

연속영상에서 학습효과를 이용한 하이브리드 제스처 인식

강용관*, 나인호**, 김동욱***

* 전주대학교 정보통신공학과

**군산대학교 전자정보공학부

***전주대학교 전자매체공학부

목 차

- | | |
|-----------------|---------------------|
| I. 서 론 | V. 영상 시퀀스의 자동 분할 |
| II. 제스처 인식의 개요 | VI. HMM을 이용한 제스처 인식 |
| III. 제스처 인식의 분류 | VII. 실험 및 결론 |
| IV. 최근의 연구동향 | |

I. 서 론

인간은 일상 생활에서 제스처, 표정과 같은 비언어적인 수단을 이용하여 수많은 정보를 교환한다. 따라서 자연스럽고 지적인 인터페이스를 구축하기 위해서는 제스처와 같은 비언어적 통신 수단에 대한 연구가 매우 중요하다. 최근에 들어 대규모 비디오 데이터베이스의 구축, 감시 시스템, 고 압축 통신 시스템의 구축을 위해 제스처 인식에 관한 연구가 활발히 진행되고 있다.

제스처를 인식한다는 것은 인체 각 부위가 시간 축에 대해 어떠한 형상 변화를 가지는가를 자동으로 알아내는 것을 의미한다. 그러나 인체는 매우 복잡한 3차원 관절 구조를 지니고 있어서 자동으로 제스처를 인식한다는 것은 어렵다.

초기에는 인체 각 부위의 관절에 부착된 센서를 통해서 형상 변위값을 입력하여 시공간적인 형상 패턴을 추출하고 제스처를 인식하였다. 이 방법은 장치를 몸에 붙이는 과정이 복잡하고 초기 교정이 어려울 뿐만 아니라 연결 케이블 때문에 자연스러운 제스처 입력이 불가능하여 현재는 거의 사용되고 있지 않다.

최근에 들어, 광학적 마커를 몸에 부착하고 카메라로 입력된 영상으로부터 마커들의 궤적을 추적하여 제스처를 인식하는 방법들이 개발되었다. 그러나 사용되는 장비가 고가이고 실시간으로 인식하기가 어려워 현실적으로는 사용이 어렵게 된다. 또 다른 방법으로 카메라를 통해 인간의 연속적 동작 정보를 분석하여 인식하는 시각 기반 제스처 인식 방법을 들 수 있다. 이는 사용자의 기구에 대한 부담이 없고 동작의 구분에는 좋은 인식을 보인다. 이 접근법은 실시간 비전 프로세싱이 가능한 고속 계산기와 고속 비전 알고리즘이 실용화된 덕택으로 구현이 가능하게 되었다.

시각 기반 제스처 인식은 3차원 모델 이용법과 2차원 모델 이용법으로 구분할 수 있다. 3차원 모델 이용법은 인체 관절의 고 자유도(Degree of freedom)로 모델링을 하기가 어렵고 매칭에 어려움이 있다. 그리고 2차원 모델 이용법은 카메라의 위치에 따라 보여지는 외관(appearance)이 다르기 때문에 수많은 모델이 필요하고 이는 많은 계산량을 요구하게 된다.

본 논문에서는 연속적인 영상 시퀀스에서 시간에 따른 신체의 형상 변화량을 복수 다차원 특

징 값으로 추출하고 주성분 분석법(PCA:Principal Component Analysis)이라는 통계적인 수법을 이용하여 신체의 외관 특징들을 표현할 수 있는 저차원 벡터 공간, 즉 파라메트릭 고유 공간에 투영시킨 후 은닉 마르코프 모델을 이용하여 제스처를 인식하는 방법을 기술한다. 제안하는 알고리즘의 특징은 배경이 고정된 영상 시퀀스 속에서, 인간의 제스처가 바뀌는 프레임들 자동으로 검출하여 이를 기준으로 모델을 생성, 인식할 수 있다는 데에 있다.

2장에서는 제스처란 무엇이며 제스처 인식은 어떠한 과정으로 이루어지는지 간략히 설명한다. 3장에서는 제스처 정보를 획득하는 과정에 따라 제스처 인식 방법을 분류해 보고, 4장에서는 제스처 인식에 대한 최근의 연구 성과를 중심으로 요소 기술의 중요 내용을 살펴본다. 5장에서는 여러 종류의 제스처가 섞여진 영상 시퀀스로부터 시간축의 변화량을 특징 값으로 추출하는 방법과 행동 패턴을 자동으로 분류할 수 있는 방법을 소개한다. 6장에서는 은닉 마르코프 모델을 이용하여 학습, 인식하는 방법을 설명한다. 마지막으로 7장에서는 실험 내용과 결과를 살펴보고 앞으로의 연구 진행방향에 대해 고찰하고자 한다.

II. 제스처 인식의 개요

2.1 제스처의 정의

일반적으로 제스처는 인간의 생각이나 감정을 표현하고 강조하기 위한 신체 또는 팔다리의 움직임이라고 정의되어진다. 이는 단순히 일상 생활에서의 의미를 나타낸 것으로, 카메라를 통하여 입력되는 2차원 영상에서의 제스처는 시공간 속에서 어떤 의미를 지닌 연속적인 패턴의 집합이라고 정의할 수 있다.

2.2 제스처의 인식 과정

제스처 인식은 모델링(modeling), 해석(analysis), 인식(recognition)의 3단계로 이루어진다. 제

스처 모델링은 인식할 제스처의 수학적 모델을 만드는 것이다. 모델링을 하는데 사용된 접근법들은 제스처 인식의 중추적 역할을 하고 제스처 해석의 성능에 영향을 준다. 모델이 만들어지면, 비디오 입력으로부터 파라미터를 추출하고 계산을 통해 제스처를 인식한다.

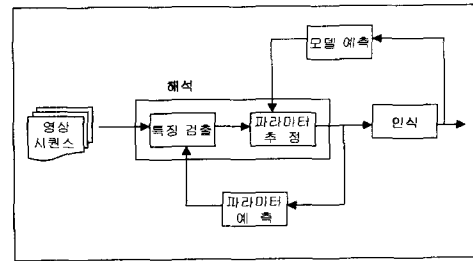


그림 1. 제스처의 해석과 인식

III. 제스처 인식의 분류

3.1 센서 부착식 제스처 인식

모션캡처에서 주로 사용되는 방식으로 인식보다는 계측, 측정의 의미가 강하다.

(1) 기계식 시스템

이 시스템은 사람의 관절 움직임을 측정하기 위해 전위차계(potentiometer)와 슬라이더(slidebar)의 복합체로 구성되어 위치 변화에 따른 변위 값을 물리량으로 직접 추출한다[18].

(2) 자기식 시스템

자기식 시스템은 연기자의 각 관절 부위에 자기장을 계측할 수 있는 센서를 부착하고 자기장 발생장치 근처에서 사람이 움직일 때 각 센서에서 측정되는 자기장의 변화를 다시 공간적인 변화량으로 계산하여 움직임을 측정하는 방식이다 [18].

(3) 광학식 방식

광학식 시스템은 사람의 주요 관절부분에 적

외선에 반응하는 적외선 마커(혹은 광 다이오드)를 부착하고 여기에 적외선 빛을 비추어 반사되는 영상을 3대에서 32대 가량의 CCD 카메라로 촬영하고 각 카메라에서 마커들의 2차원 좌표를 생성한다[18]. 각 독립된 카메라에서 캡처된 2차원 데이터는 소프트웨어로 분석되어 3차원 공간상의 좌표를 계산한다.



그림 2. 광학식 시스템

(4) 음향식 시스템

이 시스템은 다수의 초음파 발생장치와 3개의 수신장치로 구성된다. 사람의 각 관절에 부착된 초음파 발생장치들은 순차적으로 초음파를 발생하고, 그 초음파가 수신장치에 수신되기까지 걸린 시간과, 그때의 소리의 속도를 이용해서 발생장치에서 수신장치까지의 거리를 계산한다. 각 전송장치의 3차원 공간상의 위치는 3개의 수신장치에서 각각 계산된 값을 이용한 삼각 측량원리에 의하여 구할 수 있다[18].

3.2 영상(시각) 기반 제스처 인식

이는 비디오 카메라와 컴퓨터 비전 기술을 이용하여 제스처의 정보를 획득하는 방식으로 인체의 동작이 갖는 추상적인 의미를 이해하려는 데에 목적이 있다.

(1) 3차원 모델 이용법

카메라를 통하여 얻은 영상으로 가상의 3차원 모델을 모델링한 후 이를 시뮬레이션(simulation)

하는 기법이다. 정교한 표현이 가능하나 많은 계산 양을 요구하므로 실시간 인식에는 부적합하다.

(2) 2차원 모델 이용법

2차원 모델을 이용하는 것은 가장 일반적이고 여러 응용 분야에서 이용되는 방식이다. 이는 다음과 같이 크게 3가지 방식으로 나누어 볼 수 있다.

- 외관 이용법
- 특징 파라미터 이용법
- 가변 템플릿 이용법

외관 이용법은 카메라를 통해 입력되는 2차원 영상에서 음영 정보를 이용하여 제스처를 인식하는 방식이다. [1]에서는 카메라를 사람의 측면과 위쪽에 설치하여 각각의 카메라로부터 얻은 영상을 이용하였다. 최근에는 MHI(Motion History Image)를 이용해 제스처의 모형을 만든다[4]. 이와는 달리 특징 파라미터 이용법은 신체의 에지(edge), 윤곽선, 영상의 모멘트와 같은 파라미터를 추출하여 모델링하는 방식이다. 그리고 가변 템플릿은 입력 영상이 모델 영상에서의 제스처와 조금 다른 형태를 가지고 있더라도 같은 제스처로 인식할 수 있는 방법으로 특징 파라미터를 이용하는 방법과 외관을 이용하는 방법의 중간 형태로 두 방법의 개념을 함께 사용한 것이다.

IV. 최근의 연구동향

4.1 3차원 모델 이용법

(1) 3차원 모델

3차원 제스처의 모델은 크게 두 가지로 구분할 수 있다.

- 체적(볼륨) 모델(volumetric model)
- 골격 모델(skeletal model)

체적 측정의 모델은 주로 컴퓨터 애니메이션에서 사용되는 것으로 컴퓨터 비전 분야에서는 analysis-by-synthesis 기법에 사용된다[15]. 이 기법은 신체의 3차원 모델을 합성함으로써 신체의

자세를 해석하고 모델과 실제의 신체가 같은 영상으로 보일 때까지 변화시키는 것이다. 이는 사실감을 줄 수는 있지만, 실시간으로 나타내기에는 너무 복잡하다. 그래서 원통 (cylinder), 표면의 2차 곡면(super-quadric)과 같이 단순한 3D의 기하학적인 구조를 사용한다[5]. 원통의 모델의 경우 높이, 반지름, 색깔의 3가지 파라미터를 가지고 표현할 수 있다. 주로 간단한 신체 일부를 표현할 때 사용하는 것으로 복잡한 신체를 나타낼 때는 간단한 신체 일부를 연결하여 사용한다.

체적 측정의 모델은 2가지 문제점을 가지고 있다. 첫째는, 파라미터 공간의 차원이 높다는 것이다. 손 하나를 나타내는 데에도 23×3 이상의 파라미터가 필요하다고 한다. 둘째는, 컴퓨터 비전 기술을 통해 파라미터를 얻는 것이 꽤 복잡하다는데 있다.

체적 측정의 모델을 사용하는 것 대신, 관절 각도의 파라미터를 사용하기도 한다. 이는 골격의 모델로 알려져 있다. 뼈를 연결하는 관절은 자유로움의 정도가 다르다. 따라서 자유로움의 정도(DoF)를 가지고 해석한다[5][15]. 골격의 모델 역시 너무 복잡하고 많은 계산 양을 필요로 하는 단점을 안고 있다.

(2) 제스처의 해석 및 인식

제스처의 3차원 정보를 쉽게 얻기 위해서 MLD (Moving Light Display)와 같은 도구를 많이 이용하고 있다. 이는 신체의 관절 부위에 마커를 부착하고 빛을 비추으로써 마커의 위치 정보를 쉽게 얻을 수 있는 장점을 가지고 있다. 어깨나 엉덩이의 관절 상태(DoF)는 고도(ϕ), 근육의 외전 운동(ψ), 비틀림(θ)의 3가지 Euler 각도를 이용하여 나타낼 수 있다[5]. 팔꿈치나 무릎은 신전(extension)의 파라미터 ρ 를 이용한다. 이렇게 얻은 관절의 정보를 바탕으로 제스처를 구별하고 새로 입력 영상이 들어오면 DTW (Dynamic Time Warping)와 같은 매칭 방법에 의해서 각 제스처와의 거리를 측정하고 인식하게 된다.

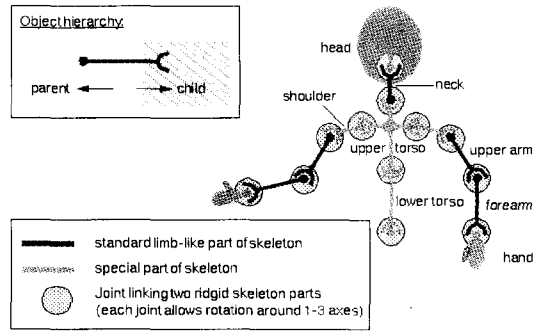


그림 3. 인간의 뼈 구조의 모델

4.2 특징 추출 기반 제스처 인식

간단한 제스처의 경우에는 복잡한 파라미터를 구하지 않고도 에지, 윤곽선, 점의 위치 등의 정보를 이용하여 쉽게 제스처를 구별할 수 있다. 먼저, 특징을 검출하기 위해서는 신체에 해당되는 영역을 배경으로부터 따로 분리해야 한다. 이를 위해 사용되는 2가지 방법이 있는데 컬러 큐(color que)와 모션 큐(motion que)가 그것이다.

컬러 큐는 피부색을 이용해 구분하는 방법으로 RGB 공간에서보다 hue- saturation 공간에서의 특징이 조명 변화에 덜 민감하다. 색깔을 이용하는 기법의 주된 단점은 밝기 조건에 따라서 피부색이 다양해진다는 것이다. 이로 인해 잘못된 세그멘테이션이 이루어질 수 있다. 특정 위치나 어떤 크기의 영역에서만 제스처가 발생한다면 그 영역만을 고려하면 되기 때문에 문제가 다소 경감될 수는 있다. 다른 해결책으로 사람이 배경과 뚜렷하게 구별되도록 배경색과 사람이 입는 옷을 정해 둘 수도 있다. 하지만 너무 부자연스럽고 절대적인 제약을 가한 것으로 근본적인 해결책은 아니다.

모션 큐는 한 사람의 제스처만 존재한다는 것과 배경이 고정되어 있다는 조건하에 움직이는 부분만을 분리하는 것이다. 이런 제약으로 배경이 고정되어 있지 않거나 한 명 이상의 제스처가 존재한 경우에는 문제가 발생한다.

4.3 외관 기반 제스처 인식

에지, 윤곽선 등의 기하학적인 정보를 이용하여 제스처 모델을 구성하고 입력 영상으로부터 이들 정보를 추출하여 모델과 매칭을 통하여 인식하는 방법은 인식 결과가 불안정한 단점이 있다. 따라서 기하학적인 특징을 이용하지 않고 영상 자체가 가지고 있는 음영 정보를 그대로 이용하고자 하는 것이 외관 기반 제스처 인식 법이다. 그러나 이것 역시 카메라의 위치에 따라 제스처 정보가 달라지고 신체 일부가 가려진 경우 잘못된 인식 결과를 얻을 수 있는 불안 요소를 안고 있다. 이를 보완하고자 하나의 카메라를 이용하지 않고 다중의 카메라를 이용하여 각각의 카메라로부터 얻은 영상을 동시에 사용한다.

전체 영상을 특징으로 사용하는 것은 MEI (Motion Energy Image), MHI(Motion History Image)와 관계가 있다[4]. MEI[4]는 영상 시퀀스의 어디에서 동작이 일어나고 있는지를 표현하는 이진화 영상으로 식 (1)로 표현되어진다.

$$E_{\tau}(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t-i) \quad \text{식(1)}$$

- $E_{\tau}(x, y, t)$: MEI
- $D(x, y, t)$: 동작 영역을 나타내는 이진화 영상 시퀀스
- τ : 시간 윈도우의 길이

MHI[4]는 더 최근에 움직인 화소들이 더 밝은 값으로 할당되어지는 영상으로

식 (2)로 표현되어진다.

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad \text{식(2)}$$

인식은 2차원 영상 클러스터링 기법을 사용하여 이루어진다. [4]에서는 각각의 제스처에 대해 여러 각도에서의 영상들을 수집하고 Hu 모멘트를 계산한다. 제스처 영상이 입력되면 이들 정보를 추출하고 저장된 제스처들의 모멘트 각각에

대해 Mahalanobis 거리를 계산하여 인식한다. 장점은 계산이 단순하다는 것이다. 그러나 모션이 누적되기 때문에 우발적인 물체의 모션이 있을 경우 문제가 생긴다.

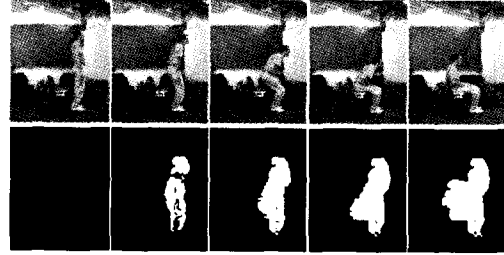


그림 4. MEI (Motion Energy Image)

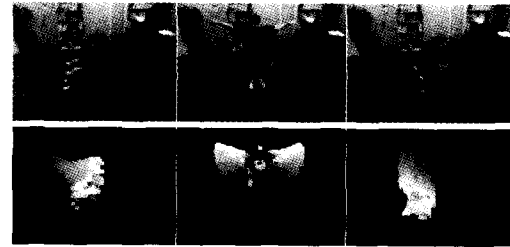


그림 5. MHI (Motion History Image)

인식에 필요한 모델 영상을 수집 정규화하고, 주성분분석법이라는 통계적인 수법을 이용하여 매칭의 기준이 되는 고유공간을 구성하는 방법이 있다. 정규화 과정을 거친 영상 집합을 이용하여 제스처들의 전체적인 외관 특징들을 표현할 수 있는 저차원 벡터공간, 즉 파라메트릭 고유공간을 생성한다. 고유 공간을 계산하기 위해서는 먼저 모든 영상에서 평균 영상을 구하여 각 영상들과의 차를 구한다. 평균 영상 c 라 할때 새로운 영상 집합 Z 는 식 (3)과 같이 나타낸다[2].

$$Z = [z_1 - c \quad z_2 - c \quad \dots \quad z_M - c] \quad \text{식(3)}$$

여기서 M 은 전체 영상의 개수이고 N 은 한 영상의 픽셀 수라 할 때 Z 의 크기는 $N \times M$ 이다. 고유 공간을 구하기 위해서는 $N \times M$ 의 크기를 지닌 영상집합 Z 를 식 (4)와 같이 계산하고 식 (5)를 만족하는 고유벡터를 구하면 된다. 즉 공분산 행렬 Q 에 대한 고유치 λ 와 고유벡터 e 를 구

한다.

$$Q = ZZ^T \quad \text{식(4)}$$

$$Qe_i = \lambda_i e_i \quad \text{식(5)}$$

고유치 분해를 위해 특이치 분해(Singular Value Decomposition)을 이용한다. 특이치 분해를 이용하면 영상집합 Z의 공분산 행렬에 대한 고유 벡터를 쉽게 얻을 수 있다. 특이치 분해 과정에서는 고유치가 큰 순서대로 고유벡터를 구할 수 있다. 각 고유 벡터가 지닌 고유치의 크기는 바로 그 고유벡터의 중요도를 의미하므로 고유 공간을 규정하는 중요 고유 벡터를 선택한다. 이제 얻어진 고유공간에 평균 영상 c에서 뺀 영상집합 x를 모두 식 (6)을 이용하여 투영시킨다.

$$\zeta_i = E^T(z_i - c) \quad \text{식(6)}$$

새로운 영상이 입력되면 평균 영상을 뺀 다음 고유공간에 투영하여 각 모델들과의 거리 중 최소 거리를 갖는 모델로 인식하면 된다. 이때 고유 공간상에서의 점들간의 거리를 이용하지 않고 스플라인 보간법(Spline Interpolation)을 이용하여 거리를 계산하면 보다 안정적인 결과 값을 얻을 수 있다[2]. [6]에서는 수많은 영상을 모델링하여 데이터베이스를 구축하고 이를 검색하는데 용이하게 하기 위한 트리(tree) 구조의 공간 분할법을 제시하였다. 모델링 된 임의의 샘플을 T라 하고 반지름을 r로 표현했을 때, 입력 벡터 V는 먼저 $\|T - V\| < r$ 을 만족하는 영역(cell)을 찾는다. 여기서 $\|\cdot\|$ 는 Euclidean 거리를 말한다. 반지름은 계층구조에서 레벨이 낮아질 수록 감소하게 되어 검색하는 영역(cell)은 점점 작아진다. 따라서 모든 공간을 검색하지 않고 일부 영역만을 검색하여도 어떤 제스처에 해당하는지를 쉽게 결정할 수 있다.

통계적 성질을 이용하여 연속 패턴을 분류하는 방법으로 은닉 마르코프 모델(HMM)이 있다. 현재 상태와 과거의 사건이 주어졌을 때, 현재 사건의 조건 확률 밀도가 과거에 발생한 사건에만 의존한다면 그 확률 과정은 마르코프의 특성

을 지닌다고 한다. 현재 사건이 가장 최근의 마지막 사건에 의존하면 그때 그 과정이 1차 마르코프 과정이다. 은닉 마르코프 모델에 대한 초기 형태는 얼마나 많은 상태들이 하나의 신호를 나타내는데 필요한지를 측정함으로써 결정되어질 수 있다. 이 형태를 잘 조율하는 것은 경험에 입각하여 이루어질 수 있다. 여러 가지의 형태들이 각 신호에 대해 형성되어질 수 있다.

은닉 마르코프 모델은 3가지 주요 단계들이 있다. 계산, 추정, 디코딩이 그것이다. 관측 결과와 모델이 주어졌을 때 계산 값은 관측된 결과가 모델($\Pr(O | \lambda)$) [16]에 의해서 생기는 확률이다. 관측된 결과들에 대해서 모든 후보 모델들의 계산이 이루어지면, 그 때 가장 높은 확률 값을 갖는 모델이 인식으로 선택되어질 수 있다. 추정은 관측 결과 O가 주어졌을 때, $\Pr(O | \lambda)$ 를 최대 화시키기 위해서 어떻게 λ 를 적용시킬 것인가 하는 것이다. 초기 모델이 주어지면, 그 모델은 균일한 확률 값들을 갖게 되고 forward-backward 알고리즘[9]은 이 확률을 계산할 수 있게 해 준다. 남은 것은 초기 모델을 진보시킬 방법을 찾는 것이다. 불행히도, 분석의 해답은 알려져 있지 않다. 그러나 되풀이하는 기법이 적용될 수 있다. 위에서 설명한 추정과 계산 과정은 은닉 마르코프 모델 시스템의 개발에 충분할 지라도 Viterbi 알고리즘[9]은 실제로 HMM's의 집합을 계산하는 빠른 방법을 제공할 뿐 아니라 디코딩 문제에 대한 해답을 제공한다. 디코딩의 목표는 관측 결과들이 주어졌을 때 상태 시퀀스를 재생하는 것이다. Viterbi 알고리즘은 각 단계에서 최대 경로가 모든 경로를 대신할 수 있다는 점에서 forward-backward 알고리즘의 특별한 형태로 보여질 수 있다. 이 최적화는 계산 양을 감소시키고 가장 유사한 상태 시퀀스를 재생한다.

파라메트릭 은닉 마르코프모델(PHMM)[10][11]은 표준 은닉 마르코프모델(HMM)의 출력 확률 안에 전체적인 파라메트릭 변화를 포함함으로써 확장한 것이다. 선형 파라메트릭 은닉 마르코프 모델(Linear PHMM)과 비선형 파라메트릭 은닉 마르코프모델(Nonlinear PHMM)로 나누어서 살펴볼 수 있다. 표준 은닉 마르코프 과정은 각 제

스처 부류의 공간적 변화를 노이즈로 간주하는 반면, 파라메트릭 은닉 마르코프 모델은 각 부류에 존재하는 공간적 변화를 복원하여 구분함으로써 표준 은닉 마르코프 과정보다 더 나은 성능 결과를 얻을 수 있다. 비선형 파라메트릭 은닉 마르코프 모델은 선형 파라메트릭 은닉 마르코프 과정보다 더 많은 제스처를 모형화 할 수 있는 장점을 갖는다.

은닉 마르코프 시스템은 2가지 큰 제약점이 있다. 첫째, 상향식 시스템 구성 방식으로 인해, 예러가 발생한 경우나 영상 특징이 빠진 경우 안정성에 문제가 생긴다. 이는 선택적 처리 기법을 적용한 하향식 시스템을 구축함으로써 해결된다. 둘째, 하나의 제스처이어야 한다는 점 또한 다중의 제스처를 동시에 인식할 수 없다는 제약을 가진다. 그러나 활성화된 상태를 색깔 있는 토큰으로 표시하는 방식으로 극복할 수 있다[14]. 다른 접근법으로는 한 쌍의 은닉 마르코프 모델(CHMM)[12][13]이 있다. 이 방식은 훈련 속도, 초기 상태에 대한 민감성 등에 있어서 은닉 마르코프 모델보다 더 우위에 있다.

4.4 가변 템플릿 기반 제스처 인식

가변 모델은 제스처의 형태와 음영 정보에서의 변화량을 나타낼 수 있다. Active Shape Model(ASM) 과 Active Appearance Model(AAM)이 여기에 속한다[20]. ASM과 AAM은 제스처의 윤곽선을 이루고 있는 특징점(landmark)을 중심으로 계산이 이루어진다. AAM은 식 (7)에서 보여지는 것처럼 형태를 나타내는 벡터 x 와 음영 정보를 나타내는 벡터 g 로 표현된다.

$$\begin{aligned} x &= \overline{x} + Q_s c \\ g &= \overline{g} + Q_g c \end{aligned} \quad \text{식(7)}$$

여기서 \overline{x} 는 제스처 형태의 평균값이고 \overline{g} 는 음영 정보의 평균값이며 Q_s, Q_g 는 변화량을 나타내는 행렬이다. AAM은 식 (8)를 이용하여 입력 영상 g_s 과 모델 영상 g_m 을 비교한다.

$$\delta g = g_s - g_m \quad \text{식(8)}$$

결국 $|\delta g|$ 의 크기가 최소가 되는 모델이 제스처로 인식된다. 검색 과정에서는 c 값을 $c - \delta c$ 로 갱신하면서 δg 를 반복적으로 계산하게 된다. δc 는 훈련 과정에서 $\delta c = A \delta g$ (A 는 행렬)의 관계로 구해진다. 이는 합성된 모델 영상과 입력 영상 사이의 차이가 최소가 되는 것을 찾는 것으로 모델의 점(landmark)과 입력 영상에서 대응되는 점 사이의 거리가 최소가 되는 것을 찾는 ASM과는 구별이 된다.

ASM은 형태 파라미터 b 등의 파라미터를 이용하여 모델에서의 점과 대응되는 입력 영상의 점의 위치를 찾는다. 매칭은 각각의 점(landmark)에 대해서 식 (9)을 계산한다.

$$f(g_s) = (g_s - \overline{g})^T S_g^{-1} (g_s - \overline{g}) \quad \text{식(9)}$$

$f(g_s)$ 의 가장 낮은 값을 갖는 모델이 제스처로 인식된다. S_g 는 공분산 행렬이다.

AAM과 ASM은 각각의 제스처에 대해 다양한 파라미터를 갖는다. 이는 모델을 구성할 때의 영상과 조금 다른 입력 영상이 있다고 하더라도 같은 제스처로 인식하는 것을 가능하게 한다. 최근에는 2차원의 가변 템플릿을 확장하여 3차원의 가변 템플릿인 Point Distribution Model(PDM)의 방법도 사용되고 있다[15].



(1) 초기 포즈 (2) 5번의 반복 후 (3) 수렴

그림 6. Active Shape Model (ASM)

V. 영상 시퀀스의 자동 분할

5.1 세그멘테이션

제스처란 추상적인 의미를 지닌 인간의 몸 동작, 손짓, 표정 등을 말한다. 따라서 제스처를 인식하기 위해서는 입력 영상으로부터 신체 부분(전경 영역)을 정확히 분리하는 것이 필요하다. 이를 분리하기 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 x 가 시간 t 에 의해서 갖는 최소 밝기 값, $N(x, t)$ 는 화소 x 가 시간 t 에 의해서 갖는 최대 밝기 값을 나타낸다.

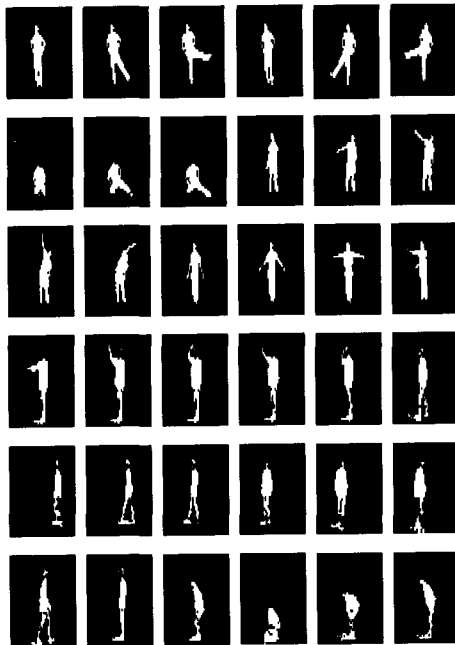


그림 7. 영상 집합의 세그멘테이션 결과

$D(x, t)$ 는 화소 x 가 가질 수 있는 최대 밝기 차이 값을 나타낸다. 전경 영역은 식 (10)에 의해서 결정되어진다[21]. 즉, 식 (10)을 만족하는 화소 x 는 모두 전경 영역으로 세그멘테이션되며 이

는 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 C 는 상수 값이다.

$$|M(x, t) - I(x, t)| \text{ OR } |N(x, t) - I(x, t)| > D(x, t) + C \quad \text{식(10)}$$

5.2 특징 추출

입력 영상 시퀀스로부터 신체 영역을 분리하는 세그멘테이션 과정이 끝나면, 전경 영역으로부터 제스처를 분석할 수 있는 신체의 외관 특징 값을 추출해야 한다. 그러나 추상적인 단순 특징을 사용하면 너무 간단하여 복잡한 형상이나 동작을 표현하기가 어렵기 때문에 복수 다차원 특징 값을 추출해야 한다.

본 논문에서는 특징 값으로 1)신체 영역의 가로축 길이(FeretX)와 세로축 길이(FeretY)의 비를 나타내는 페렛 비(Feret_ratio), 2)무게 중심의 x좌표, 3)무게 중심의 y좌표, 4)조밀성(Compactness), 5)모멘트의 주축, 6)모멘트 주축의 수직인 축이 사용되었으며 식 (12)로 표현된다. 또한 2차원 영상에서 움직임을 알아야 제스처 분석이 가능하므로, 연속된 영상의 차분 정보로부터 움직임 정보를 얻어야 한다. 따라서 그림 8에서 보여주는 것처럼 식 (11)에 의해 구해진 각 영상에 대해 복수 다차원 특징 값을 계산한다.

$$I_t^{t+n} \quad (1 \leq t \leq T-5, 0 \leq n \leq 5) \quad \text{식(11)}$$

위 식에서 각 영상(I)은 5.1절의 세그멘테이션 과정을 거친 이진 영상이고 T 는 영상 시퀀스의 총 길이를 나타낸다. 즉, 시간에 따른 신체 움직임의 변화량을 특징 값에 의해서 형상화한 것이다.

사용한 특징 값들은 수치적으로 동일한 단위를 가지고 있지 않기 때문에 정규화 과정을 거쳐야 한다. 정규화 된 특징의 집합 \mathbf{x} 는 식 (12)과 같다. 여기서 N 은 6개의 영상으로 구성된 그룹의 수로 N 은 $T - 5$ 이다. 그리고 $x_i (i = 1, \dots, N)$

는 36개의 특징 값으로 구성된다.

$$\mathbf{x} = [x_1, x_2, x_3, \dots, x_N]^T \quad \text{식(12)}$$

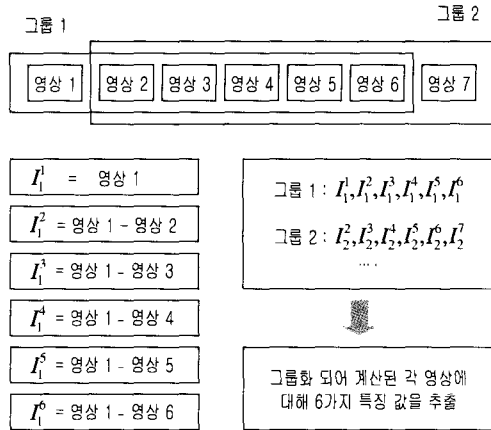


그림 8. 시간에 따른 영상 그룹화



그림 9. 시간에 따른 영상그룹화의 예 (걸는 동작의 group 1)

5.3 주성분 분석

전절에서 논의한 바와 같이 특징 집합을 이용하여 신체의 전체적인 외관 특징들을 표현할 수 있는 저차원 벡터공간, 즉 파라메트릭 고유공간을 생성한다. 고유공간을 계산하기 위해서는 먼저 모든 특징 벡터에서 평균 벡터를 구하여 각 특징들과의 차를 구한다. 평균 벡터 c 와 새로운 특징 집합 \mathbf{X} 를 식 (13)과 식 (14)와 같이 나타낸다[24].

$$c = (1/N) \sum_{i=1}^N x_i \quad \text{식(13)}$$

$$\mathbf{X} \triangleq [x_1 - c, x_2 - c, x_3 - c, \dots, x_N - c]^T \quad \text{식(14)}$$

고유공간을 구하기 위해서는 $36 \times N$ 의 크기를 지닌 특징 집합 \mathbf{X} 를 식 (14)와 같이 계산하고 식 (15)를 만족하는 고유벡터를 구하면 된다 [24]. 즉, 공분산 행렬 \mathbf{Q} 에 대한 고유치 λ 와 고유벡터 e 를 구한다.

$$\mathbf{Q} \triangleq \mathbf{X} \mathbf{X}^T \quad \text{식(15)}$$

$$\lambda_i e_i = \mathbf{Q} e_i \quad \text{식(16)}$$

고유치 분해를 위하여 특이치 분해(singular value decomposition)을 이용한다. 특이치 분해를 이용하면 특징 집합 \mathbf{X} 의 공분산 행렬에 대한 고유벡터를 쉽게 얻을 수 있다[24]. 이제 얻어진 고유공간에 평균 벡터 c 에서 뺀 특징 집합 \mathbf{x} 를 모두 식 (17)을 이용하여 투영시킨다.

$$m_i = [e_1, e_2, e_3, \dots, e_k]^T (x_i - c) \quad \text{식(17)}$$

5.4 클러스터링

영상 시퀀스에서 제스처가 바뀌는 프레임들을 자동으로 검출하여 모델을 구성하고 해석하기 위해서는 같은 제스처에 속한 영상끼리 분류하는 작업이 필요하다. 그러나 파라메트릭 제스처 공간에 투영된 점들을 클러스터링 알고리즘에 의해서 분류할 때, 영상 시퀀스 안에 몇 개의 제스처 패턴들이 존재하는지를 알 수 없기 때문에 클러스터의 개수를 몇 개로 지정할 것인가에 관한 문제가 발생하게 된다. 본 논문에서는 다변량 분산 분석법[25]을 이용하여 이런 문제를 해결하였다. 즉, 관찰 값들의 클러스터 내, 그리고 클러스터 사이의 흩어진 정도를 분산의 항목으로 측정하고 '클러스터간의 분산'이 상대적으로 '클러스터 내 분산'보다 충분히 큰 클러스터 개수를 채택하면 되는 것이다. B 를 클러스터 간 분산, W 를 클러스터 내의 분산이라고 하면 이는 각각 식 (18), 식

(19)와 같다.

$$B = \sum_{i=1}^g N_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T \quad \text{식(18)}$$

$$W = \sum_{i=1}^g \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T \quad \text{식(19)}$$

여기서 N_i 는 클러스터 i 에 포함된 점의 수, \bar{X}_i 는 클러스터 i 의 평균, \bar{X} 는 제스처 공간상의 전체 점들의 평균, g 는 클러스터 개수를 나타낸다. 그리고 전체 분산 T , 클러스터 간의 분산과 전체 분산의 비(Λ)는 각각 식 (20)과 식 (21)에 의해서 구할 수 있다.

$$T = B + W = \sum_{i=1}^g \sum_{j=1}^{N_i} (X_{ij} - \bar{X})(X_{ij} - \bar{X})^T \quad \text{식(20)}$$

$$\Lambda = \frac{|B|}{|B+W|} \quad \text{식(21)}$$

여기서 작은 Λ 값은 클러스터 내의 변동이 전체(클러스터 간)의 변동에 비해 상대적으로 작다는 것을 의미하며, 이는 클러스터간의 차이가 유의한지를 판정하는데 통계적 근거가 된다. 클러스터의 개수가 정해지면, 계층적(hierarchical) 클러스터링 알고리즘에 의해 식 (22)와 같이 제스처 영상들이 분류되어진다. 이 때 g 는 식 (22)의 판단 기준에 의해 정해지는 제스처의 종류를 나타낸다.

$$J = \{J_i\} \quad (1 \leq i \leq g) \quad \text{식(22)}$$

VI. HMM을 이용한 제스처 인식

6.1 은닉 마르코프 모델

클러스터링을 통해 식 (22)와 같이 영상 시퀀스가 여러 개의 제스처로 분류되어지면, 각각의 클러스터 J_i 는 코드북(code book)에 의해서 심

볼로 형상화되어지고 이는 은닉 마르코프 모델의 입력으로 사용된다. 은닉 마르코프 과정은 다음의 5개 요소로 정의되어진다.

S : 상태의 유한 집합 ; $S = \{s_j\}$

Y : 출력 심볼의 집합

A : 상태 천이 확률의 집합 ; $A = \{a_{ij}\}$

a_{ij} : 상태 s_i 에서 s_j 로 천이할 확률

$$\sum_j a_{ij} = 1$$

B : 출력 확률의 집합 ; $B = \{b_{ij}(k)\}$

$b_{ij}(k)$: 상태 s_i 에서 s_j 로 천이할 때 심볼 k 를 출력할 확률

$$\sum_k b_{ij}(k) = 1$$

Π : 초기 상태 확률의 집합 ; $\Pi = \{\pi_i\}$

π_i : 초기 상태가 s_i 일 확률

$$\sum_j \pi_j = 1$$

6.2 제스처 모델의 학습

은닉 마르코프 모델의 각 제스처 모델 $\{\pi, A, B\}$ 은 Baum-Welch 알고리즘에 의해서 추정되어지고 식(23)과 식(24)에 의해서 계산된다 [22].

$$\begin{aligned} \xi_t(i, j) &= \frac{P(s_t = i, s_{t+1} = j, Y | \lambda)}{P(Y | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}{P(Y | \lambda)} \\ &= \frac{\alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(y_{t+1}) \beta_{t+1}(j)} \end{aligned} \quad \text{식(23)}$$

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad \text{식(24)}$$

여기서 $\xi_t(i, j)$ 는 시간 t에서는 상태 i, 시간 t+1에서는 상태 j일 확률이고 $\gamma_t(i)$ 는 전체 관측 시퀀스와 λ 가 주어졌을 때, 시간 t에서 상태 i일 확률을 나타낸다. 식 (23), 식 (24)를 이용하여 제스처 모델은 식 (25), 식 (26), 식 (27)로 추정되어진다.

$$\bar{\pi}_j = \gamma_1(i) \quad \text{식(25)}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad \text{식(26)}$$

$$b_j(k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{s.t. o_t = v_k}^T \gamma_t(j)} \quad \text{식(27)}$$

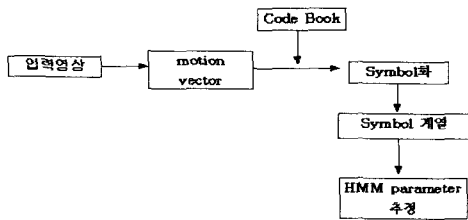


그림 10. 학습의 구성도

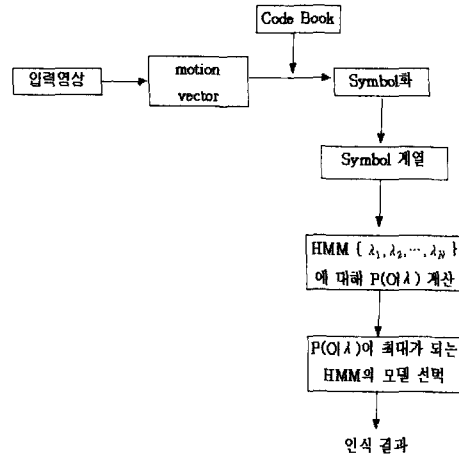


그림 12. 인식의 구성도

입력의 심볼 시퀀스(Y)가 주어지면 모델 λ_i 에 대한 확률 값은 forward 변수인 $a_t(i)$ 와 backward 변수인 $\beta_t(i)$ 를 이용하여 식 (28)과 같이 구하고 가장 높은 확률 값을 갖는 모델로 인식하게 된다.

$$P(Y|\lambda_i) = \sum_T \sum_j a_t(i) a_{ij} b_{ij}(y_{t+1}) \beta_{t+1}(j) \quad \text{식(28)}$$

6.3 제스처 인식

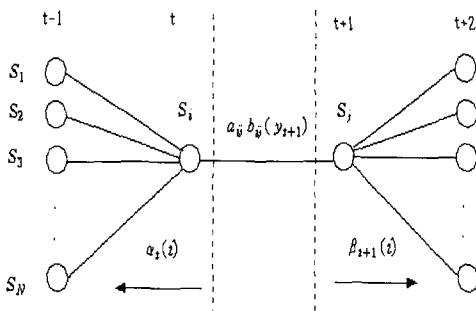


그림 11. Forward-Backward 알고리즘

VII. 실험 및 결론

실험에 사용한 제스처 영상은 걷는 동작, 앉는 동작, 일어서는 동작과 같이 일상 생활에서 우리가 매일 취하는 동작뿐만 아니라 맨손 체조에서 하는 다리 운동, 옆구리 운동, 제자리에서 걷는 운동 등의 제스처를 사용하였다. 각 영상의 크기는 320×240을 사용하였고 총 8개의 제스처 시퀀스를 모델로 구성하였다. 그 결과 모델로 구성된 제스처들은 40개의 클러스터로 분류되어졌고 HMM을 통하여 입력 영상에 대한 인식 결과를 확인하였다. 모델을 구성했던 영상 시퀀스들과 모델과 동일한 속도로 동일한 동작을 취한 영상 시퀀스들에 대해서는 거의 대부분 올바르게 인식됨을 알 수 있었고 모델과 좀 다른 동작을 취한

영상 시퀀스들에 대해서는 일부가 다른 시퀀스로 인식하는 오류를 범하기도 하였다. 그 원인들을 분석해 본 결과 우리가 사용한 특징 값들은 전체적인 외형의 변화 량 만을 보고 있기 때문에 한쪽 팔을 올리는 포즈와 다리 하나를 올리는 포즈들을 같은 포즈들로 분류한 것을 알 수 있었는데, 이는 앞으로 새로운 변화량 값을 특징으로 사용함으로써 보완해 나갈 계획이다. 우리가 실험에 사용한 데이터만 가지고는 입력되는 모든 영상을 분류해 내기란 무척 어려운 일이다. 따라서 우리가 사용한 방법이 보다 일반성을 갖기 위해 가능한 모든 영상을 수집하여 분석해 나가야 할 것이다.

그런데 우리가 취하고 있는 방법은 에지나 코너와 같은 기하학적인 특징을 이용하는 것이 아니고 면적, 영역의 가로 세로 비, 이동 량 등의 매우 추상적인 수치적 양을 이용하고 있다. 따라서 인간의 행동 전부를 인식할 수 없지만

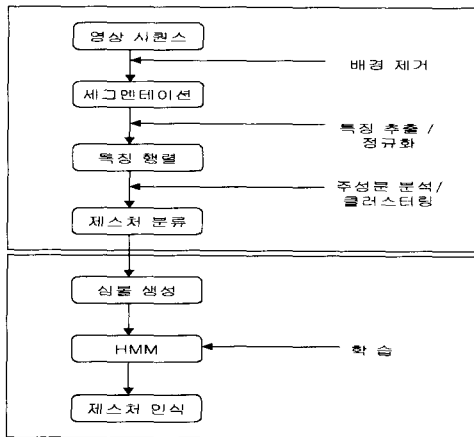


그림 13. 시스템 전체 구성도

예를 들어 공항 대합실에서 행동의 분석, 또 학교 교실 내에서의 학생들의 행동 분석, 운동장에서 행동 분석, 스포츠 분야에서의 행동 분석, 슈퍼마켓에서의 행동 분석 등 특정 좁은 분야에서 사용이 가능하다. 또 지금까지의 연구 결과에 의하면 제스처 인식에 있어서는 정적인 양보다는 변화량, 즉 사람의 키보드는 키의 변화량이 훨씬 인식에 유리하다는 결론을 내릴 수가 있었다. 전체 시스템의 구성도는 그림 13과 같다.

7.1 모델 시퀀스 영상 집합

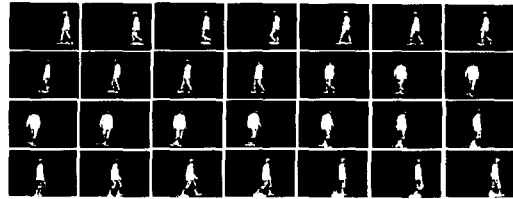


그림 14. 모델 시퀀스 1
(걸다가 방향을 바꾸어서 걷기)

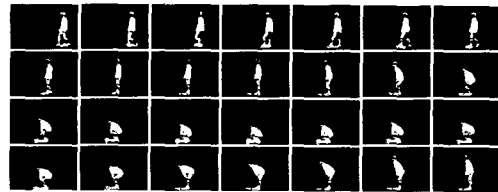


그림 15. 모델 시퀀스 2
(걸다가 앉고 일어서기)

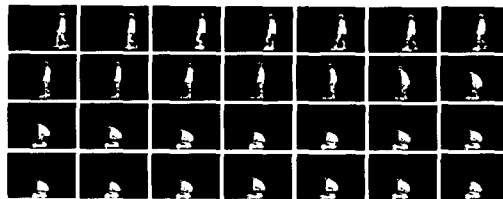


그림 16. 모델 시퀀스 3
(걷기, 앉기, 다리운동)

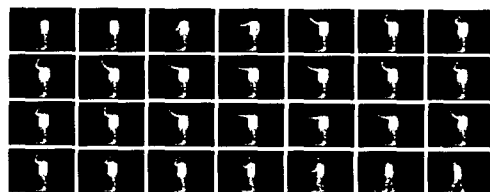


그림 17. 모델 시퀀스 4
(손을 흔들다가 걷기)

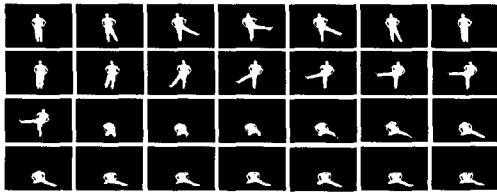


그림 18. 모델 시퀀스 5
(다리 벌리기와 다리 펴기)

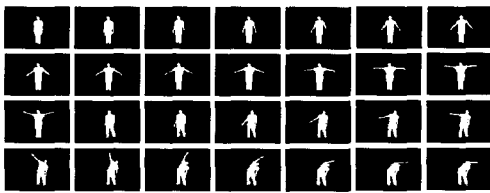


그림 19. 모델 시퀀스 6
(양팔 올리기와 옆구리 운동)

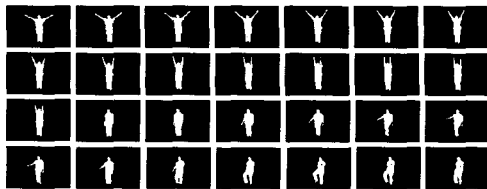


그림 20. 모델 시퀀스 7
(양팔 올리기와 제자리 걷기)

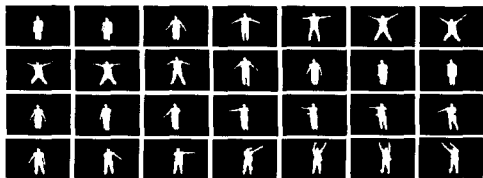


그림 21. 모델 시퀀스 8
(팔 벌려 뛰기와 팔 돌리기)

7.2 모델 시퀀스를 제스처 공간으로 투영시킨 결과

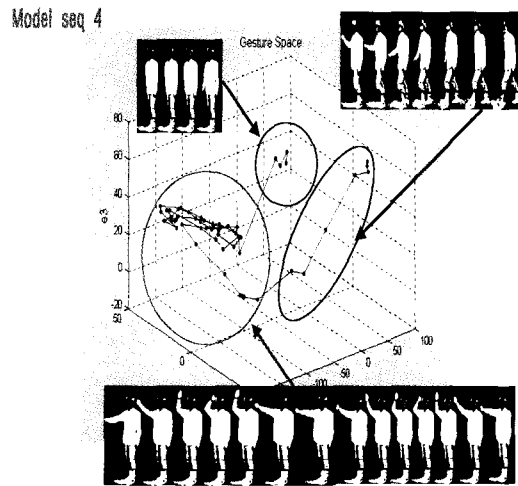


그림 22. 모델 시퀀스 4를 제스처 공간에 투영한 결과

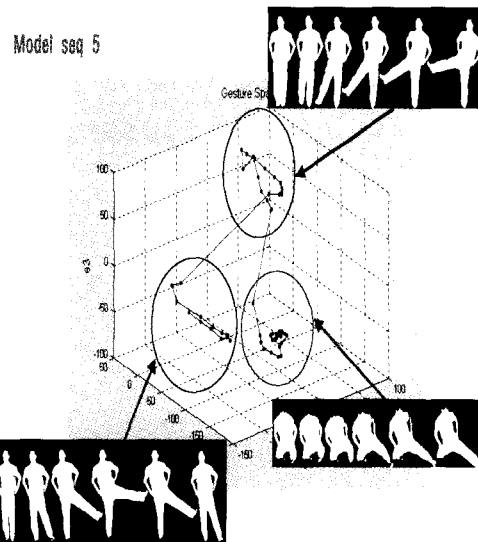


그림 23. 모델 시퀀스 5를 제스처 공간에 투영한 결과

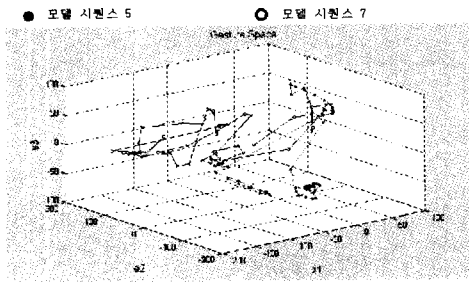


그림 24. 두 개의 영상 시퀀스를 제스처 공간에 투영한 결과 (모델 시퀀스5와 모델 시퀀스 7)

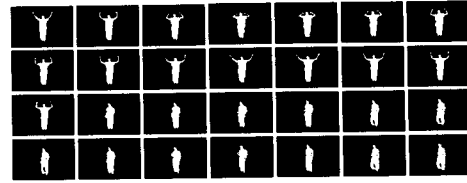


그림 28. 모델 시퀀스 7의 변형된 입력 시퀀스

7.4 입력 영상 시퀀스와 모델 영상 시퀀스를 투영한 결과

7.3 입력 영상 집합

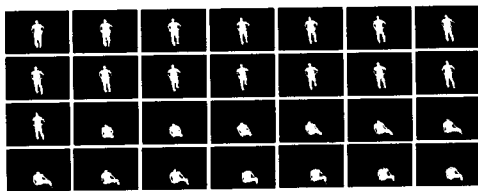


그림 25. 모델 시퀀스 5의 변형된 입력 시퀀스

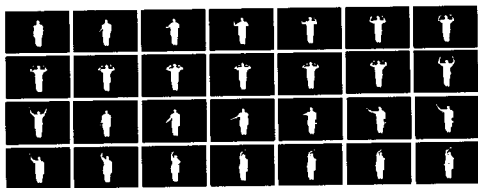


그림 26. 모델 시퀀스 6의 변형된 입력 시퀀스

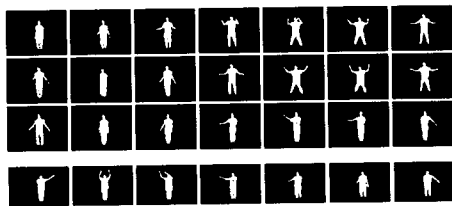


그림 27. 모델 시퀀스 8의 변형된 입력 시퀀스

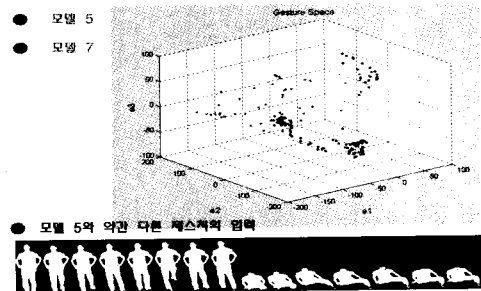


그림 29. 입력 시퀀스의 제스처 공간으로의 투영 1

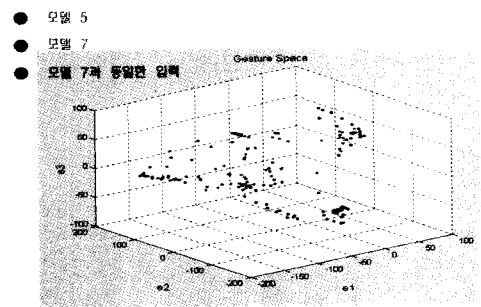


그림 30. 입력 시퀀스의 제스처 공간으로의 투영 2

참 고 문 헌

- [1] Takahiro Watanabe and Masahiko Yachida, "Real Time Recognition of Gesture and Gesture Degree Information Using Multi Input Image Sequence", ICPR, 1998
- [2] Shigeyoshi Hiratsuka, Kohtaro Ohba, Hikaru Inooka, Shinya Kajikawa, and Kazuo Tanie, "Stable Gesture Verification in Eigen Space", LAPR Workshop on Machine Vision Application, Nov. 17-19, 1998
- [3] Jakub Segen, Senthil Kuma, "Fast and Accurate 3D Gesture Recognition Interface", ICPR, 1998
- [4] James W. Davis, Aaron F. Bobic, "The Representation and Recognition of Action using Temporal Templates", CVPR, 1997
- [5] D.M. Gavrilu, L.S. Davis, "Towards 3D model-based tracking and recognition of human movement: a multi-view approach", Int. Workshop on Face and Gesture Recognition, 1995
- [6] Yuntao Cui, Daniel L. Swets, and John J. Weng, "Learning-Based Hand Sign Recognition Using SHOSLIF-M"
- [7] A. F. Bobick, Y. A. Ivanov, "Action Recognition using Probabilistic Parsing", CVPR, 1996
- [8] Trevor J. Darrel , Alex P. Pentland, "Recognition of Space-Time Gesture using a Distributed Representation", Technical Report
- [9] Thad Starner, Alex Pentland, "Real-Time American Sign Language Recognition from Video using Hidden Markov Models", ISCV, 1995
- [10] Andrew D. Wilson, Aaron F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition", IEEE Transaction on PAMI, Vol. 21, No. 9, September 1999
- [11] Andrew D. Wilson, Aaron F. Bobick, "Recognition and Interpretation of Parametric Gesture", ICCV, 1998
- [12] Matthew Brand, Nuria Oliver, and Alex Pentland, "Coupled Hidden Markov Models for complex action recognition", CVPR, 1997
- [13] Christian Vogler, Dimitris Metaxas, "ASL Recognition Based on a Coupling Between HMMs and 3D Motion Analysis", ICCV, 1998
- [14] Toshikazu Wada, Takashi Matsyama, "Appearance Based Behavior Recognition by Event Driven Selection Attention", CVPR, 1998
- [15] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Transaction on PAMI, Vol. 19, No. 7, July 1997
- [16] X. Huang, Y. Ariki, and T. O'Rourke. A Basic Course in American Sign Language. T. J. Publ.,Inc., Silver Spring, MD, 1980
- [17] Tony Jaara, Alex Pentland, "Action Reaction Learning: Analysis and Synthesis of Human Behavior", CVPR, 1998
- [18] 이인호, 박찬중, "모션캡처 기술의 현황과 응용 분야", 멀티미디어학회지, 1999
- [19] James W.Davis, Aaron F. Bobick, " Virtual PAT: A Virtual Personal Aerobics Trainer", 1998
- [20] T.F. Cootes, G. Edwards and C.J. Taylor, "Comparing Active Shape Models with Active Appearance Models", Proc. British Machine Vision Conference, Vol. 1, 1999, pp173-182
- [21] Ismail Haritaoglu, David Harwood and Larry S. Davis, "W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People", International Conference on Face and Gesture Recognition, 1998, pp. 14-16
- [22] Yoshio IWAI, Tadashi HATA, and Masahiko YACHIDA, "Gesture Recognition based on Subspace Method and Hidden Markov

Model", IEEE, 1997, pp. 960-966

[23] Ismail Haritaoglu, Ross Cutler, David Harwood and Larry S. Davis, "Backpack: Detection of People Carrying Objects Using Silhouettes", IEEE International Conference on Computer Vision (ICCV), 1999

[24] 이용재, 이철우, "외관 기반의 파라메트릭 고유 공간을 이용한 물체인식", 정보과학회, 1999

[25] 김기영, 전명석, "다변량 통계 자료 분석", 자유 아카데미



김 동 욱

- 1987년 2월 성균관대학교 전자공학과 공학사
- 1992년 2월 중앙대학교 대학원 전자공학과 공학석사
- 1996년 8월 중앙대학교 대학원 전자공학과 공학박사
- 1997년 3월~1998년 2월 청운대학교 전임강사
- 1998년 3월~현재 전주대학교 전자매체공학부 조교수

저 자 소 개



강 응 관

- 1993년 2월 중앙대학교 전자공학과 졸업(공학사).
- 1995년 2월 중앙대학교 대학원 전자공학과 졸업(공학석사).
- 1995년 1월~1997년 5월 현대전자정보통신연구소 연구원.
- 2001년 2월 중앙대학교 첨단영상대학원 영상공학과 졸업(공학박사).
- 2001년 9월~현재 전주대학교 정보기술컴퓨터공학부 멀티미디어전공(조교수).
- 관심분야: 동영상 부호화, 멀티미디어 신호처리, 멀티미디어 데이터베이스, 영상통신 등임



나 인 호

- 1998년 2월 울산대학교 전자계산학과(공학사)
- 1991년 2월 중앙대학교 대학원 전자계산학과(공학석사)
- 1995년 8월 중앙대학교 대학원 전자계산학과(공학박사)
- 1995년 9월~현재 : 군산대학교 전자정보공학부 조교수
- 1997.7.~현재:전주 첨단영상산업 추진협의회 위원
- 1997.7.~현재:한국 해양정보통신학회 편집 위원
- 1997. 12.~1998.3:전자통신연구원 초빙 연구원
- 1999. 10.~2000.10:전주국제컴퓨터게임 축제조직위원
- * 관심분야 : 멀티미디어 통신시스템, 분산시스템, 병렬처리