Journal of Korea Multimedia Society Vol. 10, No. 8, August 2007(pp. 981-996)

주 인자 분석을 이용한 제스처 인식에 관한 연구

이용재[†], 이칠우[††]

요 약

본 논문에서는 연속적인 제스처 영상으로 부터 주 인자 분석을 통해 얻어진 동작 특징 정보를 이용하여 제스처를 인식하는 방법에 대해 기술한다. 제안된 방법은 먼저, 인간의 신체 영상이 포함된 연속적인 입력영상에서 2차원 실루엣 제스처 영역을 분할한 다음 전역특징정보와 지역특징정보를 추출한다. 여기서 전역특징정보는 요인 분석을 통하여 제스처를 효과적으로 표현하는 의미 있는 소수의 핵심 특징을 선택하여 이용한다. 추출 된 특징정보로 부터 제스처의 시간 변화를 나타내는 특징히스토리정보를 얻어 저 차원 제스처공간을 구성한다. 마지막으로 제스처 공간상에 투영된 모델 특징 값은 은닉마르코프 모델의 입력 기호로 이용되기 위해 군집화 알고리즘을 통해 특정한 상태 기호로 구성되며 임의의 입력 동작은 확률 값이 가장 높은 해당 제스처 모델로 인식된다. 주 인자 분석으로부터 제스처에 기여도가 높은 특징인자로 모델을 구성하기 때문에 외관기반방법에서 몸의 형상 정보만을 특징 값으로 이용하거나 직관적인 방법으로 특징을 추출하는 방법보다 복잡한 동작에서 비교적 우수한 인식률을 나타낸다.

A Study on Gesture Recognition Using Principal Factor Analysis

Yong-Jae Lee[†] and Chil-Woo Lee[††]

ABSTRACT

In this paper, we describe a method that can recognize gestures by obtaining motion features information with principal factor analysis from sequential gesture images. In the algorithm, firstly, a two dimensional silhouette region including human gesture is segmented and then geometric features are extracted from it. Here, global features information which is selected as some meaningful key feature effectively expressing gestures with principal factor analysis is used. Obtained motion history information representing time variation of gestures from extracted feature construct one gesture subspace. Finally, projected model feature value into the gesture space is transformed as specific state symbols by grouping algorithm to be use as input symbols of HMM and input gesture is recognized as one of the model gesture with high probability. Proposed method has achieved higher recognition rate than others using only shape information of human body as in an appearance-based method or extracting features intuitively from complicated gestures, because this algorithm constructs gesture models with feature factors that have high contribution rate using principal factor analysis.

Key words: Gesture Recognition(제스처인식), Principal Factor Analysis(주인자분석), Global Feature Information(전역특징정보), HMM

※ 교신저자(Corresponding Author) : 이칠우, 주소 : 광주시 북구 용봉동 300 전남대학교(500-757), 전화 : 062)530-1803, FAX : 062)530-1759, E-mail : leecw@chonnam.ac.kr
접수일 : 2007년 5월 2일, 완료일 : 2007년 6월 28일
[†] ㈜삼성테크윈 정밀기기연구소
(E-mail : yj0907.lee@samsung.com)

[††] 종신회원, 전남대학교 전자컴퓨터공학부
※ 본 연구는 정보통신부 및 정보 통신 연구진흥원의 IT신성장 동력 핵심기술개발사업의 일환으로 수행하였음. [2006-S-028-01, 상호 협력하는 분신형 네트워크 기반 휴머노이드 기술]

1. 서 론

최근에 들어 그래픽 유저 인터페이스를 비롯한 컴퓨터의 인터페이스 기술이 급속한 진전을 거듭함에 따라 컴퓨터를 이용한 다양한 인터페이스 응용시스템이 연구되고 있다. 이러한 분위기 속에, 보다 인간이 사용하기 쉬운 새로운 사용자 인터페이스 제작 기술로서 제스처 인식기술이 주목을 받고 있다. 그 이유는 인간은 일상생활에서 제스처, 표정과 같은 비언어적인 수단을 이용하여 수많은 정보를 전달하기 때문이다. 손의 움직임을 비롯하여 시선방향, 머리의 각도, 전신의 행동 등은 인간들 사이의 커뮤니케이션에 있어서 매우 주요한 역할을 한다. 따라서 자연스럽고 지적인 인터페이스를 구축하기 위해서는 제스처와 같은 비언어적 통신 수단에 대한 연구가 매우 중요하다. 또한, 대규모 비디오 데이터베이스의 구축, 감시 시스템, 고 압축 통신 시스템의 구축을 위해 제스처 인식에 관한 연구가 활발히 진행되고 있다.

일반적으로 제스처는 인간의 생각이나 감정을 표현하고 강조하기 위한 신체 또는 팔다리의 움직임이라고 정의되어진다. 이는 단순히 일상생활에서 의미를 나타낸 것으로, 카메라를 통하여 입력되는 2차원 영상에서의 제스처는 시공간 속에서 어떤 의미를 지닌 연속적인 패턴의 집합이라고 정의할 수 있다. 따라서 제스처를 인식한다는 것은 인체 각 부위가 시간축에 대해 어떠한 형상 변화를 가지는가를 자동으로 알아내는 것을 의미한다. 그러나 인체는 매우 복잡한 3차원 관절 구조를 지니고 있어서 자동으로 제스처를 인식한다는 것은 매우 어렵다.

동작을 인식하기 위한 제스처 인식방법은 인간의 신체 범위에 따라 손 인식, 상반신 행동인식, 전신 동작 인식으로 나뉘어 질수 있다. 첫째, 전통적으로 많이 연구되고 있는 손 동작인식으로 사용목적에 따라 수화인식, 지시형인식, 가상 물체를 조작할 수 있는 제어 형과 대화형 인식으로 나눌 수 있다. 상반신에서의 팔이나 손동작 움직임영역을 추적하거나 스킨 색상이나 형상모델을 이용하여 모양을 검출하는 방법을 많이 이용하고 있으며 최근에는 여러 방향에서 획득한 영상을 3차원 특징 모델로 구성하여 인식에 이용하는 방법이 있다[1].

둘째, 전신 동작 인식은 주로 걷기 동작과 손이나 발, 몸의 형상 전체가 바뀌는 제스처를 대상으로 한다. 걷기 동작은 직립 인간의 가장 기본적인 동작으로 많은 연구 활동이 이루어 졌으며 현재는 걷기의 속도나 방향을 고려한 걷기 제스처 데이터베이스가 구축이 되어 있어 많이 활용되고 있다[2,3]. 또한 사람마다 걷는 방식이 다르기 때문에 키네메틱 특징 벡터를 이용하여 개별적인 걸음걸이를 인식하는 방법도 연구 되고 있다[4].

전신의 동작을 인식하는 방법은 앞서 설명한 손동작을 포함한 걷기, 뛰기 등 테니스 동작이나 발레와 같은 다양한 동작이 인식 대상에 포함되어 있으며 다중 시각 정보를 이용하여 시점변화에도 안정적인 인식을 하는 방법이 연구 되고 있다[5]. 특히 동작 인식에 대한 객관적인 성능 평가가 어렵기 때문에 최근에는 일상적인 동작과 비정상적으로 일어날 수 있는 동작에 대한 데이터베이스를 구축하여 이용하고 있으며 2차원뿐만 아니라 3차원 데이터도 이용되고 있다[6].

상반신 영역의 인식방법은 주로 상체나 팔의 움직임에서 외곽선, 좌표 영역, 마스크를 적용하여 특징을 추출한다[7-9]. 컴퓨터와의 인터페이스 시 대부분의 동작이 상반신에서 이루어지기 때문에 가장 접근하기 용이한 인식영역이며 실시간에 적용이 쉬운 간단한 특징추출을 이용하고 있다. 하지만 외곽선이나 마스크를 적용한 추출 방법은 별도의 분할알고리즘이 적용하시 않을 경우 정확한 패턴이나 외곽선이 추출하기 어렵기 때문에 제한적인 동작에만 적용할 수 있다[10]. 피부 색 이나 손의 형상을 특징으로 사용하는 경우는 비교적 분할인식에 유리하나 형상의 겹침이나 동작의 빠르기가 다를 경우 오 인식 할 위험이 있다[11]. 상반신 인식에서 모델 구성방법 으로는 일반적으로 많이 쓰는 형판 매칭이나 벡터 기호의 거리비교, HMM 등이 주로 사용 되고 있다.

3차원을 이용한 경우 여러 시점에서 촬영한 손의 실루엣 영상을 합성시켜 볼륨모델 형판으로 매칭하는 방법이 있으며 수화동작을 스테레오 비전을 통해 깊이 정보를 추출하여 인식하는 방법이 있다. 전자의 경우 다양한 시점 변화에도 인식하는 장점이 있으나 한쪽 손의 단순한 동작에 제한적으로만 적용되고 있고 실시간 시스템에 적용을 위해서는 좀 더 빠른 계산 속도가 요구 된다. 후자는 수화라는 복잡하고 다양한 형상을 3차원으로 인식할 수 있으나 손 형상의 특징을 얻기 위해 칼라 장갑을 이용해야 하며 안정적

인 분할을 위해 특수제작 된 상의를 착용해야 하기 때문에 응용에 제한적 일 수 있다[12].

전신 영역에서 걷기 동작의 주요 추출 특징은 대부분 그림자영상으로부터 구하여 진다. 2차원 인식의 경우 반복되는 동작이며 형상도 복잡하지 않기 때문에 상하좌우로 투영시켜 얻어진 특징이나, 정면/측면 시점에서 구한 인체 모델, 신체 자유도에 대해 미리 정의된 키네메틱 골격구조로부터 기하학적 특징 변위(곡률정도)를 특징으로 사용하고 있다[13, 14]. 동작의 단순성에 비해 응용에 있어 다양한 환경에서 적용되어야 하며 거리, 방향, 개인적인 동작차이 등으로 인식률이 비교적 낮은 편으로 조사되었다[15]. 전신 영영의 동작 인식에서는 주로 실루엣 신체 영역의 위치, 형상, 모션 영역을 특징화 하였으며 복잡한 화면에 여러 명의 신체 형상을 분할인식하기 위해 윤곽선모델(머리와 어깨선, 몸통선, 두 다리 선의 각도)을 이용 하였다[16,17].

제스처인식의 특성상 적용 분야나 실험 환경에 따라 유사한 특징이나 모델구성 방법을 이용하여도 다른 결과를 얻을 수 있다는 것을 알 수 있었다. 특히 어떤 동작에서 다른 동작으로 전환 시 전후 동작의 경계 구분도 중요한 연구가 될 수 있다. 이 경우 기준 모델을 사람의 수작업을 통해 만들어 놓거나 수학적인 알고리즘으로 정한다음 두 결과를 비교하는 방법도 이용되고 있다. 하지만 대부분의 제스처 인식방법들은 특징을 구하기 위해 응용분야에 적합하거나 직관적으로 선택하게 된다. 특히 다양한 특징들을 조합하여 사용할 경우 중복되거나 적당한 특징을 선택하는데 어려움이 있다는 것을 알 수 있었다. 따라서 본 연구에서 인식해야할 동작에 대한 다양한 특징의 인자를 찾아 주요 인자에 대해 상관관계가 높은 특징에서 대표 특징을 선택하여 핵심인자를 특징으로 이용함으로써 주 인자에서 중복 될 수 있는 특징 수를 줄일 수 있을 뿐 만 아니라 의미 있는 특징을 선택할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 연속적인 제스처 영상에서 배경을 제외한 신체영역만을 추출하여 주 인자분석 이용하여 여러 가지 2차원 전역 특징으로부터 핵심특징을 선택하는 방법과 제스처 탐색 윈도우를 이용하여 특징을 추출하는 방법을 설명하고, 3장에서는 주성분 분석법을 이용하여 제스처 공간을 구성하고 K-평균 알고리즘으로 제스처를 재구성하여 제스처를 인식하는 방법을 설명하였다. 4장에서는 실험결과를 보이며 마지막 5장에서는 결론과 향후 연구내용에 대해 설명하였다.

2. 주 인자 특징 추출과 영상군집화

2.1 주 인자 분석을 이용한 제스처 특징 추출

제스처란 추상적인 의미를 지닌 인간의 몸동작, 손짓, 표정 등을 말한다. 따라서 제스처를 인식하기 위해서는 입력 영상으로부터 신체 지역(전경 영역)을 정확히 분리하는 것이 필요하다. 카메라를 통하여 얻은 영상 시퀀스는, 일반 실내 환경에서 취득한 것으로 영상에는 제스처 인식에 방해요소가 되는 많은 오브젝트들(배경)이 포함되어 있고 조명의 밝기가 일정하지 않고 수시로 변하기 때문에 같은 카메라로 일정 시간 동안 똑같은 배경을 촬영한다고 할지라도, 모두 동일하지 않아 안정적인 배경 모델을 얻는데 어려움이 따른다. 인식에 필요한 영역은 신체 영역(전경)이므로 우선 배경과 신체 영역을 분리하는 작업이 필요하고 이를 위해서는 먼저 배경 모델을 생성해야 한다. 배경 모델(Background Model : BM)은 전경 영역을 포함하지 않은 영상 시퀀스로부터 얻어지는 것으로, $M(x,t)$, $N(x,t)$, $D(x,t)$의 3가지 정보에 의해 계산되어질 수 있다. 여기서 $M(x,t)$는 화소 x가 시간 t에 의해서 갖는 최소 밝기 값, $N(x,t)$는 화소 x가 시간 t에 의해서 갖는 최대 밝기 값을 나타낸다. $D(x,t)$는 화소 x가 가질 수 있는 최대 밝기 차이 값을 나타낸다. 전경 영역은 식 (1)에 의해서 결정되어진다. 즉, 식 (2)를 만족하는 화소 x는 모두 전경 영역으로 분할되며 이는 조명의 변화로 생기는 밝기 차이는 무시하고 이보다 큰 밝기 차이 값을 갖는 영역을 전경 영역으로 분리하는 것을 뜻한다. 여기서 $I(x,t)$는 입력 영상이고 C는 상수 값이다.

$$|M(x,t)-I(x,t)|>D(x,t)+C \ OR \\ |N(x,t)-I(x,t)|>D(x,t)+C \quad (1)$$

본 논문에서는 조명 변화로 인한 배경의 밝기 변화를 측정하기 위해 시간 요소(t)를 고려해서 일정 시간 T_1 동안 배경 영상 I_t을 취득한 다음, 영상 전체 영역 R 내에 있는 각 픽셀(x)들의 밝기 값 I(x)들을 분석하여 조명이 가장 밝았을 때의 화소값 M(x)와 가장 어두울 때의 화소값 N(x)을 얻는다. 결국, 이

두 화소값의 차이 $D(x)$는 조명의 변화로 나타날 수 있는 밝기의 임계치로, 이 3가지 요소를 이용해 배경 모델(Background Model : BM)을 구성한다. 이와 같은 내용을 수식으로 표현하면 식(2)과 같다[18].

$$\begin{aligned} BM &= M(x), N(x), D(x)_{x \in R} \\ M(x) &= Max I_t(x), (1 \le t \le T_1) \\ N(x) &= Min I_t(x), (1 \le t \le T_1) \\ D(x) &= M(x) - N(x) \end{aligned} \tag{2}$$

일단 배경 모델이 만들어지면, 이진 영상 B(x)는 식 (3)에서 보여주는 것처럼 입력 영상 I(x)와 가장 밝은 화소값 M(x)와 가장 어두운 화소값 N(x)의 차분 연산을 통해 얻은 차이 값이 임계치 D(x) 보다 크면 255의 화소값을, 그 외에는 0의 화소값을 갖는다.

$$B(x) = \begin{cases} 255 & \text{if}(|M(x) - I(x)| > D(x) \\ & \text{or}\,|N(x) - I(x)| > D(x)) \\ 0 & otherwise \end{cases} \tag{3}$$

식(3)은 조명으로 인해 생길 수 있는 밝기 차이는 무시하고 신체의 움직임으로 차이를 갖는 영역만 분리하는 기준이 된다. 그러나 이 방법 역시 조명으로 인해 그림자가 생기는 경우, 그림자도 전경 영역으로 분리되어 정확한 신체 영역을 얻을 수 없다는 단점을 가지고 있다. 그림 1은 최대 휘도치 배경영상집합과 최소 휘도치 배경영상집합의 차분영상과 Exclusive OR에 대한 영상을 나타내었다. XOR영상을 보면 배경 간 영상변화가 화면전체에서 많이 일어날 수 있다는 것을 알 수 있다. 그림 1에서 (e)는 배경모델 파라메타를 통해 신체 영역만을 분할하여 얻어진 영상을 나타내었다. 하지만 이렇게 얻어진 영상에서도 완전하게 제거 되지 않는 잡음이 발생하였으며 발생된 잡음을 완전하게 제거하기 위해 침식연산을 수행하고 이 때 신체영역도 같이 줄어드는 현상을 막기 위해 팽창연산을 적용한다[19].

신체 특징은 신체 형상 전체가 갑작스럽게 변하는 전역 특징과 머리, 손, 발과 같이 주된 신체 부위의 일지역만 변하는 지역 특징으로 구분할 수 있다. 이와 같은 사실을 실험을 통해 분석하고 증명할 수는 없지만, 특징을 신체 포즈의 변화라고 가정했을 때 인간의 다양한 동작들을 관찰함으로써 쉽게 이해할 수 있다. 예를 들어, 사람이 앉는 동작을 취했을 때 이 동작은 서 있는 포즈에서 앉은 포즈로 변하는 것

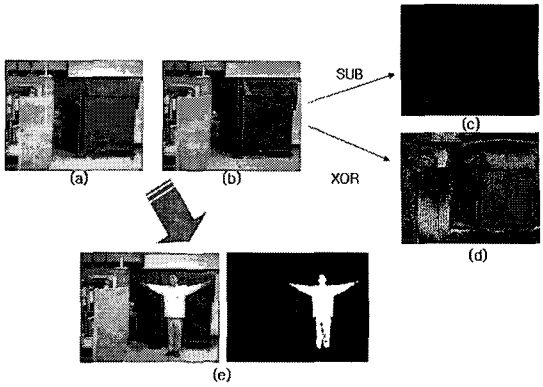


그림 1. (a)는 배경영상의 최대 밝기 값을 가지는 영상,(b)배경영상에서 최소 밝기 값을 가지는 영상,(c)와(d)는 최대 휘도치 영상과 최소 휘도치 영상과의 차분영상과 Exclusive OR 영상,(e)입력영상을 배경모델 파라메타를 이용해 추출한 실루엣 제스처 영상

으로, 서 있는 포즈는 수직으로 긴 직사각형 형태이며 앉은 포즈는 거의 정사각형의 형태이다. 즉, 움직임을 고려하지 않았을 때 앉는 동작은 두 팔의 직사각형의 형태에서 정사각형의 형태로의 포즈 변화로 해석 할 수 있다. 그러나 서서 손을 흔드는 동작과 같은 경우에는 서 있는 동작과 비교했을 때 손의 위치와 움직임이 매우 중요한 의미를 갖게 된다.

전역 특징과 지역 특징은 서로 영향을 미치기 때문에 따로 분리해서 생각한다는 것은 매우 어려운 일이다. 따라서 이 두 가지 특징 정보를 결합하여 특징 히스토리 정보를 얻고 이를 인식에 사용함으로써 보다 안정적인 인터페이스 구현이 가능하다.

시각적인 방법으로 얻은 영상으로부터 특징 값을 추출하여 신체의 포즈와 특징을 효과적으로 표현하고 인식하는 연구는 전역 특징 정보로부터 특징을 추출하는 방법과 머리, 손, 발과 같이 신체의 특정 부위에 의미가 있는 지역 특징 정보로부터 특징을 추출하는 방법으로 나눌 수 있다.

제안한 방법은 2차원 그림자 영상인 제스처 영상을 이용하여 특징을 얻기 때문에 2차원 패턴으로부터 얻을 수 있는 다양한 특징을 추출해야 한다. 특히 제스처에서 추출된 개별 특징들은 단독으로 사용 될 때는 여러 동작을 표현하는 데는 한계가 있으나 특징 결합 집합을 이용 할 경우 다양한 특징들이 상호 보완적인 역할을 하기 때문에 복잡한 동작을 인식하는 데 많이 이용 되고 있다. 하지만 너무 많은 특징을 이용 할 경우 특징 정보 간의 간섭으로 인해 애매모

호한 인식결과를 나타낼 수 있으며 계산 량이 많아지기 때문에 실시간 시스템 적용이 어려울 수 있다. 수많은 특징 중에 최적의 특징을 선택하기는 쉽지 않으며 일반적으로 동작에 따라 사람의 직관을 통해 적절한 특징을 선택해야 만 한다. 인식대상이 되는 동작 집합에서 적절한 특징을 선택하기 위해서는 각 특징이 동작집합에서 어느 정도 기여도를 가지고 있으며 특징 값 사이에서 상관관계 정도를 알아야 한다. 특징 값 간의 상관관계가 큰 특징 집합을 군집 화 하여 대표 특징을 선택 할 수 있다면 효과적으로 특징의 수를 줄일 수 있을 뿐 아니라 중복 되지 않은 특징을 이용 할 수 있다.

본 논문에서는 주 인자 분석을 이용하여 다양한 특징으로부터 최적의 특징 집합을 구성하여 이용하였다. 주 인자 분석이란 주어진 변수들을 가상적인 공통인자(변수)들의 일차결합으로 나타내고, 이들 인자들 가운데 중요한 몇 개의 인자만 선택하여 전체의 변동을 설명하고자 하는 것이 주요 목적이다. 이때에 각 인자에 대응되는 계수들의 크기를 이용하여 원래 변수들을 몇 개의 집단으로 나누고 각 집단의 특성에 따라 공통인자의 의미를 해석 할 수 있으며 이는 변수 축소(Variable Reduction), 중요 특징들을 선택하는 기준이 될 수 있다. 여기에서는 인자분석에서 가장 기본이 되는 분석법인 공분산행렬의 주성분법을 이용하였다[20].

여기서 p개의 성분을 가진 관찰 가능한 확률벡터 즉 추출된 전역특징 집합 $X=[X_1,X_2,...,X_p]$가 있고 평균벡터 $\mu = (\mu_1,\mu_2,..,\mu_p)'$와 공분산행렬을 Σ로 한다면 다중 인자 모형에서 X는 공통요인인 저변에 m ($\ll$p)개의 인자라 부르는 관찰 할 수 없는 확률변수 $F_1,F_2,...,F_m$의 선형결합과 그 변수에만 영향을 미치는 특수 인자인 $\varepsilon_1,\varepsilon_2\varepsilon_3$의 합으로 식(4)와 같이 표현 될 수 있다. 여기서 선형결합에 사용된 가중계수 λ_{ij}를 인자적재라고 부르는데, 이는 인자 모형에서 고려된 i번째 변수 X_i에 관한 j번째 인자 F_j의 중요성을 나타낸다.

$$X_{(p\times 1)}-\mu_{p\times 1}=A_{(p\times m)}F_{(m\times 1)}+\varepsilon_{p\times 1} \quad (4)$$

단, $\Lambda=(\lambda_{ij})$

공분산 구조의 인자모형을 위하여 사용하고 모든 I 에 대하여 특수인자변수가 0이고 관측 불가능한 확률벡터의 공분산이 $Cov(F) = I$ 이라면 다음과 같이 나타낼 수 있다.

$$\begin{aligned}\Sigma_{(p\times p)} &= A_{(p\times p)}A'_{(p\times p)}+0_{(p\times p)} \quad (5)\\ &= AA'\\ &= \lambda_1\beta_1\beta_1'+\lambda_2\beta_2\beta_2'+...+\lambda_p\beta_p\beta_p'\end{aligned}$$

따라서 적재행렬은 A는

$$A=[\sqrt{\lambda_1}\beta_1\ \sqrt{\lambda_2}\beta_2\cdots\sqrt{\lambda_p}\beta_p] \quad (6)$$

A의 모든 변수를 다 포함하며 특수인자벡터 분산은 전혀 인정되지 않기 때문에 실제적으로 유용하게 쓰이지 못한다. 특수요인벡터는 무시될 만큼 중요하지 않다고 가정하고 m개의 인자를 선택하여 이용하면 인자패턴은 다음과 같다.

$$\Sigma=AA'+\Psi \quad (7)$$

$$A=[\sqrt{\lambda_1}\beta_1\ \sqrt{\lambda_2}\beta_2\cdots\sqrt{\lambda_m}\beta_m] \quad (8)$$

$$\Psi=\begin{pmatrix}\Psi_1 & 0 & \bullet & \bullet & 0\\ 0 & \Psi_2 & \bullet & \bullet & 0\\ \bullet & \bullet & & & \\ 0 & 0 & \bullet & \bullet & \Psi_p\end{pmatrix} \quad (9)$$

$$Cov(x)=PDP^T \quad (10)$$

여기서 Cov(x)는 고유벡터 P와 고유치 D의 대각행렬로 분해 될 수 있다. 전체 전역특징 정보는 전체 23개와 전체 동작프레임 수 (784 = 98 X 8)로써 공분산행렬 Cov(x)는 23 x 23이 된다. 통계 분석도구로는 SAS(Statical Analysis System)을 사용하였으며 요인 추정방법은 주성분법을 이용하여 요인적재량을 추정하였다[20].

표 1에서와 같이 초기 특징의 값들은 일반적으로 2차원 패턴 인식에 이용되는 특징을 이용하여 결과를 나타낸 것이다. 전체 전역특징의 종류는 23가지이며 그중 19가지는 형상, 면적, 각도와 같은 제스처 정보를 이용하였으며 나머지 4가지는 신체 외곽의 좌표특징 값을 이용하였다. 전체 특징 정보에 대해F 시간에 따른 변화량도 특징으로 이용하기 때문에 현재 특징 값(T), 이전 프레임과 현재 프레임 특징 차분 값(T-1), 이전프레임 전 프레임과 현재 프레임 특징 차분 값(T-2)도 인자 분석의 입력 값으로 포함 하였다. 입력 동작은 본 연구에서 인식하고자 하는 8가지 동작(총 800프레임)으로 한 동작을 100프레임 씩 구

표 1. 2차원 실루엣 패턴에서 이용되는 특징 값

No.	Features	Definition
1	Width	폭
2	Height	높이
3	Gravity X	무게중심 X
4	Gravity Y	무게중심 Y
5	Compactness	원형도
6	Axis Principal Angle	주축
7	Axis_Secondary_Angle	주축
8	Perimeter	둘레
9	Roughness	거친정도
10	Area	면적
11	Feret_Max_Angle	최대길이의 각도
12	Feret_Min_Angle	최소길이의 각도
13	Elongation	가로 VS 세로비율
14	Number_of_Runs	수평String 수
15	Feret_Mean_Diameter	평균 직경
16	Feret_Min_Diameter	최소 직경
17	Feret_Max_Diameter	최대 직경
18	Intercept Vertical	수평방향 전경영역으로 변이하는 횟수
19	Intercept Horizenral	수직방향 전경영역으로 변이하는 횟수
20	Box_X_Min	신체영역의 최소 X좌표
21	Box_X_Max	신체영역의 최대 X좌표
22	Box_Y_Min	신체영역의 최소 Y좌표
23	Box_Y_Max	신체영역의 최대 Y좌표

성하여 이용하였다. 또한 시간별 특징의 인자 분석을 개별적으로 시행하였으며 표 2에서 그 결과를 나타내었다. 표 2에서처럼 주요 인자에 대해 특징들은 서로 높은 상관을 가지는 것으로 군집화 될 수 있으며 중복된 특징을 제외한 대표 특징으로 선택 되어 질수 있다. 표 3은 특징별 전체 인자들이 군집 화되어 최적의 특징을 선택한 결과를 나타낸 것이다. 여기서 여러 변수에 공통으로 작용하는 요인들을 찾아내고, 특징수를 축소하여 전체 23가지에서 7가지만 선택할 수 있었다. 이렇게 선택 된 특징들은 전역특징 정보로 이용되며 전체적인 신체 특징을 형상화하기 위해 7가지 전역 특징 정보; 1)신체 영역의 가로축 길이(width), 2)세로축 길이(height), 3)무게 중심의 x좌표, 4)무게 중심의 y좌표, 5)조밀성(compactness), 6)모멘트의 주축(Principal) 7)모멘트 주축(Secondary)을 추출한 후, 이들 특징 값의 시간적인 변화량을 계산한다. 이 7가지 특징 정보는 동작의 변화로 인해 신체의 형상 변화가 생겼을 때 의미 있게 변하는 특징 값들을 관찰하여 채택한 것들이다.

그림 2는 인식하고자 하는 제스처 특징 정보의 의미를 나타낸 것이다. 그림에서처럼 걷는 동작의 경우 의미 있게 변하는 특징 값은 무게 중심의 x 좌표이고 서있다 앉는 동작은 신체 영역의 세로축 길이와 무게 중심의 y 좌표가 많은 변화를 보였다. 그리고 손과 발을 벌리는 동작은 신체 영역의 가로축 길이와 조밀성(compactness), 몸을 한쪽으로 기울이는 동작은 모멘트의 주축 값이 의미 있게 변함을 확인할 수 있다.

이들 특징 값들의 변화량은 전체적인 외형의 변화만을 설명해 줄뿐, 신체의 어느 부위가 움직이고 있는지는 보여주지 않는다. 그러나 제스처는 때때로 신체 특정 부위의 특징의 변화에 따라 움직임의 많은 차이를 보일 수 있다. 우리는 정확한 제스처 인식을 위해 신체의 어느 부위가 움직이고 있는지 알 필요가 있고 이를 위해서는 부분적으로 움직임을 알 수 있는 지역 특징 정보가 필요하다. 이를 위해 손이나 발의 정확한 위치를 정확하게 계측하려면, 너무도 많은 계산 량이 필요하다. 따라서 우리는 매우 간단한 특징 데이터(신체의 지역 영역의 무게 중심 좌표, 지역 영역의 면적)를 도입하여 이 문제를 해결하고자 한다. 신체 영역은 전역 특징으로 구한 무게중심을 기준으

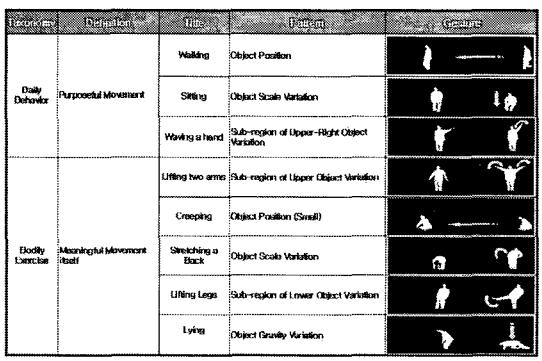


그림 2. 제스처 영상패턴에 따른 동작의 움직임 분석

표 2. 시간 T에 대한 특징 집합의 인자 패턴

Index	특 징	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
V13	Compactness	0.95718	0.00282	0.05131	0.07701	0.10758	0.07855
V5	Elongation	0.95689	0.00383	0.05213	0.07806	0.10810	0.07985
V8	Perimeter	0.87757	0.37174	0.12710	-0.04930	-0.11385	0.04575
V14	Number_of_Runs	0.78120	0.48873	-0.12117	0.05526	-0.02290	-0.04233
V19	Intercept_H	0.78120	0.48873	-0.12117	0.05526	-0.02290	-0.04233
V9	Roughness	0.76579	-0.25633	-0.41369	0.09422	-0.16518	-0.06066
V18	Intercept V	0.72757	0.00514	0.39463	0.02240	-0.20838	0.21910
V2	Height	0.20168	0.80378	-0.16780	-0.10063	0.37439	-0.09950
V17	Feret_Max_Diameter	-0.00714	0.65399	0.33667	0.24582	0.48317	0.17924
V10	Area	0.16047	0.64470	0.38822	0.09505	-0.41629	-0.02335
V4	Gravity Y	-0.10306	-0.81155	-0.03954	0.10857	0.11417	-0.01521
V22	Box_Y_Min	-0.09218	-0.90744	0.14785	-0.06969	-0.12717	0.09607
V1	Width	-0.05955	-0.13387	0.92056	0.08604	-0.00311	0.21795
V15	Feret_Mean_Diameter	0.03566	0.54872	0.73846	0.12637	0.13561	0.21581
V16	Feret_Min_Diameter	0.26661	0.00053	0.68134	-0.09565	-0.40198	0.08936
V20	Box_X_Min	0.13616	0.04713	-0.86085	-0.03353	-0.06563	0.38314
V12	Feret_Min_Angle	0.07349	-0.02171	0.03848	0.87814	0.10333	0.07874
V7	Axis_secondary_angle	0.07447	0.05876	0.04984	0.86973	-0.04959	-0.20708
V6	Axis Principal Angle	-0.22142	0.01725	0.02210	-0.43590	0.31256	0.43490
V23	Box_Y_Max	0.20239	0.28096	0.05451	-0.08275	0.70256	-0.16049
V11	Feret_Max_Angle	-0.12308	-0.05351	-0.08725	0.05905	0.65898	-0.04034
V21	Box_X_Max	0.13763	-0.22583	0.45344	0.10745	-0.22112	0.73335
V3	Gravity X	0.27684	0.08863	-0.15327	-0.42171	-0.19009	0.72479

표 3. 특징별 전체 인자들이 군집화 되어 최적의 특징을 선택한 결과

No	Features	T						T-1							T-2						
		F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6	F7	F1	F2	F3	F4	F5	F6	F7
V1	Width			3					2							2					
V2	Height		2									5									
V3	Gravity X						6						6							6	
V4	Gravity Y										5						3				
V5	Compactness	1						1							1						
V6	Axis Principal Angle						6														7
V7	Axis_Secondary_Angle				4									7							
V8	Perimeter	1						1							1						
V9	Roughness	1						1							1						
V10	Area		2					1													
V11	Feret_Max_Angle					5						5							5		
V12	Feret_Min_Angle				4					3								4			
V13	Elongation	1						1							1						
V14	Number_of_Runs	1						1							1						
V15	Feret_Mean_Diameter			3						3						2					
V16	Feret_Min_Diameter			3					2							2					
V17	Feret_Max_Diameter		2							3								4			
V18	Intercept V	1						1							1						
V19	Intercept_H	1						1							1						
V20	Box_X_Min																				
V21	Box_X_Max			3			6						6							6	7
V22	Box_Y_Min										4						3				
V23	Box_Y_Max					5						5							5		

로 4개의 지역 영역으로 나눌 수 있고 각각의 영역에 속하는 블랍들의 중심 좌표와 면적의 변화량을 살펴봄으로써 지역 특징 정보를 얻을 수 있다. 또한 지역 특징 정보에서 각 지역 영역의 무게중심 좌표를 통해 특징정보를 추출하고 있으나 손이나 발동작의 움직임을 표현위해서는 보다 정확한 움직임 영역을 검출하여한 한다. 인간 신체에서 팔이나 발이 움직일 경우 신체 중심인 몸통 부위에서 벗어나기 때문에 그때의 손, 발의 좌표 값을 특징 값으로 이용하는 것이 필요하다. 따라서 본 논문에서는 인간의 신체 변화 비율을 통해 정의함으로써 손이나 발 부위의 움직임을 비교적 정확하게 검출하고 추적할 수 있는 제스처 탐색 윈도우를 이용하는 방법을 제안한다. 그림 3의 (a)처럼 인간의 움직임 영역은 양팔을 좌우로 벌렸을 때 신장과 1:1의 비율을 가지며 일정한 영역 안에서 대부분의 동작이 이루어진다는 것을 알 수 있다. 동작성이 없는 자세의 기준을 서있는 자세로 정하고 그때의 팔, 다리의 움직임의 비율을 측정한 결과 개별적으로 움직임의 차이나 사람의 신체 크기를 고려한 모델을 그림 3 (b)에서 나타낸 것처럼 얻을 수 있었다. 이러한 신체 제스처 측정 결과를 기준으로 그림 4에서처럼 제스처 탐색 윈도우를 만들어 인간의 신체 부위 움직임의 끝점을 찾아냄으로써 동작성의 주요 특징을 찾아 지역 특징 정보로 이용 할 수 있다. 제스처 탐색 윈도우에서 손과 발, 머리 등의 움직임을 추출 하는 방법은 입력 영상에서 얻어진 실루엣 영상에서 중심좌표와 가장 밑 지역에서 추출된 좌표를 기준으로 임의로 정해진 신체 영역비로 각 윈도우 영역을 설정한다. 설정된 각 영역에서 탐색하여 얻어진 특징 좌표를 모든 프레임마다 적용하여 특징 좌표를 추적한다. 이렇게 얻어진 특징좌표는 지역 특징 정보로 이용 되어 진다. 만약 손이나 발, 머리(대상자가 엎드릴 때)의 영역이 발생 하지 않을 경우는 탐색 윈도우 값들을 몸 중심 값으로 이용하였다.

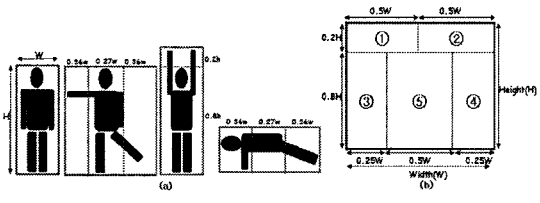


그림 3. 실험 대상자들의 신체 움직임 영역을 측정하여 얻어진 제스처 윈도우

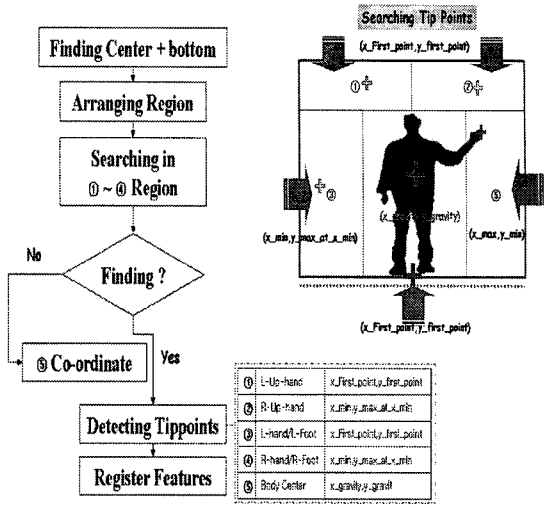


그림 4. 제스처 탐색 윈도우 실행 순서

이러한 특징 정보는 독립적으로 이용되기 보다는 상호 의존적으로 이용되어야 하며 전체 특징간의 종합적인 분석을 통해 제스처로 부터 특징을 추출 할 수 있었다. 그림 5는 원 영상으로부터 분할된 실루엣 제스처 영상과 전역과 지역특징정보(제스처 탐색 윈도우)를 추출한 결과를 나타낸 것이다.

2.2 영상군집화

행동이나 제스처는 연속적인 신체 부위 또는 전체적인 신체의 움직임으로 이루어지기 때문에 제스처 인식에서 고려해야 하는 중요한 사항은 특징 히스토리 정보(특징 값들의 시간적인 변화량)를 구하여 이를 인식에 이용하는 것이다. 앞 절에서 구한 특징 값들의 히스토리 정보를 구하기 위해 연속하는 3개의 영상을 하나의 군집으로 간주하고 이들 특징 값들의 시간적인 변화량을 계산하는 영상 군집화 방법을 사용한다.

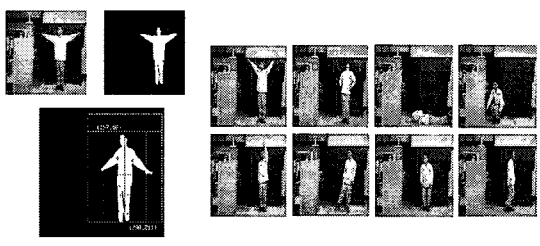


그림 5. 원 영상으로 부터 전역과 지역특징 정보를 추출 과정과 결과 영상

$$F(t) = F_g(t) + F_p(t) \quad (1 \le t \le T) \tag{11}$$

$$F'(y) = F(t-2), F(t-1) - F(t), F(t) \quad (3 \le t \le T,\ 1 \le y \le G) \tag{12}$$

식(11)에서 F(t)는 T를 영상 시퀀스의 총 길이라고 했을 때 시간 t에서의 특징 집합을 나타내는 것으로, 신체 전체의 포즈 정보를 담고 있는 전역 특징 벡터 Fg(t)와 신체의 특정 부위의 특징 특징을 나타내는 지역 특징 벡터 Fp(t)의 합으로 표현할 수 있다. 따라서 F(t)는 29개의 특징 값(7개의 전역 특징 + 22개의 지역 특징)으로 구성 되고 시간 t에서의 영상 It는 이웃하는 영상 It-1, It-2와 함께 하나의 군집으로 묶인다. 따라서 식 (12)에서 보여주는 것처럼 G(=T-2)를 전체 영상 시퀀스에서 얻을 수 있는 군집의 수라고 할 때, y번째 영상 군집의 특징 벡터 F'(y)는 식(11)를 이용해 구한특징 벡터 F(t)뿐만 아니라 F(t-1), F(t-2)와의 차분을 통해 얻은 특징 값들의 변화양의 합으로 구성 된다. 결국, 한 군집의 전체 특징 정보는 87개가 되고 이와 같이 시간적 영상 군집화 알고리즘은 그림 6에서와 같이 나타낼 수 있다. 하지만 각 특징들은 서로 영역, 각도, 좌표 등 단위 체계가 서로 다르기 때문에 이 값들을 그대로 사용하면 특징변화의 중요도와 관계없이 단위가 큰 특징 값들의 영향이 크게 작용 될 수 있기 때문에 모든 특징 값들은 양자화를 통해 동일한 단위 기준을 적용하였다. 본 연구에서는 1~100으로 전역/지역 특징 정보들을 양자화 하였다.

3. 제스처 특징 기호 생성과 제스처 인식

3.1 주성분 분석법을 이용한 기호생성

주성분 분석법(Principal Component Analysis)은 고차원의 입력 데이터 집합을 저 차원의 의미 있는

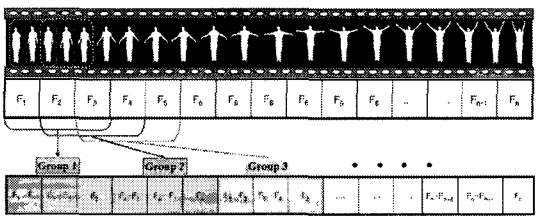


그림 6. 2진 영상에서 추출된 특징으로 부터 시간에 따른 특징 정보를 구성함.

데이터 집합으로 줄일 수 있다. 제스처 영상 데이터의 경우 하나의 동작을 구성하는 프레임(frame)의 수가 많고 특징을 추출하기 비교적 어렵기 때문에 빠른 인식속도와 효과적인 특징 추출이 가능한 방법을 적용해야 한다. 따라서 앞에서 추출한 연속적인 동작의 선형적 특징을 이용하여 저 차원 벡터로 표현하는 방법에 대해 기술 한다.

앞 절에서 구한 특징 벡터 x는 식 (13)과 같이 표현될 수 있고 이 벡터의 고유공간을 계산하기 위해서는 먼저 모든 특징 벡터의 평균 벡터를 구하여 각 특징 벡터와의 차를 구한다. 평균 벡터 c와 새로운 특징 집합 X는 식 (14)과 식 (15)와 같다. 그런 다음, 식 (16)을 만족하는 고유벡터를 구하기 위해 공분산 행렬 Q에 대한 고유치 λ와 고유벡터 e를 구한다.

$$x = [x_1, x_2, \cdots, x_N]^T \tag{13}$$

$$c = (1/N)\sum_{i=1}^{N} x_i \tag{14}$$

$$X \triangleq [x_1 - c, x_2 - c, \ldots\ldots, x_N - c]^T \tag{15}$$

$$Q \triangleq X \cdot X^T \tag{16}$$

$$\lambda_i \cdot e_i = Q \cdot e_i \tag{17}$$

이 때, 고유치 분해를 하지 않고 특이치 분해(Singular Value Decomposition)를 이용함으로써 특징 집합 X의 공분산 행렬에 대한 고유벡터를 쉽게 얻을 수 있다. 이렇게 얻어진 고유공간에 평균 벡터 c에서 뺀 특징 집합 X를 모두 식 (18)을 이용하여 투영시킨다.

$$m_i = [e_1, e_2, \ldots\ldots, e_k]^T (x_i - c) \tag{18}$$

이와 같이 얻어진 저 차원 벡터 공간, 즉 파라메트릭 고유공간을 제스처 공간이라 부른다[21]. 이미 설명한 바와 같이 주성분 분석은 몇 개의 주성분 벡터를 유도하여 이를 통해 차원의 축소와 자료의 요약을 주목적으로 하고 있다. 따라서 전체 변이의 대지역을 적절히 설명하기 위하여 보유해야 할 주성분의 수를 결정해야 한다. λi는 i번째 고유값, p는 전체 고유 값 개수라고 할 때 전체 분산 중 주성분 Ci가 설명할 수 있는 비율은 λi/p이다. 우리는 처음 k개 주성분들이 설명할 수 있는 누적비율이 70% 이상일 때의 개수를 선택한다. 이를 수식으로 표현하면 식 (19)와 같다.

$$(\sum_{i=1}^{k}\lambda_i / p)\times 100 \geq 70 \tag{19}$$

그림 7은 실험에 사용한 제스처들의 특징 집합으로부터 구한 고유치 개수에 따른 주성분의 누적 기여도를 보여주고 있다. 그림 8은 동작별 제스처 공간에 투영된 결과를 나타내었다.

3.2 K-Means 클러스터링을 이용한 특징 기호 군집화

그림 8에서 보듯이 개별 동작으로 투영된 제스처의 경우 완전한 선형적인 형태로 표현되지만 한 공간상에 다른 여러 동작을 투영한 결과 동작간의 간섭이 존재한다는 것을 알 수 있었다. 다른 동작 간에도 유사한 포즈로 인해 특징 값 비교 시 애매한 결과가 나올 수 있기 때문에 투영 된 전체 포즈를 K-means 클러스터링 알고리즘을 통해 비슷한 포즈끼리 군집화 하였다. 데이터를 비교하거나 통합 하는데 가장

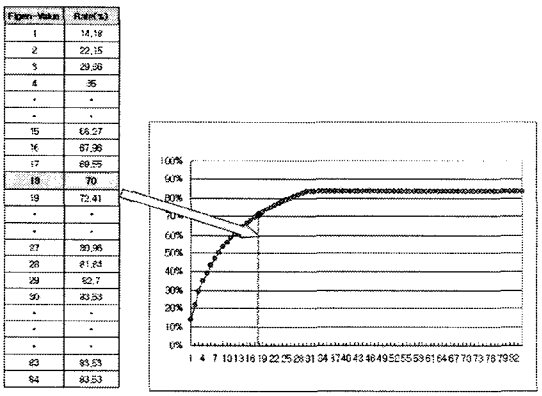


그림 7. 고유치의 갯수에 따른 누적 기여도

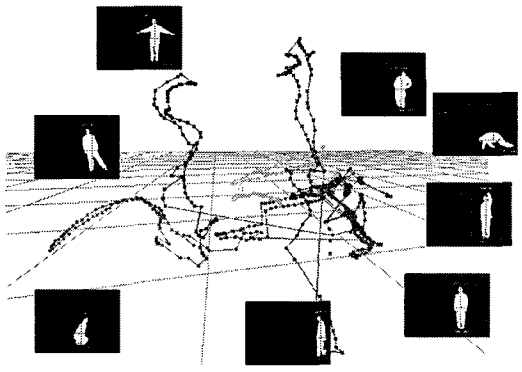


그림 8. 제스처 공간상에 투영 된 동작집합

우선적으로 적용되어야 하는 것은 데이터의 핵심적인 특성을 잃지 않으면서 데이터의 집합을 줄이거나 대표가 되는 데이터를 선정하거나 계산하는 것이다. 계산적으로 효율성이 있고, 원래의 데이터를 대표할 수 있는 데이터 집합을 찾아가는 방법이 필요하다. 클러스터링은 중첩되는 군집이나 집합, 점이 없도록 데이터를 나누는 것이다. 제스처 데이터 집합이 군집화 될 때, 모든 벡터는 어떤 하나의 대표집합에 속해야 한다. 모든 군집들은 하나의 데이터로 표현되는데, 이러한 대표적인 데이터를 대표중심 값이라 하며, 보통 집합 내의 데이터의 평균으로 구해진다. K-means 클러스터링 알고리즘은 크게 두 단계로 나누어 볼 수 있다. 첫째는, 각 군집들과 초기 중심 값과 수가 정해지면 모든 데이터를 가장 가까운 초기 중심 값으로 할당하여 군집화 한다. 두 번째는 각 제스처 간의 군집들에서 새로운 중심 값을 계산하여 다시 모든 데이터와의 거리를 계산하여 가까운 중심 값과 가까운 데이터끼리 서로 군집화 한다. 그러나 K-means 클러스터링 알고리즘은 유일한 최적해가 존재하지 않으므로 여러 번 반복실험을 통해 가장 작은 전체 평균 왜곡을 가진 코드 북을 선택한다. 본 연구에서는 초기 군집의 수를 30개로 정하였으며 반복횟수는 9회에서 수렴 한다는 것을 알 수 있었다[20].

그림 9는 K-mean 클러스터링 알고리즘을 이용하여 유사 동작별로 기호들을 재구성 한 제스처집합을 나타낸다. 투영된 제스처 공간상에서 모델로 이용되는 전체 제스처 특징기호를 포즈별로 클러스터링 하여 30개의 인덱스 값들로 군집화 하였다. 각 모델 동작의 인덱스 값을 분석한 결과를 그림 10에 나타내었듯이 한 가지 동작에도 몇 가지의 포즈가 연속적인 선형상태로 존재 한다는 것을 알 수 있으며 시간에 따라 변화하는 기호들은 다음 절에서 HMM의 입력 기호로 이용 된다.

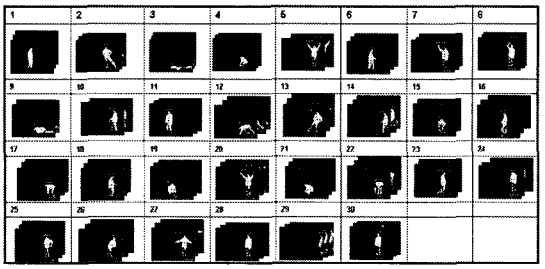


그림 9. 전체 기호를 30개의 군집으로 나눈 결과

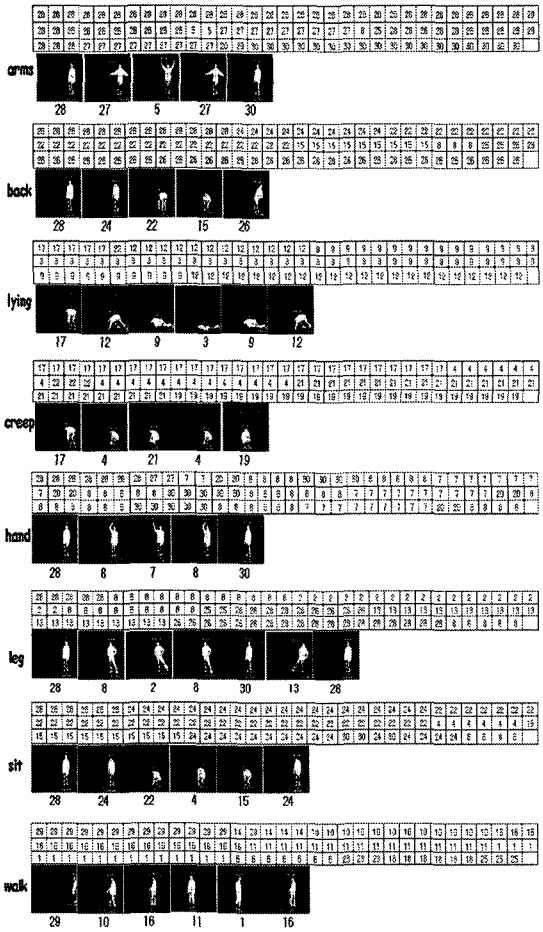


그림 10. 전체 모델 영상들의 시간에 따라 변화하는 기호들의 집합

3.3 HMM을 이용한 제스처 인식

파라메트릭 제스처 공간에 투영된 점들은 클러스터링 알고리듬에 의해서 몇 개의 제스처 패턴으로 분류되어질 수 있다. 이렇게 분류된 제스처 패턴들에 대해 특정한 심볼을 부여함으로써, 제스처 시퀀스는 심볼 시퀀스로 형상화되어지고 이를 은닉 마르코프 모델의 입력으로 사용한다. 우리가 앞 절에서 구한 저 차원의 특징 데이터 값들을 기호로 바꾸기 위해서 클러스터링(클러스터링) 알고리즘을 이용해 몇 개의 제스처 군집(Cluster)으로 나누고, 각 군집에 대해 특정 기호(숫자)을 할당한다. 그리고 각 클러스터의 중심 좌표 값은 코드 북으로 저장되어 새로운 특징 값이 들어왔을 때 기호를 할당하는 기준이 된다. HMM에서 상태 천이 확률 a_{ij}는 상태가 i로부터 j로 변화하는 확률을 의미한다. 그리고 확률 $b_{ij}(y)$는 출력 심볼 y가 상태 i로부터 j로 천이되면서 관측될 수 있는 확률, π_i는 초기 상태 확률 값을 나타낸다. HMM의 학습은 {π, A, B}의 파라미터들을 추정하기위해 봄-웰치(Baum-Welch) 알고리즘을 이용하였다[22].

입력의 기호 시퀀스(Y)가 주어지면 모델 λ_i에 대한 확률 값은 forward 변수인 $\alpha_t(i)$와 backward 변수인 $\beta_t(i)$를 이용하여 식(20)과 같이 구하고 가장 높은 확률 값을 갖는 모델로 인식하게 된다.

$$P(Y|\lambda_i) = \sum_i \sum_j \alpha_t(i) a_{ij} b_{ij}(y_{t+1}) \beta_{t+1}(j) \qquad (20)$$

4. 실험 결과

사람의 정면에 위치한 비디오카메라를 통하여 입력되는 256 계조의 흑백 영상을 실험 영상으로 사용하였다. 인위적인 스튜디오가 아닌 일반 사무실 환경에서 비교적 낮은 해상도의 고정된 1대의 흑백카메라를 이용하였으며 실험대상 들은 별도의 장신구(모자, 가방 등)는 착용하지 않았으며 신체 전체영역 촬영을 위해 카메라와 약 2m의 간격을 두고 동작의 빠르기나 개별동작간이 동작유형이 유사하도록 인위적인 사항은 배제하여 촬영하였다. 실험동작은 일상생활에서 주로 동작하는 걷기, 의자에 앉기, 손 흔들기 같은 일상행동과 엎드려 팔굽혀펴기, 등배운동, 쪼그려 걷기, 다리 펴기, 양팔 들어올리기 같은 움직임 자체의 의미를 갖는 의도된 동작 즉 체조동작들로 구성하였기 때문에 가혹조건을 부여하였다. 일상행위와 체조동작들은 의미적으로는 분류가 서로 다르나 신체 움직임의 변화부위는 비슷하기 때문에 구분하기 어렵다. 따라서 본 연구에서는 이러한 유사한 움직임의 각 각의 포즈를 기호화한 다음 기호들의 시간적인 변화를 모델화하였기 때문에 유사한 포즈가 많은 동작 간에서도 분류가 용이하다는 것을 알 수 있었다. 그림11은 실험에 사용 된 원 영상 집합을 나타낸 것이다. 인식대상의 위치, 크기, 부분 영역 좌표 이동 등에 따라 제스처 패턴 변화는 화면에서 다양한 형태로 나타날 수 있기 때문에 다각적인 특징정보를 추출 하여 이용할 필요가 있다. 걷기 동작이나 쪼그려 앉아서 이동하는 동작은 패턴의 시간에 따라 좌표의 이동이 나타나며 앉거나 등배운동의 경우는

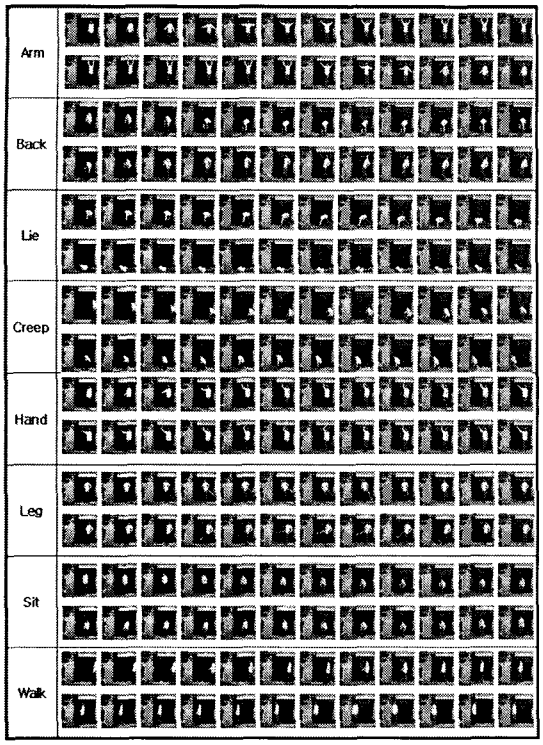


그림 11. 실험에 사용된 제스처 집합의 원 영상 집합(a)팔운동 (b)다리 펴기,(c)앉기 (d)걷기,(e)한팔 흔들기 (f)다리운동(g)의자에 앉기,(h)걷기

패턴의 크기의 변화가 주요 요소로 작용 한다. 제자리에서 팔이나 다리를 움직이는 동작과 엎드리는 동작에서는 부분 패턴의 변화와 손, 발, 머리의 위치가 주요 핵심 특징으로 나타난다는 것을 알 수 있었다.

입력영상의 크기는 320×240이고 프레임 입력 속도는 초당 20프레임을 기준으로 각각의 제스처는 100프레임으로 구성되어있다. 따라서 한 사람 당 총 800프레임(8×100)의 영상을 이용하였고, 실험에 참가한 사람은 4명으로 총 3200 프레임을 획득하여 이중 800프레임을 이용하여 모델 제스처로 구성하였다. 입력실험 영상은 실험에 참가한 4명이 10회 반복적으로 동작한 영상을 다른 시간대 별로 나누어서 사용하였다.

표 4는 전체 특징 정보 중에 전역 특징의 수를 달리하여 인식속도를 계산하여 비교한 것으로 입력영상의 동작별 입력 단위는 100Frames 이며 제스처를 동영상 파일(avi)로 저장 후 파일을 읽는 방식으로 실험한 것이다. 처리 시스템의 환경은 일반 노트북 PC(Core(TM)2 CPU 2.00GHz, RAM 1GB)를 이용하였으며 (a)에서는 전처리부에서 제스처 영역을 분할하고 전역과 지역특징정보를 추출 할 때의 처리시간을 나타내었고 (b)는 HMM을 통해 최종 인식결과까지의 처리속도를 계산한 것이다. 따라서 한 프레임 당 평균 처리 속는 전역 특징 수 전체 23개를 이용 하였을 때 보다 7개를 이용한 방법이 약간 빠른 것으로 나타났으며 프레임 당 최대 0.02초로 두 방법 모두 실시간 시스템에 적용이 가능하다는 것을 알 수 있다.

표 5에서는 특징별 동작의 인식성능을 비교하여 나타내었다. (a),(b) 경우 본 연구의 동작과 유사한 제스처를 이용한 타 연구논문과 인식성능을 비교하였으며 (c)~(h)는 실험에 이용된 8가지의 동작에 대

표 4. 전역특징 수에 따른 인식속도 비교.

단위 : Sec

Gesture(100Frames)	(a) Segmentation Extracting Features		(b)Recognition		Total(sec/frame)	
	Global(23)	Global(7)	Global(23)	Global(7)	Global(23)	Global(7)
Raising Arms	1.72	1.47	0.26	0.25	0.0198	0.0172
Bending a back	1.80	1.33	0.27	0.22	0.0207	0.0155
Waving a hand	1.85	1.53	0.28	0.23	0.0213	0.0176
Raising legs	1.56	1.16	0.26	0.26	0.0182	0.0142
Lying on the ground	1.31	1.06	0.22	0.19	0.0154	0.0125
Sitting on the chair	1.52	1.38	0.25	0.25	0.0176	0.0162
Creeping	1.34	1.20	0.21	0.22	0.0156	0.0142
Walking	1.69	1.38	0.23	0.23	0.0192	0.0161
Average	1.60	1.31	0.25	0.23	0.0185	0.0154

*()는 사용된 특징의 수

표 5. 인식에 사용된 특징에 따른 인식률 비교 단위 : %

Method / Gesture	(a) 3D Structural feature points [1]	(b) Contour signature [7]	(c) Appearance	(d) Global (7)	(e) Global(23) +Local	(f) Global(7) +Local	(g) Global(7)+ Local(GSW)	(h) Failure Rate
Raising Arms	96.7	-	75 (leg,walk)	99	82 (leg)	98	98	2
Bending	100.0	96.0	71 (sit)	49	76 (sit)	70 (sit)	75	25 (sit)
Waving a hand	100.0	96.0	62 (arm)	52	73 (leg)	77 (arm,leg)	81	19 (arm,leg)
Raising legs	-	-	98 (arm,hand)	99	79 (arm)	96	98	2
Lying on the floor	92.0	-	99	62	99	95	100	0
Sitting on the chair	89.2	-	81 (back)	41	93	79 (bend)	78	22 (bend)
Creeping	-	-	80 (walk)	49	96	99	98	2
Walking	91.6	-	83 (arm)	98	100	100	100	0
Average	94.9	96.0	81	69	87	88	91	9

*[ ]참고인용논문, ()는 사용된 특징의 수, ( )인식오류 동작

해 (c)외관특징 정보 (d)전역특징정보 (e)전체 23개의 전역과 지역특징정보 (f)주 인자 분석을 통해 얻어진 7개의 전역특징 정보와 지역특징정보 (g)는 (f)와 같은 전역과 지역특징에 제스처 탐색 윈도우를 적용한 방법을 이용하여 인식 결과를 나타낸 것이다. (a)3차원 특징정보를 이용한 방법은 다리 동작과 엎드린 동작을 제외한 동작에서 평균 인식성능이 94.9%로 동작의 수에 비해 가장 높은 인식 성능을 보였으며 (e)전역 특징 정보만을 이용한 방법은 69%로 가장 낮은 인식률을 보인다. 단순히 전역 특징정보 만을 이용한 방법 보다 외관 기반 방법이 인식 결과가 우수하였으나 전역과 지역 특징 정보를 결합한 방법이 전체적으로 우수하다는 것을 알 수 있었다.

(c)의 외관기반 제스처인식 방법은 특별한 특징을 추출하지 않고 사람의 실루엣 형상자체를 모델로 구성하기 때문에 전처리 과정이 간단하며 주성분분석법을 통해 저 차원 벡터로 영상을 표현할 수 있어서 영상 압축 효과를 통해 빠른 인식속도가 가능하였다. 또한 모델과 입력 영상간의 투영된 벡터를 비교하여 동작에 대한 구체적인 정보 또한 인식에 이용 할 수 있는 장점이 있다. 하지만 전체적인 형상에 대한 정보는 신체 움직임의 주요 요소가 되는 손과 다리와 같은 움직임을 구별하는 데는 어려움이 있었으며 모델 동작의 수가 많거나 유사한 형상의 동작일 경우는 많은 간섭으로 인해 애매한 결과를 얻는 다는 것을 알 수 있었다. 특히 다리운동이 오류율이 가장 높게 나타난 것은 팔 운동과 움직임의 변화율이 서로 유사한 값을 가지고 있기 때문이다. 이는 모델을 생성할 때 평균영상과의 거리를 기준으로 제스처 공간을 생성하여 영상간의 변이(영상 차분)값이 동작을 구분하는 주요 요인으로 작용함을 나타낸다. 따라서 신체를 움직이는 부위가 비슷한 유사 동작뿐만 아니라 움직임의 위치와 상관없이 외관상의 동작의 크기나 움직임의 범위가 비슷한 동작에서는 전혀 다른 신체 부위를 움직였더라도 해당 동작으로 오 인식 할 수 있는 문제가 있다. 이는 단순히 움직임의 변화 즉 영상의 밝기 변화의 상관관계 정보를 매칭에 이용하기 때문에 움직임의 위치에 변화가 있는 경우에도 움직임의 크기가 유사사한 동작일 경우는 잘못된 인식이 이루어진다는 것을 알 수 있었다.

직관적으로 23개의 전체 전역 특징정보를 이용한 방법 보다 주 인자 분석을 통해 7개의 특징 방법의 인식률이 약간 높은 성능을 보였으며 이는 전체 특징을 사용한 방법이 중복된 특징이 포함한다는 것을 알 수 있다. 결과적으로 (f)방법을 보안할 수 있는 특징을 이용한 (g)방법이 가장 우수한 인식률을 나타내는 것을 확인 할 수 있었다. 하지만 (h)에서 보듯이 굽히는 자세나 앉는 동작처럼 팔, 다리의 움직임 없고 몸의 움직임만 있는 동작의 경우는 각 각 오류율이 25%와 22%로 높게 나타났으며 두 동작은 서로 인식 확률 값이 타 동작에 비해 매우 높다는 것을 알 수 있었다. 이는 동작 특성상 신체 형상의 변화가 적고 손, 발의 좌표를 얻을 수 없고 전체적인 특징 변화가 적기 때문에 모델과의 매칭 시 애매한 결과를 나타낸 것이다.

지금까지의 실험결과에서 보였듯이 제시한 인식 방법은 복잡한 동작에서도 비교적 우수한 인식률을 나타내었으며 외관기반 방법보다 유사 동작 간 간섭이 적다는 결과를 나타내었다. 손 흔들기 동작에서와 같이 피 실험자의 개별 동작이나 빠르기의 차이가 있어도 인식률에 크게 영향을 미치지 않는다는 것을 알 수 있었다. 이는 기하학적 방법과 같은 신체의 외곽선이나 에지 정보를 정확하게 얻을 필요가 없고 패턴 자체의 특징 값과 국부적인 움직임 변화 정도를 특징으로 이용하기 때문에 잡음의 영향에 민감하지 않으면서 세부적인 움직임 부위도 인식이 가능 할 수 있다. 특징 기호의 시간 변화에 따른 모델을 구성하여 일련의 기호 변화상태가 보기 때문에 움직임의 길이나 빠르기가 다를 경우에도 빠르기별 모델을 구성할 필요가 없다. 따라서 신체 움직임 변화에 대해 복합적인 시공간적인 특징정보로 걷기, 팔 운동, 다리 운동, 눕는 동작 등 다양한 신체 움직임을 인식할 수 있는 것을 알 수 있다.

기존의 대부분 제스처 인식방법에서 모델로 구성할 특징 선택이 개발자의 직관적인 방법이나 경험이나 실험을 통해서만 이루어 졌지만 본 연구에서는 주 인자 분석을 이용하여 다양한 특징으로부터 주요 특징 인자들을 계산하여 인자 특징들을 군집화하고 중복특징을 모델 구성 시 배제함으로써 소수 핵심 특징을 추출 하여 전역 특징으로 이용하였다. 실험에서는 총 23개의 특징 중 단지 7개의 특징만 선택 추출하여 이용하였기 때문에 계산 량도 줄일 수 있었다. 하지만 이런 방법을 제스처 외관정보나 기하학적 특징에 적용할 경우 계산해야 될 특징벡터의 수가 너무 많고 단순 밝기 차 변화에 대한 변화 값들의 대응을 찾기 어려워 의미함축이 효과가 적고 적용하는데 어려움이 있다.

5. 결 론

논문에서는 하나의 카메라를 통해 입력되는 동작들을 전역과 지역특징정보를 이용하여 인식하는 방법에 대해 기술하였다. 카메라를 통해 입력된 영상에서 동작 요소만을 정확하게 분리하기위해 시간에 따른 여러 프레임의 배경 영상들을 모델링 하는 방법을 이용하여 휘도치가 변위 값이 일정 임계값보다 크거나 작게 변하는 영역을 제스처 영역으로 분할하였다.

외관 기반의 특징을 통한 방법을 이용하여 동작의 빠르기와 크기 등의 동작 정보에 대한 인식도 가능하였지만 평균영상과의 거리 값을 특징으로 모델링 하였기 때문에 움직임 영역이 유사한 제스처를 구별하기 어려운 한계가 있으며 어느 부위를 움직였는지 알 수 있는 특징정보가 부족하기 때문에 동작의 중요한 요인이 되는 손이나 발에 대한 움직임을 포함하는 동작의 경우 인식률이 매우 낮았다.

신체 포즈에 대해 다양한 특징을 얻기 위해 신체의 어느 부위가 움직이는지를 나타내는 지역(local) 특징 정보와 전체적인 신체의 형상을 표현하는 전역(global) 특징 을 이용하여 모션 히스토리 정보를 구성하였다. 또한 이러한 다양한 특징정보는 요인 분석을 통하여 제스처를 효과적으로 표현하는 의미있는 소수의 핵심 특징을 선택하여 이용하여 데이터 마이닝 효과뿐만 아니라 계산 량도 줄일 수 있었다.

혼합특징 정보로 영상내의 제스처의 공간적 특징 요소를 추출 한 후 동작의 시간별 변화 요소를 추출하기 위해 프레임간의 변화정도를 추출 하여 특징에 이용하였다. 매칭 속도 향상과 특징 값을 함축적인 기호로 구성하기 위해 주성분 분석법을 이용하여 저차원 특징 기호를 생성하였으며 서로 다른 동작 간에도 유사한 포즈가 존재하기 때문에 전체 동작을 포즈별로 재구성하기 위해 K-Means 클러스터링 알고리즘을 이용하였다. 각 동작은 재구성된 기호의 흐름으로 표현 되었고 포즈 즉 상태 변화를 모델로 구성하는 방법 중 하나인 은닉 마르코프 모델을 이용하여

모든 동작에 대한 확률 모델을 생성하여 인식에 이용하였다.

이 방법의 특징은 복잡한 계산을 이용하거나 기하학적인 특징인 에지나 코너를 얻기 위해 별도의 알고리즘을 사용하지 않고 인간의 신체 동작 모델을 이용하여 제스처 탐색 윈도우를 통해 손이나 발, 머리의 움직임을 비교적 정확하게 추출 할 뿐 아니라, 영상에서 쉽게 계산이 가능한 특징 값들을 이용하여 구한 모션 히스토리 정보를 인식과정에 사용한다는 점이다. 따라서 구체적인 동작을 인식하면서 많은 계산량이 요구되지 않기 때문에 실세계에서 구현이 용이하고 실시간 시스템 구축에 적합하다.

참 고 문 헌

[1] 노명철, 장혜민, 강승연, 이성환, "휴먼-로봇 상호작용을 위한 비전 기반 3차원 손 제스처 인식," *제 33회 한국정보과학회 추계 학술발표회 발표 논문집*, 서울, pp. 421-425, Oct. 2006.

[2] Ekinci, M., "Gait Recognition Using Multiple Projections," Automatic Face and Gesture Recognition, 2006. *FGR 2006. 7th International Conference*, pp. 517-522, April 10-12, 2006.

[3] L. Wang, T. Tan, and H. Ning, "Silhouette Analysis-Based Gait Recognition for Human Identification," *IEEE Trans. on PAMI*, Vol. 25, No. 12, Dec. 2003.

[4] Chan-Su Lee and A. Elgammal, "Gait Tracking and Recognition Using Person-Dependent Dynamic Shape Model," *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference*, pp. 553-559, April 10-12, 2006.

[5] Neil Robertson and Ian Reid, "Behaviour understanding in Video: a combined method," *ICCV'05*, Vol1., pp. 808-815, Oct. 2005.

[6] Bon-Woo Hwang, Sung-Min Kim and Seoung-Whan Lee, "A Full-Body Gesture Database for Human Gesture Analysis," *International Journal of Pattern Recognition and Artificial Intelligence*, August 4, 2006.

[7] Peixoto, P. Goncalves, and J. Araujo, H., "Real-time gesture recognition system based on contour signatures," *Pattern Recognition, 2002. Proceedings. 16th International Conference*, Vol. 1, pp. 447-450, 2002.

[8] G. Awad, T. Coogan, J. Hann, and A. Sutherland, "Real-Time Hand Gesture Segmentation Tracking and Recognition," *9th European Conference on Computer Vision*, Graz Austria, May 7-13, 2006.

[9] Kirishima, T. Sato, and K. Chihara, "Real-time gesture recognition by learning and selective control of visual interest points," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, Vol. 27, No. 3, pp. 351-364, March 2005.

[10] Ishihara. T and Otsu. N, "Gesture recognition using auto-regressive coefficients of higher-order local auto-correlation features," *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference*, pp. 583- 588, May 17-19, 2004.

[11] Kazuhiko Takahashi, Tatsumi Sakaguchi, and Jun Ohya, "Real-Time Estimation of Human Body Postures Using Kalman Filter," *ROMAN'99 8th International Workshop on Robot and Human Interation*, September 27-29, 1999.

[12] Qi Wang, Xilin Chen. Liangguo Zhang, Chunli Wang, and Wen Gao, "Viewpoint Invariant Sign Language Recognition," *International Conference on Image Processing(ICIP2005), Genova*, pp. 281-284, September 11-14, 2005.

[13] M. Dimitrijevic, V. Lepetit, and P. Fua, "Human Body Pose Recognition Using Spatio-Temporal Templates," *ICCV'05 workshop on Modeling People and Human Interaction, Beijing, China*, October 2005.

[14] Das, S.R. Wilson, R.C. Lazarewicz, M.T. Finkel, L.H., "Gait Recognition by Two-Stage Principal Component Analysis," *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference*, pp. 579-

584, April 10-12, 2006.

[15] Jian Li, Shaohua Kevin Zhou, and Rama Chellappa, "Appearance Modeling under Geometric Context," *International Conference on Computer Vision (ICCV), Workshop on Dynamical Vision, Beijing, China*, October 2005.

[16] Furukawa, M. Kanbara, Y. Minato, and T. Ishiguro, "Human behavior interpretation system based on view and motion-based aspect models," *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference*, Vol. 3, pp. 4160-4165 September 14-19, 2003.

[17] Cristian Sminchisescu, Atul Kanaujia, Zhiguo Li, and Dimitris Metaxas, "Conditional Models for Contextual Human Motion Recognition," *International Conference on Computer Vision (ICCV), Workshop on Dynamical Vision, Beijing, China*, October 2005.

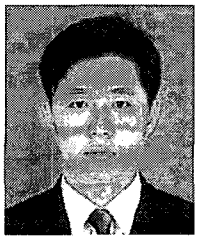
[18] Ismail Haritaoglu, David Harwood, and Larry S. Davis, "W4: Who? When? Where? What? A Real-time System for Detecting and Tracking People," *Third Face and Gesture Recognition Conference*, pp. 222-227, 1998.

[19] Rafael C.Gonzalez and Richard E.Woods, *Digital Image Processing 2/E*, Prentice Hall, pp. 519-532, 2001.

[20] 송문섭, 조신섭, SAS를 이용한 통계자료 분석, 자유아카데미, 2002.

[21] Shigeyoshi Hiratsuka, Kohtaro Ohba, Hikaru Inooka, Shinya Kajikawa, and Kazuo Tanie, "Stable Gesture Verification in Eigen Space," *LAPR Workshop on Machine Vision Application*, Vol. 0000141262, pp. 119-122, 1998.

[22] Caelli. T and McCane. B, "Components analysis of hidden Markov models in computer vision," *Image Analysis and Processing, 2003. Proceedings. 12th International Conference*, pp. 510-515, September 17-19, 2003.



이 용 재

1998년 호원대학교 전자계산학과 학사
2000년 전남대학교 컴퓨터공학과 석사
2007년 전남대학교 컴퓨터정보통신공학과 박사졸업예정
2004년 1월~9월 ㈜어플라이드비전테크 연구원
2004년 10월~현재 ㈜삼성테크윈 정밀기기연구소 선임연구원
관심분야 : 컴퓨터비전, 제스처인식, 반도체 검사장치

이 칠 우

1986년 중앙대학교 전자공학과 학사
1988년 중앙대학교 대학원 전자공학과 공학 석사
1992년 동경대학 대학원 전자공학과 공학 박사
1992년~1995년 이미지 정보과학연구소 수석 연구원 겸 오사까대학 기초공학부 협력연구원
1995년 리츠메이칸대학 특별초빙강사
1996년~현재 전남대학교 전자컴퓨터공학부 교수
관심분야 : 컴퓨터비전, 지능형 휴먼인터페이스, 멀티미디어 데이터베이스, 컴퓨터그래픽스