

매개변수 변형 모델과 회귀분석을 이용한 비디오로부터의 3D 인체 복원

한지수, 박인규

인하대학교 정보통신공학과

{jshan0799@gmail.com, pik@inha.ac.kr}

요약

본 논문은 비디오로부터 추출된 다중 프레임에 대하여 3D 인체 복원을 수행하고 인체의 형상 및 움직임을 부드럽고 사실적으로 표현하도록 보정하는 기법을 제시한다. 영상으로부터 매개변수 기반의 3 차원 morphable 모델을 사용하여 인체의 자세 및 체형을 복원하도록 접근한다. 인체 복원은 자세 복원과 체형 복원으로 나뉘게 되며 입력영상으로부터 측정된 관절의 위치 정보와 윤곽 정보에 의해 최적의 매개변수 값을 찾게 된다. 관절의 위치 정보는 CNN 을 기반으로 측정하며 3D 모델로부터 2D 영상으로의 투영을 통해 거리를 비교하고 최소화되는 값을 찾아 복원한다. 윤곽 정보는 전경 객체 추출 기법을 적용하여 인체의 윤곽 마스크를 측정하고 3D 로부터의 윤곽 정보와 정합 코스트 값을 측정하고 최소 값을 갖는 체형 매개변수 값을 찾는다. 비디오로부터 추출된 다중 프레임에 대한 3D 인체 복원은 각 매개변수가 연속성 및 일관성을 갖고 있으며 전체 복원 매개변수에 대한 분위 회귀 분석을 적용하여 자연스럽게 사실적인 모델을 생성한다.

1. 서론

2D 영상으로부터 인체의 3D 모델을 복원하는 문제는 오랜 기간동안 연구되어왔다. 매개변수 기반의 모델인 SMPL(Skinned multi-person linear model)[1] 이 제시되고 머신 러닝의 발전으로 인체의 정보를 보다 정확하게 측정할 수 있게 되어 인체 복원 연구는 보다 빠르게 발전하고 있다. 하지만 이러한 기존의 연구는 단일 영상으로부터의 복원에 초점이 맞춰져 있으며 비디오로부터 추출된 다중 프레임에 대해서는 부자연스러운 결과를 나타낸다. 이러한 부자연스러움은 자세의 복잡성과 가려지는 현상에 의해 정확한 인체 정보 측정에 어려움이 있으며 카메라의 위치 및 각도에 의한 오차에 의해 발생한다. 이에 따라 비디오로부터 복원된 각 프레임 사이의 모델들은 일관성과 연속성을 잃게 된다.

본 논문에서는 비디오로부터 추출된 각 프레임에 대한 3D 인체 복원을 수행하고 시간의 흐름에 따른 각 모델들의 매개변수 값들의 상관 관계를 분석하여 통일 성과 연속성을 갖는 보정된 3D 모델을 생성하는 기법을 제시한다. 매개변수 기반의 모델은 다양한 자세와 체형을 갖는 인체 데이터의 학습에 의해 만들어 진다. 단일 영상에 대한 복원은 2D 영상으로부터 측정된 관절 정보와 체형 정보는 3D 로부터 2D 로의 투영된 모델과의 정합을 통해 최적의 매개변수 값을 찾게 된다[2][3]. 하지만 이러한 단일 영상 복원 기법은 초기 측정 정보 및 복원 과정에서의 오차에 의해 오류 프레임이 생성되고 복원된 결과에 대한 재생에 대해 부자연스러운 모습을 나타낸다. 체형 매개변수는 동일 개체에 대해 일관

성을 갖고 자세 매개변수는 프레임 사이의 연속성을 갖는다. 이러한 가정에서 전체 매개변수에 대한 회귀 분석을 통해 오류 프레임을 측정 및 보정하여 부드러운 3D 모델의 움직임을 재현할 수 있다.

2. 제안하는 기법

비디오로부터의 인체 복원 기법은 단일 영상 복원기법을 기반으로 비디오로부터 추출된 각 프레임에 대해 인체 데이터를 측정하고 해당 데이터와 3D 모델의 매개변수의 상관관계를 찾음으로써 자세와 형상을 복원한다. 단일 영상에 대한 복원은 관절 위치 정보를 이용한 자세 복원과 윤곽 정보를 이용한 체형 복원으로 나뉜다. 전체 프레임에 대해 복원된 모델의 매개변수의 일관성과 연속성을 찾기 위해 회귀 분석을 수행하고 오류 프레임의 제거 및 보정을 통해 시간의 흐름에 따른 자연스러운 변화를 나타낸다.

2.1 단일 영상 복원 모델

본 논문은 매개변수 기반의 변형 가능한 SMPL 모델을 사용한다[1]. 이는 72 개의 관절 매개변수와 10 개의 체형 매개변수에 의해 6890 개의 정점으로 이루어진 메쉬가 변형되며 다음과 같이 정의된다.

$$M(\vec{\beta}, \vec{\theta}; \Phi) =$$

$$W(T_p(\vec{\beta}, \vec{\theta}; \bar{T}, S, P), J(\vec{\beta}; J, \bar{T}, S), \vec{\theta}, W) \quad (1)$$

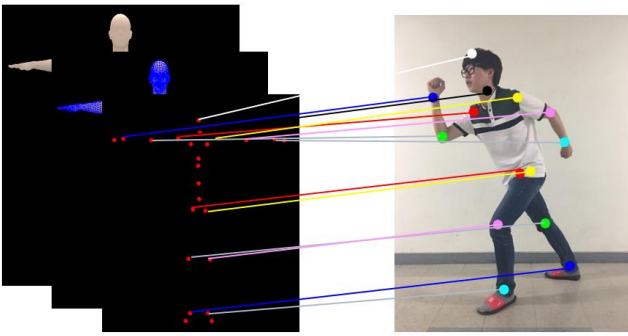


그림 1. 관절 위치 측정을 통한 자세 복원.

여기서 θ 는 자세에 대한 매개 변수로 23 개의 관절에 대해 $72(23 \times 3 + 3)$ 개의 매개변수 변화에 의해 자세를 변형시키고, β 는 다양한 체형의 데이터셋을 주성분 분석 (principal component analysis, PCA)를 통해 얻어진 매개 변수로서 10개의 변수 변화에 의해 체형을 변형시킨다.

인체의 자세 복원은 먼저 관절의 위치를 찾을 필요성이 있다. 관절 위치 측정은 인체의 각 관절 위치를 측정하는 기법을 사용하여 13 개의 주요 관절을 측정한다[4]. 해당 기법은 각 신체의 부위를 CNN을 이용하여 인체의 파트를 추정하고 한 사람에게 속한 후보들로 분류한다. 분류된 결과로부터 각 부위에 대해 라벨을 부여하여 한 명에게 속한 관절의 위치를 추정할 수 있다. 추정된 관절의 데이터를 이용하여 자세 매개변수를 추정하기 위해 3D 모델로부터 2D로 관절을 투영하고 목적함수를 최소화하는 방법을 통해 최적의 자세 매개변수 값을 찾는다[3].

$$E(\beta, \theta) = E_j(\beta, \theta; K, J_{est}) + \lambda_\theta E_\theta(\theta) + \lambda_a E_a(\theta) + \lambda_{sp} E_{sp}(\theta; \beta) + \lambda_\beta E_\beta(\beta) \quad (2)$$

목적함수는 식 (2)와 같이 정의되며 3D로부터 2D로 투영된 SMPL 모델의 3D 관절과 2D 영상으로부터 추정된 관절 사이의 거리를 측정하여 최소화한다. 큰 부위의 자세를 복원함에 따라 측정된 매개변수 값은 Rodrigues 공식에 의해 각 관절의 회전 반경을 결정하게 되고 3D 모델의 자세를 변형한다.

체형 복원은 2D 영상으로부터 측정된 윤곽 정보와 3D 모델의 윤곽정보 정합을 통해 최적의 체형 매개변수 값을 찾는다. 윤곽 정보를 측정하기 위해 CNN을 사용하여 ROI 영역에 대한 분할 마스크를 추정 마스크를 취득하고 GrabCut 기법을 사용하여 정교하게 가공한다[5][6]. 2D 영상으로부터 측정된 윤곽 마스크와 3D 모델로부터 측정된 윤곽 마스크를 비교하기 위해서는 EMD 공식을 사용하여 두 윤곽 데이터 사이의 최소 비용을 계산한다. 10 개의



그림 2. 마스크를 이용한 윤곽 취득.

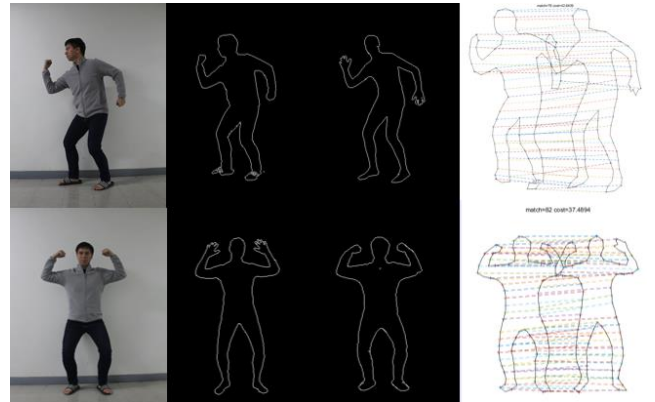


그림 3. EMD 정합을 통한 인체 윤곽 비교.

매개변수에 대해 최소 비용을 갖는 체형을 결정하기 위해서 Levenberg-Marquardt(LM) 알고리즘을 이용하여 최적의 체형 매개변수를 찾는다[7].

2.2 다중 영상 복원 모델

비디오로부터 추출한 각 프레임의 단일 영상 복원은 초기 인체 데이터 측정의 오류 및 각 프레임의 환경의 차이로 인해 부자연스러운 모습을 나타낸다. 다중 영상에 대한 인체 복원은 동일한 객체에 대한 체형 매개 변수의 일관성과 자세 매개변수의 연속성을 전제로 한다. 측정 대상은 동일하며 시간의 흐름에 따라 움직임이 나타난다. 이러한 전제조건은 전체 프레임에 대한 복원된 매개변수들을 분석하여 기존의 문제점을 해결하고 프레임 사이의 자연스러운 변화를 나타낸다.

전체 프레임에 대한 체형 매개변수는 이상점을 검출하여 배제하고 참값에 대해 빈도에 따른 가중치를 부여하여 가중치 평균값을 찾는다. 이에 따라 비디오 전체의 체형은 동일한 체형을 갖게 된다. 전체 프레임에 대한 10 개의 체형 매개변수는 설정된 임계값에 의해 이상점이 제거 및 빈도를 측정하게 된다. 이러한 과정은 입력 프레임의 수에 따라 정확도의 차이를 보인다.

자세 매개변수에 대해서는 각 프레임에 대해 매개변수 값을 관측 값으로 갖는다. 인체의 움직임에 의한 자세 매개변수의 변화는 각 프레임 사이에 연속성을 갖는다. 이에 대해 분위 회귀분석(quantile regression)을 적용한다. 시간을 독립변수 x 로 갖고

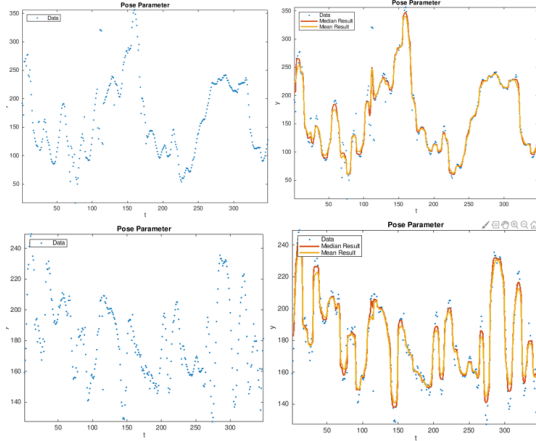


그림 4. 자세 매개변수의 분위 회귀 모델

매개변수 값은 종속 변수 y 로 갖는 분위 회귀분석은 인체의 움직임이 일정한 변화를 나타내지 않고 변화하게 됨에 따라 구간에 따라 분석을 적용해야 하는 자세 분석에 용이하다. 이러한 접근법은 기존의 회귀분석을 적용하였을 때 이상점의 영향을 받아 참값의 결과가 오류로 측정되는 문제점을 해결할 수 있다. 또한 기존에 무시되던 짧은 움직임의 변화로 인한 오차를 줄여준다. 분위 회귀분석 모델을 적용하여 이상점을 제거하고 추정 모델 값을 적용하여 자연스러운 움직임을 갖는 모델을 복원한다.

3. 실험결과

본 논문은 Intel i7-7700 3.6GHz CPU 와 NVIDIA GeForce GTX 1080 Ti 를 장착한 컴퓨터에서 수행하였다. 단일 영상에서 3D 복원은 약 4 분의 수행 시간 소요되었고 다중 영상을 이용하여 모델 보정은 평균 30 초의 수행 시간이 소요되었다.

그림 4는 3D 모델의 우측 무릎 관절에 대한 7번과 8번의 자세 매개변수에 대한 분위회귀분석 결과이다. 무릎 관절에 대해 균일한 변화를 나타내는 roll 회전 방향에 대한 결과와 역동적인 변화를 나타내는 pitch 회전 방향에 대한 결과를 확인 할 수 있으며 좌측의 초기 측정 결과와 우측의 분위 회귀 모델을 적용한 결과를 확인했을 때 이상점을 확인할 수 있다. 하지만 이러한 모델은 연속적인 오류에 대해서는 참 값으로 인식하는 문제점이 있다. 분위 회귀모델을 적용한 결과를 3D 모델에 적용한 결과는 그림 5에서 확인할 수 있다. 비디오부터 추출한 연속적인 동작에 대해 자연스러운 변화를 나타내는 것을 확인할 수 있다. 이러한 결과는 초기인체 정보 추출과정에서 해상도의 영향을 받으며 또한 입력 프레임의 수에 영향을 받는다.



그림 5. 비디오로부터 인체 움직임 복원 결과.

4. 결론

본 논문은 단일 영상에 대한 3D 복원을 다중영상으로 확장하며 이상점 검출 및 회귀분석을 통해 인체 모델을 복원하는 기법을 제안하였다. 이러한 접근 방법은 비디오 입력 영상에 대해 기존의 기법에 비해 부드럽고 자연스러운 움직임을 재현할 수 있다. 또한 2D 영상으로부터 측정된 오류를 보정할 수 있어 다중영상을 통한 기존 복원 결과의 보완이 가능하다. 또한 새롭게 제시되는 단일 영상 복원 연구에 적용이 가능하다.

감사의 글

본 논문은 CoAsia Holdings 의 지원을 받아 수행된 연구입니다.

참고문헌

- [1] M. Loper et al., "SMPL: A skinned multi-person linear model," ACM Trans. on Graphics, vol. 34, no. 6, Article No. 248, November 2015.
- [2] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," Proc. of ICCV, pp. 1381-1388, September 2009.
- [3] F. Bogo et al., "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image," Proc. of ECCV, October 2016.
- [4] L. Pishchulin et al., "DeepCut: Joint subset partition and labeling for multi person pose estimation," Proc. of CVPR, June 2016.
- [5] K. He et al., "Mask R-CNN," Proc. of ICCV, October 2017.
- [6] C. Rother et al., "GrabCut: Interactive foreground extraction using iterated Graph Cut," ACM Trans. on Graphics, vol 23, no. 3, pp. 309-314, August 2004.
- [7] K. Grauman and T. Darrell, "Fast Contour Matching Using Approximate Earth Mover's Distance," Proc. of CVPR, June 2004.