

FISH 세포영상에서의 군집세포 분할 기법

정미라^{*} · 고병철^{**} · 남재열^{***}

요약

본 논문에서는 입력된 FISH 세포영상을 군집세포영역과 독립세포영역으로 분류하고, 군집세포영역에 대해서는 하나의 세포를 분리하는 알고리즘을 제안한다. 먼저 입력된 영상에 대해서 가우시안혼합모델과 세포의 명암도 값에 대한 최대 우도 함수를 사용하여 세포영역과 배경영역을 분할해줄 임계값을 정의하게 된다. 이렇게 얻어진 전경세포영역에 대해서 보다 정확한 세포 분석을 위해서 군집세포와 독립세포를 분류하게 된다. 세포 영역의 분류과정을 위해서는 베이지안 네트워크와 확률밀도함수를 사용한다. 학습데이터로부터 밀집도(compactness), 평활도(smoothness), 후-모멘트(Hu-moment)에 대한 형태학적 특징값을 추출하여 확률밀도함수를 구성하고, 이를 기반으로 베이지안 네트워크를 사용하여 두 영역을 분류하게 된다. 군집세포로 분류된 영역에 대해서는 그 군집세포를 구성하고 있는 독립세포로 각각 분리한다. 먼저, 명암도 기울기 변환(intensity gradient transform) 영상과 워터셰드 알고리즘을 이용하여 군집세포 영역을 작은 영역으로 분할하게 된다. 작게 분할된 영역을 하나의 세포영역으로 병합시키기 위해서, 군집세포에 존재하는 독립세포의 수만큼의 마커를 결정 침식 연산을 사용하여 추출하고, 추출된 마커를 중심으로 단계적 병합 알고리즘을 제안한다. 본 논문에서 제안한 방법은 166개의 FISH 세포를 사용하여 테스트한 결과 99.29%의 정확한 분리결과를 보여줬으며 기존의 다른 알고리즘보다도 뛰어난 성능과 빠른 실행시간을 보여주었다.

키워드 : FISH 영상, 세포 분류, 세포 분할, 명암도 기울기 변환, 마커, 단계적-병합

Segmentation Method of Overlapped nuclei in FISH Image

MiRa Jeong^{*} · ByoungChul Ko^{**} · Jae-Yeal Nam^{***}

ABSTRACT

This paper presents a new algorithm to the segmentation of the FISH images. First, for segmentation of the cell nuclei from background, a threshold is estimated by using the gaussian mixture model and maximizing the likelihood function of gray value of cell images. After nuclei segmentation, overlapped nuclei and isolated nuclei need to be classified for exact nuclei analysis. For nuclei classification, this paper extracted the morphological features of the nuclei such as compactness, smoothness and moments from training data. Three probability density functions are generated from these features and they are applied to the proposed Bayesian networks as evidences. After nuclei classification, segmenting of overlapped nuclei into isolated nuclei is necessary. This paper first performs intensity gradient transform and watershed algorithm to segment overlapped nuclei. Then proposed stepwise merging strategy is applied to merge several fragments in major nucleus. The experimental results using FISH images show that our system can indeed improve segmentation performance compared to previous researches, since we performed nuclei classification before separating overlapped nuclei.

Keywords : FISH Image, Chromosome Classification, Chromosome Separation Division, Intensity Gradient Transform, Marker, Stepwise-Merging

1. 서론

최근 의료 기술의 발달과 함께 대용량의 의료영상이 급속하게 증가함에 따라, 이를 효과적으로 저장하고 관리할 수 있는 시스템의 필요성이 증대되었다. 이에 따라 대용량 의료 영상을 디지털화하고 하나의 데이터베이스에서 통합 관리하는 시스템인 PACS(Picture Archiving Communication

System)의 관심이 급증하고 있다. PACS와 같이 다양한 의료영상들이 통합 관리되기 위해서는 자동적으로 영상에 대한 분석이 이뤄지고, 그 분석을 바탕으로 해서 각 영상들이 분류하여야 한다. 하지만, 영상의 분석 작업 이전에 반드시 선행되어야 할 과정이 영상에서 관심영역을 정확하게 분할하는 것이다. 특히 X-ray, MRI, CT 영상과는 달리 세포 및 분자영상은 다양한 세포질과 세포핵으로 구성되고 여러개의 세포들이 군집을 이루어 존재함으로써 군집세포영역에서 그 군집세포를 구성하고 있는 독립세포의 정확한 분리선을 찾는 것은 쉽지 않다. 이는 여러 개의 세포가 맞닿아져 군집세포를 구성할 경우에 그 독립세포가 가지던 고유한 모양이 변형되기 때문이다.

*준회원: 계명대학교 컴퓨터공학과 석사과정
**정회원: 계명대학교 컴퓨터공학과 조교수
***정회원: 계명대학교 컴퓨터공학과 교수
논문접수: 2008년 8월 11일
수정일: 1차 2009년 1월 15일
심사완료: 2009년 2월 12일

세포영역을 자동으로 분할하기 위한 기존의 연구들을 살펴보면 서미숙 [1]등은 일반 세포영상에서 중요한 세포영역들만 자동으로 추출하기 위해 관심윈도우와 병합 기법을 이용하여 세포를 분할하는 알고리즘을 제안하였다. Umesh Adiga [2]등은 독립세포의 추출을 독립세포영역과 군집세포영역의 분류 없이 전체 세포영상에 대해서 워터셰드 알고리즘을 적용하고, 분할된 영역에 대해서는 일정 크기보다 큰 분할영역은 하나의 세포영역으로 정의하고, 그 일정 크기보다 작은 분할영역에 대해서는 그 주위의 가장 많이 맞닿아 있는 큰 영역과 병합시키는 과정을 반복하여 하나의 세포영역을 추출한다. 하지만, 이 방법은 독립세포영역과 군집세포영역의 구분 없이 전체 세포영역에 대하여 독립세포추출 알고리즘을 실행함으로써 시간과 자원을 낭비하게 되고, 너무 간단한 병합 알고리즘으로 그 성능이 저하된다. Glory [3]등은 세포영역을 세포의 밀집정도에 따라 분류하고, 워터셰드 알고리즘과 간단한 규칙을 통해서 군집세포영역을 독립세포영역으로 분할한다. 하지만 세포영역을 밀집정도에 따라 분류하는 것은 세포의 크기가 일정했을 때만 가능한 것이기 때문에 그 한정성을 가진다. 또한 워터셰드 알고리즘을 통한 과분할 문제를 간단한 하나의 수식으로 병합하는 것은 그 정확성을 많이 낮추게 한다. Raman [4]등은 세포영역의 볼록성(Convexity) 특징값으로 군집세포와 독립세포를 분류하고, 군집세포로 분류된 영역에 대해서는 세포윤곽의 포인터 중 최상의 곡률(curvature)값을 가지는 몇 개의 포인터를 독립세포의 분할을 위한 초기포인터로 정한다. 초기포인터 중에서 역평행성(Antiparallel), 비교차(Non-intersecting) 제약조건을 만족하는 포인터가 최종적으로 군집세포를 독립세포로 분할하는 분할 포인터가 된다. 하지만 이 방법도 하나의 특징값만으로 세포영역을 분류하는 것은 정확도가 떨어지고, 군집세포의 분할포인터를 결정하는 과정에서도 초기포인터를 설정하는 곡률값의 임계값 설정은 매우 난해하며, 역평행성과 비교차 조건도 너무 단순하고 일정 군집세포모델에 한정되어 있어 정확한 독립세포영역의 추출성능을 저하시킨다. Chang [5]등은 정확한 세포영역을 추출하기 전에 대충적인 세포영역을 찾아서 반복적인 방사형 보팅(Radial Voting)알고리즘을 사용하여 세포의 중심위치를 얻고, 이 얻어진 각 세포의 위치로부터 보로노이 배열(Voronoi tessellation)을 이용하여 세포영역의 보로노이 영역(Voronoi region)을 정의하고, 변형된 동적 윤곽선 모델(Active Contour model)을 적용하여 최종적인 하나의 세포영역을 추출하게 된다. 즉, 이 알고리즘은 세포의 중심위치를 먼저 찾은 후에 세포의 중심위치를 기반으로하여 군집세포영역에서 하나의 세포영역을 추출하고 있다. 하지만 이것은 세포영상에서 배경과 세포영역을 정확한 분리 후에 세포의 중심을 찾는 것이 아니기 때문에 불분명하게 명암값을 가지는 영역에 대하여 잘못된 세포영역으로 인식하고 세포의 중심을 찾게 되는 오류가 발생한다. 따라서 본 논문에서는 좀더 정확한 세포영역의 분류와 분할방법을 제안하고자 하였다.

본 논문에서는 이전 연구[6]의 결과성능을 개선하여 여러

가지 세포영상 중 FISH(Fluorescence in situ hybridization) 영상에서 독립세포와 군집세포를 구분하고 군집세포영역을 독립세포영역으로 분할하는 알고리즘을 제안한다. FISH는 염색체 이상 질환을 정확하게 진단할 수 있는 진단법으로 특정 DNA 서열에 형광으로 표지된 탐침(Probe)을 결합시켜 염색체의 이상 유무를 관독하는 검사법이다. 이로 인해서 FISH영상은 염색체에서 여러색으로 염색된 부분이 존재하며 (그림 1)과 같은 형태를 가진다. 세포영역에 존재하는 점들은 염색체가 염색된 결과로, 이 결과에 따라서 다운증후군, 에드워드 증후군, 파타우 증후군등 염색체 이상으로 생긴 유전적인 질병들을 미리 예측하게 된다. 이 질병들은 염색체의 수적으로 문제가 발생하는 질병이다. 현재의 FISH영상의 분석은 병리전문의가 영상을 직접보고 염색된 염색체를 관찰 후에 질병유무를 판단하지만, 세포영역을 하나의 독립세포로 자동적으로 분할할 수 있다면 추출된 독립세포에서 염색된 염색체의 수를 자동적으로 체크하여 질병유무를 판단할 수 있다.

이전 연구에서 문제점으로 지적된 분할 시간과 정확율을 높이기 위해서 본 논문에서는 독립세포영역과 군집세포영역의 분류과정을 거친 후, 독립세포를 추출하는 알고리즘을 적용하게 된다. 먼저 세포영상에서 정확한 독립세포의 추출을 위해서 세포영역과 배경영역을 분할하고, 추출된 세포영역들을 군집세포영역과 독립세포영역으로 분류하기위해서 형태학적 특징벡터로부터 확률모델을 생성하고 이를 기반으로 베이지안 네트워크를 이용하게 된다. 여기서 군집세포로 분류된 영역에 대해서는 군집세포를 구성하고 있는 각각의 독립세포영역을 추출하기 위해서 먼저 해당 독립세포의 마커를 찾고, 해당영상에 전처리과정을 거친 후 워터셰드 알고리즘을 이용해서 군집세포영역을 작게 분할한다. 분할된 영역은 각 마커를 중심으로 몇 가지의 규칙기반 병합 알고리즘을 이용하여 하나의 독립세포를 추출하게 된다.

본 논문의 구성은 다음과 같다. 2장에서는 입력된 세포영상으로부터 배경영역을 분리하고, 추출된 전경영역에 대해서 세포영역의 분류과정을 서술한다. 3장에서는 각 세포의 마커추출, Watershed알고리즘, 단계적 병합 알고리즘을 사용하여 군집세포영역에서 독립세포를 추출하는 과정을 설명



(그림 1) FISH 영상의 예

한다. 또한 4장은 성능평가를 위해서 다른 세포분할 알고리즘과 비교하여 본 논문에서 제안된 알고리즘의 우수성을 입증하고, 마지막으로 5장에서는 제안된 알고리즘에 대한 전체적인 정리를 하면서 본 논문을 마무리 할 것이다.

2. 확률 밀도 함수와 베이지안 네트워크를 이용한 FISH 영상 분류

2.1 세포영역 분할

FISH영상에서 세포는 B채널에서 나타나고, 세포내의 염색된 염색체 부분은 R채널과 G채널에서 나타난다. 하지만 FISH영상에서 염색된 부분은 차후의 세포를 분석을 위해서는 중요한 정보이지만, 본 논문에서 제안하는 하나의 독립세포를 추출하는 과정에서는 오히려 세포를 분류하고 분할하는 과정을 방해하는 요소이다. 최적의 결과를 얻기 위한 입력영상은 염색된 부분은 작으면서 세포의 중심으로 갈수록 명암도가 밝아지고, 세포 간의 맞닿는 부분에서의 명암도의 차이가 클 때 최적의 결과를 얻을 수 있다. 이점을 고려한 최적의 영상을 구성하기 위해서 본 논문에서는 입력된 컬러 영상의 R, G, B 채널 중에서 G, B채널만을 이용하여 그레이 영상을 구성하고, 그레이 영상에서의 염색된 부분과 노이즈를 줄이기 위해서 미디언 필터(Median Filter)와 침식(Erosion) 연산을 사용하여 앞으로 사용되어질 영상을 구성하게 된다. 이렇게 구성된 그레이 영상에서 세포영역과 배경영역을 분할하기 위한 임계값 설정은 가우시안 혼합 모델(Gaussian Mixture Model)과 최대 우도 함수(Maximum Likelihood Function)[7]를 이용한다. 최대우도함수를 이용하여 결정된 임계값을 중심으로 (그림 2)와 같은 이진 영상을 구할 수 있다.

2.2 형태학 특징 추출과 파즌 창(Parzen window)을 이용한 확률 밀도 추정

앞서 추출된 세포영역은 군집된 세포영역과 독립된 세포영역으로 구성된다. 기존연구에서는 군집세포와 독립세포의 구분없이 모든 세포에 대해서 세포분할을 시도함으로써 분할시간이 오래 걸리고, 결과 성능이 떨어지는 단점이 발생한다. 본 논문에서는 이전 방법과는 달리 영상 내에 독립세포와 군집세포를 미리 분류하고 군집세포영역에 대해서는 분할알고리즘을 수행하여 실행속도를 향상시키고 분할정확

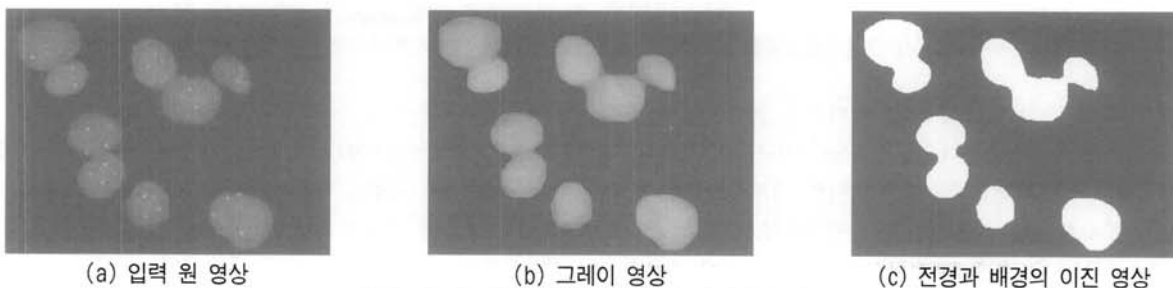
도를 높일 수 있는 방법을 제안하고자 한다. 세포영역의 분류과정의 학습데이터로부터 6개의 형태학 특징을 추출하고, 이 특징들을 기반으로 계층적 베이지안 네트워크를 사용하여 군집세포영역과 독립세포영역을 정확하게 분류하게 된다.

형태학적 특징은 객체의 모양이나 크기 등을 나타내는 특징으로서 객체의 원주, 면적, 둘레, 밀집도(compactness), 평활도(Smoothness), 볼록성(Convexity), 이심률(Eccentricity) 등이 이에 속한다. 본 논문에서는 FISH 영상의 일정하지 않은 크기와 둥근모양의 세포의 특징을 이용하여, 군집세포와 독립세포를 가장 이상적으로 분류해줄 특징 6개를 실험을 통하여 밀집도, 평활도, 후-모멘트(Hu-moment) 1, 2, 3, 4 [8]를 결정하였다. 먼저 각각의 25개의 군집세포와 141개의 독립세포로 이루어진 데이터군집으로부터 총 6개의 특징을 추출하게 되고, 수집된 각 특징벡터는 데이터의 분포 형태가 일정하지 않고 일반적으로 많이 사용되는 가우시안 분포 형태를 보인다고 할 수 없으므로, 본 논문에서는 각 데이터로부터 추출된 특징벡터에서 확률분포를 추정하기 위해 비모수(non-parametric)방법인 히스토그램을 구성하고, 파즌 창 중 가우시안 스무드 커널(Gaussian Smooth-Kernel)을 적용하여 보다 정밀한 확률밀도를 추정한다. 가우시안 스무드 커널을 적용하기에 앞서 추출된 군집세포영역과 독립세포영역의 특징벡터를 분리하여 각각 특징벡터를 0~1의 값으로 정규화 시킨다. (그림 3)은 특징벡터를 가우시안 스무드 커널을 이용해 얻어진 확률 밀도 추정 결과를 보여주고 있다.

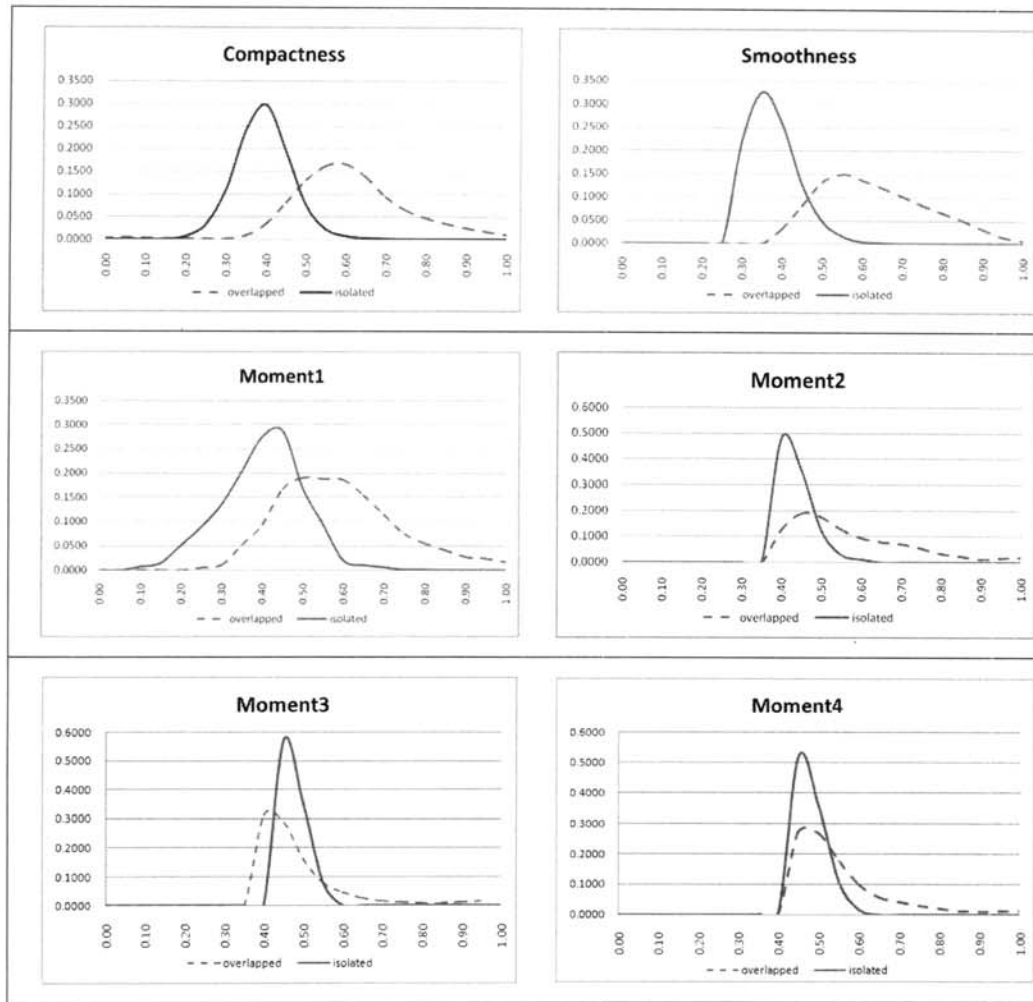
2.3 베이지안 네트워크를 이용한 세포영역의 분류

베이지안 네트워크는 기본적으로 이산 확률 변수들의 조건부 확률 관계를 방향성 비순환 그래프(DAG: directed acyclic graph)와 조건부 확률 테이블로 가지고 있으며 이 확률 테이블은 파라미터간의 확률적 의존 관계를 나타내는 그래프와 각 파라미터별 조건부 확률로 구성된다. 군집세포와 독립세포를 분류하기 위해 본 논문에서 제안하는 계층적 베이지안 네트워크의 구성은 (그림 4)와 같다.

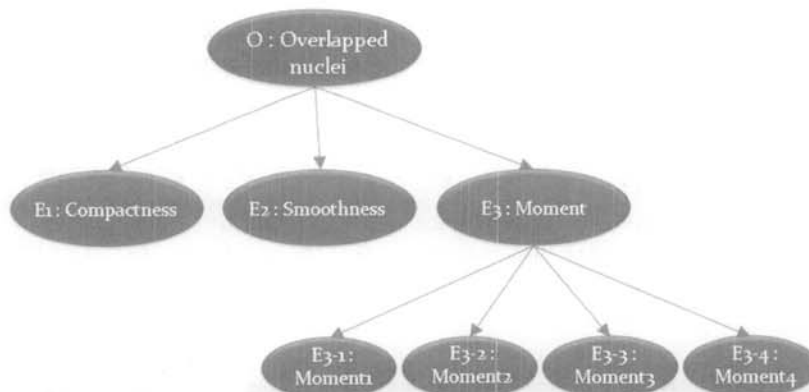
(그림 4)에서와 같이 본 논문에서 제안하는 계층적 베이지안 네트워크는 6개의 관찰노드(observation-node : $E_1, E_2, E_3, E_{3-1}, E_{3-2}, E_{3-3}, E_{3-4}$)와 1개의 서브네트워크(sub-network)로 구성된다. (그림 4)에서 루트 노드 O 는 3개의 자식 노드 E_1, E_2, E_3 를 가지며, 그 중 자식 노드 E_3



(그림 2) 원 영상에서 세포영역과 배경영역의 분할



(그림 3) 밀집도, 평활도, 모멘트 1,2,3,4의 확률밀도 그래프



(그림 4) 군집세포영역을 분류하기 위한 계층적 베이저안 네트워크의 흐름 그래프

는 3개의 자식 노드로 구성된다. 베이저안 분류기에 의한 루트 노드의 확률 밀도 $P(O|E_1, E_2, E_3)$ 는 (그림 4)와 같이 자식 노드의 조건부 확률로부터 추측되어 지고, 주어진 E_1, E_2, E_3 은 주어진 O 와 조건적으로 독립적이다. 그것은 베이저안 규칙에 의해서 군집세포영역의 사후 확률 값이 추론되어 지고, 노드 E_3 의 조건부 확률은 (그림 4)의 자식 노드로

부터 유추되어 진다. 서브네트워크의 목표노드인 E_3 는 '군집 후보(E_3)'와 '독립후보($\neg E_3$)'의 두가지의 상태를 가지며, 각 상태의 조건부 확률을 가지는 식(3)에 의해 구해진다. 만약 루트노드 O 가 경쟁적으로 대응하는 요소를 가지지 않는다면, 루트 노드의 반대는 $\neg O$ 가 고려되어 질 수 있다. 그리고 그것은 주어진 조건부 확률과 다음 식(1)에 의한 사전확률

을 이용해서 루트노드 O 의 사후확률을 계산할 수 있다. $P(O)$ 는 군집세포영역에 대한 사전 확률이고, $P(\neg O)$ 는 독립세포일 때의 사전확률이다.

$$P(O | E_1, E_2, E_3) = \frac{P(E_1, E_2, E_3 | O) P(O)}{P(E_1, E_2, E_3 | O) P(O) + P(E_1, E_2, E_3 | \neg O) P(\neg O)} \quad (1)$$

식(1)은 각 사건들이 상호 배타적이므로 다음과 같이 다시 나타낼 수 있다.

$$P(O | E_1, E_2, E_3) = \frac{\prod_{i=1}^3 P(E_i | O) P(O)}{\prod_{i=1}^3 P(E_i | O) P(O) + \prod_{i=1}^3 P(E_i | \neg O) P(\neg O)} \quad (2)$$

식(2)의 파라미터 $P(E_1 | O), P(E_2 | O), P(E_3 | O)$ 와 $P(E_1 | \neg O), P(E_2 | \neg O), P(E_3 | \neg O)$ 는 앞서 제안된 훈련 데이터로부터 학습된 비모수적 확률 밀도 추정 모델에서 구해질 수 있다. 중간노드 E_1, E_2, E_3 의 조건부 확률로부터 군집세포영역에 대한 사후 확률이 추측되어지고, 그것이 0.5이상이라면 실제 군집세포영역으로 분류하게 된다. 이때의 E_3 의 확률 분포는 자식 노드 $E_{3-1}, E_{3-2}, E_{3-3}, E_{3-4}$ 의 결합 확률로부터 식(3)과 같이 추측할 수 있고, 자식노드 $E_{3-1}, E_{3-2}, E_{3-3}, E_{3-4}$ 의 확률 분포는 식(4)에서 유추할 수 있다.

$$P(E_3 | E_{3-1}, E_{3-2}, E_{3-3}, E_{3-4}) = \frac{\prod_{i=1}^4 P(E_{3-i} | E_3) P(E_3)}{\prod_{i=1}^4 P(E_{3-i} | E_3) P(E_3) + \prod_{i=1}^4 P(E_{3-i} | \neg E_3) P(\neg E_3)} \quad (3)$$

$$P(E_{3-1:4}, E_3) = \frac{P(E_{3-1:4} | E_3) P(E_3)}{P(E_{3-1} | E_3) P(E_{3-2} | E_3) P(E_{3-3} | E_3) P(E_{3-4} | E_3)}$$

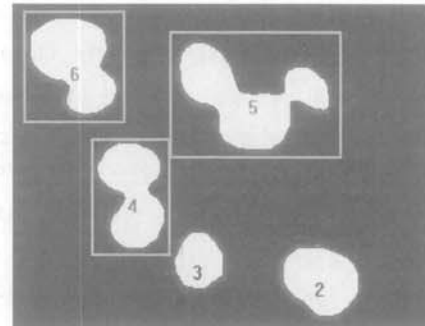
$$P(E_{3-1:4}, \neg E_3) = \frac{P(E_{3-1:4} | \neg E_3) P(\neg E_3)}{P(E_{3-1} | \neg E_3) P(E_{3-2} | \neg E_3) P(E_{3-3} | \neg E_3) P(E_{3-4} | \neg E_3)} \quad (4)$$

<표 1>은 계층적 베이지안 네트워크를 이용한 각각의 노드에 대한 설명이다.

최종적으로 입력된 세포영역으로부터 6가지의 특징을 추출하고 확률 모델로부터 각 특징 값의 확률밀도를 추정하여 식(2)를 통해 군집 혹은 독립세포를 구분하게 된다. (그림 5)는 베이지안 네트워크를 이용하여 군집세포영역을 분류한 결과를 보여주며, 사각박스로 표시된 영역이 군집세포로 분류된 세포를 의미한다.

<표 1> 계층적 베이지안 네트워크에 사용되는 각 노드의 설명

Node	Description	
O	군집 세포	
$\neg O$	독립 세포	
E_1	밀집도(Compactness)	
E_2	평활도(Smoothness)	
E_3	Hu-Moment for Overlapped nuclei	
	E_{3-1}	모멘트1
	E_{3-2}	모멘트2
	E_{3-3}	모멘트3
	E_{3-4}	모멘트4



(그림 5) 베이지안 네트워크를 사용한 배경 분할과 최종 군집세포영역의 분류 결과

3. 군집세포영역에서 독립세포 추출

세포영역의 분류 후, 영역기반 접근을 사용하여 군집세포영역을 독립세포영역으로 분리한다. 이를 위해, 첫 번째로 각 군집세포영역으로부터 개선된 Duchene[9]의 알고리즘을 사용하여 마커가 추출된다. 두 번째로 명암도 기울기 변환(Intensity Gradient Transform)을 이용하여 전처리과정을 거친 영상에 대해서 워터셰드 알고리즘을 적용하여 군집세포영역을 작게 분할한다. 세 번째로 작게 분할된 군집세포영역을 앞서 찾은 마커를 중심으로 제안된 병합 알고리즘을 이용하여 군집세포영역에서 보다 정확한 독립세포영역을 분리하게 된다.

3.1 마커(Marker) 추출

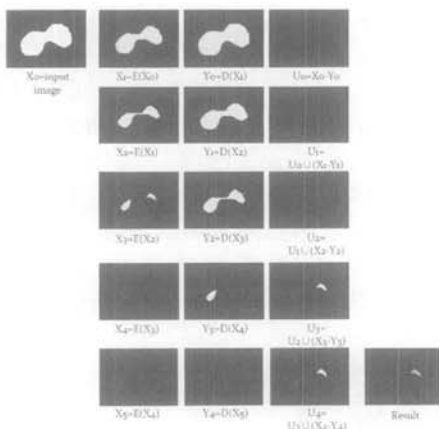
성공적인 군집세포영역을 독립세포영역으로 분리하는데 있어서 정확한 마커의 추출은 필수적이다. 하나의 군집세포영역 안에는 하나 이상의 독립세포가 존재하므로 군집세포가 포함하는 독립세포 수만큼의 마커를 찾아야 한다. 정확한 마커 추출을 위해서 본 논문에서는 Duchene[9]의 결정 침식(Ultimate Erosion) 연산을 적용하였다. 결정 침식 연산은 반복적인 침식연산 과정으로 세포영역에 침식 연산을 반복적으로 적용하게 되면 어느 단계에서는 하나의 세포영역

이 사라지는 시점이 생기게 된다. 영역이 사라지기전의 영역을 마커라 정의하고 이를 찾는 과정은 다음과 같다. 결정 침식 연산을 적용하기 전, 분할된 이진 세포영상에서 거리 지도(Distance Map)를 구성하고, 명암임계값 이하를 가지는 픽셀은 버린다. 이 결과 영상에 대해서 다음과 같은 결정 침식 연산을 적용하게 된다.

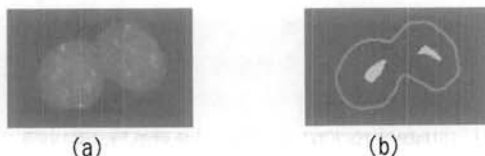
- 1) $X_0 = \text{Input Image}$
- 2) $X_{i+1} = E(X_i)$
- 3) $Y_i = X_i \cap D(X_{i+1})$
- 4) $U_i = U_{i-1} \cup (X_i - Y_i)$ (5)

X_{i+1} 는 영상 X_i 에 침식연산(E)을 적용한 영상이고, Y_i 는 X_{i+1} 와 X_i 의 팽창(D) 연산 결과와 겹쳐지는 영역이다. 이것은 X_{i+1} 영상을 팽창 연산 후의 영역이 원래의 X_i 영상의 영역보다 커지면 X_i 영상과 Y_i 영상의 차의 결과가 팽창된 영역이 나올 수 있기 때문이다. U_i 는 이전 단계의 U_{i-1} 집합과 현재 X_i , Y_i 의 차영상의 집합으로, X_i 영상에서 Y_i 영상을 제외하여 추출되는 결과영역들이 마커가 된다. 여기서 U_{i-1} 의 초기값은 공집합이다. 최종적으로 마커영역이 남게 되고 이 영역들은 U_i 집합의 원소가 된다. 위의 과정을 군집 세포영역에서 세포영역의 픽셀이 사라질 때까지 2, 3, 4의 과정을 반복하여 입력된 군집세포영상으로부터 최종 마커를 획득하게 된다. (그림 6)에서는 결정 침식 연산의 과정을 자세히 보여주고 있다.

(그림 7)은 군집세포영역에서 최종 마커를 추출한 결과를 보여준다.



(그림 6) 결정 침식 연산을 이용한 마커 추출 과정



(그림 7) 결정 침식 연산을 이용한 마커 추출 결과 화면: (a) 원영상, (b) 마커 추출 결과

3.2 명암도 기울기 변환 영상을 이용한 워터셰드 알고리즘
워터셰드 알고리즘을 이용하여 군집세포영역을 작게 분할할 때 FISH영상의 염색된 DNA로 인해서 군집세포영역에 존재하는 각 독립세포영역의 경계부분이 정확하게 분리되지 못하는 문제가 야기된다. 따라서 본 논문에서는 워터셰드 알고리즘을 적용하기 전, Gang Lin[11]의 전처리 알고리즘을 개선하여 문제를 해결하였다.

배경 분할알고리즘에 의해 분할된 전경영상에서 G, B 채널과 가우시안 스무드필터를 이용하여 구성된 그레이 영상에 대해서 다음과 같은 알고리즘들을 적용하게 된다.

먼저 전경영상(I)에 대해서 팽창(Dilation)연산과 침식(Erosion)연산의 차이를 이용하여, 명암도 기울기(Intensity Gradient), G 를 정의하고 $I, G \in \{R^d | 0, 1, \dots, 255\}$ 과 같은 조건을 만족한다.[11]

$$G = |\text{GrayDilation}(I) - \text{GrayErosion}(I)| \quad (6)$$

일반적인 그레이-팽창 연산과 그레이-침식 연산의 마스크는 사각형을 사용하지만, 본 연구의 테스트 영상의 객체는 둥근 원모양을 가지므로 사각형 마스크는 적합하지 않다. 그러므로 각각의 연산을 위한 마스크는 (그림 8)과 같이 둥근 마스크로 구성된다. 명암도 기울기 G 를 생성한 뒤에 Gang Lin의 방법에서 사용한 거리 변환(Distance Transform) 대신에 동일한 값을 가지는 실수 δ 를 결합하여 식(7)을 통해 G' 영상을 구성하게 된다. 이렇게 구해진 G' 영상은 (그림 9)-(a)와 같다. 아래 식(7)에서 G_{\max} 와 G_{\min} 은 G 로부터 구해진 최대값 및 최소값을 의미한다.

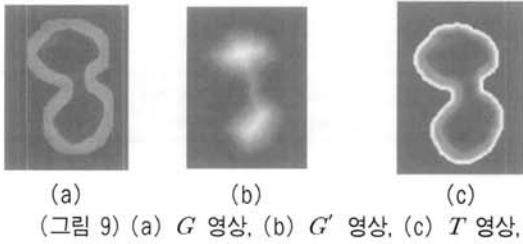
$$G' = \delta \times \exp(1 - [(G - G_{\min}) / (G_{\max} - G_{\min})]) \quad (7)$$

워터셰드 알고리즘의 특성상 군집세포영역을 분리하기 위해서는 세포중심의 명암도 가장 낮고 그 곳으로부터 물을 채워나가면서 담수벽을 형성해야 한다. 따라서 G' 영상은 (그림 9)-(c)와 같이 반전을 필요로 하게 되고, 식(9)를 통해서 반전된 영상 T 를 구성하게 된다.

$$\begin{aligned} \text{if } G'[i] = Bg \\ \text{then } G'[i] = [MAX(G') + 20] \end{aligned} \quad (8)$$

0	0	1	1	1	0	0
0	1	1	1	1	1	0
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
0	1	1	1	1	1	0
0	0	1	1	1	0	0

(그림 8) 그레이-팽창 연산과 그레이-침식 연산을 위한 7x7 마스크



(그림 9) (a) G 영상, (b) G' 영상, (c) T 영상.

$$T = S_g(\text{MAX}(G') - G') \quad (9)$$

식(8)에서 우선 i 번째 좌표의 $G'[i]$ 값이 배경 영역(Bg)인지 우선 살펴보고, 만약 배경 영역에 해당된다면 해당 좌표를 반전시키기 전에 식(8)을 적용하여 새로운 $G'[i]$ 값을 생성한다. 세포영역을 제외한 배경 영역(Bg)에 대해서 G' 의 최고값에 20을 더한 값을 대입시키는 이유는 반전 시켰을 경우 전경과 구분되는 0에 가까운 값을 갖게 하기 위함이다. 배경 영역에 대해 식(8)을 적용시킨 후에, 모든 G' 에 대해 식(9)을 적용시켜 반전된 영상을 얻게 된다. 식(9)에서 S_g 는 가우시안 스무드 연산을 나타내고, 이는 영상에서 작은 노이즈를 제거하기 위해서 적용하게 된다. 이렇게 구성된 T 영상은 (그림9)-(c)와 같고, T 영상에 대해 최종적으로 워터셰드 알고리즘을 적용하게 된다.

3.3 단계적 병합 알고리즘을 통한 독립세포영역 분리

앞서 추출된 T 영상에 워터셰드 알고리즘을 적용하게 된다. 하지만 워터셰드 알고리즘은 과분할 문제점을 가지고 있다.[11] 본 논문에서는 이를 해결하기 위해서 다음과 같은 단계적-병합 기법을 제안한다. 단계적-병합 기법은 워터셰드 알고리즘으로 생긴 각 영역들에 대해서 각각 다음 3가지 조건을 기반으로 하여 병합시키면서 하나의 독립세포를 구성하게 된다.

워터셰드 분할 영역의 집합을 R 이라 하면 ($R = \{r_1, r_2, r_3, \dots, r_m\}$) 집합 R 의 영역 중에서 앞서 찾은 마커와 겹쳐지는 영역의 집합을 $M(M = \{m_1, m_2, \dots, m_k\})$ 을 마커영역으로 정의한다. 이때 m 은 분할된 영역의 개수, k 는 이전단계에서 찾은 마커의 개수(군집세포영역의 개수)이다. 첫 번째 단계로 하나의 영역 R 에 대해서 다음 조건들을 적용시키게 된다.

(조건 1)

$$h = \text{MAX}(\text{Hist}_k[w])$$

$$\text{if } (h < \rho) \theta_k = h$$

$$\text{else } \theta_k = \rho$$

$$\text{if } (d_{km} < \theta_k), m_k = r_m \cup m_k$$

첫 번째 조건은 마커를 중심으로 영역 r_m 까지의 일정 거리 임계값(θ_k)을 설정하고, 그 거리 임계값에 속하는 워터셰드 분할에 대해서 해당 마커영역으로 병합시키게 된다. 임계값 θ_k 의 설정은 각 마커를 중심으로 군집세포영역의 경계면에 대한 거리(h)를 측정하고, 그 거리에 대한 히스토그램 $\text{Hist}_k[w]$ 를 계산하여 최대 빈도수를 정의한 후, 가장 많은 빈도를 가지는 거리를 해당 마커 m_k 의 거리 임계값으로 설정한다. (그림 10)은 마커를 중심으로 군집세포영역의 경계면에 대한 거리 히스토그램을 나타낸다.

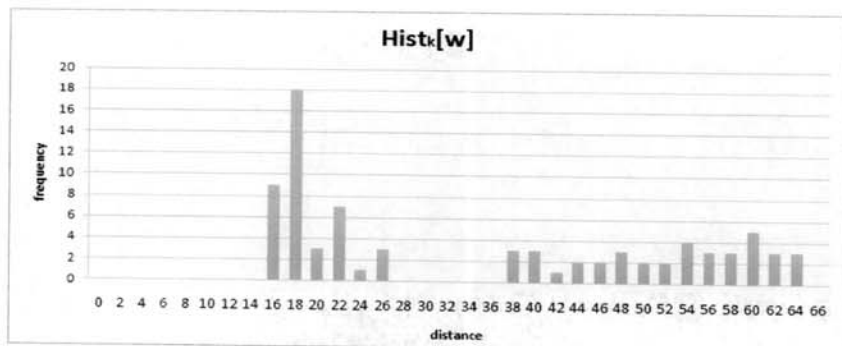
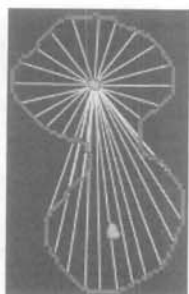
(그림 10)에서 보듯이 마커를 중심으로 각 경계면의 거리에 대한 히스토그램이 만들어지고 가장 빈도수가 많은 거리 값 18이 마커 m_k 를 위한 임계값 θ_k 로 결정된다. 여기서 이 임계값은 $\rho(40)$ 이상이 되어서는 안된다. 이것은 본 논문에서 사용한 FISH 영상의 세포의 최대 반경이 40을 넘지 않기 때문이다. 각 마커와 워터셰드 분할 영역까지의 거리값 d_{km} 이 θ_k 에 속한다면 현재 영역 r_m 은 마커 m_k 로 병합되고 마커 m_k 의 영역은 다시 갱신하게 된다.

영역 r_m 에 대해서 첫 번째 조건을 만족하지 않았을 경우, 두 번째 조건은 영역 r_m 에서 마커 m_k 와의 맞닿은 픽셀을 조사하게 되고, 현재 영역 r_m 과 더 많이 맞닿은 마커 m_k 로 영역 r_m 을 병합하게 된다.

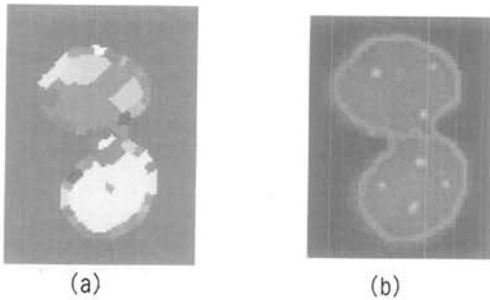
(조건 2)

$$m_k' = m_k \cup r_m, k = \{\forall k | \text{MAX}(\text{touchPixel}(r_m \cap m_k))\}$$

임의의 순서로 체크되는 워터셰드 분할 영역 r_m 이 일정 거리 임계값 안에 속하지 않고, 맞닿은 픽셀도 없을 경우가 발생할 수 있다. 이 경우는 (조건 3)을 기반으로 하여 현재 r_m 영역에서 가장 가까운 마커 m_k 로 병합하여 m_k' 을 갱신한다.



(그림 10) 마커를 중심으로 군집세포영역의 경계면에 대한 거리 히스토그램



(그림 11) (a) 워터셰드 알고리즘으로 생긴 과분할된 영역, (b) 단계적-병합 기법을 적용한 결과 영상

(조건 3)

$$m_k' = m_k \cup r_m, k = \{\forall k | \min(d_{km})\}$$

각 영역 r_m 에 대해서 위 조건을 집합 R 의 원소 개수와 집합 M 의 원소 개수가 동일해 질 때까지 반복시키면서 적용시킨다. 위의 과정을 반복하게 되면 (그림 11)의 (a)와 같이 워터셰드 분할된 영역들이 위 병합과정을 통하여 (b)와 같이 정확하게 추출된 독립세포영역을 얻게 된다.

(그림 12)는 최종적으로 독립세포를 추출한 결과화면이다. (그림 12)는 독립세포로 분류된 영역은 회색선으로 나타내었고, 군집세포로 분류된 영역은 흰색선으로 나타내었다.

4. 실험 결과 및 분석

본 논문에서 제안하는 방법의 실험은 개인용 펜티엄 PC(CPU 2.40GHz)와 Windows기반의 Visual C++ 6.0을 이용하여서 개발하였다. 실험은 320x240사이즈의 FISH(Fluorescence in situ hybridization) 영상 데이터를 사용하였다. FISH영상 안에는 25개의 군집세포와 141개의 독립세포를 포함한다.

본 논문이 제안하는 방법의 세포분할 성능을 평가하기 위해서 세포영역의 분류과정이 없는 Gang Lin의 알고리즘 <표 2>과 본 논문에서 제안한 방법 <표 3>을 비교하였다. 각 알고리즘의 평가 항목으로는 'Correctly segmented', 'Incorrectly merged', 'Over-segmented', 'Fused', 'Ignored', 'Segmentation Time'을 기반으로 해서 비교하게 된다. 'Incorrectly Merged'는 마커를 정확하게 찾았지만, 좋지 못한 병

<표 2> 세포분류 과정을 제외한 Gang Lin의 기율기 가중치 거리 변환 영상을 사용한 세포분할 성능

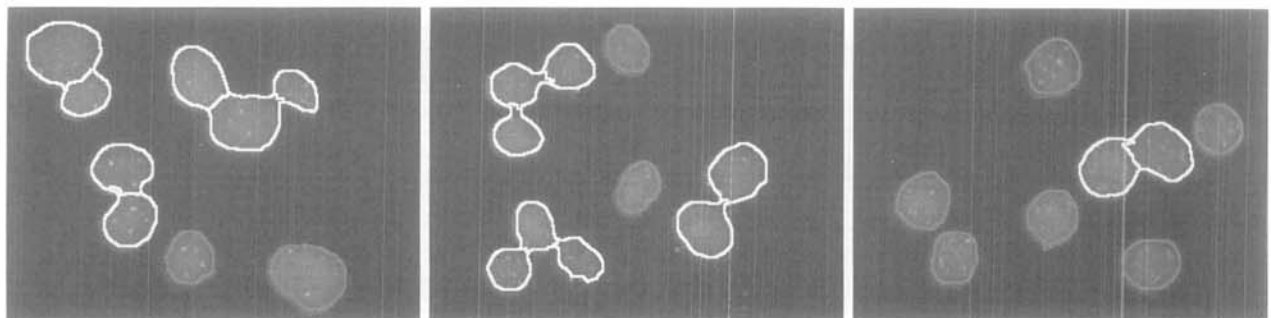
Number of nuclei	Isolated nuclei	Overlapped nuclei	All cases
Analyzed	141	25	166
Correctly segmented	138(97.87%)	10(40%)	148(89.16%)
Incorrectly merged	0(0%)	9(36%)	9(5.42%)
Over-segmented	3(2.13%)	0(0%)	3(1.81%)
Fused	0(0%)	6(24%)	6(3.61%)
Ignored	0(0%)	0(0%)	0(0%)
Segmentation Time	13.984 sec		

<표 3> 본 논문에서 제안하는 알고리즘의 세포분할 성능

Number of nuclei	Isolated nuclei	Overlapped nuclei	All cases
Analyzed	141	25	166
Correctly segmented	140(99.29%)	21(84%)	161(96.99%)
Incorrectly merged	0(0%)	0(0%)	0(0%)
Over-segmented	1(0.71%)	1(4%)	2(1.20%)
Fused	0(0%)	3(12%)	3(1.81%)
Ignored	0(0%)	0(0%)	0(0%)
Segmentation Time	4.748 sec		

합 결과를 가져온 경우를 나타낸다. 'Over-segmented'는 하나의 독립세포영역에서 두 개 이상의 마커를 찾을 경우를 나타내고, 'Fused'은 군집세포영역에서 실제로 존재하는 독립세포의 수보다 작은 수의 마커를 찾을 경우를 나타낸다. 또한 'Ignored'는 마커를 찾지 못했을 경우를 나타낸다[10].

<표 2>와 <표 3>을 비교해보면 본 논문에서 제안한 방법이 모든 부분에서 뛰어난 성능과 빠른 실행시간을 가지며, 특히 군집세포영역을 분할하는데 있어 훨씬 높은 성능을 보임을 알 수 있다.



(그림 12) 세포영상에서 독립세포영역의 추출 결과 화면

5. 결 론

본 논문에서 제안한 알고리즘은 세포영상에서 세포영역을 배경과 분리하고, 분리된 세포영역에 대해서 독립세포와 군집세포로 분류한다. 군집세포로 분류된 영역에 대해서는 다시 그 군집세포영역을 구성하는 독립세포를 하나하나씩 분리하는 알고리즘을 제안하고 있다. 본 논문에서 중점적으로 제안하는 것은 군집세포를 독립세포로 더욱 정확한 독립세포간의 분할선을 추출해내는 것이다. 각 독립세포들이 모여서 군집영역을 구성하므로써 독립세포의 고유한 모양을 유지하지 못한 채 군집영역을 이루게 된다. 또한 각 독립세포가 맞닿은 부분에서의 명암도가 일정하게 변하거나 뚜렷하지 못하므로 군집세포영역에서 독립세포를 분리하는 것은 쉬운 작업이 아니다. 이에 따라서 본 논문에서는 명암도 기울기 변환 영상을 구성하여 워터셰드 알고리즘을 적용한 결과 세포와 세포가 맞닿은 부분을 정확하게 구분해주면서 작은 영역들로 분할되게 하였다. 또한 작게 분할된 영역들이 각 마커를 중심으로 정확한 독립세포를 추출하기 위해서 3가지의 규칙을 기반으로 단계적으로 병합되는 알고리즘을 제안하였다. 이는 군집세포에서 독립세포를 추출하는 기존에 제안된 논문보다 빠른 시간 안에 정확한 분할선을 찾는 것을 확인할 수 있었다.

감사의 글

본 연구는 지식경제부 지방기술혁신사업(RTI04-01-01) 지원으로 수행되었습니다.

참 고 문 헌

- [1] 서미숙, 고병철, 남재열, "적응적 관심윈도우 기반의 세포영상 분할 기법," 한국정보처리학회논문지B, 제14-B권, 2호, pp.99-106, 2007.
- [2] P. S. Umesh Adiga, B. B Chaudhuri, "An efficient method based on watershed and rule-based merging for segmentation of 3-D histo-pathological images," Pattern Recognition, Vol.34, pp.1449-1458, 2001.
- [3] E. Glory, V. Meas-Yedid, G. Stamon, Ch. Pinsest, J.-Ch. Olivo-Marin, "Automated Image-Based Screening of Cell Cultures for Cell Therapy," IEEE International Symposium, pp.259-262, 2006.
- [4] S. RAMAN, C. A. MAXWELL, M. H. BARCELLOS-HOFF, B. PARVIN, "Geometric approach segmentation and protein localization in cell culture assays," Journal of Microscopy, Vol.225, pp.22-30, 2007.
- [5] H. Chang, Q. Yang, and B. Parvin, "Segmentation of heterogeneous blob objects through voting and level set formulation," Pattern Recognition Letters, Vol.28, No.13, pp.1781-1787, 2007.
- [6] 정미라, 심정희, 고병철, 남재열, "형태학적 특성과 베이지안 네트워크를 이용한 자동 세포 분할 및 분류," 한국화상학회 논문지, 제14권, 1호, pp.31-38, 2008.
- [7] N. Otsu, "A threshold selection method from gray level histograms," IEEE Trans. On Systems, Man and Cybernetics, Vol.9, pp.62-66, 1979.
- [8] M. K. Hu, 'Pattern recognition by moment invariants', IRE Trans. Inform. Theory, Vol.8, Issue2, pp.179-187, 1962.
- [9] P. Duchene, D. Lewis, 'Visilog 5 Documentation', Noesis Vision Inc., Canada, 1996.
- [10] C. Restif, W. F. Clocksin, "Comparison of Segmentation Methods for Cytometric Assay," International Conference on Medical Image Understanding and Analysis, pp.153-156, 2004.
- [11] Gang Lin, Umesh Adiga, Kathy Olson, John Guzowski, Carol Barnes, and Badrinath Roysam. "A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks," Cytometry, Vol.56A, pp.23-36, 2003.

정 미 라

e-mail : jungmr@kmu.ac.kr

2007년 계명대학교 컴퓨터공학과(학사)

2007년 3월~현재 계명대학교 컴퓨터공학과 석사과정

관심분야 : 컴퓨터 비전 및 패턴인식



고 병 철

e-mail : niceko@kmu.ac.kr

1998년 경기대학교 전자계산학과(이학사)

1998년~2000년 연세대학교 컴퓨터과학(공학 석사)

2000년~2004년 연세대학교 컴퓨터과학(공학 박사)



2004년 3월~2005년 8월 삼성전자 통신연구소 책임연구원

2005년 9월~현재 계명대학교 컴퓨터공학과 조교수

관심분야 : 내용기반 영상검색, 컴퓨터 비전 및 패턴인식



남 재 열

e-mail : jynam@kmu.ac.kr

1983년 경북대학교 전자공학과(공학사)

1985년 경북대학교 전자공학(공학석사)

1991년 University of Texas at Arlington
전기공학(공학박사)

1985년 5월~1987년 7월 한국전자통신연
구소 연구원

1991년 9월~1995년 2월 한국전자통신연구소 선임연구원

1995년 3월~현 재 계명대학교 컴퓨터공학과 교수

관심분야 : 영상압축, 영상통신, 멀티미디어 시스템