

구조적 특징 정보를 이용한 복잡한 세포영상 분할

김성곤*

Complex Cell Image Segmentation via Structural Feature Information

Seong-gon Kim*

요 약

본 논문에서는 과분할 방지와 복잡한 현미경 세포영상의 자동 분할을 위한 새로운 마커 기반의 워터셰드 알고리즘을 제안한다. 워터셰드 방식은 접촉 또는 겹침으로 인한 복잡한 대상들을 분할하기 위해 효과적이며 보다 성공적인 적용을 위해 정확한 마커 추출이 선행되어야 한다. 세포 영상의 정확한 마커 추출을 위해 본 논문에서는 방사형 대칭성을 이용한 반복 보팅 알고리즘을 이용하였다. 합성 영상과 실제 영상에서도 기존의 다른 방식들에 비해 양호한 분할 결과를 보였다.

▶ Keywords : 워터셰드 알고리즘, 방사형 대칭, 반복 보팅, 마커

Abstract

We propose a new marker driven Watershed algorithm for automated segmentation of clustered cell from microscopy image with less over segmentation. The Watershed Transform is able to segment extremely complex objects which are highly touched and overlapped each other. The success of the Watershed Transform depends essentially on the finding markers for each of the objects of interest. For extracting of markers positioning around center of each cell we used radial symmetry and iterative voting algorithms. With synthetic and real images, we quantitatively demonstrate the performance of our method and achieved better results than the other compared methods.

▶ Keywords : Watershed Algorithms, Radial Symmetry, Iterative Voting, Marker

· 제1저자 : 김성곤 · 교신저자 : 김성곤

· 투고일 : 2012. 05. 23, 심사일 : 2012. 07. 04, 게재확정일 : 2012. 09. 19

*부산가톨릭대학교 컴퓨터공학과(Department of Computer Engineering, Catholic Uni. of Pusan)

※ 본 논문은 2010학년도 부산가톨릭대학교 교내연구비에 의하여 지원되었습니다.

I. 서론

인체 조직의 세포 영상을 분할하는 작업은 다음과 같은 이유로 상당히 어려움이 있다. 살아있는 세포를 과도하게 염색하여 대조(contrast)가 좋은 영상을 만들기도 어렵고, 주변 환경의 영향을 받아 세포의 형태가 변화기 쉽고, 특히 세포 운동 등으로 인한 세포와 세포 사이의 경계나 세포와 배경과의 구분이 명확하지 않기 때문이다.

세포 영상을 분석하기 위해 우선적으로 정확한 세포의 분할이 필요하다. 디지털영상처리 기술을 활용하여 세포영상을 정확히 분할하기엔 많은 어려움이 있다. 잡음의 영향과 영상에서 분할 대상인 세포들이 접촉 또는 겹치거나 특히 밀집하여 모여 있기 때문에 정확한 분할이 쉽지 않다.

세포분할에 많이 적용되는 대표적 방법으로는 스네이크(snakes), 레벨세트(level set), 모폴로지(morphology) 그리고 워터셰드(watershed) 방법 등이 있다.

Kass 등이 제안한 스네이크 모델[1][2]은 영상에서 특정 대상물의 경계를 추출하거나, 특정한 한 개의 움직이는 물체를 추적하는 경우엔 추출속도나 성능도 양호하지만 추출대상이 여러 개일 경우 위상 처리가 어려워 추출에 실패하기 쉽다.

Sethian 등이 제안한 레벨세트 알고리즘[3][4]은 설정된 초기 곡선을 안정된 속도로 확장 또는 축소시켜가면서 영상의 원하는 대상들의 경계들을 추출해낼 수 있다. 스네이크 방법에 비해 초기곡선 설정의 제약이나 위상처리도 간단하지만 계산량이 많고 특히 세포영상 같이 세포들이 겹치거나 밀집된 경우엔 각각의 세포들을 분할하기엔 어려움이 많다.

워터셰드 방식은 계산도 복잡하지 않고 전체 영상을 분할하는 경우 추출 성능도 양호하지만 대체적으로 잡음이나 영상 밝기 변화에 민감하여 과분할(over segmentation) 문제점을 항상 가지고 있다. 워터셰드의 과분할 문제를 해결하기 위해 분할될 각 영역의 일부를 예측하여 마커(marker)를 지정하고 이를 기점으로 워터셰드 분할을 수행하는 방법이 많이 이용되고 있다[5][6][7].

본 논문에서는 복잡한 세포 영상을 분할하기 위해 마커 기반 워터셰드 방식을 활용하였으며, 특히 정확한 마커를 추출하기 위해 분할 대상의 구조적 특징 정보를 이용한 반복 보팅(iterative voting)방법을 이용하였다[8].

본 논문에서 다루는 실험영상은 대장의 일부인 결장(colon) 암세포 조직의 형광현미경(fluorescent microscopy)세포영상이다. 논문의 구성은 다음과 같다. II장에서는 워터셰드 분할 알고리즘을 이용한 분할방법을 설명

하고 III장에서는 제안모델의 이용한 마커 추출 알고리즘을 상세히 기술한다. IV장에서는 복잡한 세포 영상을 이용한 제안 모델의 유효성을 실험하며 결과에 대해 고찰한 후, 마지막으로 V장에서 본 논문의 결론을 맺는다.

II. 워터셰드 분할

2.1 워터셰드 분할 알고리즘

워터셰드 분할 알고리즘은 지형학적 관점에서 설명하면 영상의 밝기 값의 변화 즉 그라디언(gradient)을 지형학에서의 고도(altitude)로 생각한다. 고도가 낮은 지역적 최소점을 기점으로 담수지역(catchment basin)을 구분하는 워터셰드 또는 댐(dam)을 생성함으로써 각 담수지역에 결합된 영역을 분리해서 찾아내는 알고리즘이다.

현재까지 워터셰드를 구하기 위해 많은 알고리즘이 제안되었으나 가장 효과적인 방법은 침수 모의실험(immersion simulation)을 이용한 방법이다[5].

침수 모의실험이란 그림 1(b)와 같이 고도가 가장 낮은 지역적 최소 영역들 즉, 마커를 기점으로 하여 침수가 시작되어 점점 침수가 진행되면서 최고 수위까지 진행된다고 가정한다. 우선 지역적 최소점을 기점으로 담수지역이 형성되어 확장되다가 언젠가는 인접한 담수지역끼리 합해질 시점에 두 담수지역의 경계 부분에 댐 즉, 워터셰드를 구축하여 담수지역을 구분해준다.

워터셰드 알고리즘은 지역적 최소점과 담수지역을 이용해 이웃하는 담수지역의 경계를 쉽게 찾을 수 있는 장점이 있지만, 잡음이나 영상의 지역적 변화가 심한 경우 지역적 최소점의 수가 많아져서 과도한 분할을 초래한다.

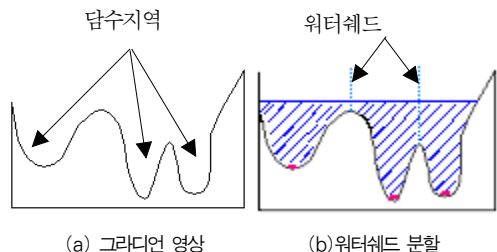


그림 1. 침수 기반 워터셰드 분할
Fig. 1. Immersion based watershed segmentation

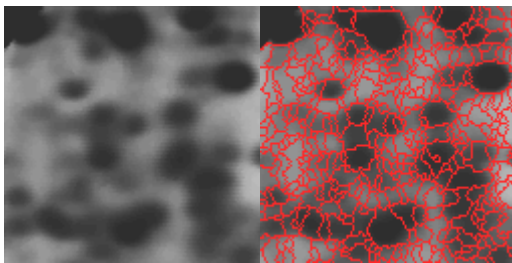
이와 같은 문제점을 해결하기 위해 분할 작업을 수행하기 전에 영상의 다양한 특징 정보들을 이용하여 분할 대상 영역

들의 일부로 간주할 수 있는 지역적 최소점들 즉 마커들을 먼저 설정해 준다.

그림 2(b)는 마커 지정 없이 원영상을 워터셰드 방식으로 분할한 경우로 영상 자체의 많은 지역적 최소영역들 때문에 과도한 분할이 발생하는 예이다.

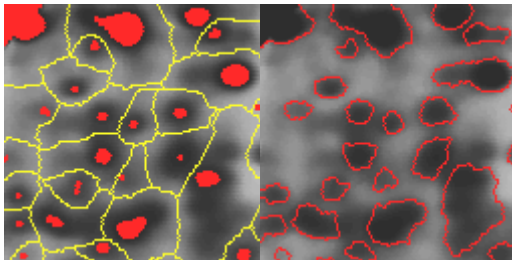
2.2 마커 기반 워터셰드 분할

그림 3.은 그림 2(a)의 원영상에 대해 마커를 이용한 분할



(a)원영상 (b)분할 결과

그림 2. 과분할 예
Fig. 2. Example of over-segmentation



(a)원영상 + 마커 (b)분할 결과

그림 3. 마커 기반 분할 예
Fig. 3. Example of mark based segmentation

결과이다. 그림 3(a)처럼 워터셰드 분할을 수행하기 전에 우선 대상 영상에서 분할될 각 영역을 대표하는 마커 추출이 이루어져야 한다. 그림 3(b)는 마커를 기반으로 한 워터셰드 분할 결과이며 그림 2(b)와 비교해 보면 과분할 문제가 개선됨을 알 수 있다.

정확한 분할은 정확한 마커 선정에 크게 좌우되며 마커 추출에 사용되는 방법은 영상의 특징에 따라 다양한 방법이 사용되고 있다. 가장 일반적인 마커 추출 방법은 거리함수와 지역적 최소화[6]를 이용하며, 텍스처 특징을 이용한 방법[7] 등이 있다.

본 논문에서는 세포영상내의 개별 세포영역을 구분하기 위

해 각 세포들의 중심을 추정하며 이를 위해 세포들의 구조적 특징 즉 방사형 대칭성을 이용한 반복 보팅 방법을 사용한다.

마커 추출 작업을 포함한 본 논문에서 제안하는 세포 분할 알고리즘의 전체 처리 순서는 그림 4.와 같다.

III. 제안하는 영상분할 모델

3.1 마커(Marker) 추출 알고리즘

워터셰드를 이용한 영상분할의 단점은 밝기 변화가 심하거나 잡음 등의 문제로 인해 과분할 현상을 유발하는 경우가 많다. 이러한 문제점을 해결하기 위해 기존에 많이 사용하는 방법이 마커를 이용한 분할법이다.

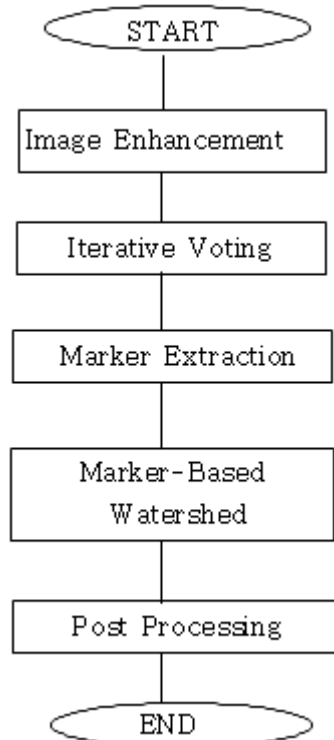


그림 4. 처리 순서도
Fig. 4. Flow diagram

기존의 마커추출 방법들은 추출할 각 대상물체의 중심영역을 구하기 위해 거리함수(distance function)를 많이 이용한다. 그러나 추출할 영상내의 물체들이 아주 근접하게 붙어 있거나 특히 물체의 겹침이 심할 경우 거리함수만 이용해서는

정확한 대상물체들의 중심영역들을 구하기가 어렵다.

본 논문에서는 정확한 마커 추출을 위해 영상의 구조적 특징을 이용하는 보팅(voting)방법을 이용하여 마커를 구한다 [8]. 세포영상 중에서 각 세포들이 겹침이 심하거나 너무 밀집된 복잡한 세포영상에서 보다 정확한 마커추출이 가능하다.

3.1.1 반복 보팅 알고리즘

일반적인 세포영상에서 개별 세포들은 불룩하거나 타원과 유사한 형태를 가진다. 각 세포들의 중심점에 근사한 후보 영역을 추출하기 위해 방사형 대칭(radial symmetry) 특징을 이용한다. 그림 5와 같이 물체의 경계점에서 일정한 크기와 범위를 가지는 커널을 반복적으로 크기와 범위를 줄여가면서 후보영역을 보팅을 해가면 세포의 중심에 근사화된 위치를 구할 수 있다. 실제 세포영상의 경우 그라디언 값이 큰 에지점에서 그라디언 방향으로 보팅을 시작하여 점진적으로 보팅방향과 보팅영역을 수정해가며 중심영역을 찾아낸다.

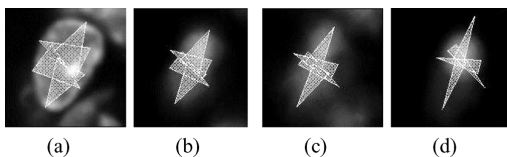


그림 5. 반복 보팅
Fig. 5. Iterative voting

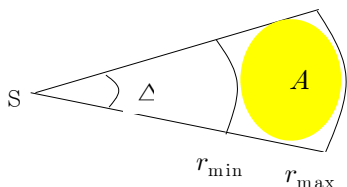


그림 6. 보팅 커널
Fig. 6. Voting kernel

본 논문에서 사용하는 커널의 구조는 그림 6과 같다. S는 보팅 적용점이며, $\{r_{min}, r_{max}\}$ 는 방사 방향으로의 보팅 범위(radial range), Δ 는 보팅각 범위(angular range)이다. $A(x, y; r_{min}, r_{max}, \Delta)$ 는 보팅영역이며 보다 중심부분에 집중적인 부드러운 보팅값을 얻기 위해 가우시안 마스크를 적용하였다.

보통 세포들의 크기가 다양할 경우엔 최소반지름과 최대반지름 값을 이용하여 r_{min} 과 r_{max} 값을 정하지만, 본 논문에서는 세포영상에서 세포들의 크기가 유사할 경우엔 평균세포

반지름 길이 \pm 편차값을 r_{min} 과 r_{max} 값으로 함으로써 보다 효율적으로 보팅이 수행되도록 하였다.

본 논문에서 사용하는 방사형 대칭 특징을 이용한 보팅 알고리즘의 전체 과정은 다음과 같다.

- 1) 파라미터를 초기화 한다: $r_{min}, r_{max}, \Delta_{max}$

$$\Delta_{max} = \Delta_n > \Delta_{n-1} > \dots > \Delta_0$$

여기서 n 은 반복 횟수이다.

- 2) 그라디언 크기 영상 $F(x, y)$ 을 구한다.

- 3) 보팅방향 $\alpha(x, y)$ 와 보팅 적용점 S 를 구한다.

$$\alpha(x, y) = -\frac{\nabla I(x, y)}{|\nabla I(x, y)|}$$

$$S = \{(x, y) | |\nabla I| > \Gamma_g\}$$

Γ_g 는 보팅 적용점을 구하기 위한 임계값이다.

- 4) S점들에 대한 보팅영상 $V(x, y)$ 를 구한다.

$$V(x, y; r_{min}, r_{max}, \Delta) = V(x, y; r_{min}, r_{max}, \Delta) + \sum_{(u, v) \in A} F(x + u, y + v) G(u, v; \sigma, \alpha, A)$$

$G(x, y, \sigma, \alpha, A)$ 는 보팅영역에 마스크되는 분산 σ^2 을 가지는 가우시안 커널이다.

- 5) V 값 중 최대값을 가지는 (u^*, v^*) 방향으로 보팅 방향을 수정한다.

$$(u^*, v^*) = \arg \max_{(u, v) \in A(x, y; r_{min}, r_{max}, \Delta_n)} V(u, v; r_{min}, r_{max}, \Delta_n)$$

- 6) $n = n - 1$, 보팅각 Δ_n 을 수정 후 $n = 0$ 때까지

4)에서 6)과정을 반복한다.

- 7) 중심 후보점 C 를 구한다.

$$C = \{(x, y) | V(x, y; r_{min}, r_{max}, \Delta_0) > \Gamma_v\}$$

그림 7의 이진영상은 세포영상의 특징을 가당한 영상으로 분할할 대상들이 일반적인 세포 영상처럼 불룩하거나 타원 형태를 하고 있다. 그리고 시각적으로 잘 분리된 세포들이 아니고 세포들이 인접하거나 겹치거나 밀집된 모양으로 이루어져 있다. 반복 보팅 처리과정과 중심 후보점을 추출한 결과를 보여준다. 세포들이 인접하거나 겹칠 경우에도 개별 세포들을 구분하는 중심점들을 정확히 추출함을 알 수 있다.

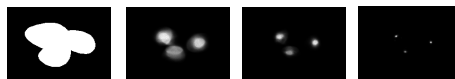


그림 7. 중심점 추출 예
Fig. 7. Example of center point extraction

3.2 워터셰드 분할

최종 마커 추출이 완료되면 최종적으로 분할 대상 영상에 마커들을 포함시키고 워터셰드 분할을 수행한다. 분할에 사용되는 에지 정보는 다음 식 (1)과 같은 모폴로지 그라디언이다.

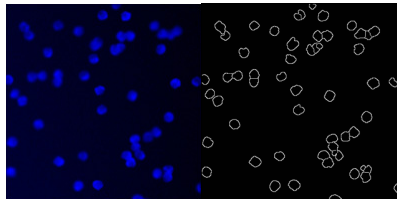
$$G(f) = (f \oplus B) - (f \ominus B) \dots\dots\dots(1)$$

\oplus : 모폴로지 팽창 연산, \ominus : 모폴로지 침식 연산.

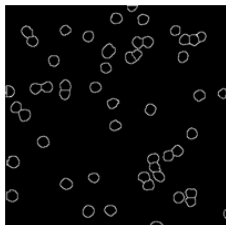
식 (1)에서 f 는 원영상, B 는 구조 요소이다.

IV. 실험 및 고찰

본 논문에서 제안하는 알고리즘의 구현과 객관적인 세포 분할 성능을 비교하기 위해 Tampere University of Technology에서 제공하는 벤치마크 데이터베이스의 영상을 이용하였다. 동일한 영상에 대해 서로 다른 분할 방식의 성능을 객관적으로 비교한다는 것은 어렵기 때문에 우선 벤치마크 용으로 제공되는 다양한 실험 영상에 대해 추출 성능을 비교해보았다. 전체 실험영상 중 세포 간 접촉이나 겹침 정도가 15% 수준 영상(그림 8(a)), 그리고 좀 더 겹침 정도가 심한 45% 수준의 영상(그림 9(a))을 이용하여 성능 평가를 수행하였다. 각 영상의 실제 크기는 950 X 950 픽셀이며 영



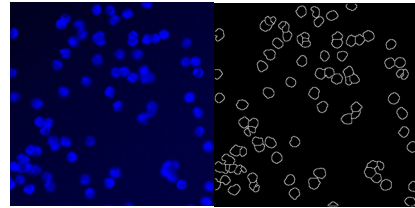
(a)원영상 (b)CellProfiler



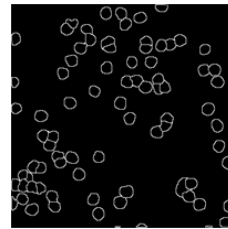
(c)제안 모델

그림 8. 세포 분할(15%)

Fig.8. Cell segmentation(15%)



(a)원영상 (b)CellProfiler



(c)제안 모델

그림 9. 세포 분할(45%)

Fig.9. Cell segmentation(45%)

상 당 세포의 개수는 300개이다. 그림 8(a)에 원영상의 일부를 표시하였다. 그림 8(b)는 의료영상 분석용 공개소프트웨어 패키지인 CellProfiler를 사용하여 분할한 결과이며 그림 8(c)는 제안 모델로 분할한 결과이다. 그림 9(a)는 겹침이나 접촉 정도가 심한 경우의 영상이다. 그림 9(b)는 CellProfiler 소프트웨어로 분할한 결과이며 그림 9(c)는 제안모델의 분할 결과이다. 제안 방식의 추출 결과가 CellProfiler 보다 시각적으로 양호하다는 것을 알 수 있으나 보다 객관적이고 정량적인 비교를 위해 식 (2)를 사용하여 추출 성능을 나타내었다. 분할 기준치를 GT로 두고 분할된 결과를 SR로 하여 기준치와 분할결과의 일치 정도를 성능으로 표시하였다. 표 1에 두 방식의 분할 결과와 추출 성능을 나타내었다.

$$\text{추출 성능} = \frac{GT \cap SR}{GT} \dots\dots\dots(2)$$

여기서 GT=Ground Truth 영역

SR=Segmentation Result 영역

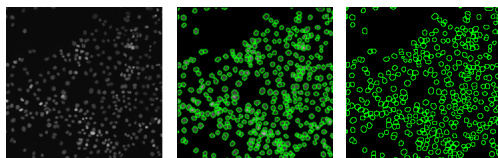
표 1. 벤치마크 데이터 집합에 대한 분할 결과
Table 1. Segmentation results on Benchmark data sets

영상 데이터	영상 #	CellProfiler	제안 모델	추출 성능
15% 정도 겹침 영상	1	308(8)	304(4)	0.975
	2	306(6)	303(3)	0.986
	3	309(9)	305(5)	0.977
	4	312(12)	307(7)	0.971
	5	310(10)	305(5)	0.978
45% 정도 겹침 영상	1	316(16)	308(8)	0.969
	2	320(20)	313(13)	0.955
	3	313(13)	307(7)	0.970
	4	315(15)	309(9)	0.966
	5	317(17)	312(12)	0.958
평균		(12.6)	(7.3)	0.971

전체 10개의 영상에 대해 실험한 결과 제안 모델이 CellProfiler의 평균 추출 오차 12.6보다 양호한 평균 7.3개 정도의 추출 오차를 보였다. 그리고 제안 모델의 추출 성능은 겹침의 정도에 따라 다소 차이는 있지만 평균 0.971의 우수한 결과를 보였다.

복잡한 실제 세포 영상에 대해 제안 알고리즘의 적용 가능성을 확인하기 위해 결장 암세포 영상(그림 10(a))을 사용하였다. 그림 10(b)에 전문의의 수작업 분할 결과를 원영상에 표시하였으며 그림 10(c)에는 제안 모델로 추출한 결과를 나타내었다. 전문의 수작업 결과와 제안 방식의 성능을 비교하기 위해 유사한 5개 데이터 집합을 사용하여 실험하였으며 표 2에 실험 결과를 나타내었다.

표 2에 나타난 추출 성능은 수작업 추출 결과를 식 (2)의 기준치 GT로 두고 제안모델의 분할 결과를 SR로 하여 백분율로 구하였다. 표 2에서 나타난 것처럼 실제 영상에서도 93.6%에서 96.9% 범위에서 평균적으로 95.5%의 성능을 보였다. 실험 환경은 Intel Pentium-IV 2GHz(256RAM) pc에서 Matlab과 C로 수행되었고 처리 시간은 영상 당 1.25초 정도가 소요되었다.



(a)원영상 (b)수작업 (c)제안 모델

그림 10. 결장 암세포 분할

Fig. 10. Human colon cancer cell segmentation

표 2. 결장 암세포 영상 집합에 대한 분할 결과.
Table 2. Segmentation results on human colon cancer cell image sets.

데이터 집합	1	2	3	4	5	평균
추출 성능	93.6	96.9	96.3	95.7	94.8	95.5 (%)

V. 결론

임상 진단에서 환자의 의료 영상을 시각적 또는 수작업으로 영상을 해석하여 진단에 이용한다. 하지만 수작업으로 수없이 발생하는 영상을 해석하는 것은 번거로운 작업이며 또한 정확도 면에서도 전문가 마다 상이할 수 있으며 시간에 따라 동일한 전문가의 수작업 결과는 다를 수 있다. 이런 관점에서 영상처리 기술을 응용한 자동화 처리의 필요성이 높고 동시에 실시간 처리로 다양한 영상의 중요 정보들을 획득하여 진단이나 치료에 요긴하게 활용할 수 있다. 본 연구는 다양한 세포 영상 중에서 세포의 형태가 타원과 유사한 구조의 세포 영상에 국한되지만 세포 각각의 구조적인 특징들을 이용함으로써 워터셰드 분할을 위한 정확한 마커를 추출할 수 있었기 때문에 세포간 겹침이나 겹침이 심한 경우에도 상당히 양호한 분할 성능을 보여 주었다.

참고문헌

- [1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models", *Int. J. Computer Vision*, vol. 1, pp. 321-331, 1987.
- [2] 이승태, 한영준, 한현수, "컬러 분산 에너지를 이용한 확장 스네이크 알고리즘", *한국컴퓨터정보학회 논문집*, 제 14권 10호, pp. 83-92, 2009.
- [3] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modelling with front propagation", *IEEE Trans. Pattern Anal. Machine Intell.*, 17, 1995.
- [4] J. A. Sethian, "Level Set Methods and Fast Marching Methods", Cambridge University Press, 1999.
- [5] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583-598, 1991.
- [6] Jierong Cheng and Jagath C. Rajapakse, "Segmentation of Clustered Nuclei With Shape Markers and Marking Function", *IEEE Transaction on Biomedical Engineering*, 56(3), pp. 741-748, March 2009.
- [7] Mohammad Ibrahim Khan, Muhammad Kamal Hossen and Md. Sabbir Ali, "Cell Segmentation from Cellular Image", *Global Journal of Computer Science and Technology*, 10(13), pp. 55-60, October 2010.
- [8] Qing Yang, Bahram Parvin, "Perceptual Organization of Radial Symmetries," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 320-325, 2004.

저자 소개



김 성 곤 (Seong-gon Kim) 正會員
 1983년 2월 :
 경희대 전자공학과(공학사)
 1985년 2월 :
 경희대 전자공학과(공학석사)
 2000년 2월 :
 동아대 전자공학과(공학박사)
 1992년 3월 ~ 현재 :
 부산가톨릭대학교 교수
 관심분야 : 신호처리,
 의료영상해석, DSP
 Email : skkim@cup.ac.kr

