비디오로부터의 움직이는 3D 인체 형상 및 자세 복원

*한지수, **조명래, *박인규
*인하대학교 정보통신공학과, **코아시아 홀딩스
{jshan0799@gmail.com, michaelmrcho@gmail.com, pik@inha.ac.kr}

Moving Human Shape and Pose Reconstruction from Video

*Ji Soo Han, **Myung Rai Cho, *In Kyu Park
*Inha University, **CoAsia Holdings

요 약

본 논문에서는 비디오로부터 추출된 프레임에서 3D 인체 모델의 복원하고 이를 부드럽게 재생될 수 있도록 보정하는 기법을 제안한다. 매개변수 기반의 모델을 사용하여 자세 및 체형을 복원하도록 접근하고 있다. 매개변수 기반의 인체 모델은 다양한 인체 데이터의 학습을 통해 만들어지며 입력 영상으로부터 최적의 자세와 체형 매개변수 값을 찾아 복원하게 된다. 자세 복원은 CNN 을 사용하여 영상으로부터 인체의 관절 위치를 추정하고 3D 모델로부터 2D 로 투영을 통해 관절 간의 거리가 최소화되는 매개변수 값을 찾아 복원한다. 형상 복원은 2D 영상으로부터 취득된 사람의 윤곽 데이터와 3D 모델의 윤곽 데이터 간의 매칭을 통해 복원된다. 이러한 단일 입력 영상에서 비디오와 같은 다중 입력 영상으로 확장하여 칼만 필터를 적용하여 오류 프레임을 검출하고 이전, 이후 프레임의 매개변수와의 보간을 통해 보다 자연스럽고 정확한 모델을 생성한다.

1. 서론

사람의 형상을 컴퓨터 상에서 재현하려는 인체 3D 복원에 대한 연구는 오랜 기간 진행되어 왔으며 매개변수 기반의 SMPL(Skinned multi-person linear model)[1] 모델이 제안된 이후로 빠르게 발전하여 현재 매우 정교한 모델을 복원할 수 있다. 최근 머신 러닝 기술의 발전으로 영상으로부터 자세와 형상과 같은 인체의 정보를 추출하는 기술이 발전하여 보다 정확한 정보를 토대로 복원이 가능하게 되었다. 하지만 기존의 연구는 움직임을 포함하고 있는 비디오의 인체를 복원하는데 자세의 복잡성과 가려지는 현상으로 인해 어려움을 나타내고 있다. 또한 기존의 연구 결과를 비디오와 같은 다중 프레임에 적용하였을 때 실제 복원된 결과는 각 프레임마다 개별적으로 모델링 되어 불연속적 결과를 보인다.

본 논문에서는 이러한 문제를 해결하기 위해 비디오로부터 추출된 각 프레임에 대해 시간의 흐름에 따른 각 모델들 간의 상관 관계를 고려한 복원 기법을 제안한다. 먼저 비디오로부터 추출된 각 프레임을 단일의 영상으로 고려하고 3D 복원을 수행한다. 매개변수 기반의 3D 인체 모델은 다양한 자세를 취하고 있는 인체 데이터와 서로 다른 체형 및 성별을 지니고 있는 인체 데이터의 학습을 통해 만들어진다. 모델의 자세는 머신 러닝을 통한 파트 측정을 통해 영상으로부터 관절의 위치를 추정하고 3D 모델로부터 2D 로 투영하여 복원한다. 체형의 경우엔 마스크를 생성하여 윤곽 정보를 추출하고 매칭을 통해 최적의 형상 매개변수를 찾는다. 단일 프레임을 기준으로 복원된 모델은 인체 정보 취득 과정에서의 오류와 복원과정에서의 오차로 인해 각 모델들 사이의 연속성이 부족하게 되고 이를 해결하기 위해 칼만 필터를 사용한다. 이전 프레임을 토대로 현재 프레임의 결과를 예측하여 오차를 줄임으로써 부드러운 3D 모델의 움직임을 재현할 수 있다. 이러한 과정을 통해 본 논문은 기존에 연구되어 왔던 단일 영상이 아닌 다중 영상에서 3D 모델을 복원하고 부드러운 움직임을 표현하는 기법을 제시한다.

2. 3D 인체 형상 및 자세 복원

3D 인체 복원은 매개 변수 기반의 인체 모델인 SMPL 을 사용한다. 다양한 사람들의 자세와 체형에 대한 데이터를 토대로 정점 기반 모델로 기존의 표준화된 모델링 방식은 꼭지점들과 골격 구조를 관계시키는 방법을 사용하였으나 사실성이 떨어지는 문제점들이 있었다. 이를 해결하기 위하여 자세 변화를 각 관절의 회전 매트릭스 요소들에 의한 선형 함수로 만들어 공식화하였다. 해당 공식은 목적이 되는 함수가 등록된 메쉬와 데이터로부터 생성된 모델 간의 차이를 없애고 이를 위해 다양한 자세 변화에 따라 생성된 스캔을 사용하여 템플릿 메쉬를 정렬하고 데이터를 최적화하였다. 메쉬 정보를 구축하고 자세 변화에 따른 근육의 변화와 같은 세부적인 부분을 보정하여 데이터 기반의 통일화 된 모델을 토대로 매개변수 값의 변경을 통해 다양한 체형과 자연스러운 자세 변화를 나타낼 수 있다. [2][3][4]

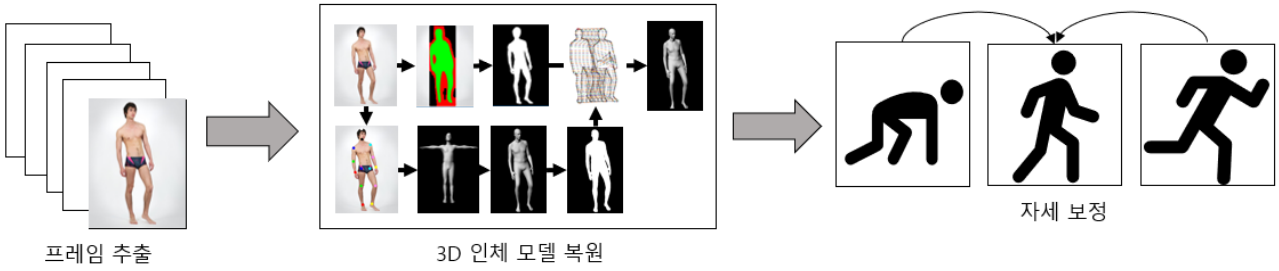

그림 1. 3D 인체 모델 복원 및 보정 파이프라인

2.1. 자세 복원

자세 측정 방법은 FR-CNN 을 사용하여 신체의 각 파트를 검출하는 DeepCut 기법을 통해 23 개로 표현된 관절 데이터 추출하여 이를 기반으로 투영된 3D 관절과 CNN 에 의해 검출된 2D 관절 위치 사이의 오차를 줄이는 목적함수를 최적화하는 과정을 통해 이루어진다. Z 축으로 나타나는 깊이에 대해선 SMPL 모델의 어깨와 골반 사이의 거리를 통한 상체 길이와 측정된 2D 관절로 정의된 유사한 삼각형의 비율을 통해 깊이를 추정한다. 이 과정에서 인체의 대략적인 체형이 측정되고 그에 따른 전체적인 비율을 도출해 낸다.

$$E(\beta,\theta) = E_{\mathrm{j}}(\beta,\theta;K,J_{est}) + \lambda_\theta E_\theta(\theta) + \lambda_a E_a(\theta) + \lambda_{sp} E_{sp}(\theta;\beta) + \lambda_\beta E_\beta(\beta) \quad (1)$$

식(1)은 그 목적함수를 나타내며 가장 최적의 자세와 체형 매개변수를 나타내는 β, θ를 찾는다. 측정된 관절의 위치 값은 매개 변수 기반의 모델에서 회전 반경에 대해 Rodrigues 공식에 의해 적용되고 이는 3D 모델로 복원되게 된다.

2.2. 체형 복원

3D 체형의 복원은 자세 복원과 밀접한 관계를 갖고 있으며 이는 서로 상호 의존적인 관계를 맺고 있다. 신체의 형상복원은 윤곽을 기반으로 매칭이 진행된다. 2D 영상으로부터 대상의 윤곽을 찾고 3D 모델의 윤곽과의 EMD(Earth mover's distance) 매칭을 통해 형상 매개변수를 찾기 때문이다 [5]. 윤곽 정보의 비교는 같은 자세를 취하고 있을 때 이루어진다. 이를 위해 자세복원을 하는 과정에서 대략적인 형상 복원을 수행하지만 이는 전체적인 비율을 나타낼 뿐 허리의 둘레나 팔목의 굵기와 같은 세부적인 복원은 이루어지지 않는다. 형상 복원에서는 이러한 부분을 윤곽 매칭을 통해 해결한다. 이러한 윤곽 매칭은 최근 CNN 의 발전으로 사람의 추정이 정확해짐에 따라 그 결과가 향상되어왔다. FR-CNN 을 확장하여 기존의 바운딩 박스 형식으로 표현하는 것과 달리 마스크로 추정한 대상을 표현하는 Mask R-CNN[6]을 사용하여 대략적인 윤곽 정보를 취득하고 GrabCut 에 의해 보다 정교하게 가공하여

그림 2. 단일 영상으로부터 3D 복원 결과

정확도를 매칭 과정에서 오차를 최소화함에 따라 보다 정확한 윤곽 데이터를 취득하였다. 윤곽 매칭은 두 개의 확률 분포 사이의 최소 비용을 계산하는 EMD 공식을 사용하여 2D 영상에서의 윤곽과 3D 모델로부터의 윤곽사이의 비용을 최소화하는 형상 매개변수를 찾도록 하였다.

3. 다중 영상으로부터 모델 보정

단일 영상으로부터 3D 인체 복원은 오랜 연구를 통해 매우 높은 정확한 결과를 나타내고 있다. 하지만 가려진 부분이나 복잡한 포즈의 복원에서 큰 어려움은 단일 영상에서의 복원에서도 여전히 과제로 남아 있으며 비디오와 같은 다중 영상에서는 움직임으로 인해 더욱 정확한 측정이 어려운 상황이다. 본 논문은 인체의 움직임에서 시간의 변화에 따른 각 관절 파트 변화의 연관성을 고려하여 칼만 필터를 적용하여 오류 프레임을 검출 및 보정하였다. 칼만 필터는 이전 프레임과 현재 프레임으로부터 복원된 모델의 자세 매개변수(θ (pitch), Φ (roll), Ψ (yaw))에 대해 다음 수식에 의해 매개변수의 상태를 추정한다.

$$\mathrm{x}_k = Ax_{k-1} + Bu_k \quad (2)$$

$$\mathrm{z}_k = Hx_k + v_k \quad (3)$$

여기서 식(2)의 x 는 상태 벡터이며 u 는 자세 매개변수 값을 입력 값으로 받는다. 여기서 계산된 자세 값은 측정된 자세 값과 비교를 통해 해당 프레임의 정확성 여부를 확인 하여 이전 프레임과 이후 프레임 값의 보간을 통해 새로운 자세 매개변수를 도출한다.

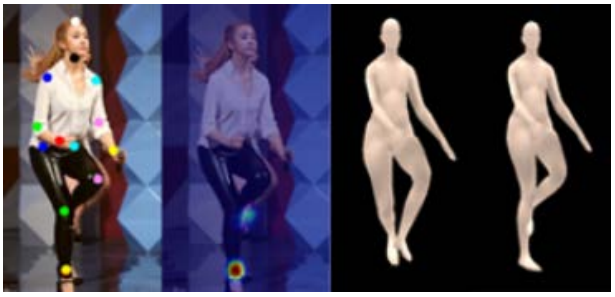

그림 3. 관절 측정 오류 및 보정

4. 실험 결과

본 논문은 Intel i7-7700 3.6GHz CPU 와 NVIDIA GeForce GTX 1080 Ti 를 장착한 컴퓨터에서 수행하였다. 단일 영상에서 3D 복원은 약 5 분의 수행시간 소요되었고 다중 영상을 이용하여 모델 보정은 약 3 분의 수행 시간이 소요되었다.

그림 2 는 단일 영상으로부터 자세와 체형 복원을 수행 한 결과로 입력 영상과 매우 유사한 결과를 확인 할 수 있다. 논문에 수록된 결과는 배경과 모델이 제한된 상황의 결과이며 옷에 의해 정확한 체형이 영상에 보이지 않거나 관절 위치의 측정이 어려운 정도에 따라 그 정확도는 점차 떨어진다. 그림 3, 4 는 비디오로부터 추출한 영상에 대해 3D 복원 및 오류 프레임의 보정을 수행한 결과이다. 그림 3 은 초기 관절 위치 측정에서 왼쪽 발목의 위치가 잘못 측정되어 잘못된 3D 복원이 수행되었으며 중앙의 결과에 비해 오른쪽 모델이 입력 영상과 유사한 모습을 확인 할 수 있다. 그림 4 는 연속적인 2D 영상에 대해 자신의 팔에 의해 가려진 결과로 인해 잘못 복원된 오류 프레임을 보정하여 위와 아래의 결과를 비교하였을 때 아래의 동작이 자연스러운 연속 동작을 보이는 것을 확인 할 수 있다.

5. 결론

본 논문에서는 매개변수 기반 모델을 이용하여 입력 영상으로부터 3D 인체 복원을 수행하고 다중 영상으로부터 오류 프레임을 검출 및 보정하는 기법을 제안하였다. 이러한 접근은 비디오 영상을 입력으로 받았을 때 기존 접근 방식에 비해 부드럽고 자연스러운 움직임을 재현할 수 있다. 또한 2D 영상으로부터 측정하기 어려운 정보를 다중 영상을 통해 보완하여 보다 정교한 인체의 복원이 가능하다.

감사의 글

본 논문은 CoAsia Holdings 의 지원을 받아 수행된 연구입니다.

참고문헌

[1] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: a skinned multi-person linear model,” *ACM Trans. on Graphics*, vol. 34, no. 6, pp. 248, Nov, 2015.

[2] L. Pishchulin, E. Insafutdinov, S. Tang, and B. Andress, “DeepCut: joint subset partition and labeling for multi person pose estimation,” *Proc. IEEE CVPR*, pp. 4929-4937, June 2016.

[3] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, “Estimating human shape and pose from a single image,” *Proc. IEEE ICCV*, pp. 1381-1388, September 2009.

[4] F. Bogo, A. Kanazawa, C. Lassner, and P. Gehler, “Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image,” *Proc. ECCV*, pp. 561-578, October 2016.

[5] K. Grauman and T. Darrell, “Fast contour matching using approximate earth mover's distance,” Proc. *IEEE CVPR*, June 2004.

[6] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” *Proc. IEEE ICCV*, pp.2980-2988, October 2017.

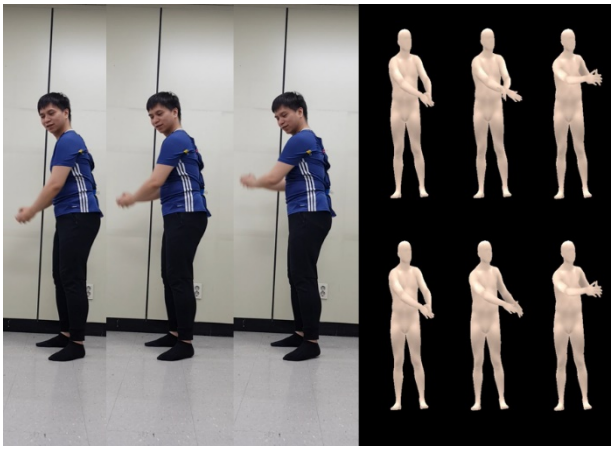

그림 4. 다중 영상으로부터 3D 인체 보정