

## 형태론적 체인코드 에지벡터를 이용한 핸드 제스처 시퀀스 인식

이강호\*, 최종호\*\*

# Hand Gesture Sequence Recognition using Morphological Chain Code Edge Vector

Kang-Ho Lee\*, Jong-Ho Choi\*\*

### 요약

최근 들어 인간의 의지를 컴퓨터에 전달하기 위한 수단으로 컴퓨터 시각기반 방식으로 제스처를 인식하고자 하는 연구가 널리 진행되고 있다. 제스처 인식에서 가장 중요한 이슈는 알고리즘의 단순화와 처리시간의 감소이다. 이러한 문제를 해결하기 위하여 본 연구에서는 기하학적 집합론에 근거하고 있는 수학적 형태론을 적용하였다. 본 논문에서 제안한 알고리즘의 키 아이디어는 형태론적 형상분해를 적용하여 얻은 원시형상 요소들의 중심점을 연결하는 궤적을 추적하는 것이다. 핸드 제스처 시퀀스의 중심점 궤적은 핸드 제스처의 형상에 관련된 중요한 정보를 내포하고 있다. 이러한 특징에 근거하여 본 연구에서는 원시형상 요소들의 중심점 궤적과 직접적으로 관련되는 체인코드 에지벡터로부터 형상의 특징벡터를 계산하여 핸드 제스처 시퀀스를 인식할 수 있는 알고리즘을 제안하고, 실험을 통하여 그 유용성을 증명하였다.

### Abstract

The use of gestures provides an attractive alternate to cumbersome interface devices for human-computer interaction. This has motivated a very active research area concerned with computer vision-based analysis and interpretation of hand gestures. The most important issues in gesture recognition are the simplification of algorithm and the reduction of processing time. The mathematical morphology based on geometrical set theory is best used to perform the processing. The key idea of proposed algorithm is to track a trajectory of center points in primitive elements extracted by morphological shape decomposition. The trajectory of morphological center points includes the information on shape orientation. Based on this characteristic, we proposed the morphological gesture sequence recognition algorithm using feature vectors calculated to the trajectory of morphological center points. Through the experiment, we demonstrated the efficiency of proposed algorithm.

▶ Keyword : Shape Function, Dynamic Thesholding, Background Subtraction

• 제1저자 : 이강호

• 접수일 : 2004.10.26, 심사완료일 : 2004.11.19

\* 한국재활복지대학 정보보안과 부교수, \*\* 강남대학교 전자시스템공학부 교수

## I. 서론

최근 들어 인간의 의지를 직감적으로 기계에 전달하기 위한 수단으로 제스처를 이용하고자 하는 연구가 널리 진행되고 있다. 일상의 대화에서 자연스럽게 사용되고 있는 제스처를 이해하는 인터페이스를 구현할 수만 있다면 제스처는 의사전달 및 감정이나 정서 등의 미묘한 표현에 대응하는 자연스러운 입력수단으로 널리 이용될 수 있을 것이다.

제스처 인식 기법(1)(2)(3)은 크게 인간의 몸체에 글러브와 같은 여러 가지 타입의 센서를 부착해서 취득한 데이터를 해석하는 기기장착 방식과 비디오 카메라로부터 취득한 영상을 처리하는 시각기반 영상처리 방식으로 나눌 수 있다. 제스처 인식 초기에는 센서글러브를 사용하는 방식이 주 관심대상이었으나 움직임이 제한되고 동작이 부자연스러우며 착용시의 불편함과 심리적 부담 때문에 취급이 매우 불편하다는 문제점이 있다. 이러한 상황에서 현재는 특별한 응용분야를 제외하고는 시각기반의 영상처리를 이용하여 비접촉 방식으로 제스처를 인식하고자 하는 분야가 주로 연구되고 있다.

시각기반 영상처리 방식은 센서를 부착하지 않아도 된다는 장점이 있지만 비디오 데이터의 분해능이 낮고, 시각기반 영상처리의 주 연구대상인 모습기반 접근(appearance-based approach) 방식에서 주로 적용하고 있는 HMM(hidden Markov model)과 신경망 모델(neural network model) 등은 인식 알고리즘이 복잡하여 하드웨어 구현 및 실시간 처리가 불가능하다는 문제점이 있다. 그러나 제스처 인식은 인간과 기계간의 인터페이스에 관련된 부분이므로 하드웨어의 실시간 제어에 관련된 부분이 그 핵심이다. 따라서 알고리즘의 복잡성과 실시간 처리 및 하드웨어화의 불가능성은 실용화 레벨의 측면에서 큰 걸림돌로 작용하고 있다.

따라서 본 연구에서는 하드웨어 구현이 용이하고 고속연산이 가능한 수학적 형태론을 적용한 형태론적 핸드 제스처 시퀀스 인식 알고리즘을 제안하였다. 화소사이의 논리적 연산에 근거하고 있는 수학적 형태론에서는 형태론적 논리연산으로 구성되는 유용한 영상처리 기술들이 다양한 형태로 개발되어 있다. 수학적 형태론은 기하학적인 특징을 정량화

하고자 하는 집합이론적 접근방식으로 미리 정의된 모양의 패턴을 가지고 영상을 변환하고 해석하고자 하는 방법으로서 영상에 내포된 형상의 기하학적인 구조를 정량화 할 수 있다는 특징 때문에 실용화에 가장 가까운 기술로 각광받고 있다.

인간이 일상생활에서 주로 사용하고 있는 제스처로 핸드 제스처, 몸짓, 표정, 시선 등을 들 수가 있다. 이중에서도 핸드 제스처는 그 표현의 다양성 때문에 중요한 의사표현 수단으로 널리 사용되고 있다. 따라서 본 논문에서는 핸드 제스처 시퀀스를 이용한 직감적이고 빠른 인터페이스 설계에 응용될 수 있는 형태론적 핸드 제스처 시퀀스 인식 기법에 관해 연구하였다. 2-D 영상내에 포함되어 있는 물체들의 복잡한 형상을 단순한 원시형상요소들로 분해하여 이들을 계층적으로 표현하는 것은 인간의 시각적인 인식이 기본 단계이므로 핸드 제스처 영상시퀀스에서 추출된 형상영역을 원시형상 요소들로 분해하였다. 이를 바탕으로 원시형상 요소간의 위치관계를 이용하여 핸드 제스처 시퀀스를 인식하는 방법을 제안하고, 제안된 이론의 유효성을 증명하기 위한 실험을 통하여 핸드 제스처 시퀀스 인식이 가능하다는 것을 확인하였다. 본 연구에서 새롭게 제안된 방법은 형태론적 형상분해를 이용하여 추출된 원시형상 요소의 중심점을 선택하여 그 궤적을 추적함으로써 형상의 특징벡터를 인식하고자 하는 것이다. 본 연구에서 제안한 방식은 비디오 데이터의 검색 및 기타 전자시스템 등의 작동에 관련된 인터페이스 설계에도 널리 응용될 수 있을 것으로 판단된다.

## II. 형태론적 형상분해

형상분해는 형상 X를 여러개의 원시형상 요소  $X = \{X_1, X_2, X_3, \dots, X_n\}$  들로 분해하고자 하는 것으로 이의 수학적 표현은 다음과 같다.

$$X = \bigcup_{i=1}^n X_i, X_i \in G(R^2) \dots \dots \dots (1)$$

여기서  $G(R^2)$ 은 2-D 유클리드 공간 R에서 정의되는 2-D 개집합(Open Set)이다.

형태론적 형상분해에서 원시형상 요소는 다음과 같은 형태론적 연산을 수행하여 얻을 수 있다[4].

$$X_i = X_{niB} = (X \ominus niBs) \oplus niB \dots\dots\dots (2)$$

여기서 B는 구성소(Structuring Element)이고, ni는 구성소의 크기이며,  $\oplus$ 와  $\ominus$ 는 각각 확장(Dilation)과 침식(Erosion) 연산자이다. 그리고 Bs는 원점에 대한 B의 반전(Reflection)이다.

식 (2)는 형상이 점 또는 선으로 축소될 때까지 형상 X를 B로 침식연산한 결과를 침식 연산한 횟수만큼 확장연산하면 원시형상 요소를 얻을 수 있다는 것을 의미한다. 이러한 처리는 X-X<sub>i</sub>에 대해서 반복적으로 수행된다. 이상에서 서술한 처리과정을 회귀형식으로 표현하면 다음과 같다[4].

$$\begin{aligned} X_i &= (X - X'_{i-1})niB \\ &= ((X - X'_{i-1}) \ominus niBs) \oplus niB \\ X_i &= \bigcup_{1 \leq j \leq i} X_j \\ X_0' &= \emptyset \\ \text{Stopping Condition: } &(X - X_k') \ominus Bs = \emptyset \end{aligned} \quad (3)$$

여기서 종료조건은 형상의 전영역을 남김없이 분해하기 위한 조건이고, k는 분해된 원시형상 요소의 총수이다.

### III. 핸드 제스처 시퀀스 인식

본 연구에서 제안한 체인코드 에지벡터를 이용한 핸드 제스처 시퀀스 인식 알고리즘은 그림 1에 나타난 바와 같이 그 처리과정을 4단계로 구분할 수 있다. 첫 번째 과정은 물체영역과 배경을 분리하는 영상분할과 잡음제거 및 문턱치 처리 과정이다. 두 번째 과정은 기하학적 패턴에 따라 예측된 구성소를 선택하고 형상을 분해하는 과정이고, 세 번째 과정은 원시형상 요소들의 중심점의 궤적을 추출하는 과정이다. 마지막으로 네 번째 과정은 체인코드 에지벡터를 계산하고 제스처 시퀀스를 인식하는 과정이다[5][6][7].

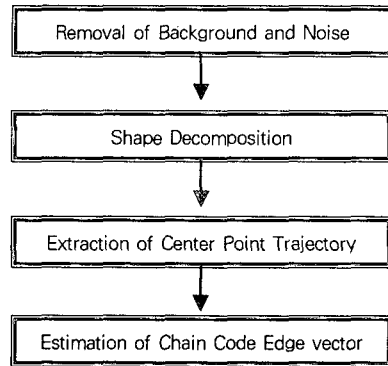


그림 1. 핸드 제스처 시퀀스 인식 알고리즘  
Fig.1 Hand gesture sequence recognition algorithm

#### 3.1 전처리

##### 3.1.1 영상분할

형태론적 영상분할 알고리즘은 동적 문턱치처리(Dynamic Thresholding)와 배경제거(Background Subtraction)에 근거한다. 그 기본원리는 주위보다 더 밝거나 혹은 어두운 형상들을 제거하여 배경을 얻고, 이를 원영상에서 분리하고자 하는 것이다. 형상분해의 대상이 되는 영상에서 형상은 배경보다 어둡거나 혹은 밝게 표현된다. 그러므로 형상영역의 그레이레벨의 최대값과 최소값을 결정하면 배경영역에 대응되는 영상을 얻을 수 있다.[8]

지형적 구조에서 열기연산은 뾰족한 봉우리를 무디게 하는 연산특성을 갖고, 닫기연산은 좁고 끝이 깊은 골짜기를 제거하는 연산 특성을 갖는다. 이와 같은 특성 때문에 대상영상을 2-D 구성소로 열기연산하면 피크부분이 잘려 나가게 되고, 닫기연산하면 영상 내에 존재하는 골짜기 부분이 소멸되게 된다. 그러므로 열기연산과 닫기연산을 수행하여 마루(peak)와 계곡(valley)에 해당하는 그레이레벨을 결정한 후에, 그에 대응되는 배경영상을 원영상에서 제거하면 형상영역만을 얻을 수 있다.

이상에서 서술한 연산과정은 다음과 같이 정의되는 Top-Hat 변환을 적용하여 수행된다.

$$\text{Peak}(f) = f - (f \circ B) \dots\dots\dots (4)$$

$$\text{Valley}(f) = (f \bullet B) - f \dots\dots\dots (5)$$

위의 과정에 의해 배경이 제거된 영상으로부터 2진영상을 얻는 과정은 식 (4)로 표현되는 문턱치 처리로 수행한다. 형상분해의 대상이 되는 영상에서는 배경과 형상의 그레이레벨의 차이가 비교적 큰 경우이므로 문턱치값은 히스토그램의 쌍봉성을 이용하여 결정한다.

$$\begin{aligned}
 f_t(i,j) &= 1 & f(i,j) &\geq t \\
 &0 & f(i,j) &= 0
 \end{aligned}
 \text{for } \dots\dots\dots (6)$$

여기서 t는 문턱치값이다.

3.1.2 잡음제거

열기연산은 대상물체의 외부에 존재하는 날카로운 모서리 부분을 부드럽게 하고 구성소보다 작은 물체를 삭제하는 필터의 성질을 갖고 있다. 그리고 닫기연산은 협곡모양의 물체에 나타나는 작은 구멍을 제거하는 성질을 갖고 있다. 이로부터 열기연산과 닫기연산은 각각 양의 잡음성분과 음의 잡음성분의 제거를 위한 필터로 사용할 수 있다는 것을 알 수 있다.

위에서 검토한 사실로부터 영상내에 존재하는 불규칙 잡음을 제거하는 저역통과 필터는 다음의 식으로 표현할 수 있다.

$$\begin{aligned}
 flp(m,n) &= f(m,n) \circ B1(m,n) \bullet B1(m,n) \circ B2(m,n) \\
 &\bullet B2(m,n) \dots\dots\dots (7)
 \end{aligned}$$

여기서 B1 과 B2는 각각 반경이 1과 2인 원 구성소(Structuring Element)이다.

3.2 형상분해

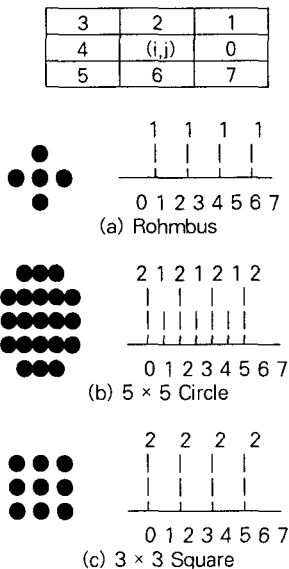


그림 2. 구성소의 형상함수

Fig. 2 Shape function of structuring elements

형상의 기하학적 특징이 강조된 원시형상 요소들로 형상을 분해하면 기술효율이 향상된다. 그리고 기술효율을 향상시키기 위하여 고려되는 대표적인 기하학적인 특징은 형상의 경계선이 갖고 있는 방향정보이다. 이와 같은 측면에서 본 연구에서는 체인코드를 이용하여 형상의 방향정보를 형상함수(Shape Function)로 정량화하고, 이를 이용하여 최적의 구성소를 선택하였다.

형상함수는 경계선상의 임의의 화소점에서 출발하여 반시계 방향으로 추적하면서 얻은 인접한 한쌍의 화소를 연결하는 각각의 선분에 (그림 2)에 표현된 방법으로 부호를 부여해서 얻은 동일부호의 갯수를 종속변수로 하는 함수로 정의한다. 여기서 독립변수는 체인코드 값이다. (그림 2)에 각각의 구성소에 대응되는 형상함수를 나타냈다.

3.3 원시형상 요소의 중심점 궤적 및 체인코드 에지벡터

제스처 시퀀스는 형상분해로부터 얻은 원시형상 요소의 중심점의 궤적으로 구성되는 체인코드 에지벡터를 정량화한 특징벡터를 이용해서 인식한다. 본 제스처 시퀀스 인식 알고리즘의 키 아이디어는 위치 및 방향 적응 영상처리를 적용하는 것으로 원시형상 요소의 방향성은 중심점의 궤적에 해당하는 체인코드를 이용하여 추출한다. 이러한 특징벡터들은 제스처 시퀀스 인식의 주 요소가 된다.

먼저 형상분해로부터 얻은 원시형상 요소들의 중심점은 원시형상요소의 추출의 위해 사용하기 위해 미리 예측된 구성소와 Streak 구성소를 사용하여 대상형상을 연속적으로 침식 연산한 결과들 중에서 1개의 화소만을 선택하여 설정하고, 원시형상 요소들의 중심점 중에서 선택한 출발 중심점과 시계방향으로 다른 중심점들과 연결되는 직선에 각도를 할당한다. 그리고 그림 3에 나타낸 바와 같이 45o의 배수로 표현되는 각도로 근사시킨 후에 {0, 1, 2, 3, 4, 5, 6, 7}중의 하나로 부호화한다. 이와 같은 방법을 사용하면 궤적의 분포도를 실제의 각도 분포 대신에 방향을 나타내는 수치로 간략하게 표현할 수 있다. 이와 같은 방법으로 얻은 8개의 값은  $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$  형식으로 표현되고, 체인코드 에지벡터는 다음과 같은 방식으로 정규화 한다.

$$X = x / (x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2)^{1/2} \dots\dots\dots (8)$$

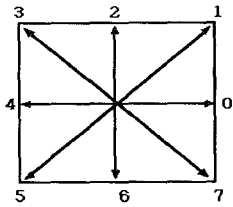
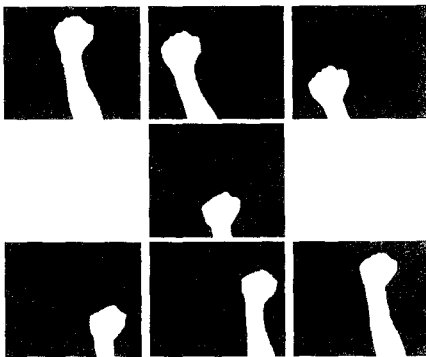


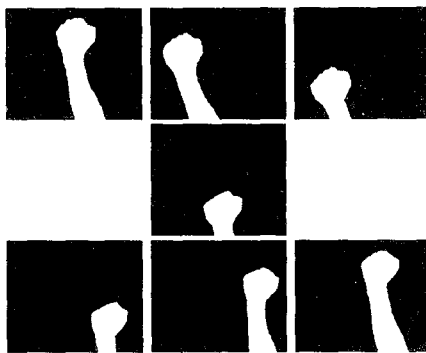
그림 3. 8 방향 체인코드  
Fig. 3. 8-directional chain code

### IV. 실험

화소수가  $256 \times 256$ 이고, 해상도가 8비트인 6개의 핸드 제스처 시퀀스 영상을 취득하여 제안된 알고리즘의 유효성을 확인하였다. 실험영상을 (그림 4)에 각각 나타냈다.



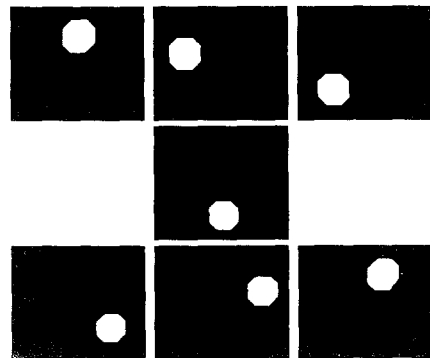
(a) Circle



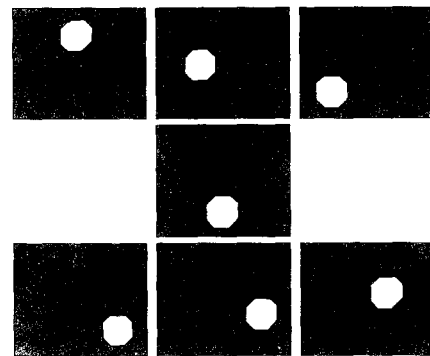
(b) Triangle

그림 4. 핸드 제스처 인식에 사용된 영상시퀀스  
Fig. 4. Image sequences for hand gesture recognition

각각의 핸드 제스처 시퀀스들은 형상에 따라 미리 예측한  $5 \times 5$  원구성소로 분해된다. 그 결과중에서 가장 큰 원시형상 요소만을 (그림 5)에 나타냈다.



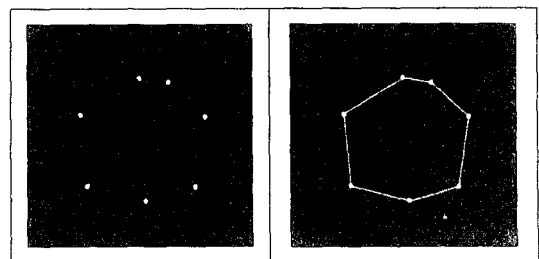
(a) Circle



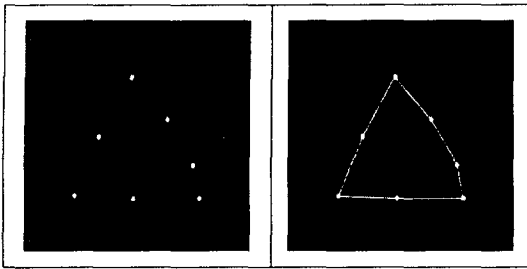
(b) Triangle

그림 5. 가장 큰 원시형상 요소  
Fig. 5. The largest primitive elements

형상분해된 원시형상 요소들의 중심점을 검출하기 위해 침식연산을 형상이 구성소보다 작게 될 때까지 반복수행한 결과를 (그림 6)에 나타냈다.



(a) Circle



(b) Triangle

그림 6. 원시형상 요소의 중심점  
Fig. 6. Center point of primitive elements

핸드 제스처 시퀀스를 형상분해 과정에서 사용하기 위해 미리 예측한 구성소를 이용하여 중심점의 궤적으로부터 얻은 체인코드 에지벡터 및 제스처 시퀀스에 해당하는 이상적인 체인코드 에지벡터와의 비를 <표 1>에 나타냈다. 에지벡터의 비로부터 핸드 제스처 시퀀스를 정확하게 인식할 수 있다는 것을 알 수 있다.

<표 1> 체인코드 에지벡터의 값  
<Table 1> Value of chain code edge vectors

Gesture	Elements of chain code edge vector							
	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>
Circle	68	34	55	36	53	36	55	32
Triangle	160	0	54	80	0	80	54	0
Gesture	Element of normalized chain code edge vector							
	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>
Circle	0.52	0.26	0.42	0.28	0.41	0.28	0.42	0.25
Triangle	0.76	0	0.26	0.38	0	0.38	0.26	0

### V. 결론

본 연구에서는 핸드 제스처 시퀀스를 이용해서 직감적이고 빠른 인터페이스 설계에 응용될 수 있는 형태론적 핸드 제스처 시퀀스 인식 기법으로 핸드 제스처 영상에서 추출된 형상영역을 원시형상 요소들로 분해한 다음, 원시형상 요소 간의 위치관계를 이용하여 핸드 제스처 시퀀스를 인식하는 방법을 제안하였다. 본 연구에서 새롭게 제안된 방법은 형태론적 형상분해를 이용하여 추출된 원시형상 요소의 중심

점을 선택하여 그 면적이 가장 큰 원시형상 요소들의 중심점의 궤적을 이용해서 계산된 특징벡터로 핸드 제스처 시퀀스를 인식하고자 하는 것이다.

본 연구에서 제안된 이론의 유효성을 검증하기 위한 실험을 통하여 비교적 단순한 핸드 제스처 시퀀스는 형태론적 연산을 적용하여 정확하게 인식할 수 있다는 것을 확인하였다. 따라서 본 연구에서 제안한 방식은 비디오데이터의 검색 및 기타 전자시스템 등의 작동에 관련된 인터페이스 설계에도 널리 응용될 수 있을 것으로 판단된다.

그러나 본 연구에서 대상으로 하고 있는 핸드 제스처 시퀀스는 비교적 단순한 형상이다. 따라서 복잡한 동작을 필요로 하는 시스템의 경우에는 더 복잡한 핸드 제스처 형상이 요구된다. 다양한 핸드 제스처 시퀀스를 이용하여 복잡한 시스템의 제어가 가능한 인터페이스를 설계하는 것이 본 연구의 다음과제이다.

### 참고문헌

- [1] Gray Bradski, Boon-Lock Yeo, and Minerva M. Yeung, "Gesture for Video Content Navigation", IS&T/SPIE Conference, California, 1999.
- [2] J. Yamato, J. Ohya and K. Ishii, "Recognition human action in time-sequential images using hiddenMarkov model", Proc. CVPR92, 1992.
- [3] T. Starner, J. Weaver and A.Pentland, "Real time American Sign Language recognition using desk and wearable computer based video", IEEE Trans. PAMMI, Vol.20, No.12, 1998.
- [4] Pitas, I. and Venetsanopoulos, A. N., "Morphological Shape Decomposition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12, No.1, 1990.
- [5] Serra, J., Image Analysis and Mathematical Morphology, Vol.1, Academic Press, New York, 1982.
- [6] Serra, J., "Introduction to Mathematical Morphology," Computer Vision, Graphics, and

Image Processing, Vol.35, No.3, 1986.

- [7] Maragos, P. and Schafer, R.W., "Morphological Skeleton Representation and Coding of Binary Images," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.ASSP-34, No.5, pp.1228-1244, 1986.
- [8] 이강호 안용학, 김학춘, "디지털 영상처리를 위한 에지 클래스의 설계" 한국컴퓨터정보학회 논문집, Vol 9, No 2, 2004.

**저 자 소 개**



**이 강 호**  
 중앙대학교 전자공학과 졸업(공학석사)  
 중앙대학교 전자공학과 졸업(공학박사)  
 현재 국립한국재활복지대학 정보보안과 교수  
 현재 한국컴퓨터정보학회 부회장  
 <관심분야> 정보보안, 디지털 영상처리



**최 중 호**  
 1982년 중앙대학교 전자공학 학사  
 1984년 중앙대학교 전자공학 석사  
 1987년 중앙대학교 전자공학 박사  
 현재 강남대학교 전자시스템공학부 교수  
 <관심분야> 컴퓨터시각, 영상컨텐츠, 제스처인식