
영상 재구성방법을 이용한 염색체 영상의 패턴 분류

김충석*, 남재현*, 장용훈**

Pattern Classification of Chromosome Images using the Image Reconstruction Method

Choung-Seok Kim*, Jae-Hyun Nam*, Yong-Hoon Chang**

요 약

본 연구에서는 염색체의 영상패턴을 인식하고 분류하는 방법을 개선하기 위해 패턴인식의 특징정보로 사용되는 비선형적인 염색체 영상을 선형적으로 재구성하는 영상 재구성 알고리즘을 사용하여 선형화된 특징정보를 추출하여 패턴분류기인 신경회로망의 입력정보로 사용한다. 중앙축 변환방법과, 영상 재구성방법을 사용하여 임상적으로 정상인으로 판명된 20명의 염색체 영상의 특징정보를 추출하였다. 중앙축 변환방법에 의하여 추출된 특징정보의 패턴조합과 영상 재구성방법에 의하여 추출된 특징정보의 패턴조합을 구성하였으며, 10명에 대하여 추출한 특징정보를 계층적인 신경회로망(Hierarchical Multilayer Neural Network : HMNN)의 학습입력으로 사용하여 염색체를 분류하기 위한 패턴인식기를 구현하였다. 그리고 나머지 10명에 대하여 학습입력과 동일하게 조합된 패턴조합을 HMNN의 분류입력으로 사용하여 수행한 결과 약 98.26%의 우수한 인식률을 나타내는 최적화된 패턴인식기를 구현할 수 있었다.

ABSTRACT

To improve classification accuracy in this paper, we proposed an algorithm for the chromosome image reconstruction in the image preprocessing part. also we proposed the pattern classification method using the hierarchical multilayer neural network(HMNN) to classify the chromosome karyotype. It reconstructed chromosome images for twenty normal human chromosome by the image reconstruction algorithm. The four morphological and ten density feature parameters were extracted from the 920 reconstructed chromosome images. The each combined feature parameters of ten human chromosome images were used to learn HMNN(Hierarchical Multilayer Neural Network) and the rest of them were used to classify the chromosome images. The experimental results in this paper were composed to optimized HMNN and also obtained about 98.26% to recognition ratio.

키워드

Classification method, reconstructed chromosome images, combined feature parameters, Hierarchical Multilayer Neural Network

*신라대학교 컴퓨터정보공학부 교수

**동주대학 네트워크전자계열 교수

1. 서 론

인간에 대한 염색체 분석은 선천성 질환, 백혈병, 방사선에 의한 장애, 생물학적 인 연구 등을 수하는데 있어 필요한 연구로 알려져 있다.[1-2] 1956년 Tjio와 Levan에 의해 인간 염색체의 수가 밝혀졌으며, 1960년 국제 염색체 회의에서 인간 염색체에 대한 명명법과 분류법에 관하여 ISCN(International system for human cytogenetic nomenclature)규약이 설정되었다[3,4]. 1964년 Ledly에 의해 컴퓨터를 이용한 염색체를 분류하는 연구가 처음으로 수행되어 컴퓨터를 사용한 시스템이 세포유전학 연구실 등에 사용할 수 있다고 보고하였다[5]. 특징정보를 추출하는 방법은 윤곽선을 이용하는 방법, 중앙축 변환(medial axis transformation : MAT)방법, 염색체마디의 패턴정보를 사용하는 방법 등이 있다[6,7]. 패턴을 인식하는 방법은 Maximum likelihood를 사용하는 방법, Nonparametric bayes' rule을 사용하는 방법, 신경회로망을 이용하는 방법들이 제안되어 연구가 진행되고 있다[6,7]. 현미경으로 관찰되는 염색체의 원형상은 비선형적인 형태로 나타나며, 이러한 영상을 사용하여 특징정보를 추출하는 대표적인 방법으로는 MAT를 이용하는 방법이 있으나, 이는 같은 번호의 염색체라도 세포분열과정에서 나타나는 요인으로 인하여 동일한 형태로 관찰하기가 매우 어려운 실정이다. 유전학 전문가가 아닌 컴퓨터에 의해 보다 정확하게 판단하기 위해서는 선형화된 특징정보추출이 선행되어야 한다. 본 연구에서는 비선형적으로 나타나는 염색체영상을 선형적으로 재구성하는 방법을 제안하였으며, 동일한 염색체에서 MAT방법을 사용하여 추출한 특징정보와 영상 재구성방법에 의해 추출된 특징정보를 본 연구에서 제안하는 패턴인식기를 사용하여 패턴인식의 정확도를 비교 하고자 한다. 본 연구에서 사용하는 패턴인식기는 신경회로망을 이용한 인식기로 염색체의 군(group)과 각 군에 속하는 염색체들을 인식하기 위하여 두 단계의 구조로 구성하여 패턴인식기의 성능을 평가하였다.

II. 본 론

1. 염색체 번호와 형태

인간 염색체에 대한 명명은 1960년 Denver, 1963년 London, 1966년 Chicago, 1971년 Paris, 1975년 Paris Supplement, 1977년 Stockholm, 그리고 1980년 Paris 국제회의를 거쳐 표 1과 같이 7개의 군으로 구분된다 [1,2,3,4].

표 1. 염색체의 특징과 번호
Table 1. The numbers and characters of chromosome

Group	Numbers	Characters
A	1~3	Large metacentric chromosome
B	4~5	Large submetacentric chromosome
C	6~12, X	Medium-size metacentric chromosome
D	13~15	Medium-size acrocentric chromosome with satellite
E	16	Relatively short metacentric chromosome
	17, 18	Submetacentric chromosome
F	19, 20	Short metacentric chromosome
G	21, 22	Acrocentric chromosome with satellite
	Y	Acrocentric chromosome

2. 염색체의 형태구조학적인 특징

염색체의 모양과 농도 등을 이용하여 표 2와 같이 동원체 지수, 상대적 길이비, 상대적 면적비 및 염색체 마디의 농도에 관한 특징정보들을 추출할수 있다[4,8].

III. 염색체 영상 재구성

1. 영상 재구성 알고리즘

영상을 재구성하기 위하여 세션화 알고리즘과 윤곽선 검출 알고리즘을 사용하여 염색체의 중앙축을 추출한다[8,9]. 추출한 중앙축에 존재하는 화소 당 32 방향으로 선을 방사하여 염색체 영역의 가장자리까지 32방향의 누적 길이를 구한다. 그리고 32개의 길이를 정렬하여 가장 짧은 선분의 끝 좌표와 중앙축의 원점과의 유클리드 거리를 계산하고, 계산치와 π 라디안만큼 떨어진 선분과의 합을 구하여 재구성을 위한 선분으로 결정한다. 그림 1은 염색체 영상을 재구성하는

과정을 나타내었다.

표 2. 염색체의 형태구조학적인 특징
Table 2. The morphological feature for chromosome

Feature Information	Equation or Density Information
Centromeric Index : C.I.	$C.I. = \frac{\ell_s}{\ell_c} \leq 0.5$
Relative Length Ratio : R.L.	$\ell_i = \sum_{c=0}^{44} \ell_c$ $R.L. = \frac{\ell_c}{\ell_i}$
Relative Area Ratio : R.A.	$S_i = \sum_{c=0}^{44} S_c$ $R.A. = \frac{S_c}{S_i}$
Density Information	Mean Density : M.D. Density Centromeric Index : M.C.I.

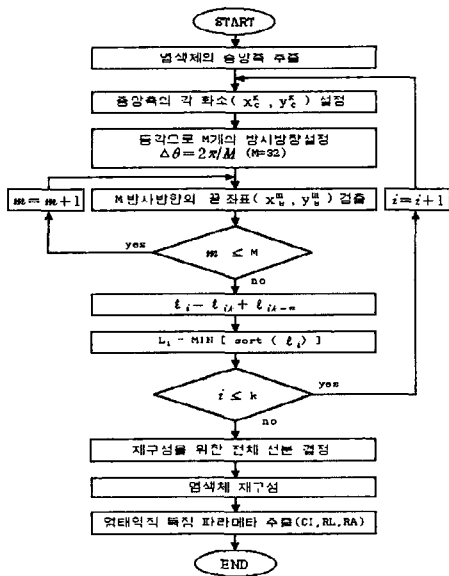


그림 1. 영상 재구성을 위한 알고리즘의 흐름도
Fig. 1 flowchart for the reconstruction algorithm

2. 패턴 인식기

염색체의 특징정보를 사용하여 염색체를 분류할 수 있는 패턴인식기를 구현하였다. HMNN1은 염색체 군

을 HMNN2는 각 군에 속하는 염색체를 분류하기 위하여 모듈화된 신경회로망으로 구성하였다. HMNN에 사용되는 7개의 염색체 군과 염색체번호는 표 3과 같으며, 표 4는 학습 및 분류입력으로 사용되는 패턴조합을 나타내며, 염색체를 분류하는 HMNN의 구성도는 그림 2와 같다.

표 3. 염색체 군과 번호
Table 3. Chromosome groups and numbers

Group	Number	Group	Number
A	1 ~ 3	E	16 ~ 18
B	4 ~ 5	F	19 ~ 20
C	6 ~ 12, X	G	21 ~ 22, Y
D	13 ~ 15		

표 4. 특징정보의 패턴
Table 4. Patterns of feature information

Combined patterns	Feature vectors	Feature number
Pattern1	C.I.+R.L.+R.A.	3
Pattern2	C.I.+R.L.+R.A.+C.L.	4
Pattern3	C.I.+R.L.+R.A.+C.L.+(D0~D9)	14
Pattern4	C.I.+R.L.+R.A.+C.L.+(C.I./R.L.)+(C.I./R.L.)+(R.L./R.A.)+(R.L./C.I.)+(R.A./R.L.)+(R.A./C.I.)	10
Pattern5	C.I.+R.L.+R.A.+C.L.+M.D.+M.C.I.+(C.I./R.L.)+(C.I./R.L.)+(R.L./R.A.)+(R.L./C.I.)+(R.A./R.L.)+(R.A./C.I.)+(D0~D9)	22

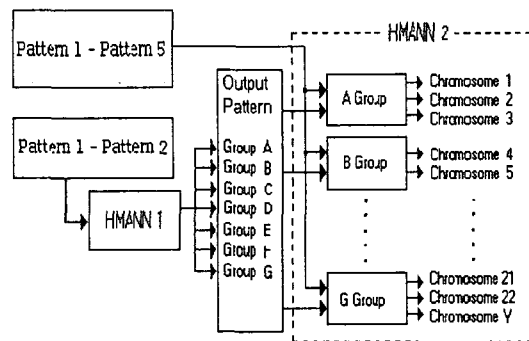


그림 2. 계층형 신경회로망의 구조
Fig. 2 The HMNN structure

IV. 실험결과

1. 특징정보 추출

1.1 MAT방법

염색체 영상에서 특징정보를 추출할 때 가장 많이 사용되고 있는 MAT방법에 의한 특징정보 추출의 과정을 그림 3에 나타내었다.

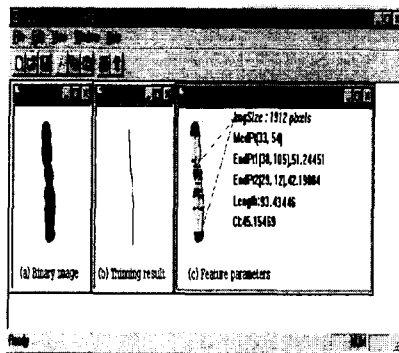


그림 3. MAT 방법을 사용한 특징정보 추출
Fig. 3 The feature information extraction using MAT

1.2 영상 재구성방법

영상 재구성방법을 사용하면 비선형적인 염색체 형태를 선형적인 형태로 재구성할 수 있다. 그림 4는 영상 재구성방법에 의하여 염색체를 재구성한 수행 과정과 특징정보의 추출 결과를 나타낸다.

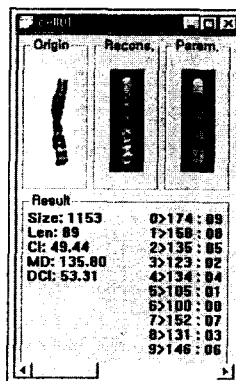


그림 4. 영상 재구성 방법을 사용한 특징정보 추출
Fig. 4 The feature information extraction using reconstructed image method

2. 패턴 벡터공간

본 연구에서는 460개의 염색체를 사용하여 MAT방법과 영상 재구성방법에 의하여 추출한 특징정보의 군집된 정도를 살펴보기 위하여 패턴 벡터공간을 구성하였다. 그림 5는 MAT방법에 의한 2차원 패턴 벡터공간을 그림 6은 영상 재구성방법에 의한 특징정보의 2차원 패턴 벡터공간을 나타내었다.

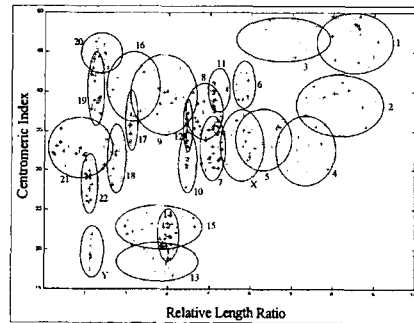


그림 5. MAT 방법에 의한 2차원 패턴 벡터
Fig. 5 Two dimensional pattern space using MAT

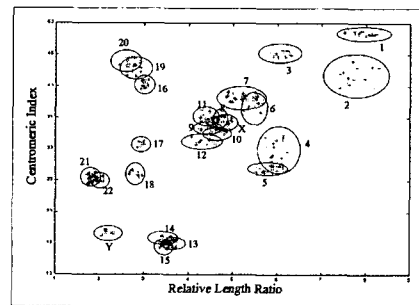


그림 6. 영상 재구성 방법에 의한 2차원 패턴 벡터
Fig. 6 Two dimensional pattern space using reconstructed image method

3. 패턴인식기 구현

표 5는 MAT방법과 영상 재구성방법을 사용하여 추출한 특징정보의 패턴조합을 나타낸다. HMNN1에는 PatternM과 PatternI의 패턴조합을 사용하여 학습 입력으로 하였고, HMNN2에는 5가지 패턴조합을 신경회로망의 학습입력으로 사용하여 패턴인식기 구현을 위한 실험을 수행하였다.

표 6은 패턴인식기를 구현하기 위하여 사용한 학습요소와 중간층뉴런의 수를 나타내었다.

표 5. 염색체 특징정보의 조합
Table 5. The combined feature of chromosome

특징정보 추출방법	특징정보 조합		패턴조합
	HMNN1	HMNN2	
MAT방법	PatternM	C.I., R.L., R.A.	PatternM
재구성 방법	Pattern1	C.I., R.L., R.A.	Pattern1
		C.I., R.L., R.A., C.L.	Pattern2
		C.I., R.L., R.A., C.L., D0~D9	Pattern3
		C.I., R.L., R.A., C.L.,(C.I./R.L.),(C.I./R.A.), (R.L./C.I.),(R.L./R.A.),(R. A./C.I.),(R.A./R.L.)	Pattern4
		C.I., R.L., R.A., C.L., M.D.,M.C.I.(D0~D9),(C.I. /R.L.),(C.I./R.A.),(R.L./C. I.)(R.L./R.A.),(R.A./C.I.), (R.A./R.L.)	Pattern5

표 6. 신경회로망의 파라미터
Table 6. Parameters of neural network
(PM : PatternM, P1 : Pattern 1, P2 : Pattern 2,
P3 : Pattern 3, P4 : Pattern 4, P5 : Pattern 5)

신경 회로망 파라미터	HMNN1	HMNN2					
		Group A		Group B		Group C	
		PM P1	P3 P2	PM P1	P3 P2	PM P1	P3 P2
함수기울기	1.0	1.2	1.2	1.1	1.1	1.0	1.0
학습상수	0.05	0.05	0.05	0.01	0.1	0.05	0.001
중간뉴런수	30	10	15	10	30	30	30

신경 회로망 파라미터	HMNN2	HMNN2							
		Group D		Group E		Group F		Group G	
		PM P1	P3 P2	PM P1	P3 P2	PM P1	P3 P2	PM P1	P3 P2
함수기울기	0.9	0.9	0.8	0.8	1.0	1.0	1.0	1.0	
학습상수	0.1	0.001	0.5	0.001	0.05	0.05	0.1	0.01	
중간뉴런수	15	20	15	30	10	30	10	30	

4. 인식 결과

염색체의 핵형을 분류하기 위하여 계층적 신경회로망으로 패턴인식기를 구현하였다. 그리고 분류오차를 감소시키기 위하여 5가지 종류의 패턴조합을 패턴인식기의 입력정보로 사용하여 염색체를 분류하는 실험

을 수행하였다.

먼저 HMNN1에 의한 군의 인식률은 MAT방법에 의한 경우 97.39%를 인식하였고 전체 염색체의 인식률은 92.83%인 반면, 영상 재구성방법을 사용한 경우 염색체 군의 인식률은 100%이고 전체 염색체의 인식률은 95.44%로 MAT방법 보다 우수한 인식률을 나타내었다. 그리고 HMNN2를 사용한 전체 염색체의 인식결과는 표 7과 같다.

표 7. 염색체 분류 결과
Table 7. The classification results of chromosome

패턴조합과 염색체 분류		A	B	C	D	E	F	G	Total	
M A T 방법	PM	오분류수	3	2	11	5	3	4	5	33
	오차율	5.0	5.0	7.1	8.3	5.0	10.0	11.4	7.17	
재구성 방법	P1	오분류수	0	0	12	6	0	3	0	21
		오차율	0	0	7.7	10.0	0	7.5	0	4.56
	P2	오분류수	0	0	9	4	0	2	0	15
		오차율	0	0	5.8	6.7	0	5.0	0	3.26
	P3	오분류수	0	0	7	3	0	1	0	11
		오차율	0	0	4.5	5.0	0	2.5	0	2.39
	P4	오분류수	0	0	13	40	0	0	0	53
		오차율	0	0	8.3	66	0	0	0	11.52
	P5	오분류수	0	0	6	2	0	0	0	8
		오차율	0	0	3.9	3.3	0	0	0	1.74

5. 최적 패턴인식기 조합

염색체를 분류하기 위하여 계층형 신경회로망을 사용하여 MAT방법에 의하여 추출한 특징정보의 패턴조합과 영상 재구성방법에 의한 5가지 종류의 패턴조합을 사용하여 표 8과 같은 분류결과를 나타내었다. 입력패턴으로 C군과 D군은 Pattern5를 사용하고 나머지 염색체 군은 Pattern4를 사용하여 최적화된 계층적 신경회로망을 이용한 패턴인식기를 구성할 수 있었다

표 8. 신경회로망의 조합
Table 8. The combine of neural network

Pattern	A	B	C	D	E	F	G
Pattern4	○	○			○	○	○
Pattern5			○	○			

IV. 결 론

본 연구에서 수행한 HMNN1에 의한 염색체 군의 인식률은 MAT방법에 의한 경우 97.39%를 인식하였고 전체 염색체의 인식률은 92.83%인 반면, 영상 재구성방법을 사용한 경우 염색체 군의 인식률은 100%이고 전체 염색체의 인식률은 95.44%로 MAT방법 보다 우수한 인식률을 나타내었다. 따라서 염색체의 군을 인식하는 입력정보로는 Pattern1을 사용하였으며, HMNN2를 사용한 전체 염색체의 분류 실험결과는 Pattern4와 Pattern5에서는 A, B, E, F 및 G군은 100%로 완전하게 분류할 수 있었으나, C와 D군은 상호 밀접하게 분포되어 있어 인식률이 다소 낮게 나타났다. 따라서 Pattern4는 A, B, E, F 및 G군을 분류하기 위한 패턴조합으로 사용하였으며, C와 D군을 분류하기 위한 패턴조합은 Pattern5를 사용하여 최적화된 패턴인식기를 구현할 수 있었다.

Pattern4와 Pattern5를 사용한 인식 실험결과 전체 염색체에 대한 오분류는 8개로써 약 1.74%의 오분류율을 나타내었으며, 인식률은 약 98.26%의 우수한 인식결과를 나타내었음을 알 수 있었다.

참고문헌

- [1] D. E. Rooney, B. H. Czepulkowski, Human cytogenetics essential data, John Wiley & Sons, 1994.
- [2] J. H. Tjio and A. Levan, "The chromosome number of man", Hereditas, Vol. 42, pp.1-16, (1956).
- [3] John R. Gosden, Chromosome analysis protocols, Humana press, 1994.
- [4] ISCN, An international system for human cytogenetic nomenclature, Karger, 1985.
- [5] Robert S. Ledly, "High-speed automatic analysis of biomedical picture", Science, Vol. 146, pp.216-223, (1964).
- [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] J. Piper, E. Granum, D. Rutovitz and H. Rutledge, "Automation of chromosome analysis", Signal Processing, Vol. 2, pp.203-221, (1980).
- [8] R. C. Gonzalez and Paul Wintz, Digital image processing, Addison- Wesley, 1987.
- [9] 김명남, 조진호, "좌심실에 대한 2D 단축 심초음파 도에서 기하학적 모델을 이용한 심내벽 윤곽선의 자동검출", 대한의용생체공학회 논문지, Vol. 15, No. 4, pp.447-453, (1994).

저자 소개

김충석(Choung-Seok Kim)

1986년 홍익대학교 전산학과 학사
 1988년 홍익대학교 전산학과 석사
 1993년 홍익대학교 전산학과 박사
 1990년 3월 ~현재 신라대학교 컴퓨터정보공학부 교수



남재현(Jae-Hyun Nam)

1989년 부산대학교 컴퓨터공학과 (공학사)
 1992년 부산대학교 컴퓨터공학과 석사

2002년 부산대학교 컴퓨터공학과 박사
 1993년 3월 ~2002년 10월 동주대학 네트워크전자계열 조교수
 2002년 11월 ~현재 신라대학교 컴퓨터정보공학부 전임강사
 ※ 관심분야 : 컴퓨터네트워크, VoIP



장용훈(Yong-Hoon Chang)

1986년 동아대학교 전기공학과 (공학사)
 1990년 동아대학교 전기공학과 석사
 1998년 동아대학교 전기공학과 박사

1989년 ~ 1994년 동아대학교 병원 전산실
 1994년 ~ 현재 동주대학 네트워크전자계열 조교수
 ※ 관심분야 : 영상처리, 패턴인식, 인공지능, 의공학