제스처 인식과 응용에 관한 연구

이 현주, 이칠우
전남대학교 컴퓨터공학과

Gesture Recognition and Applications

Hyun-Ju Lee, Chil-Woo Lee
Dept. of Computer Engineering, Chonnam National Univ.
Jeehj98@hanmail. net

요약

인간은 일상 생활에서 제스처와 같은 비언어적 수단을 이용하여 수많은 정보를 전달한다. 따라서 인간과 정보 시스템이 자연스럽게 대화할 수 있는 지적인 인터페이스를 구축하기 위해서 제스처 인식에 관한 연구는 필수적이다. 그러나 인체는 매우 복잡한 구조를 지닌 3차원 물체이므로 제스처를 자동으로 인식하는 것은 매우 어려운 일이다. 본 논문에서는 제스처 인식을 위한 중요 기술을 중심으로 최근의 연구 동향에 대해 기술한다.

1. 서론

컴퓨터 기술의 발달과 함께 정보 시스템이 복잡하게 되면서 인간과 정보 시스템 사이에 자연스럽게 정보를 교환할 수 있는 지적 인터페이스에 관한 관심이 날로 커지고 있다. 인간은 일상 생활에서 제스처, 표정과 같은 비언어적인 수단을 이용하여 수많은 정보를 전달한다. 따라서 자연스럽고 지적인 인터페이스를 구축하기 위해서는 제스처와 같은 비언어적 통신 수단에 대한 연구가 매우 중요하다. 최근에 들어, 대규모 비디오 데이터베이스의 구축, 감시 시스템, 고압축 통신 시스템의 구축을 위해 제스처 인식에 관한 연구가 활발히 진행되고 있다.

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 인식하는 것은 매우 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 연결케이블 때문에 자유스런 제스처 입력이 불가능하여 현재는 거의 사용되지 않고 있다. 최근에 들어, 광학적 마커를 몸에 부착하고 카메라로 입력된 영상으로부터 마커들의 궤적을 추적하여 제스처를 인식하는 방법들이 개발되었다.

본 논문에서는 제스처 인식에 대한 최근의 연구성과를 중심으로 요소 기술의 중요 내용과 응용에 대해서 설명한다.

2. 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 제스처 인식 과정

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

3. 제스처 인식의 분류

3.1 센서 부착식 제스처 인식

모션캡쳐에서 주로 사용되는 방식으로 인식보다는 계측, 측정의 의미가 강하다.

3.1.1 기계식 시스템

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

3.1.2 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[1읽.

3.1.3 광학식 방식

광학식 시스템은 사람의 주요 관절부분에 적외선에 반응하는 적의선 마커(혹은 광 다이오드)를 부착하고 여기에 적의선 빛을 비추어 반사되는 영상을 3대에서 32대 가량의 CCD 카메라로 촬영하고 각 카메라에서 마커들의 2차원 좌표를 생성한다[18]. 각 독립된 카메라에서 캡쳐된 2차원 데이터는 소프트웨어로 분석되어 3차원 공간상의 좌표를 계산한다.

3.1.4 음향식 시스템

이 시스템은 다수의 초음파 발생장치와 3개의 수신장치로 구성된다. 사람의 각 관절에 부착된 초음파 발생장치들은 순차적으로 초음파를 발생하고, 그초음파가 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.

3.2.1 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 부적함하다.

3.2.2 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 템플리트 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 템플리트는 입력 영상이 모델영상에서의 제스처와 조금 다른 형태를 가지고 있더라고 같은 제스처로 인식할 수 있는 방법으로 특징파라미터를 이용하는 방법과 의관을 이용하는 방법의 중간 형태로 두 방법의 개념을 함께 사용한 것이다.

4. 최근의 연구동향

4.1 3차원 모델 이용법

4.1.l3차원 모델

3차원 제스처의 모델은 크게 두 가지로 구분할 수 있다.
· 체적(볼륨) 모델(volumetric Model)
· 골격 모델(skeletal Model)
체적 측정의 모델은 주로 컴퓨터 에니메이션에서 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 3가지파라미터를 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. 둘 째는, 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.

4.1.2 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 비춤으로써 마커의 위치 정보를 쉽게 얻을 수 있는 장점을 가지고 있다. 어깨나 엉덩이의 관절 상태(DoF)는 고도($\varnothing$), 근육의 의전 운동(Ψ), 비틀림(Θ)의 3가지 Euler 각도를 이용하여 나타낼 수 있다[5]. 팔꿈치나 무릎은 신전(extension)의 파라미터 p를 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.

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 평 이상의 제스처가 존재한 경우에는 문제가 발생한다.

4.3 의관 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 보안하고자 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.

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

$$\text{- } E_\tau(x,y,t) : MEI$$

-D(x,y,t) : 동작 영역을 나타내는 이진화 영상 시퀀스

-τ : 시간 원도우의 길이

MHI[4]는 더 최근에 움직인 화소들이 더 밝은 값으로 할당되어지는 영상으로 식 (2)로 표현되어진다.

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

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.

인식에 필요한 모델 영상을 수집 정규화하고, 주성분 분석법이라는 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. 고유공간을 계산하기 위해서는 먼저 모든 영상에서 평균영상을 구하여 각 영상들과의 차를 구한다. 평균 영상 c라 할때 새로운 영상 집합 Z는 식 (3)과 같이 나타낸다[2]

$$Z=[z_1-c \;\; z_2-c \;\cdots\; z_M-c] \qquad (3)$$

여기서 M은 전체 영상의 개수이고 N은 한 영상의 픽셀 수라 할 때 Z의 크기는 N×M이다. 고유 공간을 구하기 위해서는 N×M의 크기를 지닌 영상집합Z를 식 (4)와 같이 계산하고 식 (5)를 만족하는 고유벡터를 구하면 된다. 즉 공분산 행렬 Q에 대한 고유치 λ와 고유벡터 e를 구한다.

$$Q=ZZ^T \qquad (4)$$

$$Qe_i=\lambda_i e_i \qquad (5)$$

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 중요고유 벡터를 선택한다. 이제 얻어진 고유공간에 평균 영상 c에서 뺀 영상집합 x를 모두 식 (6)을 이용하여 투영시킨다.

$$\zeta_i=E^T(z_i-c) \qquad (6)$$

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. 모델링된 임의의 샘플을 T라 하고 반지름을 r로 표현했을 때, 입력 벡터 V는 먼저 $//T-V//<r$을 만족하는 영역(cell)을 찾는다. 여기서 ‖·‖는 Euclidean 거리를 말한다. 반지름은 계층구조

에서 레벨이 낮아질 수 록 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 이용한 것으로 은닉 마르코프 모델(HMM)이 있다. 현재 사건들, 과거의 사건들 모두 주어졌을 때 현재 사건의 조건 확률 밀도가 단지 가장 최근에 발생한 j번째 사건에만 의존한다면 시간도메인 과정은 마르코프의 특성을 지닌다고 말할 수 있다. 현재 사건이 가장 최근의 마지막 사건에 의존하면 그때 그 과정이 1차 마르코프 과정이다. 은닉마르코프 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. 여러가지의 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, $Pr(O1\lambda)$를 최대화시키기 위해서 어떻게 λ를 적용시킬 것인가 하는 것이다. 초기 모델이 주어지면, 그 모델은 균일한 확률 값들을 갖게 되고 forward-backward 알고리듬[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 알고리듬[9]은 실제로 HMM's의 집합을 계산하는 빠른 방법을 제공할 뿐 아니라 디코딩 문제에 대한 해답을 제공한다. 디코딩의 목표는 관측 결과들이 주어졌을 때 상대 시퀀스를 재생하는 것이다. Viterbi 알고리듬은 각 단계에서 최대 경로가 모든 경로를 대신할 수 있다는 점에서 forward-backward 알고리듬의 특별한 형태로서 보여질 수 있다. 이 최적화는 계산양을 감소시키고 가장 유사한 상태 시퀀스를 재생한다.

파라메트릭 은닉 마르코프 모델(PHMM)[10][11]은 표준 은닉 마르코프 모델(HMM)의 출력 확률 안에 전체적인 파라메트릭 변화를 포함함으로써 확장한 것이다. 선형 파라메트릭 은닉 마르코프 모델(Linear PHMM)과 비선형 파라메트릭 은닉 마르코프 모델(Nonlinear PHMM)로 나누어서 살펴볼 수 있다. 표준 은닉 마르코프 과정은 각 제스처 부류의 공간적 변화를 노이즈로 간주하는 반면, 파라메트릭 은닉마르코프 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 은닉마르코프 과정보다 더 많은 제스처를 모형화 할 수 있는 장점을 갖는다.

은닉 마르코프 시스템은 2가지 큰 제약점이 있다. 첫째, 상향식 시스템 구성 방식으로 인해. 에러가 발생한 경우나 영상 특징이 빠진 경우 안정성에 문제가 생긴다. 이는 선택적 처리 기법을 적용한 하향식시스템을 구축함으로써 해결된다. 둘째, 하나의 제스처이어야 한다는 점 또한 다중의 제스처를 동시에 인식할 수 없다는 제약을 가진다. 그러나 활성화된 상태들을 색깔 있는 토큰으로 표시하는 방식으로 극복할 수 있다[l14]. 다른 접근법으로는 한 쌍의 은닉마르코프 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 템플리트 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[22]. ASM과 AAM은 제스처의 윤곽선을 이루고 있는 특징점(landmark)을 중심으로 계산이 이루어진다. AAM은 식 (7)에서 보여지는 것처럼 형태를 나타내는 벡터 x와 음영 정보를 나타내는 벡터 g로 표현된다.

$$x = \overline{x} + Q_s c$$
$$g = \overline{g} + Q_g c \qquad (7)$$

여기서 $\overline{x}$는 제스처 형태의 평균값이고 $\overline{g}$는 음영정보의 평균값이며 Q_s ,Q_g는 변화량을 나타내는 행렬이다. AAM은 식 (8)를 이용하여 입력 영상 g_s과 모델 영상 g_m을 비교한다.

$$\delta g = g_s - g_m \qquad (8)$$

결국 $|\delta g|$의 크기가 최소가 되는 모델이 제스처로 인식된다. 검색 과정에서는 c값을 $c-\delta c$로 갱신하면서 δg를 반복적으로 계산하게 된다 δc는 훈련과정에서 $\delta c = A\delta g$ (A는 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.

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

$f(g_s)$의 가장 낮은 값을 갖는 모델이 제스처로 인식된다. S_g는 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 템플리트를 확장하여 3차원의 가변 템플리트인 Point Distribution Model(PDM)의 방법도 사용되고 있다[15].

5. 응용 예

5.1 가상 지휘자 시스템

이 시스템[1]은 지회자의 제스처를 인식하고 제스처의 속도와 크기 정보를 얻는다. 속도 정보는 컴퓨터에 의해서 연주되는 음악의 속도를 제어하고 크기정보는 소리를 제어한다. 제스처의 속도는 단위 시간 동안 이동한 거리로 얻을 수 있다. 이는 무게 중심을 이용한다. 제스처의 크기는 하나의 제스처를 취하는데 이동한 거리의 함으로 계산된다.

5.2 The KIDSROOM

이 시스템[4]은 대화를 할 수 있는 어린이용 놀이장소이다. 그 방은 최대 4평의 어린이를 감지하고 어린이들이 하는 행동에 의해 영향을 받는데, 이야기를 통해서 진행이 된다. 컴퓨터는 조명, 음향 효과, 점수 그리고 두 벽에 투사되는 장면들을 제어한다. 현재 시나리오는 Monsterland의 어드벤쳐 여행이다.

5.3 Virtual PAT: A Virtual Personal Aerobics Trainer

이 시스템[19]은 보통의 체조 비디오 테이프나 TV의 운동 쇼와는 달리, 에어로빅 세션을 사용자의 요구대로 만들 수 있다. 다양한 미디어 기술과 컴퓨터 알고리듬으로 가상의 강사와 대화할 수 있는 시스템을 구축한 것이다.

5.4 Multiple-Perspective Interactive Video (MPI -Video)

보통의 비디오는 작동. 되감기, 빠르게 감기 등의 매우 기본적인 제어만 가능하다. 또한 사용자는 비디오 에디터가 조작하는 장면만 볼 수 있다. 이에 반해 MPI-Video는 사용자가 직접 다중의 카메라 중 보고자 하는 카메라 영상을 선택해서 볼 수 있고 질의에 의해서 비디오 시퀀스 중에 원하는 내용을 선택할 수 있다.

5.5 기타

Artificial Life Interactive Video Environment는 사람이 가상의 창조물과 대화할 수 있는 시스템이다. 이 외에도 비디오게임 제어 시스템, 가상 비행 시뮬레이션 시스템, City of News, Dance Space 등의 응용들이 존재한다.

6. 결론

지금까지 제스처 인식의 최근 연구 동향과 그 응용에 대해서 알아보았다. 어떤 의미를 지닌 시공간상의 패턴을 인식하기엔 인체가 매우 복잡한 구조를 갖고 있어서 인식 과정에는 여러 제약 사항들이 수반된다. 그러나 인간 정보 시스템과 자연스럽게 대화할 수 있는 지적인 인터페이스를 구축하고자 한다. 향후 연구에서는 기존의 요소 기술을 서로 접목시켜 복잡한 배경 속에서도 제스처를 인식할 수 있도록 다방면으로 연구를 해야 할 것이다.

[참고문헌]

[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

[4l James W. Davis, Aaron F. Bobic, "The Representation and Recognition of Action using Temporal Templates", CVPR, 1997

[5] D.M. Gavrila, L.S. Davis,"Towards 3D model- based tracking and recognition of human movement: a multi-view approach", mt. Workshop on Face and Gesture Recognition, 1995

[6] Yuntao Gui, Daniel L. Swets, and John J. Weng, "Learning-Based Hand Sign Recognition Using SHOSLIP-M"

[7] A. F. Bobick, Y. A. Ivanov, "Action Recognition using Probabilistic Parsing", CVPR, 1996

[8] Trevor J. Dan-el Alex P. Pentland, "Recognition of Space-Time Gesture using a Distributed Representation", Technical Report

[9] JThad Stainer, 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] Mattew 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. Paviovic, 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] JTony Jaara, Alex Pentland, "Action Reaction Learning: Analysis and Synthesis of Human Behavior", CVPR, 1998
[18] 이인호, 박찬종, “모션캡쳐 기수의 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-l82