Journal of Korea Multimedia Society Vol. 8, No. 10, October 2005(pp. 1337-1346)

지능형 로봇 제어를 위한 제스처 인터페이스

오재용[†], 배기태[††], 김만진[†††], 이칠우[††††]

요 약

본 논문에서는 지능형 로봇의 구현을 위한 효율적인 제스처 인식 방법에 대하여 기술한다. 기존의 2차원 기반 제스처 인식 방법들은 제스처의 3차원 특성을 정확하게 표현할 수 없으며, 3차원 정보를 추가로 사용하는 경우 3차원 데이터의 에러와 시스템의 복잡성 때문에 그 활용에 제약이 따랐다. 본 논문에서는 이들 단점을 보완하기 위하여, 3차원 공간에서의 제스처를 효율적으로 정량화하는 방법으로 2차원 형상 정보와 3차원 깊이 정보를 동시에 포함하는 제스처 캡처링 모델 *APM(active plane mode)*을 제안한다. *APM*은 외관 특성 및 영상의 노이즈에 덜 민감한 특징이 있으며, 행위자의 변화에도 보다 안정적인 인식을 수행할 수 있는 장점이 있다. 이렇게 추출된 제스처 특징은 주성분 분석법(PCA)과 은닉 마르코프 모델(HMM)을 이용하여 분석되고, 최종적으로 제스처를 인식하게 된다. 본 방법은 서로 다른 15명의 제스처에 대해 실험한 결과 90% 이상의 인식 결과를 보였으며, 지능형 로봇뿐만 아니라 지적 인터페이스 시스템과 같은 여러 응용 시스템에 적용될 수 있을 것이다.

Gesture Interface for Controlling Intelligent Humanoid Robot

Jae Yong Oh[†], Ki Tae Bae[††], Man Jin Kim[†††], Chil Woo Lee[††††]

ABSTRACT

In this paper, we describe an algorithm which can automatically recognize human gesture for Human-Robot interaction. In early works, many systems for recognizing human gestures work under many restricted conditions. To eliminate these restrictions, we have proposed the method that can represent 3D and 2D gesture information simultaneously, *APM*. This method is less sensitive to noise or appearance characteristic. First, the feature vectors are extracted using *APM*. The next step is constructing a gesture space by analyzing the statistical information of training images with PCA. And then, input images are compared to the model and individually symbolized to one portion of the model space. In the last step, the symbolized images are recognized with HMM as one of model gestures. The experimental results indicate that the proposed algorithm is efficient on gesture recognition, and it is very convenient to apply to humanoid robot or intelligent interface systems.

Key words: Gesture Recognition(제스처 인식), Principle Component Analysis(주성분 분석법), HMM(은닉마르코프 모델), Motion Analysis(동작분석)

※ 교신저자(Corresponding Author) : 오재용, 주소 : 광주광역시 북구 용봉동(500-757), 전화 : 062)530-0258, FAX: 062)530-1759, E-mail : ojyong@image.chonnam.ac.kr
접수일 : 2005년 3월 21일, 완료일 : 2005년 5월 23일
[†] 준회원, 전남대학교 대학원 컴퓨터 정보통신공학과 박사과정
[††] 준회원, 전남대학교 대학원 컴퓨터정보통신공학과 (E-mail : bkt2002@empal.com)
[†††] 전남대학교 대학원 컴퓨터정보통신공학과 (E-mail : mjkim@image.chonnam.ac.kr)
[††††] 종신회원, 전남대학교 전자컴퓨터정보통신 공학부 교수 (E-mail : leecw@chonnam.ac.kr)
※ 본 연구는 전남대학교 "고품질 전기전자부품 및 시스템 연구센터" 및 KIST "네트워크 기반 휴머노이드기술 개발사업"의 연구비 지원에 의해 수행되었음.

1. 서 론

최근 인간의 제스처에 대한 관심이 높아지고 컴퓨터 시스템이 급속도로 발달하면서 인간 친화적 인터페이스와 같은 분야에 인간의 제스처를 기반으로 하는 기술이 다각도로 응용되고 있다. 특히 일상생활에서 가사 혹은 엔터테인먼트 등을 목적으로 하는 생활 지원 로봇들의 개발이 활발히 진행됨에 따라 인간과 닮은 로봇, 인간처럼 행동하는 로봇을 구현하기 위해서 인간과 로봇간의 자연스러운 의사소통 수단의 확보가 중요한 문제가 되었다.

인간은 80%이상의 정보를 시각을 통하여 획득한다고 알려져 있다. 다시 말해서, 시각정보는 일상생활에서 매우 많은 비중을 차지하며, 이를 통한 의사소통이 가장 자연스러운 것임을 알 수 있다. 인간의 눈과 대응하는 것이 로봇의 카메라이며, 로봇은 카메라를 통해 입력되는 영상을 분석하여 외부 상황을 자율적으로 판단하게 되는 것이다. 또한 인간은 언어 이외에도 제스처와 같은 비언어적 수단을 이용하여 의사소통을 하며, 이러한 비언어적 의사소통 수단을 로봇이 이해한다면, 로봇은 인간과 보다 친숙한 대상이 될 수 있을 것이다. 그러나 인체는 매우 복잡한 3차원 관절 구조를 가지고 있을 뿐 아니라 신체 부위에 따라 각기 다른 의미를 표현 할 수 있기 때문에 로봇이 자동으로 제스처를 인식하는 것은 매우 어려운 일이다.

최근 신체에 특별한 장치를 부착하지 않고 제스처를 인식하는 방법으로 카메라를 이용하는 방법이 많이 연구되고 있으며, 주로 2차원적 제스처 특징을 사용한다[1]. 그러나 이러한 방법들은 제스처의 3차원 특성을 정확하게 표현할 수 없으며, 서로 다른 동작이라 할지라도 유사한 모습으로 보일 수 있기 때문에 안정적인 모델 구성이 어렵다는 단점이 있다. 이러한 단점을 보완하기 위하여 여러 가지 방법으로 추출된 제스처의 3차원 특징을 사용하지만, 3차원 데이터의 에러와 시스템의 복잡성 때문에 그 활용에 제약이 따른다[2].

본 논문에서는 위와 같은 단점을 보완하고, 제스처의 특징을 효율적으로 추출하기 위하여 APM을 제안한다. 제안하는 알고리즘은 제스처의 2차원 형상 정보와 3차원 깊이 정보를 동시에 표현할 수 있는 특징이 있으며, 행위자 독립적인 제스처 특징을 추출할 수 있다는 장점이 있다. 또한 행위자의 의도가 포함되지 않은 제스처의 영향을 줄여 안정적인 모델을 구성할 수 있는 효과를 얻을 수 있다.

본 논문은 다음과 같은 순서로 구성된다. 먼저 2절에서는 제스처 인식 기술의 연구 동향을 살펴보고, 3절에서는 본 논문에서 제안하는 제스처 인식 시스템에 대하여 기술한다. 4절에서는 다양한 환경에서의 실험을 통하여 제안하는 알고리즘의 타당성을 확인하고, 5절에서 결론을 정리하고 앞으로 개선될 연구 방향에 대하여 기술한다.

2. 제스처 인식 기술의 연구 동향

2.1 제스처 인식 기술의 개요

'제스처'의 사전적 의미는 "1)표현의 수단으로서 팔다리 또는 신체의 사용, 2)생각, 감정, 태도를 표현하거나 강조하는 신체나 팔다리의 움직임"으로 정의된다[3]. 마찬가지로 HCI(human and computer interaction)의 관점에서 제스처의 의미도 무심코 행한 움직임이 아닌, 의미를 전달하는 움직임이나 기계와 컴퓨터를 조작하기 위한 모든 움직임을 일컫는다. 이와 같이 컴퓨터나 로봇이 자율적으로 인간의 행동을 분석하고 인지하는 기술을 제스처 인식 기술이라고 한다.

최근 인간과 정보 시스템 간에 자연스럽게 정보를 교환할 수 있는 지적 인터페이스에 대한 관심이 높아지고, Smart Environment, 웨어러블 컴퓨터, 유비쿼터스 컴퓨팅과 같이 제 4세대 정보기술의 중요성이 강조되면서, 제스처 인식은 컴퓨터 비전 연구자들의 많은 주목을 받고 있다. 제스처를 인식한다는 것은 인체 각 부위가 시간의 경과에 따라 어떻게 변화하는가를 자동으로 분석하고, 그 변화를 추상적인 의미로 해석하는 것을 의미한다. 즉 동영상으로부터 신체 영역을 추출한 다음 각 부분들이 하나의 의미를 갖기 위해 어떤 변화를 거치는지를 알아내는 것이다. 그러나 인체는 고자유도를 지닌 매우 복잡한 3차원 관절 물체로 동영상으로부터 인체부위를 안정적으로 분리해 내고 그 내용을 인식한다는 것은 매우 어려운 일이다. 또한 행위자에 따라 착용하는 의복이 다르므로 특징정보를 안정적으로 추출하기가 어려워 많은 노력에도 불구하고 만족할 만한 결과를 얻기가 어려웠다.

일반적인 제스처 인식 기술은 그림 1과 같이 특징

추출, 분석, 학습, 인식의 단계로 구성된다. 카메라 혹은 센서로부터 움직임 정보를 취득하고, 이로부터 제스처를 구별할 수 있는 특징정보를 추출한다. 이렇게 추출된 특징정보와 미리 학습된 모델 제스처를 비교하여 행위자가 어떤 제스처를 행했는지를 인식하게 된다.

2.2 센서 기반 제스처 인식 기술

인간의 제스처는 3차원 공간에서 매우 복잡한 구조를 가지고 있기 때문에 그 움직임을 수치적으로 정량화하는 일은 매우 어렵다. 그러나 변화량을 측정할 수 있는 부위에 물리적인 센서나 마커를 부착하면 움직임의 위치와 방향을 정확히 추출할 수 있다[4]. 인간의 제스처를 수치적으로 표현하는 작업을 모션 캡처라고 하고, 이를 위한 장비는 작동 방식에 따라 크게 기계식, 자기식, 광학식으로 분류할 수 있다. 주로 자기식과 광학식 방법을 많이 사용하며, 손, 발, 팔꿈치 등과 같이 동작의 주가 되는 신체 부위에 기구를 부착하여 데이터를 얻어낸다. 그러나 이러한 방법들은 센서를 연결하는 장치들을 모두 몸에 부착해야 하므로, 동작의 제약이 있을 수 있다. 한편 광학식의 경우 별다른 장비 없이 적외선 카메라에 반응하는 마커를 몸에 부착하고, 그 마커의 궤적을 추적함으로서 정확한 모션 데이터를 추출할 수 있기 때문에, 캐릭터 애니메이션이나 컴퓨터 그래픽스 분야에서 최근 많이 사용되고 있는 추세이다. 그러나 광학식 모션 캡쳐 시스템의 경우 다른 마커에 가려서 마커가 보이지 않는 마커들 간의 중첩(occlusion) 문제가 발생하여 3차원 좌표를 얻는 것이 불가능하고, 이 때문에 많은 후처리 과정을 필요로 하게 되며, 그 결과 실시간 처리가 불가능하게 되거나 모션 캡쳐 성능을 떨어뜨리는 요인이 되기도 한다.

2.3 영상 기반 제스처 인식

앞 절에서 언급한 센서 기반의 방법들은 신체에 많은 장비들을 부착해야 한다는 큰 단점을 가지고 있다. 특히 기계식 혹은 자기식 방법의 경우 케이블의 길이에 따른 행동의 제약을 받게 된다. 이러한 문제점들을 해결하기 위하여 센서를 부착하지 않고 카메라를 통해 입력되는 영상을 분석하여 제스처를 인식하는 방법이 제안되었다. 객체 추출 및 추적 등의 영상처리 기술을 제스처 인식에 응용함으로서, 환경의 제약을 완화시키고, 보다 자연스럽게 인간과 컴퓨터 간에 의사소통을 할 수 있게 되었다. 영상 기반 제스처 인식 방법은 크게 특징기반, 외관 기반, 3차원 모델 기반의 3가지로 분류될 수 있다.

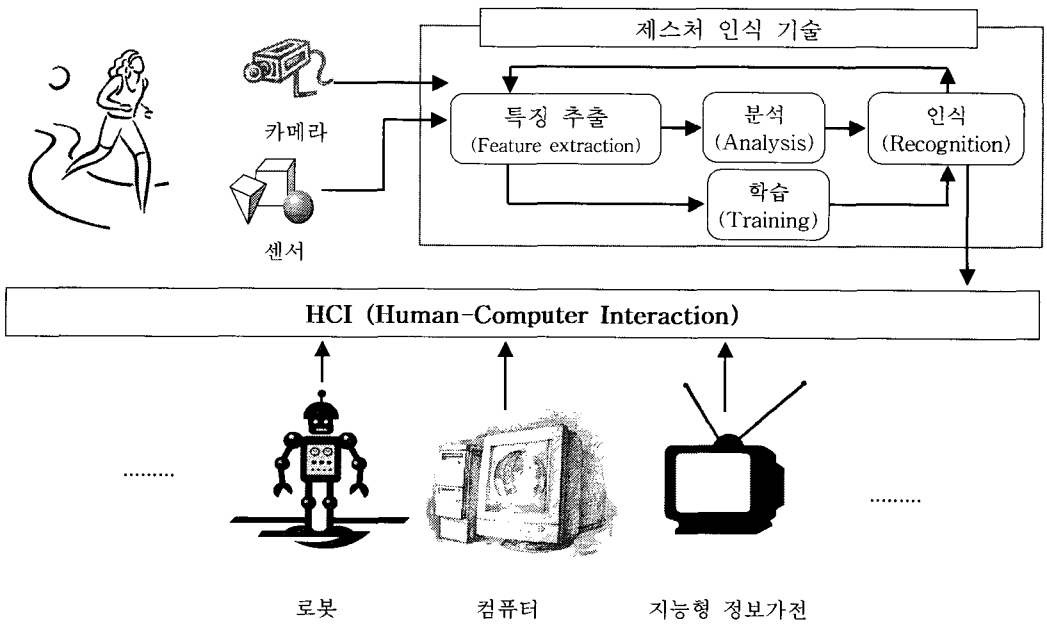


그림 1. 제스처 인식 알고리즘의 개요와 응용

간단한 제스처의 경우에는 복잡한 파라미터를 구하지 않고도 에지, 윤곽선, 특징점의 위치 등의 정보를 이용하여 쉽게 구별할 수 있다. 이 방법에서는 명확한 특징 추출의 여부가 관건이며, 추출된 특징을 이용하여 제스처 모델을 구성하고 입력 영상과 제스처 모델의 비교를 통하여 제스처를 인식한다. 그러나 특징으로 사용되는 에지 및 윤곽선 정보가 주위 환경에 매우 민감하기 때문에, 일반적으로 환경에 대한 제약조건과 함께 사용된다.

에지나 윤곽선 이외에 제스처의 특징으로 손이나 발, 얼굴 등의 위치 정보도 사용된다[5]. 인간은 비언어적 정보를 전달하는 수단으로 손, 발, 얼굴 등의 움직임을 많이 사용하고 있기 때문에, [5]에서는 제스처를 시·공간상에서 정의되는 특정 부위 움직임 조각의 집합이라고 가정한다. 그 후, 얼굴과 양손의 공간상 위치를 dynamic k-means 클러스터링 방법을 이용하여 자동으로 분류한 뒤, 분류된 클러스터의 시간에 따른 상태 천이를 유한 상태 기계(finite state machine : FSM) 알고리즘을 이용하여 제스처를 학습하고, 인식하게 된다. 이와 같은 방법은 계산량이 적고, 처리 속도가 빠르기 때문에 실시간 인식 시스템에 적합하지만, 주위 환경의 영향에 민감하고 특징점의 중첩이 발생할 경우 인식이 어렵다는 단점이 있다.

이와 같은 특징 기반의 제스처 인식 방법은 인식 결과가 주위 환경의 영향을 많이 받기 때문에 매우 불안정하다. 이러한 단점을 보완하고자 영상의 기하학적 특징을 이용하지 않고 영상 자체가 가지는 음영 정보를 그대로 이용하고자 하는 방법이 외관 기반(appearance based) 제스처 인식 방법이다. 실루엣(silhouette)이라고 표현되는 입력 영상의 음영 정보는 배경 이미지를 제거한 전경의 이진화 영상이며, 이를 분석함으로서 제스처를 인식할 수 있다. 대표적인 외관기반 제스처 인식 방법은 MHI(motion history image)다[1]. MHI는 입력 영상 시퀀스에서 더 최근에 움직인 영역의 화소들을 더 밝은 값으로 나타낸 영상으로서, 이 영상을 다시 9개의 세부 영역으로 나누고, 각 영역에서의 히스토그램을 모델 제스처와 비교하여 제스처를 인식하는 방법이다.

Ismail Haritaoglu의 알고리즘은 인간의 신체를 6개의 영역(cardboard model)으로 나누어 이를 분석하는 방법을 사용한다[6]. 이 시스템에서는 똑바로 선 상태에서의 제스처로 모델을 제한하여, 실루엣 영상의 볼록한 부분(convex hull)을 이용하여 다양한 형태의 제스처를 인식할 수 있는 시스템으로, 실루엣 영상에서 수직 및 수평 히스토그램을 이용하여 대략적인 제스처를 찾아 낸 뒤, Recursive Convex hull 알고리즘과 위상적인 분석을 통하여 최종 신체 특징점을 결정한다. 이 방법은 제스처 영상의 형상 정보와 세부 정보를 조합하여 사용하고, 신체 영역에 따른 계층적 분석방법을 사용했다는 점에서 주목할 만하다.

한편, 2차원의 영상정보는 3차원의 제스처를 표현하는데 한계가 있다. 앞 절에서 언급한 특징점의 중첩과 같은 문제는 2차원 영상이 갖는 모호성 때문에 발생되는 것이며, 이를 해결하기 위하여 3차원 시각정보를 사용한다[2]. 3차원 시각정보는 스테레오 카메라 영상의 깊이(depth)정보를 이용하여 획득할 수 있다. [2]에서는 깊이 정보를 이용하여 손 제스처 영역을 추출한 뒤, 손 영상의 기하학적 형태와 움직임 궤적을 분석하여 제스처를 인식하여, 로봇에게 의사를 전달하는 수단으로 사용한다. 또한, 머리와 양손의 3차원 데이터를 추출하고, 이를 제스처 인식에 응용하기도 한다[7]. 머리와 양손의 3차원 위치는 스테레오 기하를 기반으로 계산되며, 블랍(blob)의 모멘트 정보를 이용하여 회전각을 계산한다. 이렇게 계산된 데이터는 가상현실에서 인간-컴퓨터간의 인터페이스에도 응용되며, 시간상의 궤적을 이용한 제스처 인식 방법에도 응용이 가능하다. 그러나 카메라 보정 등의 사전작업이 필요하며, 복잡한 3차원 계산 과정에서 오차가 발생할 수 있다는 단점이 있다.

제스처 인식에서 가장 어려운 작업은 제스처를 수치적으로 잘 표현할 수 있는 특징을 선택하는 일이다. 그러나 3차원의 다관절체로 구성된 인간의 움직임을 표현하는 일은 매우 어려운 일이며, 모든 데이터를 제스처에 사용할 수도 없다. 이러한 배경에서 인간의 골격 모델을 기반으로 단순화된 3차원 모델을 생성하고, 이를 기준으로 움직임을 분석하는 연구도 진행되고 있다[8]. [8]에서는 원통형의 단순화된 모델을 입력 영상과 비교하여 자세를 추정하고, 이 데이터를 바탕으로 제스처를 인식한다. 이 시스템은 양 손을 사용하는 9가지 정도의 간단한 제스처를 인식하여 HCI로 응용된다. 그러나 이 방법은 단순화된 3차원 모델을 사용하기 때문에 정교한 제스처 인식

에는 부적합하며, 영상 기반 시스템이 갖는 배경 및 조명에 따른 문제점을 가지고 있다.

3차원 제스처 정보를 획득하는 다른 방법으로 여러 대의 카메라를 이용한 체적(volumetric) 모델을 사용하기도 한다[9]. [9]에서는 여러 대의 카메라로부터 입력받은 영상을 합하여 3차원 모델로 복원함으로서 제스처를 해석한다. 여러 각도에서 입력되는 영상을 사용하기 때문에 정교한 움직임을 추출할 수 있다는 장점이 있지만, 실시간으로 3차원 복원을 수행하는데 많은 계산량이 필요하다. 실제로 이 시스템에는 9 대의 클러스터링 된 컴퓨터가 사용되었다.

3. 제스처 인식 시스템의 구성

3.1 제스처의 정의

인간의 동작은 매우 다양하기 때문에 본 논문에서는 그림 2와 같이 지능형 로봇의 제어에 필요한 기본적인 동작을 정의하고 인식의 대상으로 정하였다.

3.2 제스처 특징 추출

일반적인 환경에서 획득된 비디오 영상에는 수많은 물체들이 포함되어 있다. 영상 내의 이러한 부분은 행위자의 제스처 정보를 추출하는 과정에서 불필요한 부분이기 때문에 반드시 제거 되어야 한다. 배경 제거에 있어서 카메라가 고정되어 있는 경우 미리 생성된 배경 모델을 이용할 수 있지만, 로봇에 장착된 카메라와 같이 카메라가 움직이는 경우에는 적용이 불가능 하다. 따라서 본 논문에서는 3차원 거리 정보를 기반으로 하는 간단한 배경 제거 방법을 사용한다.

배경제거의 첫 번째 단계로 얼굴 영역을 탐색한다[10]. 탐색된 얼굴영역을 기준으로 행위자까지의 거리를 구하며, (1)에서와 같이 행위자의 앞쪽에 위치하는 영역을 전경 영역으로 선택한다. 이와 같은 방법은 카메라와 행위자 사이에 장애물이 없어야 하며, 두 명 이상의 행위자가 존재하지 않는다는 제약 조건을 가진다.

$$F(x,y)=\begin{cases} 0 & : \ D(x,y) > D_f + c \\ F(x,y) & : \ D(x,y) \le D_f + c \end{cases} \quad (1)$$

(1)에서, $F(x,y)$는 스테레오 카메라로부터 입력되는 시차(disparity)영상이고, $D(x,y)$는 카메라로부터의 거리이며, D_f는 얼굴 영역의 카메라로부터 거리이다.

앞서 기술한 바와 같이 제스처의 2차원적 특징이 갖는 모호성과 3차원적 특징의 복잡성은 제스처 인식을 수행하는데 있어서 많은 문제점들을 야기 시킨다. 이러한 문제점들을 보완하고자 본 논문에서는 제스처의 2차원 형상정보와 3차원 거리정보를 동시에 표현하는 APM 모델을 제안한다.

APM은 행위자의 앞쪽에 3차원의 변형이 가능한 가상의 평면을 위치시킨 뒤, 행위자의 제스처에 따라 변형되는 평면의 형태 정보를 제스처 특징으로 사용한다. 그림 3과 같이 APM 알고리즘은 전처리 단계에서 탐색된 얼굴영역을 기준으로 APM을 위치시킨 뒤, 추출된 전경 영역에 의해 모델을 변형시키고, 변형된 모델을 해당 제스처의 특징으로 결정한다.

APM의 변형에 사용되는 영상은 카메라로부터의 거리 정보를 포함하는 시차 영상이며, 이전 단계에서 탐색된 얼굴 영역을 중심으로 전경 영역을 둘러싼 3차원의 외곽선을 구성하고, 각 APM 노드에서 주위 노드와의 무게중심 위치 정보를 이용하여 최종적으로 내부노드를 재배치한다. APM의 노드수가 많아

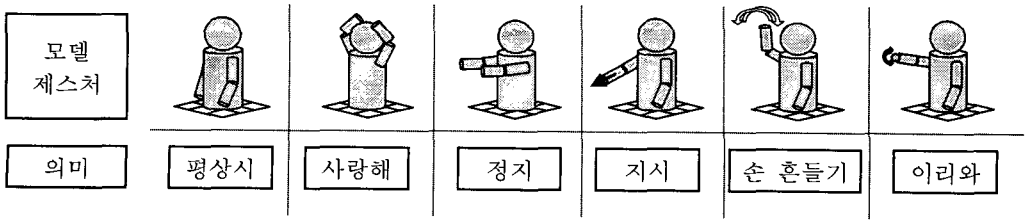


그림 2. 로봇 제어를 위한 기본 제스처의 분류

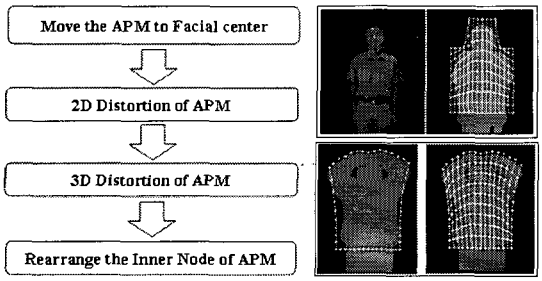


그림 3. APM 알고리즘

질수록 보다 자세한 제스처를 표현할 수 있지만, 계산 시간이 급격히 증가하거나 노이즈 발생 확률이 높아질 수 있다. 또한 노드수가 너무 작으면 제스처의 특징을 표현하지 못하기 때문에 본 논문에서는 다양한 실험을 통해 최적의 성능을 갖는 15X10 해상도의 APM을 사용한다.

APM은 그림 3과 같이 여러 개의 격자 형태의 노드로 구성되며, 각 노드는 3차원 공간에서의 위치로 표현된다. 임의의 시간 t에서의 특징벡터 F_t는 (2)와 같이 표현된다.

$$F_t = \{N_1, N_2, ..., N_n\} \tag{2}$$

$$N_i = \{x_i, y_i, z_i\} \tag{3}$$

$$(0 \le t \le T, \quad 1 \le i \le n, \quad n \ge 4)$$

F_t 는 임의의 시간 t 에서의 특징 벡터 집합, N_i는 i 번째 APM 노드, n은 APM의 총 노드 수이다. 그림 4는 정의된 제스처에 대한 APM의 변형 예를 보여준다.

또한, APM은 행위자의 제스처 특징을 추출하기 위하여 정형화된 모델을 사용한다는데 그 특징이 있다. 행위자의 의도가 포함되지 않은 제스처의 경우 특징벡터의 변화가 최소화 되어야 함에도 불구하고, 행위자의 외관 특성 및 노이즈의 영향으로 그렇지 못한 경우가 대부분이며, 이것은 인식률을 저하 시키는 큰 원인이 된다. 이와 같은 문제점을 해결하기 위하여 APM 알고리즘은 그림 4의 (a)에서와 같이 고정된 형태의 변형 최소 모델을 사용하여, 행위자의 외관 특성 차이 및 노이즈의 영향을 최소화 할 수 있는 효과를 얻을 수 있다. 또한, (4)와 같이 APM 노드 중 변형이 가장 많은 노드로의 방향 벡터를 이용하여 행위자의 개략적인 지시 방향을 쉽게 추정할 수 있는 장점이 있다.

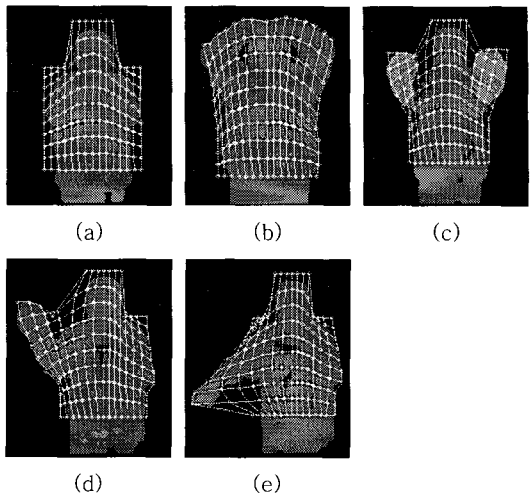


그림 4. 제스처의 APM 변형 예: (a) 평상시, (b) 사랑해, (c) 정지, (d) 손 흔들기, (e) 지시.

$$N_{peak} = \arg max_{N_i} \|N_i - N_o\| \tag{4}$$

$$v_{peak} = N_{peak} - N_f$$

(N_o: APM의 초기 위치, N_f : 얼굴 영역의 위치, v_{peak}: 지시 방향 벡터)

3.3 주성분 분석법을 이용한 포즈의 심볼화

일반적으로 제스처 데이터는 고차원의 특성을 가지며, 데이터의 특성상 직관적이지 못한 경우가 많다. 따라서 이를 효율적으로 분석하는 방법이 필요하며, 본 논문에서는 통계학적 접근 방법인 주성분 분석법 (principle component analysis: PCA)을 사용한다. 앞 절에서 언급한 바와 같이 APM를 통해서 추출된 특징 벡터는 PCA를 이용하여 파라메트릭 제스처 공간으로 구성되며[11], 이 공간에서의 고유벡터의 크기는 고유공간의 중요도를 의미하게 된다. 따라서 고유벡터의 크기에 따라 고유벡터를 선택함으로써 고차원의 제스처 데이터를 저차원의 벡터를 이용하여 표현할 수 있다. 그림 5는 제스처 공간에 투영된 모델 제스처의 분포를 나타낸다. 이렇게 구성된 제스처 공간은 새로운 입력을 비교하는 기준이 되며, 새로운 입력과 각 모델 동작과의 거리를 비교함으로써 포즈를 구별할 수 있다. 각각의 동작은 포즈 심볼로 표현될 수 있으며, 연속적인 제스처는 포즈 심볼의 집합으로 표현된다.

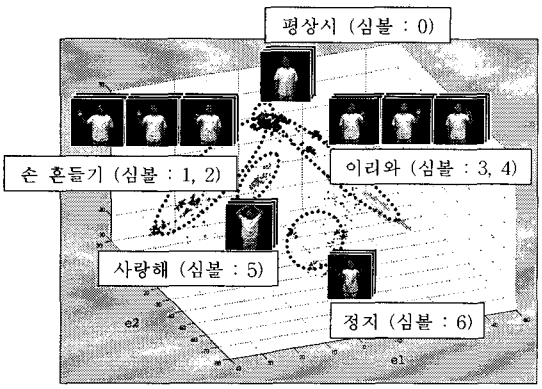


그림 5. 모델 제스처의 제스처 공간으로의 투영결과

3.4 포즈 심볼을 이용한 제스처 인식

본 논문에서는 제스처를 시공간상의 연속된 동작군으로 가정하고, 제스처의 시간적인 개념을 표현하기 위하여 은닉 마르코프 모델(HMM)을 이용한다. 은닉 마르코프 모델(HMM)은 은닉상태(hidden status)와 관측가능상태(observable status)로 이루어진 확률적 네트워크를 이용한 통계적인 인식 방법이며, 공간적인 개념과 시간적인 개념을 동시에 표현할 수 있다. 새로운 포즈 심볼 시퀀스 가 주어질 때, 각 HMM의 확률은 (5)와 같이 계산된다.

$$P(Y \mid \lambda_i) = \sum_i \sum_j \alpha_t(i) a_{ij} b_{ij}(y_{t+1}) \beta_{t+1}(j) \quad (5)$$

학습 모델 제스처로부터 얻어진 제스처 공간을 통하여 입력 제스처는 포즈 심볼의 집합으로 표현될 수 있으며, 그림 6과 같이 큐에 입력된 일정 시간 동안의 포즈 심볼 시퀀스에 대한 각 HMM의 확률을 계산하고, 최대 천이 확률을 갖는 제스처로 최종 결정된다. 단, 지시 제스처의 경우 동작의 형태가 고정적이지 않기 때문에, APM의 변형 정도를 이용하여 제스처를 인식한다.

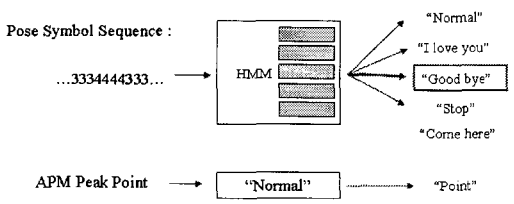


그림 6. 포즈 심볼 시퀀스를 이용한 제스처 인식

4. 실험 및 고찰

그림 7은 제스처 인식 과정을 도식화 한 것이다. 본 논문에서는 얼굴 영역 탐색을 위하여 Gaussian Mixture Model (GMM) 기반의 방법을 사용하며, 스테레오 정보 추출을 위하여 PointGrey사의 Bumblebee 스테레오 카메라를 이용하였다[12].

본 실험에서는 15×10 해상도의 APM을 사용하여 매 프레임 450차원의 특징 벡터가 추출되며, 주성분 분석법을 이용하여 5차원의 제스처 공간을 구성하였다. 또한, 다양한 실험을 통하여 결정된 8개의 상태를 가진 HMM을 사용하였으며, HMM의 파라미터 추정을 위하여, Forward-backward 알고리즘을 사용하였다. 실험을 위한 제스처 영상 데이터베이스는 학습영상과 실험영상으로 구분하여 촬영하였으며, 가

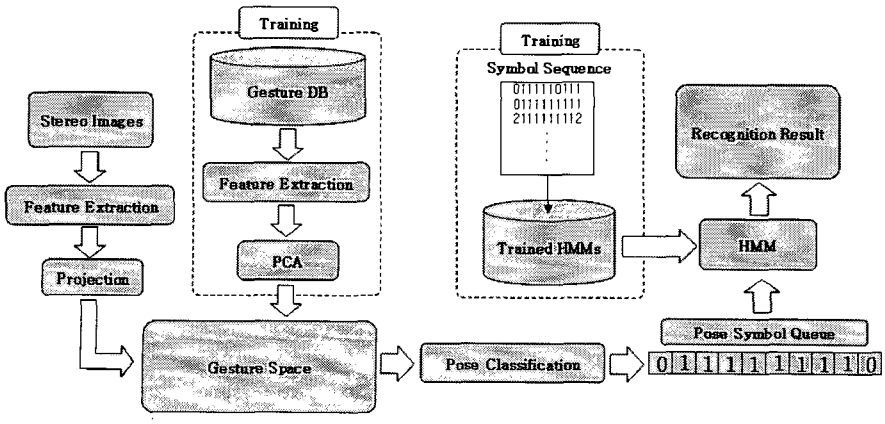


그림 7. 제스처 인식 시스템의 구성

로 320 픽셀과 세로 240 픽셀의 해상도를 갖는 컬러 영상으로 구성된다. 학습 영상은 제한된 배경에서 서로 다른 10명의 행위자의 제스처 영상을 사용하였으며, 실험영상은 일반적인 배경에서 촬영된 15명의 제스처를 사용하였다.

그림 8은 실험에 사용된 제스처 데이터베이스의 예를 보여주며, 표 1은 15명의 제스처 영상으로 구성된 입력 제스처에 대한 인식률(%)을 나타낸 것이다. 실패 란에는 오인식 된 동작과 비율을 표시하였다.

인식에 실패한 경우는 크게 세 가지 경우로 분류할 수 있다. 첫째, 그림 9의 (a)에서와 같이 얼굴영역과 손 영역이 겹쳐서 정확한 얼굴 영역을 찾지 못하는 경우이며, 둘째, 그림 9의 (b)에서와 같이 실내조명 조건에 의해 시차영상(disparity image)을 구하지 못한 경우이다. 이외에도 그림 9의 (c)와 같이 제스처의 움직임이 너무 작아서 제스처를 인식하지 못하는 경우도 발생하였다. 본 논문에서 제안하는 알고리즘은 행위자의 얼굴 영역 탐색과 시차영상 생성에 영향을 받으며, 정확한 데이터가 없는 경우 제스처 인식이 어렵다는 제약을 가진다.

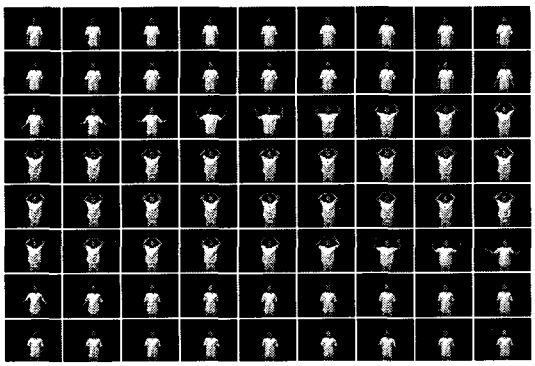


그림 8. 제스처 데이터베이스의 예 (사랑해)

표 1. 실험 영상에 대한 제스처 인식률

	성공률 (%)	오인식 결과 (실패율:%)
평상시	98	이리와 (2)
사랑해	95	평상시 (5)
손 흔들기	85	평상시 (10), 이리와 (5)
정지	80	이리와 (20)
이리와	85	평상시 (15)
지시	100	-
평균	90.5	-

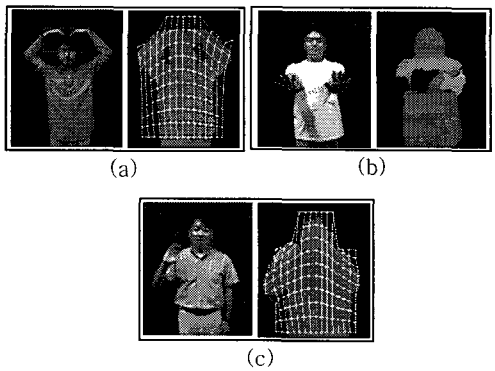


그림 9. 제스처 인식의 실패 예: (a) 얼굴 탐색 실패, (b) 시차 계산 실패, (c) 움직임 검출 실패.

그림 10은 바닥의 고정된 사각형 형태의 지점에 대한 지시 방향의 실험 결과를 행위자의 위쪽 방향에 대해서 표시한 그래프이다. 그림에서도 알 수 있듯이, APM을 통한 지시 방향은 행위자를 중심으로 약 10 cm 의 오차를 보였으며, 정확한 지시 위치보다는 행위자의 개략적인 지시 방향을 빠르게 계산할 수 있는 장점을 가진다.

또한, 그림 11은 본 논문에서 정의한 기본적인 제스처 인터페이스를 이용하여 가상의 로봇 캐릭터를 제어하는 예를 보여준다. 제스처 인터페이스에 대응하는 가상 로봇의 움직임을 미리 입력한 뒤, 인식 결과에 따라 반응하도록 하였으며, 펜티엄4 2.0 GHz 의 환경에서 초당 약 10 프레임의 처리 속도를 보였다.

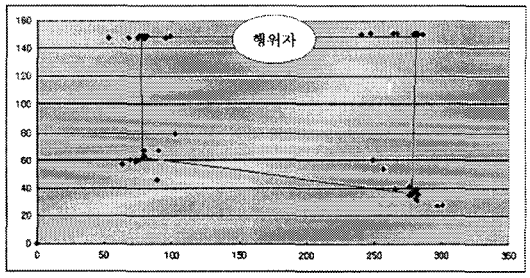


그림 10. 지시 제스처의 지시 방향

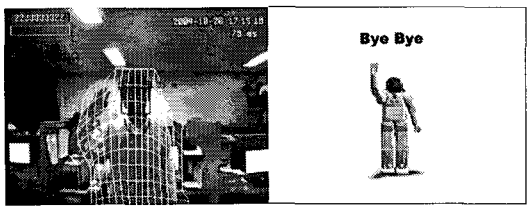


그림 11. 제스처를 이용한 가상 로봇 제어

5. 결 론

본 논문에서는 효율적인 제스처 특징 추출을 위한 방법으로 APM 모델을 제안하였다. 제스처 인식을 위한 기존의 방법들은 제스처의 3차원 특성을 정확하게 표현할 수 없었으며, 3차원 정보를 추가로 사용하는 경우 3차원 데이터의 에러와 시스템의 복잡성 때문에 그 활용에 제약이 따랐다. 본 논문에서 제안하는 APM은 기존의 방법들과는 달리 2차원 형상정보와 3차원 거리 정보를 동시에 표현할 수 있는 장점이 있으며, 행위자의 외관 특성과 영상의 노이즈에 영향을 덜 받는 특징이 있다. 또한, APM을 통해 행위자의 개략적인 지시방향을 쉽고 빠르게 계산할 수 있다. APM을 통해 추출된 특징 벡터는 주성분 분석법(PCA)과 은닉 마르코프 모델의 통계학적 방법을 이용하여 분석되며, 학습된 제스처를 효율적으로 인식할 수 있었다.

그러나, 소수의 학습된 모델 제스처를 이용하여 인간의 모든 제스처를 분류한다는 것은 매우 어려운 일이다. 본 논문에서 제안하는 방법은 실험결과 90% 이상의 인식 성공률을 보였지만, 학습된 제스처와 매우 상이한 제스처의 경우 인식이 어렵다는 단점이 있으며, 움직임이 작거나 주위 환경의 변화가 심한 경우 인식하지 못한 경우도 발생하였다. 따라서 많은 수의 모델 제스처 데이터베이스를 확보하고, 이를 체계적으로 분석한다면, 지능형 로봇뿐만 아니라 지적 인터페이스 시스템과 같은 여러 응용 시스템에 제스처 인식 기술을 적용할 수 있을 것이다.

참 고 문 헌

[1] James Davis, "Recognizing Movement using Motion Histograms," *MIT Media Lab. Technical Report* No. 487, 1999.

[2] Xia Liu and Kikuo Fujimura, "Hand Gesture Recognition using Depth Data," *Automatic Face and Gesture Recognition, Proceedings. Sixth IEEE International Conference*, pp. 529-534, 2004.

[3] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interpretation: A Review," *IEEE Transaction on PAMI*, Vol. 19, No. 7, pp. 677-695, 1997.

[4] Baihua Li, Holstein, H., and Qinggang Meng, "Articulated Point Pattern Matching in Optical Motion Capture Systems," *Control, Automation, Robotics and Vision, ICARCV 2002. 7th International Conference*, Vol. 1, pp. 298-303, 2002.

[5] Pengyu Hong, Turk, M., and Huang, T.S., "Gesture Modeling and Recognition using Finite State Machines," *Automatic Face and Gesture Recognition, Proceedings. Fourth IEEE International Conference*, pp. 410-415, 2000.

[6] Ismail Haritaoglu, David Harwood, and Larry S. Davis, "W4: Who? When? Where? What? A Real-time System for Detecting and Tracking People," *Automatic Face and Gesture Recognition, Proceedings. Third IEEE International Conference*, pp. 222-227, 1998.

[7] Davis, J.W. and Bobick, A.F., "The Representation and Recognition of Human Movement using Temporal Templates" *Computer Vision and Pattern Recognition, Proceedings. IEEE Computer Society Conference*, pp. 928-934, 1997.

[8] Amat, J., Casals, A., and Frigola, M., "Stereoscopic System for Human body Tracking," *Modelling People, Proceedings. IEEE International Workshop*, pp. 70-76, 1999.

[9] Wada, T., Xiaojun Wu, Tokai, S., and Matsuyama, T., "Homography Based Parallel Volume Intersection," *Computer Architectures for Machine Perception, Proceedings. Fifth IEEE International Workshop*, pp. 331-339, 2000.

[10] F. J. Huang and T. Chen, "Tracking of Multiple Faces for Human-Computer Interfaces and Virtual Environments," *International Conference on Multimedia and Expo.*, Vol. 3, pp. 1563-1566, 2000.

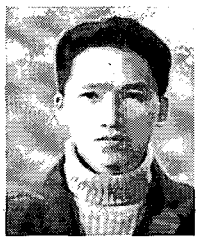
[11] Chil-Woo Lee, Hyun-Ju Lee, Sung H. Yoon,

and Jung H. Kim, "Gesture Recognition in Video Image with Combination of Partial and Global Information," *in Proceedings*. VCIP, pp. 458-466, 2003.

[12]. Point Grey Inc. (http://www.ptgrey.com).

오 재 용

2000년 전남대학교 컴퓨터공학과 졸업(학사)
2002년 전남대학교 대학원 컴퓨터공학과(공학 석사)
2002년~현재 전남대학교 대학원 컴퓨터정보통신공학과 박사과정
관심분야 : 컴퓨터 비젼, 제스처 인식, 컴퓨터 그래픽스



배 기 태

1997년 호원대학교 전자계산학과 졸업(학사)
1999년 전남대학교 대학원 컴퓨터공학과(공학 석사)
2001년~현재 전남대학교 대학원 컴퓨터정보통신공학과 (박사 수료)
관심분야 : 컴퓨터 비젼, 멀티미디어 데이터베이스, 패턴 인식, 컴퓨터 그래픽스, 영상 처리

김 만 진

1998년 호원대학교 전자계산학과 졸업(학사)
2000년 전남대학교 대학원 컴퓨터공학과(공학 석사)
2000년~현재 전남대학교 대학원 컴퓨터정보통신공학과 (박사 수료)
관심분야 : 컴퓨터 비젼, 머신 비젼, 패턴 인식, 컴퓨터 그래픽스

이 칠 우

1986년 중앙대학교 전자공학과 졸업(학사)
1988년 중앙대학교 대학원 전자공학과(공학석사)
1992년 동경대학교 대학원 전자공학과(공학박사)
1992년~1995년 이미지 정보과학 연구소 수석 연구원 겸 오사까 대학 기초공학부 협력연구원
1995년 리츠메이칸 대학 특별초빙강사
1996년~현재 전남대학교 전자컴퓨터정보통신 공학부 교수
관심분야 : 컴퓨터 비젼, 멀티미디어 데이터베이스, 컴퓨터그래픽스