

게임 인터페이스를 위한 모션인식 기술의 현황과 활용 방안

남 양 희 / 이화여자대학교 디지털미디어학부 교수*

목 차

- I. 서론
- II. 모션 인식 기술의 분류와 현황
- III. 게임에서의 모션 인식 기술 적용방안
- IV. 결론

인간의 모션을 포착하고 인식하는 것은 키보드나 조이스틱과 같은 인공적인 인터페이스를 사용하지 않아도 된다는 점에서 주목받고 있는 차세대 휴먼-컴퓨터 인터페이스 기술이다. 그러나, 사람의 동작이라는 것이 3차원의 공간에서 행해지고, 동일한 사람이라도 때에 따라 같은 동작을 다르게 취할 수 있으며, 서로 다른 사람의 경우는 의미상 같은 동작이라도 스케일이나 행하는 패턴이 크게 차이가 날 수 있기 때문에 일반적인 게이머 모션을 '인식'하는 인터페이스는 아직 찾아볼 수 없다.

본 논문에서는 인간의 모션을 추적하고 인식하는 컴퓨터 비전 기술의 현황을 소개함과 더불어, 대상 모션의 종류 및 복잡도가 인식 기술이나 정확도에 어떠한 영향을 끼치는지 설명함으로써, 향후 게임 인터페이스에 모션 인식 기술을 채택할 때 모션 종류와 인식엔진의 설계 및 적용을 위한 방안을 제시하고자 한다.

Key word : 모션 인식, 게임 인터페이스, 영상 처리, HCI

* 남양희는 KAIST에서 공학박사 학위를 취득하고, 인공지능연구센터(CAIR), 스위스 EPFL 및 일본의 ATR 연구소의 선임연구원으로서 제스처 인식기술 기반의 상호작용과 가상현실 분야를 연구해왔다. 『Recognition of Hand Gestures with 3D, Nonlinear Arm Movements』 등 다수의 국내외 저널 논문을 저술하였으며, 세종대학교를 거쳐 현재 이화여대 디지털미디어학부에 재직 중이다. yanghee@ewha.ac.kr

I. 서론

인간이 특별한 장치의 장착 없이 컴퓨터와 상호작용하기 위해서는 지능적인 컴퓨터 기술에 의한 모션의 포착과 인식이 필요하다. 이에 컴퓨터 비전 분야에서는 차세대 휴먼-컴퓨터 인터페이스 기술로서 인간의 동작과 인식 방법을 연구해왔으나, 3차원의 모션이 내포하는 데이터의 복잡성 및 오류 패턴의 다양성, 그리고, 동일한 모션이라도 때와 장소와 사람에 따라 패턴이 달라짐으로 인해 모션들을 일반화하여 인식하는 것이 매우 어렵다고 알려져 있다.

최근에는 DDR(Dance Dance Revolution)이나 EyeToy와 같이 마치 인간의 동작을 인식이라도 하는 듯한 게임 인터페이스들이 등장하고는 있으나, 전자의 경우는 손으로 키버튼을 누르는 대신 발판의 각 방향 센서를 누르는 것이므로 내부에서 모션을 포착하거나 인식하는 인터페이스라고 보기 어렵다. 후자, 즉, EyeToy와 같은 경우는 카메라로 모션 센싱을 하는 것이므로 모션 기반 인터페이스를 채택했다고 할 수 있으나, 모션을 행하는 대상이나 행위의 내용을 파악하지 않고 단순한 영상 차이 여부만을 추적하는 방법을 이용하였다. 3년간 연구개발 끝에 채택한 EyeToy의 기술은 사실상 90년대 초반에 이미 캐나다의 Vivid 그룹이 Mandala 시스템을 통해 선보인 것과 유사한 스트림 영상의 합성과 영역 추적 방법을 개선하여 게임에 적절히 결합한 것이라 할 수 있다. 즉, 간단한 움직임 추적 기술을 효과적으로 게임에 도입한 사례라고 할 수 있으며, 이와 더불어 소니사는 차세대 플레이스테이션에서 모션센서를 내장하고 사용자들의 손동작이나 눈의 움직임, 혹은 감정을 사용하여 컴퓨터와 상호 동작하는 개념을 도입할 예정이라고 한다.

인간의 모션과 게임을 결합시키는 위와 같은 시도에서 중요한 것은 복잡한 기술의 채택보다는 모션의 복잡도에 따른 인식 기술의 가능성과 한계점을 알고, 그에 걸맞은 효과적인 게임 인터페이스를 설계하는 것이다. EyeToy의 주 개발자인 Ron Festejo 역시 기술 개발 자체도 어렵긴 하지만 그런

〈그림 1〉 Mandala 시스템의 상업적 광고에 사용된 소개 이미지



〈그림 2〉 플레이스테이션 용 EyeToy 게임 소개 광고 이미지



기술을 어떤 게임에 적용해서 게이머들에게 어떻게 다가가야 하는지를 설정하기가 더욱 어려웠다고 한다. 따라서, 이러한 시점에 모션 인식의 세계적인 기술 현황과 제한점을 분명히 함으로써 향후 게임에의 활용 방안을 분석해볼 필요가 있겠다.

이에, 본 소고에서는 컴퓨터 비전 분야의 실시간 모션 인식 기술의 분류와 현황 및 한계점을 소개하고, 이를 바탕으로 게임에서의 적용 가능성, 활용 방안을 논하고자 한다.

II. 모션 인식 기술의 분류와 현황

게임을 위한 모션 인터페이스에서 모션 인식이라 함은 컴퓨터가 해당 모션을 다른 모션과 구분되게 파악한다는 것이다. 흔히 EyeToy와 같이 사람의 모션에 따라 시스템이 반응을 하는 경우 이를 모션 인식으로 착각하는데, 이는 모션의 종류나 의미를 분석하지 않고 몸체형상의 실루엣을 추적하는 것이므로 모션 인식이 아니라 모션 추적 기술이라고 언급하는 것이 정확하다.

이와 같이, 추적된 실루엣이나 3차원 동선 데이터로부터 그 의미나 부류를 추정해 내는 것이 인식 기술이며, 모션 추적은 모션 인식에 포함되는 하나의 단계라 하겠다. 모션 인식 분야의 국제적인 기술 현황을 명확히 분류하기는 어렵지만, 하나의 고정된 카메라 뷰에서 취득된 2차원의 영상 시퀀스로부터 모션을 인식하는 것과 센서를 사용자의 몸에 장착하거나 혹은 카메라를 여러 대 이용하여 데이터 자체를 3차원 값으로 얻어 이를 분석하는 기술들로 크게 분류할 수 있다. 현재는 센서 부착없이 카메라 영상을 기반으로 한 모션 분석이 주류를 이루므로 이를 중심으로 모션 인식 기술들을 소개하기로 하겠다.

1. 하나의 고정된 카메라 뷰에 기반한(view-specific) 모션 인식

고정된 카메라 뷰로부터 인식하는 경우, 당연한 것이지만 해당 카메라 뷰에서 잘 구분되지 않는 3차원 동작은 인식이 어렵게 된다. 예컨대, 다리를 굽히는 동작은 옆 면에서 실루엣의 변화로 구분 가능하지만 정면에 고정된 카메라를 두었다면 다리를 굽힌 것인지 여부를 구분하기 어려워진다.

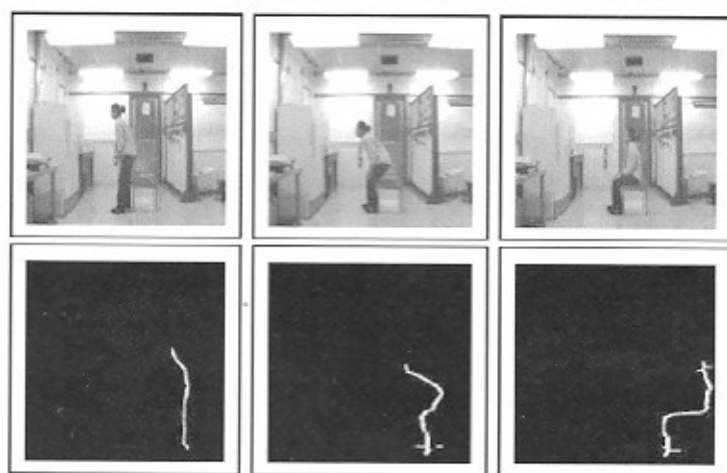
그럼에도 불구하고, 인식할 대상 모션의 종류가 고정 카메라 시선방향에서 확연히 구분되는 경우, 멀티 카메라 뷰를 통해 3차원 모션을 추적하는 것보다 기술적인 측면에서 더 용이하게 모션 인식을 할 수 있다. 예컨대, 몸 전체를 굽혀 인사하거나 앉고 서는 동작의 경우는 측면 뷰에서 전신의 모양에 차이가 크므로 구분하기 쉽다. 이러한 경우, 고정된 카메라 뷰를 통한 2차원적인 영상 시퀀스의 분석이 오히려 문제의 복잡도를 줄이고 인식률을 높이는 방법이 될 수 있다. 즉, 모션 인터페이스의 설계자가 대상 모션의 종류를 어떻게 설정하는가에 따라 하나의 카메라뷰로 충분히 인식할 수 있는 모션 인터페이스를 설계할 수 있다.

이러한 모션 기반 인터페이스에 대해, MIT의 A. Bobick, A. D. Wilson, A. Pentland 팀, U. Penn의 D. Metaxas 팀(현재는 Rutgers 대학으로 옮김), Texas Austin의 Aggarwal 팀, 로체스터 대학의 D. H. Ballard 팀, 독일 뮌헨 공대, KAIST 등을 비롯하여 국내외 우수 기관에서 10년 여에 걸쳐 많은 연구 결과를 발표해왔다. 이들의 연구 결과는 대개 동작 가운데 가장 복잡하다는 손 제스처어만을 대상으로 한 것들이 많고, 일부는 전신 동작을 대상으로 기술 개발한 것이다. 분리된 동작의 자세나 모션 패턴을 인식하는 것과 연속되게 행해지는 동작을 인식하는 것에 따라 방법과 난이도가 크게 달라짐을 알 수 있다.

1) 분리된 단일 모션의 인식

분리된 단일 모션은 시작과 끝이 명확하게 주어지고 그 사이에 행해진 하나의 동작을 말한다. 이러한 단일 모션 인식은 연속 모션 인식에 비해 상대적으로 간단한 편이다. [그림3]과 같이 전신 동작의 경우 고정 카메라 영상으로부터 몸체 실루엣과 뼈대구조를 추출하고 관절각의 특성에 따라 기하형상을 비교함으로써 대상 모션들을 구분짓는 것이 한가지 방법이며(Ali, 2001), 영상을 구성하는 픽셀조합 전체를 템플릿으로 하여 템플릿 매칭 결과를 추적하는 방법을 쓰기도 한다(Bobick, 2001). 이 때 대상 모션들은 '앉기', '걷기', '일어나기', '웅크리기', '굽히기' 등의 제한된 일상 동작인 경우가 많으며(Ali, 2001), 에어로빅 동작과 같이 특수한 모션 18종을 다룬 경우도 있다(Bobick, 2001). 그러나, 후자의 경우, 에어로빅 전문가사 1인의 모션만을 대상으로 하였다라는 점에 주목할 필요가 있는데, 이는 다수의 사람을 대상으로 할 경우 동종의 모션이라도 패턴의 변화가 너무 다양하기 때문이

〈그림 3〉 전신동작의 정지영상으로부터 뼈대구조의 추출(Ali, 2001)



다. 따라서, 고정된 카메라 영상 뷰에 기반한 모션 인식의 경우, 다수의 사람을 대상으로 모션 인식을 적용하려면 구분할 동작의 개수를 7~8개 정도로 제한하는 것이 현실성이 있음을 알 수 있다.

고정 카메라 뷰 영상으로부터 모션을 인식하는 일반적인 방법은 [그림4]와 같이 도식화할 수 있다. 각 과정들을 상술하면 다음과 같다.

(1) 물체 영상 또는 실루엣 분리

먼저, CCD 영상으로부터 대상자의 물체 영역을 분리해내는 과정이 있게 되는데, 블루 스크린 등 특별한 장치 없이 사람의 영역을 추출하는 것은 어려운 일 중 하나이므로, 종종 이 과정을 생략하고 영상 전체를 픽셀 시퀀스로만 보아 양자화하는 방법으로 건너뛰기도 한다. 일반 영상에서 사람의 영역을 추출할 때는 대개 사람이 없는 원래의 기본 배경 영상과 사람이 들어온 이후의 영상에서 픽셀 값의 차이를 구해 전신 영역을 추정한다. 이 방법은 계산시간이 빠르지만 사람이 여럿 등장하거나 빈번하게 오고 가는 복잡한 상황에서는 잦은 오류를 발생시킨다. 좀 더 시간이 걸리는 방법으로는 다이내믹 프로그래밍을 적용한 SNAKE(active contour models - M. Kass, A. Witkin, D. Terzopoulos, 1987) 알고리즘 등의 윤곽 추적법이 있다. 그러나, 이 경우에도 채광 조건 등이 달라지면 정확도가 낮아지므로 사람의 얼굴색이나 관절구조에 대한 지식을 바탕으로 연속 영상 간에 미세한 정보변화를 계속 추적하지 않으면 대상 영역을 놓치기 쉽다.

(2) 영상처리 및 특징추출

분리된 물체 영역에 대해 골조화(skeletonization)를 적용하여 전체적인 기하 구조를 패턴화하여

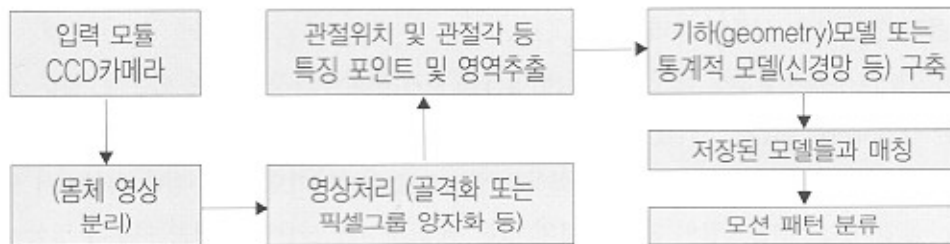
나, 전체 영상의 픽셀 그룹들에 대해 양자화(quantization)하는 등 영상처리 기법을 적용하여 특징을 추출한다. 전자의 경우는, 골격 관절이 이루는 모양(configuration)의 기하학적 관계를 특징벡터들로 계산할 수 있으며, 영상의 픽셀값 대신 FFT 등에 의한 주파수 공간에서 영상 신호를 분석하는 방법으로 특징을 추출한다. 그 밖에 비디오 영상의 각 화면을 여러 시간과 뷰에 대해 합한 누적 이미지(cumulative image)를 기준으로 특징을 추출하기도 한다(Bobick, 2001).

(3) 패턴 매칭

전체 혹은 추출된 이미지 영역에 대한 특징 벡터들을 신경망의 입력으로 하거나 벡터 양자화 방법, 유전자 알고리즘 등 다양한 방법들로 패턴화 할 수 있다. 이들을 미리 학습된 모델들과 비교하게 되는데, 많은 경우에 모션의 다양성으로 인해 정확한 매칭이 불가능하므로, 통계적, 확률적으로 해당 모션일 가능성을 산출하여 모션을 판정하게 된다.

분리된 단일 동작 인식의 응용 사례로는 5가지 정도의 모션으로 로봇 동작을 제어하는 것(Corradini, 2000), 테니스의 포핸드, 백핸드 스트로크 영상을 미리 주어진 키프레임과의 형상 매칭을 통해 인식한 것 등이 있다(Carlsson, 2002).

〈그림 4〉 모션 인식 시스템의 전반적인 구조



2) 연속 동작의 인식

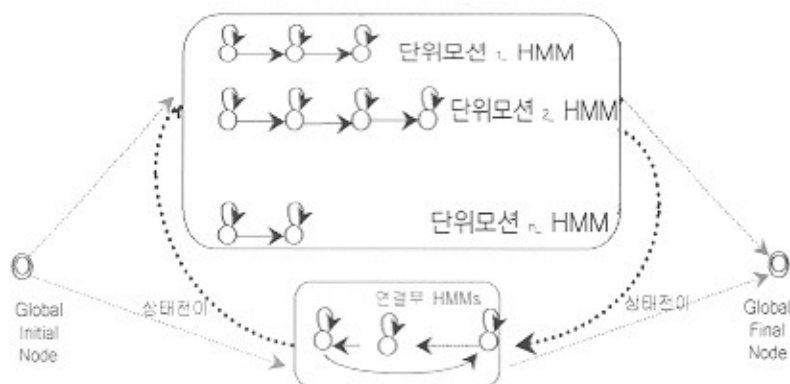
한 동작을 취한 후 차렷 자세나 기본 자세를 취하기 보다는 여러 가지의 의미있는 동작을 연속해서 행하는 경우가 보다 일상적이라고 볼 수 있다. 이렇게 여러 모션이 연속해서 행해지는 경우는, 단일(시작과 끝이 분리된) 모션을 인식하는 문제에 대해 여러 연속 동작들을 분리해내야 하는 문제를 추가로 안고 있는 것이다.

연속 동작을 인식하는 방법들은 대개 분리 동작 인식 방법을 사용하면서, 인식되어지는 패턴의 시간적 상태 변화를 분석하여 여러 모션 시퀀스 중 가장 가능성이 높은 시퀀스로 해석해내는 기술을

쓴다. 즉, 각각의 영상에서 자세(pose)를 인식하고 기본 동작을 모션음절(movelet)로 하여 단어(약속 어휘, codeword)를 찾아내듯이 인식하는 방법을 쓰거나(Feng, 2002), 상태 변환 표현과 분석을 위한 오토마타 방법들을 사용한다(Iwai, 1999, Nam, 1999). 특히, HMM(Hidden Markov Models)은 상태변화와 각 상태에서의 패턴 발생 확률을 표현할 수 있는 수학적 모델로서 모션 패턴의 학습이 간편하고 인식률이 높기 때문에 연속 동작의 인식에 많이 쓰인다.

HMM을 통한 인식의 경우, 하나의 동작을 표현하는 [그림5]와 같이 단위 HMM들이 각 단위 모션 종류별로 모델링된다. 이는 학습을 통해 모션의 특성을 반영할 수 있도록 상태 변화와 패턴 발생의 확률을 파라미터화 한 것이며, 연속 동작의 인식을 위해서는 단위 모션 인식용 HMM들이 서로 네트워크를 통해 연결되도록 하면 된다. 즉, 연속 모션에 대해서는 하나의 HMM 모델로부터 다른 HMM 모델로의 연결부 전이(transition)을 통한 확률 탐색을 계속함으로써 결과적으로 가장 가능성이 있는 HMM 모델 시퀀스를 찾아낸다.

〈그림 5〉 연속모션 인식을 위한 단위모션 HMM들의 네트워크 구조의 예



HMM의 각 상태는 패턴의 종류별 발생 확률을 표현하는데, 각 패턴 자체의 표현 복잡도가 높아지면 각 상태와 연결부 전이를 거치면서 패턴들의 가능한 조합이 배가 되므로, 패턴 인코딩을 어떻게 효율적으로 할 수 있는지에 대한 인식 단계의 시간적 효율에 큰 영향을 끼친다. 따라서, 고정된 카메라 뷰의 영상 시퀀스를 다루는 경우는 다차원의 특징 벡터를 HMM 입력값들(패턴값들)로 하지 않고, 영상을 벡터 양자화하여 스칼라 코드값을 입력으로 하는 경우가 많다. HMM과 같은 잘 정의된 확률 모델을 사용하는 것은 비교적 실시간의 안정적인 인식률을 얻을 수 있는 방법이나, 이와 같이 영상 인코딩에 한계가 주어지기 때문에 여전히 특징 벡터와 오토마타, 시간 지연 신경망 등을 사용한 방법들이 병행 연구된다.

고정 카메라 뷰의 연속 동작 인식에 관한 응용 사례로는, 회의 등 한정된 시나리오 상에서의 모션 인식이나, 악기 연주 또는 지휘 동작, 걷기 동작 등을 대상으로 한 모션 인식, 그리고, 비디오 데이터를 분석하여 스퀘시 동작의 히트(hit) 순간을 찾아내어 게임에 응용한 것이 있다(Ben-Arie, 2002).

문제는, 대부분의 연구가 한번에 분류할 동작들을 5~7개 내외로 한정하는 경향이 있으며, 이 경우에도 인식률이 66%~88% 정도에 그치므로 상업적 인터페이스에 채택하기에는 매우 부족한 경향이 있다는 것이다. 드물게 90% 대의 연속 동작 인식 결과들이 보고되고 있으나, 이들의 경우는 뚜렷이 구분되는 소수의 동작들로 모션 종류를 구성했거나 특정 응용의 맥락(Context)을 효과적으로 표현한 후, 이러한 맥락과 결합하여 상황 인식을 수행한 경우이다. 기존 연구 결과들을 볼 때, 아직까지 일반적인 모션들을 모두 수용할 수 있는 범용 인식 엔진은 없기 때문에, 상업적 응용을 위해서는 대상 모션과 응용에 대해 충분히 튜닝된 별도 인식 엔진을 개발해야 한다. 또한, 어떤 응용 맥락에서 모션의 종류를 어떻게 설정하는가가 인식률에 크게 영향을 끼침을 알 수 있다.

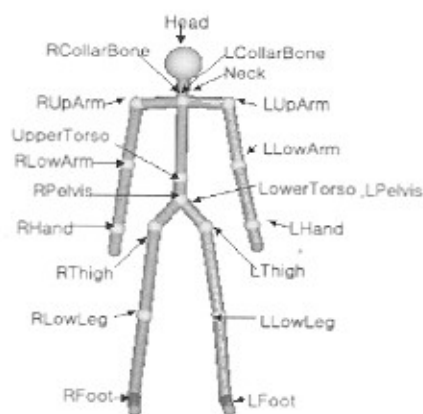
2. 멀티 카메라 뷰 또는 센서에 기반한 3차원 데이터의 인식

3차원의 데이터 입력 값을 얻을 수 있는 경우는 한 차원의 정보가 추가된다는 이득이 있기도 하지만, 추가된 한 차원으로 인해 오히려 복잡도와 모호성이 증가되기도 한다. 즉, 모션의 특성상 한정된 축을 따라 발생하는 정보만이 가치 있는 정보일 수 있는데, 이 경우 추가된 차원의 데이터로 인해 오히려 데이터 분석의 혼란을 가져올 수 있다. 따라서, 이러한 입력 데이터의 특성을 잘 이해하고 그에 따라 인식 대상 및 방법을 결정해야 한다.

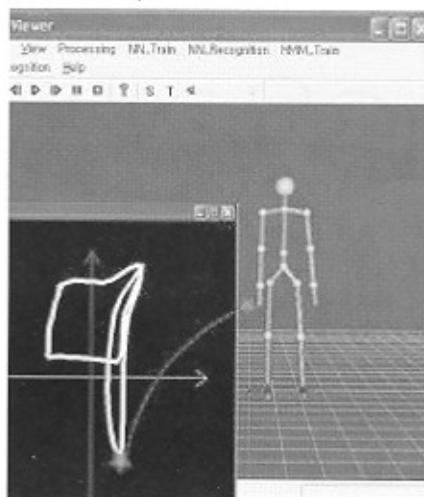
3차원의 데이터 값은 여러 카메라를 사용하거나 센서를 사용하여 얻을 수 있다. 카메라를 여러 개 사용하는 경우 2차원 영상을 그대로 사용하지 않고 캘리브레이션을 통해서 3차원의 대상체 위치값을 추정해 낼 수 있다. 한편, 센서를 이용하는 경우 역시 3차원 데이터를 얻을 수 있는데, 위치와 관절 각도값을 얻는 정확도에 따라 장치의 비용이 억대에 이르므로, 대중적 게임 인터페이스에 채택하기는 어렵고, 모션 캡처의 정확도가 다소 떨어지더라도 대중화된 PC 카메라를 이용한 3차원 모션 추적 인터페이스를 개발하여야 할 것이다.

멀티 카메라 뷰 또는 센서에 기반한 3차원 데이터는 전신 모션의 경우 대개 [그림6]과 같이 신체 각 주요 부위에 대한 위치값, 각도값을 포함하게 된다. 대부분의 신체 부위는 관절부이며, 인간의 모션을 해석하는 데 있어서 각 관절이 이루는 회전각들이 중요한 근거가 되기 때문이다. 머리(Head)와 같은 경우는 관절은 아니지만 시선방향, 고개 돌림을 알아내기 위해 중요한 부분인데, 카메라 기반의 센싱으로는 어렵기 때문에 카메라로 모션을 추적하는 경우에도 머리 부분에는 별도로 6자유도의

〈그림 6〉 20개의 신체 주요 부위



〈그림 7〉 손목 관절의 위치 변화(궤적)

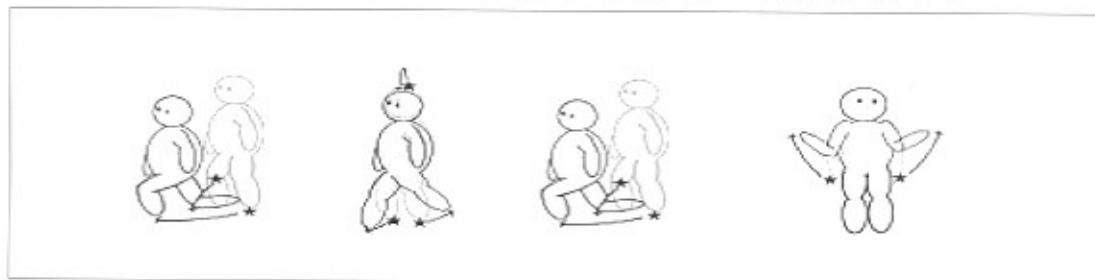


센서를 부착하여 회전 여부를 알아내는 경우가 많다. 나머지의 관절부들은 대개 3자유도의 위치 값, 3자유도의 각도 값으로 이루어진 6자유도 데이터들로 표현되나, 어깨와 같은 특수 관절 등은 별도의 축으로 표현되기도 한다. 또한, 각 관절마다 특정 축의 회전각이 일정 범위를 벗어날 수 없는 경우 해당 자유도의 값에 제약범위가 부과된다.

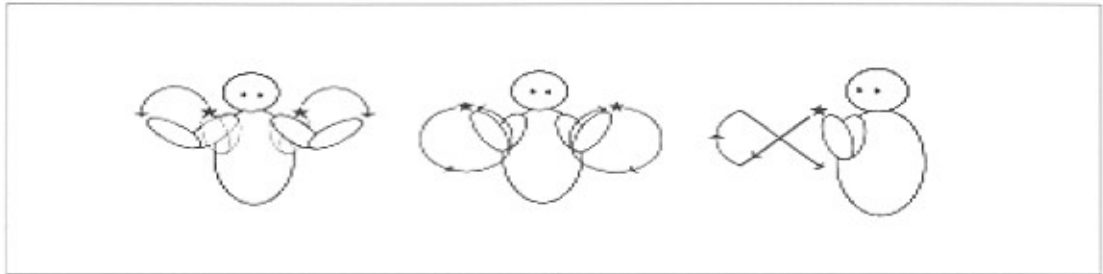
3차원의 전신 동작은 위와 같이 파라미터의 수가 너무 많고(대략 관절수 X 6자유도 X 시간구간), 동작의 유형이나 발생 부위에 따라 모션 패턴의 해석 방법이 전혀 달라질 수 있다. 예컨대, [그림8]의 예시와 같은 전신동작들은 시간 구간 내에서 특정 자세가 발생하였는가에 따라 인식되는 모션들이다. 따라서, 자세 인식 또는 특정 자세 발생 여부를 검출을 연속 적용하여 인식하게 된다. 이 경우는 형상 정합(shape matching)을 위한 신경망이나 특징 비교 등을 매 순간 적용하여, 임계치 이상으로 부합되는 자세들이 발생했을 때 이들간의 선후 관계를 분석하여 모션을 판정하게 된다.

반면, [그림9]와 같이 시간 구간 내의 특정 순간의 자세로는 모션이 구분되지 않고 연속되는 궤적

〈그림 8〉 특정 자세(pose) 발생 여부를 관찰에 의해 인식되는 동작들의 예



〈그림 9〉 연속 궤적(trajjectory) 분석에 의해 인식되는 동작들의 예



상의 흐름을 통해 패턴을 알 수 있는 모션들이 있다. 팔을 움직여서 이모티콘이나 수신호를 보내는 경우가 대표적인 경우이다. 이런 모션들에 대해서는 자세 분석을 통한 인식이 어려우므로, 앞 절에서 언급한 HMM 등 시간 시퀀스 해석 방법들을 통해 궤적 패턴의 발생 확률을 계산하게 된다.

그런데, HMM과 같이 패턴 심볼을 대상으로 발생 확률을 구하는 방식에서, 3차원 데이터 자체를 HMM의 입력으로 할 경우 계산 복잡도가 너무 커지기 때문에 실시간 계산에 장애 요소가 될 수 있다. 따라서, 일반적으로 다차원 데이터를 스칼라 값에 매핑하여 심볼화하는 인코딩 단계가 선행된다. 그러나, 전신 동작의 경우는 20개 관절 각각에 대해 6자유도 정도의 데이터가 매 시간 발생하는 것이므로, 이들을 한데 묶어서 심볼화 하기는 어렵다. 이러한 전신 동작의 경우는 주성분 분석(principal component analysis) 등을 거쳐서 모션들의 주요 움직임 파트를 추출하고, 부분별로 발생하는 모션 패턴들을 병렬적으로 인식한 다음에 전신에 대해 동시 발생된 모션 패턴들의 조합을 분석하는 게 한 가지 방법이다. 또한, 최근에는 HMM 대신 특징 분석과 신경망, 베이시언 네트워크 등을 융합하여 단계별로 적용하거나 시간 오토마타에 따라 분석하는 방법들이 다수 등장하고 있다(Emering, 1997, Nam, 2003).

그러나, 일부 제한된 실험 환경에서의 연구들이 높은 인식률을 보고한다고 해도 일반적으로 적용하기에는 정확도가 매우 떨어지는 경향이 있어서 실용성에 의문이 있다. 예컨대, 90%대의 인식률을 보고하는 연구들은 대개 정확도가 높은 센서를 사용하고 크레인 수신호나 발레 동작 등 대상 동작의 한계와 컨텍스트가 명확하게 제한되는 것이 보통이다.

따라서, 실제 응용에 적용함을 목적으로 하는 모션 인식 연구에서는 응용 맥락(context)의 적극적인 도입과 모델링을 패턴 인식 결과에 커플링하여 분석 성공률을 높이는 추세를 보인다(Ayers, 1998, Wren, 2000, Yu, 2002). 또한, 일반인을 대상으로 한 범용 모션 인식은 아직 그 사례가 없을 만큼 사람 간의 동작 차이가 크기 때문에, 모션의 종류를 인식하기보다 동일 모션에 대해 특징을 분석하여 사람을 구분함을 목적으로 하는 경우도 있다.

Ⅲ. 게임에서의 모션 인식 기술 적용 방안

앞서 언급한 모션 인식 기술의 여러 연구 결과는 실험실 연구가 아닌 상업적 응용에 모션 인식이 활용되기까지 많은 제약들이 있음을 시사한다. 따라서, 게임 인터페이스에 모션 인식을 적용하기 위해서는 다음과 같은 사항들을 고려할 필요가 있다.

- 인식의 관점에서 볼 때 모션은 시간 구간 내에서 특정 자세들의 검출을 통한 인식이 가능한 것과, 궤적 패턴을 분석하여 인식 가능한 것의 2가지로 구분된다. 이 두 종류는 서로 다른 종류의 인식 엔진을 채택하는 것이 일반적이므로, 통합된 하나의 방법으로 인식하고자 한다면 모션 인식 선정시 해당 카테고리들을 고려하여야 한다.
- 단위 모션의 시작과 끝이 명확한 경우(단일 분리 동작) --> 연속 동작이면서 시간 구간 내 자세 전이의 추적이 가능한 모션(연속 자세 전이 동작) --> 특정 자세 시퀀스로 분석해내기 어려운 연속 궤적 모션 순으로 인식의 난이도가 증가함을 고려한다.
- 일반화된 모션 인식이 용이하지 않으므로, 인식 성공률을 높이기 위해서는 모션 간의 차이가 명확한 소수(7~8개 이내)의 동작들에 대해 충분한 모집단의 인원에 대한 수백 샘플의 훈련 데이터들을 얻어야 한다. 그것이 어려울 경우 혼동 매트릭스의 분석을 통해 서로 혼동되기 쉬운 모션들을 배제할 필요가 있다.
- 응용 맥락을 모델링하고 각 맥락에서 발생 가능한 모션 패턴에 커핑링하여 인식을 향상을 도모할 필요가 있다.
- 일반 사용자를 대상으로 한 범용 모션 인식이 목적인 경우, 개인별 데이터로 모션 모델을 따로 구축하고, 이들간의 차이 분석에 의해 개인별 차이가 크게 발생하지 않는 모션 종류를 응용에 채택하는 것이 좋다.
- 일반 사용자가 아닌 특정인에게 커스터마이징 되는 인터페이스의 경우 인식의 정확도를 향상시킬 수 있지만, 연속궤적 인식과 같은 경우는 개인의 패턴을 충분히 학습시키기 위해서는 백 여 샘플의 데이터가 필요하고, 사용자에게 소수의 샘플 모션만을 요구하는 경우는 파라미터 보정에 한계가 있음을 주지할 필요가 있다.
- 현재까지의 기술 수준으로는 3차원 궤적에 의한 모션 패턴에는 데이터의 다양성 및 연결 모션의 예측 어려움으로 인한 문제가 존재하기 때문에, 범용 사용자 대상 시스템인 경우 시간 구간 내에서의 타겟 자세 검출에 의한 모션 인식을 게임 시나리오에 접목하는 방법이 가장 인식 성공률이 높은 방법이다. 예컨대, 7~8개 내외의 특정 자세들로 구분 가능한 전투, 액션 게임 등에 모션 인식을 채택하는 것은 비교적 현실화 가능한 모션 인터페이스 형태라 하겠다.

IV. 결론

모션 인터페이스를 구축하기 전에, 모션 추적과 실루엣 추출로 충분한 것인지 모션의 인식이 필요한 것인지 판단해야 하며, 모션 인식의 경우 어느 레벨의 인식이 필요한지 판단할 필요가 있다. 실험실 환경에서의 모션 인식과 높은 인식을 취득은 어려운 일이 아니지만, 상업적 응용 시스템에 채택되기에는 기술적으로 풀어야 할 문제들이 다수 존재한다. 통계적 방법들이나 확률적 모델링, 시간에 따른 상태 변화 모델링 등이 데이터의 노이즈나 다양성을 일부 흡수할 수 있지만, 인간의 모션의 다양성의 정도가 매우 크기 때문에 패턴 모델링의 한계 내에서 해결하려는 시도는 무리이다. 따라서, 게임 인터페이스에 채택시 모션 발생의 관점에서 응용의 맥락을 정의하고 이러한 부가 정보를 바탕으로 모션 해석의 범위를 최대한 줄이는 것이 사용 가능한 모션 인터페이스를 만드는 방법이라 하겠다.

참고문헌

- Mandala System, <http://www.vividgroup.com/>, Vivid Group.
- EyeToy, Playstation Games, Sony Computer Entertainment Inc.
- Ali, A., Aggarwal, J. K.(2001), "Segmentation and Recognition of Continuous Human Activity", IEEE Detection and Recognition of Events in Video.
- Bobick, A., Davis, J. W.(2001), "The Recognition of Human Movement Using Temporal Templates", IEEE PAMI.
- Ben-Arie, J., Wang, Z., Pandit, P., Rajaram, S.(2002), "Human Activity Recognition Using Multidimensional Indexing", IEEE PAMI.
- Kass, M., Witkin, A., Terzopoulos, D.(1987), "Snakes: Active Contour Models", Int. Journal of Computer Vision, p.321-331.
- Corradini, A., Gross, H. M.(2000), "Camera-based Gesture Recognition for Robot Control", IJCNN.
- Carlsson, S., Sullivan, J.(2002), "Recognition of Gestures using Morphological Features of Networks Made of Gesture Motion Image and Word Sequences", IEEE WMECV.

- Feng, X., Perona, P.(2002), "Human Action Recognition By Sequence of Movelet Codewords", IEEE 3DPVT.
- Iwai, Y., Shimizu, H., Yachida, M.(1999), "Real-time Context-based Gesture Recognition using HMM and Automaton", IEEE WRATFG.
- Nam, Y., Lee-Kwang, H., Wohn, K.(1999), "Modeling and Recognition of Hand Gestures using Colored Petri Nets", IEEE SMC.
- Ayers, D., Shah, M.(1998), "Recognizing Human Actions in a Static Room", IEEE WACV.
- Wren, C., Clarkson, B. P., Pentland, A. P.(2000), "Understanding Purposeful Human Motion", IEEE ICAFGR.
- Yu, C., Ballard, D. H.(2002), "Learning to Recognize Human Action Sequences", IEEE ICDL.
- Emering, L., Boulic, R., Thalmann, D.(1997), "Live Participant's Action Recognition for Virtual Reality Interactions, IEEE CG&A.
- Nam, Y.(2003), "안정적인 모션복원을 위한 동작 인식 기술 개발에 관한 연구", 한국전자통신연구원 제출 최종연구보고서.
- Yu, C., Ballard, D. H.(2002), "Learning to Recognize Human Action Sequences", IEEE ICDL.
- Emering, L., Boulic, R., Thalmann, D.(1997), "Live Participant's Action Recognition for Virtual Reality Interactions, IEEE CG&A.
- Nam, Y.(2003), "안정적인 모션복원을 위한 동작 인식 기술 개발에 관한 연구", 한국전자통신연구원 제출 최종연구보고서.

Motion Recognition Techniques for Game Interface

Yang-Hee Nam

Division of Digital Media, Ewha Womans University

Human motion interpretation is one of important future HCI techniques in respect that it might reduce the need of artificial interfaces. However, we can still hardly find such game interfaces recognizing general gamer's motion because there are many technical difficulties. The main difficulty comes from the fact that human motions are conducted in 3D space and thus the motion patterns have 3D spatial data complexity added by temporal variation. This paper introduces how to deal such difficulty in the state of the art techniques. It is pointed that the recognition method and its accuracy greatly vary according to the target motion set and the types of constituting motions. Paying regard to the technical aspects, this paper finally suggests what to be considered for motion based game interface design.