

PCR Gel 전기영동 이미지 분석을 위한 레인검출 및 추적 알고리즘

이복주*, 문혁*, 박종훈+, 최영규*
*한국기술교육대학교 컴퓨터공학과, +(주)넷코덱
e-mail : bokju618@koreatech.ac.kr

Lane detection and tracking algorithm for PCR gel electrophoresis image analysis

Bok-ju Lee*, Hyuck Moon*, Jong-Hoon Park+, Young-Kyu Choi*
*School of Computer Science and Engineering, Korea University of
Technology and Education, +Netcodec Corp.

요 약

중합 효소 연쇄 반응 (PCR) 젤 전기영동 이미지에서 DNA 지문을 분석하기 위한 새로운 레인 검출 및 추적 알고리즘이 제안하였다. 이전에 여러 연구 결과가 보고되었지만 갑작스런 배경 밝기 차이와 구부러진 레인이 있는 이미지에서 레인을 정확하게 추출하는 것은 여전히 어려움이 있다. 우리는 평균 레인 폭과 레인 주기를 계산하기 위한 에지 기반 알고리즘을 제안한다. 본 논문에서 제안한 방법은 k-means 클러스터링 알고리즘을 이용하여 상승 에지와 하강 에지를 정확하게 추출하는 부화소 (sub-pixel) 알고리즘을 적용하여 레인 폭과 주기를 추정한다. 구부러진 레인을 처리하기 위해 젤 이미지를 정상영역과 비정상영역으로 분할하고, 각 분할 된 이미지의 레인 중심을 추적한다. 우리가 제안한 방법의 성능을 평가하기 위해 534 레인을 포함한 32 개의 젤 이미지가 사용되었다. 실험 결과는 우리의 방법이 전처리 과정 없이 배경 차이와 구부러진 레인을 갖는 이미지에 강인함을 보여 주었다.

1. 서론

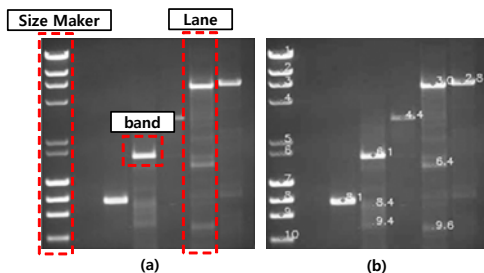
PCR(Polymerase Chain Reaction : 중합효소 연쇄 반응) 젤 전기영동(Electrophoresis)은 크기에 기초하여 생체 분자들을 분리하고, 효율적으로 정량적인 정보를 추출하는 가장 일반적인 방법 중의 하나이다. 생체분자는 젤 전기영동 실험 동안에 특정 효소의 작용에 의해 다수의 DNA 단편들로 분해될 수 있다. 아가로스 젤에 전류를 흘려주면, DNA 단편들이 각각 다른 속도로 분산되고, 시간이 지나면 젤이 염색된다[1]. 그림 1에서 보이는 밴드들의 수직 나열 모음을 레인(lane)이라 부르고 검사대상 물질들을 구별하는 기준이 된다. 그리고 레인 내에서 보이는 작은 사각형 모양의 수평 밴드(band)들의 위치는 검사 대상 물질을 구성하는 분자들의 이동거리 값을 표현하고, 수평 밴드의 넓이는 이동된 구성 분자들의 축적량을 나타낸다.

염색된 젤은 디지털 영상으로 기록되어 많은 분자 생물학 응용프로그램에서 생물학적 물질로부터 다양한 정보를 추출하는데 사용된다. 이때, 사람의 육안을 통한 영상의 수동적인 평가는 대량의 반복적인 전기영동 젤 분석 작업을 위해 많은 시간이 소요된다. 또한, 평가자의 개인 경험에 의존하기 때문에 단조롭고, 주관적이고, 오류를 발생하기 쉽다. 이에 따라, 부정확한 육안 검사의 문제점을 해결하기 위한 다양한 자동화 방법들이 연구되고 있다.

2. 관련연구

자동화의 가장 중요한 단계는 레인을 정확하게 식별하는 것이다. 레인의 식별을 위해 영상에서 레인을 정확히 검출하는 레인 검출(lane detection)이 반드시 필요하다[2]. 또한, 휘어져 있는 레인들의 정확한 검출을 위해 레인의 중심을 추적하는 레인추적(lane tracking)단계도 필요하다.

그림 1에서 볼 수 있듯이 젤 영상에 있는 레인들은 일반적으로 폭과 간격이 일정하고 밴드와 배경간의 밝기에 어느 정도 차이가 있는 특성을 가지고 있다. 이 특성을 이용한 프로젝션 프로파일 기반의 방법들이 연구되고 있는데, 레인의 중심축만을 추출하며, 다양한 영상처리에 의해 밴드 밝기에 왜곡이 생길 가능성이 크다는 문제점을 가지고 있다[4-5]. Park 등은 휘어져 있는 레인을 검출하기 위해, 전체 영상의 수직 프로젝션 프로파일 계산을 통해 평



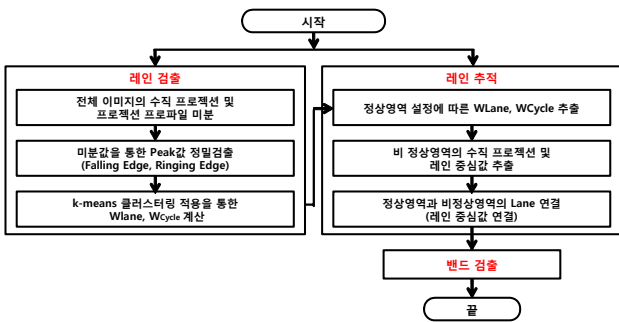
(그림 1) (a)전기영동 작업이 완료된 이미지의 예, (b) 분석 결과의 예

균 라인 폭을 추정하고, 이 값을 이용하여 레인을 추적한다[3]. 그러나 이 방법은 배경영역이 밝기차이를 가지는 영상에서는 정확한 결과를 갖기 어렵다는 단점을 가지고 있다.

본 논문에서는 PCR 전기영동 영상에서 자동으로 레인을 안정적으로 검출하고 중심을 추적하는 방법을 제안한다. 특히, 에지 기반의 알고리즘을 통해 배경영역의 밝기 차이가 있어도 안정적으로 중심을 추출할 수 있으며, 부화소 알고리즘을 사용하여 보다 정밀한 계산이 가능하며 휘어진 레인에 대해서도 잘 대응할 수 있도록 하였다.

3. 제안된 레인검출 알고리즘

그림 2는 제안된 레인 검출 및 추적 알고리즘을 보여주고 있다. 레인 검출을 위해서는 입력 영상에서 평균 라인 폭(W_{Lane})과 주기(W_{Cycle})를 정확히 검출해야 한다.



(그림 2) 제안된 알고리즘의 Flow Chart

3.1. 수직 프로젝션 및 미분

수직 프로젝션은 각 열의 화소 값을 누적하여, 2차원 정보를 1차원으로 변환한다. 그림 3 (b)는 이 영상의 수직 프로젝션 프로파일을 보여주는데, 값들이 급격하게 증가하거나 감소하는 부분들이 반복적으로 나타난다. 이것은 영상에서 밝기 변화가 큰 에지(edge) 영역에 해당하는데, 이들이 레인의 시작과 끝 정보를 나타내고, 이들 사이의 거리를 이용하여 레인 폭과 주기를 추정할 수 있다. 프로젝션 프로파일에서 에지의 위치를 추출하기 위해 프로젝션 프로파일의 미분을 구한다. 그림 3 (c)는 미분 결과를 보여주고 있다.

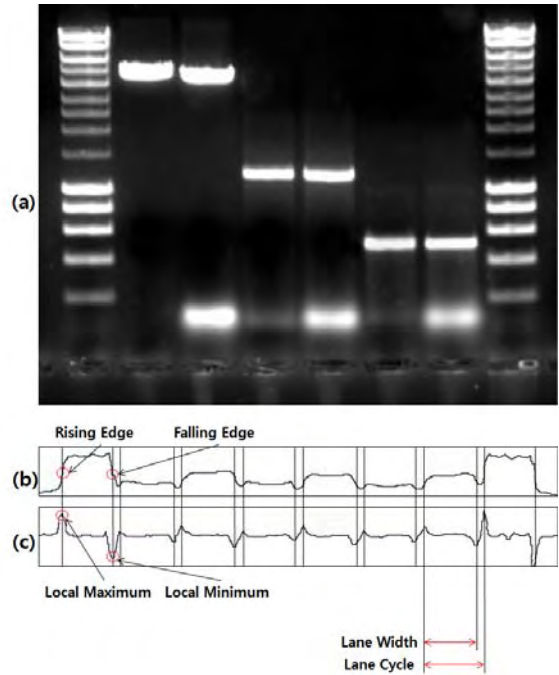
3.2. 부화소 알고리즘을 통한 에지 위치의 정밀 검출

미분 프로파일이 만들어지면 에지 위치를 검출해야 한다. 에지는 미분 프로파일에서 국부 최대치와 최소치(local maximum, minimum)로 나타나는데 이를 피크(peak)라 하자. 그림 3을 보면 국부 최대치는 레인의 시작 에지이고, 국부 최소치는 레인이 끝나는 부분이다. 미분 프로파일의 피크 위치는 정수값으로 나타난다. 따라서 보다 정밀도를 향상시키기 위해 보간법(interpolation)을 사용하였다. 이것은 부화소(sub-pixel) 처리라고 하는데, 화소 단위의 정밀

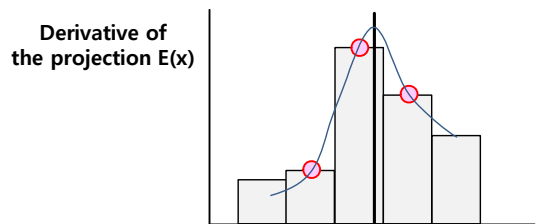
도에 비해 5배 이상 정밀한 것으로 알려져 있다.

그림 4는 이러한 부화소 처리의 개념을 보여주고 있다. E(i)가 최대값이라고 할 때, 실제의 최대값은 i가 아니라 i와 i+1사이에 존재하는데 샘플링에 의해 이 위치가 사라진 것으로 생각할 수 있다. 따라서 다음과 같이 더 정밀한 최대값의 위치를 계산할 수 있다.

$$i_{\max} = i - \frac{E(i+1) - E(i-1)}{4E(i) - 2E(i+1) - 2E(i-1)} \quad (1)$$



(그림 3) 하나의 영상에 대한 레인검출 과정 1, (a) 원래 영상, (b) 수직 프로젝션 프로파일, (c) 프로젝션 미분



(그림 4) 보간법의 개념

3.3. K-means 알고리즘을 통한 레인 폭과 주기 계산

에지 위치가 정밀하게 검출되면 이들을 이용해 평균 라인 폭과 주기를 계산해야 한다. 검출된 에지 위치는 다양한 잡음을 포함하고 있다. 누락 레인을 포함하는 영상이나, 잡음에 의해 불필요한 에지가 발생할 수 있다. 따라서 안정적인 레인의 폭과 주기의 계산을 위해, 우리는 k-means 알고리즘을 사용한다. 상승 에지와 하강 에지의 쌍들을 모두 찾아 각각의 길이를 구하여, k-means 알고리즘에 적용하고 가장 큰 그룹의 평균값을 평균 라인 폭으로 간주한다.

4. 제안된 레인추적 알고리즘

전기영동 과정을 통해서 젤 영상을 추출할 때, 젤의 온도, 젤의 농도, 전압, 완충용액의 조성, 공기와의 노출 여부 등 많은 요인들에 의해 영상의 추출을 방해하기 때문에 휘어진 레인을 포함하는 영상을 추출하기도 하며, 자동화 시스템에서는 이와 같은 영상에 대해서도 정확한 분석 결과를 도출해야 한다. 하지만 앞의 3절에서 제안한 방법으로 구한 평균 레인 폭과 주기만을 이용하여 모든 영역의 레인을 정확히 분리하기는 어렵다. 따라서 이러한 영상에서는 앞에서 구한 레인 정보를 바탕으로 레인의 중심을 추적할 필요가 있으며, 이를 레인 추적(Lane Tracking)이라고 한다.



(그림 6) 정상영역 설정에 따른, 탐색 구간의 설정

4.1. 정상영역 설정

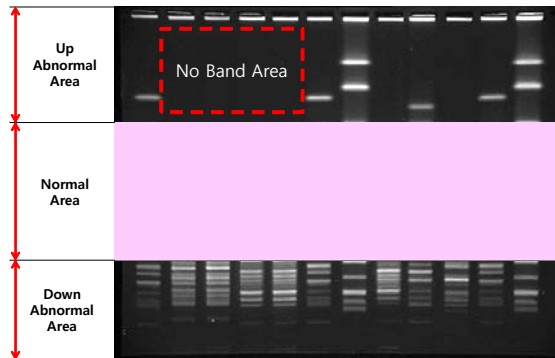
레인 추적은 먼저 영상을 정상영역과 비정상영역으로 분리하는 것으로 시작한다. 정상영역을 선택하는 과정은 자동화과정과 반자동화과정의 두 가지로 나뉠 수 있다. 본 논문에서는 용자로 하여금 정상영역을 지정하는 반자동화 과정을 선택하였다. 정상영역은 중심이 휘어져 있지 않고, 밴드가 집중되어 있는 지역을 선택하는 것이 바람직하다. 정상영역을 설정되면, 앞의 3절에서 설명한 방법을 통해, 정상영역의 레인 폭과 주기를 추출하고, 각 레인의 피크(에지위치)를 저장하여, 초기 레인을 확정한다.

정상영역 설정을 통해 정상영역의 평균 레인 폭 (W_{Lane})과 각 레인의 시작위치(x_1), 끝 위치(x_2)가 구해지면, 비정상영역의 피크 탐색 구간의 시작 위치(x'_1)와, 끝 위치(x'_2)를 다음과 같이 정할 수 있다.

$$\begin{aligned} x'_1 &= x_1 - W_{Lane} \times \alpha, \\ x'_2 &= x_2 + W_{Lane} \times \alpha, \end{aligned} \quad (2)$$

실험에서는, 파라미터 α 의 값을 0.5로 두었다. 즉, 정상영역 각 레인의 상승-하강 피크를 기준으로 좌측으로 정상 영역 평균 레인 폭의 절반만큼, 우측으로 정상 영역 평균 레인 폭의 절반만큼의 영역을 탐색한다. 상승 피크 구간에서는 상승 피크만을 검색하고, 하강 피크 구간에서는 하강 피크만을 검색하여, 상승-하강 피크 쌍을 구하고, 구해진 상승-하강 피크의 중심위치를 비정상영역의 레인 중심값으로 설정한다. 이 과정을 반복하여 영상의 모든 비정상영역들을 처리한다. 이때, 비정상영역 안에 흰 밴드 영역이 있을 수도 있고, 없을 수도 있다. 밴드가 존재하는 경우에는 문제가 없지만, 밴드가 존재하지 않는 경우에는 프로젝션 프로파일을 수행하더라도 상승-하강 피크 쌍이 나타나지 않는다. 이러한 경우에는 앞에서 구한 정상영역의 레인 정보를 유지하도록 한다.

4.2. 비정상영역의 레인 정보 추출



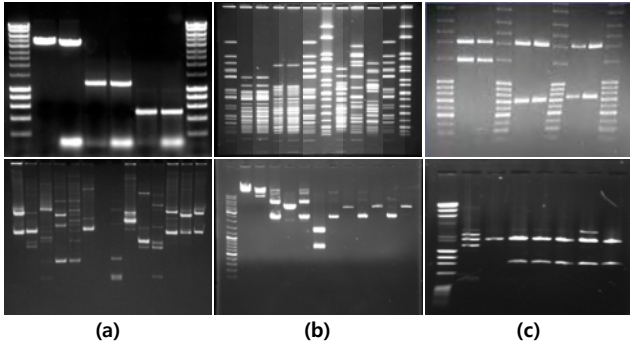
(그림 5) 정상영역 설정에 따른 위, 아래쪽 비정상영역의 설정

정상 영역이 선택되면 그림 5처럼 정상 영역을 기준으로 위쪽 비정상영역과 아래쪽 비정상 영역으로 분리되고, 위, 아래쪽의 비정상영역에서의 레인 중심값을 갱신한다. 이를 위해 위아래 각각의 비정상 영역에 대한 프로젝션 프로파일을 구하고, 미분 프로파일을 구한 후 피크를 찾는다. 이때, 상승 피크와 하강 피크를 구분해야 한다. 다음으로 인접한 정상 영역에서의 상승, 하강 피크 쌍과 비교하여 레인의 이동 정도와 레인 중심을 계산한다. 이때, 앞에서 구한 정상영역의 각 레인의 정보와 비교하기 위한 피크 탐색 구간의 설정은 그림 6과 같다.

5 실험결과 및 분석

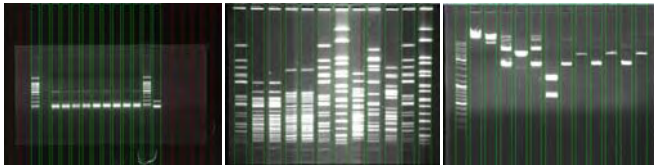
그림 7은 실험을 위해서 사용된 영상들을 보여주고 있다. 실험에서는 534개의 레인을 가지는 32개의 PCR Gel 영상을 사용하였다. 전체 영상은 기본적으로 검은 배경에 흰색 객체를 가지는 영상을 선택하여 실험하였다. 일부 젤 영상은 그림 7 (b)와 같이 배경영역의 밝기가 부분적으로 변하는 경우가 있다. 이런 영상을 처리하기 위해서 배경제거 기법을 사용하는 전처리를 사용할 수도 있지만, 제안된 방법에서는 에지 기반 알고리즘을 통해 전처리 없이도 안정적으로 레인을 검출한다. 또한, 레인 추적의 성능을 평가하기 위해 그림 7 (c)와 같이 일부 젤 영상은 휘어져 레인을 포함하였다. 전체 실험 영상들에는 배경영역이 밝기

차이를 가지는 10개의 영상과 휘어져 있는 레인을 가지는 6개의 영상이 포함되었다.



(그림 7) 실험 영상의 예, (a) 정상 영상, (b) 배경에 밝기 차이가 있는 영상, (c) 구부러진 레인을 포함하는 영상

그림 8은 배경영역이 밝기차이를 가지는 PCR Gel 영상에 대한 레인 자동 검출결과를 보여주고 있다. 그림 8의 영상은 모두 배경영역이 밝기차이를 가지고 있는 영상이지만, 제안된 방법은 이러한 영상에 대해서도 안정적으로 레인을 검출하는 것을 알 수 있다. 또한, 사용자의 간섭 없이도 자동적으로 레인을 검출한다.



(그림 8) 레인 자동검출 결과

그림 9는 구부러진 레인을 포함하는 PCR Gel 영상에 대한 레인 추적 결과를 보여주고 있다. 휘어져 있는 레인을 포함하는 영상은 레인의 구부러짐 현상이 영상의 어느 부분에서 나타날지 예측할 수 없기 때문에 사용자로 하여금 정상영역을 선택하게 한다. 그림 8 (b)처럼 정상영역을 설정하게 되면 위, 아래의 비정상영역의 레인중심을 찾아가며, 레인을 안정적으로 추적하는 것을 확인할 수 있다

표 1은 실험에 사용된 영상들의 레인 검출 및 레인 추적, 밴드검출 성공 비율을 보여주고 있다. 16개의 Normal 영상에 대해서는 97%의 확률로 레인 검출을 성공하였으며, 10개의 배경영역의 밝기차이가 존재하는 영상에 대해서도 94%의 확률로 안정적으로 레인을 검출하고, 휘어져 있는 레인을 포함하는 6개 영상의 레인 추적은 90%의 확률로 레인을 검출하는 것으로 나타났다. 밴드검출에 대해서는 95%의 확률로 밴드를 검출하는 것으로 나타났다. 입력 영상의 레인 개수가 너무 적거나, 레인의 밴드영역이 너무 적은 경우에는 상승-하강 에지 계산을 위한 데이터가 충분하지 않기 때문에 레인검출 및 추적에 오류가 발생하였다.

(그림 9) (a) 구부러진 레인의 레인검출 실패, (b) 정상영역의 설정, (c) 레인 추적 결과

<표 1> .레인검출, 레인추적, 밴드검출 성공비율

구 분		전체 개수	검출 개수	비 율
레인 검출	정상 영상	276	268	97.0%
	배경에 밝기차이를 가지는 영상	177	167	94.3%
레인 추적	레인 추적 전	276	268	97.0%
	레인 추적 후	276	274	99.2%

이와 같은 실험을 통해 제안된 방법이 PGR Gel 영상에서 안정적으로 레인을 검출하고, 추적하는 매우 효과적인 기법임을 확인할 수 있었다.

참고문헌

[1] Machado, Alexei, et al. "An iterative algorithm for segmenting lanes in gel electrophoresis images." Computer Graphics and Image Processing, 1997. Proceedings., X Brazilian Symposium on. IEEE, (1997).

[2] Wong, Richard TF, et al. "LaneRuler: automated lane tracking for dna electrophoresis gel images." IEEE transactions on automation science and engineering, 7.3, pp. 706-708, (2010).

[3] Park, Sang Cheol, et al. "Lane detection and tracking in PCR gel electrophoresis images." Computers and electronics in agriculture, 83, pp. 85-91, (2012).

[12] Kim Seung Il, et al. "Automatic Alignment of Electrophoresis Gel Images Based on Standard Deviation of Vertical Profiles." Journal of KISS Software and Applications, 39.8, pp. 631-638, (2012).

[13] Lee, Wan Yeon, et al. " Automatic Analysis Scheme for Multiple Images of Ongoing Electrophoresis Gel." Journal of KISS Software and Applications, 39.8, pp. 672-677, (2012)