비디오 컨텐츠 검색을 위한 형태론적 손짓 인식 알고리즘

김 정 훈, *최 종 호, 최 종 수
중앙대학교 첨단영상전문대학원 영상공학과, *강남대학교 전자공학과
전화 : 02-820-5295 / 핸드폰 : 016-654-4737

Morphological Hand-Gesture Algorithm for Video Content Navigation

Jeong-Hun Kim, Jong-Ho Choi*, Jong-Su Choi
Dept. of Image Eng., GSAIM, Chung-Ang Univ., *Dept, Electronic Eng., Kang-Nam Univ.
E-mail : jhop@imagelab.cau.ac.kr

Abstract

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 real-time processing. A key idea of the algorithm proposed in this paper is to apply morphological shape decomposition. The primitive elements extracted from a hand gesture have very important information including the directivity of the hand gestures. Based on this algorithm, we proposed the morphological hand-gesture recognition algorithm using feature vectors extracted from lines connecting the center points of a main-primitive element and sub-primitive elements. Through the experiments, we applied to the video contents browsing system with natural interactions and demonstrated the efficiency of this algorithm.

I. 서론

인간과 컴퓨터의 상호 의사전달의 수단으로 제스처를 이용하고자 하는 연구가 최근 널리 진행되고 있다. 제스처를 이해하는 컴퓨터 인터페이스를 구현할 수만 있다면 제스처는 의사전달 및 감정이나 정서 등의 미묘한 표현에 대응하는 자연스러운 입력수단으로 널리 이용될 수 있을 것이다.

제스처 인식 기법은 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. 부착된 센서를 사용하는 방식은 움직임이 제한되고 동작이 부자연스러우며 착용시의 불편함과 심리적 부담 때문에 취급이 매우 불편하다는 문제점이 있다. 이러한 이유로 현재는 특별한 응용분야를 제외하고 시각기반의 영상처리를 이용하여 비접촉 방식으로 제스처를 인식하고자 하는 분야가 활발히 연구되고 있다.

시각기반 영상처리 방식은 센서를 부착하지 않아도 된다는 장점이 있지만 비디오 데이터의 분해능이 낮고, 시각기반 영상처리의 주 연구 대상인 모습기반 접근 방식에 주로 적용되고 있는 HMM과 신경망 모델 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[1][2][3].

따라서 본 연구에서는 하드웨어 구현이 용이하고 고속연산이 가능한 수학적 형태론을 적용한 형태론적 손짓 인식 알고리즘을 제안한다. 화소 사이의 논리적 연산에 근거하고 있는 수학적 형태론에서는 형태론적 논리연산으로 구성되는 유용한 영상처리 기술들이 다양한 형태로 개발되어 있다. 2차원 영상 내에 포함되어 있는 물체들의 복잡한 형상을 단순한 원시형상 요소들로 분해하여 이들을 계층적으로 표현하는 것은 인간의 시각적인 인식의 기본 단계이므로 손짓영상에서 추출된 형상영역을 원시 형상 요소들로 분해하였다. 이를 바탕으로 원시형상 요소간의 위치관계를 이용하여 손짓을 인식하는 방법을 제안하고, 제안된 이론의 유용성을 증명하기 위한 실험을 통하여 비디오 컨텐츠 탐색을 위한 손짓인식이 가능하다는 것을 확인하였다.

II. 형태론적 형상분해

2.1 형태론적 연산

형태론적 영상처리는 영상의 기하학적 특징을 해석하고자 하는 비선형처리 방식으로 집합론에 근거하고 있는 수학적 형태론에서의 영상은 평행이동(translation), 합집합(union), 교집합(intersection)등과 같은 집합 연산들을 수행할 수 있는 점들의 집합이다. A, B, C, K를 2차원 유클리드 공간 Z^2에서 정의되는 개집합(*open set*)으로, 그리고 O를 Z^2의 원점으로 표현하면 형태론적 영상처리 분야에서 널리 사용되고 있는 연산들은 다음과 같다[4][5][6].

확장(dilation) : $\boldsymbol{A \oplus B}$ (1)

침식(erosion) : $\boldsymbol{A \ominus B}$ (2)

열기(open) : $\boldsymbol{A \circ B = (A \ominus B) \oplus B}$ (3)

닫기(close) : $\boldsymbol{A \bullet B = (A \oplus B) \ominus B}$ (4)

여기에서 구성요소(structuring element)로 언급되는 B는 영상을 변환할 목적으로 사용하는 영상 패턴이다. 열기 연산은 침식연산을 수행한 결과를 다시 확장 연산하는 연산이며, 닫기 연산은 이와 쌍대 관계에 있는 연산 즉 확장연산을 수행한 결과를 다시 침식 연산하는 연산이다. 열기연산은 대상 물체의 외부에 존재하는 날카로운 모서리 부분을 부드럽게 하고 구성 요소보다 작은 물체를 삭제하는 필터의 성질을 갖고 있다. 그리고 닫기 연산은 협곡모양의 물체에 나타나는 작은 구멍을 메우는 성질을 갖고 있다. 이로부터 열기연산과 닫기 연산은 각각 양의 잡음 성분과 음의 잡음 성분의 제거를 위한 필터로 사용할 수 있다는 것을 알 수 있다.

2.2 형상 분해

2차원 영상 내에 포함되어 물체들의 복잡한 형상을 단순한 원시형상 요소들로 분해하여 이들을 계층적으로 표현하는 것은 인간의 시각적인 인식의 기본단계에 대응되는 처리이다. 본 연구에서는 제스처 형상을 형태론적으로 형상 분해하여 얻은 원시형상 요소들을 이용하여 제스처 형상을 표현하였다.

형태론적 연산을 사용하여 2차원 형상 $\boldsymbol{X \in Z^2}$를 여러 개의 집합 즉 원시형상 요소인 $\{\boldsymbol{X_i}\}$로 분해하고자 하는 형상분해의 형태론적 표현은 다음과 같다.

$$\boldsymbol{X = \bigcup_{i=1}^{n} X_i, \quad X_i \in G(Z^2)} \tag{5}$$

여기서 $G(Z^2)$는 2차원 유클리드 공간 Z에서 정의되는 2차원 개집합(open set)이다.

X_i를 생성하기 위한 원시형상 요소를 Y_i라고 하고 Y_i에 대응되는 구성소를 B로 표현하면 형태론적 형상분해 알고리즘은 다음과 같이 표현되며 원시형상 요소 Y_i의 가장 간단한 예는 단위 면적을 갖는 원판이나 정사각형 등으로 선택되는 구성소 B의 스칼라 배인 n_iB이다.

$$X_i = X_{n_i B} = (X \ominus n_i B^s) \oplus n_i B \tag{6}$$

여기서 B는 구성소이고, n_i는 구성소의 크기이며, B^s는 원점에 대한 B의 반전(reflection)이다.

식 (6)은 형상이 점 또는 선으로 축소될 때까지 형상 X를 B로 침식 연산한 결과를 침식 연산한 횟수만큼 확장연산하면 원시형상 요소를 얻을 수 있다는 것을 의미한다. 이러한 처리는 $X - X_K$에 대해서 반복적으로 수행된다. 이상에서 서술한 처리과정을 회귀형식으로 표현하면 다음과 같다.[7][8]

$$X_i = (X - X_i) n_i B, \quad X_i' = \bigcup_{j=1}^{i} X_j, \quad X_0' = \emptyset$$

- Stopping condition : $(X - X_K') \ominus B^s = \emptyset$

여기서 종료조건은 형상의 전영역을 남김없이 분해하기 위한 조건이고, k는 분해된 원시형상 요소의 총수이다. 형상분해 알고리즘을 적용하여 분해된 각각의 원시형상 요소는 점 또는 선으로 표현되는 궤적 L_i를 따라 n_iB를 평행 이동할 때 생기는 영역이다. 원시형

상 요소에 내접하는 최대 구성소의 궤적 L_i는 다음과 같이 쓸 수 있다.

$$L_i=\left(X-\bigcup_{0\le j\le i-1}^{i}(L_i\oplus n_iB)\right)\ominus n_iB^s \quad (8)$$

위의 식 (7)과 식 (8)을 결합하면 원시형상 요소와 궤적의 관계를 얻을 수 있다.

$$X_i=L_i\oplus n_iB \quad (9)$$

III. 형상 인식

3.1 형상 표현

형태론적 형상분해 과정에서 원시형상 요소들은 큰 것부터 작은 것의 순서대로 추출된다. 이러한 원시형상 요소들을 다시 인접관계에 따라 계층적으로 표현하는 과정은 다음과 같다.

(1) 형상분해에서 첫 번째로 추출되는 가장 큰 원시형상 요소들은 X_i을 주요소로 설정한다.

(2) 형상분해에서 추출되는 $X_i(i=2,3,\cdots,$끝$)$그 순서에 따라 제 $1,2,3,\cdots,$끝요소로 설정한다.

(3) 그림(1)과 같이 상위 두 부요소 사이에 위치한 원시형상요소는 손짓 방향성 인식에 필요한 부분이 아니므로 형상표현에서 제거한다. 이 과정에서 상위 두 요소에 위치한 원시형상 요소는 바로 위의 상위 요소보다 주 요소까지의 거리가 짧은 경우에 해당하는 원시형상 요소이다.

(4) 위의 과정을 $i=2$부터 n까지 수행한다. 여기서 n은 손짓 인식이 가능한 특징벡터를 얻을 수 있을 때까지의 형상분해 횟수이다.

다음 단계는 원시형상 요소의 중심점을 추출하는 것이다. 중심점은 침식연산을 구성소보다 작아질 때까지 반복수행하여 얻어낸다.

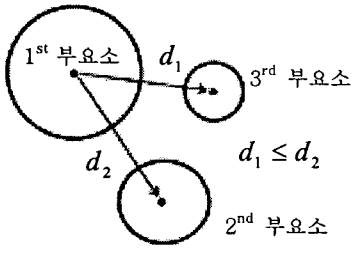

그림 1. 상위 두 부요소에 위치한 원시형상 요소

3.2 특징 벡터 추출 및 인식

형상 분해 및 형상 표현 과정을 거쳐 2차원 공간상에 위치한 주 원시형상 요소의 중심점과 부 원시형상 요소들의 중심점을 연결하는 선의 각도를 특징 벡터로 설정한다.

$$\boldsymbol{x}=\{\theta_1,\theta_2,\cdots,\theta_{n-1}\} \quad (10)$$

여기서 n은 형상분해 과정에서 얻은 원시형상 요소의 개수이며, θ_{n-}은 주 원시형상 요소의 중심점이 (x,y)이고 $n-$번째 주 원시형상 요소의 중심점이 $(x_{n-1},y_{n-}$일 경우 그 값이 다음과 같이 주어진다.

$$\theta_{n-1}=\tan^{-1}\left(\frac{|y-y_{n-1}|}{|x-x_{n-1}|}\right) \quad (11)$$

주 원시형상 요소와 부 원시형상 요소의 중심점을 연결하는 각도들로 구성된 특징벡터의 평균치는 다음과 같다.

$$x=\frac{(\theta_1+\theta_2+\theta_3+\cdots+\theta_{n-1})}{(n-1)} \quad (12)$$

식 (10)과 (11)을 적용하여 계산된 특징벡터의 계산치는 손짓에서 중요한 의미를 갖는 엄지 손가락과 방향성에 관련된 정보이다.

3.3 손짓 인식

그림 (2)와 같이 특징 벡터의 평균치가 45° ~134° , -315° ~44° , 135° ~224° 에 따라 각각 영역 2, 3, 4로 인식하며 영역 1인 부 원시형상 요소들이 주 원시형상 요소 부근에 위치하는 경우에는 주 원시형상 요소만이 존재하는 것으로 가정하여 부 원시형상 요소를 제거하는 경우에 해당하므로 주 원시형상 요소만이 존재하는 경우로 인식한다.

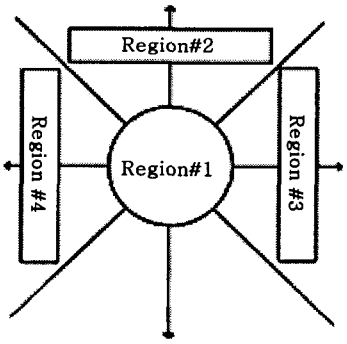

그림 2. 특징 벡터 인식 영역

IV. 실험 결과

그림 3은 취득한 영상(320x240 RGB)을 YCbCr 모델로 변환하여 사람의 피부색을 검출한 영상이다 (77<Cb<127, 133<Cr<173). 그림 (4)의 (a)와 (c)는 형상분해를 이용해 상위 3요소를 나타낸 결과를, (b)와 (d)는 특징 벡터를 이용하여 인식한 결과를 나타낸다.

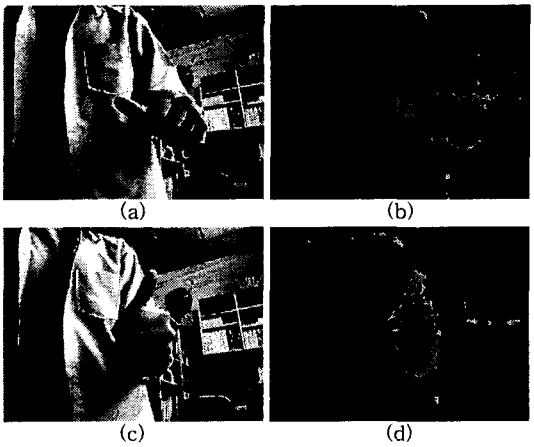

그림 3. (a) (c) 원 영상, (b) (d) 피부영역 검출 결과

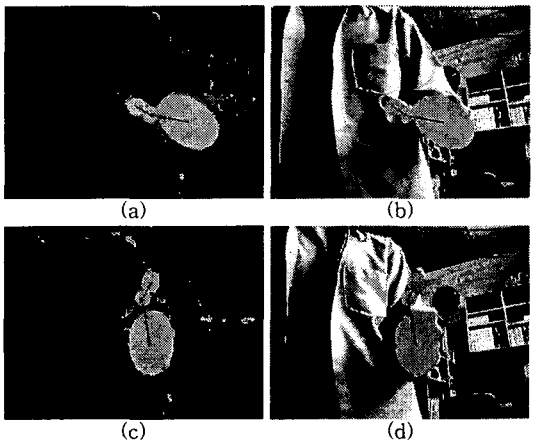

그림 4. (a)(b) 형태분해 결과, (c)(d) 손짓 인식 결과

다음의 표 1은 비디오 컨텐츠 검색을 위한 영역과 그의 기능을 나타내며 각 영역별의 손짓 인식한 결과에 따른 특징 벡터의 각도 평균값을 보인다.

표 1. 손짓 영역별 기능 및 특징 벡터 각도 평균값 결과

손직 인식 영역	기능	평균값
Region 1	Play	93.2°
Region 2	Forward	172.8°
Region 3	Rewind	23.4
Region 4	Stop	x

V. 결론

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 비디오데이터의 검색 및 기타 전자시스템 등의 작동에 관련된 인터페이스 설계에도 널리 응용될 수 있을 것으로 판단된다.

Acknowledge : 본 연구는 교육부의 두뇌한국21 사업(BK21) 및 과기부의 국가지정 연구실(NRL) 사업의 연구비 지원으로 수행되었습니다.

[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] Serra, J., "Introduction to Mathematical Morphology", *Computer Vision, Graphics, and Image Processing*, Vol. 35, No. 3, 1986.

[5] Serra, J., *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1982.

[6] Maragos. P, "A Representation Theory for Morphological Image and Signal Processing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 6, 1989.

[7] Pitas, I. And Venesanopoulos, A. N., "Morphological Shape Decomposition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No. 1, 1990.

[8] Pitas, I. And Venetsanopoulos, A. N., "Morphological Shape Representation", *Pattern Recognition*, Vol. 25, No. 6, 1992.