad x < 0 \text{ and } y < 0$$

$$r' = \sqrt{\frac{(x-x_0)^2}{D_{xp}} + \frac{(y-y_0)^2}{D_{ym}}}, \quad x > 0 \text{ and } y < 0$$

수식(5)에서 $S(x,y)$ 는 반점 S 내의 화소 (x,y) 의 농도(즉, 강도), B 는 배경의 강도, C_0 는 정점 (x_0,y_0) 의 초기농도(즉, 고도), erf 는 가우스 오차함수(즉, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$),

$a' = \sqrt{\frac{D}{t}} a$, $D'_{xp} = D_{xp}t$, $D'_{xm} = D_{xm}t$, $D'_{yp} = D_{yp}t$, $D'_{ym} = D_{ym}t$, a 는 초기 원의 반지름, D 는 초기 원의 확산 계수, D_{xp} 는 양의 x -축 방향 확산계수, D_{xm} 은 음의 x -축 방향 확산계수, D_{yp} 는 양의 y -축 방향 확산계수, D_{ym} 은 음의 y -축 방향 확산계수를 의미한다.

비대칭 확산 모형은 D_{xp} 와 D_{xm} 이 같은 값이고, D_{yp} 와 D_{ym} 이 같은 값이면 확산 모형과 동일하다. 즉 비대칭 확산 모형은 확산 모형을 포함한다.

비대칭 확산 모형을 매개변수벡터로 표현하면 수식(6)과 같다. 즉, 비대칭 확산 모형은 배경의 강도가 B 이며, 중심점이 (x_0, y_0) 이고 반지름이 a 인 원이 농도가 C_0 에서 0이 될 때까지 확산계수 D_{yp} , D_{ym} , D_{xp} , D_{xm} 에 따라 확산되어 반점을 형성하는 것으로 요약하여 표현할 수 있다.

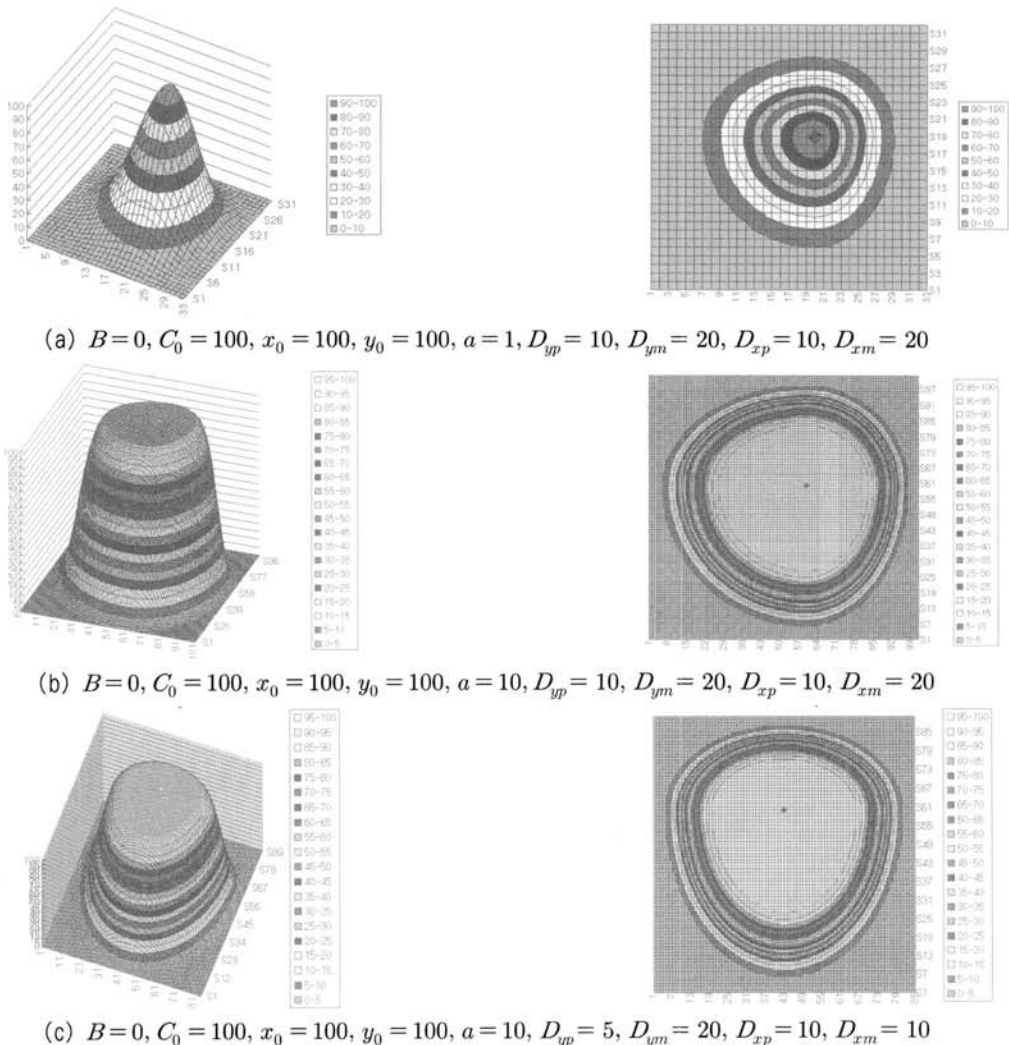
$$A_s = (B, C_0, x_0, y_0, a, D_{yp}, D_{ym}, D_{xp}, D_{xm}) \quad (6)$$

(그림 7)에 비대칭 확산 모형의 예를 그림으로 보인다.

(그림 7)의 각각의 예는 반지름이 a 인 원으로부터 확산 계수 D_{yp} , D_{ym} , D_{xp} , D_{xm} 에 따라 확산이 완료된 반점 형태를 보이며, 그림에 표기된 눈금의 단위는 화소이다. 이 그림에서는 비대칭 확산 모형의 형태를 비교하여 보이기 위해서, (a), (b), (c) 예 전체에 대해서 매개변수 B 는 0, 매개변수 x_0 는 100, 매개변수 y_0 는 100, 매개변수 C_0 는 100으로 고정시키고, 매개변수 a , 매개변수 D_{xp} , 매개변수 D_{xm} , 매개변수 D_{yp} , 매개변수 D_{ym} 의 값을 다르게 정하여 비대칭 확산 모형의 예를 보인다. 이 그림에서 볼 수 있듯이 비대칭 확산 모형은 확산 모형으로써 모형화하기 힘든, x -축과 y -축에 대해서 비대칭적인 확산 형태를 모형화하기에 적합하다.

4. 실험 및 분석

이 장에서는 비대칭 확산 모형을 실험으로써 검증한다. 먼저 실험을 위한 반점 검출 시스템을 기술하고, 세 개의 반점 모형인 가우스 모형, 확산 모형, 비대칭 확산 모형의

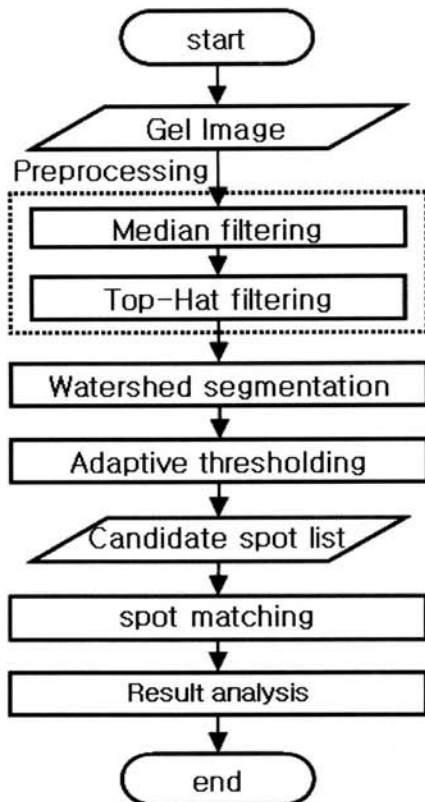


(그림 7) 비대칭 확산 모형

정합 실험과 실험 결과에 대해서 기술한다. 반점 검출 시스템을 위한 실험용 목적 시스템은 Microsoft Windows XP 기반의 PC이며, 전체 알고리즘은 Microsoft Visual C++ 6.0 을 사용해서 구현하였다.

4.1 반점 검출 시스템

본 논문에서 설계하고 구현한 반점 검출 시스템은 (그림 8)과 같다. 겔 영상을 입력받아 전처리를 수행한 다음, 분수령 영상분할을 수행하여 반점 영역을 분할한다. 다음 단계로 분할 영역에 적응적 문턱값 결정 기법을 적용하여 후보 반점을 결정하고, 후보 반점 목록(candidate spot list)을 생성한다. 마지막 단계에서 후보 반점들에 대해서 세 모형별로 반점 정합을 수행하여 결과를 분석한다.



(그림 8) 반점 검출 시스템

4.1.1 겔 영상

겔 영상의 입력 형식은 BMP(Bitmap) 형식이며, 입력된 겔 영상은 256 레벨의 그레이스케일 영상으로 변환하여 처리한다.

4.1.2 전처리

전처리 단계는 잡음을 제거하는 단계이다. 잡음은 백색잡영(white noise)과 비백색잡영(non-white noise)으로 구분할 수 있다. 백색잡영은 화소단위로 독립적으로 발생하는 잡영이며, 잡영의 강도가 무작위적이다. 즉 모든 강도의 발생 확률이 동일하다. 비백색잡영은 강도가 인접 화소들과 연관된

잡영 즉, 화소의 군집으로 구성된 잡영이다.

본 논문에서는 백색잡영을 제거하기 위해서 9×9 중간값 필터를 사용하고, 비백색잡영을 제거하기 위해서 반지름 10인 반구형 구조화 요소(structuring element)를 사용한 top-hat 필터[17]를 사용한다.

실험의 한 예로 (그림 9)에 전처리 단계 처리 후의 겔 영상 DICTYSLUG의 모습을 보인다.



(그림 9) 전처리 결과 : DICTYSLUG (영상크기: 1004×1109)

4.1.3 분수령 영상분할

분수령 영상분할 단계는 겔 영상을 반점 영역들로 분할하는 단계이다. 분수령 알고리즘은 지형학 분야에서 연구된 것으로 영상을 화소 값을 높이로 하는 3차원 지형으로 간주하고 이 지형에 물을 채웠을 때 하나의 윤곽선으로 둘러싸여진 호수를 분할 영역으로 판단하여 영상을 분할하는 알고리즘이다.

본 논문에서는 그레이스케일 영상의 각 화소의 강도를 지형적인 고도(topographic relief)로 간주하여 분수령 알고리즘을 적용하며, 분수령 알고리즘 중에서 시공간적으로 아주 효율적이라고 알려진 Vincent와 Soille[13]가 제안한 침수(immersion) 방식 분수령 알고리즘을 사용한다. 침수 방식은 범람(flooding) 방식이라고도 하는데 지형의 국소 극소점들에 구멍이 있다고 가정하고 물을 가장 낮은 곳부터 점차적으로 채워나가서 호수를 만든다. 호수에 물이 계속 차오르면 어느 시점에서 호수들은 꼭 차서 넘쳐날 지경이 되는데 이때 호수들 사이의 물이 만나는 지점에 댐을 쌓아서 물이 합쳐지지 못하도록 한다. 결과적으로 댐들을 연결한 것이 분수령이며 영역 분할의 경계선이 된다.

실험의 한 예로 (그림 10)에 분수령 영상분할 단계 처리

후의 겔 영상 DICTYSLUG의 모습을 보인다. 분수령 영상 분할 결과인 영역의 수는 6821개이다.



(그림 10) 분수령 수행결과: DICTYSLUG (영상크기: 1004×1109)

4.1.4 적응적 문턱값 결정 적용

적응적 문턱값 결정 적용 단계는 분수령 영상 분할 단계의 결과인 반점 영역들 중에 배경일 가능성이 높은 영역들을 제거하고, 제거되지 않은 나머지 영역들로 후보 반점을 결정하는 단계이다.

분수령 알고리즘의 수행에서 나온 각각의 분할 영역들을 각각의 반점 영역들로 간주할 수 있다. 하지만 top-hat 필터로 배경을 보상하여도 구조화 요소보다 작은 부분은 보상하지 못하므로 분수령 알고리즘에 의해서 영역으로 분할되어서, 결과적으로 과분할 문제를 초래한다. 이러한 과분할 문제는 전체 반점 검출 알고리즘의 시공간적 비효율성을 초래하며, 또한 배경 영역에는 형태 정보가 부족하므로 이후의 단계에서 필요한 형태 분석 과정에서 오류 발생 가능성을 높인다. 따라서 영역 중에서 배경이라고 판단되는 영역을 구분하여 제외시키는 기법이 필요하다. 과분할된 배경 영역을 구분하고 제외시키기 위해서 일정한 강도 이하의 영역을 제거하는 고정 문턱값을 사용한다면, 대다수의 겔 영상에 존재하는 미량의 단백질을 나타내는 육안으로는 잘 보이지 않는 반점들을 제거할 수도 있는 가능성이 많으므로 적합하지 않다.

본 논문에서는 영상에 따라 문턱값을 자동적으로 계산하는 기법인 적응적 문턱값 결정 기법[18]을 사용한다.

적응적 문턱값 결정 기법은 다양한 겔 영상에 대한 실험을 통계적으로 분석한 결과, 영역의 고도가 매우 낮은 경우를 제외하고, 영역의 고도에 따른 침도의 평균곡선이 지수

분포의 특성을 가짐에 착안한 기법이다. 적응적 문턱값 결정 기법은 봉우리의 삼차원적 형태를 구별하기 위한 형태적도인 침도(prominence)를 각 영역별로 구하고, 고도에 따른 영역들의 침도의 평균 곡선을 지수함수로 회귀분석하여 지수함수의 매개변수를 구한 다음, 오차의 확률분포를 이용해서 매개변수들로 문턱값을 구하여, 잡영 판단의 기준으로 사용한다.

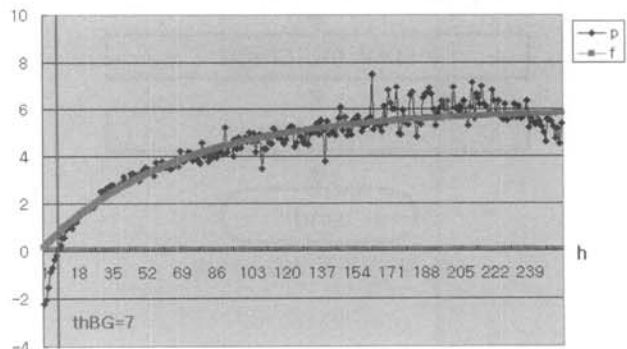
높지만 넓은 불균질한 배경영역으로부터 낮고 작아서 눈에 보이지 않지만 삼차원적 형태가 뚜렷한 봉우리를 구별할 수 있는 형태적도로서 영상 내 영역의 집합을 R 이라 할 때 영역 $r \in R$ 의 침도 p_r 을 수식(7)과 같이 정의한다.

$$p_r = \log(h_r^2/A_r) \tag{7}$$

수식(7)에서 h_r 은 영역의 고도, A_r 은 영역의 정점에서 ah_r (여기서, $0 < a < 1$) 아래의 밑면적을 나타낸다.

영역의 고도가 $thBG$ 보다 높은 영역에 대해서 영역의 고도에 따른 침도의 평균 곡선을 지수함수 $f(h) = a[1 - \exp(-bh)] + c$ 에 회귀분석한다. $thBG$ 는 영역이 배경에 가까운지를 판단하기 위한 문턱값으로서 영역의 고도가 $thBG$ 보다 높은 영역의 개수가 전체 영역 개수의 반이 되는 값으로 결정된다.

실험 예로서 (그림 11)에 겔 영상 DICTYSLUG의 회귀분석 결과를 보인다. 이 그림에서 p 선은 고도에 따른 침도의 평균 곡선이며, f 선은 회귀분석 결과인 지수함수의 곡선이다. 실험에서 $thBG$ 는 7이며, 지수함수로의 회귀분석한 결과로는 매개변수 a 는 5.836135, 매개변수 b 는 0.015, 매개변수 c 는 0.1이다.



(그림 11) 회귀분석 : DICTYSLUG

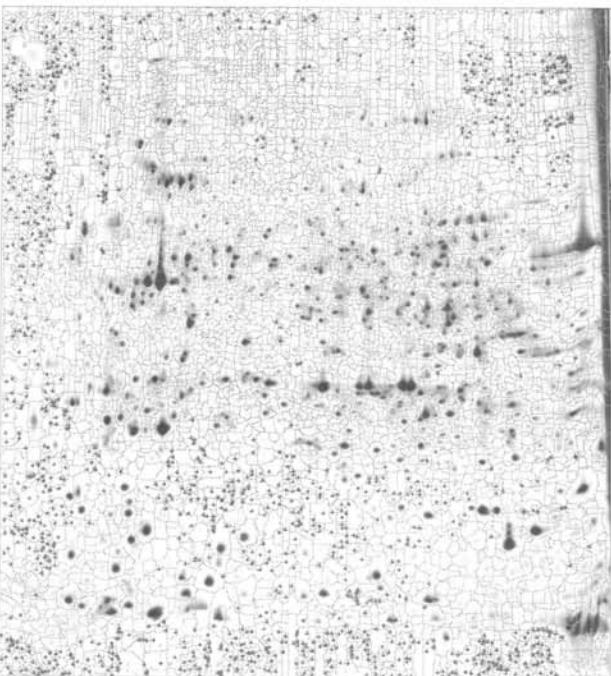
문턱값은 오차의 확률분포에 따라 결정한다. 영역의 침도 (p_r)와 영역의 고도(h_r)을 지수함수에 적용한 값($f(h_r)$)과의 오차를 e_r , 전체 영역에 대한 오차의 평균을 \bar{e} , 전체 영역에 대한 오차의 편차를 σ 라 정의할 때, 오차의 확률분포에 따르면 배경에 포함되지 않은 영역들의 99.7%는 $\bar{e} - 3\sigma < e_r \leq \bar{e} + 3\sigma$ 범위에 속함을 추정할 수 있다. 따라서 영역 중에서 $p_r \leq f(h_r) - \bar{e} - 3\sigma$ 인 영역은 배경에 속할

가능성이 매우 높으므로 제외할 수 있다. 문턱값 적용의 오류 발생 확률은 0.075%이므로 적응적 문턱값 적용의 신뢰도는 99.925%로 추정된다.

실험 예인 겔 영상 DICTYSLUG를 예로 들어 문턱값을 계산해 보이면, 문턱값은 각 영역의 고도에 따라 다른 값으로 결정된다. 즉, 해당 영역의 고도 h_r 에 의한 $f(h_r)$ 즉, $5.836135[1 - \exp(-0.015 \times h_r)] + 0.1$ 에서 \bar{e} 와 3σ 를 뺀 값이 해당 영역에 대한 문턱값이 된다. 따라서 해당 영역의



(그림 12) 후보반점 : DICTYSLUG (영상크기: 1004×1109)



(그림 13) 탈락반점 : DICTYSLUG (영상크기: 1004×1109)

침도 p_r 이 계산된 문턱값보다 작은 경우는 잠영영역으로 판단하며, 작지 않은 경우는 반점영역으로 판단한다.

(그림 12)에 겔 영상 DICTYSLUG에 적응적 문턱값 결정 기법을 적용한 결과 문턱값 이상인 반점들 즉, 후보반점들의 영역들을 나타낸다. 후보반점들의 영역의 정점들이 '+' 기호로 표기되어 있다. 분할 결과인 6821개의 영역 중에 5596개의 영역이 후보반점들의 영역임을 보인다.

(그림 13)에 적응적 문턱값 결정에 의해서 탈락된 반점들, 즉 단백질일 가능성이 적은 반점들을 나타낸다. 탈락된 반점들의 영역의 정점들이 '+'로 표기되어 있다. 분할 결과인 6821개의 영역 중에 1225개의 영역이 후보반점들의 목록에서 탈락되었음을 보인다.

4.1.5 후보 반점 목록

후보 반점 목록은 겔 영상의 반점 정보를 관리하는 자료 구조이다. 후보 반점 목록에 포함되는 반점 정보는 번호, 정점의 위치, 범위, 고도, 넓이, 체적, 형태 모형 매개변수 등으로 구성된다.

4.2 반점 정합 : 비대칭 확산 모형의 검증 실험

4.2.1 실험 방법

비대칭 확산 모형을 검증하기 위해서 겔 영상을 입력받아서 적응적 문턱값 결정 기법 적용까지 수행한 결과인 후보 반점들에 대해서 세 가지 반점 형태 모형별로 반점 정합을 수행하고, 수식(8)에 정의한 SNR을 구하여 세 모형별 반점 정합 결과를 비교한다.

$$SNR = 10 \log_{10} \frac{\sum_{p \in R} I(p)^2}{\sum_{p \in R} \{I(p) - f(X)\}^2} \quad (8)$$

수식(8)에서 $I(p)$ 는 화소 p 의 강도, f 는 모형, X 는 매개변수 벡터, R 은 고도가 영역 R 의 고도의 중간값 이상인 부분영역이다.

반점 정합 알고리즘은 영역 중에서 영역 고도의 중간값 이상인 고도를 가지는 부분 영역의 전체 화소에 대해서 각 화소의 실제 강도와 모형정의식에 (반복제어변수로써) 매개변수벡터를 적용하여 계산한 강도의 차들의 합이 최소가 되는 매개변수벡터를 구하는 반복알고리즘이다. 한 영역의 전체가 아닌 부분 영역을 정합한 이유는 봉우리를 구성하지 않는 산자락 부분을 제외시켜 반점 정합 결과를 보다 의미 있게 만들기 위함이다.

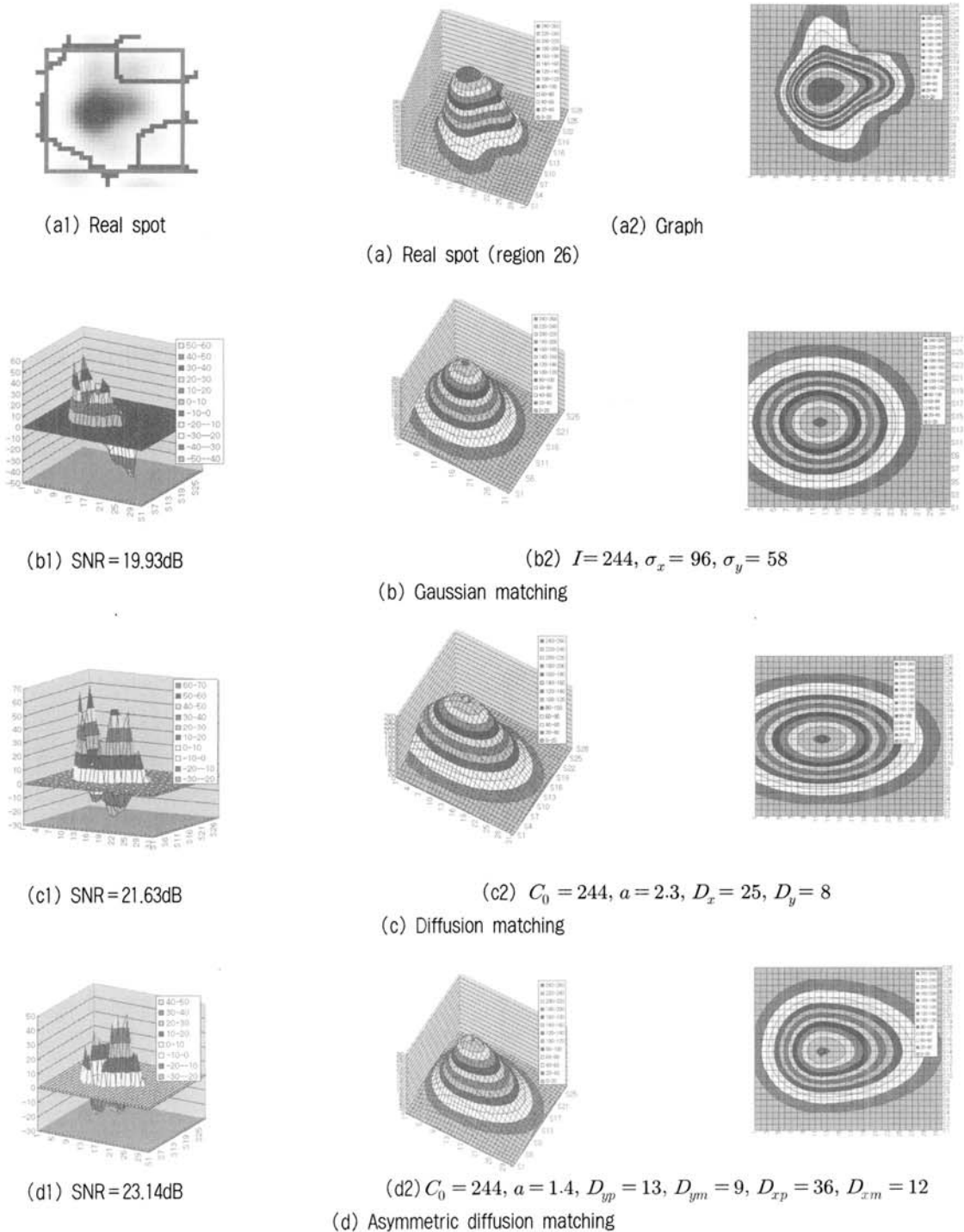
4.2.2 실험 예

반점 정합 예로 (그림 5)의 26번 영역에 대한 반점 정합을 (그림 14)에 보이며, (그림 5)의 55번 영역에 대한 반점 정합을 (그림 15)에 보인다. (그림 14)와 (그림 15)에서 세 모형에 대한 매개변수 B 는 전처리 과정에 의해서 배경을 보정한 시점이므로 0으로 설정하였으며, 매개변수 I 와 C_0 는

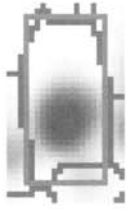
영역의 고도로 설정하였으며, 각 모형의 확산계수를 의미하는 매개변수들은 정합알고리즘 수행에 의한 결과들이다. 정점의 위치에 대한 매개변수 x_0 와 y_0 는 큰 의미가 없으므로 표기하지 않았다.

(그림 14)의 예는 육안으로도 반점의 형태가 비대칭 확산 모형의 형태에 가까우며, 정합 결과인 SNR의 평균을 비교

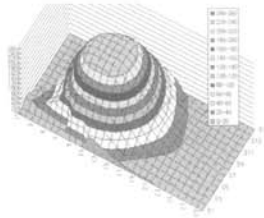
하면 비대칭 확산 모형에 의한 정합이 뚜렷하게 큼을 볼 수 있다. (그림 15)의 예는 육안으로는 반점의 형태가 확산 모형의 형태로 보일 수 있는 예이다. 하지만 실제 반점 정합 실험 결과를 보면 (그림 15)에서 확산 모형에 의한 정합과 비대칭 확산 모형에 의한 정합에 의한 SNR의 평균은 비대칭 확산 모형에 의한 것이 좀 더 큼을 알 수 있다.



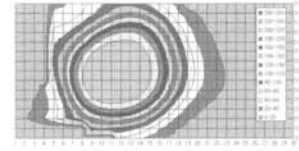
(그림 14) 반점 정합 예 : 영역 26



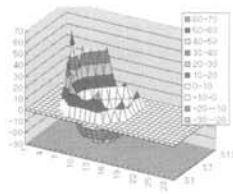
(a1) Real spot



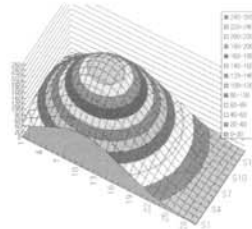
(a2) Graph



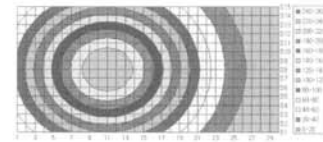
(a) Real spot (region 55)



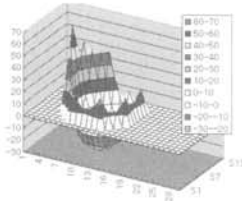
(b1) SNR = 21.97dB



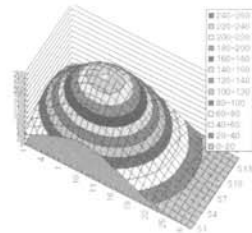
(b2) $I = 240, \sigma_x = 96, \sigma_y = 70$



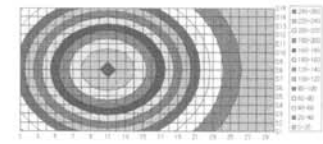
(b) Gaussian matching



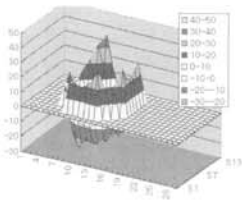
(c1) SNR = 22.12dB



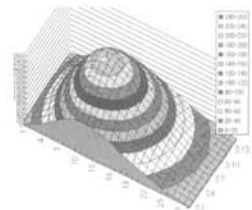
(c2) $C_0 = 240, a = 2.7, D_x = 13, D_y = 8$



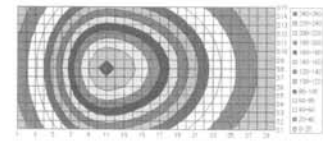
(c) Diffusion matching



(d1) SNR = 23.10dB



(d2) $C_0 = 240, a = 2.3, D_{yy} = 19, D_{yx} = 7, D_{xp} = 10, D_{xm} = 11$



(d) Asymmetric diffusion matching

(그림 15) 반점 정합 예 : 영역 55

4.2.3 결과 및 분석

<표 1>에 반점 정합 실험 결과를 보인다. 이 표에서 반점의 개수(no. of spots)는 후보 반점 영역의 개수를 의미한다. 19개의 실험용 겔 영상은 SWISS-2DPAGE[19]의 은염(silver-stained) 겔 영상들이다.

<표 1>을 보면 19개 영상 각각에 대한 SNR의 평균이 가우스모형, 확산모형, 비대칭확산모형 순이다. 또한 19개 영

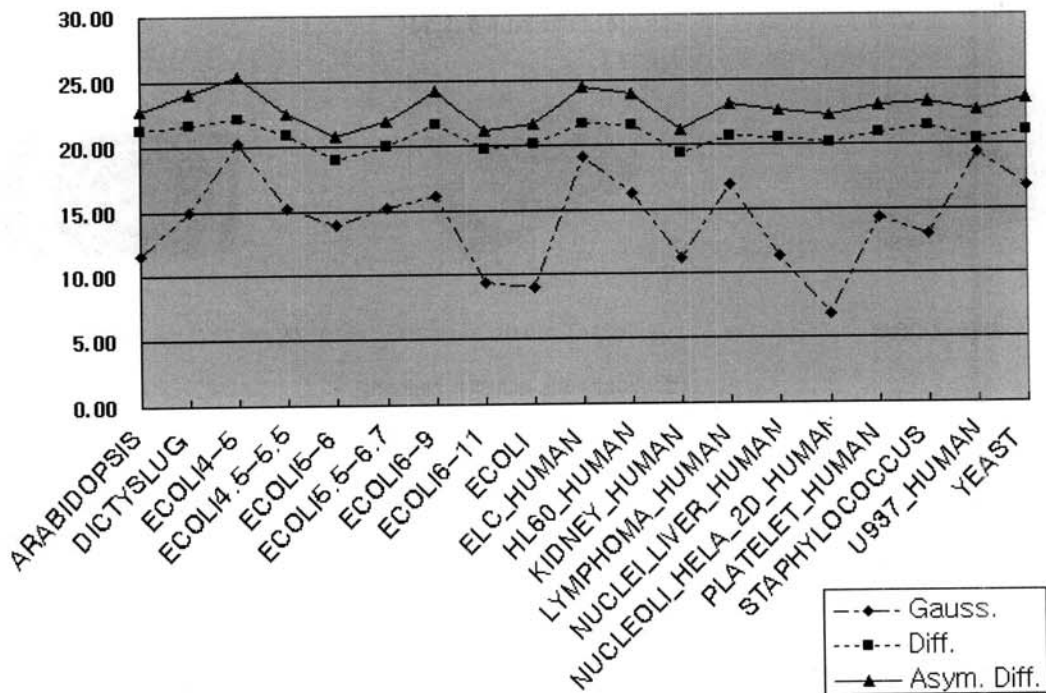
상 전체에 대한 SNR의 평균도 가우스 모형이 14.22dB, 확산 모형이 20.72dB, 비대칭 확산 모형이 22.85dB이어서 비대칭 확산 모형이 가장 크다. 따라서 실험결과로써 세 모형 중에 비대칭 확산 모형이 반점 정합에 가장 적합한 모형임을 확인할 수 있다.

(그림 16)은 <표 1>을 그래프로 나타낸 것으로서 비대칭 확산 모형이 다른 두 개의 모형에 비해서 뚜렷이 나은 정합

〈표 1〉 반점 정합 실험 결과

(단위 : dB)

	Gel image	Model	SNR	No.of Spots		Gel image	Model	SNR	No.of Spots
1	ARABIDOPSIS	Gauss.	11.64	5705	11	HL60_HUMAN	Gauss.	16.18	5946
		Diff.	21.26				Diff.	21.49	
		Asym.	22.77				Asym.	23.94	
2	DICTYSLUG	Gauss.	14.90	5596	12	KIDNEY_HUMAN	Gauss.	11.22	8389
		Diff.	21.68				Diff.	19.35	
		Asym.	24.05				Asym.	21.06	
3	ECOLI4-5	Gauss.	20.15	4137	13	LYMPHOMA_HUMAN	Gauss.	16.84	2792
		Diff.	22.17				Diff.	20.54	
		Asym.	25.33				Asym.	23.14	
4	ECOLI4.5-5.5	Gauss.	15.20	8321	14	NUCLEL_LIVER_HUMAN	Gauss.	11.36	5024
		Diff.	20.92				Diff.	20.53	
		Asym.	22.48				Asym.	22.64	
5	ECOLI5-6	Gauss.	13.86	4746	15	NUCLEOL_HELA_2D_HUMAN	Gauss.	6.80	3978
		Diff.	18.89				Diff.	20.05	
		Asym.	20.70				Asym.	22.24	
6	ECOLI5.5-6.7	Gauss.	15.21	4321	16	PLATELET_HUMAN	Gauss.	14.29	3005
		Diff.	20.01				Diff.	20.83	
		Asym.	21.85				Asym.	23.01	
7	ECOLI6-9	Gauss.	16.12	4364	17	STAPHYLOCOCCUS	Gauss.	13.00	6015
		Diff.	21.63				Diff.	21.42	
		Asym.	24.20				Asym.	23.33	
8	ECOLI6-11	Gauss.	9.37	6777	18	U937_HUMAN	Gauss.	19.25	6669
		Diff.	19.75				Diff.	20.32	
		Asym.	21.17				Asym.	22.66	
9	ECOLI	Gauss.	8.97	3489	19	YEAST	Gauss.	16.78	6189
		Diff.	20.13				Diff.	21.04	
		Asym.	21.60				Asym.	23.57	
10	ELC_HUMAN	Gauss.	19.09	3982		Average	Gauss.	14.22	99445
		Diff.	21.62				Diff.	20.72	
		Asym.	24.42				Asym.	22.85	



(그림 16) 반점 정합 실험 결과

결과를 보임을 확인할 수 있다.

5. 결 론

본 논문에서는 겔 영상 분석 프로그램의 반점 검출 단계에 필요한 형태 모형으로서 비대칭 확산 모형을 제안하였다. 비대칭 확산 모형은 겔 영상 내의 반점의 형태가 비대칭 형태를 가지는 것에 착안한 모형이며 기존의 가우스 모형이나 확산 모형보다 반점의 형태에 가까운 모형이다. 19개의 겔 영상에 대한 반점 정합 실험을 수행한 결과, 19개 겔 영상 전체에 대한 SNR의 평균이 가우스 모형이 14.22dB, 확산 모형이 20.72dB, 비대칭 확산 모형이 22.85dB이어서 비대칭 확산 모형이 가장 나은 결과임을 보았으며, 19개 겔 영상 각각에 대한 SNR의 평균도 비대칭 확산 모형이 가장 나은 결과임을 보았다. 따라서 이와 같은 실험결과로써 비대칭 확산 모형이 가우스 모형과 확산 모형에 비해서 반점 정합에 보다 더 효율적이며 적합한 모형임을 확인하였다.

본 논문에서 구현한 반점 검출 시스템은 반점 검출 단계에서 처리해야 할 문제점 중에 중첩 반점을 처리하기 위한 기법은 고려되지 않았으며, 또한 실험 영상으로 은염 겔 영상만을 사용하여 실험하였으므로 다른 염색기법을 사용한 겔 영상에 대한 실험이 필요하다. 따라서 중첩 반점을 처리하기 위한 기법, 다른 염색기법을 사용한 겔 영상에 대한 실험 등이 우선적인 향후과제이며, 그 후에 반점검출 단계 이후의 단계들인 겔 정합 단계와 정량적 비교 분석 단계에 대한 연구를 수행하고자 한다.

참 고 문 헌

- [1] P. H. O'Farrell, "High Resolution Two-Dimensional Electrophoresis of Proteins," *Journal of Biological Chemistry*, Vol.250, Issue 10, pp.4007-4021, May, 1975.
- [2] A. W. Dowsey, M. J. Dunn and G. Z. Yang, "The Role of Bioinformatics in Two-Dimensional Gel Electrophoresis," *Proteomics*, Vol.3, Issue 8, pp.1567-1597, Aug., 2003.
- [3] L. Addario-Berry, "2D Gel Electrophoresis - An Overview," <http://www.mcb.mcgill.ca/~hallett/GEP/PLectures/Plecture3.pdf>, May, 2002.
- [4] 유혜경, 이성환, "스팟 윤곽의 기술품을 이용한 단백질 2차원 전기영동 영상에서의 스팟 검출 방법", *한국정보과학회 가을 학술발표논문집*, Vol.31, No.2, Oct., 2004.
- [5] 김영호, 김정자, 김대현, 원용관, "Watersheds 기반 계층적 이진화를 이용한 단백질 반점 분할 알고리즘", *한국정보처리학회 정보처리논문지*, 제12-B권, 3호, pp.239-246, Jun., 2005.
- [6] L. Pedersen, "Analysis of Two-Dimensional Electrophoresis Gel Images," PhD Thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Feb., 2002.
- [7] K. P. Pleissner, F. Hoffman, K. Kriegel, C. Wenk, S. Wegner, A. Sahlstrom, H. Oswald, H. Alt and E. Fleck, "New Algorithmic Approaches to Protein Spot Detection and Pattern Matching in Two-Dimensional Electrophoresis Gel Databases," *Electrophoresis*, Vol.20, pp.755-765, Apr., 1999.
- [8] A. Efrat, F. Hoffman, K. Kriegel, C. Shultz and C. Wenk, "Geometric Algorithms for the Analysis of 2D-Electrophoresis Gels," *Journal of Computational Biology*, pp.114-123, Sept., 2002.
- [9] K. Takahashi, M. Nakazawa, Y. Watanabe and A. Konagaya, "Fully-Automated Spot Recognition and Matching Algorithms for 2-D Gel Electrophoretogram of Genomic DNA," *Genome Informatics*, Yebisu garden place, Tokyo, Japan, pp.161-172, Dec., 1998.
- [10] 심정은, 이원석, "다차원 클러스터링 기반의 단백질 2DE 이미지에서의 자동화된 기준점 추출 방법", *한국정보처리학회 정보처리논문지*, 제12-D권, 5호, pp.719-728, Oct., 2005.
- [11] M. Rogers and J. Graham, "Robust and Accurate Registration of 2-D Electrophoresis Gels using Point-Matching," *IEEE Transactions on Image Processing*, Vol.16, No.3, pp.624-635, Mar., 2007.
- [12] A. Roy, K. R. Lee, Y. Hang, M. Marten and B. Rauman, "Analyzing Two-Dimensional Gel Images," *Technical Reports*, Dept. of Mathematics and Statistics, University of Maryland, Sept., 2003.
- [13] L. Vincent and P. Soille, "Watershed in Digital Spaces: An Efficient Algorithm based on Immersion Simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, No.7, pp.583-598, Jul., 1990.
- [14] R. D. Appel, D. F. Hochstrasser, M. Funk, J. R. Vargas, C. Pellegrini C, A. F. Muller and J. R. Scherrer, "The MELANIE Project: from a Biopsy to Automatic Protein Map Interpretation by Computer," *Electrophoresis*, Vol.12, No.10, pp.722-735, Oct., 1991.
- [15] E. Bettens, "Peak Characterization using Parameter Estimation Methods," PhD thesis, University of Antwerpen, May, 1999.
- [16] 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, Issue 6, pp.887-896, Jun., 2003.
- [17] S. Beucher, "Recent Advances in Mathematical Morphology," Invited lecture, Société Italienne de Microscopie Electronique, Taormina, Sicile, Sept., 1992.
- [18] 최관덕, 김미애, 윤영우, "2차원 전기영동 영상에서 잡음을 제거하기 위한 적응적인 문턱값 결정", *신호처리·시스템학회 논문지*, 제9권, 1호, pp.1-9, Jan., 2008.
- [19] C. Hoogland, K. Mostaguir, J. C. Sanchez, D. F. Hochstrasser and R. D. Appel, "SWISS-2DPAGE, Ten Years Later," *Proteomics*, Vol.4, Issue 8, pp.2352-2356, Apr., 2004.



최관덕

e-mail : kdchoi@tsc.ac.kr

1989년 영남대학교 전산공학과(공학사)

1991년 영남대학교 대학원 전산공학과(공학석사)

2008년 영남대학교 대학원 전산공학과(공학박사)

1996년 3월~2008년 2월 대구과학대학 컴퓨터계열 조교수

2008년 3월~현 재 대구과학대학 의무기록정보과 조교수

관심분야 : 바이오인포매틱스, 영상처리, 프로그래밍 언어 등



윤영우

e-mail : ywyoong@ynu.ac.kr

1972년 영남대학교 전자공학과(공학사)

1983년 영남대학교 대학원 전자공학과(공학석사)

1988년 영남대학교 대학원 전자공학과(공학박사)

1988년 9월~현 재 영남대학교 전자정보공학부 컴퓨터공학전공 교수

관심분야 : 바이오인포매틱스, 영상처리, 컴퓨터비전, 컴퓨터설계, 프로그래밍 언어 등