

핵형분석을 위한 염색체 영상 표본의 자동 추출

°장용훈, 이권순, *전 계록
(동아대학교 전기공학과, *부산대학교 병원 의공학과)

Automatic Extraction of Chromosome Image Samples for the Karyotype Analysis

°Yong Hoon Chang, Kwon Soon Lee, *Gye Rok Jeon
Dept. of Electrical Eng., Dong-A Univ.,
*Dept. of Medical Eng. Pusan National Univ. Hospital

Abstract

Chromosome analysis is an important and difficult task for clinical diagnosis, for mutagen dosimetry, and for biological research. It is expensive, time consuming and imprecise when performed manually. Efforts to automate some or all of the procedures have continued for more than 30 years, with only limited success. An acquiring sample from chromosome group is not solved with automatic method. It is still performed by user. This paper represents the method of an automatic chromosome sample extraction which based on region splitting, and scan converted method.

I. 서 론

염색체(Chromosome)의 세포유전학(Cytogenetics)적인 해석은 인체의 유전학적 진단(Diagnosis) 및 일반 동식물에 관한 유전학 연구를 위하여 널리 사용되고 있다. 이러한 해석은 일반적으로 염색체의 핵형분석(Karyotype)을 통하여 이루어지며, 특히 산부나 태아의 핵형분석은 이상염색체를 일신 초기에 판별하여 태아의 이상유무를 진단하기 위하여 임상에서 널리 사용하고 있다. 즉, 유사분열 중기(Mitosis Metaphase)의 염색체를 해석하여 정상적인 염색체와 비교하고, 그 변이-염색체의 수, 형태학적 특징(Morphological Feature), 염색마디(Chromosome Bend)-를 조사하여 진단에 사용하고 있다. 그러나 이러한 염색체 해석은 고도로 숙련된 세포유전학 전문가에 의해 수행되며, 세포 배양, 염색(Stain), 슬라이드 제작, 관찰하기 용이한 염색체군(Chromosome Group)의 선택 및 사진 촬영, 인화된 염색체의 수자위에 의한 분리, 그리고 염색체 해석 및 핵형 작성이라는 매우 복잡하고 많은 작업이 요구된다. 이와같은 수작업은 염색체를 해석하기 위해 많은 시간이 소요되고 있다.

컴퓨터를 이용한 염색체의 자동 추출 및 해석은 1964년 Ledly[1]에 의하여 최초로 시도되어 임상 세포유전학 연구실에 유용하게 사용할 수 있음을 제시하였고, Neurath [2]-[4]는 혈미경의 화상데이터로 부터 염색체의 유파선(Contour)을 추출하고, 동원체의 위치를 계산하는 간단한 알고리즘을 도입하여 염색체를 해석하는 시스템을 구성하였다. 1989년 Piper와 Granum[5]은 면적, 상대농도, 길이 외에 가중농도 분포함수(Weighted Density Distribution function)를 사용하여 동원체의 위치를 검출하는 새로운 알고리즘을 제안하여 염색체를 분류하였고, 1990년 Young, Mayall 및 Delft 그룹의 Lucas 등[6]은 염색체의 길이, 동원체 지수(Centromeric Index) 및 염색마디에 관한 특징 벡터를 사용하는 염색체 해석 시스템을 개발하였다. 그러나 이러한 연구 노력의 대부분은 고가의 대형 컴퓨터를 이용한 경비이며, 전체 처리 과정 중 염색체 표본들을 추출하기 위하여 운용자의 개인 비중이 상당히 높게 차지하였다[7][8]. 그러므로 운용자의 개인요소를 배제하기 위한 염색체 표본의 자동 추출에 관한 연구는 아직까지 미진한 실정이다.

염색체 표본의 자동 추출은 전자 현미경 상에 나타나는 염색체 군의 영상을 영상 입력장치를 사용하여 컴퓨터로 취득하고, 취득

된 염색체 영상에서 각 염색체들의 동원체 지수(Centromeric Index), 염색체의 상대적 길이 비 및 염색마디에 대한 농도치(Gray Level) 등 염색체의 분류에 필요한 특정 파라메타들을 전처리 과정을 통하여 찾아낸 후 각각의 염색체들을 추출하게 된다. 특정 파라메타들을 구하기 위한 전처리 과정은 먼저 추출한 염색체의 표본들을 이용하여 하며, 각 염색체의 영상 정보들이 소실됨이 없이 원래의 영상정보를 추출하여야 정확하게 염색체의 특정 파라메타들을 구할 수 있다[9]. 그러나 염색체 군의 영상에서 개개의 염색체들의 분리를 용이하지 않으며, 더구나 겹쳐 있거나 구부러진 형태로 존재하기 때문에 정확한 개개의 영상정보의 추출은 더욱 어렵다.

염색체 표본을 추출하기 위한 방법은 대개 유파선 추출 기법을 사용하고 있으며[10], 최근에는 영역 분할법에 의한 연구도 진행되고 있다[11]. 영상에서 영역을 분할하는 방법으로는 명암도의 불연속성과 유사성에 근거하여 다음과 같이 크게 두 가지로 나누어 진다. 첫번째는 에지 기반 방식(Edge Based Method)으로 Sobel등의 에지 연산자를 적용하여 적절한 임계치(Threshold)를 설정하여 찾은 에지로부터 영역을 분할하는 방식이며, 두번째는 히스토그램 기반 방식(Histogram Based Method)으로 영역 확대(Region growing), 영역 분할 및 합성(Region splitting and merging)등의 방식으로 영역을 분할한다[12]. 이러한 방법을 이용하여 염색체 영상의 영역을 분할하여 사용할 경우 임계치의 설정 또는 2차화된 영상의 조작 등은 여전히 운용자의 개인요소들이 아주 많이 포함되므로 각 염색체의 표본을 자동추출하는 과정에서는 비실용적이다.

따라서 본 논문에서는 영역채움기법(Region filled Method)과 스캔변환기법(Scan Conversion Method)을 이용하여 염색체의 표본을 자동으로 추출하는 알고리즘을 제안한다. 또한, 현재 사용하고 있는 염색체 표본 추출 알고리즘의 결과와 본 논문에서 제안한 알고리즘을 사용하여 염색체 표본을 자동 추출한 결과를 비교하였다. 신경망 Borland C 4.5를 사용하여 프로그램 되었으며, IBM PC 586에서 수행되었다. 본 논문에서 제안한 알고리즘을 이용하여 실험한 결과 월 영상의 정보만을 추출하는데 있어 기준의 방식을 보다 더욱 우수한 특성을 보였으며, 운용자의 개인이 각 염색체 표본들을 자동적으로 추출할 수 있었다.

II. 본 론

1. 염색체 표본 추출

염색체의 표본을 추출하기 위하여 먼저 세포 배양, 염색, 슬라이드 제작을 거쳐 현미경을 통하여 관찰하기 용이한 염색체군을 선택하고 이를 활영하여 표본의 추출에 사용한다. 또한 현미경상에 나타나는 영상은 형광 물질을 투여하여 염색체들을 가시화하는 분위법을 이용한 것이다. 이 영상에서 염색체들은 여러 가지 밝고 어두운 부분들이 농도에 따라 염색마디(Stained bend)를 이루고 있으며, 염색체 이외의 배경 부분도 미세하지만 영상에 포함되어 나타난다. 이러한 염색체 이외의 배경 부분은 실제로 염색마디 중 밝은 염색마디들과 비슷한 농도를 나타내므로 구별이 용이하지 않다.

현미경 상에 나타나는 염색체군의 원영상에서 각 염색체의 표본추출은 아직 자동화가 이루어 지지 않아 상당부분 또는 전과정이 운용자의 개인을 요구하고 있다. 즉, 전체 염색체의 영상에

서 마우스 등과 같은 지시장치(Pointing Device)를 이용하여 각 염색체를 잘라내는 방법을 사용하고 있다. 따라서 개개의 염색체로 분리하는 데에는 많은 시간이 소요되며, 운용자의 판단에 의한 영상정보의 추출은 영상정보의 소실이 발생하거나, 염색체 이외의 배경부분이 많이 포함되어서 추출된다. 이와같이 지시장치를 사용하여 추출한 염색체는 염색체 자동분류에 사용하는 특정 파라메타의 정확한 추출에 상당한 문제점을 가지게 된다. 염색체 표본 추출시 영상정보의 오류를 최소화하여 정확한 개개의 염색체를 얻고, 시간이 많이 소요되는 작업을 단축하기 위하여 염색체 표본 추출의 자동화 작업이 필요적이다.

본 연구에서는 염색체 표본 자동 추출 알고리즘을 제안하고 이를 사용하여 운용자의 개입을 배제시켜 염색체 분류에 소요되는 시간을 단축시키고, 각 염색체에 대한 영상정보의 소실을 최소화하는 장점을 지니고 있음을 보이고자 한다.

2. 지시장치를 이용한 염색체 표본 추출

지시장치를 이용하여 염색체 표본을 추출하는 방법은 마우스 등을 이용하여 원영상을에서 각 염색체별로 직접 염색체의 가장자리를 따라 표본을 추출하는 것으로, 이와 같은 방법을 사용하기 위해서는 컴퓨터와 운용자 간의 인터페이스(Interface)에 아주 많은 기능들이 요구된다. 먼저, 작업을 원활하게 할 수 있도록 영상의 축소, 확대가 자유롭고, 편리한 자르기(Cutting-up), 붙이기 기능, 표본 보관 기능 등의 많은 그래픽 편집 기능을 가져야 한다. 또한, 이러한 기능이 모두 내포된 편집기(Editor) 기능이 있다 하더라도 각 염색체의 표본을 추출하는 데에는 상당한 시간이 소요되며, 정확도에 있어서는 정밀하지 못한 지시장치로 인하여 발생하는 각 표본당 영상정보의 소실 부분에 대한 보상작업이 수행되어야 하는 단점을 가지고 있다. 이와같은 지시장치를 이용한 염색체 표본의 추출 작업은 운용자의 개입이 필요적이므로 보다 정밀한 염색체 표본의 추출에 소요되는 시간을 단축하기 위하여 염색체 표본 추출 작업의 자동화가 필요하다.

3. 윈도우를 이용한 염색체 표본 추출

일반적인 디지털 영상(Digital Image)은 사각형인 공간영역(Spatial Domain)적 영상정보 특성을 가지고 있으며, 이를 기준으로 한 영상 처리 방식은 항상 사각형(Square)의 윈도우(Window) 내에서 이루어 진다. 그러나 현미경 상에 나타나는 염색체 군의 영상은 각각이 따로 존재하지 않으며, 여러 염색체가 근접하거나 겹쳐져 있는 형태이다. 이와 같은 염색체는 불규칙한 형태로 잘라내는 방식을 사용하여야 하며, 사각의 윈도우를 이용하여 따로 분리할 경우에는 많은 불필요한 배경영상이 포함되므로 상당히 곤란한 문제가 발생한다. 따라서, 배경영상이 어느 정도 포함된 개개의 염색체 영상을 윈도우를 사용하여 잘라낸 후 배경 영상을 해당하는 부분을 제거하는 작업이 필요하다.

본 논문에서 사용한 배경 영상 제거 방법은 다음과 같다.

- 1) 먼저 분리된 윈도우 내의 영상 중에서 염색체의 영상만을 추출하기 위하여 윈도우 전체에 대한 높도 히스토그램을 구하고, 임의의 높도 임계치(Gray Threshold Level)를 설정하여 임계치 이하는 배경(White)으로, 그 이상은 염색체(Black)로 간주하여 2차화 영상을 구한다.
- 2) 2차화된 염색체 영상에서 경계선 검출(edge detection)을 수행하면 각 염색체의 경계선이 폐곡면으로 나타난다. 이렇게 나타난 경계선의 좌표를 사용하여 원 영상에서 경계선 좌표가 이루는 폐곡면 이내의 영상정보를 읽어낸다.
- 3) 경계선 좌표 이내의 영상정보를 재구성하면 각 염색체의 표본을 얻을 수 있다.

이상과 같은 배경 영상 제거 방법은 임계치의 설정에 따라 노이즈 성분의 포함 정도가 달라지며, 오히려 원영상의 영상 정보가 소실될 위험성을 내포하고 있다. 또한, 윈도우 내의 염색체 표본의 영상 정보는 배경과 같은 무효영역(White)이 포함되어 있으므로 높도분포의 휘도차에 의한 임계치 설정이 곤란하므로 정확한 염색체 표본 추출이 어렵다.

윈도우를 이용한 염색체 표본 추출 방식은 운용자의 관찰에 의하여 각 염색체의 윈도우를 분리하여야 하며, 분리된 윈도우마다 배경 영상 제거 과정을 거쳐야 하고, 이 과정에서 염색체 임계치를 일일이 확인해야 하므로 자동화가 거의 불가능하다.

III. 제안한 염색체 표본 자동 추출 알고리즘

제 II장에서 실명된 염색체 표본 추출 방법은 운용자의 기술과 판단 능력이 중요한 부분을 차지하고 있고, 이에 소요되는 시간

이 염색체 해석에 소요되는 시간의 상당한 부분을 차지하고 있다.

본 논문에서는 운용자의 개입을 최소로 줄일 수 있으며, 염색체 분류과정의 소요시간을 단축시킬 수 있는 염색체 표본 자동 추출 알고리즘을 제안한다. 그림 1은 제안한 염색체 표본 자동 추출 알고리즘이다.

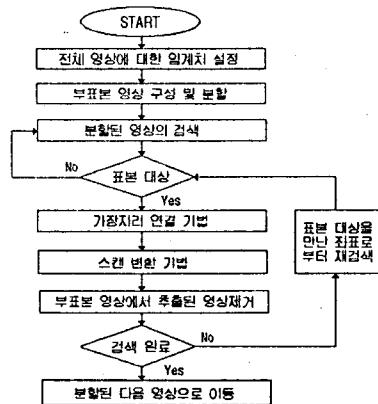


그림. 1. 염색체 표본 추출 알고리즘

이 방법은 원 영상을 이용함으로써 높도분포의 휘도차가 크지 않아 임계치 설정이 용이하고 정확도가 높으며, 전체 영상 검색 방법에 비해 속도가 빠르다. 또한, 자동적으로 표본을 그래피화 일 형태로 보관하므로 표본관리가 더욱 편리하다는 장점이 있다.

가) 영역 채움 기법

영역 채움 기법에는 터보 C나 MS-C의 그래픽 라이브러리 함수 중에 홍수 채움(Floodfill), 다각형 채움(fillpoly) 등 여러 가지 종류가 있다. 이 중 홍수 채움은 씨앗 채움(Seed fill) 기법의 일종으로 시작점을 지정하면 그 주변을 중심으로 퍼져 나가고, 지정된 테두리를 만나면 멈추는 형식이다.

이 기법은 세기호출을 이용하여 간편하고 이해하기도 쉽다. 하지만 이런 식으로 프로그램을 짜면 속도가 느리고, 스택도 넘쳐버릴 위험성이 크다. 따라서 보통 이 기법을 토대로 수평선을 y 축 방향으로 한 줄씩 검색하며, 한 줄 단위로 칠하는 방법을 사용하는 홍수 채움 기법을 많이 이용한다. 그러나, 염색체 표본 추출에 홍수 채움을 사용하면, 하나님의 폐곡선이다. 씨앗점의 데이터가 추가되어야 하고, 세기호출로 인한 스택부족으로 실행이 불가능하다. 이런 이유로 본 연구실에서는 다각형 채움 방식의 기법의 변형으로서, y 축 방향으로 한 줄씩 체워나가는 스캔 변환(Scan conversion)기법을 사용했다. 이 기법을 이용한 함수로는 블랜드 C의 Fillpoly와 MS C의 _polygon이 있다. 물론 이 함수를 그대로 사용하여 다각형 내부를 채울 수는 있으나 염색체 표본 추출을 위해서는 채우는 형식으로 스캔해 나가며 읽는 방식이 필요하다. 때문에 이 함수들을 구현하면 변형할 필요가 있다.

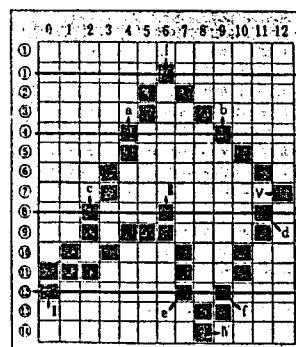


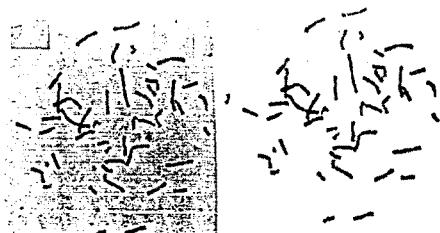
그림. 2. 스캔 변환 기법

나) 스캔 변환 기법

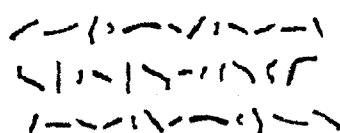
스캔 변환 기법은 다각형이 있을 때, 다각형이 가지하는 최소 $y_{\text{값}}$ 과 최대 $y_{\text{값}}$ 사이의 수평선을 하나씩 살펴 가며 유효영역을 결정해 값을 읽는다. 이때 어느 부분이 유효영역인지 판단하는 것이 가장 중요한데, 그림 2의 ①번 스캔 라인을 예로 든다. 이 경우는 꼭지점과 만나며 점 1 하나만이 유효하다. ④번 스캔 라인의 경우는 두 개의 모서리와 만나며, 직선 ab가 유효영역이다. ⑧번 스캔 라인의 경우는 c와 III 사이, III과 d 사이에 직선이 유효영역이다. 이러한 모든 경우를 고려하면 어떠한 모양의 다각형이라도 내부의 유효영역을 확인할 수 있다. C를 이용한 코드를 다음에 나타내었다. 단, 여기서 색을 칠하는 부분을 실험에 사용한 것으로 실제 사용시에는 색을 읽는 부분으로 대체하면 된다.

IV. 실험결과 및 고찰

영상정보의 전처리는 폐된인식에 사용되는 특징 파라메타를 추출하는 아주 중요한 영상처리 기법이다. 본 논문에서는 염색체의 폐된인식에 사용되는 특징 파라메타를 추출하기 위하여 제안한 알고리즘을 사용하여 각 염색체의 표본을 자동추출할 수 있었다. 그리고 염색체 영상에서 지시장치와 원도우기법을 이용하여 추출한 염색체 표본의 영상정보 특성을 비교하였다.



(a)



(c)

그림 3. 염색체 표본 자동 추출 처리과정
(a) 원영상 (b) 이지화된 영상
(c) 염색체 표본 추출 결과

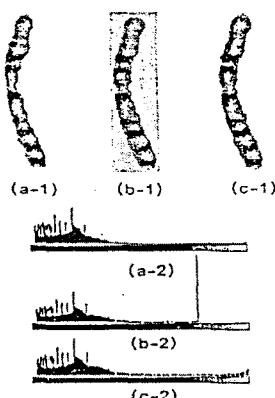


그림 4. 추출된 염색체 표본과 히스토그램
(a) 수작업 추출결과 (b) 원도우를 이용한 추출 결과
(c) 제안한 알고리즘을 이용한 추출결과

제안한 알고리즘에 의해 염색체 표본을 자동추출한 과정을 그림 3에 나타내었다. 그럼 3(a)는 현미경 상에 나타나는 염색체군을 CCD를 통하여 PC에 취득한 전체 영상을 나타내고 있고, 그림 3(b)는 원 영상을 이지화 시킨 결과이며, 그림 3(c)는 이지화 시킨 영상에서 유효영역의 좌표와 일치하는 원 영상의 염색체 영상정보 성분만을 추출한 결과를 보이고 있다. 각각의 염색체 성분을 그래픽 화일 형태로 저장할 수 있어 관리가 편리할 뿐만 아니라 자동분석 과정에서 처리가 용이하다는 장점을 가지고 있다.

본 논문에서 제안한 방법이 기존의 방법과 비교하여 우수함을 그림 4와 같이 나타내었다. 그림 4(a)는 1번 염색체를 마우스로 추출한 염색체의 표본 결과와 히스토그램을 그림 4(b)는 같은 염색체를 원도우 기법을 이용하여 추출한 결과와 히스토그램을 보이며, 그림 4(c)는 스캔 변환 기술을 이용하여 추출한 결과와 히스토그램 분포를 나타내고 있다. 각 히스토그램의 분포는 차이를 보이고 있다. 그림 4(a-2)와 그림 4(b-2)을 비교하면, 전자는 여러부분에서 원 영상이 소실된 결과를 나타내고 있다. 그림 4(b-2)는 배경 부분의 농도차 정보가 과다하게 포함되어 있음을 알 수 있다. 그림 4(c)는 제안한 알고리즘을 이용하여 배경 성분을 제외한 염색체의 영상만을 소실분 없이 추출한 결과를 나타낸다. 본 연구에서 제안한 방법을 사용하여 염색체 표본을 영상정보의 소실 없이 자동 추출할 수 있었으며, 표본 추출 처리 속도를 향상 시켰다.

V. 결 론

본 연구에서는 염색체 표본의 자동 추출을 위한 영역 분할 기법을 제안하였다. 제안된 영역 분할 기법은 영역 채움 기법과 스캔 변환 기법으로 나뉘어 수행되며, 수행 결과를 기준에 사용되는 방법과 비교, 검토를 하였다. 기존의 방법에서는 운용자가 염색체의 영상을 눈으로 관찰하면서 표본을 추출하기 때문에 수행시간이 많이 소요되었으며, 영상정보의 추출이 부정확하였다. 본 연구에서 제안한 알고리즘을 사용하였을 경우 염색체 표본 추출 수행시간이 기존의 방법보다는 현격히 짜르며, 염색체 표본의 영상정보만을 자동 추출함으로서 염색체의 진단에 효과적으로 사용할 수 있다.

현재까지 본 연구에서 진행된 결과는 염색체가 인접하거나 겹쳐지지 않은 경우에만 적용하여 실험하였다. 염색체가 인접하거나 겹쳐진 경우의 표본을 자동 추출하는 연구가 계속 진행중이다. 이것이 성공적으로 수행되었을 경우 제안한 알고리즘은 염색체 자동분류 시스템에 적용하였을 때 기존의 시스템들보다 성능과 기능이 탁월한 시스템을 개발할 수 있을 것이라 생각된다.

Reference

- [1] Robert S. Ledley, "High-speed automatic analysis of biomedical picture," *Science*, vol. 146, pp. 216-223, 1964.
- [2] Peter W. Neurath, Barkey L. Babluzain, Tom H. Warms, Russel C. Serbagi, "Human Chromosome analysis by computer—an optical pattern recognition problem," *Annals New York Academy of Sciences*, vol. 128, pp. 1013-1028, 1965.
- [3] Peter W. Neurath et al., "Individualized human karyotyping through quantitative analysis," *Comput. Biol. Med.* vol. 2, pp. 181-193, 1972.
- [4] G. Gallus and P. W. Neurath, "Improved computer chromosome analysis incorporating preprocessing and boundary analysis," *Physics in Medicine and Biology*, vol. 15, no. 3, pp. 435-445, 1970.
- [5] Jim Piper and Erik Granum, "On fully automatic features measurement for banded chromosomes classification," *Cytometry*, vol. 10, pp. 242-255, 1989.
- [6] Lucas J. van vliet, Ian T. Young, and Brian H. Mayall, "The Athens semi-automated karyotyping system," *Cytometry*, vol. 11, pp. 51-58, 1990.
- [7] Brian H. Mayall, James D. Tucker, Mari L. Christensen Lucas J. van Vliet, and Ian T. Young, "Experience with the Athens semi-automated karyotyping system," *Cytometry*, vol. 11, pp. 59-72, 1990.
- [8] C. Bruschi, F. Tedeschi, P. P. Puglisi, and N. Marmiroli, "Computer-assisted karyotyping system of banded chromosomes," *Cytogenetics and Cell Genetics*, vol. 29, pp. 1-8, 1981.