

반복 점진적 방법에 의한 2차원 단백질 분리 영상의 반점 정합

김정자¹, Minh T. Hoang², 김동욱¹, 김남균¹, 원용관²

¹전북대학교 공과대학 바이오메디컬공학부

²전남대학교 공과대학 컴퓨터공학과

(Received May 22, 2007. Accepted August 6, 2007)

An Iterative Spot Matching for 2-Dimensional Protein Separation Images

Jung-Ja kim¹, Minh T. Hoang², Dong Wook Kim¹, Nam Gyun Kim¹, Yonggwon Won²

¹Division of Biomedical Engineering, Collage od Engineering, Chonbuk National Univ

²Department of Computer Engineering, Chonnam National Univ.

Abstract

2 Dimensional Gel Electrophoresis (2DGE) is an essential methodology for analysis on the expression of various proteins. For example, information for the location, mass, expression, size and shape of the proteins obtained by 2DGE can be used for diagnosis, prognosis and biological progress by comparison of patients with the normal persons. Protein spot matching for this purpose is comparative analysis of protein expression pattern for the 2DGE images generated under different conditions. However, visual analysis of protein spots which are more than several hundreds included in a 2DGE image requires long time and heavy effort. Furthermore, geometrical distortion makes the spot matching for the same protein harder. In this paper, an iterative algorithm is introduced for more efficient spot matching. Proposed method is first performing global matching step, which reduces the geometrical difference between the landmarks and the spot to be matched. Thus, movement for a spot is defined by a weighted sum of the movement of the landmark spots. Weight for the summation is defined by the inverse of the distance from the spots to the landmarks. This movement is iteratively performed until the total sum of the difference between the corresponding landmarks is larger than a pre-selected value. Due to local distortion generally occurred in 2DGE images, there are many regions in which many spot pairs are miss-matched. In the second stage, the same spot matching algorithm is applied to such local regions with the additional landmarks for those regions. In other words, the same method is applied with the expanded landmark set to which additional landmarks are added. Our proposed algorithm for spot matching empirically proved reliable analysis of protein separation image by producing higher accuracy.

Key words : 2 dimensional gel electrophoresis(2DGE), digital image processing, protein spot segmentation, protein spot matching

1. 서론

2 DGE는 단백질의 전기적 특성과 질량정보를 이용하여 생물체 내에 혼합되어 존재하는 단백질을 분리하는 실험 방법이다. 단백질 분리의 결과는 특정 단백질의 발현 정도나 변형 상태 등을 파악할 수 있으며, 이러한 정보를 바탕으로 각종 질병의 발생원인, 진행 상태, 생리적인 변화 등을 분석할 수 있다. 이러한 전

기영동 방법은 단백체학(Proteomics) 연구에 절대적으로 필요한 중요한 수단이며, 이를 위하여 분리된 단백질의 검출 및 상호 비교는 중요한 절차이다. 즉, 2DGE로 얻어진 분리된 단백질의 반점 (spot)에 대한 정합(matching)은 비교의 기준이 되는 참조 (reference) Gel과 비교의 대상이 되는 대조(matched) Gel에 존재하는 동일한 단백질 반점을 1:1 상호 대응시켜 단백질의 발현양상(expression pattern)을 파악하는 필수적 과정이며, 이를 위한 효율적 방법론에 대한 연구는 생물정보학 및 단백체학 분야의 중요한 흥미를 제공하고 있다[1].

이러한 단백질 발현 양상의 비교 과정은 2DGE 실험의 결과로부터 얻어진 디지털 영상에서 이루어지게 된다. 즉, 단백질 분리의

이 논문은 2006년도 전북대학교 지원 신임교수연구비에 의하여 연구되었음

Corresponding Author : Jung-Ja Kim, Professor
Division of Biomedical Engineering, Collage of Engineering,
Chonbuk National University
664-14 Duckjin-Dong, Jeonju-si, Jeonbuk, 561-756, South Korea
Tel : +82-63-270-4102 / Fax : +82-63-270-4062
E-mail : jungjakim@chonbuk.ac.kr

결과는 Gel위에 시료를 투여(loading)하고 일정 시간이 지난 후에 디지털 영상으로 얻어지게 되며, 먼저 이러한 디지털 영상으로부터 단백질 반점의 추출(segmentation)이 이루어진다. 불행히도 2DGE 영상에는 최소 수백 개 이상의 단백질 반점을 포함하고 있으며, 배경(background)과 밝기의 차이, 두 개 이상의 반점이 겹쳐 있는 등의 영상 내에 존재하는 다양한 유형의 잡음으로 인하여 수작업으로 처리할 경우 많은 경험과 시간을 필요로 하는 문제점을 안고 있다[2]. 이것이 바로 컴퓨터에 의한 디지털 영상 처리가 절대적으로 필요한 이유이며, 그 동안 많은 연구가 이루어져 왔지만 아직도 절대적 자동화의 단계를 달성하지 못하고 있다.

단백질 반점이 추출된 후에는 각 반점으로부터 크기, 모양, 위치 등으로 해당 단백질에 대한 고유한 정보 추출하게 된다. 이러한 정보를 기반으로 참조 Gel과 대조 Gel 영상 내의 반점들을 상호 대응시키는 단백질 반점의 정합을 수행하게 되며, 그 동안 반점 정합을 위한 많은 방법들이 제안되었지만 대부분은 컴퓨터에 의한 자동화 과정만으로는 만족할 만한 정합의 결과를 얻을 수 없으며, 정합 이전 및 이후에 사용자의 관여를 요구하고 있다[3-5].

단백질 반점 정합의 핵심 관점은 참조 Gel과 대조 Gel의 두 영상에 존재하는 동일 단백질에 대한 최적의 정합을 얻어 내는 것이다. 이를 위하여 사용자의 관찰에 의하여 수 개의 정합 쌍 반점을 지정하게 되는데, 이러한 정합 쌍 반점을 기준반점(landmark)라고 한다. 즉, 기준점이 되는 반점 쌍은 사용자에 의하여 참조 Gel과 대조 Gel에 존재하는 반점 쌍으로 동일한 단백질의 반점이며, 기준반점 쌍의 선정은 정합 성공률의 1차적인 중요한 요인이 된다. 정합 성공에 관한 2차적인 중요한 요인은 기준반점을 기반으로 다른 반점들을 정확히 짝 짓기 하는 방법이다.

본 논문에서는 효율적인 단백질 반점 정합을 위해 기준반점들과의 기하학적인 거리에 기반 한 2단계 반복적 정합 알고리즘을 제안한다. 제안 알고리즘은 모든 기준반점들 상호 간의 차이를 줄이기 위한 정보를 종합하여 자신의 이동 방향과 이동양을 결정한다. 이러한 과정을 마친 후, 다음 단계로는 정확하게 정합이 되지 않은 부분 영역에 대하여 동일한 방법으로 정합을 수행하여 최적의 결과를 얻는다. 제안된 알고리즘의 효용성을 검증하기 위하여, 인공적으로 생성한 영상과 실제 단백질 분리 영상에 적용하였으며, 그 결과는 부분 영역 정합을 수행하지 않는 경우에 비하여 더 좋은 성능을 보였다.

II. 2차원 Gel 전기영동 영상처리 (2DGE Image Processing)

A. 2DGE 영상 분석 단계

2DGE 는 단백질 분리에 사용되는 기본적인 핵심 요소 기술로서 1차원으로는 등전점(isoelectric: PI)에 의하여 분리하고, 2차원으로는 단백질의 무게를 나타내는 분자 질량(molecular weight)에 따라 분리한다. 분리과정의 첫 번째는 다양한 단백질이

섞여 있는 복합체를 특수하게 처리된 Gel 위에 올려놓고 양쪽에 전하를 걸면 전위차가 발생하여 미리 Gel 상에 처리된 화학적 반응에 의하여 단백질의 등전점에 따라 분리가 된다. 두 번째는 또 다른 방향으로 전하를 걸면 각 단백질의 무게 차이에 의해 Gel 상에서 움직이는 속도가 달라져 일정 시간을 두면 질량에 따라 분리가 된다. 따라서 이러한 과정을 거쳐 2차원적인 분리의 결과를 얻게 된다.

분리된 단백질들에 대해서는 질량분석 기법으로 단백질의 특성을 규명하게 된다. 분리과정에 의해 등전점과 질량으로 2차원 Gel 상의 다른 위치에 단백질들이 존재하여 영상으로 획득되어진다. 2DGE 영상 분석방법은 단백질 분리 영상으로부터 단백질 반점을 추출하는 단계(Protein spot segmentation)와 복수 영상에 대하여 단백질 반점들을 상호 정합하는 단계로 구성된다. 이러한 과정을 통하여 단백질의 발현 현상에 대한 정보를 얻게 된다. 이것을 2D Gel 영상처리 기술에 적용하면 2D Gel 영상의 획득과 단백질 반점의 추출 그리고 분리된 단백질 반점의 상호 정합을 통한 비교의 과정을 거쳐 단백질의 발현 양상을 분석한다. 그 과정을 구체적으로 살펴보면 첫 단계인 영상 획득은 2차원 전기영동의 결과를 스캐너 또는 디지털 카메라를 이용하여 디지털 영상으로 얻는다. 이러한 영상은 배경 변화 및 Gel의 밀집 현상 등과 같은 잡음의 제거 및 반점의 두드러짐을 처리하기 위해 영상 향상(enhancement)을 통해 단백질 반점을 추출한다. 이 과정에서는 영상의 질(Quality)을 높이기 위한 다양한 영상 처리 기법들이 적용된다[2,6]. 두 번째 단계는 영상 분할(segmentation)으로 하나 이상의 단백질 반점이 중첩되어 있는 경우, 잡음으로 인식되는 지나치게 작거나 지나치게 큰 반점들에 대한 처리가 중요한 관건 중의 하나가 된다. 대부분 사용자의 관여가 요구되는 2D gel 영상의 특성은 반점의 최소 크기와 최대 크기에 대한 정의, 단백질의 발현 양(amount)이 영상 정보로 나타나는 가장 흐린 반점과 가장 강한 반점에 대한 정의이다. 세 번째 단계는 참조 Gel과 대조 Gel 간의 동일한 반점을 정합하는 과정이다[5-9]. 이 과정은 두 Gel들에 존재하는 몇 개의 반점 쌍들의 정합으로써 이 반점 쌍들은 동일한 단백질임을 나타내고 이들을 기준반점이라 한다. 그리고 다른 반점들은 이들 기준반점들이 정합되는데 필요한 정보를 기반으로 상호 정합이 되게 된다. 따라서 이 단계의 목표는 기준 반점의 선정이다.

서로 다른 조건으로부터 얻은 두 Gel 영상의 반점 정합의 목표는 어느 한쪽에만 존재하거나 반점의 크기와 강도로 나타나게 되는 발현 양의 차이를 식별하는 것이다. 예를 들어, 정상인과 암환자의 혈청을 이용하여 획득한 2D Gel 영상 사진의 경우, 두 Gel 간에 매치되지 않은 반점들은 환자와 정상 그룹 간을 분류하기는 생물학적 표지자(Bio-Marker)들이다. 이 표지 자는 진단, 치료, 단백질 분석 및 신약 개발 등의 연구에 중요한 단서를 제공함으로써 표지자 단백질의 탐지는 대단히 중요한 과정이며, 이러한 목표를 달성하기 위한 중요한 기술 중의 하나가 2D Gel의 단백질 반점 추출 및 상호 정합이다.

B. 관련 연구

단백질 반점의 탐지와 정합 기술은 2D Gel 실험 방법이 가진 오랜 역사와 함께 오랜 동안 널리 사용되어 왔으며, 그 중요성 역시 오랜 역사를 가지고 있다. 더구나, 수년 전부터는 실험의 양이 방대해지면서 반점 추출 및 정합을 위한 정교하고 고속처리 방법이 요구되고 있다. '점 정합(Point matching)'은 2차원 행렬의 Affine 변환(Transform)을 이용하여 참조 영상과 대조 영상 사이의 크기(Scale)와 회전(Rotation)에 의해 변환되는 위치 정보를 기반으로 정합을 처리하는 방법이다[10]. '유사성 기반 방식'은 임의의 반점이 주변 반점과 형성하는 기하학적 배치에 대한 정보를 근거로, 참조 Gel과 대조 Gel 사이의 반점들에 대해 상호 유사성이 높은 반점들에 대해 쌍을 이루어 나가는 방법이다[11]. Rangarajan은 지능적인 점 정합 알고리즘을 개발하였는데, 각 반점 쌍들이 가지고 있는 차이를 에너지로 표시하고 이 에너지의 차이를 낮추어가는 데 담금질(annealing) 방법을 이용하여 점진적으로 정합을 이루어 나간다. 담금질 기법은 반복적인 과정을 거쳐야 하는데, 이를 위해 Shinkhorn의 기법은 행렬의 정규화를 통하여 정합을 위한 담금질 계산을 쉽게 해준다[12]. 이 외에도 탄성력 에너지를 이용해서 정합하는 에너지 모델로서 탄성 그래프 정합(elastic graph matching) 및 Delaunay 삼각형 그래프(triangle graph)를 이용하여 단백질 반점 좌표를 2차원 또는 3차원의 물체로 나누고 재구성하여 분석하는 방법들이 제안되었다[13].

단백질 반점 정합에 있어서 문제의 주된 관심은 '어떻게 하면 기하학적인 일그러짐을 잘 처리할 수 있을 것인가?'이며, 이를 위한 다양한 방법들이 제안되었다[14-16]. CAROL 소프트웨어의 경우는 기준반점 쌍을 이용하지 않고 점 정합 문제를 부분선분(line segment) 정합 문제로 변환하기 위하여 Affine 변환을 사용한다[17]. Smilansky의 경우는 원래의 영상으로부터 정합을 처리하는 순수한 영상기반의 정합 기법이다[18].

지금까지 개발되거나 연구된 대부분의 방법들은 Gel 간의 정합

에서 충분한 수준의 성능과 결과를 보이지 못하여 본 논문에서는 참조 Gel 과 대조 Gel 간의 정교하고 간단한 정합 방법을 제안하여 앞서 발생된 문제점을 해결하고자 한다. 이 방법은 사용자에 의하여 선정된 기준반점 쌍들 간의 기하학적 차이를 줄이기 위하여 각 기준반점들 간의 차이 정보를 이용한다. 모든 기준반점 쌍들이 상호 간에 정합거리를 줄이기 위해 움직여야 하는 방향으로 대조 Gel 에 존재하는 모든 반점들이 움직인다. 이때 각 기준반점들이 움직여야 하는 방향과 거리에 반비례하여 각 반점에 영향을 주게 된다. 3 장에서는 이에 대한 알고리즘의 상세한 설명이 이루어진다.

III. 기준점 기반의 점진적 단백질 반점 정합 방법

본 논문과 관련된 선행연구로서 반점 정합 이전 단계인 '영상의 질 향상(quality enhancement)'과 '반점의 추출(spot segmentation)' 방법을 개발하였으며, 'ProteomBASE'라는 상용화 제품[8]으로 판매되고 있다. 그림 1은 해당 제품의 반점 추출 결과를 보여주는 예시이다.

단백질 반점 추출과 정합 과정에서 가장 어려운 문제점은 2D Gel의 특수성으로 인하여 '고무판 일그러짐 (Rubber-Sheet Distortion)' 현상을 갖는다[3]. 예를 들어, 참조 Gel과 대조 Gel의 반점들이 전역적으로뿐만 아니라 지역적으로도 서로 정합되지 않는 부분영역이 발생하는 것이다. 이는 사용자로부터 기준반점이라는 반점 쌍을 정의하여 줌으로써 어느 정도 해결 할 수 있다. 기존의 많은 연구에서는 다양한 형태의 복잡한 일그러짐 때문에 사용자에 의해 정의된 기준반점 쌍을 사용하였다. 이는 정합 과정에 어느 정도의 사용자의 개입을 요구하는 것이며, 이때 기준반점 쌍의 선택은 정합의 결과에 대한 매우 중요한 문제이다. 그림 2는 참조 Gel과 대조 Gel의 반점 추출 결과를 입력으로 하여 전역적 정합과 지역적 정합 과정을 거쳐 최종적으로 정합 결과를 얻는 흐름도이다.

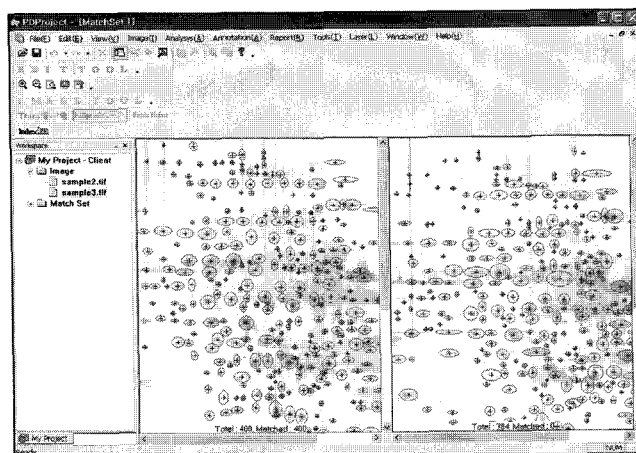


그림 1. 단백질 반점 영상 처리 및 반점 추출
Fig. 1. Protein spot image processing and spot segmentation

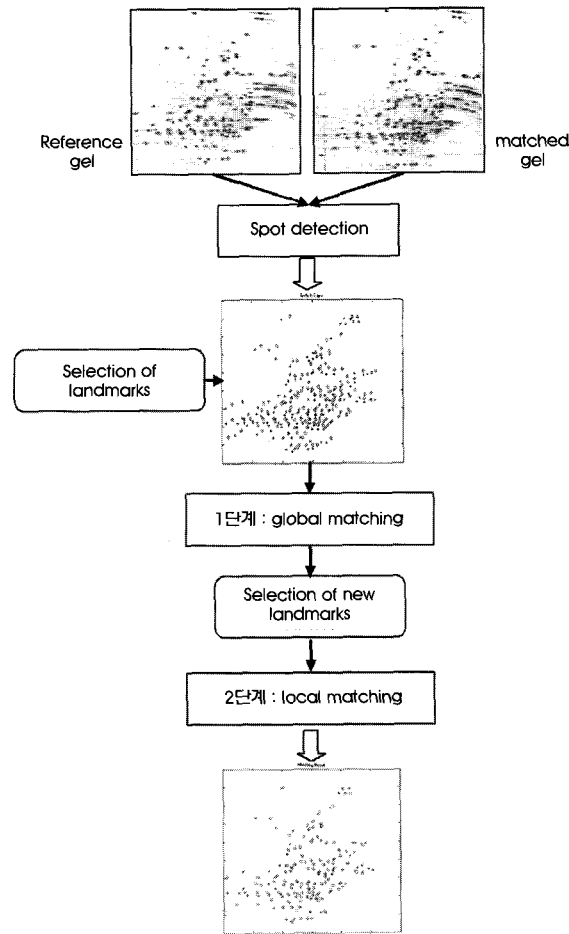


그림 2. 전역 및 국부 정합 과정으로 구성된 반점 정합 흐름도
 Fig. 2. Flow of spot matching with global and local matching procedures.

A. 단백질 반점의 점진적 전역 정합

본 논문에서는 참조 Gel과 대조 Gel 간의 최적의 정합 도를 달성 위해서 유클리디안 좌표계로 표현되는 방향과 거리에 근거한 기준 반점 쌍 간의 기하학적인 차이를 점진적이고 단계적으로 줄이는 정합 알고리즘을 제안한다. 즉, 본 논문에서는 단백질 반점의 등전 점과 분자량을 2차원 공간의 좌표 값으로 나타나는 정보로 표현하고, 참조 Gel과 대조 Gel에 존재하는 서로 상응되는 반점들끼리 최대한 동일한 좌표를 갖도록 정합을 수행하는 알고리즘을 연구하였다. 제안하는 알고리즘은 1차적으로 Gel 영상에 대해서 전역적으로 단백질 반점 쌍들을 서로 정합을 하고, 2차적 단계에서는 정합이 충분하지 않는 지역에 대하여 기준반점을 선택하고, 이를 기준으로 하여 국부적으로 부분정합을 수행하여 전반적으로 최적의 정합 결과를 얻게 된다.

반점 정합의 최종 목표는 두 Gel 영상에 있는 반점들의 분포 형태가 최대한 동일한 유형을 갖도록 하는 것으로, 특정 단백질을 나타내는 대조 Gel 영상 내의 반점 위치를 참조 Gel 영상 내의 반점이 갖는 좌표 위치로 이동시켜 최종적으로 두 대응 반점들이 다른

어떤 반점들보다 가까운 위치가 되도록 하는 것이다. 2D Gel 영상의 극심한 왜곡으로 인하여 사용자에게 의한 기준반점 쌍의 선택이 요구된다. 사용자가 지정한 두 반점 쌍이 동일하다고 판단될 수 있도록 전체의 반점들을 이동하게 되면, 다른 일반 반점들도 적절하게 자신과 대응되는 반점과 정합이 이루어진다. 이상적으로 두 Gel이 정합되면 사용자가 지정한 기준점들은 동일한 좌표를 가지게 되는 것이다. 그리고 일반 반점들은 기준반점들이 이동하여야 하는 방향 및 거리(이동 량)는 모든 기준반점들의 이동 방향 및 거리에 대한 가중치 합에 비례하며, 가중치는 해당 기준반점까지의 거리에 반비례한다. 즉, 한 반점이 이동하여야 하는 척도는 기준반점들의 이동 척도에 의하여 결정이 되는데, 가까운 기준반점에 대해서는 더 큰 영향을 받고 멀리 있는 기준반점에 대해서는 적은 영향을 받는 것이다. 이는 지역적으로 기하학적인 왜곡 현상이 발생하는 2D Gel 영상의 특성을 고려할 때 논리적으로 타당한 방법이다. 그림 2에서 어떤 반점의 이동은 기준반점 쌍(L_{r1}, L_{m1}) 간의 이동에 대해서는 d1만큼, 기준반점 쌍(L_{r2}, L_{m2})에 대해서는 d2만큼 영향을 받아 이동하게 됨을 도식하고 있다.

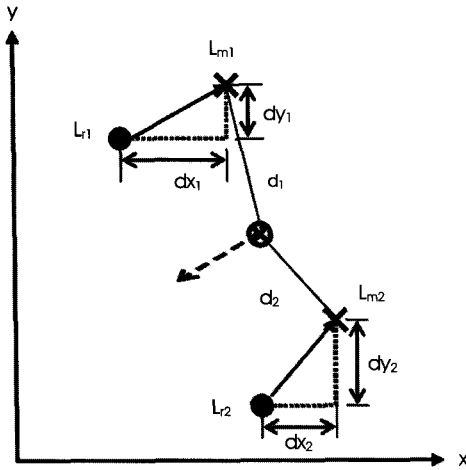


그림 3. 기준반점 쌍과 반점 이동간의 상관 관계
Fig. 3. Relation between landmark spot pair and spot movement

그림 3에서 (dx, dy) 는 기준 Gel에 존재하는 기준반점(●)과 대조 Gel에 존재하는 기준반점(X)간의 기하학적 차이를 의미하며, 는 어떤 임의의 반점을 나타낸다. 여기에서 두 기준반점 쌍 간의 기하학적인 차이는 다음 식 (1)과 같다.

$$(dx, dy) = (x^r - x^m, y^r - y^m) \quad (1)$$

이때 (x^r, y^r) 는 기준 Gel에 위치한 기준반점의 위치 좌표이고, (x^m, y^m) 은 대조 Gel에 위치한 기준반점의 좌표 값이다. 따라서, 최적의 정합 결과는 $dx = x^r - x^m = 0$ 및 $dy = y^r - y^m = 0$ 이며, 정합 과정은 기준반점 쌍들 간의 간격을 줄이는 방향으로 대조 Gel의 반점을 이동하는 것이다. 즉, 그림 3에서 반점 X는 $(dx \rightarrow 0, dy \rightarrow 0)$ 가 되도록 반점 ●을 향하여 이동하게 된다.

한편, Gel 영상의 국부 영역이 서로 다른 일그러짐 특성을 갖고 있으므로, 가까운 기준반점의 이동에 더 큰 영향을 받게 되며, 전체 이동은 모든 기준반점의 영향의 합으로 결정된다. 그림 3에서 로 표시된 임의의 한 반점은 주변 기준반점들에 의하여 영향을 받는데 거리에 반비례하여 그 영향력이 결정된다. 다시 말하면, 기준반점 쌍 (L_{r1}, L_{m1}) 에 대해서는 d_1 에 반비례하여, 기준반점 쌍 (L_{r2}, L_{m2}) 에 대해서는 d_2 에 반비례하여 이동 영향을 받는다. 여기에서 대조 Gel에 존재하는 임의의 반점 p 에 대한 이동은 식(2)와 같다.

$$(dx_p, dy_p) = \left(\sum_{j=1}^n \frac{dx_j}{d_j^p}, \sum_{j=1}^n \frac{dy_j}{d_j^p} \right) \quad (2)$$

이때 n 은 사용자가 지정한 기준반점 쌍들의 개수이고, 임의의 기준반점 j 와 반점 p 간의 거리는 식(3)과 같다.

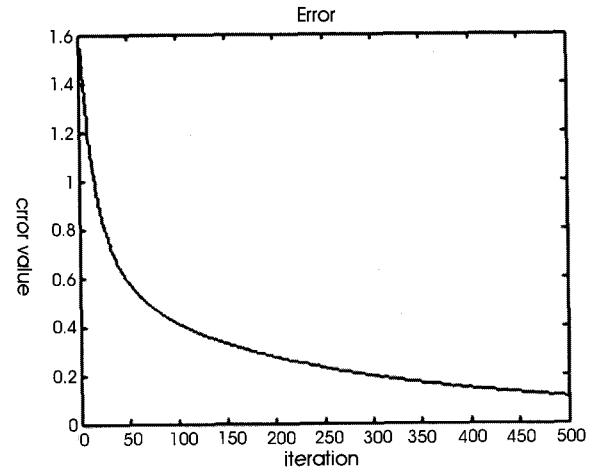


그림 4. 반복적 정합 과정 동안의 오차 값 변화
Fig. 4. Error change during iterative matching process

$$\begin{aligned} d_j^p &= \sqrt{(dx_j^p)^2 + (dy_j^p)^2} \\ &= \sqrt{(x_j^m - x_p)^2 + (y_j^m - y_p)^2} \end{aligned} \quad (3)$$

로 정의된다.

임의의 반점 p 의 새로운 좌표는 원 위치에 식 (2)에 의하여 계산된 이동량을 더하여 구하며 식 (2)에 의한 새로운 위치로의 이동량은 식 (4)와 같다.

$$\begin{aligned} x_p^{new} &= x_p^{old} + \alpha \cdot dx_p = x_p^{old} + \alpha \cdot \sum_{j=1}^n \frac{dx_j}{d_j} \\ y_p^{new} &= y_p^{old} + \alpha \cdot dy_p = y_p^{old} + \alpha \cdot \sum_{j=1}^n \frac{dy_j}{d_j} \end{aligned} \quad (4)$$

이때 α 는 1보다 작은 양수로 이동량의 일부만을 반영하기 위한 인자이다.

본 논문에서 제안하는 반점 정합 알고리즘은 식 (4)과 같이 한 시점에서 계산된 이동량의 일부만을 반영하여 반점을 이동하며, 이러한 과정은 지속적으로 반복되게 된다. 또한, 식 (4)에 의하여 모든 반점에 대하여 정합이 수행될 때 기준반점들도 일반 반점과 동일하게 취급되어 정합 과정을 거치게 된다. 여기에서 식 (5)는 한 시점에서 두 Gel 영상의 정합 오차를 나타내는 함수로 사용자가 지정한 기준반점 쌍들 간의 기하학적 차이의 합이다.

$$E = \frac{1}{n} \sum_{j=1}^n \sqrt{(x_j^r - x_j^m)^2 + (y_j^r - y_j^m)^2} \quad (5)$$

식 (4)의 정합을 반복적으로 수행하는 알고리즘은 두 개의 종료 기준을 갖는다. 첫 번째 기준은 '최대 반복 횟수'이며, 두 번째 기준

```

Set Max_Itr, T, Landmarks, E=0, i=0;
DO
  Matching(Reference, Matched);
  Compute Error E
  Increase i by 1 (i = i + 1);
UNTIL (i <= Max_Itr) AND E <= T

Matching(Ref_Img, Mat_Img)
FOR all spots in matched gel
  FOR all Landmarks
    Compute moving amount by Eq. 2.
  END FOR
  Compute new location by Eq. 4.
END FOR
END matching
    
```

그림 5. 반점의 반복 점진적 전역 정합 알고리즘
Fig. 5. Iterative Global spot matching algorithm

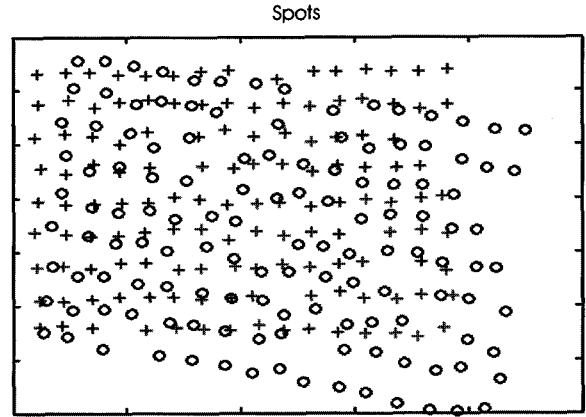


그림 6. 성능 평가를 위하여 인위적으로 생성된 반점 쌍들
Fig. 6. Synthetically generated spot pairs for performance evaluation

은 식 (5)로 정의된 정합 오차 함수 E 를 위한 임계값으로 E 값이 최소일 때 두 Gel 영상의 반점에 대한 최적의 정합 도를 가질 것이라 예상하며, 가장 이상적인 경우의 E 값은 0이 될 것이다. 그림 4는 반복적 정합 과정을 통하여 정합 오차를 감소시킨다. 이 그림은 본 논문에서 제안하는 알고리즘으로 반복 횟수와 오차 간의 상관관계 그래프를 나타내고 있다. 그리고 그림 5는 지금까지 설명한 단백질 반점의 점진적 전역 정합 알고리즘을 요약한 것이다.

B. 단백질 반점의 점진적 국부 정합

전역적 반점 정합의 결과를 분석해 보면, 오차 함수 E 값이 더 이상 줄어들지 않은 상황과 줄어든 이후에도 정합이 적절하게 이루어지지 않은 국부적 영역이 발생한다. 이는 2차원 Gel이 갖는 국부적 기하학적 왜곡이 그 원인이다. 특히, Gel의 가장자리 영역은 갈라짐 또는 밀림 현상 등이 자주 발생하게 되는데, 이는 정합의 전체적인 만족도를 떨어뜨리는 주요 원인으로 작용한다. 따라서, 정합이 불완전한 영역에 대하여 정합 도를 높이기 위한 추가적인 처리가 요구된다.

전역 정합 과정에서 정합이 불완전하게 이루어진 국부적 영역에 대해 해당 국부 영역을 집중하여 정합이 요구되며 정합의 결과는 기준반점의 선택이 가장 중요한 요인이 된다. 즉, 특정 국부 지역의 정합이 불완전하게 된 가장 중요한 원인은 그 지역에서의 기준반점 선택이 잘못 되었거나 또는 기준반점 선정이 이루어지지 않은 것이다. 전역적 정합 과정 이후에 수행되는 국부적 정합의 목표는 정합이 불완전한 영역에 대하여 적절한 수준의 정합을 달성하는 것이다. 이는 단지 정합이 불완전한 국부적 영역에 대해 기준반점을 선택하여 전역적 정합의 방법을 동일하게 적용하는 것이다. 그림 5와 같이 전역 정합의 방법을 동일하게 적용하되 단지 정합이 불완전한 국부 영역을 중심으로 기준반점을 새로이 추가 선정하는 것이며 Max_Itr , T , E 및 i 와 같은 변수들을 새로운 값으로 설정한다.

IV. 실험적 분석 및 고찰

A. 성능 평가 데이터

단백질 반점 추출 및 정합은 Gel의 질에 많이 의존하므로 정합 과정의 성능에 대한 명백한 기준을 설정하여야 한다. 본 논문에서는 정합의 성능에 대한 합리적 비교 기준의 설정을 위하여 인위적으로 생성한 가상 반점 데이터를 사용하였다. 인위적으로 생성한 반점은 140개의 반점 쌍으로 구성된 2개의 가공 Gel로 생각할 수 있다. 먼저 기준 Gel의 반점들은 일정 크기의 격자와 유사한 형태의 좌표에 임의적 잡음을 가하여 2차원 가상 반점 영상을 생성하였다. 이렇게 생성된 기준 Gel의 반점에 크기, 회전 및 국부적 왜곡을 가하여 대조 Gel의 반점을 생성하였다. 따라서, 생성 과정에서 이미 정합이 이루어져야 하는 반점 쌍들이 결정되어 있는 것이다. 이렇게 인위적으로 생성된 가상 반점 영상의 사례가 그림 6에 나타나 있다.

B. 전역적 및 국부적 정합 결과 및 분석

전역적 반점 정합만을 수행한 경우, 국부적인 왜곡을 갖는 2차원 전기영동 Gel 영상의 특성으로 인하여 정합이 불완전한 국부적 영역이 발생한다.

그림 7은 그림 6의 반점들에 대하여 전역적 정합을 수행한 결과이다. 이 결과를 살펴보면 좌측 하단 모퉁이 영역, 좌측 상단 모퉁이 영역 및 우측 상단 모퉁이 영역의 정합 결과가 부족함을 볼 수 있다. 이는 기준반점의 선정 단계에서 이 영역을 대표하는 기준반점의 선정이 부족하였거나, 지역적 왜곡이 심한 영역이다.

전역적 정합의 결과를 보여주는 그림 7에서 좌측 상단이 불완전 정합 영역에서 하나의 기준반점 쌍을 추가적으로 설정하여 2차적인 정합을 수행하였다. 그 결과는 그림 8에 도식 되어있으며, 그림 7과 비교하여 불매 새로운 기준반점을 설정한 좌측 상단의 국부적 영역의 정합이 개선되었음을 관찰할 수 있다. 특히, 좌측 하단 및

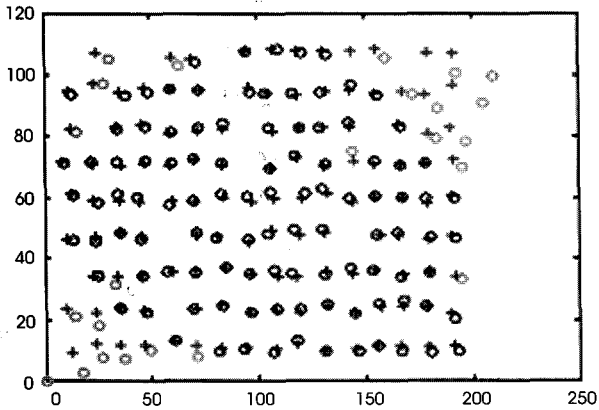


그림 7. 일부 모퉁이 영역의 불완전 정합을 갖는 전역적 정합 결과
Fig. 7. Global matching result having miss-matches in some corners

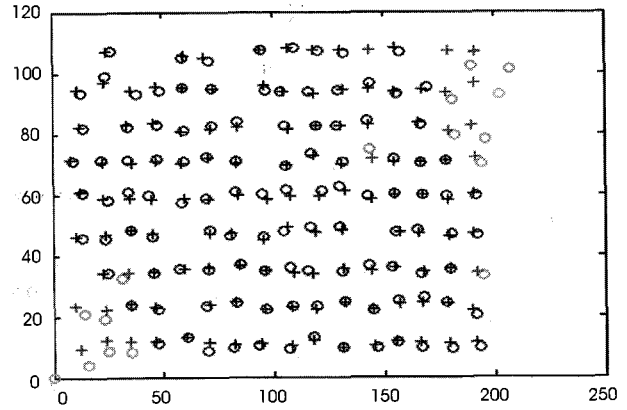


그림 8. 새로운 기준반점을 이용한 2차적인 국부적 정합의 결과
Fig. 8. Local matching result with new landmarks following global matching

우측 상단의 정합은 원래의 정합 결과와 비교해 볼 때 미미한 차이만을 보이고 있는데, 이는 새롭게 추가한 기준반점이 두 국부 영역으로부터 멀리 떨어져 있기 때문이다. 즉, 이 두 지역은 원래의 기준반점들에만 의존하여 정합이 이루어지게 되므로 2차적인 국부적 정합 과정에서도 정합이 충분하게 이루어 지지 못하게 되는 것이다.

C. 기준반점 수와 정합 성능과의 관계

기준반점의 선택은 두 가지 사항을 고려하여야 한다. 첫째는 왜곡의 특성이 다른 지역들을 대표할 수 있는 반점들의 선택이고, 두 번째는 얼마나 많은 수의 기준반점을 선택하느냐이다. 첫 번째 사항은 다분히 사용자의 개인적 주관과 경험에 의존하게 되는 요인으로 기준반점의 위치적 설정을 위한 객관적인 방법의 수립이 곤란하다. 다만, 어느 한쪽 영역에 치우치지 않도록 균형적으로 산발되게 선정하는 정도가 객관적인 방법일 것이다. 두 번째 사항에 대해서는 많은 기준반점이 더 좋은 정합 결과를 제공할 것이라는 예측을 쉽게 할 수 있다. 표 1은 기준반점의 수에 따른 정합 성공율을 나타내고 있다.

V. 결론과 향후 연구

2차원 전기영동방법은 단백질 분석에 가장 필수적이며 널리 사용되는 방법임에도 불구하고 실험 과정에서 발생하는 수많은 변화로 인하여 재현성에 많은 문제점을 안고 있다. 이를 해결하기 위하여 다양한 디지털 영상처리 방법들이 기존에 연구되었지만 더욱 효

율적인 알고리즘에 대한 연구가 끊임없이 고찰되고 있는 실정이다.

본 논문에서는 2DGE 영상에 대해서 반복 점진적 과정을 통하여 두 Gel 영상의 반점들을 상호 정합하는 효과적이고 정교한 방법을 소개하였다. 제안하는 알고리즘의 첫 단계는 전역정합 과정으로 사용자에게 의하여 미리 지정된 기준반점들 사이의 기하학적인 차이를 줄이는 방향으로 모든 반점들을 이동 시킨다. 즉, 각 반점의 이동 방향 및 거리는 모든 기준반점들의 이동 방향 및 거리에 대한 가중치 합에 비례하며, 가중치는 해당 기준반점까지의 거리에 반비례한다. 이러한 기준반점 간의 오차를 근거로 각 반점을 이동시키는 과정은 기준반점들 간의 오차의 합이 일정한 값 이하에 도달할 때까지 반복적으로 지속된다.

기준반점을 기반으로 하여 1차적으로 수행한 전역적 정합의 결과는 일부 국부적 영역에 대한 불완전 정합을 야기한다. 이러한 문제를 해소하기 위하여 정합이 불완전한 영역에 추가적인 기준반점을 설정하고, 동일한 정합 방법을 적용한다. 이러한 국부적 정합의 결과는 정합의 성능을 훨씬 향상시켰으며, 이는 사용자 개입을 피할 수 없는 2DGE 영상 처리 및 반점 정합에 사용자의 개입에 좀 더 좋은 편리성을 제공할 수 있는 성공 요인으로 평가할 수 있을 것이다.

기준반점의 선정은 그 위치와 개수가 반점 정합의 성공 요인에 대한 핵심적인 요소이다. 더구나 현재와 같이 사용자의 관여가 필수적인 점은 정합 성능의 객관성 확보 및 일반화를 위하여 큰 걸림돌로 작용한다. 따라서, 향후에는 좀 더 지능적인 방법으로 기준반점을 설정하는 방법이 연구되어야 할 것이다.

표 1. 기준반점 수에 따른 정합 성공율

Table 1. Matching Rates with different number of landmarks

기준반점 수	17	11	9	7	5
정합도(%)	84.96	80.45	79.69	78.19	74.43

참고문헌

- [1] O'Farrell, "High resolution two-dimensional electrophoresis of proteins," *The Journal of Biological Chemistry*, vol. 250, pp. 4007-4021, 1975.
- [2] A. Roy, F. Seillier-Moiseiwitsch, K. R. Lee, Y. Hang, M. Marten, and B. Raman, "Analyzing two-dimensional gel images," *Technical Report*, 2003.
- [3] M. Rogers, J. Graham, and R. P. Tonge, "2 dimensional electrophoresis gel registration using point matching and local image-based refinement," in *Proc. the British Machine Vision conference*, Kingston University, London, 2004.
- [4] A. Efrat, F. Hoffmann, K. Kriegel, C. Schultz, and C. Wenk, "Geometric algorithms for the analysis of 2D-electrophoresis gels," presented at Annual Conference on Research in Computational Molecular Biology - *Proceedings of the fifth annual international conference on Computational biology*, Montreal, Quebec, Canada, 2001.
- [5] Pleibner, F. Hoffmann, 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, 1999.
- [6] Addario-Berry, "2D gel electrophoresis - an overview," 2002.
- [7] Ye, C. Y. Suen, M. Cheriet, and E. Wang, "A recent development in Image Analysis of Electrophoresis Gels," presented at Vision Interface, Trois - Rivières, Canada, 1999.
- [8] Kim, J. Kim, Y. Won, and Y. In, "Segmentation of protein spots in 2D gel electrophoresis images with watersheds using hierarchical threshold," *LNCS - Computer and Information Sciences - ISICIS 2003*, vol. 1, pp. 389-396, 2003.
- [9] Chen, T.S. Chen, T.H. Tsai, H.F. Tsai, M. Hsieh, and S. S. Tang, "Using the ACM technique to refine protein spots in 2DGE images," in *Proc. the 19th International Conference on Advanced Information Networking and Applications (AINA'05)*, pp. 256-261, 2005.
- [10] Chang, S.H., Cheng, F.H., Hsu, W.H., Wu, G.Z., "Fast algorithm for point pattern-matching: Invariant to translations, rotations and scale changes," *Journal of PR* 20, pp. 311-320, 1997.
- [11] J Panek and J Vohradsky, "Point pattern matching in the analysis of two-dimensional gel electropherograms," *Electrophoresis*, vol. 20, pp. 3483-3491, 1999.
- [12] A Rangarajan and J Duncan, "Matching point features using mutual information," in *Proc. the Workshop on Biomedical Image Analysis*, pp.172-181, 1998.
- [13] F Aurenhammer. Voronoi diagram 13m "A survey of a fundamental geometric data structure," *ACM Computing Surveys*, pp. 25-32, 1991.
- [14] A. Potra, X. Liu, F. Seillier-Moiseiwitsch, A. Roy, Y. Hang, M. R. Marten, B. Raman, and C. Whisnant, "Protein image alignment via piecewise affine transformations," *Journal of Computational Biology*, vol. 13, pp. 614-630, 2006.
- [15] Smilansky, "Automatic registration for images of two-dimensional protein gels," *Electrophoresis*, vol. 22, pp. 1616-1626, 2001.
- [16] P. F. Lemkin, G. C. Thornwall, and J. Evans, *Comparing 2-D Electrophoretic Gels Across Internet Databases: An Open Source Application*, The Proteomics Protocols Handbook, J. M. Walker, Ed. Totowa, NJ: Human Press Inc., 2005, pp. 279-306.
- [17] F. A. Potra, X. Liu, F. Seillier-Moiseiwitsch, A. Roy, Y. Hang, M. R. Marten, B. Raman, and C. Whisnant, "Protein image alignment via piecewise affine transformations," *Journal of Computational Biology*, vol. 13, pp. 614-630, 2006.
- [18] Z. Smilansky, "Automatic registration for images of two-dimensional protein gels," *Electrophoresis*, vol. 22, pp. 1616-1626, 2001.