

통계적 산출방법을 이용한 염색체 위치 탐색

송 준 영 ° 김 주 병 윤 영 로 이 윤 선
연세대학교 보건과학대학 의용전자공학과

Searching Location of Chromosome Using Statistical Method

J.Y. Song ° J.B. Kim Y.R. Yoon Y.S. Lee

Dept of Biomedical Eng., Yonsei Univ

Abstract

In this paper, we classify between the chromosome and blood cell, and find the location of chromosome. First, the gray level images be the binary images using the threshold method. Then, the spot noises are removed by the morphological filtering. Features are obtained using the updated Run length(RL) coding and are classified using the Bayes decision rule. The performances of classification are 83.3% in chromosome and 93.3% in blood cell. Because each sub-images(256×256) is obtained from the full image(512×512), we realize the location of chromosome if we get the corrected chromosome classifications.

I. 서론

염색체분석은 여러가지 유전질환을 분석하는 세포 유전학중에 하나이다. 분석 작업은 중기 찾기(metaphase finding)와 karyotyping등이 있다. 근래에 들어 염색체 분석의 자동화된 시스템이 등장하게 됨으로써 염색체 분석 작업을 자동화하였다.[1][2] 염색체 인식에는 목적에 따라 여러가지 방법을 이용하고 있다. 띠 패턴(band pattern)을 분석하여 염색체의 centromere를 인식하는 방법이 있으며[3][4] 또한 여러가지 경계선 검출 알고리즘을 적용해서 formal linguistic method를 사용하고 있는 등의[5] 여러가지 방법을 사용하고 있는데 이러한 방법들은 염색체 분석을 위한 연구이다.

*이 연구는 한국전력공사 원자력 발전처 및 한일병원의 연구비 지원을 받아 이루어진것임

본 연구에서는 염색체 분석에 적합하도록 염색체의 위치(site)를 판별하는 것이 주목적이다. 저배율에서 더 많은 범위의 영상을 취할 수 있기 때문에 저배율(400배)에서 512×512영상을 획득하고 이 영상에서 다시 중첩되는 9개의 256×256 부영상(subimage)을 획득하여 실험하였다. 각각의 9개의 부영상은 저배율에서 얻었기 때문에 염색체의 크기가 작고 영상이 뚜렷하지 않아 경계선을 검출한다는 것은 어렵다. 따라서 9개의 gray scale 부영상을 모두 2진 영상으로 변환한후 고주파 성분의 점 잡음들을 제거하기 위해 morphological filtering을 사용했다. 그 다음 RL coding(Run Length coding)을 응용한 새로운 coding방법으로 염색체와 혈구의 feature를 얻은후 Bayes decision rule을 적용해서 얻은 문턱치(decision threshold)를 사용하여 염색체와 혈구를 구별한 후 염색체가 발견된 부영상이 512×512영상에서의 염색체의 위치가 된다. 본 연구에서 사용하고 있는 영상 개선 방법은 contrast stretching을[6] 이용하였고 threshold 기법을 이용하여 2진 영상을 얻었다.[7][8]

II. 시스템구성

그림1의 시스템 구성도에 대한 설명은 다음과 같다. 현미경(Olympus BX40)에서 얻는 염색체 영상은 배율 400배에서 CCD camera (Toshiba IK-637K)로 얻었다. 이 영상신호를 영상 획득 보드(Image signal overlay card)를 이용하여 512x512영상으로 디지털화하였다. 영상 획득 보드는 PC(pentium)와 인터페이스되어 있으며 프로그램은 윈도우즈 환경하에서 구현했다. 영상 획득 보드는 실시간으로 영상을 모니터에 중첩(overlay)시키며 영상중 한 영상을 저장하여 처리하였다.

III. 전처리 과정 및 화면 스캔 과정

영상 획득 보드에서 얻은 512×512 영상에서 256×256 크기의 윈도우를 취하여 9개의 부영상(sub-image)을 다시 획득하고 256×256 크기의 9개의 부영상에 대해서 전처리 과정을 취한다.

A. Scan 과정

영상 획득 보드에서 얻은 512×512영상을 중첩되는 9개의(3×3) 구역으로 나누어 다시 256×256 영상을 획득한다. 이 9개의 영상에 대해 번호를 매겨서 각각에 대해서 검색체 판별을 한다. 검색체가 있다고 판정된 영상의 번호를 기억한다. 부영상 획득 과정이 그림6에 나와있다.

B. 대조도 확장 처리(Contrast stretching)

영상에서 검색체와 배경간의 대조가 좋지않기 때문에 contrast stretching 기법을 이용해서 배경과 검색체와의 구분을 명확하게 하였다. 주어진 gray level $u \in [0, L]$ 을 gray level $v \in [0, L]$ 로 변환하는 과정이다.[9].

$$v(u) = \begin{cases} \alpha u, & 0 \leq u < a \\ \beta(u-a) + v_a, & a \leq u < b \\ \gamma(u-b) + v_b, & b \leq u < L \end{cases} \quad (1)$$

즉 [a,b]구간의 level의 범위를 확장해서 대조를 넓힌다. 어두운 부분을 stretching 할때는 α 의 기울기를 크게하고 중간 부분을 stretching 할때는 β 의 기울기를 크게 하며 밝은 부분을 stretching 할때는 γ 의 기울기를 크게 하고 따라서 나머지 두 기울기는 상대적으로 작아지게 된다. 그림2(a)는 스캔 과정에서 얻은 부영상이고 그림2(b)는 대조도 처리의 결과 영상이다.

C. 문턱치 처리(Thresholding)

Thresholding 기법을 사용하여 배경과 검색체, 혈구등을 구분해냈다. 문턱치는 histogram을 육안으로 보고 구했다. 이 시스템에서 현미경의 밝기와 외부 환경의 밝기가 고정되어 있으므로 문턱치(threshold)는 고정해놓았다. 그림3(a)는 대조도 확장 처리의 결과영상이고 그림3(b)는 문턱치 처리의 결과영상이다.

D. 형태학적 처리(Morphological filtering)

대부분의 morphological operation은 기본적으로 erosion과 dilation의 두가지 operation을 사용한다. X를 object라고 하고 B를 structuring element 라고 할때 erosion과 dilation은 다음과 같이 표현된다.[10][11]

$$\text{Erosion: } X \ominus B \triangleq \{x: B_x \subset X\} \quad (2)$$

$$\text{Dilation: } X \oplus B \triangleq \{x: B_x \cap X \neq \emptyset\} \quad (3)$$

위의 기본 operation을 사용하여 영상에서 검색체와 혈구 이외의 다른 물체들을 제거한다. 검색체보다 작은 점 잡음들은 erosion을 반복하여 제거한다. 이때 검색체도 크기가 작아지지만 다시 erosion의 시행 횟수 만큼 dilation을 반복하면 완벽하지는 않지만 원래의 크기대로 회복된다. 이때 검색체의 회복되는 형태는 structuring element의 형태와 그것의 원점에 따라 달라진다. morphological filtering의 structuring element는 3×3 윈도우를 사용하였다.[12] 3×3보다 큰 5×5의 structuring element를 사용하면 점 잡음들의 제거에 대한 성능은 매우 좋으나 검색체들도 많은 부분이 소실되어 부적합하다. 그림5(a)는 원의 모양을 근사화하여 표현했으며 복원되는 물체의 형태는 경계선 모양이 곡선을 그리게 된다. 여기서 erosion의 효과를 높이기 위해서 그림5(b)를 선택하였다. 그림3(b)를 erosion한 것이 그림4(a)이고 그림4(a)를 dilation한 것이 그림4(b)이다.

IV. 검색체 판별 알고리즘

RL Coding은 영상의 압축, 영역 표현(region representation)등에 사용된다.[9] 앞에서 설명한 전처리 과정을 통해 얻은 256×256 2진 영상에 대해서 RL coding을 응용한 새로운 coding방법을 수행한다. 이러한 coding방법은 그림7에서와 같이 256×256 영상의 한 줄씩을 스캔하면서 RL coding을 하게 되는데 256줄을 모두 실행한다. 이때 영상의 각 줄마다 RL coding된 횟수를 구하여 원소수가 256인 벡터를 구성한다. 이 벡터를 다시 RL coding하여 coding의 빈도수를 측정하게 되면 최대수치가 256을 넘지않는 빈도수 데이터를 얻게되는데 이것이 검색체 판별의 feature가 된다. 검색체가 들어있는 영상은 복잡하여 coding횟수가 높은 반면 혈구가 들어 있는 영상은 간단해서 coding횟수가 적다. 결국 물체 빈도수 측정 알고리즘은 영상의 RL coding횟수를 평가하는 것이다. 검색체가 들어있는 90개의 영상과 혈구가 들어있는 90개의 영상에 대해 실시하여 각 90개씩의 데이터를 얻게된다.

V. 통계적 산출

검색체, 혈구 각각의 90개씩의 데이터에 대해서 확률밀도함수를 구하여 판별(decision)의 문턱치(threshold)를 구하게 된다.[5] 이 문턱치(threshold)를 기준으로 256×256영상안에 검색체의 존재여부를 판가름하게 된다. 검색체 부류(class)와 혈구 부류(class)의 분포가 그림7에 나타나 있

다. 잘못 판단할 수 있는 확률이 가장 작은 문턱치를 계산한다. c_1 을 염색체에 대한 부류(class), c_2 를 혈구에 대한 부류(class)라 하고 P_1, P_2 를 각 부류(class)의 a priori 확률, $p_1(X), p_2(X)$ 를 conditional probability라 한다. L_1 과 L_2 는 X 가 부류 c_1 과 부류 c_2 로 분류될때의 영역을 나타낸다. Bayes decision rule을 사용하였다. Bayes error를 구하면,

$$\begin{aligned} \epsilon &= P_1 \int_{L_2} p_1(X) dX + P_2 \int_{L_1} p_2(X) dX \\ &= P_1 \epsilon_1 + P_2 \epsilon_2 \end{aligned} \quad (4)$$

ϵ_1 은 부류 c_1 을 부류 c_2 로 판단했을때의 오류이고 ϵ_2 는 부류 c_2 를 부류 c_1 으로 판단했을때의 오류이다. 여기서 Bayes error ϵ 을 최소로 하는 문턱치를 결정한다.

VI 결과

염색체와 혈구에 대한 90개씩의 training 영상에 대한 결과는 다음과 같다.

(표 1)

| | 염색체 | 혈구 |
|------|---------|---------|
| 평균 | 67.78 | 17.1556 |
| 표준편차 | 14.5122 | 8.9341 |

위의 통계적 수치를 이용하여 판별의 문턱치를 48로 정했으며 이 값은 Bayes error를 최소로 한다.

염색체와 혈구에 대한 30개씩의 test 영상에 대한 결과는 다음과 같다.

(표 2)

| | 판별갯수 | 오류갯수 | 성능평가 |
|-----|------|------|-------|
| 염색체 | 25 | 5 | 83.3% |
| 혈구 | 28 | 2 | 93.3% |

VII. 결론 및 고찰

광학 현미경에서 획득한 염색체 영상에서 염색체의 위치(site)를 찾는 시스템 연구에 중점을 두었다. 염색체 영상에 대해서 여러가지 영상 개선(image enhancement)과 영상 압축 방법중의 하나인 RL coding을 응용하여 염색체를 구분해내었다. 형태학적 처리를 이용하여 점 잡음을 제거하

므로써 물체 빈도수 측정 알고리즘의 효과를 높였고 처리 순서는 erosion 2번, dilation 2번을 순차적으로 실행하였다. 염색체와 혈구에 대한 training set을 가지고 시스템을 training하여 두 class를 구분하기 위한 문턱치(threshold)를 구하였는데 문턱치는 48에서 판별 오류의 최소치를 갖는다. 염색체 분석의 전단계인 염색체의 위치(site)판별을 자동화함으로써 많은 시간과 비용을 줄일 수 있을 것이다. 앞으로의 발전방향은 현미경의 선반(stage)를 모터 구동에 의해 자동화함으로써 염색체 위치(site) 판별에 완벽한 자동화를 시도할 것이다.

참고 문헌

- [1] D.E.Rooney, B.H.Czepulkowski, Human Cytogenetics a practical approach vol.1, Constitutional analysis: Oxford University press
- [2] D.E.Rooney, B.H.Czepulkowski, Human Cytogenetics a practical approach vol.2, Malignancy and acquired abnormalities: Oxford University press
- [3] 조 종 만, 홍 승 홍 "염색체 자동분류 시스템의 구현" 의공학회지: 제9권, 제2호, 1988
- [4] Jenes Gregor "Finding chromosome centromeres using band pattern information", Comput. Biol. Med. vol.21, No.1/2, pp.55-67, 1991
- [5] Morton Nadler, Eric P.Smith Pattern Recognition Engineering, John Wiley & Sons INC. Wiley-Interscience Publication
- [6] Franck Neycenssac "Contrast enhancement using the Laplacian-of-a-Gaussian filter" CVGIP: GRAPHICAL MODELS AND IMAGE PROCESSING vol.55, No.6, November, pp 447-463, 1993
- [7] John F. Haddon "Generalised threshold selection for edge detection" Pattern Recognition, vol.21, No.3, pp 195-203, 1988
- [8] A.K.Jain, Fundamental of Digital Image Processing, Englewood Cliffs, NY: Prentice Hall, INC. 1989

[9] E.R. Giardino, E.R. Dougherty Morphological Methods in Image and Signal Processing Prentice Hall, Englewood Cliffs, New Jersey

[10] E.A. Glasbey "An analysis of histogram-based thresholding algorithms" CVGIP: GRAPHICAL MODELS AND IMAGE PROCESSING vol.55, No.6, November, pp. 532-537, 1993

[11] Harley R. Myler, Arthur R. Weeks Computer Imaging Recipes in C, Englewood Cliffs, NJ: Prentice Hall

[12] R.C. Gonzalez and R.E. Woods, Digital Image Processing, Addison Wesley, 1992

[13] Keinosuke Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, INC 2nd edition

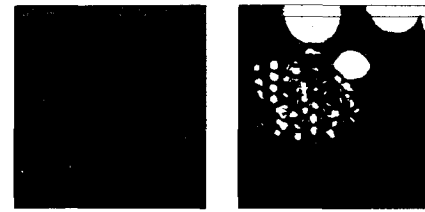


그림 3 Thresholding
(a)는 contrast stretching을 한 영상이고 (b)는 thresholding 한 영상이다.

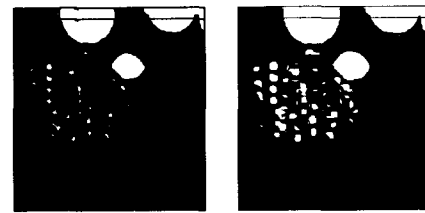


그림 4 Morphological filtering
(a)는 erosion, (b)는 dilation 한 것이다.
(a)는 erosion, (b)는 dilation 한 것이다.

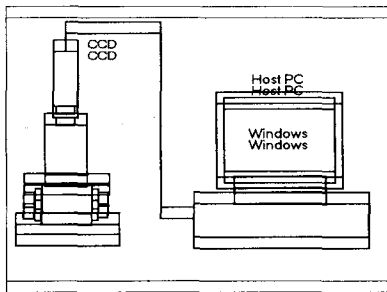


그림 1 시스템 구성도



그림 2 Contrast stretching
(a)는 CCD camera를 얻은 영상이고 (b)는 contrast stretching 한 결과 영상이다.

| | | |
|---|---|---|
| 0 | 1 | 0 |
| 1 | 1 | 1 |
| 0 | 1 | 0 |

(a)

| | | |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 1 | 1 |
| 1 | 1 | 1 |

(b)

그림 5 Morphological filter의 Structuring element

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |

그림 6 영상들을 판정
512x512영상을 9개의 256x256영상을
얻는 과정이다.

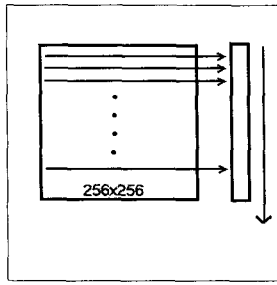


그림7 물체 빈도수 측정 방법
RL coding 방법을 응용하여 검색체
영상에서의 물체의 빈도수를 측정하
는 과정이다.

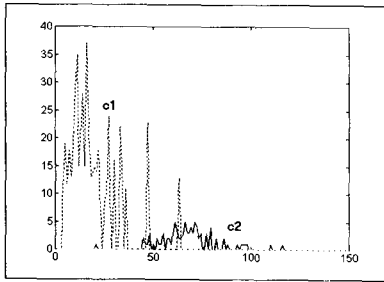


그림8 검색체와 혈구 class의 분포
c₁은 혈구의 부류이고 c₂는 검색체의 부류
이다.