

가상 공간에서 에이전트 생성을 위한 실시간 마커프리 모션캡처 시스템

김성은, 이란희, 박창준, 이인호
한국전자통신연구원 가상현실연구센터 모션정보연구팀
전화: 042-860-1057

Real-time Marker-free Motion Capture System to Create an Agent in the Virtual Space

Sung-Eun Kim, Ran-Hee Lee, Chang-Joon Park, and In-Ho Lee
Motion Information Research Team, VR center, ETRI
E-mail : sek62958@etri.re.kr

Abstract

We described a real-time 3D computer vision system called MIMIC(Motion Interface & Motion Information Capture system) that can capture and save motion of an actor. This system analyzes input images from vision sensors and searches feature information like a head, hands, and feet. Moreover, this estimates intermediated joints as an elbow and knee using feature information and makes 3D human model having 20 joints. This virtual human model mimics the motion of an actor in real-time. Therefore this system can realize the movement of an actor unaffectedly because of making intermediated joint for complete human body contrary to other marker-free motion capture system.

I. 서론

모션캡처 기술은 신체와 같은 복잡한 관절체의 움직임을 시간에 따른 공간상의 주요 위치 데이터로 기록하는 기술이다. 행동을 취할 동작자의 몸에 많은 수의 마커를 부착한 후, 동작자를 둘러싼 여러 대의 카메라를 이용해 마커들의 움직임을 촬영한다. 마커의 움직임을 잘 분석할 경우, 동작자의 각 관절의 움직임에 대한 정보를 얻을 수 있고, 이 정보를 가상 개릭터에게 적용하면 실제 동작자와 같이 자연스럽게 움직이는 결과를 얻을 수 있다. 그러나, 이러한 모션캡처 기술은 동작의 복잡성이나 정밀도에 대해서는 아주 높은 성능을 발휘하지만, 높은 컴퓨팅 파워, 불편한 장비의 부재에 따른 동작자의 행동에 대한 제약, 그리고 고가의 장비에 따른 재정적인 어려움 등의 문제점이 있어서 쉽게 사용할 수 없는 단점이 있다. 본 논문에서는 이러한 문제점으로 인한 동작자의 불편함이 없이, 동작을 획득할 수 있는 마커프리 모션캡처 기술을 소개한다.

초기의 마커프리 분야 연구는 신체 중에서 의사 전달 정보를 비교적 많이 가진 얼굴과 손의 해석에 대한 것이었다. 컴퓨팅 파워의 급속한 발전으로 영상을 처리하는 속도가 증가하자 신체의 특정 부분만을 해석하는 쪽에서 다양한 부분을 동시에 실시간으로 해석하는 쪽으로 연구 방향이 확대되어지고 신체의 상반신 또는 전체에 대해 해석하게 되었다. 최근에는 많은 연구자들이 다양한 연구 결과를 발표하고 있다[1,2,3,4].

3차원 공간에서의 동작자의 동작을 얻기 위한 마커프리

모션캡처 시스템인 미믹(MIMIC: Motion Interface & Motion Information Capture system)은 동작자의 동작에 불편함을 주는 별도의 마커나 센서와 같은 장비 대신 동작자의 좌, 우측 전방에 설치되어있는 동기화 된 두 대의 CCD카메라를 이용해 동작자의 동작 영상을 획득한다. 그리고, 이들 영상으로부터 몸의 중심과 머리, 두 손, 두 발의 특징을 추출한다. 추출된 특징점들은 좌, 우 영상의 정합과 특징 추적 알고리즘에 의해 3차원 좌표값을 가지는 형태로 만들어지고, 이 값들에 역운동학(Inverse Kinematics) 이론을 적용하여 중간 관절의 위치를 추정함으로써 어깨와 팔꿈치, 무릎 등 전체 20개의 관절점을 가진 인체 모델을 실시간으로 생성한다. 그래서, 기존의 마커프리 모션캡처 시스템과 달리 완벽한 인체를 구성하기위한 중간 관절까지 생성함으로써 동작자의 동작을 더욱 자연스럽게 구현할 수 있다.

II. 미믹(MIMIC)

2.1 시스템 구성

미믹은 그림 1와 같이 크게 서버 부분과 2개의 클라이언트 부분으로 나누어진다. 각 클라이언트는 하나의 칼라 CCD카메라를 가지고 있으며, 카메라로부터 입력된 영상 정보를 해석한 후, 특징 추출 결과를 서버로 전송한다. 카메라는 동작자의 전방 좌, 우측에 설치되어 있다. 서버는 각 클라이언트로부터 특징 정보를 받아 3차원 정합 및 추적을 하고, 인체 모델을 구성하기위한 중간 관절을 구한다. 그리고, 그 결과를 화면에 출력한다.

2.2 신체 특징점 추출

영상 취득 및 특징 추출 단계에서는 동작자의 좌, 우 전방에 위치한 2대의 컬러 CCD 카메라를 이용하여 동시 동기 방법으로 동작자 움직임에 대한 영상을 취득한다. 배경 영상과 동작자 초기 자세 영상의 차 영상을 컬러 영역별로 분할하여 동작자의 몸통과 인체 끝 부분의 색 범위와 영역 속성을 추출한다. 그리고, 연속 입력되는 동작 영상에 대해 색 범위 안에 있는 후보 화소들을 선별한 후, 고속 연속부 길이(run-length) 레이블링을 통해 그룹화하여 영역의 크기와 페리미터(perimeter) 특징, 영역의 최소 외접

사각형에 대한 화소 수 비율을 추출하여 그림 2에 보이는 것처럼 동작자의 손, 발, 머리와 같은 특징을 실시간으로 검출한다.

이와 같이 적은 계산량을 가지는 선별 작업을 넓은 임계 조건을 부가하여 영상 전역에 대해 먼저 수행한 후, 선별된 소수의 후보 화소들에 대해 영역 특징을 정밀하게 추출하여 신체 끝 부분과 몸통을 검출함으로써 동작을 정밀하게 획득하기 위한 인체 정보를 복잡한 배경에서도 실시간으로 검출할 수 있도록 한다.

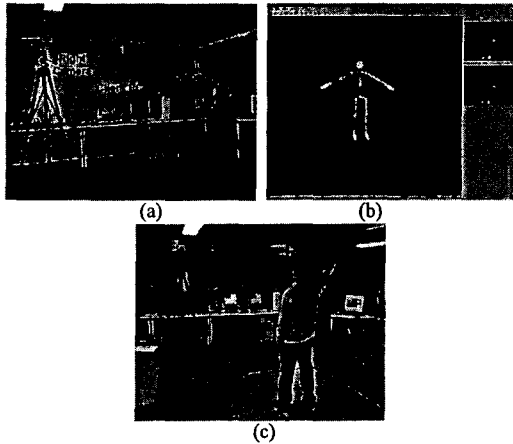


그림 1. (a) 미믹의 클라이언트 부분, (b) 서버측 실행 화면, (c) 동작자의 동작 모습

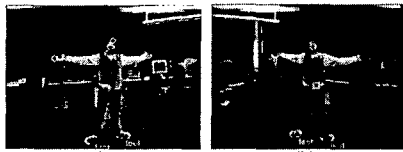


그림 2. 신체 특징점 추출

2.3 예측값과 2차원 특징 추출 데이터간의 매칭

두 개의 2차원 입력 영상에 각각 맺힌 머리, 양 손, 양 발에 해당하는 3차원 좌표는 좌측과 우측 카메라의 보정 결과로서 얻어진 두 개의 사영 행렬(perspective projection matrix)을 이용해 구할 수 있다. 이 사영 행렬은 카메라 자체의 내부 변수와 카메라와 카메라간의 이동과 회전에 대한 값을 담고있다[7]. 두 카메라의 입력 영상에 포함된 각 특징들은 3차원 공간의 한 점에서 만나므로, 각 카메라의 원점과 특징의 2차원 영상 좌표를 잇는 두 개의 3차원 직선이 교차하는 위치를 구하면, 그 점이 3차원 공간에서 특징의 위치가 된다. 그림 3에서, 두 개의 3차원 직선간의 가장 가까운 거리를 지나는 점 P , Q 의 중점 M 이 공간상의 특징의 위치가 된다.

동작이 이루어지는 동안 특징들은 공간을 자유롭게 움직이기 때문에, 양쪽 영상에 각각 맺힌 특징들이 서로 올바르게 짝지어지기 위해서는 각 특징들의 움직임을 추적할 필요가 있다. 이것을 위해 칼만 필터(Kalman filter)를 사용한다[5,6]. 미믹의 입력은 2차원 영상이지만, 내부 연산과 출력은 3차원 데이터이기 때문에, 3차원 칼만 필터를 사용한다. 현재 시간에서 얻어진 머리와 양 손, 양 발에 대한 3차원 좌표로부터, 다음 시간에서의 3차원

예측값을 구한다. 이 값을 2차원 영상에 맺힌 특징에 적용하기 위해, 이 예측값을 사영 행렬을 통해 2차원 평면으로 역투영하면 2차원 예측값을 얻을 수 있다.

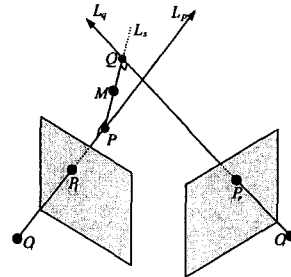


그림 3. 3차원 공간상의 정합점.

특징 추출 알고리즘으로부터 얻어진 2차원 특징의 데이터는 잡음이 포함되어 있다고 간주하고, 앞서 얻어진 2차원 예측값과 비교해 각 특징의 최종 위치값을 구한다. 하나의 예측값 주위에는 이 값에 매칭될 수 있는 특징들이 여러 개 존재할 수 있고, 반대로, 하나의 특징에 여러 개의 예측값이 매칭될 수도 있다. 미믹에서는 각 특징의 현재 위치와 예측 위치간의 3차원 거리를 계산함으로써 좌, 우 영상에서 얻은 특징들을 연결한다. 각 특징은 모든 시간의 입력 영상에 대해 겹쳐지거나 사라지는 현상 없이 2차원 영상 평면 위에 존재한다고 가정한다. 즉, 머리, 두 손, 두 발을 제외한 동작자의 모든 신체는 투명하다고 가정하면, 하나의 특징이 몸 뒤에 숨겨지더라도 카메라에서는 보이게 된다. 각 영상의 마커 (영상으로부터 얻은, 아직 이름이 결정되지않는 특징 데이터)를 모두 한번씩 조합하여 3차원 정합을 하고, 이 값과 한 특징의 예측 위치와의 3차원 거리를 계산한다. 그 결과, 그 특징에 대한 가장 가까운 거리부터 가장 먼 거리를 가지는 좌, 우 영상의 마커들의 조합을 얻을 수 있다. 가장 가까운 거리를 가지는 좌, 우의 마커들이 그 특징과 매칭되는 것들이지만, 여러 개의 특징이 동일한 마커 데이터를 가질 수도 있고, 가장 가까운 거리가 아주 멀리 떨어진 거리값일 수도 있다.

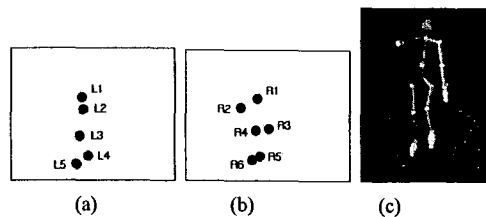


그림 4. 영상으로부터 추출된 특징의 매칭: (a) 왼쪽 카메라 영상, (b) 오른쪽 카메라 영상, (c) 재구성된 인체 모델

각 특징에 대해 가장 가까운 거리를 가지는 마커 데이터가 다른 특징의 것과 중복이 될 경우에는, 마커 데이터와의 거리가 더 짧은 특징을 우선적으로 선택해서 중복되는 현상을 없앤다. 반면, 각 특징에 대해 예측값간의 거리가 매우 가깝고 그 특징의 마커 데이터가 서로 중복되었을 경우에는, 특징들은 겹쳐져있다고 간주하고 중복을 허용한다. 가장 짧은 거리가 실험적으로 15cm이상일 경우에는, 그 특징과 어울리는 마커 데이터는 없다고 간주하고(영상에서 특징이 사라졌다고 간주), 특징의 예측값을 현재의 위치값으로 간주한다. 결과적으로, 미믹은 입력 영상에서 마커의 정보가 추출되지 않았다

하더라도 예측값을 이용해 마커의 정보를 생성하게 된다. 그림 4는 동작자가 앞으로 걷는 동작을 하는 과정 중의 한 프레임에 대한 매칭의 예를 보여준다. 그림 4(a)는 왼쪽 카메라 영상으로부터, (b)는 오른쪽 카메라 영상으로부터 추출된 마커 데이터의 결과이다. 머리의 예측값과 가장 가까운 거리를 가지는 마커의 조합은 (L1, R1)이 되었다. 또한, 루트는 (L3, R4), 왼발은 (L4, R5), 그리고, 오른발은 (L5, R6)의 조합으로 나타났다. 손의 경우, 오른손의 예측값과 가장 가까운 거리를 가지는 마커의 조합은 (L2, R2)로 예측값과 10.78Cm의 거리차를 가졌다. 왼쪽 영상은 오른쪽 영상보다 추출된 마커의 개수가 하나 작다. 왼손의 경우, 가장 가까운 거리를 가지는 마커의 조합은 오른손과 같이 (L2, R2)로 나타났지만, 예측값과의 거리는 53.92Cm로 실험적으로 설정한 임계치(15Cm)보다 크기때문에 예측값을 현재의 위치값으로 그대로 간주하였다. 이것은 실제 동작자의 왼손이 왼쪽 영상에서 사라져버린 결과에 대한 처리이다. 그림 4(c)는 재구성된 인체모델의 결과를 보여준다.

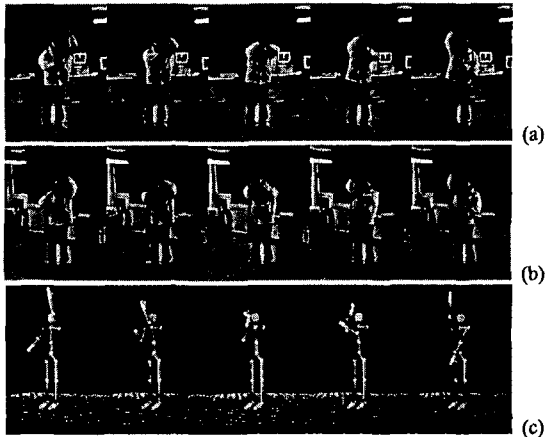


그림 5. 두 손과 머리 영역이 겹치는 경우에 대한 처리 예: (a) 왼쪽 영상, (b) 오른쪽 영상, (c) 3차원 신체 모델로 재구성된 결과

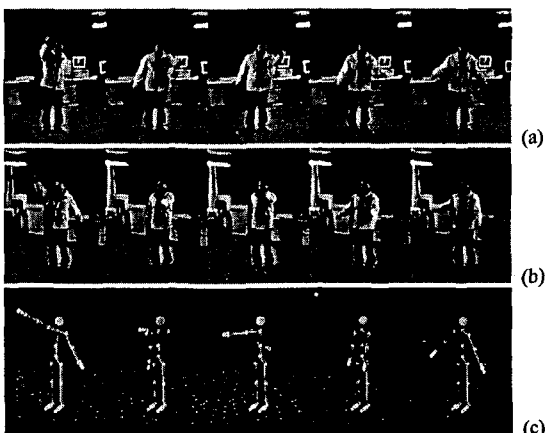


그림 6. 카메라 영상에서 특징이 사라졌을 경우에 대한 처리 예: (a) 왼쪽 영상, (b) 오른쪽 영상, (c) 3차원 신체 모델로 재구성된 결과.

그림 5는 두 손과 머리 영역이 겹쳐지는 과정을 포함한

연속 동작에 대한 것이다. 그림 5(a)는 좌측 카메라로부터, (b)는 우측 카메라로부터 입력된 영상이다. 그리고, (c)는 인체 모델의 동작 결과를 보여준다. 여기서, 동작 중 두 영역 이상이 겹쳐진 후, 다시 분리되더라도 예측값을 이용해 올바르게 추적하고 있음을 확인할 수 있다.

그림 6은 마커가 사라졌을 때의 처리 과정을 나타낸다. 첫번째 신체 모델에서 왼쪽 카메라 영상에서 왼손이 추출되지 못하였고, 세번째 신체 모델에서는 오른쪽 카메라 영상에서 동작자의 몸 뒤로 숨어버린 오른손이 추출되지 못하였지만 이런 경우에도 제대로 추적하고 있음을 확인할 수 있다.

2.4 중간 관절의 위치 추정

머리, 손, 발, 루트의 3차원 위치 정보와 각 관절과 관절을 연결하는 세그먼트의 길이 정보는 팔꿈치와 무릎, 목과 같은 중간 관절의 위치를 추정하기 위한 기준 정보가 된다. 이 정보를 역운동학 이론에 적용하면 관절의 접힌 각을 구할 수 있고, 이 각 정보를 이용해 3차원 위치를 계산한다[8,9,10,11].

사람의 신체 구조를 이루는 관절은 뼈로 연결되어 있으며, 각 관절은 계층적 트리 구조를 이루는 상대 좌표계로 표현할 수 있는데, 이는 각 관절의 자세를 상위 부모 관절의 좌표계에 대한 상대적인 값으로 표시하는 것이다. 인체 구조가 계층적 트리 구조로 되어 있기 때문에, 상위 관절의 회전 각도와 뼈(세그먼트)의 길이를 안다면 하위 관절의 위치를 추정할 수 있다. 미믹에서는 3차원 정합 및 추적 알고리즘으로부터 얻어진 특징점(인체 모델의 끝점(end-effectors))의 3차원 위치 정보와 세그먼트 길이 정보를 시간별로 입력하고, 이 끝점의 축과 활동 반경을 따라서 중간 관절의 움직임은 구한다. 현재 구현된 역운동학 알고리즘은 2축 링크(2-link) 관절만을 대상으로 하며, 다음과 같이 3단계의 과정으로 이루어진다.

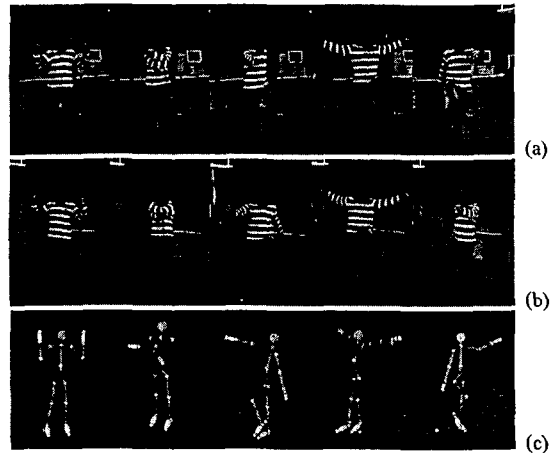


그림 7. 자연스러운 자세를 위한 중간 관절의 추정: (a) 왼쪽 영상, (b) 오른쪽 영상, (c) 3차원 신체 모델로 재구성된 결과

만약, 팔의 움직임에 대한 팔꿈치의 위치 추정에 대해 생각한다면, 첫번째로, 손으로 추정된 끝점의 상위 관절이 되는 어깨 관절의 위치를 추정한다. 추정된 상위 관절의 위치 정보와 기준 정보를 코사인 제 2법칙에 적용하여 중간 관절의 접힌 각을 계산한다. 두번째로, 계산된 접힌

각과 어깨, 팔꿈치, 손목이 이루는 평면의 법선 벡터를 이용하여 중간 관절의 3차원 위치를 추정한다. 즉, 영상으로부터 획득된 기준 정보를 이용하여 구하여진 회전 각도를 연속적으로 행렬 연산을 함으로써 중간 관절의 위치를 추정할 수 있다. 마지막으로, 중간 관절의 위치는 이전 프레임에서의 위치와의 상관 관계를 고려하여 결정된다. 또한 모든 관절의 움직임은 해부학적으로 허용이 가능한 범위 내에서 그 크기가 결정되어진다. 결정된 중간 관절의 위치로 움직이기 위한 회전각을 구면 좌표계를 이용하여 계산하며, 계산된 회전 각도를 계층적 트리 구조를 이루는 상대 좌표계로 표현하기 위한 회전 각도로 다시 계산된다. 이렇게 계산된 회전 각도를 가상 캐릭터에 적용함으로써 동작자와 유사한 움직임을 생성하게 된다. 다양한 동작에 따른 중간 관절의 추정 결과는 그림 7과 같다.

III. 실험 결과

미믹에 사용된 컴퓨터 환경은 서버와 클라이언트 부분 모두 Intel 펜티엄IV-2GHz, OS는 Microsoft사의 Windows 2000을 사용하였고, Visual C++ 6.0과 OpenGL을 이용하여 프로그래밍 되었다. 영상의 입력을 위해 사용된 CCD카메라는 Jai사의 CV-M70이고, 영상의 처리는 Matrox사의 Meteor II/Multi-Channel 프레임 그레버를 이용하였다. 서버와 클라이언트는 네트워크로 연결되어 TCP/IP통신을 한다. 결과적으로, 미믹은 초당 평균 30프레임의 처리 시간을 가지며, 조명은 사무실 천정에 설치된 일반적인 60Hz 형광등을 사용하였다.

카메라 간의 거리는 300cm, 카메라의 높이는 210cm, 동작자는 카메라로부터 약 450cm정도 떨어진 거리에서 동작을 하였다. 카메라의 보정을 위해 Tsai모형을 이용하였고, 두 카메라에 의한 3차원 정합 결과 평균 1.02cm의 오차를 가졌다. 실험은 서로간에 키가 상이한 동작자들의 다양한 동작을 통해 이루어졌다. 실험에 앞서, 동작자들의 뼈 길이를 먼저 측정 한 후, 이 값을 3차원 인체 모델 구성에 사용하였다. 동작자가 카메라 방향으로부터 회전을 하는 동작과 두 다리를 꼬아서 서 있는 자세 및 거꾸로 서는 자세와 같은 정상적인 자세가 아닌 동작은 제한을 두었다. 또한, 배경은 동작 중에는 고정되어 있다는 가정을 두었고, 동작자는 자신의 피부와 차이 나는 색의 옷을 착용하도록 하였다.

좌, 우 영상에서 획득한 특징을 3차원 정합하고, 이 값을 기준으로 중간 관절을 생성한 결과 그림 7에서 보는 것과 같이 자연스럽게 나타나고 있음을 확인할 수 있었다. 또한, 마커가 사라지거나 겹쳐질 경우에도 시스템은 특징을 제대로 추적하고 있음을 확인할 수 있었다.

IV. 결론

미믹은 동작자의 좌, 우측 전방에 설치된 두 대의 CCD카메라를 이용해 동작자의 동작 영상을 획득한다. 그리고, 이들로부터 몸의 중심과 머리, 두 손, 두 발의 특징을 추출한다. 추출된 특징점들은 좌, 우 영상의 정합과 특징 추적 알고리즘에 의해 3차원 좌표를 가지는 형태로 만들어지고, 이 값들에 역운동학 이론을 적용하여 중간 관절의 위치를 생성한다. 생성된 데이터는 범용적으로 사용되는 모션캡처 데이터 포맷(예를 들면, htr포맷)으로 쉽게 전환이 가능하도록 저장이 된다.

본 연구실에서 개발한 미믹은 여러 동작자에 대해

실험해 본 결과, 복잡한 배경 하에서도 동작자의 신체 끝 부분에 해당하는 머리, 두 손, 두 발 및 몸의 중심의 위치를 실시간으로 검출할 수 있었으며, 역운동학 이론을 기반으로 한 중간 관절의 위치 추정 결과로 모션 데이터를 생성할 수 있었다.

아직, 미믹은 특징의 추출 및 겹침 발생시에 대한 처리, 3차원 인체 모델 생성에 대해 개선할 점이 많이 남아 있다. 현재, 조명에 민감하지 않는 환경에서도 동작자의 복잡한 움직임을 보다 정밀하게 추정할 수 있고, 더욱 안정되고 자연스럽게 따라 움직일 수 있도록 시스템을 개선 중이다. 또한, 개발 중인 동작 인식 모듈과의 통합을 통해 동작자의 동작을 인식하고 가상 공간에서의 또 다른 객체(avatar)와의 직접적인 상호 작용이 가능하도록 할 예정이다.

참고문헌

- [1] K.H. Jo, Y. Kuno, and Y. Shirai, "Manipulative hand gesture recognition using task knowledge for human computer interaction", *Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp.468-473, 1998.
- [2] Ismail Haritaoglu, David Harwood, and Larry S. Davis, "W4: Real-Time Surveillance of People and Their Activities", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, August 2000
- [3] Pattie Maes, Bruce Blumberg, Trevor Darrell, and Alex Pentland, "The Alive System: Wireless, Full-body Interaction with Autonomous Agents", *ACM Multimedia Systems*, 5:105-112, 1997
- [4] S.E. Kim, I.H. Lee, and K.H. Jo, "Interaction with Agents in the Virtual Space by Recognition of Face Direction and Hand Gestures", *Proc. of the IASTED, Visualization, Image Processing*, pp.1-6, September 3-5, 2001, Marbella, Spain
- [5] Greg Welch and Gray Bishop, "An Introduction to the Kalman Filter", *TR 95-041*, University of North Carolina at Chapel Hill, February 8, 2001
- [6] C.H. Sul, S.K. Jung, and K.Y. Wahn, "Synthesis of Human Motion using Kalman Filter", *Proceedings of CapTech'98, First International Workshop on Modeling and Motion Capture Technologies for Virtual Environments*, 26-28 November 1998, Geneva, Switzerland.
- [7] Emanuele Trucco and Alessandro Verri, "Introductory Techniques for 3-D Computer Vision", *Prentice Hall*, New Jersey, 1998
- [8] N.I.Badler, M.Hollick and J.Granieri, "Real-time control of a virtual human using minimal sensors.", *Presence*, 2(1):82-86, 1993
- [9] D.Tolani and N.I.Badler, "Real-time inverse kinematics of the human arm", *Presence*, 5(4): 393-401, 1996
- [10] J. Zhao and N. Badler, "Inverse Kinematics positioning using nonlinear programming for highly articulated figures", *ACM Transactions on Graphics*, 13(4),313-316,1994
- [11] Vladimir M.Zatsiorsky, "Kinematics of Human Motion", 1998