

시-공간 특징을 이용한 효율적인 3차원 인체 자세 재구성

양희덕, 아마드, 이성환
고려대학교 정보통신대학 인공시각연구센터/컴퓨터학과
{hdyang, mohi, swlee}@image.korea.ac.kr

Efficient Reconstruction of 3D Human Body Pose Using Spatio-Temporal Features

Hee-Deok Yang, Mohiuddin Ahmad and Seong-Whan Lee
Dept. of Computer Science and Engineering, Korea University

요 약

본 논문에서는 스테레오 영상에서 깊이 정보를 추출하여 사람의 자세를 학습된 2차원 깊이 영상들의 선형 결합으로 표현하여 3차원 인체 모델을 재구성하는 방법을 제안한다. 한 장의 2차원 깊이 영상으로 최소 제곱법을 이용하여 프로토타입 깊이 영상의 선형 결합으로 표현되는 최적의 계수를 찾을 수 있다. 입력된 깊이 영상의 3차원 인체 모델은 프로토타입 깊이 영상에서 예측된 계수를 적용하여 생성한다. 학습 단계에서는 데이터를 계층적으로 나누어 모델을 생성한다. 또한, 재구성 단계에서는 실루엣 영상과 깊이 영상으로부터 계층적으로 나누어진 학습 데이터를 이용하여 3차원 인체 자세를 재구성한다. 학습 및 재구성의 마지막 단계에서는 실루엣 영상 대신 깊이 영상을 이용하여 3차원 인체 모델을 재구성한다. 한 장의 실루엣 영상을 이용하면 영상의 노이즈에 민감하기 때문에 재구성 단계의 상위 레벨에서는 실루엣 영상의 누적 영상을 이용한다. 실험 결과는 제안된 방법이 효율적으로 3차원 인체 자세를 재구성함을 보여준다.

1. 서론

사람과 컴퓨터간의 상호작용을 효율적이고 편안한 방법으로 개선시키려는 노력이 계속되고 있다. 특히, 사람과 컴퓨터가 사람과 사람이 의사소통하는 것과 유사하게 의사소통을 할 수 있는 방법에 대한 필요성이 날로 증가하고 있다. 그래서, 많은 연구가 사람이 일반적으로 사용하고 있는 의사소통 수단인 음성, 제스처, 수화 인식 등에 집중적으로 이루어지고 있다.

특히, 인체의 구성요소를 검출하여 제스처를 인식하는 기술은 사람과 컴퓨터간의 상호작용을 위해서 필요한 기술 중 하나이다. 그래서, 2차원 또는 3차원 인체 구성요소 검지 및 예측 기술에 대한 많은 연구가 수행되고 있다[1, 2, 4, 5].

Bowden은 2차원 실루엣의 좌표 정보와 이에 대응되는 3차원 인체 골격의 좌표 정보를 특징으로 선택하고, 선택된 특징으로부터 비선형 통계모델을 생성하여 인체 구성요소를 검출하였다[1]. 통계모델은 2차원 실루엣 좌표 정보와 3차원 골격 좌표 정보를 주성분 분석(PCA)을 이용하여 모델링하였다. 3차원 정보의 좌표는 머리, 손, 상반신으로 제한되어있다. Song은 Hidden Markov Model과 에너지 최소화 방법을 이용하여 인체 구성요소의 움직임,

속도, 위치 변화의 다양성에 대한 확률밀도함수를 계산하여 인체 포즈의 변화를 모델링하였다[5].

Rosales는 SMA(Specialized Mapping Architecture)를 사용하였다. SMA는 특화된 매핑 함수와 하나의 피드백 매핑 함수로 이루어져 있으며, 함수들은 학습 데이터로부터 예측된다[4].

Heap은 인체 구성요소의 좌표를 예측하기 위해 학습방법을 이용하였다. 이들은 통계적인 방법으로 접근을 시도하여 인체 자세의 3차원 모션을 예측하는 연구를 수행하였다[5].

본 연구에서는 실루엣 영상 및 에지 특징을 이용했을 때 발생하는 인체 구성요소간의 전/후 관계를 정확히 검출하지 못하는 문제를 해결하기 위해서 실루엣 영상과 스테레오 영상에서 생성된 깊이 영상을 이용하는 방법을 제안한다. 제안된 방법은 실루엣 영상 추출 시 발생하는 노이즈를 공간적 특징 정보와 현재 자세는 이전의 자세와 비슷하다는 시간적 특징 정보를 사용한다. 시간적 특징 정보를 위해 실루엣 누적 영상을 생성하여 노이즈를 효율적으로 제거하고 현재 자세와 유사한 자세에서만 재구성한다. 또한, 하나의 문제를 여러 개의 하위 군집으로 계층적으로 나누어서 많은 데이터를 작은 데이터로 소규모화하여 처리하여, 검색속도를 개선한다.

2. 3차원 인체 모델

3차원 인체 모델링은 인체 구성요소의 모양 및 움직임을 제어한다. 인체의 모양을 위해서, 인체를 관절과 마디로 구성된 3차원 인체 모델을 생성하였다. 생성된 3차원 인체 모델은 17개의 마디로 구성되어 있으며, 37개의 자유도를 갖고 있다. 또한, 손, 발, 머리에 관절의 끝을 알려주는 표시자를 두었다. 이 표시자는 예측된 관절의 위치 정보로부터 각도 정보를 추출할 때 사용된다. 그림 1은 제안된 3차원 인체 모델의 형태와 계층적 구조를 보여주고 있다. 각 인체 구성요소들은 실제의 인체 모양과 비슷한 모양을 갖도록 superquadrics를 사용하여 표현하였다.

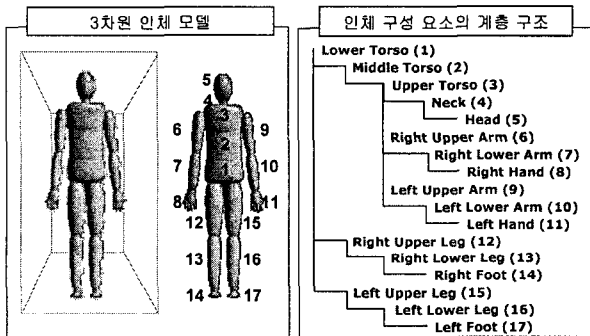


그림 1. 제안된 3차원 인체 모델

비록 제안된 3차원 인체 모델이 실제 사람의 모양과는 많이 다르지만, 인체 자세의 표현과 인체 구성요소간의 충돌 검출 방식을 충분히 할 수 있다.

인체 구성요소간의 움직임을 위해서, 전방향/역방향 운동학을 생성하였다. 인체 구성요소의 상위 마디의 움직임은 마디와 연결되어 있는 모든 하위 마디의 움직임에 영향을 준다. 이를 kinematics chain이라 한다. 예를 들면, 그림 1에서 Upper torso, neck, head가 하나의 kinematics chain이다.

또한, 실 환경의 원/근을 표시할 수 있는 Perspective Camera Model을 이용하여 깊이 영상과 실루엣 영상을 생성한다. 그림 2는 제안된 3차원 인체 모델로부터 생성된 깊이 영상과 실루엣 영상을 보여주고 있다.



그림 2. 깊이 영상, 실루엣 영상 생성

3. 3차원 자세 재구성

3.1 3차원 자세 표현

사람의 자세를 2차원의 깊이 영상과 이에 대응되는 3차원 좌표 정보로 표현하였다. 다양한 자세의 2차원 깊이 영상들의 선형 결합으로 입력된 2차원 깊이 영상이 표현 가능하다면, 다양한 자세들에 대응되는 3차원 좌표 정보의 선형 결합으로 입력된 2차원 깊이 영상에 대응되는 3차원 모델을 재구성할 수 있다. 깊이 정보를 적용함으로써 실루엣 영상만으로는 문제가 발생하는 인체 구성요소간의 전/후 관계를 검지할 수 있다.

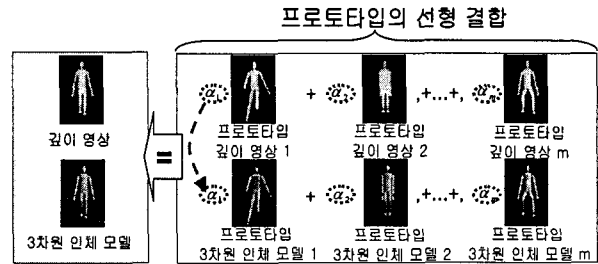


그림 3. 3차원 자세의 표현

깊이 영상 $s = (s'_1, \dots, s'_n)^T$ 는 0~255의 값을 갖는 벡터로 표시되며, n 은 픽셀의 개수이며, 3차원 인체 모델은 $P = ((x_1, y_1, z_1), \dots, (x_q, y_q, z_q))^T$ 로 표시되며, x, y, z 는 인체 모델의 관절의 3차원 좌표이고, q 는 관절의 개수이다. 전체 데이터는 아래와 같이 표시된다.

$$S = (s_1, \dots, s_m), P = (p_1, \dots, p_m) \quad (1)$$

여기서, m 은 전체 프로토타입의 개수이다.

3.2 계층 구조

다양한 형태의 인체 자세의 모델을 학습 데이터로 사용하기 위해서 3차원 인체 모델을 이용하여 100,000 개 이상의 데이터를 생성하였다. 생성된 데이터는 사람이 표현할 수 있는 많은 형태의 자세를 갖고 있으며, 각 인체 구성요소 별로 움직일 수 있는 범위에 대한 제약을 적용하여 생성되었다. 많은 양의 데이터로 인하여 검색 속도가 느려지는 단점을 해결하기 위해서, 학습 데이터를 3단계의 계층 구조를 가지는 형태로 분류하였다. 학습 데이터를 분류하기 위해서 K-means 군집화 알고리즘을 사용하였다. 클러스터의 1, 2 단계 분류를 위해서는 실루엣 영상을 이용하였고, 3 단계에서는 깊이 영상을 이용하였다. 1, 2 단계에서는 자세의 모양을 분류하였고, 3 단계에서는 비슷한 자세를 갖는 데이터의 전/후 관계를 분류하였다. 각 클러스터들은 2차원 실루엣 영상 공간에서 유사한 모양을 갖는다. 계층적 분류를 위해서 하위 단계는 상위 단계의 클러스터의 평균값을 갖도록 구성하였다.

3.3 시-공간 특징 정보

한 장의 실루엣 영상을 이용하면 현재 영상의 노이즈에 민감하기 때문에 재구성 단계의 상위 레벨에서는 실루엣 영상의 누적 영상, $H(x,y,t)$ 를 이용하였다. 실루엣 영상의 누적 영상을 이용함으로써, 현재 자세와 비슷한 영역에서만 자세를 재구성할 수 있다.

$$H(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t)=1 \\ \max(0, H(x,y,t-1) - \lambda) & \text{otherwise} \end{cases} \quad (2)$$

여기서, τ 는 실루엣 영상이 영향을 주는 주기, $D(x,y,t)$ 는 시간 t 에서 실루엣 영상의 x, y 좌표, λ 는 이전 실루엣 영상의 소멸을 위한 상수이다.

4. 실험 결과 및 분석

4.1 실험 환경

제안된 방법의 성능평가를 위해서, KU 제스처 데이터베이스[3, 6]의 일부인 정상적 제스처 데이터베이스를 이용하였다. 정상적 제스처 데이터베이스는 10명의 남자, 10명의 여자에 대해서 14가지의 제스처를 3차원 모션 데이터와 2차원 스테레오 데이터, 2차원 실루엣 영상의 형태로 구성되어있다.

4.2 실험 결과

그림 4는 KU 제스처 데이터베이스의 인사하기 동작을 이용하여 실험한 결과이다. 입력 영상은 깊이 영상과 스테레오 영상이며, 결과 영상은 재구성된 3차원 인체 모델의 인체 구성 요소 별 각도 정보이다.

실험 결과의 정확성을 검증하기 위해서 KU 제스처 데이터베이스의 Ground Truth와 예측된 3차원 인체 모델의 각 인체 구성 요소의 평균 오류율을 계산하였다. 그림 5에서 θ 는 x, y 축에 투영된 각의 평균차를 ψ 는 y, z 축에 투영된 각의 평균차를 나타낸다. 실험 결과 움직임이 많은 다리 부분에서는 다른 부분보다 오류율이 높음을 볼 수가 있다.

본 논문에서는 깊이 영상과 이에 대응되는 3차원 인체 모델의 선형 결합에 기반한 3차원 인체 자세 재구성 방법을 제안하였다. 제안된 방법은 미리 주어진 깊이 영상의 선형 결합으로 입력된 영상을 근사화하는 방법을 사용하였다. 실험 결과, 입력된 영상으로부터 재구성된 3차원 인체 모델의 자세가 원영상의 자세와 유사함을 확인할 수 있었다. 제안된 방법은 정면에서 입력된 실험 데이터에서만 실험이 실행되었다. 앞으로의 연구는 이러한 제약 조건을 해결할 수 있는 방법을 개발하여 시점에 무관한 자동화된 알고리즘을 개발하는 것이다.

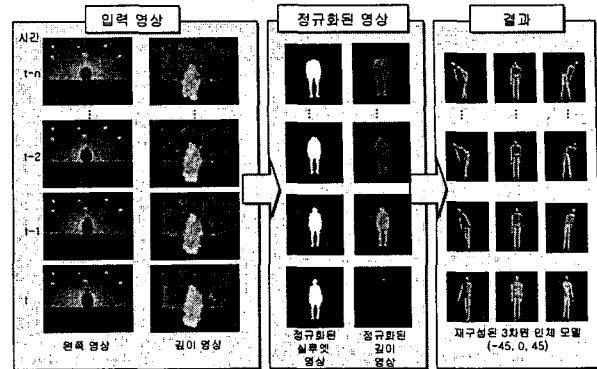


그림 4. 재구성된 3차원 인체 모델의 예

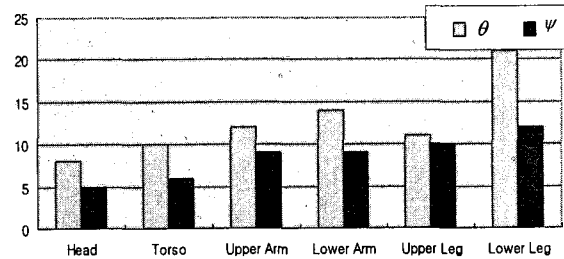


그림 5. 인체 구성 요소 별 평균 오류율

참고문헌

- [1] R. Bowden, T. A. Mitchell, and M. Sarhadi, "Non-linear Statistical Models for 3D Reconstruction of Human Pose and Motion from Monocular Image Sequences," *Image and Vision Computing*, Vol. 18, No. 9, 2000, pp. 729-737.
- [2] T. Heap and D. Hogg, "Improving Specificity in PDMs using a Hierarchical Approach," *Proc. of 8th British Machine Vision Conference*, Colchester, UK, Sep. 1997, pp. 590-599.
- [3] B.-W. Hwang, S. Kim, and S.-W. Lee, "Full-Body Gesture Database for Analyzing Daily Human Gestures," *Advance in Intelligent Computing, Lecture Notes in Computer Science*, Vol. 3644, 2005, pp. 611-620.
- [4] R. Rosales and S. Sclaroff, "Specialized Mapping and the Estimation of Human Body Pose from a Single Image," *Proc. of IEEE Workshop on Human Motion*, Texas, USA, Dec. 2000, pp. 19-24.
- [5] Y. Song, X. Feng, and P. Perona, "Towards Detection of Human Motion," *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, South Carolina, USA, June 2000, pp. 810-817.
- [6] KU Gesture Database, <http://GestureDB.korea.ac.kr/>.