

## 구성요소기반 온라인 학습을 이용한 인체 자세 추정

이경미, 김혜정  
 덕성여자대학교 컴퓨터공학부  
 {kmlee, kimhj}@duksung.ac.kr

### Estimation of human posture using component-based online learning

Kyoung-Mi Lee, Hye-Jung Kim  
 Dept. of Computer Science, Duksung Women's University

#### 요약

주어진 영상에서 인체를 찾고 그 자세를 인식하기 위해 자세나 조명 조건의 변화에 덜 민감한 방법으로 구성 요소에 기반한 접근이 있다. 본 논문에서는 10개의 구성요소와 그들간의 유연한 연결로 구성된 인체모델을 사용한다. 각 구성요소는 기하학적, 명시적, 다른 구성요소와의 연결요소에 대한 정보로 구성되어 있다. 인체 구성요소 사이의 계층적 연결은 일반-상세 탐색으로 시간효율적인 인체 매칭을 가능케 한다. 본 논문에서는 새로운 인체를 찾을 때마다 인체 구성요소를 갱신함으로써 자세 및 조명 변화에 보다 적응적으로 자세를 추정하는 방법을 제안한다.

#### 1. 서론

최근 인체 자세를 추정하는 것이 컴퓨터 시각 분야에서 중요한 문제로 대두되고 있다. 인체의 전신을 이용하여 자세를 추정하는 방법은 자세나 조명 변화에 영향을 매우 많이 받기 때문에, 인체를 표현하기 위해 인체의 각 부위와 그들 사이의 연결 관계를 고려하는 방법이 널리 적용되고 있다. Fischler 과 Eischlager 는 물체를 구성하는 구성요소를 사이의 기하학적 배치로 연결하는 회화구조를 소개하였다[1]. Felzenszwalb 와 Huttenlocher 은 회화구조를 이용하여 구성요소들의 위치 관계를 다양한 구성요소로 표현하였다[2]. 구성요소 쌍은 그들 사이의 상대적 위치에 근거하여 만들어진 시각적 스프링으로 연결된다.

본 논문에서는 영상에서 인체의 주요부위와 연결부분을 찾은 후에 인체 구성을 매칭함으로써 인체의 자세를 추정한다. 인체의 각 부분은 형태, 위치, 연결을 포함하는 구성요소로 표현되고, 각 부위는 계층적인 방법으로 연결된다. 다음엔, 효율적인 인체 매칭을 위해 일반-상세 탐색이 사용된다. 또한, 본 논문은 새로운 예제 데이터가 들어올 때마다, 각 구성요소를 적응적으로 갱신시킴으로써 자세와 조명의 변화에 관련된 인체자세를 추정하였다.

#### 2. 구성요소 모델을 이용한 매칭

본 장에서는 계층적 방법으로 인체 부위를 연결하는 인체모델과 그 인체모델로 영상의 인체를 매칭하는 방법을 설명한다. 인체모델의 각 구성요소는 일반 탐색을 통해 머리, 몸통, 팔/다리 등으로 분류된 후, 상세한 분류를 통해 인체부위의 매칭을 완성한다. 또한, 구성요소들은 새

로운 인체 데이터가 들어올 때마다 온라인으로 갱신된다.

#### 2.1 구성요소를 이용한 인체 모델

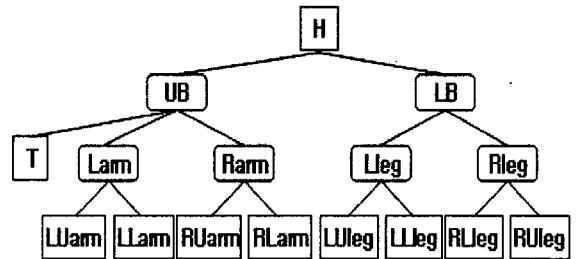


그림 1 구성요소를 이용한 계층적 인체 모델은 머리(H), 몸통(T), 왼쪽 팔 상단(LUarm), 왼쪽 팔 하단(LLarm), 오른쪽 팔 상단(RUarm), 오른쪽 팔 하단(RLarm), 왼쪽 다리 상단(LUleg), 왼쪽 다리 하단(LLleg), 오른쪽 다리 상단(RUleg), 오른쪽 다리 하단(RLleg)으로 구성되어 있다. 각각은 10개의 실제적인 구성요소들을 의미한다. 동근 각각은 계층적 관계를 정의하기 위한 가상적 구성요소들로 상반신(UB), 하반신(LB), 왼쪽 팔(Larm), 오른쪽 팔(Rarm), 왼쪽 다리(Lleg), 오른쪽 다리(Rleg)이다.

본 논문에서 사용되는 인체모델은 구성요소로서 10개의 인체 부위로 구성되고, 계층적 방법으로 구성요소들을 연결한다(그림 1). 각 구성요소는 기하학적 정보(위치, 상대적인 크기, 형태)와 외형적인 정보(구성요소 내 화소 값에 대한 평균 및 표준편차)를 포함한다. 게다가, 각 부분은 각 부분을 표현하는 4 부분 중 연결 가능한 면, 연결

되어지는 다른 부분의 이름, 연결 각도, 연결 거리와 같은 연결에 대한 정보를 가지고 있다. 인체 모델은 아래와 같이 표현된다:

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

여기서 Human 는 I (=10)개의 부분들로 구성된 인체 모형이며,  $g_i$ 와  $a_i$ 는 각 부위의 기하학적, 외형적 정보를 나타낸다.  $R_i$ 는 각 부위 사이의 연결정보를 의미한다.

본 논문에서 사용되는 인체 모델은 계층적 관계로 구성된다. 이런 계층적 관계를 이용하는 것은 매칭 탐색공간을 제한하기 때문에 시간적인 효율을 향상시킨다. 특히, 논문 [2]에서 사용된 별 구조의 모델과는 달리, 일반적인 트리구조의 인체모델을 사용한다. 우선 인체모델은 일반 탐색을 통해 머리, 몸통, 팔, 다리 등으로 분류되고, 다음으로 보다 자세한 분류로 인체 부위의 매칭을 완성한다.

### 2.2 계층적 탐색을 이용한 매칭

구성요소를 이용한 인체모델로 자세를 추정하는 것은 찾은 구성요소를 결합하여 적당한 인체모델을 구성하는 매칭 과정이다. 인체모델 (식 (1))은 매칭되지 않은 확률을 최소로 하는 결합을 찾아 매칭된다.

$$d_i = \sum_j \min_j \|g_i - p_j\| + \sum_j \min_j \|a_i - q_j\| + \sum_j \min_j \|R_i - r_j\| \quad (2)$$

$$HumanScore = \sum_{i=1}^I d_i$$

여기서  $(p_j, q_j, r_j)$ 는  $j$ -번째 구성요소의 기하학적, 외형적 정보와 다른 구성요소와의 관계를 나타낸다.

본 논문에서는 일반-상세 탐색을 적용한다. 이 탐색법은 탐-다운 모델인 트리 구조로 구성요소들 사이의 관계를 제한하므로, 다른 구성요소에 탐색 범위를 제한시킨다. 예를 들어, 몸통을 찾은 후, 왼쪽 팔의 상단과 하단에 대한 상세한 탐색이 이루어진다. 왼쪽 팔의 매칭 확률,  $d_{Larm}$ ,은 왼쪽 팔의 상/하단에 대한 매칭 확률 ( $d_{LUarm} / d_{LDarm}$ )의 합으로 구해질 수 있다:

$$d_{Larm} = d_{LUarm} + d_{LDarm}$$

다른 부위들도 이러한 계층적 기법을 반복적으로 적용하여 찾을 수 있다.

### 2.3 구성요소기반 온라인 학습

인체 모델의 각 구성요소는 자세나 조명의 변화에 따라 바뀔 수 있으므로, 기하학적 정보와 외형정보가 변경되는 것이 필요하다. 인체 데이터들이 주어진다면, 모델은 간단하게 각 구성요소의 평균,  $\mu$ ,과 표준편차,  $\sigma$ ,를 계산함으로써 갱신될 수 있다. 표준 편차를 이용하여 모델의 각 구성요소와 연결에 대한 불확실성을 구할 수 있으므로 [7], 인체모델(식(1))을 불확실성을 포함한 유연한 모델이 될 수 있다.

그러나, 이런 배치형태의 갱신은 온라인으로 모델 매칭을 처리하는 경우에는 적당하지 않다. 따라서, 새 데이터가 들어올 때 마다 모든 이전의 데이터들을 모으는 대신

에, 새 데이터만으로 인체 모델을 확장시킬 수 있는 방법이 유용하다.  $n$ -번째 데이터,  $h_n$ ,가 모델화되면, 인체모델의  $i$ -번째 구성요소의 평균과 표준편차는 다음과 같이 갱신된다:

$$\mu_{i,n} = \frac{\mu_{i,n-1} + h_{i,n}}{n} \quad \text{과} \quad \sigma_{i,n} = \sqrt{\frac{(n-1)\mu_{i,n-1} + (\mu_{i,n-1} - h_{i,n})^2}{n}}$$

$$u_{n-1} = (\sigma_{i,n-1})^2 + (\mu_{i,n-1})^2$$

여기서  $h_i$ 는  $i$ -번째 구성요소의 기하학적,  $g_i$ , 또는 외형적,  $a_i$ , 정보를 의미한다. 식 (1)과 (2)는 각각 아래와 같이 변경된다:

$$Human_i = [(\mu_i^g, \sigma_i^g), (\mu_i^a, \sigma_i^a), R_i] \quad \text{과} \\ d_i = \sum_j \min_j \left\| \frac{\mu_i^g - p_j}{\sigma_i^g} \right\| + \sum_j \min_j \left\| \frac{\mu_i^a - q_j}{\sigma_i^a} \right\| + \sum_j \min_j \|R_i - r_j\|$$

따라서,  $n$  개의 데이터를 학습시킨 후에, 각 구성요소는 구성요소 공간에서 대응되는 인체부위의 통계적 정보를 나타낸다. 이러한 구성요소의 학습은 자세와 조명이 변하는 경우에 인체 자세를 적응적으로 추정하는 것을 가능하게 만든다.

### 3. 인체 자세 추정

구성요소 모델을 이용한 자세 추정은 인체의 각 부위에 할당된 구성요소를 포함하고 있는 모델을 만들어가는 과정이다. 본 장에서는 영상에서 인체를 찾아 부위를 매칭 시켜가는 과정을 소개한다.

영상이 주어지면, 사람을 찾는데 가장 유용한 정보인 피부색을 찾는다. 그러나 사람들은 서로 다른 의상과 장신구를 착용할 수 있기 때문에, 실제로 피부색만으로 인체 자세를 추정하는 것은 매우 어려운 일이다. 본 논문에서는 최소한 얼굴은 피부색을 이용하여 검출할 수 있다고 가정한다. 컬러 영상에 중간값 필터와 피부색 필터[3]을 차례로 적용하여 피부색 화소와 그렇지 않은 화소로 구성된 이진 영상을 만든다. 이런 이진 영상은 연결된 피부색 화소로 그룹 지어져 영역으로 분할된다 [4]. 그룹화로 만들어진 피부색 영역들 중에서 인체가 될 확률이 적은 불필요한 영역들을 제거시킨다.

인체는 피부색 영역들 주변에 위치할 것이다; 이것은 전체 영상에서 관련 있는 부분만 탐색하게 하여 보다 효율적으로 탐색할 수 있다. 인체 탐색 범위는 모든 피부색 영역을 포함하는 면적의  $\alpha$  ( $\alpha > 1.0$ )배로 정의될 수 있다.

정확한 인체자세를 추정하기 전에, 연결 관계로 구성요소를 결합하고, 인체 부위들에 대한 대략적으로 탐지하게 된다. 만일 의상에 대한 정보가 주어진다면, 인체모델은 추정된 인체 영역에서 의상 색상을 찾는다. 의상색상을 갖는 영역들을 그룹화로 찾은 후, 면적, 크기, 사각형 정도 등의 조건을 만족시키지 못하는 불필요한 영역을 제거시킨다.

머리는 보통 영상의 위 부분에 위치하며, 팔의 전체길이보다 짧고 다른 인체부위에 비교하여 조밀한 피부색 영역을 가진다. 영상에서 머리를 찾기 위해, 형태 데이터와

몸통과 가까이 위치하고 있다는 연결 데이터가 사용된다. 몸통은 다른 후보영역들에 비해 영상의 위에 존재하고 주어진 크기 조건보다 영역의 크기가 더 크다. 인체 후보영역들을 탐색하면서 몸통에 연결된 팔/다리를 찾는다. 몸통과 네 팔/다리를 연결하여 몸통에 연결된 두 팔과 두 다리를 찾는다.

대략적인 인체모델을 찾은 후에, 인체 부위의 길이와 연결 각도 등을 더해 보다 상세한 자세 추정이 이루어진다. 대부분의 경우에 상세한 탐색은 팔과 다리를 찾기 위해 필요하다. 본 논문에서는 팔을 상세한 탐색으로 찾는 경우만을 다룬다.

팔은 어깨와 팔의 끝, 그리고 이 둘을 연결하는 팔꿈치로 구성된다. 이들 세 부분을 팔로 결합하는 것은 팔의 상단과 하단의 자세를 추정하게 한다. 팔꿈치는 몸통에서 가장 가까운 곳에서 팔의 상단의 길이만큼 떨어진 곳까지 탐색 공간에서 찾는다. 팔꿈치 탐색은 다음 두 단계로 구성된다: (i) 피부색 영역을 안에서 바깥으로, 위에서 아래로 움직이며 대략적인 경계선을; (ii) 그리고 경계선의 중심점의 위치를 찾는다. 팔 자세의 보다 정확한 추정은 다섯 경우로 나누어 생각할 수 있다- 각 팔을 개별적으로 올리거나 내린 네 가지 경우와 손이 다른 쪽 팔 위에 겹치는 경우. 팔의 위치는 팔의 네 꼭지점으로 나타난다. 몸통에서 가장 가까운 점은 어깨일 것이고, 반면 가장 먼 쪽은 팔의 끝일 것이다. 몸통에 연결되면, 한 손이 다른 팔 위에 놓일 때 몸통의 상위 좌측 점이 왼쪽 어깨가 되고, 상위 우측 점이 오른쪽 어깨가 된다. 다리에 대해서도 이러한 상세 탐색을 적용할 수 있다.

4. 실험 결과 및 결론

제안된 알고리즘은 Sony DSC-P10로 획득된 109개의 영상(420×316)으로 테스트되었다. 제안된 방법은 계층적 탐색으로 인체 자세를 추정하는데, 109개의 영상에 대해 평균 0.615 초의 처리 시간을 가졌다. 이는 3-5초가 걸린 논문 [5]와 10초가 걸린 논문 [6]과 비교하여 약 4.88~16.26배 시간 효율적이다.

매칭 정확도는 매 데이터가 갱신된 후에 측정되었다. 그림 2는 인체자세를 추정하기 위해 인체부위를 매칭한 결과를 보여주고 있다. 머리, 몸통, 팔, 다리에 대해 각각 100%, 97.22%, 89.81%, 92.59%의 매칭율을 얻었다.

모든 인체 부위가 옳게 매칭된 경우만 올바르게 자세가 추정되었다고 가정한다면, 88.07%의 인체 자세 추정율을 구했다. 그림 2는 시스템이 보다 많은 데이터를 처리할수록, 영상에서 인체부위를 옳게 매칭할 확률이 증가하게 된다. 처음 몇 개의 데이터를 처리한 후에는 매칭율이 안정됨을 보여준다.

본 논문에서는 구성요소를 이용한 인체모델의 계층적 관계와 학습을 통해 자세와 조명변화에 적응적으로 인체의 자세를 추정하는 방법을 제안한다. (i) 구성요소 방법은 부분 왜곡 등의 이유로 모든 구성요소를 찾지 못하여도 적절한 인체 자세를 추정하는데 유용하다. (ii) 계층적

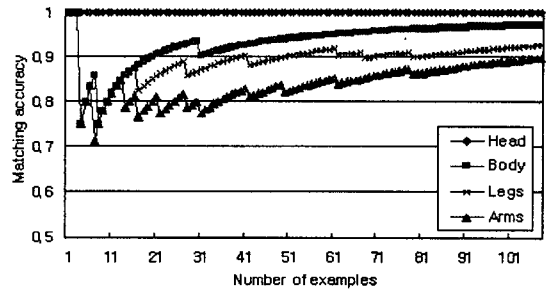


그림 2 자세 추정을 위한 인체 각 부위의 매칭율

자료구조는 가능한 결합에 대한 탐색 범위를 제한하기 때문에 인체자세 추정을 시간 효율적으로 수행할 수 있다. (iii) 게다가, 제안하는 구성요소 기법은 조명 환경이나 자세에 따른 변화에 덜 영향을 미치기 위해 많은 데이터를 학습시킬수록 매칭이 향상되는 적응적 모델을 사용한다.

Acknowledgement

본 연구는 한국과학재단 목적기초연구(R04-2003-000-10092-0 (2004)) 지원으로 수행되었음.

References

1. Fischler, M.A. and Elschlager, R.A., The representation and matching of pictorial structures, IEEE tran. on Computers, 22(1) (1973) 67-92.
2. Felzenszwalb, P.F. and Huttenlocher, D.P., Efficient matching of pictorial structures, in Proc. of CVPR (2000) 66-73.
3. Garcia, C. and Tziritas, G., Face detection using quantized skin color regions merging and wavelet packet analysis, IEEE tran. on Multimedia, 1(3) (1999) 264-276.
4. Lee, K.-M., Component-based online learning and its application to face detection, in Proc. of CGIM (2004) 285-290.
5. Holstein, H. and Li, B., Low density feature point matching for articulated pose identification, in Proc. of BMVC (2002) 678-687.
6. Mittal, A., Zhao, L., and Davis, L.S., Human body pose estimation using silhouette shape analysis, in Proc. of International conference on Advanced video and signal based surveillance (2003) 263-270.
7. Kyoung-Mi Lee and W. N. Street, Model-based detection, segmentation, and classification using on-line shape learning, Machine Vision and Applications, 13(4) (2003) 222-233.