

염색체 마디 추출을 위한 최적타원 추정기법

o 이상렬, 이권순, *전계록, 장용훈, 임상희
(동아대학교 전기공학과, *부산대학교 의공학과)

The Optimal Ellipse Estimation Method for Chromosome Bands Extraction

o Sang-Yeol Lee, Kwon-Soon Lee, *Gye-Rok Jeon, Yong-Hoon Chang
(Dept. of Electrical Eng., Dong-A Univ., *Dept. of Biomedical Eng., Pusan National Univ.)

Abstract

This paper attempts to examine an optimal method for the chromosome specific vector extraction. Usually, represented method are used with a line segmentation on a chromosome image. It is not inaccurate but also needs a long time for the analysis. This paper purpose to acquire specific vector in the image with a using optimal ellipse estimation method.

Normally, shapes of chromosomes are curved and too difficult to analyze automatically. A chromosome has a lot of band which looks like an ellipse. If we can estimate their bands with an ellipse, we can reconstruct the sample which is straight and can be analyzed easily. We have rearranged a chromosome image with above proposed. Result shows a reconstructed sample which is simple for chromosome analysis.

서론

염색체의 핵형분석(karyotype)은 태아의 유전학적 진단을 위해 현재 임상에서 널리 사용하고 있다. 분석과정에 필요한 작업과 그에 따르는 소요시간이 많아 자동화의 필요성이 대두되었다. 컴퓨터를 이용한 자동화에 관한 연구가 1964년 Ledly에 의해 시작되었고, 자동분류의 결과는 임상적으로 사용이 가능하다고 확인되었다[1]. 그 이후에 많은 연구가 진행되었으나 전체 과정에 대한 자동화는 이루어지지 않고 있으며 사용자의 개입을 필요로 하는 요소가 많고, 정확도가 낮다. 그 예로 1990년 Young, Mayall 및 Delft 그룹의 Lucas 등은 Athena라는 염색체 해석 시스템을 발표하였는데, 하나의 염색체군 처리에 소요된 9분 중에서 7분은 사용자에 의한 것이었다. 또한, 정확도가 4.0에서 11.5%에 불과하여 임상에 사용하기는 어려웠다[2][3]. 분류에 대한 정확도가 점차 높아지고는 있으나 현재까지도 개선의 여지가 많이 남아있다.

염색체의 분류를 위해서는 각 염색체의 동원체 지수(centromeric Index), 염색체의 상대적 길이비율 및 염색체 마디에 대한 농도치에 대한 정보가 필요하다[4]. 기존의 방법들은 이들 정보를 얻기 위해 취득된 염색체 영상에서 중심축을 구하고 축을 따라 일정간격으로 일정한 갯수만큼 계속되는 수직선을 이용하여 폭과 농도치를 구한다[5]. 그 후에 폭과 농도치 각각에 대한 프로파일(profile)을 구성하고 이를 통해 자동분류를 하는 것이다. 이 방법에서 염색체 영상의 중심축에 대한 수직선을 규정할 때 상호교차가 되지 않고 최대한 전체 영상을 모두 지나는 수직선들의 집합이어야 한다는 조건에 만족해야만 분

류정보의 정확성이 높아진다. 하지만, 보통의 염색체는 직선 형태가 아니고 구부러진 형태이기 때문에 기하학적으로 이 두가지 조건은 상호 배타적이다. 정확도를 높이기 위해 어느 정도의 손실본은 감수하고, 최대한 상호교차하지 않는 수직선을 이용하여 그 결과를 신경 회로망을 통해 학습시켜 분류하는 방법이 제시되었다[6]. 이 방법은 신경 회로망에 입력되는 입력 특징벡터의 수가 많아 계산에 소요되는 시간이 많고, 특징벡터 추출시에 손실본이 많아 정확도를 향상 시킬 수 없다.

염색체 각각은 타원형태의 염색체 마디를 가지고 있다. 이에 근접하는 타원을 계산해내고 이를 이용하여 염색체 샘플을 일직선 형태로 재구성하면 기존의 방법보다 정확한 접근이 가능하다. 재구성에 염색체 마디의 농도치에 대한 정보를 농도치 총합으로 부터 구하거나, FFT(Fast Fourier Transform)를 통해 얻은 주파수 특성으로 부터 구하면 더욱 정확한 분석이 가능하리라고 예상한다. 현재 이에 대한 부분은 따로 진행 중이라 논외로 하고, 본 논문에서는 최적타원 추정기법을 이용하여 염색체 샘플을 재구성하는 것을 목적으로 하였다.

본 논문에서는 염색체 샘플로 재구성하기 위한 윤곽선 검출(edge extraction)기법, 윤곽선 연결(edge linking)기법, 그리고 최적타원 추정기법을 다룬다.

본론

1. 처리과정

본 논문에서의 염색체 분류를 위한 특징벡터 추출 과정은 다음과 같다.

- ① 윤곽선 검출기법을 이용하여 염색체의 윤곽선 마디가 명확하게 구분되도록 한다.
- ② 검출된 염색체 마디의 윤곽선을 이용하여 최적타원을 추정, 폭과 넓이를 구한다.
- ③ 검출된 염색체 마디의 영역안에 있는 픽셀들을 이용하여 농도치 총합을 구한다.
- ④ 얻어진 결과를 이용하여 염색체를 재구성하고, 특징벡터를 구한다.

1.1. 윤곽선 검출기법

윤곽선 검출에 관하여는 상당히 많은 연구가 이루어져 있다. 윤곽선 검출 방법으로는 경사(gradient)를 이용하는 방법, 픽셀 공간에서의 명암도를 면함수로 보고 적당한 면함수에 적합시키는 방법(surface fitting), 2계 미분에 의한 방법(second-

derivative methods), 그리고 색상을 이용하는 방법 등이 있다. 전체적인 처리속도 증가를 위해 가장 간단한 경사를 이용하여 윤곽선 검출을 한다. 이는 경사가 어떤 역채, 즉 Sobel과 Robert 등의 윤곽선 연산자를 초과하는지의 여부에 달려 있다 [7][8][9]. 먼저, 검색채 마디부분을 관심영역(ROI)으로 하여 히스토그램 조작(histogram operation)을 하고 식(1)의 연산을 통하여 잡음이 제거된 영상을 얻는다.

$$f = (X - B) + B \quad (1)$$

여기서 X 는 원 영상이며 B 는 형틀(template)이다. 잡음을 제거한 영상에 윤곽선 검출 알고리즘을 사용하여 검색채 마디의 윤곽선을 검출하였다.

1.2. 최적타원 추정

1.2.1. 초기중심 자동검출 기법

검색채 마디내의 초기중심을 자동으로 검출하기 위하여 먼저 원 영상을 부표본화하고, 부표본화된 영상의 매 화소에 초기중심 추정 기법을 적용한다[10]. 이는 원 영상의 매 화소에 초기중심 기법을 적용할 경우에 연산시간이 오래 걸리는 단점을 극복하면서 동시에 계산결과에 큰 오차를 유발하지 않도록 하기 위함이다. 초기중심을 찾기 위해 사용되는 초기중심 추정 기법은 다음과 같다.

- ① 배열 $count(i, j)$ 와 $area(i, j)$ 의 모든 요소를 0으로 초기화 한다.
- ② 원점을 화소 (ic, jc) 로 둔다.
- ③ 원점에서 $\Delta\theta = 2\pi/M$ 의 등각도로 M 개의 반경을 계산한다.
- ④ 반경 i 에 대해 다음의 식을 이용하여 $redge$ 값을 계산한다.

$$redge = \frac{g(r^i + 1) + g(r^i) - g(r^i - 1) - g(r^i - 2)}{2} \quad (2)$$

여기서 r^i 는 반경 i 에 대한 방사방향의 증분값이고, $g()$ 는 각각의 화소에서의 그레이 레벨치이다.

⑤ 반경 i 에 따른 $redge$ 값이 문턱치(Rth) 보다 더 클 경우, 반경 i 에 대한 진행을 멈추고 $count(i, j)$ 의 값을 1 증가시키고, $area(i, j)$ 의 값에 원점으로 부터의 거리 li 를 더한다. 즉,

$$\begin{aligned} count(i, j) &= count(ic, jc) + 1 \quad \text{if } redge > Rth \\ area(i, j) &= area(ic, jc) + li \end{aligned} \quad (3)$$

- ⑥ 모든 M 개의 반경에 대하여 ④ - ⑤를 반복 한다.
- ⑦ 모든 화소에 대하여 ② - ⑥을 반복한다.
- ⑧ $count(ic, jc)$ 를 3×3 마스크로써 가중평균을 구한다.
- ⑨ $area(ic, jc)$ 에 문턱치를 취하여 초기 중심을 결정한다.

위에서 단계 ⑧은 $count$ 배열에서 임펄스적인 값을 제거한다.

1.2.2. 최적 타원모델 추정 기법

Friendland 등은 수직된 타원모델만을 추정하였으나, 이 경우 검색채가 구부러져 있거나 기울어진 영상에 대해서는 모델의 오차가 커지게 된다. 실제 검색채 윤곽선 검출시 타원모델로 부터의 거리가 멀어짐에 따라 그 화소가 윤곽선으로 결정될 확률이 감소하도록 코스트 함수를 정의하기 때문에 모델 오차의 증감은 윤곽선 오차의 증감에 직결된다. 주어진 영상에 대한 수학적 모델을 찾아내는 가장 좋은 방법으로는 Hough 변환이 알려져

있고 최적 타원모델 추정이 Ballard에 의해 수행되었다[11]. 그러나, 이 모델 정합방법은 기억용량이 커야하고 시간이 오래 걸리는 단점이 있을 뿐 아니라 인식 대상의 정확한 규격이 요구되며 결정적인 모양에 대해서만 적용이 가능하다. 따라서 경계가 뚜렷하지 못하여 정확한 윤곽을 검출할 수 없는 경우에는 적용할 수 없다.

중심추정 기법에서 구한 M 개의 화소를 사용하여 총 $N/2$ 개의 축 길이 LX_k 를 계산한다.

$$LX_k^2 = (x_k - x_{k+32})^2 + (y_k - y_{k+32})^2$$

$$k = 0, \dots, N/2 - 1$$

그 중에서 최대값 3개를 선택하여 각각에 대하여 장축, 단축의 길이, 중심점, 그리고 각도를 계산한다. 먼저 선택된 LX_k 의 1/2을 장축 길이로, 이 축에 수직하는 축 길이의 1/2을 단축 길이로 정한다. 그리고 장축의 1/2지점을 타원의 중심으로 정한다. 그 다음 $LX_{N/4}$ 과 LX_k 사이의 각을 계산하면 타원의 기울어진 각도 θ 가 된다. 본 연구에서 제안한 방법은 이미 검색채의 윤곽선을 검출하는 방법을 전처리 과정에서 거치기 때문에 이 방법의 적용이 가능하다.

1.3. 검색채 마디별 농도치 총합 검출

윤곽선 검출을 거친 검색채 마디의 내부에서 모든 픽셀들의 농도치 총합을 구한다. 이때, 마디내부에서의 좌표이동은 다각형 채움기법의 변형으로서 y 축 방향으로 한 줄씩 채워나가는 스캔변환(scan conversion)기법을 이용하였다. 스캔변환 기법은 다각형이 있을 때, 다각형이 차지하는 최소 y 값과 최대 y 값 사이의 수평선을 하나씩 살펴가며 유효영역을 결정하는 기법이다 [12].

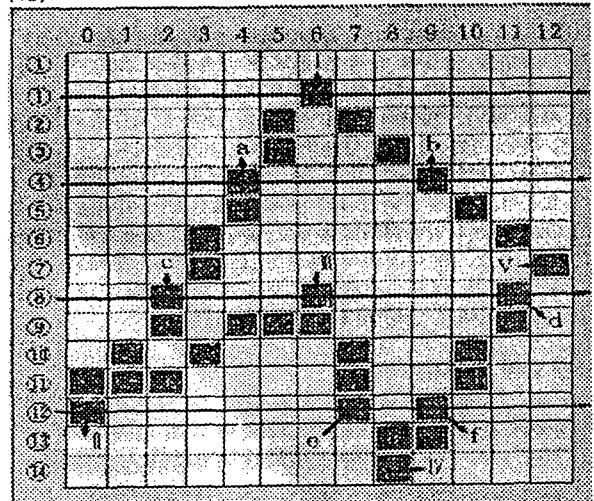


그림 1. 스캔변환 기법
Fig 1. Scan conversion method

1.4. 검색채 샘플 재구성

검색채 마디 각각에 대한 최적타원과 농도정보를 이용하여 검색채 마디에 관한 정보, 즉 각 검색채 마디의 중심점과 장·단축의 길이, 그리고 농도치 총합을 이용하여 검색채 샘플을 재구성한다. 고정된 x 좌표에 최적타원의 중심점을 위치시키고 y 축을 따라 각각의 검색채 마디의 단축을 기준으로 하여 배치시킨다. 그리고 각각의 내부에 스캔변환 기법을 이용하여 해당 농도치 총합을 농도 특성값으로 별도로 처리한다.

실험결과 및 고찰

결론

현미경으로 부터 해상도가 512x512인 CCD 카메라를 통해 전달된 염색체의 영상을 Frame grabber(Data Translation, DT-2855)를 거쳐 PC에서 취득하고, Windows Bitmap 화일 형태로 저장했다. 그 후에 1번 염색체를 디지털타이저로 배경에서 샘플을 분리함으로써 실험에 사용하였다. 이 과정에서 마우스를 이용해도 가능하나 정확성을 높이기 위해서 디지털타이저를 이용하였다. 본 논문에서 제안한 방법의 과정은 크게 세 가지로 구분된다. 첫째, 염색체 마디의 윤곽선을 추출하고, 이를 이용하여 최적타원을 계산한다. 둘째, 추출된 염색체 마디의 영역내부의 농도치 총합을 스캔변환 기법을 이용하여 구한다. 셋째, 염색체 샘플을 계산된 최적타원과 농도치 총합을 이용하여 재구성하는 과정이다.

본 논문에서 제안한 방법의 타당성을 검증하기 위하여 기존의 특징벡터 추출방법을 그림 2에 나타내었고, 샘플 중에서 1번 염색체로 분류된 것을 선택하여 최적타원 추정기법에 적용한 실험결과를 그림 3에 보였다. 그림 3(a)에서는 샘플로 사용한 염색체 1번의 영상을 보였으며, 그림 3(b)에서는 염색체 마디의 윤곽선을 추출한 결과를 나타냈고, 그림 3(c)에서는 염색체 마디에 대한 최적타원 추정의 결과, 그리고 그림 3(d)에서는 염색체 샘플을 재구성한 결과를 보였다.

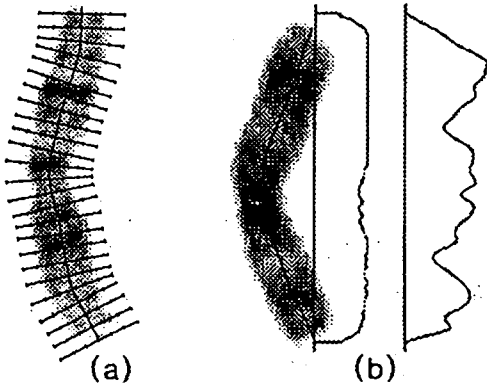


그림 2. 기존의 특징벡터 추출방법
Fig 2. Conventional method

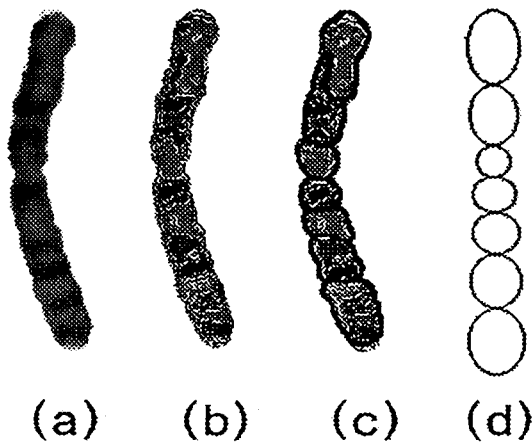


그림 3. 1번 염색체에 대한 제안된 방법의 결과 (a) 원 영상 (b) 윤곽선 추출결과 (c) 최적타원 추정 결과 (d) 염색체 샘플 재구성

Fig 3. Result of proposed method for 1st chromosome (a) an original image (b) the result of edge extraction (c) the result of optimal ellipse estimation

본 논문에서는 염색체 마디의 윤곽선을 검출하고 이와 근접하는 최적타원을 계산한 후, 염색체 마디의 농도치 총합을 이용하여 염색체 샘플을 재구성하는 방법을 제안하였다. 기존의 방법이 기하학적으로 기울어지거나 굽어진 염색체의 특징벡터 추출이 복잡하고 정확하지 않은 반면, 본 논문에서 제안한 최적타원 추정기법은 염색체 마디를 재구성 하여 염색체를 일직선 형태로 나타낼 수 있었으며, 각 염색체 마디의 농도특성을 구할 수 있음을 보였다.

참고 문헌

- [1] Robert S. Ledly, "High-speed automatic analysis of biomedical pictures," *Science*, vol. 146, pp. 216-223, 1964.
- [2] Lucas J. van Vliet, Ian T. Young, and Brian H. Mayall, "The Athena semi-automated karyotyping system," *Cytometry*, vol. 11, pp. 51-58, 1990.
- [3] Brian H. Mayall, James D. Tucker, Marl L. Christensen, Lucas J. van Vliet, and Ian T. Young, "Experience with the Athena semi-automated karyotyping system," *Cytometry*, vol. 11, pp. 59-72.
- [4] David G. Harnden, B.Sc., Ph.D., F.R.C. Path. Harold P. Klinger, Md.D., Ph.D. "An International System for Human Cytogenetic Nomenclature," *KARGER*, pp. 114-115, 1985.
- [5] John R. Gosden, "Chromosome Analysis Protocols," *Humana Methods in Molecular Biology*, Vol. 29, pp. 164-173, 1994.
- [6] 조종만, '신경회로망을 이용한 염색체의 핵형분류에 관한 연구,' 인하대학교 전자공학과 박사 학위논문, pp. 103-125, 1994.
- [7] 김희승, '영상인식,' 생능 출판사, pp. 133-159, 1993.
- [8] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing," *Addison Wesley*, pp. 416-423, 1992.
- [9] Dwayne Phillips, "Image Processing in C," *R&D Tech. Books*, pp. 165-250, 1994.
- [10] 김명남, 조진호, '가중모델 Hough 변환을 이용한 2D 심초음파도에서의 좌심실 윤곽선 자동검출,' 의공학회지 제 15권, 3호, 1994.
- [11] Ioannis Pitas, "Digital Image Processing Algorithms," *Prentice Hall*, pp. 231-239, 1993.
- [12] 임채성, 김응석, '효과적인 래스터라이저와 윤곽선 폰트 개발,' 마이크로 소프트웨어, 10월호, pp. 199-206, 1993.