
형태론적 손짓 인식 알고리즘

최종호*

Morphological Hand-Gesture Recognition Algorithm

Jong-Ho Choi*

요 약

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 중요한 아이디어는 형태론적 형상분해를 적용하여 제스처를 인식하는 것이다. 손짓 형상으로부터 얻은 원시형상요소들의 방향성은 손짓에 관한 중요한 정보를 내포하고 있다. 이러한 특징에 근거하여 본 연구에서는 주 원시형상요소와 부 원시형상요소의 중심점을 연결하는 직선으로부터 특징벡터를 이용한 형태론적 손짓 인식 알고리즘을 제안하고 실험을 통하여 그 유용성을 증명하였다. 자연스러운 손짓을 이용한 인터페이스 설계는 TV 스위치 조정이나 비디오 컨텐츠 검색용 시스템으로 널리 이용할 수 있을 것으로 판단된다.

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. A key idea of proposed algorithm in this paper is to apply morphological shape decomposition. The primitive elements extracted to a hand gesture include in very important information on the directivity of the hand gestures. Based on this characteristic, we proposed the morphological gesture recognition algorithm using feature vectors calculated to lines connecting the center points of a main-primitive element and sub-primitive elements. Through the experiment, we demonstrated the efficiency of proposed algorithm. Coupling natural interactions such as hand gesture with an appropriately designed interface is a valuable and powerful component in the building of TV switch navigating and video contents browsing system.

키워드

Morphology, Shape Decomposition, Hand Gesture Recognition, Navigation

I. 서 론

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[1-4]은 크게 인간의 물체에 글러브와 같은 여러 가지 타입의 센서를 부착해서 취득한 데이터를 해석하는 기기장착 방식과 비디오 카메라로부터 취득한 영상을 처리하는 시각기반 영상처리 방식으로 나눌 수 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)[5] 등은 인식 알고리즘이 복잡하여 하드웨어 구현 및 실시간 처리가 용이하지 않다는 문제점이 있다. 그러나 제스처 인식은 인간과 기계간의 인터페이스에 관련된 것이므로 하드웨어의 실시간 제어에 관련된 부분이 그 핵심이다.

따라서 본 연구에서는 직감적이고 빠른 인터페이스 설계에 응용될 수 있는 형태론적 손짓인식 기법에 대해 연구하였다. 2-D 영상내에 포함되어 있는 물체들의 복잡한 형상을 단순한 원시형상요소들로 분해하여 이들을 계층적으로 표현하는 것은 인간의 시각적인 인식의 기본단계이므로 손짓영상에서 추출된 형상영역을 원시형상요소들로 분해하였다. 이를 바탕으로 원시형상요소간의 위치관계를 이용하여 손짓을 인식하는 방법을 제안하고, 제안된 이론의 유효성을 증명하기 위한 실험을 통하여 손짓인식이 가능하다는 것을 확인하였다. 본 연구에서 새롭게 제안된 방법은 형태론적 형상분해를 이용하여 추출된 원시형상요소의 중심점을 선택하여 그 면적이 가장 큰 원시형상요소의 중심점과 그 밖의 원시형상요소의 중심점을 연결하는 특징벡터로 손짓을 인식하고자 하는 것이다. 본 연구에서 제안한 방식은 비디오 컨텐츠 검색 및 기타 전자시스템 등의 작동에 관련된 인터페이스 설계[6]에도 널리 응용될 수 있을 것으로 판단된다.

II. 손짓영상

본 연구에서는 제스처를 이용한 인터페이스 설계를 목적으로 표 1의 의미를 갖는 그림 1에 나타낸 4개의 손짓영상을 대상으로 하였다. 이러한 손

짓영상은 최소의 손운동으로 자연스럽게 동작하면서 각 형상들이 서로 뚜렷하게 구별될 수 있도록 고려하여 설정하였다[7].

표 1. 손짓 영상의 의미

Table 1. Meaning of hand gestures

Hand-Gestures	Functions
Play	Play video in the selection
Stop	Stop playing video
Left	Step to previous scene
Right	Step to next scene

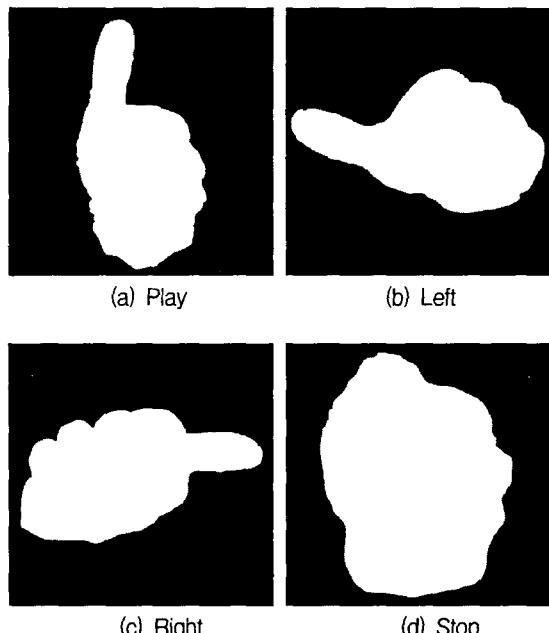


그림 1. 손짓영상

Fig. 1 Hand gesture images

III. 형상분해

2-D 영상내에 포함되어 있는 물체들의 복잡한 형상을 단순한 원시형상 요소들로 분해하여 이들을 계층적으로 표현하는 것은 인간의 시각적인 인식의 기본단계에 대응되는 처리이다. 본 연구에서는 제스처 형상을 형태론적으로 형상분해하여 얻은 원시형상 요소들을 이용하여 제스처 형상을 표현하였다.

형태론적 연산을 사용하여 2-D 형상 $X \ominus Z_2$ 를 여

러개의 집합 즉 원시형상 요소인 $\{X_i\}$ 로 분해하고자 하는 형상분해의 형태론적 표현은 다음과 같다.

$$X = \bigcup_{i=1}^n X_i, X_i \in G(Z2) \quad (1)$$

여기서 $G(Z2)$ 은 2-D 유클리드 공간 Z 에서 정의되는 2-D 개집합(open set)이다.

X_i 를 생성하기 위한 원시형상요소를 Y_i 라고 하고, Y_i 에 대응되는 구성소를 B 로 표현하면 형태론적 형상분해 알고리즘은 다음과 같이 표현된다. 원시형상요소 Y_i 의 가장 간단한 예는 단위 면적을 갖는 원판이나 정사각형 등으로 선택되는 구성소 B 의 스칼라 배인 $n_i B$ 이다.

$$X_i = X \circ Y_i = X \circ n_i B = (X \ominus n_i B_s) \oplus n_i B \quad (2)$$

여기서 B 는 구성소이고, n_i 는 구성소의 크기이며, \circ 및 \ominus 와 \oplus 는 각각 열기 및 확장과 침식연산자이다. 그리고 B_s 는 원점에 대한 B 의 반전(Reflection)이다.

식 (2)는 형상이 점 또는 선으로 축소될 때까지 형상 X 를 B 로 침식연산한 결과를 침식연산한 횟수 만큼 확장연산하면 원시형상 요소를 얻을 수 있다.는 것을 의미한다. 이러한 처리는 $X - X_i$ 에 대해서 반복적으로 수행된다. 이상에서 서술한 처리과정을 회귀형식으로 표현하면 다음과 같다.

$$\begin{aligned} X_i &= (X - X'_{i-1}) \circ n_i B \\ &= ((X - X'_{i-1}) \ominus n_i B_s) \oplus n_i B \\ X'_i &= \bigcup_{1 \leq j \leq i} X_j \\ X'_0 &= \emptyset \end{aligned} \quad (3)$$

Stopping Condition: $(X - X_{k'}) \ominus B_s = \emptyset$

여기서 종료조건은 형상의 전영역을 남김없이 분해하기 위한 조건이고, k' 는 분해된 원시형상 요소의 총수이다.

로 표현하는 과정은 다음과 같다.

- (1) 형상분해에서 첫 번째로 추출되는 가장 큰 원시형상요소 X_1 을 주요소로 설정한다.
- (2) 다음의 형상분해에서 추출되는 $X_i(i=2,3,\dots,n)$ 를 그 순서에 따라 제 1, 2, 3, ..., n 부요소로 설정한다.
- (3) 그럼 2에 나타낸 바와 같은 상위 두 부요소 사이에 위치한 원시형상요소는 손짓의 방향성 인식에 필요한 부분이 아니므로 형상표현에서 제거한다. 이 과정에서 상위 두 요소에 위치한 원시형상요소는 바로 위의 상위요소보다 주요소까지의 거리가 짧은 경우에 해당하는 원시형상요소이다.
- (4) 위의 과정을 $i=2$ 에서 n 까지 수행한다. 여기서 n 은 손짓인식이 가능한 특징벡터를 얻을 수 있을 때까지의 형상분해 횟수이다.

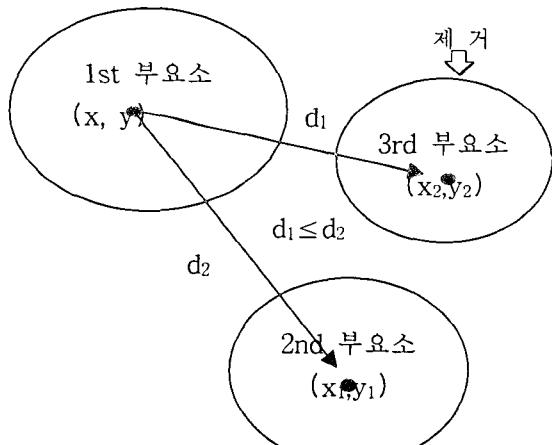


그림 2. 상위 두 부요소에 위치한 원시형상요소

Fig. 2 Primitive element in 2 sub-elements

2. 중심점 추출

본 연구에서 제안한 손짓 인식 알고리즘은 형태론적 형상분해로 추출한 원시형상요소들의 중심점을 연결하는 특징벡터를 이용하여 손짓을 인식하고자 하는 것이므로 형상분해 및 형상표현 다음 단계는 원시형상요소의 중심점을 추출하는 것이다.

형상의 중심점은 침식연산을 수행하여 결정한다. 먼저 Rohmbus 구성소를 사용하여 침식연산한 결과를 다시 3×3 정사각형 구성소로 침식연산하는 과정을 형상이 구성소보다 작아질 때까지 반복하여 수행하고, 그 결과를 다시 $90^\circ, 180^\circ, 45^\circ, 135^\circ$ 방향의 Streak 구성소로 침식연산하여 형상의 중심점을 결정한다.

IV. 손짓 인식

1. 형상표현

형태론적 형상분해 과정에서 원시형상요소들은 큰 것부터 작은 것의 순서대로 추출된다. 이러한 원시형상요소들을 다시 인접관계에 따라 계층적으로

3. 특징벡터 추출 및 인식

형상분해 및 형상표현 과정을 거쳐 2-D 공간상에 위치한 주 원시형상요소의 중심점과 부 원시형상요소들의 중심점을 연결하는 선의 각도를 특징벡터로 설정한다.

$$x = \{\theta_1, \theta_2, \theta_3, \dots, \theta_{n-1}\} \quad (4)$$

여기서 n 은 형상분해 과정에서 얻은 원시형상요소의 갯수이며, θ_{n-1} 은 주 원시형상요소의 중심점이 (x, y) 이고, $n-1$ 번 째 부 원시형상요소의 중심점이 (x_{n-1}, y_{n-1}) 인 경우 그 값은 다음과 같이 주어진다.

$$\theta_{n-1} = \tan^{-1}(|y - y_{n-1}| / |x - x_{n-1}|) \quad (5)$$

주 원시형상요소와 부 원시형상요소의 중심점을 연결하는 각도들로 구성된 특징벡터의 평균치는 다음과 같다.

$$x = (\theta_1 + \theta_2 + \theta_3, \dots, \theta_{n-1}) / (n-1) \quad (6)$$

식 (5)와 (6)으로부터 계산된 특징벡터의 계산치는 손짓에서 중요한 의미를 갖는 엄지손가락의 방향성에 관련된 정보이다. 그러므로 그림 3에 나타낸 바와 같이 특징벡터의 평균치가 $450\sim1340^\circ$ 인 경우에는 손짓 1(Play)으로, $1350\sim2250^\circ$ 인 경우에는 손짓 2(Left)로, 그리고 $2240\sim430^\circ$ 인 경우에는 손짓 3(Right)로 인식한다. 마지막으로 손짓 4(Stop)는 부 원시형상요소들이 주 원시형상요소 부근에 위치하는 경우이므로 주 원시형상요소와 부 원시형상요소간의 거리가 문턱치 이하인 경우에는 주 원시형상요소만이 존재하는 것으로 가정하여 부 원시형상요소를 제거하는 경우에 해당하므로 주 원시형상요소만이 존재하는 경우로 인식한다.

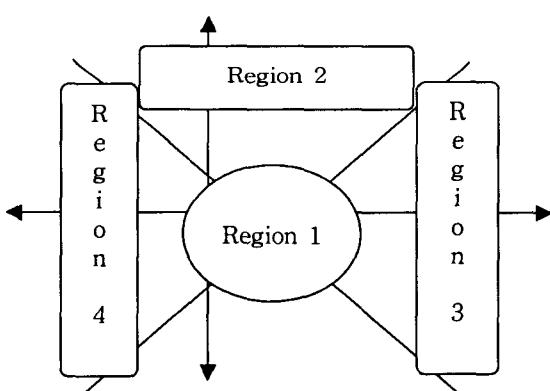


그림 3. 특징벡터 인식
Fig. 3. Recognition of feature vectors

V. 실험

손짓인식에서는 정확한 에지정보가 특징벡터의 계산에 영향을 미치지 않는다. 따라서 에지의 세밀한 변화는 불규칙 잡음으로 간주할 수 있다. 불규칙 잡음이 포함된 형상을 대상으로 형상분해를 수행하면 피크들의 양단에 존재하는 잡음 스파이크 때문에 기술효율이 떨어지게 된다. 이러한 문제를 해결하는 방법으로 형태론적 저역통과 필터를 사용하는 것을 들 수 있다.

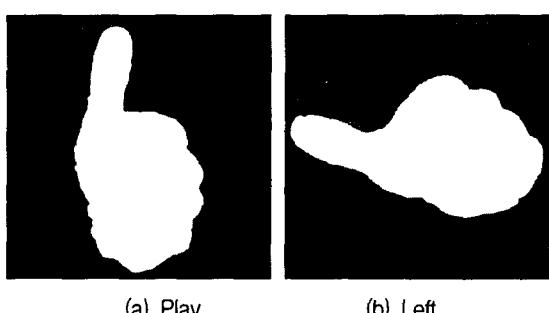
열기연산은 대상물체의 외부에 존재하는 날카로운 모서리 부분을 부드럽게 하고 구성소보다 작은 물체를 삭제하는 필터의 성질을 갖고 있다. 그리고 닫기연산은 협곡모양의 물체에 나타나는 작은 구멍을 메우는 성질을 갖고 있다. 이로부터 열기연산과 닫기연산은 각각 양의 잡음성분과 음의 잡음성분의 제거를 위한 필터로 사용할 수 있다는 것을 알 수 있다.

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

$$f_{lp}(m,n) = f(m,n) \circ B1(m,n) \bullet B1(m,n) \\ B2(m,n) \bullet B2(m,n) \dots \dots \quad (7)$$

여기서 $B1$ 과 $B2$ 는 각각 반경이 1과 2인 구성소이다.

CCD Video Color Module(TMC-7, PULNiX)과 개인용 컴퓨터에 내장된 Frame Grabber를 사용하여 화소수가 256×256 이고, 해상도가 8비트인 4개의 손짓 영상을 취득하여 제안된 알고리즘의 유효성을 확인하였다. 형태론적 저역통과필터를 사용하여 불규칙 잡음을 제거하고 에지를 평활화한 실현영상을 그림 4에 나타냈다.



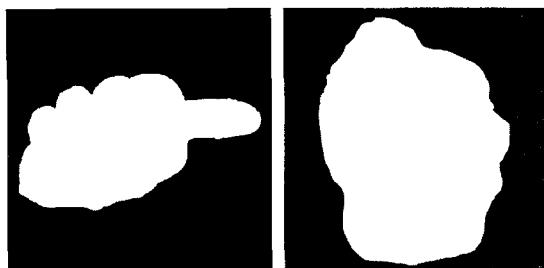


그림 4. 에지 평활화 및 불규칙 잡음제거 영상
Fig. 4 Edge smoothing and random noise reduction images

형상이 점 또는 선으로 축소될 때까지 형상 X 를 각각의 구성소로 침식연산한 결과를 침식연산한 횟수만큼 확장연산하여 원시형상 요소들을 얻는 처리를 $X \cdot X'^{-1}$ 에 대해서 반복적으로 수행함으로써 각각의 손짓 형상을 5×5 원 구성소로 분해하였다. 5×5 원 구성소를 사용하여 얻은 손짓 형상의 원시형상 요소들 중에서 제 5 부 원시형상요소까지의 원시형상 요소를 그림 5에 나타냈다.

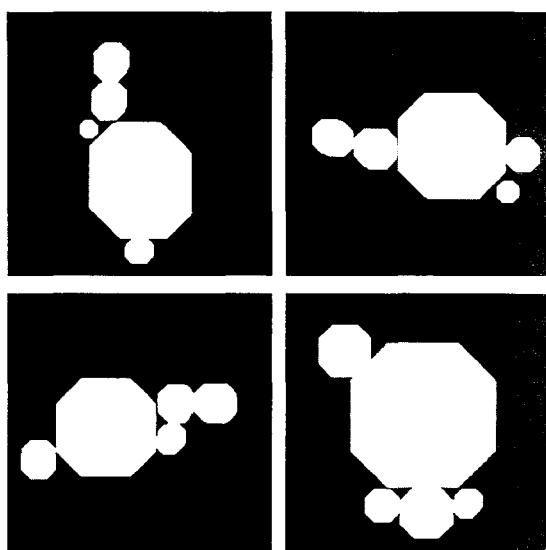


그림 5. 형상표현 결과
Fig. 5 Results of shape representation

특징벡터를 추출하기 위하여 각각의 손짓 형상에서 주 원시형상요소의 중심점을 부 원시형상요소들의 중심점을 연결하는 직선을 그림 6에 나타냈다.

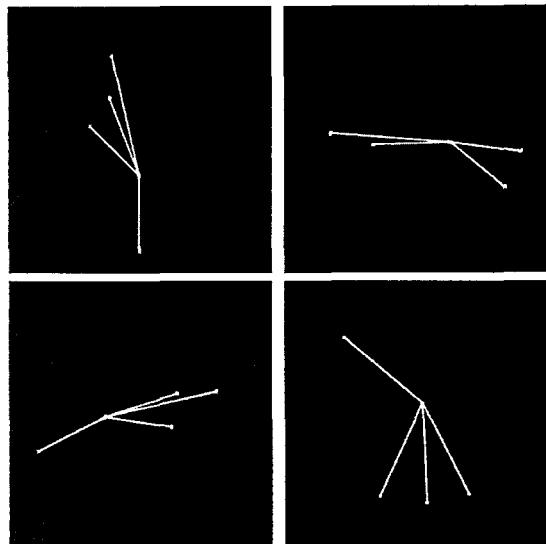


그림 6. 형상요소 연결선
Fig. 6. Connected line of shape elements

그림 6에서 원시형상요소의 중심점과 부 원시형상요소들간의 중심점을 연결하는 직선이 수평선과 이루는 각도를 계산한 다음 그 평균치를 계산한 결과를 표 2에 나타냈다.

표 2. 손짓 인식 결과
Table 2. Recognition result of hand gestures

Gesture	Mean Value	Class	Recognition
Play	115.72°	2	Play
Left	177.54°	3	Left
Right	15.24°	4	Right
Stop	Origin	1	Stop

표 2에서 각각의 손짓으로부터 계산된 특징벡터의 평균값이 각각 115.72° , 177.54° , 15.24° 로 각각의 이상적인 값이 90° 와 180° , 0° 인 점을 감안하면 손짓이 사용자마다 조금 다른 경우에도 설정범위에 충분히 포함될 것으로 판단되므로 정확한 인식이 가능하다는 것을 알 수 있다. Stop에 해당하는 주 원시형상요소의 중심점과 주 원시형상요소의 중심점을 연결하는 직선의 유클리드 거리가 설정된 임계값보다 작은 경우이므로 주 원시형상요소만 존재한다. 따라서 부 원시형상요소가 존재하지 않는 경우를 Stop으로 인식하면 4개의 손짓 형상을 정확하게 인식할 수 있다.

본 논문에서 제안한 알고리즘을 실용화레벨에서 검토하기 위해 실생활 환경을 고려하여 영상카메라로부터 취득한 RGB 영상을 YCbCr 공간으로 변환하고, $77 \leq Cb \leq 127$, $133 \leq Cr \leq 173$ 문턱치 값을 적용하여 손 형상을 추출한 후에 손짓 인식 알고리즘을 적용하였다. 그림 7에 손 형상 추출 결과를 나타냈다. 그림 8은 본 논문에서 제안한 형태론적 저역통과 필터 및 손짓 인식 알고리즘을 적용하여 얻은 형상분해 및 형상인식 결과이다.

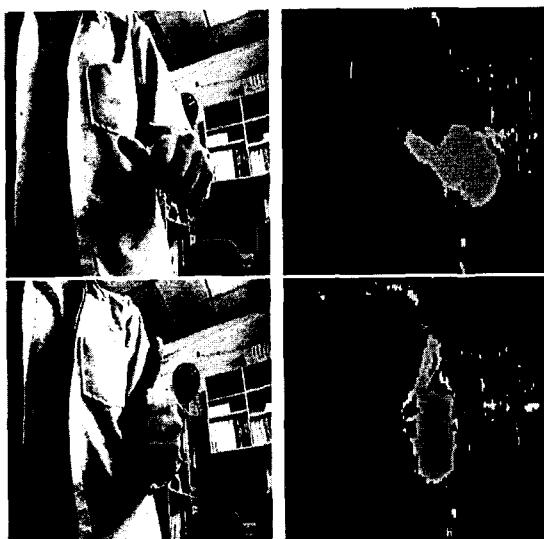


그림 7. 실험영상과 손 형상 추출 영상
Fig. 7 Source images and skin region images

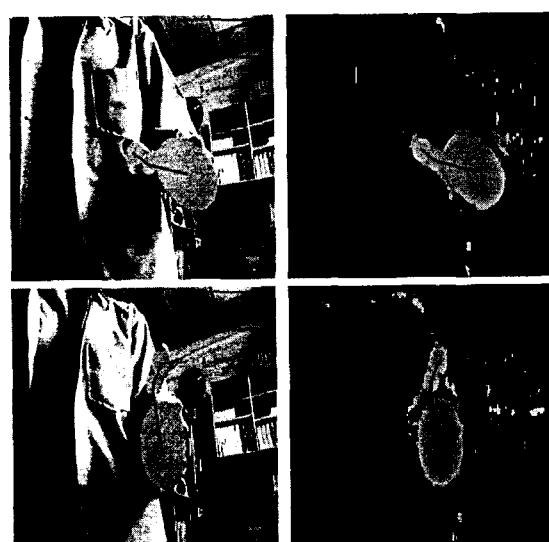


그림 8. 손짓 인식 결과
Fig. 8 Results of hand gesture recognition

VI. 결 론

본 연구에서는 손짓을 이용해서 직감적이고 빠른 인터페이스 설계에 응용될 수 있는 형태론적 손짓 인식 기법으로 손짓 영상에서 추출된 형상영역을 원시형상요소들로 분해한 다음 원시형상요소간의 위치관계를 이용하여 손짓을 인식하는 방법을 제안하였다. 본 연구에서 새롭게 제안한 방법은 형태론적 형상분해를 이용하여 추출된 원시형상요소의 중심점을 선택하여 그 면적이 가장 큰 원시형상요소의 중심점과 그 밖의 원시형상요소의 중심점을 연결하는 특징벡터로 손짓을 인식하고자 하는 것이다.

본 연구에서 제안된 이론의 유효성을 검증하기 위한 실험을 통하여 4개의 손짓 동작은 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 손짓 형상은 4개에 불과하다. 물론 본 연구에서 대상으로 하고 있는 비디오 컨텐츠 검색 및 가정용 전자시스템의 제어에는 쉽게 응용될 수 있지만 보다 복잡한 동작을 필요로 하는 시스템의 경우에는 더 많은 손짓 형상이 요구된다. 다양한 손짓을 이용하여 복잡한 시스템의 제어가 가능한 인터페이스를 설계하는 것이 본 연구의 다음 과제이다.

참고문헌

- [1] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Trans. PAMI, Vol. 19, No. 7, 1997.
- [2] T. Ahmad, C.J. Taylor, A. Lanitis, T.F. Cootes, "Tracking and Recognising Hand Gestures, using Statistical Shape Models", Image and Vision Computing 15, Elsevier, 1997.
- [3] Andrew D. Wilson and Aaron F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition", IEEE Trans. PAMI, Vol. 21, No. 9, 1999.

- [4] Min C. Shin, leonid V. Tsap, and Dmitry B. Goldgof, "Gesture recognition using Bezier curves for visualization navigation from registered 3-D data", Pattern Recognition, 37, 2004.
- [5] Feng-Sheng Chen, Chih-Ming Fu, and Chung-Lin Huang, "Hand Gesture Recognition using a Real-time Tracking method and Hidden Markov Models", Image and Vision Computing 21, 2003.
- [6] Tatio Frantti and Sanna Kallio, "Expert System for Gesture Recognition in Terminal's User Interface", Expert Systems with Applications 26, 2004.
- [7] Gray Bradski, Boon-Lock Yeo, and Minerva M. Yeung, "Gesture for Video Content Navigation", IS&T/SPIE Conference, California, 1999.

저자소개



최종호(Jong-Ho Choi)

1982년 중앙대학교 전자공학 학사
1984년 중앙대학교 전자공학 석사
1987년 중앙대학교 전자공학 박사
현재 강남대학교 전자시스템공학
부 교수

※ 관심분야 : 컴퓨터시각, 영상컨텐츠, 제스처인식