

# 구성요소 기반 인체 추적을 이용한 실시간 아바타 애니메이션<sup>☆</sup>

## Real-time Avatar Animation using Component-based Human Body Tracking

이 경 미  
Kyoung-Mi Lee

### 요 약

인체 추적은 차세대 인간과 컴퓨터사이의 상호작용 (HCI)의 필수요소이다. 본 논문에서는 구성요소에 기반을 둔 인체 모델을 이용하여 인체의 각 부위를 검출하여 자세를 추정하고 아바타의 동작을 구현하는 방법을 제안한다. 인체 각 부위의 색상정보와 함께 연결정보, 위치정보 등을 이용하여 인체 각 부위가 검출되고 인체 모델의 각 구성요소에 매칭된다. 이렇게 구한 2D의 인체 자세 정보는 다음 프레임에서 유사도 계산을 통해 사람 추적에 이용될 수 있다. 이 때, 각 구성요소의 상대적인 위치 관계를 이용하여 깊이 정보를 추출하고 이를 움직임 방향으로 변환하여 2-1/2D 인체 모델을 구한다. 인체 각 부위는 자세와 방향성으로 모델링 되고 연결된 3D 아바타의 각 부분은 인체 모델로부터 넘겨받은 정보를 이용하여 3D 회전을 적용함으로써 실시간 아바타 애니메이션을 구현하게 된다. 다양한 자세를 포함하고 있는 실험 동영상에 대해 90%의 인체 추적율을 얻었고, 처리된 프레임 수가 늘어남에 따라 모델이 체계화되어 추적율이 꾸준히 증가하는 결과를 나타냈다.

### Abstract

Human tracking is a requirement for the advanced human-computer interface (HCI). This paper proposes a method which uses a component-based human model, detects body parts, estimates human postures, and animates an avatar. Each body part consists of color, connection, and location information and it matches to a corresponding component of the human model. For human tracking, the 2D information of human posture is used for body tracking by computing similarities between frames. The depth information is decided by a relative location between components and is transferred to a moving direction to build a 2-1/2D human model. While each body part is modelled by posture and directions, the corresponding component of a 3D avatar is rotated in 3D using the information transferred from the human model. We achieved 90% tracking rate of a test video containing a variety of postures and the rate increased as the proposed system processed more frames.

☞ Keyword : Component-based model, human tracking, animation, HCI

## 1. 서 론

컴퓨터 애니메이션에 대한 동작 캡처의 사용은 1970년 후반에 시작되어 최근에야 널리 이용된 비교적 새로운 학문분야이다. 동작 캡처는 분석을 위해 인체 움직임을 기록하는데, 캡처된 정보는 공간에서 인체의 단순한 위치에서 얼굴이나 근육의 변

형처럼 복잡할 수도 있다. 컴퓨터 애니메이션을 위한 동작 캡처는 사람의 동작을 컴퓨터 아바타의 동작에 매핑하는 과정을 포함한다. 이러한 매핑은 사람의 팔 동작으로 아바타의 동작을 제어하는 것처럼 직접적일 수도 있고, 사람의 손이나 손가락 패턴이 아바타의 피부색이나 감정 상태를 조절하듯이 간접적일 수도 있다.

지금까지 동작 캡처에 의한 애니메이션은 모션 캡처와 같은 장비를 갖추고 마커(marker)를 부착하여 동작을 시뮬레이션하는 것이다. 그러나 모션 캡처 장비는 고가일 뿐만 아니라 마커를 부착하는 것으로 사용자의 편의성에 장애가 되고 있다. 최근

\* 정 회 원 : 덕성여자대학교 컴퓨터공학부 교수  
kmlee@duksung.ac.kr

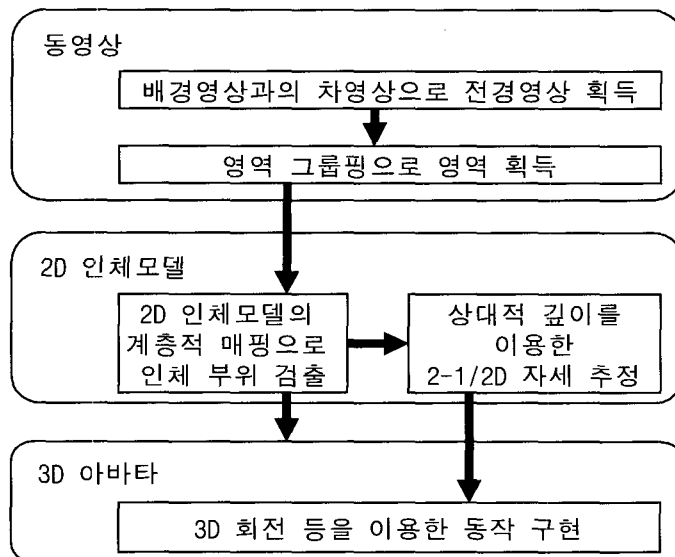
[2005/01/28 투고 - 2005/04/26 1차 심사 - 2005/09/12  
2차 심사 - 2005/11/18 심사완료]

☆ 이 논문은 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(R04-2003-000-10092-0 (2005)).

수십 년 동안 인체를 캡처해서 동작을 분석하는 여러 가지 방법들이 제안되었는데, Gavrilin은 이들 방법들을 사용된 모델의 종류 (막대형[11], 공간적[12], 통계적[10])와 추적 공간 (2D[10,11], 3D[12])에 따라 분류하였다[4]. 특히, 아바타의 동작을 제어하기 위해 필요한 3D 공간에서의 동작 캡처에 대한 연구는 인체의 3D 모델을 직접 이용하거나[8], 영상으로부터 훈련된 인체의 2D 모델들을 이용하는 방법들로 분류될 수 있다. 또한, 동영상에서 동작을 캡처하기 위한 시각기반 방법으로 색상을 이용하여 영상을 분할한 후 윤곽선을 분석하여 비슷한 색상으로 연결된 블랍들의 집합으로 인체를 표현하고 단일 카메라로 추적하는 방법이 소개되었다[8,9]. 또한, 네 대의 카메라에서 획득된 영상에 템플릿 매칭이 이용되거나[5], 직각으로 배치된 세 대의 카메라에서 획득된 영상에 윤곽선 분석을 실행하는 방법[6] 등이 소개되었다. 특히, Colombo et. al은 두 대의 카메라에서 얻은 2D 영상에서 3D 동작을 추출하여 아바타를 애니메이션 시키는 저렴한지만 강인한 시각기반 방법을 제안하였다[1]. 이렇게 마커를 부착하지 않고 동작을 애니메이션 시키는 방법은 가상 회의 및 원격 회의와 같은 실제적인 분야에 응용될 수 있다. 그러나 그들이 제안한 방법

은 머리와 손만을 피부색으로 제한할 뿐만 아니라, 인체 모델과 동작 추적이 머리와 손, 발에 대해서만 이루어지고 있다.

본 논문에서는 마커를 부착하지 않고(marker-free), 한 대의 카메라에서 획득한 2D 동영상으로부터 3D 아바타의 동작을 애니메이션 시키는 방법을 제안한다. 우선, 인체의 각 부위를 구성요소로 정의하고 색상정보, 연결정보, 위치정보 등을 포함하는 구성요소 모델을 이용하여 단일 카메라로부터 획득된 2D 동영상에서 인체의 각 부위를 검출하고 동작을 추적한다. 검출된 부위는 역시 구성요소로 이루어진 3D 아바타의 각 구성요소에 매칭될 것이다. 연속된 프레임들과 구성요소들 간의 상대적인 관계를 이용하여 2-1/2D 움직임이 획득되고 3D 회전과 같은 3D 동작을 이용하여 아바타의 3D 움직임이 구현된다. 그림 1은 제안하는 시각기반 3D 애니메이션 구현에 대한 순서도를 보여준다. 본 논문에서 제안하는 구성요소 기반 모델과 3D 아바타에 대해서는 2장에서 설명하고, 3장에서는 동영상에서 전경영상을 분리하여 인체 영역을 획득하고, 인체 부위를 검출하여 자세를 추정하며, 아바타의 3D 동작으로 구현되는 과정을 소개한다. 실험 결과와 결론은 4장과 5장에서 보여준다.



〈그림 1〉 제안된 방법에 대한 순서도

## 2. 구성요소 기반의 인체 모델

### 2.1 구성요소 기반의 모델

인체를 검출하거나 동작을 추적하기 위해서, 인체 전신보다 인체 부위들이 자세나 조명 변화에 덜 민감하고 각 부위들이 독립적으로 검출될 수 있기 때문에, 구성요소에 기반한 인체검출 방법의 성능이 더 좋다고 알려져 있다. 그림 2는 본 논문에서 제안하는 구성요소에 기반한 인체 모델을 보여준다. 인체는 10개의 부위로 구성되어 진다: 머리, 몸통, 왼쪽 팔 상단, 왼쪽 팔 하단, 오른쪽 팔 상단, 오른쪽 팔 하단, 왼쪽 다리 상단, 왼쪽 다리 하단, 오른쪽 다리 상단과 오른쪽 다리 하단 그림 2(a). 본 논문에서는 3D 공간에 존재하는 아바타를 모델링하기 위해 Maya를 사용하여 그림 2(a)에 제안된 인체 모델의 각 구성요소를 모델링하였다(그림 2(b)).

그림 2(a)의 인체 모델의 각 구성요소  $p_i$ 는 기하 (geometrical) 정보  $g_i$ 와 외형(appearance) 정보  $a_i$ 을 포함한다. 또한 각 구성요소는 다른 구성요소와의 연결(connection) 정보  $c_i$ 와 구성요소간의 상대적인 깊이(depth) 정보  $d_i$ 을 포함한다. 따라서 각 구성요소  $p_i$ 는 다음과 같이 표현된다:

$$p_i = (g_i, a_i, c_i, d_i) \quad (1)$$

여기서 기하정보는 위치와 상대적 크기 등, 외형 정보는 평균 색상과 표준편차 등, 연결정보는 각 구성요소를 표현하는 네 면 중 연결 가능한 면과 연결되는 부위의 명칭, 연결 각도, 연결 거리 등을 사용한다.

구성요소에 기반을 둔 인체 모델을 이용하여 동작을 추적하는 것은 검출된 구성요소들을 조합하여 최적의 인체 모델을 구성하는 매칭 과정이다. 식(1)에 정의된 구성요소  $p_i$ 은 아래와 같은 매칭식을 이용하여 가장 작은 차이  $D_i$  또는 가장 큰 매칭 확률을 가진 조합을 최적의 매칭으로 여긴다:

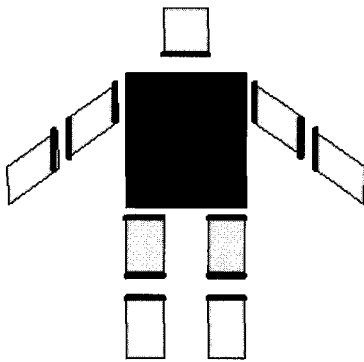
$$D_i = \sum_j^{\min} \| p_i - p_j \|$$

$$Human.Score = \sum_{i=1}^I D_i$$

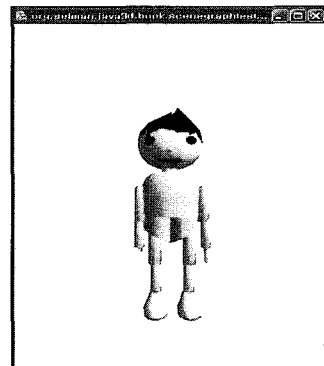
여기서  $p_j$ 는  $j$  번째 구성요소,  $I$  는 구성요소의 개수(본 논문에서는 10)를 의미한다.

### 2.2 계층적 연결을 이용한 모델

구성요소에 기반한 방법은 매칭을 위한 탐색시간을 줄임으로써 시간 효율적인 알고리즘을 작성하기 위해 구성요소들을 계층적으로 연결할 수 있다[2]. 그러나 논문[2]에서 사용된 별구조 형태의 트리구

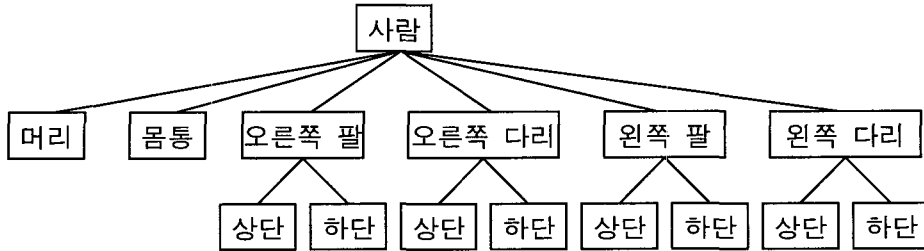


(a)



(b)

<그림 2> 구성요소에 기반한 인체 모델: (a) 2D 인체 모델, (b) 3D 아바타. 인체 모델은 머리, 몸통, 왼쪽 팔 상단, 왼쪽 팔 하단, 오른쪽 팔 상단, 오른쪽 팔 하단, 왼쪽 다리 상단, 왼쪽 다리 하단, 오른쪽 다리 상단, 오른쪽 다리 하단의 구성요소로 이루어진다. (a)의 굵은 선은 다른 구성요소와 연결될 수 있는 면을 나타낸다.



〈그림 3〉 인체 구성요소 모델의 계층적 관계

조와는 달리, 본 논문에서는 일반 3계층의 트리구조를 가진 모델을 사용한다(그림 3). 가장 상위 노드인 사람은 머리, 몸통, 오른쪽 팔, 왼쪽 팔, 오른쪽 다리, 왼쪽 다리의 자식 노드를 가진다. 다시, 두 팔과 두 다리는 각각 상단/하단의 두 자식 노드를 가진다.

상하관계로 표현된 인체 모델은 일반-상세(coarse-to-fine) 탐색법에 의한 매칭을 적용한다. 즉, 팔(arm)을 먼저 찾은 후, 팔의 영역 안에서 왼쪽/오른쪽 팔을 정하고 팔의 상단(LUarm/RUarm)과 하단(LLarm/RLarm) 등의 순서로 탐색을 수행한다. 예를 들어, 왼쪽 팔 상/하단을 상세 탐색하기 전에 왼쪽 팔을 먼저 매칭시키고 왼쪽 팔에 대한 매칭 확률  $D_{Larm}$ 을 왼쪽 팔 상/하단 매칭 확률( $D_{LUarm}/D_{LLarm}$ )의 합으로 여긴다.

$$D_{Larm} = D_{LUarm} + D_{LLarm}$$

다른 부위에 대해서도 이와 같은 계층 구조를 반복적으로 적용한다.

3D 아바타의 구성요소(그림 2(b))도 부모-자식의 계층적 연결을 이용하여 정의되어야 한다. 계층 구조의 중요한 원칙은 부모 노드의 움직임이 자식 노드의 움직임에 영향을 미친다는 것이다. 예를 들면, 오른쪽 팔의 하단의 위치는 오른쪽 팔의 상단의 회전에 의해 우선 결정된 후, 오른쪽 팔의 하단의 세부 움직임에 대해 회전이 이루어진다.

### 3. 구성요소 기반 인체 추적

본 장에서는 주어진 동영상에서 사람 검출, 대략

적인 인체 부위 검출, 팔과 다리의 상세한 인체 부위 검출, 3D 동작을 위한 2-1/2D 움직임 추출 등을 소개한다.

#### 3.1 인체 추정 영역 검출

##### (1) 배경 제거

사람을 검출하기 위해서, 우선 배경 프레임과 사람이 등장한 영상과의 차영상을 구해 전경 영상을 추출할 수 있다. 그러나 조명 및 잡음 등의 상황에 매우 민감하므로 올바른 배경 추출이 어려우며, 결국 인체 추적을 어렵게 만드는 원인이 된다.

본 논문에서는 지난 배경 프레임의 통계적 정보인 표준편차( $\sigma^t$ )와 평균값( $\mu^t$ )을 이용하여 현재의 영상에 대한 차연산을 적용시킴으로써 조명 변화에 강건한 전경 영상의 추출이 가능하다. 우선, 첫 번째 프레임을 배경으로 사용시킴으로써 통계적 정보를 0으로 초기화 한다 ( $\mu^0 = \sigma^0 = 0$ ). 시간 t 동안의 배경과 현재 영상( $C^t$ )과의 차이를 이차 편별함수인 마할라노비스 거리( $\sigma^t$ )를 이용하여 다음과 같이 구한다:

$$\sigma^t = \sqrt{\frac{(C^t - \mu^{t-1})^2}{(\sigma^{t-1})^2}} \quad (2)$$

새로운 프레임이 들어올 때마다, 식 (2)에 의해서 구해진 값이 미리 정해진 임계치보다 큰 화소는 전경으로 분리된다. 배경과의 차이가 임계치보다 적다면 평균( $\mu^t$ )과 표준편차( $\sigma^t$ )는 식 (3)을 이용하여 갱신된다[7].

$$\mu^t = a^{t-1} \mu^{t-1} + (1 - a^{t-1}) C^t \text{ 과}$$

$$\sigma^t = \sqrt{a^{t-1} W + (1 - a^{t-1})(\mu^t - C^t)^2} \quad (3)$$

여기서  $W = (\sigma^{t-1})^2 + (\mu^t - \mu^{t-1})^2$ ,  $a^{t-1} = \frac{t-1}{t}$ 이다.

(2) 인체 부위 후보 영역 검출

주어진 컬러영상에서 사람을 검출하는데 있어서 가장 유용한 정보는 피부색을 이용하는 것이다. 사람은 다양한 의상 및 장신구를 착용할 수 있으므로 피부색만으로 인체 자세를 추정하는 것은 실제로 많은 어려움이 있다. 본 논문에서는 최소한 얼굴이 피부색으로 검출가능하다고 가정한다. 컬러영상이 주어지면 (그림 4(a)), RGB 색공간에서 미디언 필터로 평활화시키고, YCbCr 색공간에서 피부색 필터인 식 (4)을 이용하여 식 (5)과 같이 이진화를 시킨다[3] (그림 4(b)).

$$\text{if } (Y > 128) \begin{cases} \theta_1 = -2 + \frac{256-Y}{16}, & \theta_2 = 20 + \frac{256-Y}{16}, \\ \theta_3 = 6, & \theta_4 = -8 \end{cases}$$

$$\text{if } (Y < 128) \begin{cases} \theta_1 = 6, & \theta_2 = 12, \\ \theta_3 = 2 + \frac{Y}{32}, & \theta_4 = -16 + \frac{Y}{16} \end{cases}$$

$$Cr \geq -2(Cb + 24), Cr \geq -(Cb + 17),$$

$$Cr \geq -4(Cb + 32), Cr \geq 2.5(Cb + \theta_1), Cr \geq \theta_3,$$

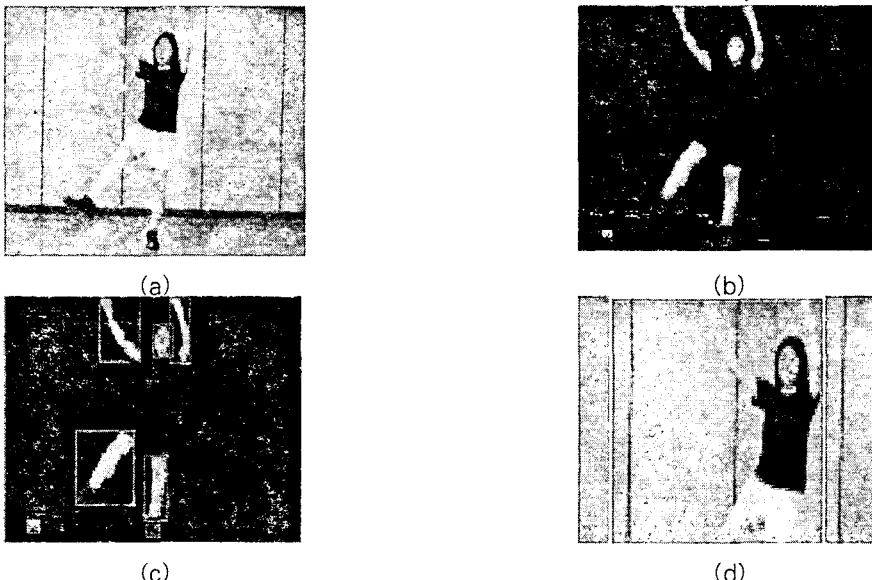
$$Cr \geq 0.5(\theta_4 - Cb), Cr \leq \frac{220 - Cb}{6}, Cr \leq \frac{4}{3}(\theta_2 - Cb) \quad (4)$$

$$I'(x, y) = \begin{cases} I(x, y) & I(x, y) \text{가 식(6)을 만족하면,} \\ 0 & \text{그렇지 않으면} \end{cases} \quad (5)$$

피부색 화소의 4-인접 화소 중 피부색 화소들을 연결하는 영역 그룹화를 통해 영상을 분할한다. 그룹화를 마치면, 주어진 영상에 피부 연결 영역들의 리스트가 만들어지고, 인체 부위 후보 영역을 결정하기 위하여 조건에 따라 인체 부위일 확률이 적은 불필요한 영역을 제거한다.

가로 또는 세로의 길이가 매우 작은 영역  
면적(= 인접 영역 내 픽셀 수)이 매우 적은 영역  
비율(= 가로/세로)이 매우 적거나 큰 영역  
밀집도(= 면적/인접 사각형의 크기)가 매우 적은 영역

그림 4(b)는 검출된 피부색 영역에서 찾아진 후보 영역들을, 그림 4(c)는 조건에 따라 불필요한



<그림 4> 인체 부위 후보 영역 검출: (a) 피부색 검출. (b) 팔, 다리, 머리 후보 영역. (c) 실험복 검출. (d) 인체 추정 영역

부분을 제거한 후 결정된 후보 영역을 보여주고 있다. 인체는 피부색 범위 주변에 존재하기 때문에, 인체의 자세 추정을 위해 전체 영상이 아닌 영상의 일부분만 검색하여 검색의 효율화를 가져온다. 인체 검색 범위는 검출된 피부색 영역들을 모두 포함하는 영역의  $\alpha (>1.0)$  배로 정의할 수 있다 (그림 4(d)).

### 3.2 일반 탐색에 의한 인체의 부위 검출

정확한 인체 자세 추정을 위한 전단계로서 색과 같은 외형 정보와 신체 비율, 인체의 위치와 같은 기하 정보와 연결 정보를 조합하여 머리, 몸통, 왼쪽과 오른쪽 팔, 왼쪽 다리, 오른쪽 다리로 이루어진 대략적인 인체 부위를 검출한다. 이 단계에서 추적된 인체의 움직임은 이전 프레임과 다음 프레임에서의 인체의 움직임의 연결 관계를 반영하고 있지 않기 때문에 추적된 움직임이 불완전하고 오차를 포함하고 있다.

영상이 주어지면, 가정한 실험 의복의 색 정보를 검출하고 영역 그룹화를 통해 영상을 분할한다. 그룹화를 마치면, 주어진 영상에 옷 색을 가진 영역들의 리스트가 만들어지고, 몸통과 다리 상, 하단

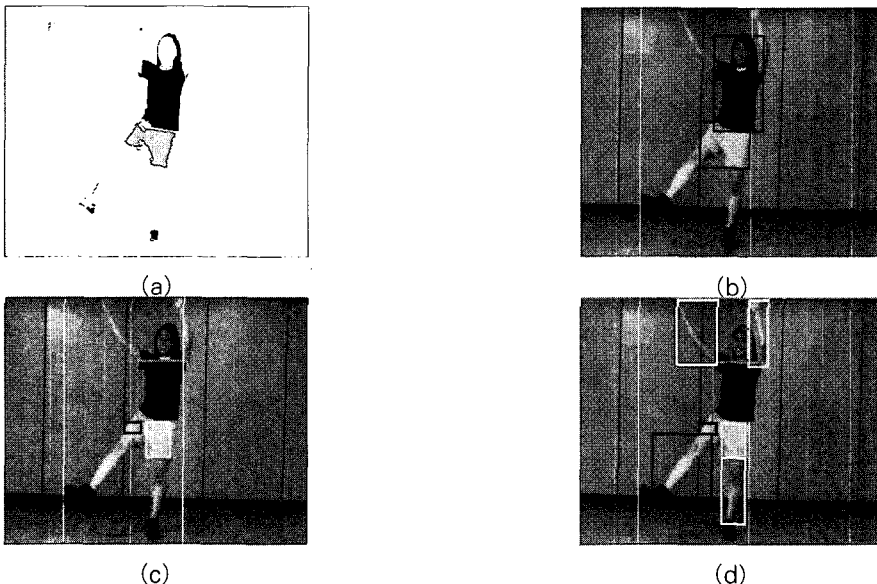
후보 영역을 결정하기 위해 영역의 면적, 크기, 압축률이 조건 이하인 불필요한 영역을 제거한다 (그림 5(a)).

머리는 영상의 상위에 있고 두 팔을 벌렸을 때, 손끝을 넘어가지 않으며 다른 인체 부위

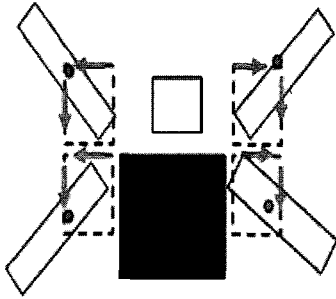
중에서도 가장 피부색이 조밀하다는 외형 정보와 몸통과 가장 가깝다는 연결 정보를 이용해 검출한다. 몸통은 후보 영역 중에서 위치적으로 상위에 위치하고 면적이 조건 이상이라는 점을 이용해 검출한다 (그림 5(b)). 검출된 인체 후보 영역 리스트를 검색하여 몸통과 연결된 팔/다리를 검출한다. 몸통과 두 다리, 두 팔의 연결 정보를 조합하여 몸통과 연결된 요소들 중 두 다리와 두 팔을 검출한다 (그림 5(c)와 (d)). 그림 5(d)는 일반 탐색에 의해 검출된 인체 모델을 보여주고 있다.

### 3.3 팔/다리의 상세한 부위 검출

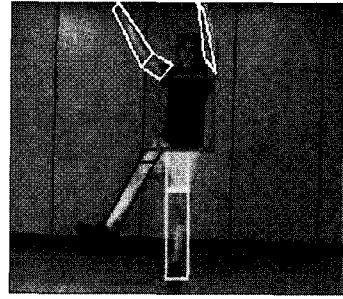
3.2 단계에서 추적된 인체의 움직임을 관절 추적과 인체 부위의 각도와 길이 정보를 조합하여 보다 상세한 인체 움직임을 추적해 낸다. 인체 부위 중 상세한 움직임 추적이 필요한 부위는 팔과 다리이다.



<그림 5> 일반 탐색에 의한 인체 부위 검출: (a) 옷 색 검출, (b) 몸통과 다리 상단 후보 영역 검출, (c) 몸통과 다리 상단 추정, (d) 대략적인 인체 모델 구성



〈그림 6〉 팔꿈치 추정



〈그림 7〉 상세한 인체 부위 검출

우선, 팔은 어깨와 손끝과 두 부위의 연결 부분인 팔꿈치로 구성된다. 이 세 부위의 움직임을 추적하고 팔의 각도를 구하여 보다 자세한 팔의 움직임 추적이 가능하게 한다. 몸통과 가장 가까운 부분이 어깨이고 몸통과 가장 거리가 먼 부분이 손끝이라고 가정한다. 팔꿈치는 어깨에서 팔 상단의 길이만큼 가로(20) × 세로(20) 영역을 검색 범위로 한다(그림 6). 팔꿈치의 검색은 안에서 밖으로, 아래에서 위로 피부색 영역을 따라가며 대략적인 경계선을 구하고 그 중심점을 구하는 것으로 이루어진다. 보다 정확한 팔 자세의 추정을 위해서 왼쪽 팔을 들었을 때와 내렸을 때, 오른쪽 팔을 들었을 때와 내렸을 때의 4가지 경우와 허리에 손을 얹은 포즈의 경우로 나누어 자세를 추정한다. 그림 6은 4가지 경우의 각각의 팔꿈치 검색 방향을 보여주고 있다. 4가지 경우의 팔 자세 중에서 본 논문에서는 왼쪽 팔의 자세 추정을 예로 들었다. 팔의 위치 정보는 팔의 모서리 네 꼭짓점을 의미한다. 네 꼭짓점 중에서 몸통의 중심과 가장 가까운 점이 어깨이고 가장 거리가 먼 점이 손끝이다. 허리에 손을 얹은 포즈의 경우는 몸통과의 연결을 고려해, 몸통의 왼쪽 상위점이 왼쪽어깨, 오른쪽 상위점이 오른쪽 어깨가 된다.

다리를 들어 올리거나 내렸을 때의 움직임은 두 다리의 y 좌표 값의 차이와 다리의 압축율의 차이를 통해 추적할 수 있다. 사람이 걷는 동작의 추적은 두 다리의 y 좌표 값의 차이가 커지게 되면 다리를 들었다는 것을 알 수 있다는 점을 이용해 이루어진다. 또한 다리를 사선으로 들었을 경우의 다리의 압축률이 다리를 곧게 세우고 있을 때보다 적

어진다는 점을 이용해 추적해 낸다. 그림 7은 팔과 다리 관절이 분리된 상세한 인체 모델의 움직임 추적을 보여준다.

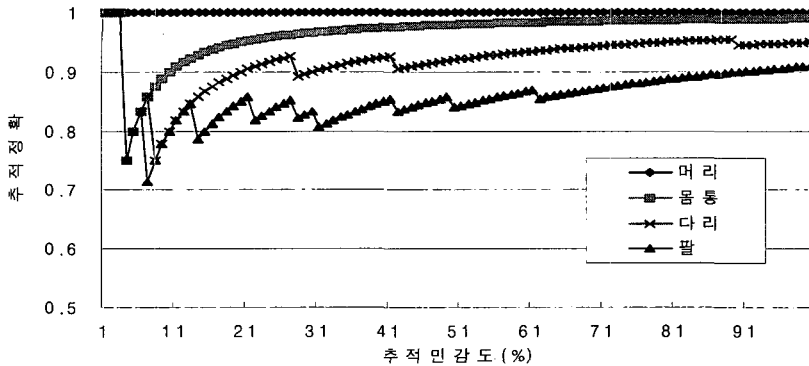
### 3.4 2-1/2차원 움직임 추출

동영상에서의 움직임 추적은 이전 프레임에서 얻어낸 정보를 이용할 수 있다는 점에서 프레임간의 연결 관계에 기반한 보다 정확한 움직임 추적이 가능하다. 현재 프레임에서 구한 인체 움직임의 위치 좌표와 이전 프레임에서의 좌표의 차이를 이용하여 각 인체 부위의 위치, 각도, 크기 정보, 색상 정보로 이루어진 2D의 인체 모델의 2D 움직임이 만들어진다.

그러나 2D 움직임으로 3D 공간에 존재하는 아바타의 동작을 애니메이션 하는 것은 매우 부자연스럽다. 본 논문에서는 깊이 정보를 구하는 대신에, 인체의 외형 정보와 연결정보를 이용하여 2-1/2D 움직임을 구한다. 일반적으로 뒤에 있는 인체 부위는 앞에 있는 인체 부위에 의해 가려질 수 있으므로 부위 매칭율이 떨어진다. 즉, 부위  $p_i$ 가 다른 부위  $p_j$ 와 겹친다면,  $p_i$ 의 깊이 정보는 아래와 같이 구해질 수 있다.

$$\begin{cases} \text{앞으로,} & \text{만일 } M(p_i) > M(p_j) \text{ 이면,} \\ \text{뒤로,} & \text{그렇지 않으면} \end{cases}$$

여기서  $M(p_i)$ 는  $p_i$ 의 매칭율로 본 논문에서는 거리 측정법을 이용하여 검출된 부위와 모델의 각 구성요소 간의 유사도를 구한다.



〈그림 8〉 인체 구성요소 모델을 이용한 추적 ROC

### 3.5 2차원 인체 움직임의 3차원 아바타 매핑

마지막으로, 동영상으로부터 구해진 인체추적 결과를 모델링한 3D 아바타에 매핑하는 단계가 있다. 우선, 추출된 사람의 크기와 위치에 따라 아바타의 크기와 위치를 정한다. 초기 프레임으로부터 인체의 부위가 추출되고 자세가 결정되면, 그림 2(b)에 보인 아바타의 자세를 초기화시킬 수 있다.

동영상에서의 추적은 각 부위의 정보가 아바타로 전달됨으로써 애니메이션으로 구현된다. 본 논문의 3D 아바타 모델의 동작을 구현하기 위해서, 각 인체 부위  $p_i$  중  $c_i$ 의 연결 가능한 면을 가진 부위들 - 목, 어깨, 무릎, 팔꿈치, 발목 -의 연결 부분이 회전 가능하여야 한다. 각 연결 부분의 회전을 위해 관절 회전값과 깊이정보  $d_i$ 을 이용하여 3D 회전을 수행함으로써 각 구성요소의 움직임을 표현한다. 이 때, 회전값과 깊이 정보에 임계치를 두어 임계치 이상의 급격한 동작이 발생하는 경우는 추적하지 않게 함으로써, 인체 부위 검출에서 생길 수 있는 오류가 아바타 동작에 영향을 미치지 않도록 한다.

## 4. 실험 결과

본 논문에서 제안된 알고리즘은 운영체제 Windows 2000XP 상에서 JAVA, JMF, JAVA 3D를 이용하여 구현되었으며, 실험은 Pentium-IV 1.8GHz인

CPU와 512MB 메모리 사양의 Windows 2000 XP 환경에서 수행되었다. 실험에 사용된 동영상은 Sony 디지털 카메라에서 획득하였다.

제안된 인체추적 알고리즘의 성능을 평가하기 위해서 추적민감도와 추적정확도로 구성된 ROC (Receiver Operating Characteristics)를 이용하였다. 추적민감도는 사람이 모델에 의해 옳게 추적될 확률을 의미하고, 추적정확도는 추적된 모델이 사람일 확률을 의미한다. 각 추적 단계에서 사람으로 추적된 모델의 수를 세고 전체 사람 수에 대한 비율을 계산하였다.

$$\text{추적민감도} = \frac{\text{추적된 사람수}}{\text{전체 추적될 사람수}}$$

$$\text{추적정확도} = \frac{\text{옳게 추적된 누적 사람수}}{\text{추적된 누적 사람수}}$$

그림 8은 인체의 각 구성요소인 머리, 몸통, 양 팔, 양 다리를 추적한 ROC 그래프를 보여준다. 각 구성요소 중 머리는 100%, 몸통은 99%, 양 팔은 91%, 양 다리는 95%의 추적율을 보이며, 전체적으로 90%의 추적율을 얻었다. 모델이 체계화 되지 않는 초기에는 그래프의 기복이 심했지만, 처리된 프레임 수가 늘어남에 따라 모델이 체계화되어 추적율이 꾸준히 증가하는 결과를 나타냈다.

표 1은 인체의 각 구성요소 별 추적율을 C. Colombo et. al[1]와 비교한 결과를 보여주고 있다. C. Colombo et. al의 추적율은 머리와 손, 팔에 대해만 측정되었으며, 신체 부위의 모양이나 회전



〈표 1〉 각 구성요소 별 추적율 비교

C. Colombo et. al [1]		제안된 방법	
머리	93.10%	머리	100%
양 손	93.05%	양 팔	91%
양 발	92.70%	양 다리	95%

을 포함하지 않은 고정된(rigid) 추적들의 평균값으로 구해졌다. 그러나 제안된 방법은 10개의 구성요소로 표현된 인체 부위 모두를 추적하였으며, 손과 발보다 복잡한 팔과 다리를 상단과 하단으로 나누어 추적하였다. 또한 제안하는 방식은 회전에 의한 모양

및 크기 변형, 겹침에 의한 변형 등을 모두 포함하고 있으므로 보다 유연하게 추적한다고 할 수 있다.

그림 9는 애니메이션 결과를 보여주고 있다. 비교적 단순한 동작에 대해 제대로 아바타의 애니메이션이 구현되었다. 그러나 인체 추적에 있어서 조명 환경이 여전히 상당한 영향을 미쳐 애니메이션 오동작의 주요 원인이 됐다.

## 5. 결 론

본 논문에서 제안하는 구성요소에 기반한 사람 동작 추적을 통한 실시간 애니메이션 구현은 마커를 부착하지 않고 동영상의 주어진 정보만으로 관절의 움직임까지 추적하였다는데 의의를 가질 수 있다. 색상 정보를 이용하여 각 부위를 검출할 때, 겹치는 부위들은 구성요소 모델을 이용하여 분리되므로 동작 추적에 있어서의 성공률을 늘릴 수 있었다.

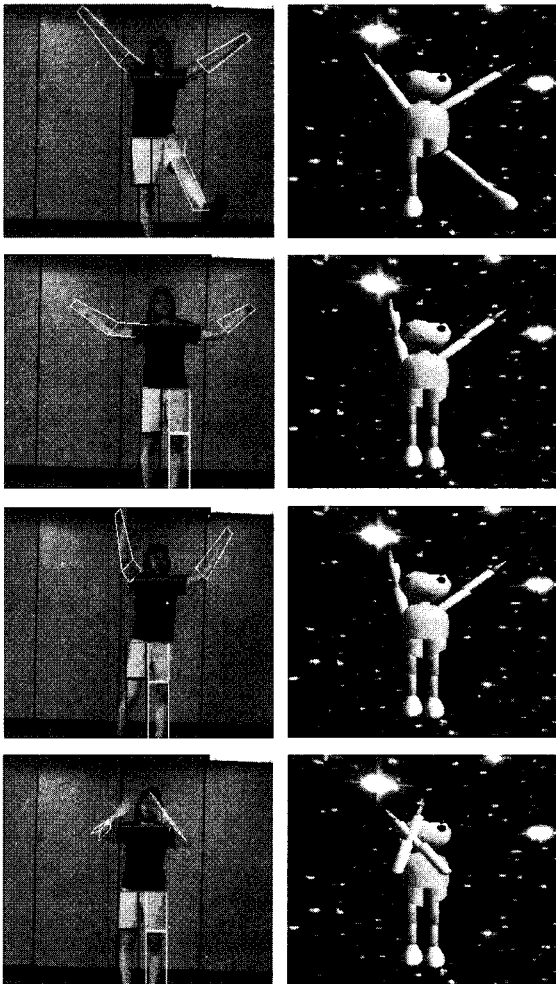
· 제안하는 구성요소에 기반을 둔 인체 모델은 인체 구성요소의 색상정보, 크기 등의 외형 정보, 유연한 연결정보, 계층적 관계정보 등을 이용한다. 색상정보를 이용하여 인체부위 또는 인체 부위 집합을 검출한 후, 외형정보와 연결정보 등을 이용하여 인체 부위를 조합하여 자세를 추정한다.

· 계층적 구조정보는 조합의 탐색 범위를 줄여 동작 추적이 시간 효율적이다.

· 조명 변화와 방향변화 때문에 인체 전신 모델을 이용하여 인체 자세를 추정하는 것은 어렵다. 제안하는 구성요소에 기반을 둔 방법은 인체 각 부위에 대한 다양한 조명이나 시점에 영향을 덜 받아 상대적으로 인체 추정이 쉽다.

· 구성요소에 기반을 둔 방법은 겹침 등으로 인하여 부분적으로 왜곡되어 모든 구성요소가 명확하게 검출되지 못하여도 적당한 인체 자세를 추정할 수 있다.

앞으로 할 일은 깊이 정보를 구하기 위해 구성



〈그림 9〉 애니메이션 결과: (a) 인체 움직임 추적, (b) 3D 아바타 매핑

요소의 위치 관계에 의해 전/후가 결정하는 대신에, 구성요소의 크기 변화 등을 이용하여 깊이 정보를 보다 정교화 시키는 것이다. 이는 아바타의 동작을 보다 사실적으로 표현하는데 도움을 줄 것이다.

## 참고 문헌

- [1] C. Colombo, A.D. Bimbo and A. Valli, "Non intrusive full body tracking for real-time avatar animation," in Proceeding of International workshop on very low bitrate video coding, pp. 119-123, 2001.
- [2] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient matching of pictorial structures," in Proceedings of CVPR, pp. 66-73, 2000.
- [3] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," IEEE Tran. on Multimedia, vol. 1, no. 3, pp. 264-276, 1999.
- [4] D.M. Gavrila, "The visual analysis of human movement: A survey," CVIU, vol. 73, no. 1, pp. 92-98, 1999.
- [5] T. Horprasert, I. Haritaoglu, C. Wren, D. Harwood, L. Davis and A. Pentland, "Real-time 3D motion capture", in Proceedings of Workshop on perceptual user interface, 1998.
- [6] S. Iwasawa, J. Ohya, K. Takahashi, T. Sakaguchi, S. Kawato, K. Ebihara and S. Morishima, "Real-time 3D estimation of human body postures from trinocular images," in Proceedings of Workshop on modelling people, pp.3-10, 1999.
- [7] K. M. Lee, and W. N. Street, "Model-based detection, segmentation and classification using on-line shape learning," Machine vision and application, vol. 13, no. 4, pp. 222-233, 2003.
- [8] G. Mori & J. Malik, "Estimating Human Body Configurations using Shape Context Matching", in Proceedings of ECCV, pp. 666-680, 2002.
- [9] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-time tracking of the human body," IEEE trans. on PAMI, vol. 19, no. 7, pp. 780-785, 1997.
- [10] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and applications," Computer vision and image understanding, vol. 61, no. 1, pp. 38-59, 1995.
- [11] M. Leung and Y. Yang, "First sight: A human body outline labeling system," IEEE trans. on PAMI, vol. 17, no. 4, pp. 359-377, 1995.
- [12] D. Gavrila and L. Davis, "3-D model-based tracking of humans in action: a multi-view approach," in Proceedings of IEEE conference on computer vision and pattern recognition, pp. 73-80, 1996.

## ○ 저자 소개 ○



### 이 경 미 (Kyoung-Mi Lee)

1993년 덕성여자대학교 전산학과 졸업(학사)  
 1996년 연세대학교 대학원 전산학과 졸업(석사)  
 2001년 Univ. of Iowa, Computer Science dept. 졸업(박사)  
 2003~현재 덕성여자대학교 컴퓨터공학부 교수  
 관심분야 : (동)영상처리, (동)영상 검색, HCI, etc.  
 E-mail : kmlee@duksung.ac.kr