

# 사람 자세 추정을 위한 모델 기반 추적

이경미, 김혜정, 이윤미  
덕성여자대학교 컴퓨터공학부  
kmlee@duksung.ac.kr

## Model-based tracking for human posture estimation

Kyoung-Mi Lee, Hye-Jeong Kim and Youn-Mi Lee  
Department of Computer Science, Duksung Women's University

### 요약

동영상에서의 움직임 추적은 이전 프레임에서 얻어낸 정보를 이용할 수 있다는 점에서 프레임간의 연결 관계에 기반한 움직임 추적이 가능하다. 그러나 사람의 신체는 고정된 형태를 가지고 있지 않기 때문에 프레임 간의 단순한 연결 관계만으로 사람의 자세를 추정하고 움직임을 추적하는 것은 매우 어려운 문제이다. 본 논문에서는 구성요소에 기반한 인체 모델을 이용하여 이전 프레임에서 찾은 블랍들을 연속된 프레임에서 찾은 블랍들로 연결함으로써, 동영상에서 사람의 자세를 추적하는 방법을 제안한다. 주어진 모델에 따라 이전 블랍은 대응되거나, 여러 블랍으로 나뉘거나, 다른 블랍들과 결합되어 사라지거나, 새로 생성되는 등의 4 가지 경우로 나뉘어 질 수 있는데, 각 경우에 대한 처리 방안을 제안하였다. 제안된 방법은 인체들과 블랍들의 리스트 처리를 간단하게 할 뿐만 아니라, 추적의 전처리 과정으로 블랍화를 옹게 수행해야 하는 부담을 덜어주어 과도한 블랍화와 부족한 블랍화 등의 문제를 해결할 수 있다.

Keyword : 구성요소 기반 모델, 인체 추적, 사람 자세 추정

## 1. 서론

최근 사람의 자세를 추정하고 추적하는 것이 컴퓨터 시각 분야에서 중요한 문제로 대두되고 있다. 시각에 기반한 인체 움직임 캡처 방법으로 색상을 이용하여 영상을 분할한 후 윤곽선을 분석하여 비슷한 색상으로 연결된 블랍들의 집합으로 인체를 표현하고 단일 카메라로 추적하는 방법이 소개되었다 [4]. 네 대의 카메라에서 획득된 영상에 템플릿 매칭이 이용되거나, 직각으로 배치된 세 대의 카메라에서 획득된 영상에 윤곽선 분석을 실행하는 방법들이 알려져 있다.

일반적으로 동영상에서의 움직임 추적은 이전 프레임에서 얻어낸 정보를 이용할 수 있다는 점에서 프레임간의 연결 관계에 기반한 보다 정확한 움직임 추적이 가능하다. 현재 프레임에서 구한 인체

의 부위에 대한 위치 좌표와 이전 프레임에서의 좌표의 차이를 이용하여 각 인체 부위의 위치, 각도, 크기 정보, 색상 정보로 이루어진 인체 모델의 움직임이 만들어진다.

그러나 사람의 신체는 고정된 형태를 가지고 있지 않기 때문에 동영상에서 사람을 추적하는 것은 여전히 어려운 문제들을 포함하고 있다. 예를 들어, 블랍을 구성한 후에 이전 프레임에서 현재영상까지 연속된 프레임들 사이의 거리를 계산함으로써 블랍을 배치하는 블랍 기반 추적 방법은 각 프레임에 서로 다른 개수의 블랍으로 인하여 사람의 자세를 정확히 추적하는 것이 매우 어렵다. 블랍은 분할 또는 합병되거나, 심지어 사라지거나 새로 만들어질 수 있다. 이런 문제를 해결하기 위해서 다대다(many-to-many) 블랍 배치가 적용될 수 있다 [2]. Y. Huang 와 T.S. Huang 는 시간  $t-1$  에서

의 블랍이 시간  $t$  에서의 블랍과 관계되는 것을 피하기 위해 동시에 다수의 블랍을 연계하는 다양한 다중 에이전트 추적 프레임워크를 적용하였다.

본 논문에서는 인체모델에 기반하여 블랍들의 배치를 초기화한 후, 블랍들이 연속된 프레임들 사이에 대응될 때 뿐만 아니라, 나뉘거나 다른 블랍과 합쳐지거나 사라지는 경우에도 처리할 수 있는 방법을 제안한다.

## 2. 구성요소 기반의 인체 모델

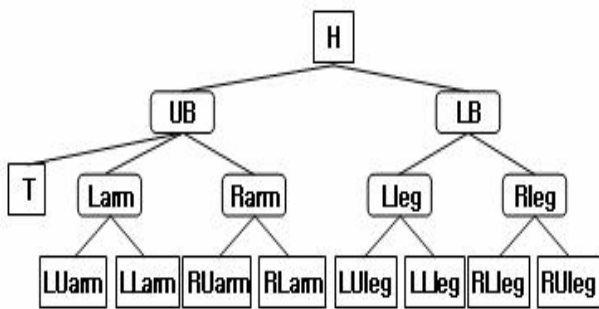


그림 1 10 개의 구성요소로 이루어진 계층적 인체 모델. 사각형은 실제적인 구성요소들을, 둥근 사각형은 계층적 관계를 정의하기 위한 가상의 구성요소를 의미한다.

본 논문에서 사용되는 인체 모델은 구성요소로서 10 개의 인체 부위로 구성되고, 계층적 방법으로 구성요소들을 연결한다 (그림 1) [5]. 각 구성요소  $C_i$  는 기하학적 정보 (위치, 상대적인 크기, 형태)  $g_i$  와 외형적인 정보 (구성요소 내 화소 값에 대한 평균 및 표준편차)  $a_i$  를 포함한다. 각  $C_i$  는 사각형으로 표현되고, 4 면 중 연결가능한 면, 연결되어지는 다른 구성요소의 이름, 연결 각도, 연결 거리 등과 같은 연결 정보  $R_i$  를 가지고 있다. 인체 모델은 아래와 같이 표현된다:

$$C_i = (g_i, a_i, R_i), \quad i=1, \dots, I$$

여기서  $I$  는 10 이다.

## 3. 모델 초기화

사람추적은 추적되는 사람이 비디오에 처음 나타

나기 시작할 때 인체모델을 초기화하는 것으로 시작될 수 있다. 본 논문에서는 전경영상으로부터 유사 색상의 영역을 구성하고, 계층적 인체 모델을 설계하여 추적에 이용할 것이다.

사람을 검출하기 위해서 우선 배경 프레임들의 평균 영상과 사람이 등장한 영상과의 차영상을 구해 전경 영상을 추출한다. 배경 프레임의 평균 영상을 차연산에 이용하여 조명 변화에 강건한 전경 영상의 추출이 가능하다.

분리된 전경 영상의 화소는 각 화소와 인접 화소간의 밝기 유사도를 계산하는 연결요소 알고리즘

을 사용하여 블랍으로 모아진다. 각 블랍  $B_j$  는

블랍의 색상, 크기, 중심점, 경계 상자, 인체를 구성하는 경계 등의 정보를 포함한다. 그러나 이렇게 밝기를 이용한 군집화는 색상과 빛의 변화로 자동 블랍 형성에 많은 문제를 야기하므로 개별적인 인체 부위를 완벽하게 구성하는 것은 매우 어렵다. 특히 초기 군집화는 거의 밀착된 블랍의 표면에 존재하는 상당한 조명 변화로 과잉 분할 (over-segment) 되기 쉬운데, 이를 해결하기 위해서 비슷한 색상을 가진 이웃 블랍들과 함께 합쳐질 것이다. 만약 생성된 블랍이 너무 작거나, 길거나, 이질적으로 비조밀하면 지워진다.

사람은 인체 부위에 대응되는 블랍의 부분집합으로 정의될 수 있으므로, 프레임 내의 블랍들은 사람의 자세를 추적하기 위해 대응되는 인체 부위에 매핑되어야 한다. 인체를 구성하는 알고리즘은 우선 블랍 인접 그래프와 각 블랍의 경계상자사이의 최소거리를 계산한다. 경계상자의 꼭지점들 사이의 거리는 화소 단위에서의 인접도보다 훨씬 빠르게 계산되어진다 [1].  $P_0$  가 블랍  $B_j$  의 부분집

합이라고 하자.  $P_0$  와 인접그래프에서 시작하여 반복적으로 만들어진다. 집합  $P_1$  은  $P_0$  의 적절한 인접 블랍과 합쳐짐으로써 얻어질 수 있다. 다음은, 합쳐진 블랍  $P_k$  의 각 새로운 집합은 초기 집합인  $P_0$  와 합쳐진 블랍  $P_{k-1}$  의 집합을 합침으로써 얻어진다. 결국 후보 인체  $CP$  은 합쳐진 블랍

$P_k$ 의 집합들을 포함한다. 즉,  $CP = \cup_{k=0}^k P_k$ 이다.

블랍을 인체 부위에 매핑시키기 위해서, 본 논문에서는 계층적 인체모델을 이용하였다 (그림 1). 인체 모델의 각 구성요소  $C_i$ 가 가지고 있는 기하학적 정보  $g_i$ 와 외형적인 정보  $a_i$ , 연결 정보  $R_i$ 를 이용해 인체 부위의 위치를 추정한다. 이 단계에서 추정된 인체의 움직임은 이전 프레임과 다음 프레임에서의 인체의 움직임의 연결 관계를 반영하고 있지 않기 때문에 추적된 움직임이 불완전하고 오차를 포함하고 있다. 특히, 팔과 다리처럼 계층적인 구조를 내포하고 있는 구성요소들은 일반-상세 탐색을 통해 분리된 관절을 중심으로 인체의 자세를 초기화 시킬 수 있다 [5].

#### 4. 적응적 인체 모델링

사람은 인체 모델을 이용하여 추적되기 때문에, 인체 모델의 정보는 다수의 사람을 추적하기 위해 저장되어야 한다. 사람의 모든 동작이 프레임마다 상대적으로 작기 때문에, 색상기반 모델에서의 큰 변화는 추적을 쉽게 실패하게 만드는 원인이 된다. 이런 색상모델 추적의 민감성을 해결하기 위해서, 프레임의 블랍들을 인체 참조모델과 비교한다.

$t-1$  번째 프레임까지 만들어진 인체 모델  $CP^{t-1}$ 은 평균 ( $\mu^{t-1}$ )과 표준편차 ( $\sigma^{t-1}$ )에 의해 표현되고,  $t$  번째 프레임은 새로운 블랍  $B^t$ 과 그들 간의 관계  $Br^t$ 로 구성된다고 하자. 인체모델  $CP^{t-1}$ 과 새 블랍  $B^t$  사이의 최소 거리는 다음과 같이 계산된다.

$$d^t = \min_{j=1 \dots k} \left( \frac{\|B^t - \mu^{t-1}\|_p}{\sigma^{t-1}} \right) + \min_{j=1 \dots l} \left( \frac{\|Br^t - \mu^{t-1}\|_p}{\sigma^{t-1}} \right) \quad (1)$$

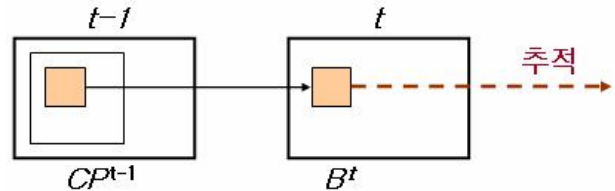
여기서  $\mu^{t-1}$ 는 시간  $t-1$ 에서  $j$  번째 신체 부위 내 블랍과 관계들의 평균을,  $\sigma^{t-1}$ 는 표준 편차를 나타낸다. 최소거리가 미리 정해진 임계치보다 적으면, 제안된 모델링 알고리즘은 블랍  $B^t$ 와 관계  $Br^t$ 를 적응적 인체모델( $\mu^{t-1}$ 와  $\sigma^{t-1}$ )에 추가하고, 그들의 중심과 불확실성을 다시 계산함으로써 적응적 모델을 갱신한다 [3]. 유사도 임계치는 실

험적으로 정해지고, 사용자에게 의해 조절될 수 있다.

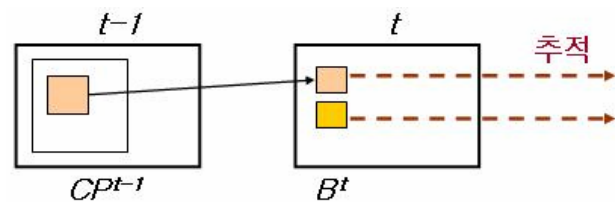
#### 5. 모델 기반 동작 추적

본 논문에서는 이전 연구에서 설계한 구성요소에 기반한 인체모델  $CP$ 을 이용하여  $t-1$  번째 프레임까지 인체  $CP^{t-1}$ 가 이미 추적되었고, 프레임  $t$ 에 새로운 블랍  $B^t$ 들이 구성되었다고 가정한다. 이후 사람의 움직임은 아래와 같은 방식으로 추적된다.

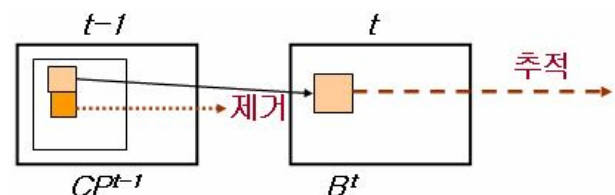
Case 1. 만일  $B^t$ 가  $CP^{t-1}$ 에 포함된다면,  $CP^{t-1}$ 내 대응되는 블랍은  $B^t$ 으로 추적된다.



Case 2. 만일  $CP^{t-1}$  내 블랍이  $t$  번째 프레임에서 여러 블랍으로 분리되었다면,  $CP^{t-1}$  내 블랍은  $t$  번째 프레임 내 한 블랍으로 추적되고  $t$  번째 프레임 내 다른 블랍들은  $CP^{t-1}$ 에 같은 구성요소로써 추가된다.

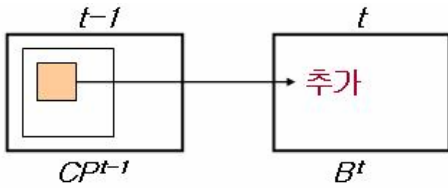


Case 3. 만일  $CP^{t-1}$  내 블랍이  $t$  번째 프레임에서 블랍  $B^t$ 으로 합쳐진다면,  $CP^{t-1}$  내 블랍은  $t$  번째 프레임 내 한 블랍으로 추적되고  $t$  번째 프레임 내 다른 블랍들은  $CP^{t-1}$ 에서 제거된다.



Case 4. 만일  $B^t$ 가  $CP^{t-1}$ 에 포함되지만 대응되는

블랍이 존재하지 않으면,  $B^t$  를  $CP^{t-1}$  에 가장 유사한 구성요소를 찾아 추가한다.



여기서 영역이 경계상자에 의한 인체에 포함된다 는 것은 그 영역이 경계상자의 90% 이상 겹침을 말한다. 블랍과 적응적 인체모델의 대응은 (식 1) 에 의해 계산된다. 제안된 방법은 인체들과 블랍 들의 리스트 처리를 간단하게 할 뿐만 아니라, 기존의 사람이 비디오 프레임에서 나가거나, 새로운 사람이 비디오 프레임 안으로 다시 들어오거나, 또 이전에 비디오 프레임에서 나갔던 사람이 다시 들어오거나를 계속해서 확인할 수 있게 한다. 제안된 방법의 또 다른 장점은 옳게 블랍화하는 것 에 대한 부담을 덜어준다. 블랍이 조명 변화에 의 해 놓쳐져도 모델에 기반한 추적은 다른 존재하는 블랍을 이용하여 개별적인 사람을 유지할 수 있게 한다. 인체가 구성된 후, 인체 내 블랍의 개수는 유연하게 변경가능하다. 제안된 방법은 과도한 블 랍화 (over-blobbing, Case 2)와 부족한 블랍화 (under-blobbing, Case 3)을 처리할 수 있다.

## 5. 실험 결과 및 결론

본 논문에서 제안된 알고리즘은 운영체제 Windows 2000XP 상에서 JAVA 와 JMF 를 이용하여 구현되었으며, 실험은 Pentium-IV 1.8GHz 인 CPU 와 512 MB 메모리 사양의 Windows 2000 XP 환경에서 수행되었다. 실험에 사용된 동영상은 Sony 디지털 카메라에서 획득하였다. 그림 2 는 사람 자세 추정 및 추적의 결과를 보여주고 있다. 본 논문에서 제안하는 사람 자세 추적을 위한 모 델기반 추적은 색상을 이용한 영상의 블랍들을 인 체 모델의 계층적 구조를 이용하여 자세를 추정하 는 초기화를 수행한 후, 프레임간의 연결관계와 모델정보를 이용하여 자세를 추적하는 방법이다. 제안하는 방법은 정확한 블랍을 구성해야 하는 부

담을 줄여줌으로써 추적의 효율을 높였다. 그러나 블랍을 인체의 부위에 매핑시키는 초기화 과정의 결과가 추적의 정확도에 영향을 미치므로 초기화 과정 없이 추적하는 방법을 연구할 계획이다.

## Acknowledgement

이 논문은 2005 년도 정부(교육인적자원부)의 재원 으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (R04-2003-000-10092-0 (2005)).

## 참고문헌

- [1] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE trans. on Multimedia*, 1(3):264-276, 1999.
- [2] Y. Huang and T. S. Huang, "Model-based human body tracking," in *Proc. of ICPR*, pp.552-555, 2002.
- [3] K. M. Lee, and W. N. Street, "Model-based detection, segmentation and classification using on-line shape learning," *Machine vision and application*, 13(4):222-233, 2003.
- [4] G. Mori & J. Malik, "Estimating human body configurations using shape context matching", in *Proc. of ECCV*, pp. 666-680, 2002.
- [5] 이주현, 김상희, 이경미, "실시간 아바타 애니 메이션을 위한 구성요소기반 인체 추적", *HCI 2005 학술대회 발표논문집*, pp. 830-835, 2005.

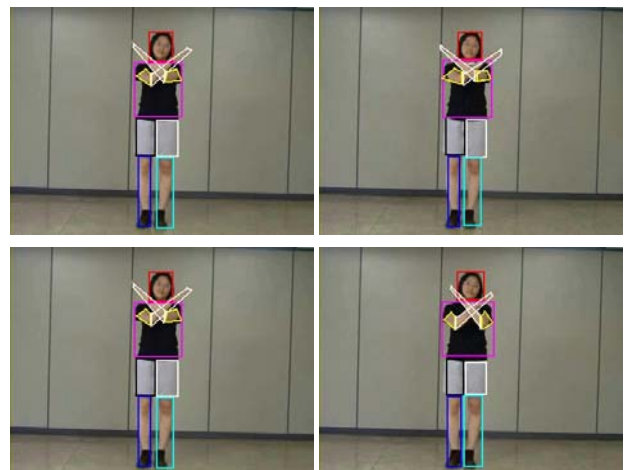


그림 2 4 개의 연속 프레임에서의 추적 결과