

마커프리 모션캡처 시스템

Marker-Free Motion Capture System

유비쿼터스 시대를 주도할
디지털콘텐츠 기술 특집

박창준 (C.J. Park)	디지털액터연구팀 선임연구원
김성은 (S.E. Kim)	디지털액터연구팀 연구원
이인호 (I.H. Lee)	디지털액터연구팀 팀장

목 차

-
- I . 기술/산업/시장 동향
 - II . 개발 현황
 - III . 개발 결과
 - IV . 결론

최근 컴퓨터비전 기술을 이용하는 새로운 패러다임의 마커프리 모션캡처 기술이 미국의 MIT, CMU, MS, 일본의 ATR, MERL, 영국의 Oxford 대학 등에서 개발되고 있다. 마커프리 모션캡처는 연기자의 몸에 마커나 센서를 부착하지 않으며 특별한 조명이 필요 없으므로, 애니메이션 제작뿐만 아니라 일반인을 대상으로 하는 동작 인터페이스 분야로의 확대 적용이 가능한 모션캡처 방식이다. ETRI에서는 여러 응용 분야에 모션 인터페이스로 활용할 수 있는 환경 변화에 강인한 마커프리 모션캡처 시스템을 개발하고 있다. 몸에 마커나 센서를 부착하지 않은 자유 복장 상태의 동작자에 대해 조명 조건 변화 및 배경 변화에 강건하게 실시간 모션캡처 할 수 있는 기술 개발을 목표로 한다. 본 연구 개발이 성공한다면, 2007년에 876억 달러 규모로 확대될 전망이다. 영화, 방송물, 게임 등을 포함한 세계 영상 콘텐츠 시장에서 핵심 요소 기술 역할을 할 것이다. 그리고, 차세대 3D OS에서는 직관적 3D 포인팅 수단으로 활용될 수 있을 것이며, 2004년에 18,600만 대가 출고된 PC 시장을 고려하면 폭발적 수요가 예측된다.

I. 기술/산업/시장 동향

최근에는 컴퓨터비전 기술을 이용하는 새로운 패러다임의 마커프리 모션캡처 기술이 미국의 MIT, CMU, MS, 일본의 ATR, MERL, 영국의 Oxford 대학에서 개발되고 있다. 마커프리 모션캡처는 연구자의 몸에 마커나 센서를 부착하지 않으며 특별한 조명이 필요 없으므로, 애니메이션 제작뿐만 아니라 일반인을 대상으로 하는 동작 인터페이스 분야로의 확대 적용이 가능한 모션캡처 방식이다.

마커프리 모션캡처 기술의 중요성을 인식한 해외 대학 및 연구소에서는 공동 연구를 통해 기술을 개발하고자 하는 시도가 있으며, 영국의 Oxford 대학과 Vicon사와의 공동 연구를 비롯해 Oxford-Microsoft, ATR-UMD-MIT 간의 공동 연구, CMU-MIT-Mitshbishi-Microsoft사 간의 공동 연구가 이에 해당한다.

기존 연구[1],[2]에서 주로 접근한 방법은 인체의 실루엣 해석에 의한 모션 복원 방법이다. 그러나, 복원 가능 동작 유형에 제약이 많이 따르고, 복원 동작의 사실성이나 정밀도가 많이 떨어지며, 많은 처리 시간을 필요로 한다. 또한, 배경이나 조명의 영향을 아직 많이 받기 때문에, 일반적인 환경에서 사용하기에는 어려운 실정이다. 이러한 제약성을 줄이며 좀 더 정확하게 캡처하기 위해 대규모 카메라 입력 시스템과 Visual Hull Carving 기법을 활용한 마커프리 모션캡처에 대한 연구가 활발히 진행되고 있다[3].

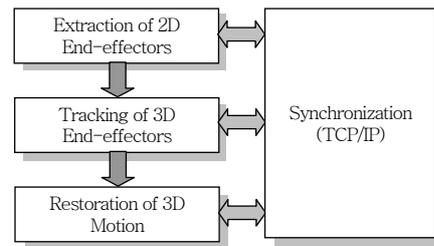
일반적인 환경에서도 구동 가능한 마커프리 모션캡처 시스템이 개발된다면, 2007년에 876억 달러 규모로 확대될 전망이다. 영화, 방송물, 게임 등을 포함한 세계 영상 콘텐츠 시장에서 핵심 요소 기술 역할을 할 것이다. 그리고, 차세대 3D OS에서 직관적 3D 포인팅 수단으로 활용될 수 있을 것이며, 2004년에 18,600만 대가 출고된 PC 시장을 고려하면 3D 인터페이스 기술을 필요로 하는 차세대 OS에서는 폭발적 수요가 예측된다.

II. 개발 현황

현재 ETRI에서는 적은 수의 카메라를 이용하면서도 조명의 변화와 배경의 변화에 대해 강건하게 인체 모션을 캡처할 수 있는 마커프리 모션캡처 시스템을 개발하고 있다.

이를 위해 기존 연구에서의 실루엣 해석에 의한 인체 동작 영상 분석 방법으로 접근하지 않고, 블롭(blob) 해석에 의해 신체 각 부위를 구분 검출하고, 이들의 3차원 위치 복원과 모션 구조화의 과정으로 인체 동작을 캡처하는 방법을 개발한다.

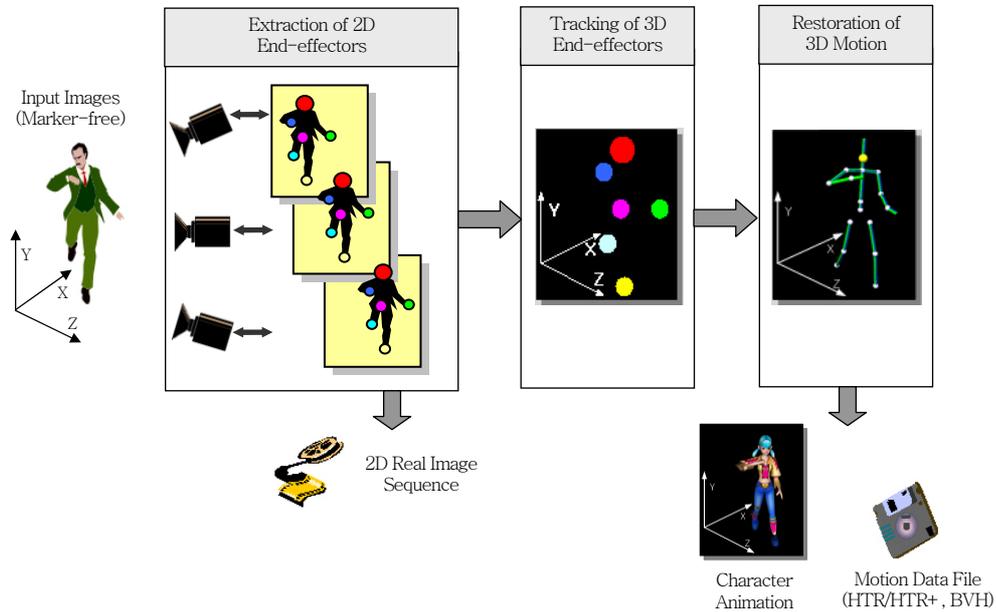
현재 개발중인 마커프리 모션캡처 시스템의 구성도는(그림 1)과 같다.



(그림 1) 개발한 시스템의 블록도

(그림 1)에서 보는 바와 같이 입력된 2차원 영상에 대해 블롭 해석으로 손, 발, 머리와 같은 인체의 끝점과 인체의 루트를 검출하는 모듈과 검출된 신체 끝점들의 시·공간적 매칭을 수행하여 3차원 좌표를 복원하고 이들 신체 끝점들을 추적하는 모듈과 역운동학을 이용하여 신체 끝점들과 루트의 3차원 좌표로부터 20개의 관절체로 구성된 인체 모델의 동작 데이터로 복원하는 모듈로 구성된다. 그리고, 이들 3개의 모듈을 동기화 시키고 파이프라인 처리가 가능하도록 하는 동기화 모듈이 전체 시스템을 제어한다. 특별한 하드웨어 가속기나 영상처리 라이브러리를 사용하지 않고 시제품을 구현하였으며, PC에 탑재된 CPU만으로도 실시간 처리가 가능하도록 알고리즘을 개발하였다.

입력된 데이터를 분석하고, 모션을 복원하는 3개의 모듈에서의 데이터 흐름은(그림 2)에 나타내었다.



(그림 2) 개발 시스템에서의 데이터 흐름도

(그림 2)에서 보는 바와 같이, 자유 복장 상태의 동작자 영상을 다수의 카메라를 제어하여 취득하고, 이들 영상을 분석하여 신체의 끝점과 루트의 2차원 위치를 검출하고 신체 부위를 구분한다. 3차원 정합 및 추적 모듈에서는 다중 영상에서의 2차원 좌표를 이용하여 3차원 위치로 복원하고, 각 신체 부위를 구분하고 추적한다. 그렇게 형성된 6개의 좌표 데이터를 이용하여 3차원 모션 복원 모듈에서는 각 중간 관절의 3차원 위치를 추정하여 20개 관절체의 모션으로 복원하고, 그 결과를 캐릭터 애니메이션으로 표시하여 사용자로 하여금 확인할 수 있도록 해준다. 또한, 캡처된 데이터를 상용 3D 소프트웨어에서 활용 가능한 형태로 출력해 주는 기능도 수행한다.

1. 인체 부위 검출 및 태깅

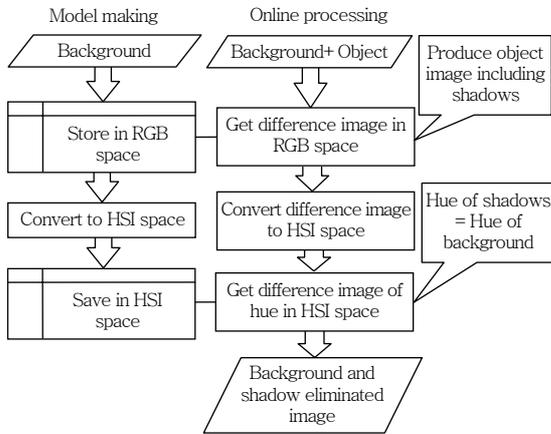
캡처된 영상에서 머리, 손, 발을 구성하는 블립을 직접 검출함으로써 외곽선 해석에 의한 방법보다 더 다양한 자세에 대해 서로 구분 검출 가능하도록 하였다. 통상적으로는 블립 해석을 하게 되면 2차원 영역 해석과 관련되는 연산이 많게 되어 외곽선을 해석하는 경우보다 많은 계산량이 필요하지만, 본

연구에서는 단계적인 후보 블립 선별 작업을 할 수 있는 방법을 개발함으로써 실시간 처리가 가능하도록 하였다.

그리고, 조명 변화와 배경 변화에도 강건하게 신체 각 부분을 구분 검출할 수 있도록, 신체 검출 및 구분을 위한 환경 적응적 모델을 고안함으로써 환경 변화에도 안정적으로 구분 검출할 수 있도록 하였다.

또한, 동작자의 동작에 의해 필연적으로 형성되는 soft shadow는 동작자 해석에 많은 영향을 주게 되는데, 이들 soft shadow의 영향을 받지 않고 동작자 영역만을 선별할 수 있도록 soft shadow 실시간 제거 방법도 개발하였다.

(그림 3)에서 보는 바와 같이, RGB 컬러 공간을 이용하여 미리 저장된 배경과 차이가 있는 영역인 동작자와 soft shadow 영역을 검출하고, 이들을 HSI 공간에서 비교를 수행하게 되면, soft shadow 영역의 hue 값은 미리 저장된 배경의 hue 값과 유사하므로 자동적으로 삭제되고 동작자 영역만 검출되게 된다. 고안한 방법은 배경과 차이가 있는 동작자 후보 화소만 변환하여 soft shadow를 제거할 수 있는 방법이므로 고속 처리가 가능하다.



(그림 3) Soft Shadow 제거 알고리즘 흐름도

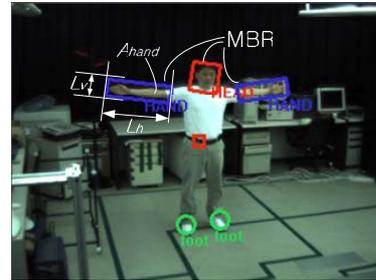
머리, 손, 발 각각을 나타낼 수 있는 대략적인 형태 모델을 정의하였고, 이들에 대한 색 범위, 신체 각 부위 간의 거리에 대한 모델을 정의하였다. 이들 모델은 동작자의 초기 자세에서 한 장의 영상을 취득하여 각 모델을 형성하도록 하였으며, 불균일 조명과 동작자의 동작에 따른 신체 부위 형태, 색, 명암 변화에 대처할 수 있도록 각 모델을 위한 가중치 요소(weighting factor)를 두었다.

(그림 4)는 형태 모델 형성을 위한 특징 요소들을 나타낸다.

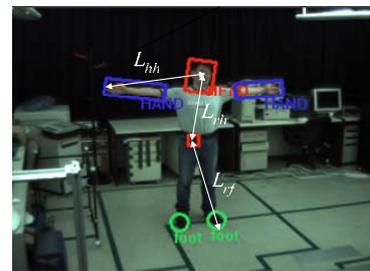
각 신체 끝점을 감싸는 MBR(최소외접사각형)의 가로 길이와 세로 길이의 비, 각 신체 부위의 크기, MBR에 대한 신체 부위 크기 비율과 같이 세 가지 요소로 정의하였다. (그림 4)에서 보는 바와 같이 형태 모델은 검출 대상 신체 부위의 대략적인 형태를 나타내도록 만들었다. 동작자의 동작에 따라 2차원 영상에 취득되는 신체 부위는 판이하게 달라질 수 있다. 따라서, 대략적인 형태만을 정의하여 유사한 형태를 가지는 화소군을 검출하고, 다른 모델과 융합하여 사용하여 안정적으로 검출하도록 하는 것이 더 현실적인 접근 방법이 될 것이다.

팔을 벌린 동작자의 초기 자세로부터 머리, 손, 발 간의 거리에 대한 모델을 형성한다. 특징 요소는 (그림 5)에 나타내었다.

거리 모델 요소는 머리와 손의 최대 거리 L_{hh} , 루트와 발의 최대 거리 L_{fp} , 루트와 머리의 최대 거리



(그림 4) 형태 모델을 위한 특징 요소들



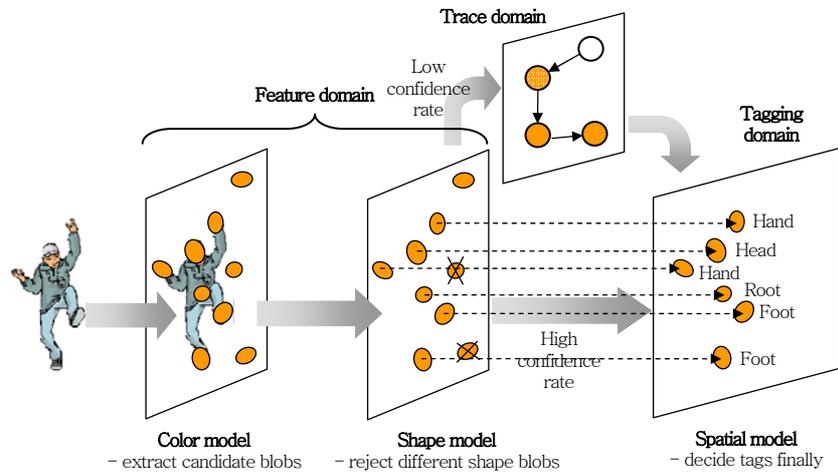
(그림 5) 신체 부위 간 거리 특징 요소

L_m , 이전 프레임의 각 신체 부위와의 거리 L_{fp} 로 정의한다.

상기와 같이 신체 부위 간의 상호 거리를 이용하여 검출된 신체 부위 위치의 타당성을 검토함으로써 신체 부위의 위치에 대한 후보 영역을 설정할 수 있다.

이와 같이 신체 부위의 대략적인 형태, 색 변화를 수용할 수 있는 색 요소, 신체 부위 간의 거리를 신체 부위 검출과 구분을 위한 모델로 정의함으로써 안정적인 신체 부위 검출과 구분이 가능하도록 하였다.

상의(上衣) 형태에 무관하게 인체 부위를 검출하기 위해 손 영역의 길이를 측정하여 각 모델에 대한 가중치를 결정한다. 손의 영역이 긴 경우에는 팔의 움직임에 따라 손 영역의 형태 변화가 심하므로 형태 모델에 대한 가중치를 줄이고, 색 모델과 거리 모델에 대한 가중치를 높인다. 또한, 손의 영역이 작은 경우에는 모두 동등한 가중치를 부여하여 검출하도록 한다. 발 영역에 대한 가중치 조절도 손에 대한 특징 가중치 조절과 동일하다. 이와 같이 각 모델에 대한 가중치를 조절함으로써 다양한 옷차림에 대해서



(그림 6) 신체 부위 검출 및 태깅 과정

도 신체 끝점들을 안정적으로 검출할 수 있게 된다. 신체 부위별 모델을 이용하여 입력된 영상에서 신체 부위를 검출하고 태깅하는 과정을 (그림 6)에 나타내었다.

(그림 6)에서는 보는 바와 같이 입력된 영상에 대해 우선적으로 컬러 모델을 이용하여 후보 화소들을 검출한다. 후보 화소를 검출하는 과정에서는 soft shadow 제거도 동시에 수행된다. 후보 화소로 검출된 화소 군들에 대해 블럽 레이블링을 수행하고, 이들에 대해 형태 특징을 추출한다. 추출된 특징과 미리 저장된 형태 모델과 유사한 블럽들을 검출하고, 각각의 모델에 대한 특징 공간에서의 거리를 이용하여 검출된 후보에 대한 신뢰도를 추출한다. 그리고, 추출된 신뢰도와 함께 신체 부위 상호간의 거리에 따른 거리 특징을 이용하여 최종적으로 신체 부위를 태깅한다. 이러한 태깅은 3차원 공간상에서의 정합과 추적에서의 정합과 대응에 대한 경우의 수를 줄일 수 있도록 한다. 그리고, 신뢰도가 낮은 경우에는 3차원 추적에서 전후간의 관계를 이용하여 태깅을 할 수 있도록 별도의 태그를 부여한다.

2. 인체 부위 3차원 공간 위치 복원

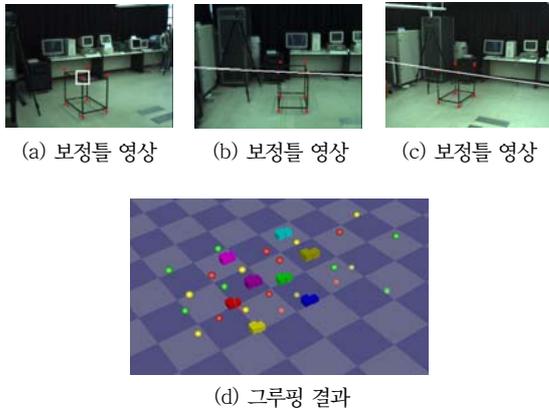
3차원 공간에서 움직이는 인체 부위가 2차원 영상에 투영되었을 때 발생하는 다양한 경우, 예를 들

면, 유사한 영역들이 인접하게 위치하거나 또는 다른 인체 영역으로 인해 특정 인체 영역이 영상에서 사라지거나, 서로간에 겹치는 경우에 대해서도 올바르게 3차원 정합을 할 수 있는 방법을 개발하였다.

가. 인체 영역 후보들의 3차원 공간 그루핑

한 카메라 영상 i 의 2차원 처리 결과로 얻은 인체 영역의 중점 $m_{i,n}$ (i 는 카메라 영상의 번호, n 은 i 에 포함된 인체 영역의 개수)이 다른 영상 i' 에서 만드는 에피플라 라인을 계산하고, 이 라인을 통과하는 i' 내의 모든 인체 영역을 찾는다[4],[5]. 찾아진 인체 영역의 중점을 $m_{i',n'}$ (i' 는 카메라 영상의 번호, n' 은 i' 에 포함된 인체 영역의 개수)라고 하면, 두 카메라의 사영 행렬을 이용해 $m_{i,n}$ 과 $m_{i',n'}$ 가 만드는 3차원 상의 한 점 $M_{(i,n)(i',n')}$ 을 계산할 수 있다. 이 과정을 모든 카메라 영상의 모든 인체 영역에 대해 반복하면 3차원 공간상에는 무수히 많은 점 M 이 생성되고, 이 중 매칭이 올바르게 이루어진 2차원 공간상의 인체 영역들의 조합은 3차원 공간에 서로 동일하거나 인접한 거리에 생성된다.

실험적으로 200mm 인근 거리에 존재하는 인체 영역들을 하나의 그룹으로 묶는다. 그룹으로 묶이지 않은 점들은 고스트로 간주한다. 실험에 사용된 3대의 카메라에서 특정 인체 영역이 모두 보일 경우 이



(그림 7) 카메라 보정틀에 대한 그루핑 알고리즘 적용 결과

인체 영역은 3차원 공간에서 3개의 점으로 구성된 그룹으로 나타난다. 어느 한 카메라에서 이 인체 영역이 나타나지 않거나 다른 인체 영역과 겹쳐져 보일 경우에는 2개의 점으로 구성된 그룹으로 나타날 경우도 있다. 3개의 점들이 모인 그룹을 편의상 G_{3b} , 2개의 점으로 구성된 그룹을 G_{2b} 이라 표기한다.

(그림 7)은 카메라 보정틀에 대해 그루핑을 적용한 결과다. (그림 7)의 (a)~(c)는 각각의 카메라로부터 입력된 영상이다. (그림 7a)의 중심에 있는 사각형으로 둘러싸인 보정틀의 한 점은 (그림 7b)와 (그림 7c)의 가로를 지나가는 에피폴라 라인으로 나타난다. 이 에피폴라 라인 위에는 후보 점이 두 개씩 있다. 이 중 하나는 올바르게 매칭된 점이고 나머지는 그렇지 않은 점이다. (그림 7d)는 에피폴라 조건에 맞는 후보 점들을 3차원 정합한 결과이다. 그림에서 구로 표현된 점은 그룹으로 묶이지 못한 고스트를, 정육면체로 표현된 점은 그룹으로 묶여진 올바른 정합 결과를 나타낸다. 3차원 공간에 그룹으로 묶인 후보들의 전체적인 배열은 8개의 꼭지점을 가진 육면체의 카메라 보정틀 형태와 동일함을 확인할 수 있다.

나. 3차원 공간에서의 인체 영역 추적

3차원 공간에서 인간의 움직임을 획득하기 위해 일반적으로 많은 수의 마커를 동작자의 몸에 부착한

후, 다중 카메라로 이들 마커의 움직임을 촬영하여 움직임을 계산한다. 이때 사용되는 마커는 인체의 표면이 아닌 뼈의 위치를 쉽게 계산하고, 몸통의 앞과 뒤 그리고 오른쪽과 왼쪽을 구별하기 위해 인체 부위마다 다양한 형태로 부착이 된다. 많은 수의 마커를 처리하는 데는 복잡한 알고리즘이 요구되지만, 부착된 마커들이 이루는 독특한 형태가 있기 때문에, 3차원 공간에서 올바른 위치에 생성된 마커의 위치 점과 그렇지 않은 점을 쉽게 구별할 수 있는 장점도 있다. 또한, 마커들의 이러한 독특한 형태를 이용해 카메라에서 관찰되지 않은 마커(인체 영역의 일부가 영상에 나타나지 않을 경우)의 위치를 역으로 추정할 수도 있다. 그러나, 본 연구에서는 마커를 전혀 사용하지 않기 때문에 일반적인 접근법으로는 올바른 3차원 공간 좌표를 계산할 수 없다.

인체 영역들의 2차원 위치 정보와 영역 태그 정보를 이용하여 일반적인 동작 외에도 유사한 인체 영역들이 인접하게 위치하거나 또는 다른 인체 영역으로 인해 특정 영역이 사라지는 경우, 서로간에 겹치는 경우에 대해서도 올바르게 인체 영역을 강인하게 구별할 수 있는 법이 필요하다. 이를 위해, 시·공간적 해석법(spatio-temporal analysis)을 제안한다. 공간적 해석법(spatial analysis)은 시간적 해석법(temporal analysis)에서 얻은 인체 부위의 예측위치와 그루핑된 3차원 인체 영역을 매칭시키는 과정이다. 시간적 해석법은 이미 얻어진 정보를 바탕으로 각 인체 부위의 위치를 최종적으로 올바르게 예측하는 과정이다. (그림 8)은 시·공간적 해석의 구성도를 나타낸다.



(그림 8) 시·공간적 해석법 구성도

1) 공간적 해석

연속된 영상에서 인체 영역의 최대 이동 거리는 200mm 내로 제한한다. 이 거리는 구현된 마커프리 모션캡처 시스템이 초당 30프레임으로 동작할 때

실험적으로 구해진 것이다. 이 거리보다 초과한 곳에 인체 영역의 후보가 존재한다면 이 후보는 고려 대상에서 제외된다. 추출된 인체 영역 그룹들 중에서 G_{3b} 가 6개 이상 존재한다면 모든 인체 영역은 사라지거나 겹쳐짐 없이 모두 추출되었다고 간주한다.

• 손에 대한 해석

손은 인체에 두 개가 존재하므로 좌, 우 방향을 확실히 구별해야 한다. 실시간 시스템에서 두 손의 좌, 우 구별은 쉽지 않다. 이 작업을 어렵게 만드는 요소는 첫째, 일반적으로는 두 손의 데이터가 서로 바뀌어도 인간의 인체 구조상 표현 가능한 자세가 되기 때문에 올바른 값으로 간주된다. 둘째, 손의 동작 범위는 아주 넓어서 사라짐과 겹침이 빈번히 발생한다. 특히 머리와 손이 겹쳐졌을 때에는 이 영역은 머리 영역으로 인식되기 때문에 손의 위치는 오직 확률적 예측 값으로 계산할 수 밖에 없다.

손의 공간적 해석에서는 좌, 우의 구별 없이 손이 될 가능성이 있는 모든 후보 영역을 설정한다. 'Hand' 태그가 3개 이상이 포함된 그룹(G_{3b})은 확실히 손의 영역이라고 간주한다. 확실히 손의 영역이라고 간주된 그룹이 2개인 경우는 겹침이나 사라짐이 전혀 발생하지 않았다고 간주하고, 좌, 우의 구별을 위해 바로 시간적 해석 루틴으로 프로세스가 넘어간다. 'Hand' 태그가 3개 이상이 포함된 그룹이 하나 이하로 존재한다면, 나머지 손 영역을 찾기 위해 'Hand' 태그가 2개 이하로 포함된 그룹(G_{2b})을 추가적으로 검색한다. 'Hand' 태그가 1개 포함된 그룹은 손 영역이 오직 하나의 카메라 영상에서만 검출된 경우이고, 이 그룹을 손의 영역으로 간주한다면 안정성은 아주 낮아지지만 검출되지 않은 손의 위치를 예측하는 데 아주 유용하게 이용되므로 확실한 손의 영역을 찾지 못했을 경우에는 이런 그룹도 손의 영역 후보에 등록한다.

• 발에 대한 해석

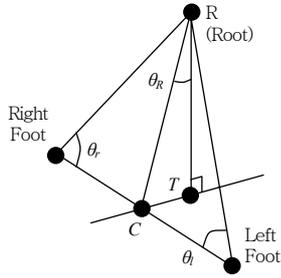
발의 공간적 해석에서도 손의 경우와 마찬가지로 좌, 우의 구별 없이 발이 될 가능성이 있는 모든 후보 영역을 설정한다. 'Foot' 태그가 3개 이상이 포함

된 그룹(G_{3b})은 확실히 발의 영역이라고 간주한다. 그리고, 'Foot' 태그가 3개 이상 포함된 그룹이 하나 이하로 존재한다면, 나머지 발 영역을 찾기 위해 'Foot' 태그가 2개 이하로 포함된 그룹(G_{2b})을 추가적으로 검색한다. 태그 정보를 이용한 해석은 손의 경우와 동일하다. 그러나 발의 영역에 국한된 두 가지의 문제점이 있어 이를 해결하기 위한 추가 조건이 필요하다.

기본적으로 두 발은 지면을 딛고 서 있고, 이 경우 두 발은 가까운 거리에 위치하므로 동작자의 정면에 위치하지 않은 카메라 영상에서는 거의 겹쳐진 상태로 나타난다. 걷거나 뛰는 일반적인 발의 동작에서도 지면 가까운 위치에서 발의 영역은 거의 겹쳐지게 되므로 정상적인 추적은 어렵다. 지면을 딛고 있는 발은 몸의 무게를 지탱하는 축이 되므로 미끄러짐이나 떨림과 같은 미세의 움직임도 발생하면 안 된다. 그러나 발의 영역이 겹쳐져서 하나로 보일 경우에는 영역의 중심이 이동하게 되어 미끄러짐이나 떨림 현상이 발생된다. 한 발은 지면에, 다른 발은 공중에 떠 있어 두 발이 모든 카메라에서 겹쳐져 보이지 않을 때에도 영상에서 발 영역을 이루고 있는 픽셀의 위치가 매 프레임마다 변하기 때문에 역시 떨림이 발생된다.

이를 보정하기 위해 몸의 무게를 지탱하는 주축을 찾아, 주축을 이루는 발의 움직임을 강제적으로 고정시키는 '발의 움직임 구속(constraint of feet's movement)' 알고리즘을 제안한다. 두 발이 모두 공중에 있다면 몸의 무게를 지탱하는 주축은 없고, 또한 이 주축에 포함된 발도 없다고 간주한다. 두 발 중 한 발은 공중에, 한 발은 지면에 있다면 지면에 있는 발이 주축 위에 있으므로 이 발의 움직임을 강제적으로 고정시킨다. 두 발이 모두 지면에 있을 경우에는 (그림 9)와 같이 두 발의 중심 C 와 몸의 중심 R , 그리고 R 을 지면에 수직으로 내렸을 때의 점 T 가 이루는 각 θ_R 을 식 (1)과 같이 계산한다.

$$\theta_R = \cos^{-1} \left(\frac{l_{RT}^2 + l_{RC}^2 - l_{TC}^2}{2l_{RT}l_{RC}} \right) \quad (1)$$



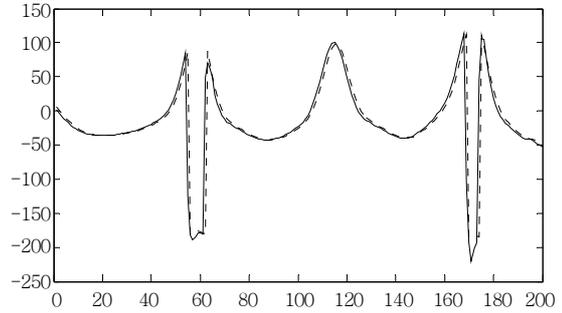
(그림 9) 몸의 주축을 찾기 위한 발의 움직임 구속

θ_r 가 실험적으로 인가된 20° 보다 크다면 R과 가까이 있는 발이 주축이 된다. 20° 보다 작다면 두 발과 몸의 중심이 이루는 각 θ_r 과 θ_l 을 구한다. θ_l 이 90° 보다 크다면 주축은 왼발, θ_r 이 90° 보다 크다면 주축은 오른발, 그렇지 않으면 두 발 모두 주축으로 간주한다. 이 알고리즘은 발의 영역이 겹쳐지거나 사라짐으로 인해 모호해져도 주축이 되는 발을 확실하게 고정시켜 주므로 추적의 안정성을 높여 준다.

2) 시간적 해석법

시간적 해석법은 이전 프레임에서 얻어진 인체 영역의 움직임을 분석하여 현재 프레임에서의 3차원 위치를 결정하고, 다음 프레임을 위한 움직임을 예측한다. 각 인체 영역의 움직임을 알고 있으므로 두 손과 두 발의 좌, 우 구분이 가능하고, 인체 영역이 사라지거나 서로간에 겹치더라도 위치의 예측이 가능하다.

움직임 예측을 위한 기본 가정으로 영상에서 얻은 특징의 데이터는 조명 조건, 동작자의 자세, 픽셀 연산의 오차 등으로 인해 잡음이 포함되어 있다고 간주한다. 이러한 환경에서 인체 영역의 움직임을 추적하기 위해 칼만 필터(Kalman filter)를 이용한다. 칼만 필터는 시간의 흐름에 대해, 역학적 모델과 관측 모델에 잡음이 존재하는 물리적인 시스템의 상태를 예측, 제어하는 데 적합하다[6]. 각 인체 영역에 대해 칼만 필터를 적용할 경우, 각 영역의 위치와 속도 값을 가진 필터링 결과를 이용해 특징의 움직임을 추적할 수 있다.



(그림 10) 오른손의 y축 움직임 궤적

(그림 10)에서 가로축은 연속된 영상열을, 세로축은 3차원 공간상의 y축을 나타낸다. 그리고, 점선은 영상으로 관측된 움직임에 대한 궤적이고, 실선은 칼만 필터를 이용해 추적된 결과를 나타낸다. 그림에서 보듯이, 60번째 영상과 170번째 영상 부근에서 동작이 급격하게 변하더라도 올바르게 추적하고 있음을 확인할 수 있다.

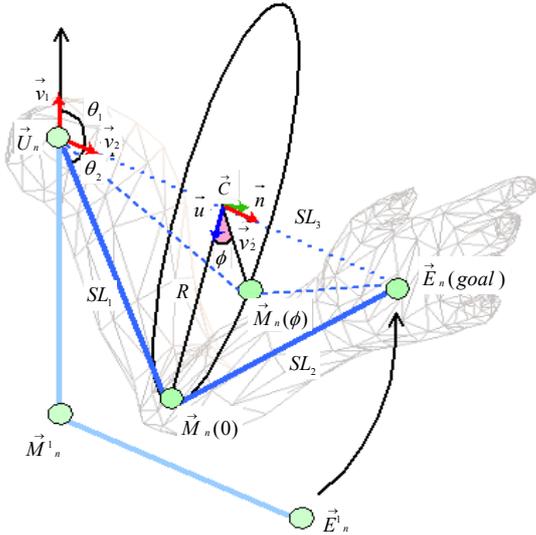
3. 인체 모션 복원

가. IK를 이용한 중간관절 위치 후보 생성

동작자의 동작을 캐릭터로 표현하기 위해서는 중간 관절들의 3차원 위치를 생성해야 하며, 중간관절의 위치를 생성하기 위해 IK 이론과 다양한 관절의 제약조건을 적용한다.

동작자의 동작과 가장 유사한 동작을 생성하는 중간관절의 위치 생성을 위해, 영상으로부터 복원된 루트 (R)와 End-effector ($E_n, n=1\sim 5$)의 3차원 위치정보와 각 관절을 연결하는 세그먼트 길이 정보 (SL_n)를 이용하여 End-effector들의 축과 활동반경을 따라 중간 관절의 위치를 계산한다.

(그림 11)에서 보는 것처럼, 두 관절 U_n 과 E_n 이 중간 관절 M_n 에 의해 연결되어 있을 경우, 관절 U_n 에서 M_n 까지의 길이를 SL_1 , M_n 에서 E_n 까지의 길이를 SL_2 , U_n 에서 E_n 까지의 길이를 SL_3 으로 정의한다. 정의된 정보를 이용하여 중간관절의 3차원 위치를 추정하기 위해서는 다음과 같은 과정을 거치게 된다.



(그림 11) 중간관절의 위치 생성

첫번째로, 중간 관절의 위치 데이터를 생성하기 위해 Y축과 벡터 \vec{v}_2 가 이루는 회전각 θ_1 과 관절 $\vec{U}_n \vec{M}_n \vec{E}_n$ 가 이루는 삼각형에서 관절 $\vec{U}_n \vec{M}_n$ 과 $\vec{U}_n \vec{E}_n$ 가 이루는 회전각 θ_2 를 계산한다. 회전각 θ_1 은 벡터의 내적을 이용하여 계산할 수 있으며, 회전각 θ_2 는 세그먼트 길이와 영상으로부터 계산된 End-effector \vec{E}_n 의 위치를 이용하여 (2)와 같이 구해진다.

$$\theta_2 = \cos^{-1} \left(\frac{SL_1^2 + SL_3^2 - SL_2^2}{2SL_1SL_3} \right) \quad (2)$$

두번째로, 계산된 회전각도와 평면의 법선벡터를 이용하여 중간관절의 3차원 위치를 추정한다. 이를 위해 본 연구에서는 [7],[8]에서 제안한 방식을 이용한다. (그림 11)에서 회전각 ϕ 는 팔꿈치의 평면 상에 놓여진 원형의 호를 따라 움직인다. 먼저, 어깨 \vec{U}_n 에서 손목 \vec{E}_n 까지의 축에 평행한 원의 법선벡터 \vec{v}_2 를 계산한다. 계산된 벡터 \vec{v}_2 를 이용하여 원의 중심 \vec{C} 와 이것의 반지름 $Radius$ 는 (3)에서와 같이 단순한 삼각법에 의해서 계산될 수 있다.

$$\begin{aligned} \vec{C} &= SL_1 \cos(\theta_2) \vec{v}_2 \\ Radius &= SL_1 \sin(\theta_2) \end{aligned} \quad (3)$$

또한 두 개의 단위벡터 \vec{h} 과 \vec{u} 가 필요하다. 벡터

\vec{h} 은 관절 $\vec{U}_n \vec{M}_n \vec{E}_n$ 가 이루는 삼각형에 수직인 벡터이고, 벡터 \vec{u} 는 벡터 \vec{h} 과 \vec{v}_2 에 수직인 벡터이다. 두 벡터 \vec{h} 과 \vec{u} 는 벡터의 외적을 이용하여 계산된다. 생성된 단위 벡터들을 이용하여 팔꿈치 위치는 (4)에 의해 계산되며, 팔꿈치 관절의 한계가 되는 ϕ 의 값을 변화시킴으로써 팔꿈치가 후보 위치를 한정시킬 수 있다.

$$\vec{M}_n(\phi) = \vec{U}_n + \vec{C} + Radius(\cos(\phi)\vec{u} + \sin(\phi)\vec{h}) \quad (4)$$

나. NURB를 이용한 모션 필터링

모션 필터링은 루트와 End-effector의 위치정보에 먼저 적용한 후 IK에 의해 중간관절의 후보 위치를 계산한다. 그리고 마지막 단계에서 모든 관절의 부드러운 움직임 표현을 위해 다시 한번 더 필터링을 한다. 필터링은 NURB 곡선을 이용한다. NURB 곡선은 B-spline 곡선에서 knot value의 간격이 일정치 않을 때 유도되는 NURB 함수를 블렌딩 함수 ($B_{i,k}(u)$)로 사용한다. 다만 NURB 곡선에서는 제어점에 호모지니어스 좌표를 사용하여 각각 $(x_i h_i, y_i h_i, z_i h_i, h_i)$ 의 형태로 표시하고 이들 네 개의 좌표를 블렌딩 함수에 섞어 준다. 각 관절의 3차원 위치 값 (x, y, z) 를 제어점으로 하여 NURB 곡선 알고리즘에 적용하여 노이즈를 제거한다.

NURB 곡선상의 점을 (xh, yh, zh, h) 라 하면 NURB 곡선상의 점 (x, y, z) 좌표는 (5)~(7)을 (8)로 나누어 주면 되므로 NURB 곡선의 방정식은 <표 1>과 같이 정의된다.

$$xh = \sum_{i=0}^n (h_i x_i) B_{i,k}(u) \quad (5)$$

$$yh = \sum_{i=0}^n (h_i y_i) B_{i,k}(u) \quad (6)$$

$$zh = \sum_{i=0}^n (h_i z_i) B_{i,k}(u) \quad (7)$$

$$h = \sum_{i=0}^n h_i B_{i,k}(u) \quad (8)$$

〈표 1〉 NURB 곡선 방정식 인자

P_i	제어점 ($i=0 \cdots n, i=n+1$)
d	- B-spline의 order. 2부터 제어점 ($n+1$)의 수 범위 내의 정수 값에서 선택함 - d 가 낮아질수록 control polygon에 근접하는 형상 ($d=3$ 을 사용)
u_i	비주기적 knot value
$u_i =$	$\begin{cases} 0 & (0 \leq i < d) \\ i-d+1 & (d \leq i \leq n) \\ n-d+2 & (n < i \leq n+d) \end{cases}$
$P(u) =$	$\frac{\sum_{i=0}^n h_i P_i B_{i,d}(u)}{\sum_{i=0}^n h_i B_{i,d}(u)} \quad (0 \leq u \leq n-d+2)$
$B_{i,d}(u) =$	$\begin{cases} 1 & : u_i \leq u < u_{i+1} \\ 0 & : otherwise \end{cases}$
$B_{i,d}(u) =$	$\frac{(u-u_i) B_{i,d-1}(u)}{u_{i+d-1}-u_i} + \frac{(u_{i+d}-u) B_{i+1,d-1}(u)}{u_{i+d}-u_{i+1}}$

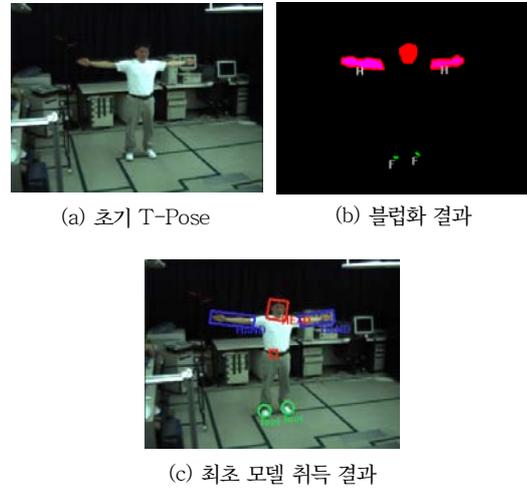
III. 개발 결과

1. 2차원 영상 해석 결과

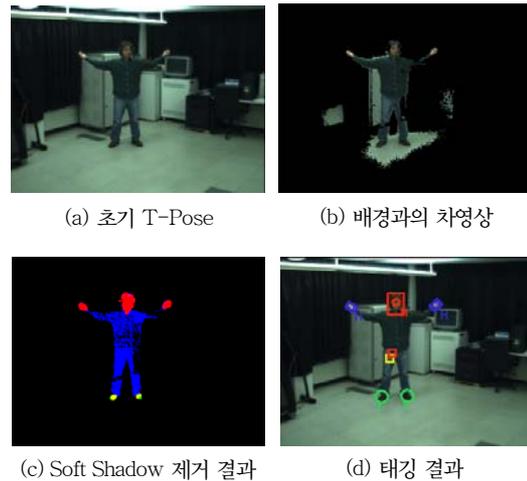
신체 크기와 복장 상태, 성별이 다른 15명의 동작자(남자 11명, 여자 4명)에 대해 모션캡처를 실행하였다. 가로 4m, 세로 4m, 높이 3m의 캡처 볼륨을 제공하는 실험실에서 수행하였으며, 일반적인 형광 등 조명을 활용하였다.

(그림 12)는 여름 옷에 대한 모델 생성 결과를 나타낸다. (그림 12b)에서 보는 바와 같이 동작자의 머리, 손, 발 끝점에서 출발하여 영역 확장을 통해, 유사한 색 분포를 가지는 영역에 대해 블러핑을 수행한다. 그리고, (그림 12c)에서 보는 바와 같이 정의된 모델을 자동으로 추출하고 태깅한다.

(그림 13)에는 동작자에 의해 필연적으로 형성되는 soft shadow를 제거하고 신체 부위를 검출 및 태깅한 결과를 나타내었다. 입력된 영상에 대해 미리 저장된 배경과 비교를 통해 차영상을 구하면(그림 13b)처럼 soft shadow 부분도 객체로 오인되어 검출된다. 본 연구에서 개발한 방법을 적용하면(그



(그림 12) 여름 옷을 입은 동작자에 대한 모델 생성 결과

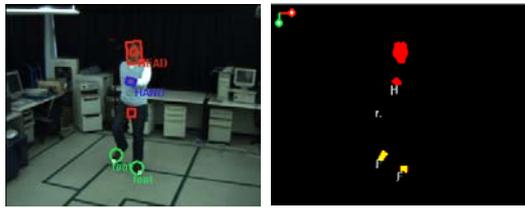


(그림 13) Soft Shadow 제거

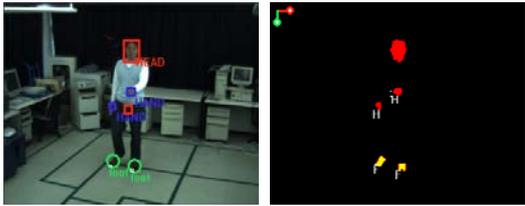
림 13c)처럼 soft shadow는 자동으로 제거되고, 동작자만 검출된다.

(그림 14)에는 동작자의 자체 동작에 의해 사라지는 신체 부위에 대한 처리 과정을 나타내었다.

(그림 14)에서 보는 바와 같이 자체 동작에 의해 신체 부위가 유실되는 경우에도 사라지기 전의 좌표를 루트와의 상대 위치로 저장하고, 다시 나타나는 경우에는 저장된 위치와 비교하여 재 태깅을 수행하게 된다. 또한, soft shadow 제거를 통해 동작자 영역만을 후보 영역으로 검출함으로써 상기와 같이 인



(a) Self-occlusion 된 손 (b) Occlusion에 의해 손실되기 전에 저장된 손 위치



(c) 다시 나타난 손 (d) 재 태깅된 결과

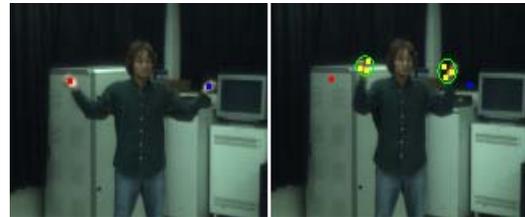
(그림 14) 동작자의 동작에 의해 유실되는 신체부위에 대한 처리

체 부위가 유실되는 경우에도 정확하게 처리할 수 있게 된다.

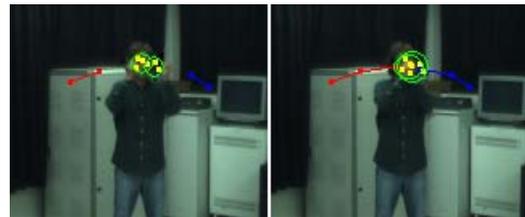
본 연구에서 개발한 인체 부위 검출 및 태깅 모듈을 실험해 본 결과 99.95%의 태깅 성공률을 보인다. 특히, 다른 신체 부위로 잘못 태깅하는 경우는 없었다. 또한, 영상에는 손, 발, 머리가 취득되었지만, 그들을 검출하지 못하는 경우는 없었다. 다만, 조명의 균일성이 매우 취약한 경우에는 초기에 취득된 모델과의 특징 공간에서의 거리가 멀어서 태그를 부여하지 않는 경우가 0.05% 정도로 나타났다. 따라서, 본 연구에서 개발한 신체 부위 검출 및 태깅 방법은 환경 변화에 매우 강인하고 적은 수의 카메라로 다양한 동작을 캡처할 수 있음을 알 수 있다.

2. 3차원 정합 및 추적 결과

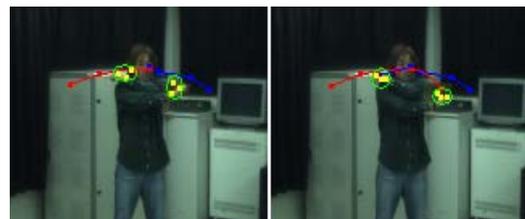
연속적으로 입력되는 영상에서 인체 영역의 3차원 매칭의 평균 오차 거리는 1.02cm이다. 인체 영역의 추적 성공률은 99.95%로 나타났다. 이것은 영역들이 서로 겹쳐졌다가 분리되거나 또는 사라진 후 다시 나타나는 동작이 포함된 다양한 동작의 실험에서 계산되었다.



(a) 손 영역 추적 (b) 손 영역 추적



(c) 손 영역 추적 (d) 손 영역 추적



(e) 손 영역 추적 (f) 그루핑 결과

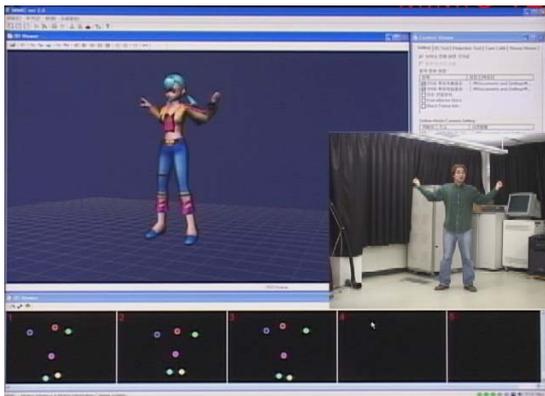
(그림 15) 연속적으로 입력되는 영상에서 두 손의 움직임을 추적하는 과정

(그림 15)는 연속적으로 입력되는 영상에서 두 손의 움직임을 추적하는 과정을 보여준다. 손 영역 위에 있는 점들은 올바르게 매칭이 되어 그룹이 형성된 손 영역 후보의 중점을 나타낸다. 고스트는 영상에 표시하지 않았다. 손 영역의 매칭 후보는 (그림 15d)와 같이 두 손이 겹쳐진 상태에서도 올바르게 그루핑됨을 확인할 수 있고, (그림 15e)와 같이 두 손의 좌, 우가 엇갈린 상태에서도 추적 알고리즘에 의해 올바르게 추적된 후 그루핑됨을 확인할 수 있다. 특히 (그림 15c~f)의 경우에는 실제로 시스템에서는 얼굴 영역에 대한 추적도 함께 이루어져 세 개의 영역이 동시에 겹치게 되어 추출된 인체 영역은 위치 오차를 포함하고 있지만 그루핑된 손 후보 영역들과 추적 알고리즘을 이용해 오류 없이 추적하고 있다.

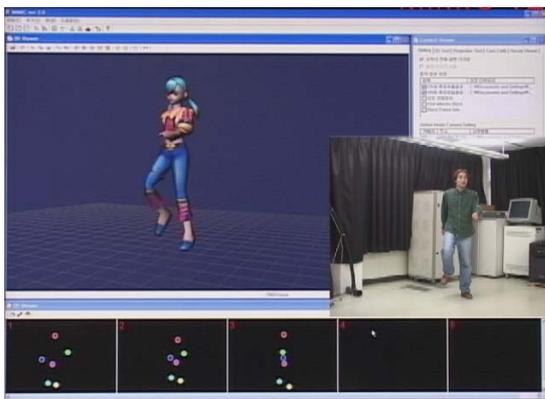
3. 모션 복원 결과

(그림 16)은 동작자의 캡처된 데이터를 실시간으로 가상 캐릭터에 적용시켜 움직임을 생성한 결과를 보여주고 있다. 중간 관절 위치 추정의 정확도는 조명 조건 및 카메라의 개수에 따라 차이가 있으나, 인위적으로 취해진 특이한 자세가 아닌 일상적인 동작들에 대해서는 정확한 결과를 보여준다.

(그림 17)은 다양한 캐릭터에 적용된 모션 복원 결과를 보여준다. 동작자의 키와 관절의 길이, 특히 세번째 캐릭터인 멧돼지와 같이 다리가 아주 짧은 경우에도 상관없이 동작은 캐릭터에 실시간으로 잘 매핑됨을 확인할 수 있다.



(a) 팔 벌린 자세에서의 팔꿈치 위치 복원 결과



(b) 걷기 동작에서의 팔꿈치, 무릎 위치 복원 결과

(그림 16) 중간 관절 위치 복원 결과



(a) 지르기 동작



(b) 앞기 동작



(c) 걷기 동작

(그림 17) 20개의 관절을 가지는 인간형 캐릭터에 적용된 다양한 모션의 복원 결과

IV. 결론

본 연구에서 개발한 모션캡처 방법 및 시스템은 실시간 그림자 제거 방법을 제공하고, 불균일 조명과 배경의 일부 변화를 허용한다. 또한, 신체 부위에 대해 정확하게 태깅을 할 수 있다. 그리고, 동작자 자체 동작에 의해 필연적으로 유실되는 신체 부위에 대해서도 지속적인 추적을 통한 자연스러운 모션 복원이 가능하다. 또한, 초기 동작자 모델 취득 과정에서 확보한 신체 손, 발, 머리의 5개의 신체 끝점과 루트를 이용하여 온라인으로 중간 관절을 추정함으로써 20개의 관절체 모션으로 타당성 있게 복원한다.

현재 개발된 시스템은 동작자의 정확한 동작을 캡처하기 위해 개선해야 하는 사항이 있으나, 현재의 기술로서도 모션 인터페이스로 이용할 수 있

며, 실시간 모션 캡처가 가능한 것을 확인하였다. 또한 앞으로 이 기술을 발전시키면 게임이나 애니메이션 분야와 같은 다양한 멀티미디어 분야에서도 이용 가능할 것이다.

약어 정리

CPU	Central Processing Unit
HSI	Hue, Saturation, Intensity
IK	Inverse Kinematics
MBR	Minimum Bounding Rectangle
NURB	Non-Uniform Rational B-spline
RGB	Red, Green, Blue

참고 문헌

- [1] Y. Li, A. Hilton, and J. Illingworth, "A Relaxation Algorithm for Real-time Multiple View 3D-tracking," *Journal of Image and Vision Computing*, Vol.20, 2002, pp.841-859.
- [2] R. Hoshino, S. Yonemoto, D. Arita, and R. Taniguchi, "Real-Time Motion Capture System Based-on Silhouette Contour Analysis and Inverse Kinematics," *7th Korea-Japan Joint Workshop on Computer Vision*, Vol.7, Feb. 2001, pp.157-163.
- [3] Lars Mündermann, "Conditions That Influence the Accuracy of Anthropometric Parameter Estimation for Human Body Segments Using Shape-from Silhouette," *SPIE Electronic Imaging*, Jan. 2005.
- [4] Emanuele Trucco and Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, New Jersey, 1998.
- [5] Gang Xu and Zhengyou Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition*, Kluwer Academic Publishers, 1996.
- [6] Greg Welch and Gray Bishop, "An Introduction to the Kalman Filter," *TR 95-041, Univ. of North Carolina at Chapel Hill*, Feb. 2001.
- [7] D. Tolani and N.I. Badler, "Real-time Inverse Kinematics of the Human Arm," *Presence*, 1996, Vol.5, No.4, pp.393-401.
- [8] J. Zhao and N. Badler, "Inverse Kinematics Positioning Using Nonlinear Programming for Highly Articulated Figures," *ACM Transactions on Graphics*, Vol.13, No.4, 1994, pp.313-316.