

실루엣 영상을 이용한 삼차원 인체 포즈인식

오치민,, 김민욱, 이철우
전남대학교 전자컴퓨터공학부
e-mail : sapeyes@image.chonnam.ac.kr

3D Pose Recognition using Body Silhouette Image

Chi-Min Oh, Chi-Min Oh, Chil-Woo Lee

*Dept. of Electronics and Computer Engineering, Chonnam National University

요 약

본 논문은 이차원 영상에 투영된 삼차원 인체의 포즈를 인식하기 위하여 이차원 영상에 투영된 인체의 실루엣 정보를 이용하였다. 인체는 삼차원 공간에서 움직이므로 이차원 영상으로 모든 정보를 알아내기에는 부족한 면이 있다. 따라서 본 논문에서는 인체 포즈의 주시 방향을 결정된 후 인체의 실루엣 영상 Convex-hull 특징점 정보를 이용하여 인체의 삼차원 포즈를 인식하였다. 인체의 포즈는 PCA 로 차원을 축소하였으며 Diffusion Distance 로 데이터베이스의 포즈모델 중 가장 가까운 모델을 선택하였다.

1. 서론

삼차원 인체 포즈 인식은 이차원 영상을 이용하여 사람의 현재 관절의 위치를 알아내는 방법으로 이루어진다. 인체 포즈는 행동의 의미를 내재하고 있으므로 제스처인식 분야에서 널리 활용되고 있으며 그 외 모션캡처, 로봇과의 의사소통, 애니메이션 제작 등 다방면에서 활용되고 있다. 특히 최근에는 로봇이 사람의 의사를 파악하고 때에 따라서 사람이 요구하는 행동을 학습하기 위해 삼차원 포즈를 인식하는 연구도 시작되고 있다.

인체 포즈를 인식하기 위해서는 인체를 구성하고 있는 다수의 관절과 강체(Rigid Object)들의 관계를 알아내야 한다. 하지만 비전을 기반으로 하는 연구분야에서는 이차원 영상으로 포즈를 인식하는 하기 때문에 삼차원에서 이차원으로 투영된 포즈 정보가 불충분한 어려움이 있다. 이러한 어려움을 극복하기 위하여 여러 연구가 진행되고 있다. 대체적으로 신체의 구성요소를 검출하는 방법 그리고 삼차원 인체모델을 이용하는 방법, 학습모델을 사용하는 방법이 있다.

신체의 구성요소를 검출하는 방법은 손과 얼굴, 다리, 몸통 등을 검출하여 이차원 영상의 사람의 포즈를 인식하는 방법이다. 색상(color)과 모양(shape) 정보를 사용하여 신체의 구성요소들을 검출한다. 삼차원 정보를 활용하지 않기 때문에 신체 부위가 겹침에 의해 영상에 나타나지 않았을 경우 검출 성능이 떨어지는 단점이 있다. [1]에서는 이러한 문제를 극복하기 위해서 손과 얼굴의 여러 주시 방향에서 검출될 수 있도록 색상과 모양의 통계적 분류방법을 사용하였다. [2]에서는 실루엣(silhouette)의 외곽선에 분포한 Convex, Concave 특징점들을 이용하여 6 가지 주요 신체부위를 결정하였으며 [1]처럼 여러 주시 방향에서 검출되는 방법을 제시하고 있다.

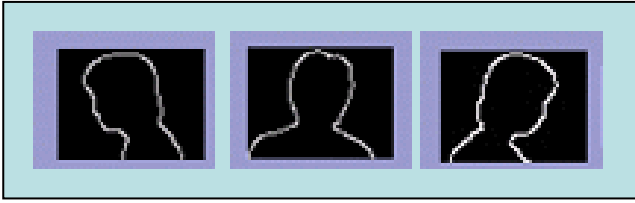
삼차원 인체모델을 사용하는 방법은 인체의 관절들이 갖는 연관성을 고려하여 포즈를 인식할 때 관절의 구조대로 탐색한다. 관절 사이의 관계는 탐색공간을 축소시키고 겹침 현상에 의해 탐색이 어려운 신체부위도 탐색공간이 적기 때문에 결정이 더 쉬운 편이다. [3,4]에서는 인체의 모델을 정의하여 제약조건에 의해 포즈를 인식하였다. Gavrilu[3]는 다수카메라의 이차원 영상에 삼차원 모델을 투영하여 Chamfer Distance 를 이용해 포즈를 인식하는 방법을 제시했다.

본 논문에서는 삼차원 포즈를 인식하기 위하여 주시 방향추정 및 실루엣 영상을 이용하였다. 2 장에서는 주시 방향 추정을 위한 템플릿 매칭 과정을 소개하고 3 장에서는 실루엣 영상을 이용한 포즈 인식 과정을 소개한다.

2. 주시 방향추정

인체는 주로 얼굴 또는 몸통의 방향에 따라 이차원 영상으로 투영된 포즈가 다르다. 즉 얼굴의 주시 방향은 이차원 영상에서는 알기 어려운 인체의 삼차원 모델 파라미터이다. 주시 방향을 알게 되면 그 외의 신체 부분들은 얼굴 또는 몸통을 중심으로 검출할 수 있다. 주시 방향을 얻기 위하여 특징이 비교적 강한 얼굴과 어깨의 에지(edge)를 이용한 템플릿 매칭 방법을 사용하였다.

주시 방향은 이차원 영상에 나타난 얼굴이 왼쪽 면일 경우 왼쪽, 오른쪽 면일 경우 오른쪽, 마지막으로 정면일 경우 정면으로 모델링하였다.. 각 주시 방향의 템플릿 모델 영상은 10 사람의 얼굴의 에지 영상을 그대로 사용하였다. 모델 영상은 상관도 분석에 의한 템플릿 매칭 방법을 이용해 입력영상과 비교한 후 위 주시 방향을 결정하였다. 그림 1 에 주시 방향 모델영상이 나타나있다.



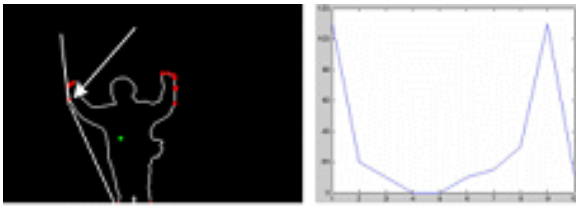
(그림 1) 주시 방향 모델 영상

주시 방향 결정은 상관도를 이용한 템플릿 매칭 방법을 사용하였다. 정규 상관도는 다음 식과 같다.

$$R(x, y) = \frac{\sum_x \sum_y T(x', y') \times I(x+x', y+y')}{\sum_x \sum_y T(x', y')^2 \times \sum_x \sum_y I(x+x', y+y')^2} \quad (\text{식 1})$$

3. 포즈 추정

포즈 추정은 주시 방향 결정 후 주시 방향에 맞는 포즈 모델의 특징과 실루엣 영상의 특징과 비교함으로써 결정하였다. 즉 포즈모델은 주시 방향 별로 분리하여 분류가 되어 있다. 포즈 모델은 다음과 같이 각 포즈를 실루엣에서 Convex-hull 점들 사이의 각도 히스토그램으로 작성하였다.



(그림 2) 포즈 실루엣 Convex-hull 각도 히스토그램

포즈모델의 각 포즈 P_i 는 10 인의 특정 포즈에 대한 각도히스토그램의 평균값이다(식 2). 특징공간 분석을 위해 PCA (Principle Component Analysis)를 사용하여 3 개의 특징벡터를 이용해 모든 포즈는 특징공간으로 3 차원 특징으로 축소하였다(식 3).

$$P_i = \frac{\sum_{i=0}^n \sum_{\theta=0,90,-90} \text{histogram}(\theta, i)}{8} \quad (\text{식 2})$$

$$pMean = \frac{1}{n} \sum_{i=1}^n P_i$$

$$COV = E[(P_i - pMean)(P_i - pMean)^T]$$

$$E = [e_1, e_2, e_3]^T = EVD(COV) \quad (\text{식 3})$$

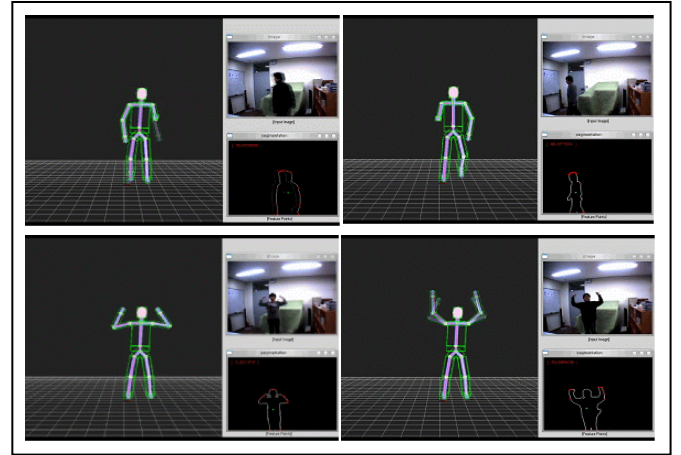
$$P_i^\circ = E \cdot P_i$$

입력영상의 포즈 추정은 위 포즈모델 구축과정과 비슷하다. 먼저 입력영상의 실루엣 외곽선을 영상을 구한다. Convex-hull 점들의 각도 히스토그램을 구한다음

특징공간으로 투영한다. 포즈 결정은 특징 공간의 P_i 와 Diffusion distance[5]를 구한 후 가장 거리가 가까운 포즈를 선택하게 된다(식 4).

$$T' = E \cdot T^T$$

$$i = \arg \min_i dd(T', P_i) \quad (\text{식 3})$$



(그림 3) 인식된 인체 삼차원 포즈

4. 결론

본 논문에서는 인체의 삼차원포즈를 인식하기 위해 템플릿 모델을 기반으로한 주시 방향결정, 실루엣 특징을 이용한 포즈 인식을 연구하였다. 기존 연구내용의 신체부위 검출방법에 삼차원 모델의 주시 방향에 불변한 특징을 추가함으로써 삼차원 포즈를 유추할 수 있었다. 향후 연구내용으로는 더 나은 포즈 인식과 포즈 모델을 만들기 위해 세밀한 삼차원 인체포즈 모델을 정의 하는 연구를 진행할 것이다.

Acknowledgement

본 연구는 정통부 및 정보통신연구진흥원의 정보통신 선도기반기술개발사업의 연구비 지원에 의해 수행되었습니다

참고문헌

- [1] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfunder: Real-time tracking of the human body. In Photonics East, SPIE, Vol 2615, 1995. Bellingham, WA.
- [2] Haritaoglu, I., Harwood, D., and Davis, L.S., Ghost: A human body part labeling system using silhouettes Fourteenth International Conference on Pattern Recognition, pp. 77-82.
- [3] D.M. Gavrila and L.S. Davis. Tracking of humans in action: A 3D model-based approach. In ARPA Image Understanding Workshop, Palm Springs, 1996
- [4] C. Bregler. Tracking people with twists and exponential maps. In CVPR98, 1998.
- [5] H. Ling., K. Okada., Diffusion Distance for Histogram Comparison., 2006 Conference on Computer Vision and Pattern Recognition. 2006