

특집논문 (Special Paper)

방송공학회논문지 제24권 제2호, 2019년 3월 (JBE Vol. 24, No. 2, March 2019)

<https://doi.org/10.5909/JBE.2019.24.2.264>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

분위 회귀 분석을 이용한 비디오로부터의 3차원 인체 복원

한 지 수^{a)}, 박 인 규^{a)†}

3D Human Reconstruction from Video using Quantile Regression

Jisoo Han^{a)} and In Kyu Park^{a)†}

요 약

본 논문은 비디오로부터 추출한 프레임으로부터 3차원 인체 형상과 자세 복원을 수행하고 이를 시간 축에서 자연스럽게 부드러운 움직임을 나타내도록 보장하는 기법을 제안한다. 제안하는 기법은 우선 비디오로부터 추출한 개별 프레임으로부터 convolutional neural network을 이용하여 관절의 위치와 인체의 윤곽을 추정한다. 인체의 형상 및 자세는 매개변수 기반의 3차원 변형가능 모델 (morphable model)을 2차원 영상으로 투영후 정합하여 최적의 매개변수 값을 추정한다. 이 때 각 프레임에 대한 복원이 개별적으로 수행되면 시간 축에서 자세의 연속성과 체형의 일관성이 보장되지 못하고 올바르게 복원된 결과가 나타난다. 제안하는 기법은 이러한 문제점을 보완하기 위하여 각 프레임으로부터 복원된 3차원 변형가능 모델의 주성분 매개변수의 분석 및 보간을 수행한다. 실험결과 3차원 인체 복원에 오류가 발생한 프레임에 대해 이전과 이후 프레임들 사이의 관계를 통해 오류가 보정되어 개선된 복원 결과를 얻을 수 있음을 보인다.

Abstract

In this paper, we propose a 3D human body reconstruction and refinement method from the frames extracted from a video to obtain natural and smooth motion in temporal domain. Individual frames extracted from the video are fed into convolutional neural network to estimate the location of the joint and the silhouette of the human body. This is done by projecting the parameter-based 3D deformable model to 2D image and by estimating the value of the optimal parameters. If the reconstruction process for each frame is performed independently, temporal consistency of human pose and shape cannot be guaranteed, yielding an inaccurate result. To alleviate this problem, the proposed method analyzes and interpolates the principal component parameters of the 3D morphable model reconstructed from each individual frame. Experimental result shows that the erroneous frames are corrected and refined by utilizing the relation between the previous and the next frames to obtain the improved 3D human reconstruction result.

Keywords : 3D human body reconstruction, human pose and shape, quantile regression, temporal consistency

a) 인하대학교 정보통신공학과(Inha University, Department of Information and Communication Engineering)

† Corresponding Author : 박인규(In Kyu Park)

E-mail: pik@inha.ac.kr

Tel: +82-32-860-9190

ORCID: <http://orcid.org/0000-0003-4774-7841>

※ 이 논문의 연구 결과 중 일부는 “2018년 한국방송-미디어공학회 추계학술대회”에서 발표한 바 있음.

※ 이 논문은 코아시아 홀딩스의 지원을 받아 수행된 연구임.

※ This work was supported by CoAsia Holdings

· Manuscript received January 15, 2019; Revised March 19, 2019; Accepted March 19, 2019.

Copyright © 2019 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

현실 세계의 인체를 3차원으로 재현하는 연구는 오랜 기간 진행되어 왔으며 매개변수 기반의 3차원 인체 변형 모델인 SMPL (Skinned multi-person linear model)^[1] 모델이 제안된 이후로 빠르게 발전하여 현재 매우 정교한 모델을 복원할 수 있게 되었다. 또한 인체 복원 연구는 딥러닝 기술의 발전으로 인해 2차원 영상으로부터 정확한 인체 정보를 측정할 수 있게 됨에 따라 보다 정교한 3차원 복원 결과를 얻을 수 있게 되었다^{[2][3]}. 하지만 이러한 기존의 연구는 단일 영상에 초점이 맞춰져 있으며, 비디오에 적용하였을 때 시간축으로의 연속성 측면에서 부정확한 복원 결과와 부자연스러운 움직임을 나타내는 한계를 가지고 있다. 비디오에 존재하는 다양한 움직임을 가진 인체는 움직임에 따라 발생하는 가려짐으로 인해 3차원 인체 형상을 정확하게 취득하기 어렵고 복잡한 인체 자세의 경우 복원과정에서 정확한 자세를 표현하지 못하는 어려움이 있다.

본 논문에서는 이러한 기존의 문제를 해결하기 위해 비디오로부터 추출된 각 프레임의 적용된 복원 결과를 분위 회귀분석 기법을 적용하여 오류 프레임을 검출 및 보정하여 자연스러운 체형 및 자세를 나타내도록 하는 기법을 제안한다. 본 논문의 초기 연구 결과는 [4]에 제시되었다. 현실의 인체는 시간의 흐름에 따라 체형은 동일하며 자세는 움직임을 나타내는 일관성과 연속성을 갖는다. 비디오로부터 추출한 다중 프레임의 인체 복원결과에 대하여 각 프레임 사이의 자연스러운 변화를 나타내기 위하여 각 프레임으로부터 복원된 매개변수 값의 분석이 필요하다. 관절 매개변수는 움직임에 따른 자세변화에 따라 일정하게 증가하거나 감소한다. 분위 회귀 분석 기법은 전체 프레임의 복원 데이터를 구간별로 나누어 분석이 가능하여 자세변화에 보다 정밀하게 보정이 가능하다. 이러한 접근 방식은 이전 프레임과 현재 프레임을 이용하여 현재 프레임의 오차 여부를 판단하는 기존의 방법에 비해 전체 프레임에 대하여 분석을 수행함에 따라 정확한 결과를 확인할 수 있다. 그 결과 분석을 통하여 오류 프레임을 보정하고 부자연스러운 인체의 변화를 나타내는 문제점을 해결하여 3차원 인체가 현실에 가까운 체형을 나타내고 자연스러운 움직임을 나타낼 수 있도록 한다.

II. 기존의 연구

기존의 많은 논문은 영상으로부터 인체를 3차원으로 복원하는 방법을 제시해 왔다. 이러한 기존의 기법들은 영상으로부터 인체의 정보를 추출하는 단계와 인체 정보를 이용하여 3차원 메쉬 모델로 복원하는 2단계로 이루어져 있다. 인체의 정보는 크게 자세 정보와 체형 정보로 나뉘며 많은 인체 복원 연구자들이 공유한 HumanEva^[5]와 Human3.6M^[6]과 같은 초기 영상의 인체 정보 데이터셋을 이용하여 2차원 관절 위치 또는 3차원 관절 위치를 추정하는 연구가 진행되어 왔으며 최근 딥러닝 기술의 발전으로 더욱 빠르게 발전되어왔다. 그러한 관절 데이터는 3차원 인체 복원을 위한 자세 정보로 사용된다. 체형 정보의 경우 특정 객체에 대한 윤곽 추정 연구는 컴퓨터 비전 분야에서 꾸준히 연구되어왔다^[7]. 그러나 인체의 경우 옷을 입고 있기 때문에따라 옷 내부의 가려진 신체 정보를 측정하기 어려운 경우가 많다. 이러한 한계로 인해 몸에 밀착하는 옷을 입고 취득된 데이터셋이 기존까지 인체 정보 취득을 위해 주로 사용되어 왔다.

영상으로부터의 인체 복원 기술은 자세와 형상 매개변수 값의 변화에 따라 3차원 모델이 변형되는 3차원 변형가능 모델 (3D morphable model)을 기반으로 발전해왔다. 3차원 스캐너로 스캔한 방대한 3차원 인체 메쉬 데이터의 주성분 분석(principal component analysis)을 통해 생성된 매개변수 모델을 이용하여 현실적인 인체의 모습을 매개변수 값의 변화에 따라 효율적으로 표현 가능하다. 이러한 점을 이용하여 앞서 취득한 2차원 인체 정보와 3차원 모델 매개변수 값 사이의 회귀 모델을 설정하고 최적화를 통해 3차원 모델을 변형함에 따라 영상 내 인체와 유사한 3차원 인체 모델을 효과적으로 복원할 수 있다^{[8][9]}. 현재의 3차원 인체 복원 연구는 딥러닝 기술의 발전으로 취득한 인체 정보의 신뢰성이 높아지고 다양한 데이터 셋으로부터 딥러닝 모델을 학습하기 용이해짐에 따라 그 결과는 더욱 현실에 가까워지게 되었다. 그러나 이러한 기존의 기법은 체형 정보 측정 및 적용의 어려움으로 인해 자세 변형에 초점이 맞춰져 있었다. 또한 인체의 자세 복원의 경우 단일 영상 기반의 연구가 주로 수행됨에 따라 비디오 같은 연속적인 프레임으로부터 나타나는 복잡한 자세를 가지는 인체의 복원의

경우 어려움이 존재한다.

III. 제안하는 기법

본 논문에서 제안하는 기법은 비디오로부터 추출한 전체 프레임에 대해 개별 2차원 영상에서 독립적으로 3차원 인체 복원을 수행하는 단계와 복원된 모델의 매개변수로부터 이상점 검출과 회귀분석을 이용하여 체형 및 자세를 보정하는 두 단계로 구성된다. 그림 1은 비디오로부터 추출한 프레임의 인체 복원에 대한 파이프라인을 제시한다.

1. 개별 2차원 영상으로부터의 3차원 인체 자세 및 체형 복원

본 논문에서 사용한 매개변수 모델은 23개의 관절에 의해 변형되며 14개의 주요 관절의 움직임에 의해 자세가 결정된다. 이러한 모델링 방식은 기존의 꼭지점과 골격 구조를 관계시키는 방식의 사실성이 떨어진다는 문제점을 해결하기 위해 자세의 변화를 각 관절들의 회전 행렬 요소들에 의한 선형 함수로 공식화하였다. 도출된 목적 함수는 등록

된 메쉬와 학습된 데이터로부터 생성된 모델 간의 차이를 없애고 정렬된 메쉬 정보와 최적화된 데이터를 기반으로 통일화 된 모델의 변형을 공식화한다. 공식화된 모델의 매개변수 값을 변형하여 다양한 체형과 자연스러운 자세변화를 나타낼 수 있도록 하였다.

2차원 영상으로부터의 3차원 인체 복원은 인체의 정보를 추출하여 매개변수 기반 모델과의 정합을 통해 수행된다. 인체의 자세정보는 Fast R-CNN^[2]을 사용하여 2차원 영상으로부터 손, 발과 같은 부위를 제외한 전체적인 자세의 변형을 나타내는 13개의 주요 부위를 추정하고 이를 그룹화하여 관절의 위치를 측정한다. 측정된 관절 위치 정보는 3차원 모델로부터 2차원으로 투영된 관절 위치 사이의 오차를 줄이는 목적함수를 최적화하는 과정을 통해 이루어진다. 자세와 체형 β, θ 매개변수 값을 찾기 위한 목적 함수는 식 (1)와 같이 표현되며,

$$E(\beta, \theta) = E_j(\beta, \theta; k, J_{est}) + \lambda_\theta E_\theta(\theta) + \lambda_{sp} E_{sp}(\theta; \beta) + \lambda_\beta E_\beta(\beta) \quad (1)$$

K 는 카메라 매개변수, $\lambda_\theta, \lambda_a, \lambda_{sp}, \lambda_\beta$ 는 스칼라 가중치, J_{est} 는 측정된 관절 위치를 나타내며 자세와 체형을 나타낸다^[9]. 측정된 관절의 정보는 2차원 평면 상의 정보이며 깊이

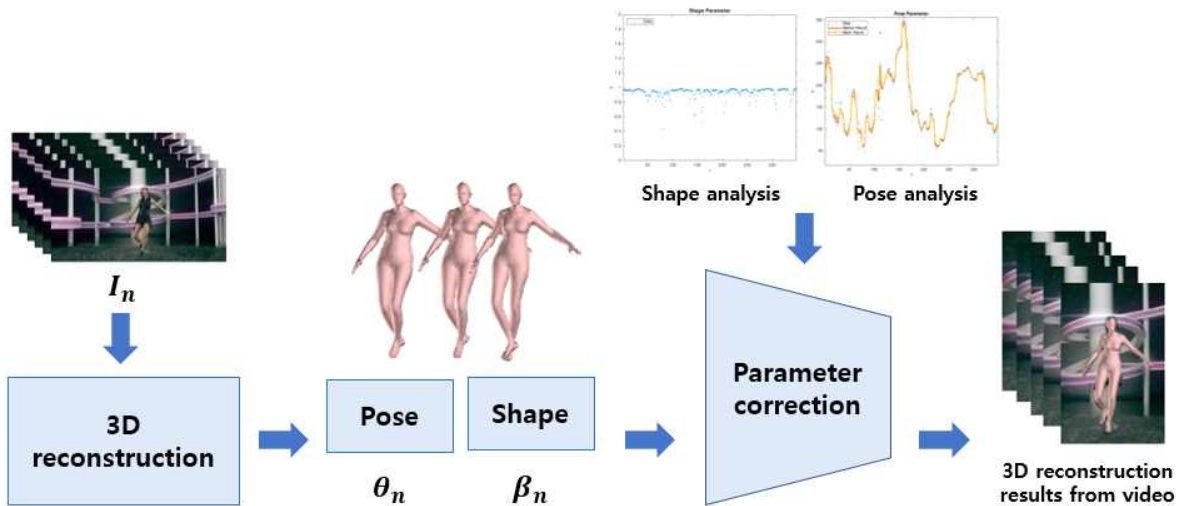


그림 1. 비디오로부터의 3차원 인체 복원 파이프라인
Fig. 1. Pipeline of 3D human reconstruction from video

를 추정하기 위해서는 매개변수 모델의 어깨와 골반 사이의 거리를 통해 상체 길이와 측정된 관절 위치에 따른 유사한 삼각 비율에 의해 관절의 깊이 정보가 추정된다. 측정된 매개변수 값은 회전 반경에 대해 Rodrigues 공식을 적용하고 회전 값으로 변환되어 3차원 변형 가능 인체 모델을 변형한다.

체형은 인체의 윤곽 정보를 기반으로 복원되지만 관절 위치를 기반으로 정의된 식 (1)에서 개략적인 체형 매개변수 값을 찾을 수 있다. 즉, 인체의 전체적인 비율을 정의하기 위해 어깨와 골반의 관절 위치 정보를 적용하며 이러한 과정에서 대략적인 체형이 복원된다. 그러나 관절로부터 복원된 형상 모델은 세부적인 형상 차이를 재현하지 못하는 한계가 있으며 본 논문에서는 윤곽 정보를 이용하여 이것을 보완한다. 윤곽 정보는 바운딩 박스에 의해 객체를 추정하는 기존의 방식을 마스크로 표현하는 Mask R-CNN 기법을 통해 얻는다.

Mask R-CNN을 통해 측정된 윤곽 마스크는 Grabcut 기법에 의해 보다 정교하게 가공된다. 측정된 윤곽 정보는 자세가 복원된 3차원 모델로부터 추출된 마스크와 정합하여 최적의 체형 매개변수를 찾게 된다. 최적의 값을 찾기 위해 두개의 확률 분포 사이의 최소 비용을 계산하는 EMD 공식을 적용하여 2차원과 3차원의 윤곽 정보 비교 값을 도출하고 Levenberg-Marquardt (LM) 함수 최적화 기법을 적용하여 적합한 체형 매개변수 값을 도출한다. 그림 2는 2차원 영상으로부터 측정된 자세와 체형 매개변수 최적화 기법을 적용하여 단일 영상으로부터 3차원 인체 복원을 수행한 결과이다.

2. 비디오로부터 복원된 인체 모델의 체형 및 자세 보정

단일 영상으로부터의 3차원 인체 복원은 오랜 기간 연구가 되어 왔지만 비디오와 같은 연속적인 영상에 적용한 경우 인체 부위의 가려짐에 의한 인체 정보 측정의 어려움이나 복잡한 자세에 의한 3차원 자세 복원의 어려움과 같은 한계가 있다. 본 논문은 비디오로부터 추출한 다중 영상에 인체 복원을 수행하고 매개변수 결과에 이상점 검출 및 분위회귀분석 기법을 적용하여 일관성 및 연속성을 유지하여 현실에 가까운 모습을 재현하는 기법을 제안한다. 단일 대상에 대한 복원의 경우 체형 매개변수는 시간의 흐름에 따라 체형의 변화가 없는 일관성을 유지한다. 전체 프레임의 체형 매개변수 결과 값을 시간의 흐름에 따라 분석하였을 때 일정 임계 값을 적용하여 체형 매개변수의 일관성에 어긋나는 이상점을 검출하여 오류 프레임을 배제할 수 있다.

$$MAD = \text{median}(|S_i - \text{median}(S)|), i = 1, \dots, N \quad (2)$$

전체 프레임 N의 체형 매개변수 S에 대하여 식 (2)와 같이 정의된 중앙값 절대 편차를 구하여 초과된 값인 이상점을 검출하여 배제한다. 이에 따라 오류 프레임을 제외한 체형 매개변수 값은 전체 프레임의 빈도를 측정하여 가중치가 부여된다. 체형 매개변수의 경우 카메라의 각도와 영상 내 인체의 방향에 따라 미세한 변화를 나타낼 수 있기 때문에 가중치에 따른 체형 매개변수들의 평균 값을 구함으로



그림 2. 단일 영상에 대한 3차원 인체 복원
 Fig. 2. 3D human reconstruction from single image

써 일관성을 유지하는 체형 매개변수 값을 구할 수 있다.

움직임을 갖는 3차원 인체 모델의 자세 매개변수는 시간의 흐름에 따라 연속성을 나타낸다. 인체의 주요 관절들은 움직임에 따라 일정한 패턴을 보이며 변화하고 시간의 흐름에 따라 자세 매개변수 값의 변화로 표현된다. 측정 대상의 움직임에 의해 변화하는 인체 관절들의 자세 매개변수 값을 통계적으로 분석하기 위해 회귀 분석을 적용한다. 하지만 인체의 움직임은 항상 일정한 움직임만을 나타내지 않고 동적인 움직임의 변화를 반영하기 위하여 전체 프레임으로부터 복원된 각각의 자세 매개변수 데이터에 대한 분위 회귀 분석을 적용한다^[10]. 분위 회귀분석 기법은 전체의 모형을 추정하기 보다 각 분위 별 변화를 추정할 때 전체 데이터에 대한 평균 베타를 구하기보다 분위에 따른 데이

터 각 구간의 분위 베타를 추정하여 움직임의 변화에 따라 일정하게 증가 및 감소하는 매개변수를 구간 별로 나누어 분석함에 따라 인체의 자세 변화를 보다 정밀하게 분석 가능한 분석 기법이다. 시간에 따른 프레임 순번을 독립 변수 x로 갖고 복원된 자세 매개변수 값을 y로 갖는 분위 회귀 분석은 인체의 움직임이 항상 일정한 변화를 나타내지 않고 변화함에 따라 각 구간에 따라 나누어 분석을 적용해야 하는 자세 분석에 용이하다. 이러한 분위 회귀 분석은 분위 베타를 구하기 위해 목적함수를 최소화하게 되는데 이는 최적화 문제로 귀결된다. 이러한 방식은 회귀 분석에서의 평균 베타를 구하는 것과 유사하며 분위 회귀 분석은 평균 값과 중간값의 차이로 인해 특이값 (outlier)에 영향을 적게 받는 장점이 있다.

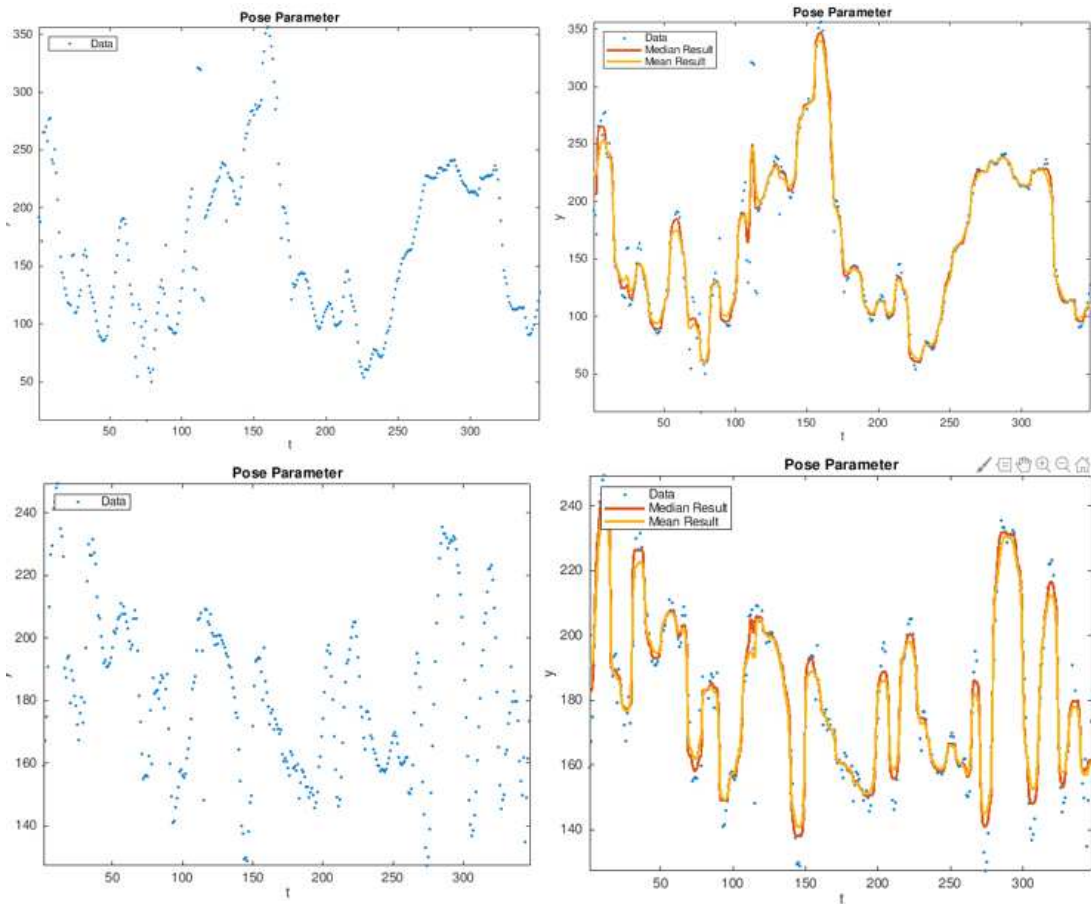


그림 3. 자세 매개변수에 대한 분위 회귀 분석
 Fig. 3. Quantile regression for pose parameters

$$E = \frac{1}{N} * \left\{ (1 - \tau) * \int |(y_i - (a_0 + a_1x + \dots + a_nx^n))| \right\} + \left\{ \tau * \int |(y_i - (a_0 + a_1x + \dots + a_nx^n))| \right\} \quad (3)$$

분위 베타는 식 (3)과 같은 목적 함수를 최소화하여 기대 값을 구할 수 있으며 τ 는 분위에 대한 백분율 값을 나타내며 10분위 베타를 구하는 경우 0.1값을 사용하며 $y_i - (a_0 + a_1x + \dots + a_nx^n)$ 은 관찰치 i 에 대한 회귀 모형에서의 에러 부분을 나타낸다. 분위 회귀 분석은 목적함수를 최소화하는 것이므로 표준 오차를 구하기가 어려움에 따라 bootstrapping을 통해 표준 오차를 구해준다. 데이터셋으로부터 추출한 샘플의 분위 베타를 구하는 과정을 반복하여 표준 편차를 표준 오차로 사용하는 방법이다.

비디오로부터 3차원 복원된 인체 모델들의 자세 매개변수 값에 적용된 회귀 분석 결과는 그림 3에서와 같이 곡선으로 나타나며 임계값을 설정하여 이상점을 제거하고 제거된 이상점을 곡선상의 근사값으로 대체하여 y 값을 예측된 자세 매개변수 값으로 가정한다. 이러한 접근 방식은 각 프레임 사이의 자세 매개변수 간의 연속성을 유지하고 인체 자세 변형 및 초기 복원 오류로 인한 분위 별 오차를 줄여 줌으로써 기존의 접근 방식인 단순히 제거된 오류 프레임의 매개변수 값을 이전 프레임과 이후 프레임 값의 보간 및 평균 값을 적용하는 것에 비해 자연스러운 움직임을 갖는 결과를 확인할 수 있다. 하지만 이러한 회귀 분석의 접근 방식은 연속적인 오류를 갖는 프레임의 복원 결과 매개변수 값을 참값으로 인식하는 문제점과 복원 결과가 입력 프

레이의 수에 영향을 받는 한계가 있다. 또한 전체 프레임에 대해 복원을 수행함에 따라 실시간 재생이 불가능하다는 한계가 있다. 이러한 한계점은 각 프레임에 대한 단일 영상 복원의 발전 및 짧은 프레임 간의 보정 기법에 의해 해결 가능할 것으로 보인다.

IV. 실험 결과

본 논문은 Intel i7-7700 3.6GHz CPU와 NVIDIA GeForce GTX 1080 Ti를 장착한 개인용 컴퓨터에서 실험을 수행하였다. 단일 영상에서 3차원 복원은 약 4분의 수행 시간이 소요되었고 다중 영상을 이용한 모델 보정은 평균 30초의 수행 시간이 소요되었다. 이용한 실험 영상으로는 26개의 유튜브 비디오로 구성된 SHPED (Stereo human pose estimation) 데이터셋을 사용하였다^[11]. 복원 결과를 정량적으로 비교하기 위해 입력 영상의 대상에 대하여 3차원 변형 모델을 이용하여 수동으로 예측 모델을 생성하고 본 논문에서 제안한 기법을 수행한 복원 결과와 기존의 기법인 SMPLify 기법으로부터의 복원 결과, 오류 프레임 검출 및 보간 기법 적용한 결과와의 3차원 관절 위치의 거리를 비교하였다. 해당 결과는 표 1에서와 같이 기존의 기법에 비해

표 1. 기존의 기법과 제안하는 기법의 3차원 관절 위치 비교
Table 1. 3D joints distance between existing technique and proposed technique

	Video no.1 Frame #5	Video no.1 Frame #6	Video no.2 Frame #5	Video no.2 Frame #9	Video no.3 Frame #5	Video no.3 Frame #12	Average
SMPLify	1596.88	1453.28	867.79	1137.67	9641.24	3796.48	3765.68
Error Frame Interpolation	1596.88	1453.28	667.79	966.24	9641.24	3796.48	3432.67
Quantile Regression	1503.45	1634.22	619.71	1083.86	1752.72	952.65	2808.98

표 2. 입력 영상에 대한 오류 프레임 검출 비율
Table 2. Error frame detection rate for input video

	Number of Error Frames	Number of Incorrectly Measured Error Frames	Percentage of Actual Error Frames
Kalman Filter	24	9	65%/
Quantile Regression Model	23	5	78%

낮은 오차를 보이고 있다. 또한 영상 3의 5번 프레임과 12번 프레임의 경우 기존 기법을 적용하였을 때와 비교하였을 때 제안하는 기법은 분위 회귀 모델로부터 매개변수 값을 예측하여 현저히 낮은 오차를 갖는 것을 알 수 있다. 이러한 결과는 단편적인 오류 프레임으로 인해 3차원 모델의 움직임의 부자연스러움을 쉽게 보정 및 검출할 수 있는 것을 나타낸다. 하지만 연속적인 오류 프레임에 대해서는 회귀 분석 과정에서 참값으로 결정됨에 따라 보정에 어려움이 있다. 이는 표 2에서 실제 오류 프레임 수에 대한 측정 오류 프레임 수의 비율을 확인할 수 있다. 입력 100개의 프레임 중 23개의 오류 프레임에 Kalman filter를 적용하여 이상치를 검출하는 기존 기법^[4]과 본 논문에서 제시하는 매개변수 분석을 통하여 이상점을 검출하는 기법을 비교하였다. 그 결과 기존 기법이 오류 프레임 검출에 보다 민감하게 반응하지만 잘못 측정되는 경우가 많은 것을 확인할 수 있

다. 또한 실제 오류 프레임에 대하여 새롭게 제시하는 기법이 이전 기법에 비해 정확한 측정 결과를 나타내는 것을 확인할 수 있다. 하지만 두가지 기법 모두 연속적인 오류 프레임에 대하여 참값으로 인식하여 오류 프레임 검출에 실패함에 따라 실제 오류 프레임에 대한 측정 비율이 떨어지는 것을 확인할 수 있다.

비디오의 인체에 대한 복원 결과는 그림 4에서 확인할 수 있다. 좌측으로부터 연속적인 움직임을 나타내는 자연스러운 인체 변형 결과를 확인할 수 있다. 또한 초기 영상을 살펴보면 발목이 겹쳐 있는데 인체 정보를 취득 과정에서 우측과 좌측이 반대로 측정되는 경우가 존재하였다. 그러한 경우는 그림 5에서 확인 가능하며 3차원 복원 결과 다리의 모습이 다르게 나타난다. 또한 입력 영상의 수에 따라 그 결과가 다르게 나오는 것을 알 수 있다. 이러한 입력 영상에 따른 제약조건은 해상도, 가려짐, 대상의 복잡도 등

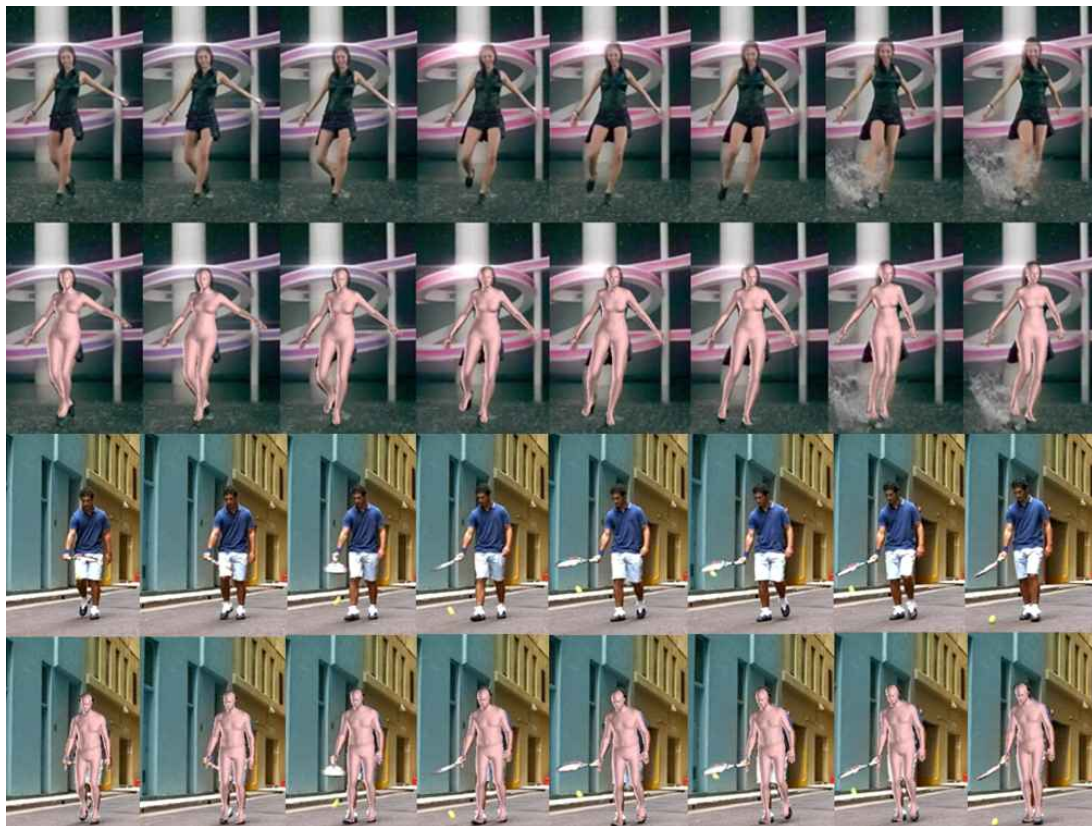


그림 4. 비디오로부터 추출한 연속적인 입력 프레임 영상과 복원된 움직임을 갖는 3차원 인체 모델

Fig. 4. 3D human body reconstruction results with continuous input frame image and motion extracted from video in order

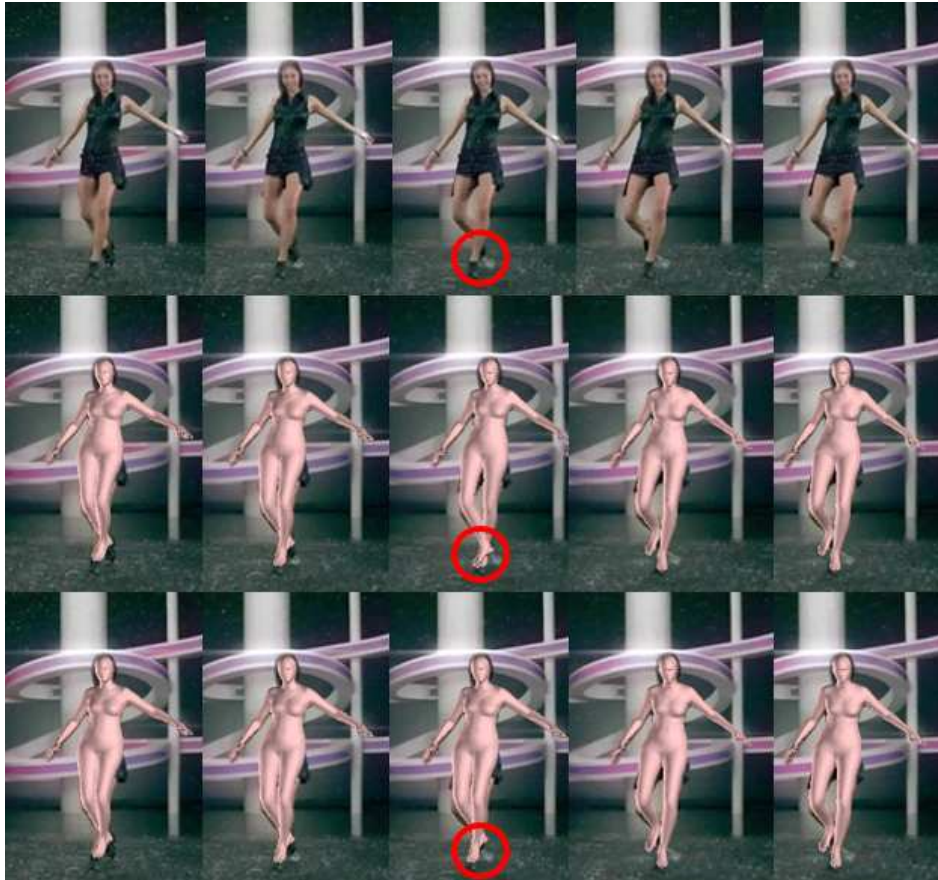


그림 5. 오류 프레임에 대한 3차원 인체 움직임 복원 결과. 위로부터 입력영상, 오류 프레임이 포함된 3차원 인체 복원 영상(중앙 3번째 프레임), 제시된 기법으로 움직임 보정된 복원 영상

Fig 5. 3D Human motion reconstruction result on error frame. From up to down: input, 3D reconstruction including error frame, motion compensated result

이 있으며 초기 측정 및 복원에 따라 비디오 전체의 복원 결과도 영향을 받는다. 또한 체형의 경우 현재 착용 중인 옷의 종류에 따라 오차가 나타날 수 있다. 치마나 부피가 큰 의상은 체형 정보를 측정하는 과정에서 오차를 야기함에 따라 복원에 어려움이 있다. 이것은 최근 제안된 의상 아래의 체형 추정 기법을 사용하여 해결할 수 있을 것으로 보인다^[2].

V. 결론

본 논문에서는 매개변수 기반 3차원 인체 복원 기법을

비디오로 확장하여 기존의 단일 영상 기반의 기법을 적용하였을 때 체형의 일관성과 자세의 연속성을 침해하는 문제를 해결하는 보정 기법을 제시하였다. 비디오로부터 추출한 영상에 단일 영상으로부터의 복원 기법을 적용하고 이상점 검출과 분위 회귀 분석을 이용하여 오류 프레임을 검출하고 보정한다. 이러한 접근 방식은 현실에 가까운 자연스럽게 정확한 인체를 복원할 수 있다. 제안하는 접근 방식은 전체 프레임을 분석하여 보정으로 입력 프레임의 수에 따라 결과가 달라진다. 이러한 부분은 향후 적은 프레임으로 복원이 가능하여 실시간으로 인체의 형상 및 움직임을 복원할 수 있도록 확장할 수 있다.

참 고 문 헌 (References)

- [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, November 2015.
- [2] L. Pishchulin, E. Insafutdinov, S. Tang, and B. Andres, "DeepCut: Joint subset partition and labeling for multi person pose estimation," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4929-4937, June 2016.
- [3] K. Grauman and T. Darrell, "Fast contour matching using approximate earth mover's distance," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 2004.
- [4] J. S. Han, M. R. Cho, and I. K. Park, "Moving human shape and pose reconstruction from video," *Proceedings of Korean Society of Broadcast and Media Engineers Fall Conference*, pp. 66-68, November 2018.
- [5] L. Sigal, A. Balan, and M. J. Black, "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion," *International Journal of Computer Vision*, vol. 87, no. 1-2, pp. 4-27, March 2010.
- [6] C. Ionescu, D. Papva, V. Olaru, and C. Sminchisescu, "Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325-1339, July 2014.
- [7] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *Proceedings of IEEE International Conference on Computer Vision*, pp.2980-2988, October 2017.
- [8] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," *Proceedings of IEEE International Conference on Computer Vision*, pp. 1381-1388, September 2009.
- [9] F. Bogo, A. Kanazawa, C. Lassner, and P. Gehler, "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image," *Proceedings of European Conference on Computer Vision*, pp. 561-578, October 2016.
- [10] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, no. 6, pp. 983 - 999, June 2006.
- [11] M. Lopez-Quintero, M. Marin-jimenez, et al. "Stereo pictorial structure for 2D articulated human pose estimation," *Machine Vision and Applications*, vol. 27, no. 2, pp. 157-174, February 2016.
- [12] C. Zhang, S. Pujades, M. J. Black, and G. Pons-Moll, "Detailed, accurate, human shape estimation from clothed 3D scan sequences," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, July 2017.

저 자 소 개



한 지 수

- 2017년 2월 : 인하대학교 정보통신공학과 학사
- 2017년 3월 ~ 현재 : 인하대학교 정보통신공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-0583-8967>
- 주관심분야 : 컴퓨터비전 및 그래픽스 (영상기반 3차원 형상 복원, 증강현실, computational photography)



박 인 규

- 1995년 2월 : 서울대학교 제어계측공학과 학사
- 1997년 2월 : 서울대학교 제어계측공학과 석사
- 2001년 8월 : 서울대학교 전기컴퓨터공학부 박사
- 2001년 9월 ~ 2004년 2월 : 삼성중흥기술원 멀티미디어랩 전문연구원
- 2007년 1월 ~ 2008년 2월 : Mitsubishi Electric Research Laboratories (MERL) 방문연구원
- 2014년 9월 ~ 2015년 8월 : MIT Media Lab 방문부교수
- 2018년 7월 ~ 2019년 2월 : University of California, San Diego (UCSD) 방문학자
- 2004년 3월 ~ 현재 : 인하대학교 정보통신공학과 교수
- ORCID : <http://orcid.org/0000-0003-4774-7841>
- 주관심분야 : 컴퓨터비전 및 그래픽스 (영상기반 3차원 형상 복원, 증강현실, computational photography), GPGPU