



특집논문 (Special Paper)

방송공학회논문지 제29권 제2호, 2024년 3월 (JBE Vol.29, No.2, March 2024)

<https://doi.org/10.5909/JBE.2024.29.2.141>

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

파라미터 기반 3차원 휴먼 모델 사전 정보 정확도의 내재 함수기반 세부적 의복 착용 3차원 휴먼 모델 생성에 미치는 영향에 관한 연구

엄기문^{a)}, 송찬들^{b)}, 김원준^{b)}, 이희경^{a)}, 양승준^{a)}, 정원식^{a)}, 추현곤^{a)}

A Study on the Impact of Parameter-based 3D Human Model Prior Information Accuracy in Generating Detailed Clothed 3D Human Models based on Implicit Function

Gi-Mun Um^{a)}, Chandeul Song^{b)}, Wonjun Kim^{b)}, Hee Kyung Lee^{a)}, Seung-Jun Yang^{a)},
Won Sik Cheong^{a)}, and Hyon-Gon Choo^{a)}

요약

단일 또는 다시점 영상에서 딥러닝 기반으로 3차원 휴먼 모델을 복원하는 대표적인 기법들인 파라미터 기반 기법과 내재 함수 기반 기법은 모두 장단점이 있다. 본 논문에서는 이 두 기법의 상호 보완적 융합을 위한 사전 연구로서, 파라미터 기반 휴먼 모델을 사전 정보로 세부적 의복 착용 휴먼 모델 생성 시 파라미터 기반 모델의 정확도가 미치는 영향을 알아보았다. 실험 결과, 파라미터 기반 휴먼 모델의 정확도가 상대적으로 낮더라도 최종 세부적 의복 착용 휴먼 모델 생성 결과에서는 대부분 영상에서 정확도가 향상되어 파라미터 기반 휴먼 모델 정확도 순서 영향이 크지 않았다. 정량적 비교 결과에서도 SOTA 기법 간 정확도 차이가 큰 경우에는 파라미터 기반 휴먼 모델의 영향이 일부 존재했지만, 정확도 차이가 적을 때에는 영향이 상대적으로 적었다. 따라서 파라미터 기반 휴먼 모델 정확도가 어느 정도 확보되면 모델 안정성을 유지하면서 세부 형태를 복원 가능한 기법 개발이 가능할 것으로 판단된다.

Abstract

Both parameter-based techniques and implicit function-based techniques that are representative techniques for reconstructing 3D human models based on deep learning from single or multi-view images have advantages and disadvantages. In this paper, as a preliminary study on the complementary fusion of these two techniques, we investigated the impact of the accuracy of the parameter-based human model when we use the parameter-based 3D human model as a prior information for reconstructing the detailed clothed human model. As results of the experiment, although the accuracy of the parameter-based human model was relatively slightly lower, the accuracy of the final detailed clothed human model was improved in most images, so the effect of the accuracy order of the parameter-based human model was not significant. In the quantitative comparison results, we found that the accuracy of the parameter-based human model had some impact on the results of the detailed clothed human model generation if the accuracy difference between SOTA techniques was large, but the impact was relatively small when the accuracy difference was small. Therefore, once the accuracy of the parameter-based 3D human model is secured to a certain degree, we believe that it is possible to develop a technique that can reconstruct detailed shapes while maintaining model stability.

Keyword : 3D Human Model, Parameter-based human model, Detailed 3D human clothed model, Implicit function

Copyright © 2024 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

2021년에 스웨덴 출신의 전설 팝 그룹인 ABBA®에서 모션 캡처 기술을 통해 4명의 남녀 멤버의 가상 캐릭터 제작과 이에 의한 가상 휴먼에 의한 콘서트가 소개되면서 실제와 가까운 휴먼 모델 제작 기술에 관한 관심과 연구가 증가하고 있다^{[1][2][3]}.

일반적으로 3차원 휴먼 모델을 복원하는 기법은 크게 두 가지로 분류할 수 있는데, 첫 번째 기법은 보통 복원할 대상인 사람을 볼륨메트릭(volumetric) 스튜디오에서 수십 대의 카메라로 사람의 동작과 형태를 촬영하여 모델을 생성하는 과정을 거쳐서 복원하는 기법이며, 이 기법의 장점은 상대적으로 고품질인 3차원 휴먼 모델을 획득할 수 있다는 장점이 있는 반면에 스튜디오 구축을 위한 공간이나 비용이 많이 소요되며, 처리 시간에 대한 제약이 많은 실정이다. 다음으로 두 번째 방법은 최근 인공지능 기술이 발전하면서 단일 또는 소수의 RGB 컬러 카메라로 촬영한 영상만을 이용하거나 부가적으로 범용 깊이 카메라^{[4][5]}로 촬영한 깊이 영상을 입력받아 인공지능 신경망을 이용하여 사람의 3차원 동작과 자세, 형태 정보를 복원하는 연구가 증가하고 있다^{[6][7][8]}. 이 방법은 첫 번째 방법에 비해 복원정확도는 상대적으로 낮지만, 적은 수의 카메라로 3차원 휴먼 모델을 생성할 수 있으며, 위치나 공간의 제약이 적다는 장점이 있다. 이 중 두 번째 분류에 속하는 기법들은 생성되는 3차원 휴먼 모델의 표현 형태에 따라 SMPL(Skinned Multi-Person Linear Model)^[9] 기법과 같이 파라미터 기반 모델을 생성하는 기법과, PaMIR(Parametric Model-Conditioned Implicit Representation) 기법^[6] 등 내재 함수(implicit function) 기반의 모델을 생성하는 기법이 있다. 먼저 파라미터 기반 모

델 생성 기법은 상대적으로 적은 계산량으로 사람의 형태와 자세를 안정적으로 생성할 수 있지만, 의복을 입지 않은 기본 모델 형태로 복원만 가능하다^[9]. 반면에 내재 함수 기반의 모델 생성 방법은 계산량이 많고, 입력 영상에 따라 생성 정확도가 안정적이지 못한 단점이 있지만 의복의 주름 등 세부적인 형태 복원이 가능하다는 장점이 있다. 파라미터 기반 모델 생성 기법으로는 ROMP(Regress One-stage Multiple 3D People)^[10], PARE(Part Attention Regressor)^[11], PyMAF(Pyramidal Mesh Alignment Feedback Loop)-x^[12] 기법 등이 속하며, 내재 함수 기반 모델 생성 기법으로는 PaMIR^[6], PIFu(Pixel-aligned Implicit Function)^[13], PIFuHD^[14] 기법 등이 포함된다. 또한 이러한 두 기법의 장단점을 보완하기 위해 최근에는 IP-Net(Implicit Part Network)^[15], ICON(Implicit Clothed humans Obtained from Normals)^[16], ECON(Explicit Clothed humans Optimized via Normal integration)^[17] 등의 기법들도 제안되고 있다.

본 논문에서는 두 기법의 장점을 활용하고, 단점을 상호 보완하기 위한 융합 기법에 관한 사전 연구로서, 세부적 의복 착용 3차원 휴먼 모델 생성을 위한 사전 정보로 파라미터 기반 휴먼 모델을 사용하는 융합 기법에서 파라미터 기반 3차원 휴먼 모델의 정확도가 세부적 의복 착용 3차원 휴먼 모델의 정확도에 미치는 영향을 비교실험으로 알아보았다. 이 때 사용된 대표적인 파라미터 기반 3차원 휴먼 모델 생성 SOTA 기법들은 사전학습(pre-trained) 모델을 이용한 생성 결과를 사용하였다. 또한 파라미터 기반 휴먼 모델을 입력으로 받는 베이스라인 신경망으로는 ECON 기법^[17]을 사용하였다.

II. 비교 대상 파라미터 기반 3차원 휴먼 모델 생성 기법과 비교 실험 신경망 구조

본 논문에서 ECON^[17] 신경망의 2.5차원 표면과 전후면 노말 맵을 복원하기 위해 입력으로 사용하기 위해 선정된 파라미터 기반 3차원 휴먼 모델 생성 기법들은 최근 컴퓨터 비전 학회에서 발표된 논문들 중에서 Github 상에 소스가 공개된 기법들 중심으로 선정하였다. 선정된 기법들은 ROMP^[10], PARE^[11], PKCN^[18], 4D Humans^[19], PyMAF-x^[12]이다. 먼저 ROMP 기법^[10]은 전체 영상으로부터 다중 분리

a) 한국전자통신연구원 실감미디어연구실(Immersive Media Research Section, Electronics and Telecommunications Research Institute(ETRI))

b) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)

‡ Corresponding Author : 엄기문(Gi-Mun Um)

E-mail: gmum@etri.re.kr

Tel: +82-42-860-1084

ORCID: <https://orcid.org/0000-0002-0929-0695>

※ 본 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2018-0-00207, 이머시브 미디어 전문연구실)

· Manuscript 22 January, 2024; Revised 20 February, 2024; Accepted 21 February, 2024.

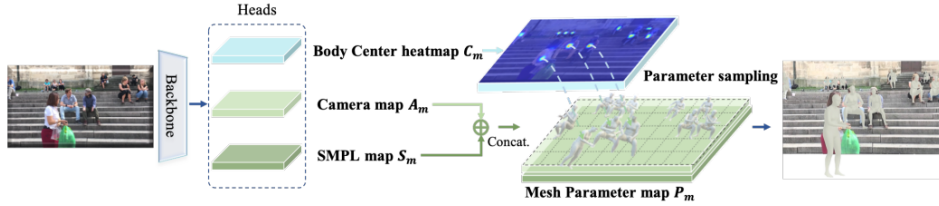


그림 1. ROMP 신경망 구조^[10]
 Fig. 1. The Structure of the ROMP Network^[10]

가능한 세가지 맵인 신체 중심 히트맵, 카메라 맵, SMPL 파라미터 맵을 추정하며, 신체 중심별로 휴먼 3D 자세와 형태를 복원하는 기법이다. 그림 1은 ROMP 기법의 구성도를 나타내고 있다.

다음으로 선정된 기법은 PARE 기법^[11]으로서 그림 2와 같이 신체 부위별로 주목되는 부위를 알아내고 주목되는 가중치를 학습하고, 보이는 영역의 특징을 취합하여 가려진 부위의 성능을 개선한 기법이다.

세 번째로 선정된 기법은 PKCN 기법^[18]으로서 가려짐이 심한 영상에서 정확도 저하되는 것을 보완하기 위해 그림 3과 같이 신체 중심 위치에 기반한 신체 부위 주목(body part attentions)을 예측하고, 그 특징 및 운동학적 연결 관계 기반 디코더를 이용하여 파라미터 기반 3차원 휴먼 모델을 추정하는 기법이다.

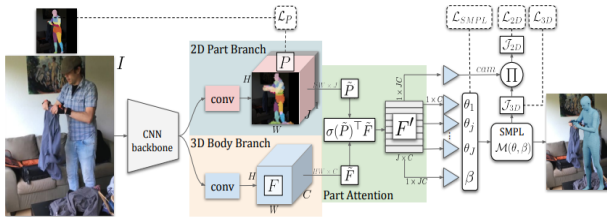


그림 2. PARE 신경망 구조^[11]
 Fig. 2. The Structure of the PARE Network^[11]

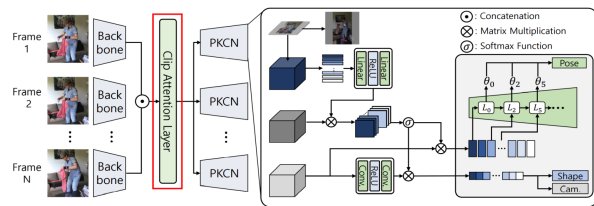


그림 3. PKCN 신경망 구조^[18]
 Fig. 3. The Structure of the PKCN Network^[18]

네 번째로 선정된 기법은 4D Humans^[19] 기법으로서 그림 4와 같이 비전 트랜스포머(vision transformer)와 교차 주목 트랜스포머 디코더(cross-attention transformer decoder)를 기반으로 3차원 휴먼 모델을 복원하며, 추가적으로 시간 축 상에서 객체 위치 추적 기능까지 제공하는 기법이다.

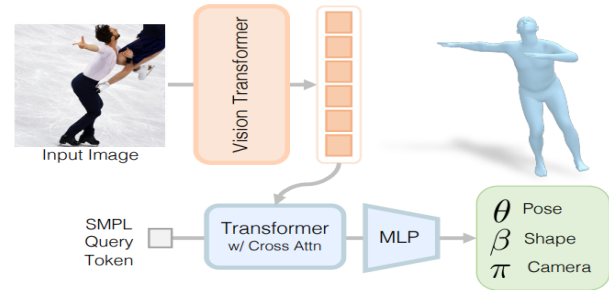


그림 4. 4D Humans 신경망 구조^[19]
 Fig. 4. The Structure of the 4D Humans HMR2.0 Network^[19]

다섯 번째로 선정된 기법은 PyMAF-x 기법^[12]으로서, 특징 피라미드를 구성하고 이로부터 예측한 파라미터의 정확도를 개선하는 regression 기반 PyMAF 기법을 확장하여 손목과 팔 부분의 미세 정확도를 추가로 개선한 기법이다. 이 기법은 그림 5에 나타난 바와 같이 몸통과 손, 얼굴 별로 PyMAF 신경망^[20]을 개별적으로 적용한 후 통합하는 구조로 되어 있다.

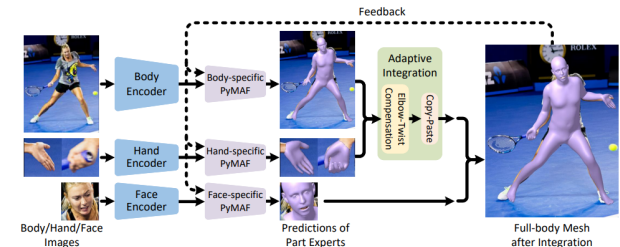


그림 5. PyMAF-x 신경망 구조^[12]
 Fig. 5. The Structure of the PyMAF-x Network^[12]

이렇게 선정된 5종의 파라미터 기반 3차원 휴먼 모델 생성 기법으로 생성된 결과를 입력으로 사용하여 세부적 의복 착용 3차원 휴먼 모델 생성을 수행하는 융합 기법의 베이스라인 신경망으로는 그림 6과 같은 구조를 가지는 ECON 신경망^[17]을 사용하였다.

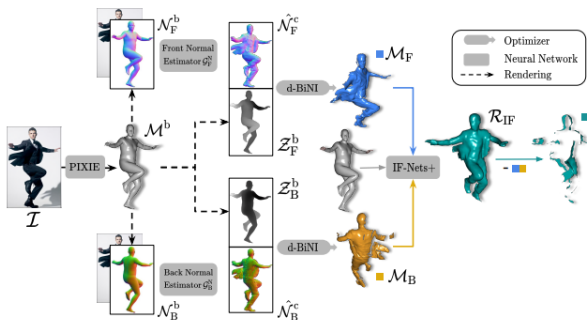


그림 6. ECON 신경망 구조도^[17]
Fig. 6. The Structure of the ECON Network^[17]

III. 3차원 휴먼 모델 생성 비교 실험 결과

먼저 II장에서 선정된 파라미터 기반 3차원 휴먼 모델 생성 기법의 정확도를 정상적으로 비교 분석하기 위해 단일 객체를 포함한 정지영상 5종을 사용하여 비교 실험을 수행하였다. 비교 실험에 사용된 신경망 학습모델은 사전 학습된 모델을 사용하였다.

그림 7~그림 16은 입력 영상 및 각 SOTA 기법에 의해 생성된 파라미터 기반 3차원 휴먼 모델 생성 결과 및 이 결과를 ECON 신경망의 입력으로 사용하여 세부적 의복 착용 3차원 휴먼 모델을 생성한 비교실험 결과의 예를 나타내고 있다.

각 그림에서 점선 부분은 다른 기법에 비해 적색의 점선으로 표시된 부분은 실제 자세나 형태에 비해 정확도가 떨어지는 부분을 나타내고 있고, 파란색 실선 부분은 다른 기법에 비해 정확도가 우수한 부분을 나타내고 있다.

먼저 그림 7의 테스트 영상 1에 관한 결과를 보면, PARE,

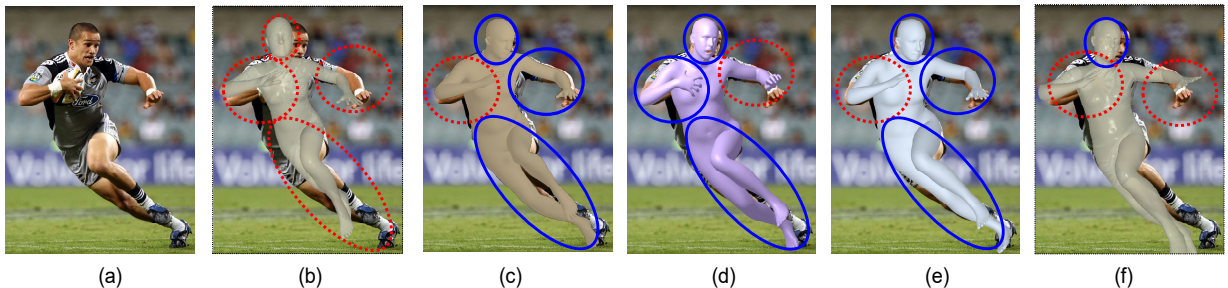


그림 7. 테스트 영상 1에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
Fig. 7. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 1 (a) Input test image (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

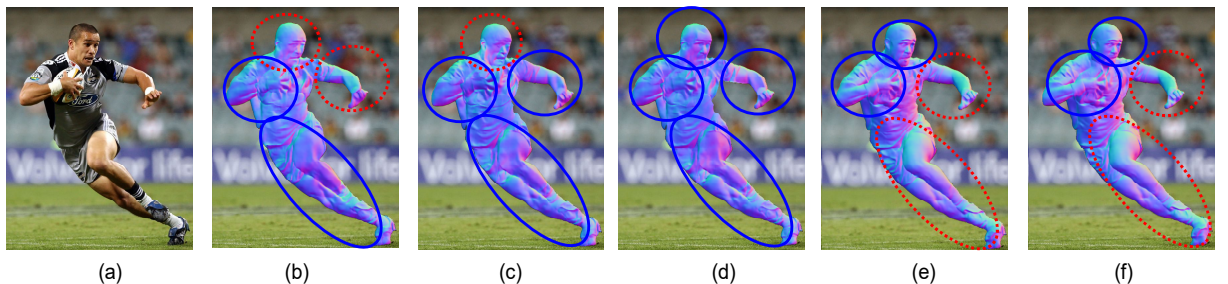


그림 8. 테스트 영상 1에 대한 각 기법의 파라미터 기반 모델을 이용한 세부적 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
Fig. 8. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 1 (a) input test image 1 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

PyMAF-x, 4D Humans 기법 적용 결과가 얼굴 방향, 손과 팔, 다리의 방향 측면에서 비교적 좋은 결과를 보여 주고 있고, PKCN 기법에 의한 생성 결과, ROMP 기법에 의한 생성 결과 순으로 정확도가 낮아지고 있다. 한편, 그림 7의 파라미터 기반 3차원 휴먼 모델을 ECON 신경망에 각각 입력하여 세부적 의복 착용 3차원 휴먼 모델을 생성한 결과가 그림 8에 나타나 있다. 그림 8의 결과를 얼굴 방향, 손과 팔, 다리 각도 측면에서 살펴보면, PyMAF-x 기법이 대부분 신체 부위에서 가장 정확하고, 디테일이 잘 표현되는 결과를 보였으며, PARE 기법에 따른 결과가 다음으로 좋은 결과를 보였고, 마지막으로 ROMP, PKCN, 4D Humans 기법 등에 의한 결과가 왼팔과 손 부분의 디테일에서 차이를 보이고는 있으나, 거의 유사한 순서로 디테일이 잘 표현되는 결과를 보였다. 이로부터 그림 7의 결과에 대한 정확도순서와 그림 8의 정확도순서가 반드시 일치하지는 않음을 알 수 있었다.

다음으로 그림 9의 테스트 영상 2에 관한 파라미터 기반

의 3차원 휴먼 모델 생성 결과를 보면, PyMAF-x 기법이 다른 기법들에 비해 상대적으로 가장 좋은 결과를 보이고 있고, PARE와 4D Humans 기법이 다음으로 좋은 결과를 보여 주고 있으며, 다음으로 PKCN 기법, ROMP 기법 순으로 좋은 결과를 보여 주고 있다. 또한 그림 10에 나타난 바와 같이 그림 9의 결과를 입력하여 세부적 의복 착용 3차원 휴먼 모델을 생성한 결과에서는 ROMP 기법에 의한 생성 결과에서 손의 한쪽 부분 모양이 약간 부정확한 것을 제외하고, PARE, PKCN, 4D Humans, PyMAF-x 기법들로부터 생성한 모델을 이용한 결과가 거의 모든 부위에서 유사하게 디테일이 잘 표현되는 결과를 보였다.

세 번째로 그림 11에 나타난 테스트 영상 3은 사람의 모습이 뒤집혀 있으며, 몸의 방향이 대각선 방향으로 되어 있는 영상으로 복원 난이도가 높은 영상임을 볼 수 있다. 먼저 파라미터 기반 3차원 휴먼 모델 생성 결과를 보면, 여기서는 4D Humans 기법이 대부분의 신체 부위에서 가장 좋은

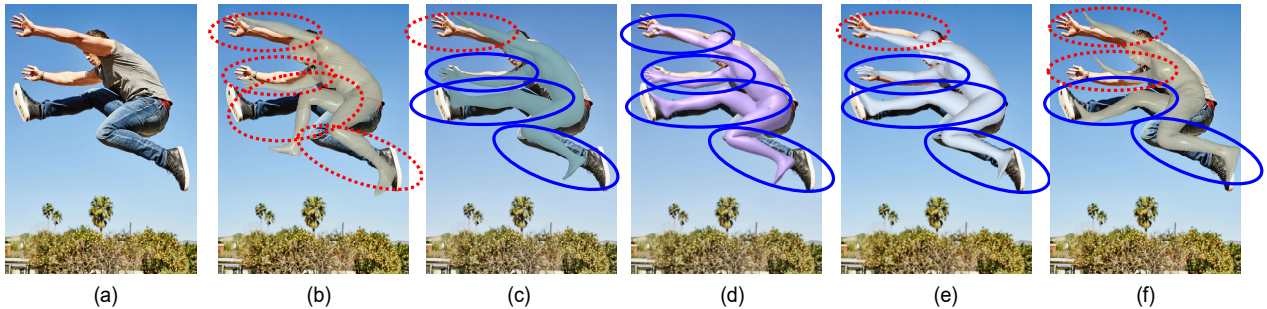


그림 9. 테스트 영상 2에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
 Fig. 9. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 2 (a) Input test image 2 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

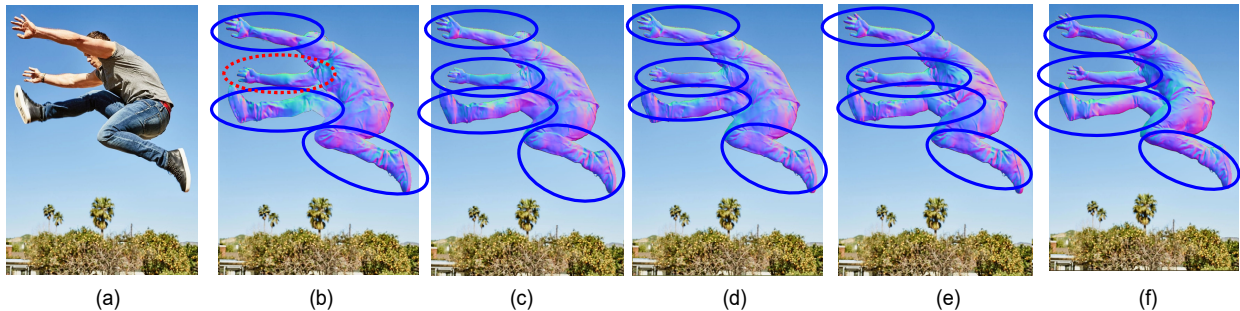


그림 10. 테스트 영상 2에 대한 ECON 기반 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
 Fig. 10. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 2 (a) input test image 2 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

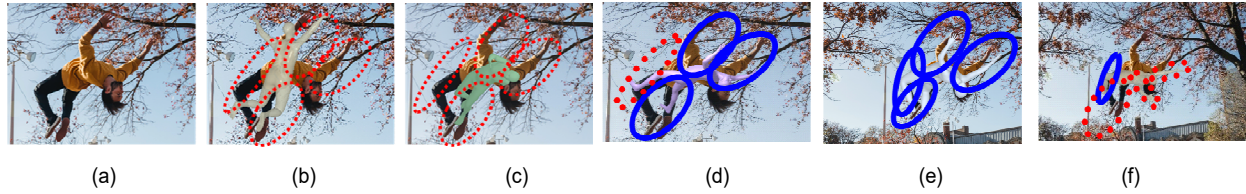


그림 11. 테스트 영상 3에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 11. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 3 (a) Input test image 3 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

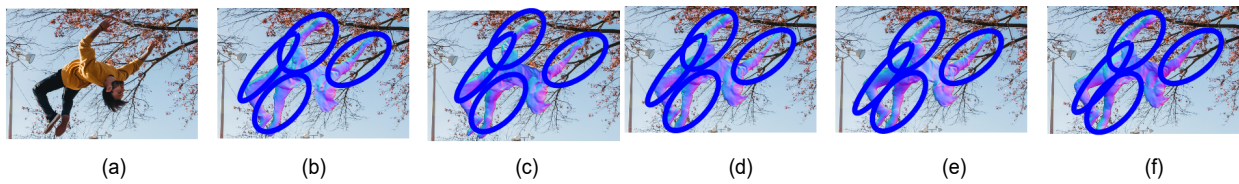


그림 12. 테스트 영상 3에 대한 ECON 기반 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 12. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 3

(a) input test image 3 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

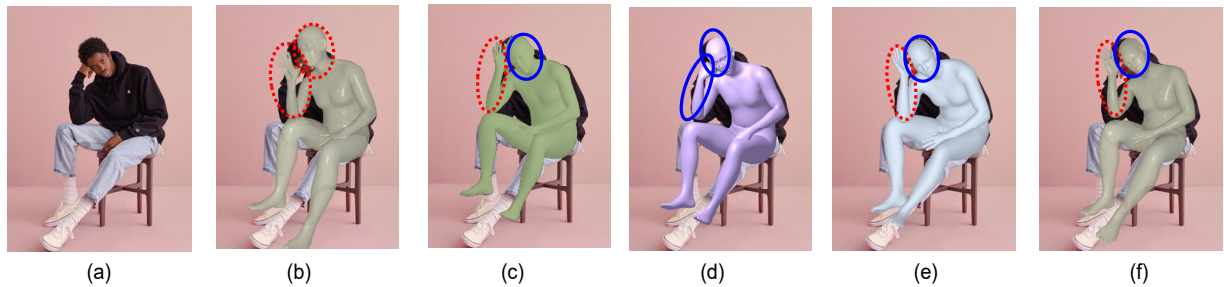


그림 13. 테스트 영상 4에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 13. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 4 (a) Input test image 4 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

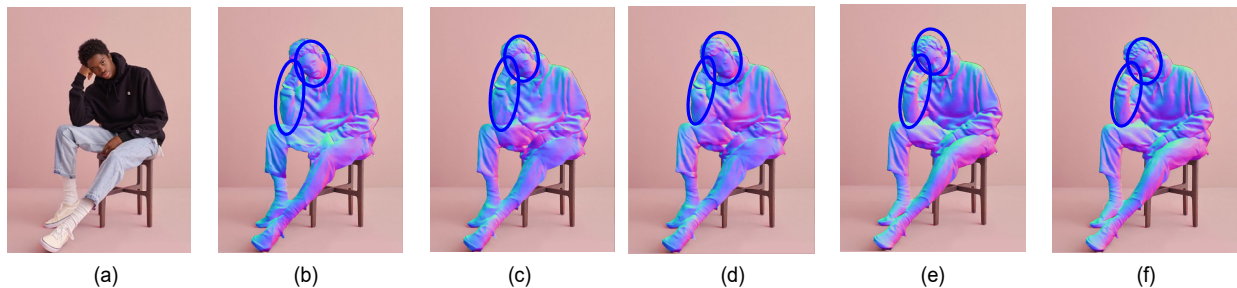


그림 14. 테스트 영상 4에 대한 ECON 기반 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 14. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 4 (a) input test image 4 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

성능을 보이고 있으며, PyMAF-x 기법이 다음으로 정확한 결과를 보이고 있다. 그리고, PKCN 기법의 경우 왼쪽 다리 부분에서만 좋은 결과를 보였고, 나머지 부분에서는 부정확한 결과를 보였다. 다음으로 ROMP와 PARE 기법의 경우에는 복원 결과가 모두 신체 방향이나 신체 부위 모양 측면에서 실제 모습과 큰 차이를 보여 주었다. 한편, 그림 12에 나타난 세부적 의복 착용 휴먼 모델 생성 결과를 보면 파라미터 기반 휴먼 모델 생성 결과에서는 좋지 않았던 ROMP, PARE 기법도 정확도가 크게 개선되어 PyMAF-x, 4D Humans, PKCN 기법, ROMP, PARE 기법 등이 거의 유사한 정확한 복원 결과를 보여주었다.

다음으로 그림 13에 나타난 테스트 영상 4에 대한 복원 결과를 비교해 보면, 대부분 결과에서 양쪽 다리와 무릎 방향과 자세는 정확하게 복원하고 있지만, 얼굴의 방향과 얼굴을 받치고 있는 오른손 모양 부분에서 복원 기법별로 차이를 보여주고 있다. 정확도 측면에서 PyMAF-x 기법이 얼굴의 방향과 얼굴을 받치고 있는 오른손 모양 부분에서 가장 좋은 결과를 보였고, PARE, 4D Humans, PKCN 기법의 경우 얼굴 방향은 비교적 정확한 결과를 보이지만, 오른손과 팔의 각도 부분에서 오차가 크게 나타나고 있다. 한편, ROMP 기법은 다리 모양을 제외하고, 나머지 부분에서는 부정확한 결과를 보였다.

다음으로 그림 14에 나타난 결과를 보면, 그림 13 결과의 정확도에 크게 영향을 받지 않고 5가지 SOTA 기법 적용 결과가 얼굴의 방향과 오른팔의 각도 측면에서 거의 유사한 결과를 보이고 있으며, 다만 기법별로 옷 주름의 모양이 약간씩 차이를 나타내고 있다.

다음 실험 결과로서 테스트 영상 5에 대한 3차원 휴먼 모델 복원 결과가 그림 15와 16에 나타나 있다. 이 영상의 경우에는 얼굴의 방향과 오른팔과 손 모양의 정확도 측면에서 각 기법의 정확도를 평가하였다. 먼저 파라미터 기반 3차원 휴먼 모델 복원 결과를 보면, 다른 테스트 영상 결과와 유사하게 PyMAF-x 기법이 얼굴의 각도, 손과 팔의 각도 측면에서 가장 좋은 결과를 보였으며, PARE, 4D Humans, PKCN 기법의 경우 손과 팔의 각도 부분에서는 오차가 있지만, 얼굴 방향은 비교적 정확하게 복원됨을 볼 수 있다. 한편 ROMP 기법은 다리 모양을 제외하고 나머지 부분은 부정확한 결과

를 보였다. 한편, 그림 16의 세부적 의복 착용 3차원 휴먼 모델 생성 결과를 보면, 테스트 영상 4의 복원 결과와 유사하게 파라미터 기반 모델의 정확도에 크게 영향을 받지 않고, 팔의 모양과 얼굴 방향이 거의 유사한 결과를 보였으며, 다만 옷 주름의 모양이 약간씩 차이를 보였다.

다음으로, 그림 17와 18의 테스트 영상 6에 대한 비교실험 결과를 보면, 이 결과에서도 PyMAF-x 기법이 얼굴의 각도, 손과 팔의 각도, 다리의 방향 및 각도 측면에서 가장 좋은 결과를 보였고, 4D Humans 기법의 경우에도 왼손 방향을 제외하고 양쪽 다리와 얼굴 방향에서 모두 가장 좋은 결과를 보였다. 또한, PARE, PKCN 기법의 경우 손의 방향과 오른쪽 다리 각도와 방향 부분에서 오차가 있었고, 얼굴 방향과 왼쪽 다리는 비교적 정확하게 복원됨을 볼 수 있다. 한편 ROMP 기법은 다리와 얼굴 방향, 손 모양 등에서 가장 부정확한 결과를 보였다.

마지막으로 테스트 영상 7의 비교실험 결과가 그림 19와 20에 나타나 있다. 그림 19의 파라미터 기반 3차원 휴먼 모델 생성 결과를 보면, 얼굴 방향과 손과 팔, 다리 모양 측면에서 PyMAF-x, PARE, 4D Humans 기법이 좋은 결과를 보인 신체 부위는 다르지만, 많은 부분에서 좋은 결과를 보였다. 한편, PKCN 기법의 경우 왼쪽 다리 방향을 제외하고 상대적으로 부정확한 결과를 보였다. 마지막으로 ROMP 기법은 다리와 얼굴 방향, 손 모양 등에서 가장 부정확한 결과를 보였다.

다음으로 그림 20의 세부적 의복 착용 3차원 휴먼 모델 생성 결과를 보면, 대부분의 비교 실험한 결과에서 파라미터 기반 모델의 정확도에 크게 영향을 받지 않고, 손과 팔의 모양과 얼굴 방향이 대체로 잘 복원되는 결과를 보였는데, 다만 옷 주름 등 디테일에서 약간씩 차이를 보였다. 또한 왼손 주먹의 세부 모양이 약간씩 차이를 보였다.

이상의 결과에서 비교 대상으로 선정된 파라미터 기반 3차원 휴먼 모델 생성 기법 5종의 기법 중에서, 대부분의 테스트 영상에서 가장 좋은 결과를 보인 기법은 PyMAF-x 기법이었으며, 4D Humans, PKCN, PARE 기법들이 PyMAF-x 기법에 이어 거의 비슷한 정확도를 보였다. 한편, ROMP 기법을 적용하여 생성한 결과가 상대적으로 가장 부정확한 결과를 보였다.

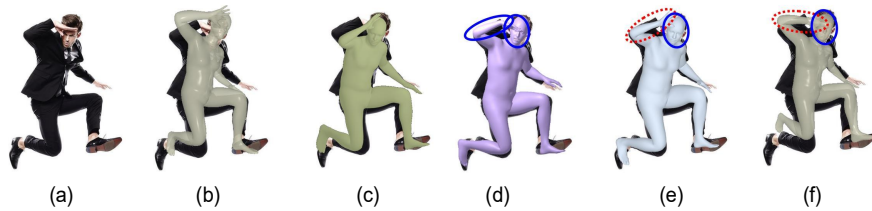


그림 15. 테스트 영상 5에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 15. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 5 (a) Input test image 5 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

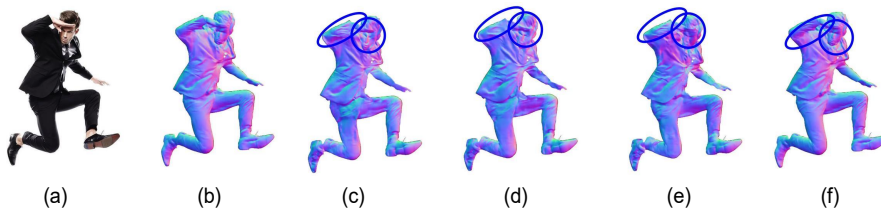


그림 16. 테스트 영상 5에 대한 ECON 기반 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
 Fig. 16. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 5 (a) input test image 5 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

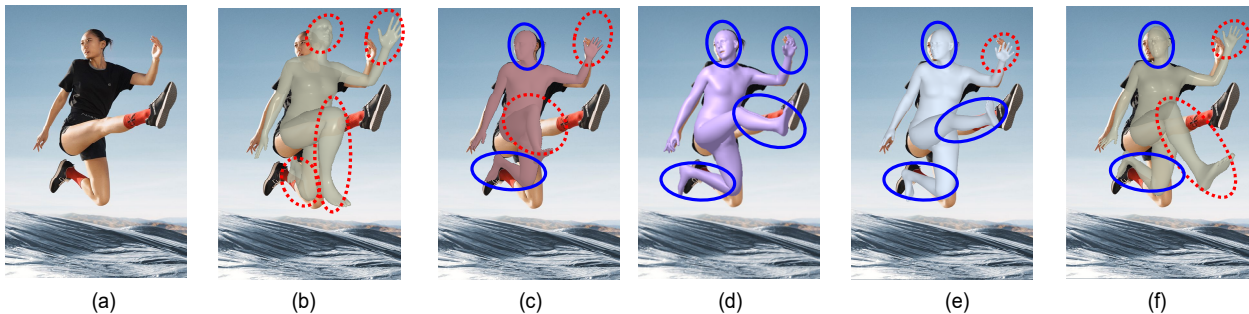


그림 17. 테스트 영상 6에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 17. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 6 (a) Input test image 6 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

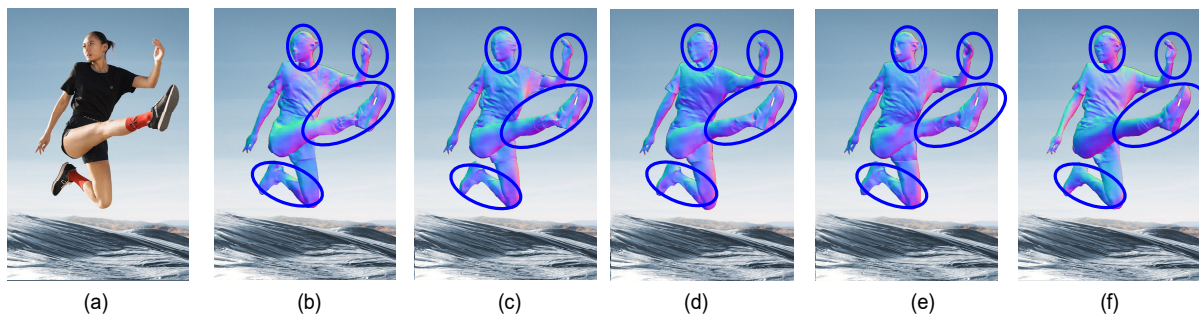


그림 18. 테스트 영상 6에 대한 ECON 기반 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
 Fig. 18. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 6 (a) input test image 6 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

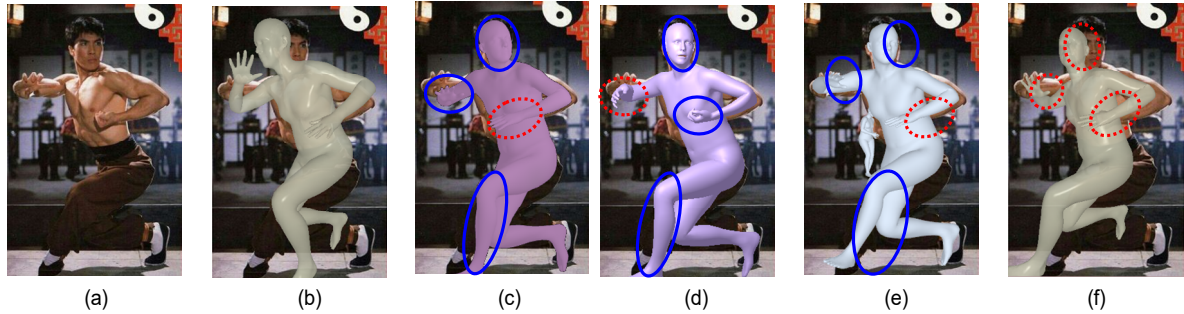


그림 19. 테스트 영상 7에 대한 파라미터 기반 3차원 휴먼 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

Fig. 19. Comparison of the Parametric 3D Human Model Generation Results of the five SOTA techniques for the test image 7 (a) Input test image 7 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

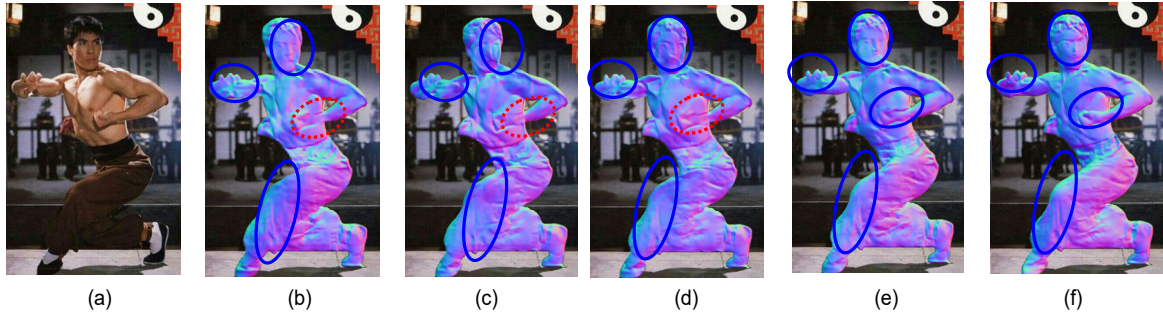


그림 20. 테스트 영상 7에 대한 세부적 의복 착용 모델 생성 결과 비교 (a) 입력 영상 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN
 Fig. 20. Comparison of the detailed Clothed Human Model Generation Results using the Parameter-based Human Model of each technique for test image 7 (a) input test image 7 (b) ROMP (c) PARE (d) PyMAF-x (e) 4D Humans (f) PKCN

다음으로 파라미터 기반 3차원 휴먼 모델을 입력하여 세부적 의복 착용 3차원 휴먼 모델을 생성한 결과의 경우에는 파라미터 기반 3차원 휴먼 모델의 정확도가 상대적으로 약간 떨어지더라도 최종 세부적 의복 착용 3차원 휴먼 모델 생성 결과에서는 시각적 또는 정성적으로 대부분의 테스트 영상에서 향상됨을 많이 볼 수 있었으며, SOTA 기법 간 정확도 차이도 크지 않았다. 따라서, 정성적으로는 파라미터 기반 3차원 휴먼 모델의 정확도 순서에 의한 영향이 크지 않음을 알 수 있다.

한편, 표 1은 5종의 SOTA 기법을 3DPW 데이터 셋에 대해 적용하여 생성한 파라미터 기반 3차원 휴먼 모델 생성 결과의 정량적 정확도 수치를 비교한 결과이다. 그리고, 표 2는 3DPW 데이터 셋이 제공하는 세부적 의복 착용 모델이 없는 관계로 대안으로 세부적 의복 모델을 제공하는 CAPE 데이터 셋 중에서 그림 21에 나타난 5종의 단일 휴먼 영상에 대해 5종의 SOTA 기법에 의해 파라미터 기반 3차원 휴

표 1. 3DPW 데이터 셋을 이용한 파라미터 기반 3차원 휴먼 모델 생성 SOTA 기법들 간의 정량적 성능 비교 결과

Table 1. Quantitative evaluation results of parameter-based 3D human model generation results for 3DPW datasets among SOTA techniques

SOTA Technique Name	MPJPE	PA-MPJPE
ROMP	85.5	53.3
PARE	74.5	46.5
PKCN	74.5	44.3
PyMAF-x	74.2	45.3
4D Humans HMR2.0	70.0	44.5



그림 21. 세부적 의복 착용 모델 생성 결과의 정량적 평가에 사용된 CAPE 데이터 셋의 5종 휴먼 영상^[21]

Fig. 21. Five Human images of the CAPE data set used for the quantitative evaluation of the detailed Clothed Human Model generation results^[21]

표 2. CAPE 데이터 셋^[21] (그림 21 참조)에 대해 5종의 SOTA 기법으로부터 생성된 파라미터 기반 3차원 휴먼 모델을 사용한 세부 의복 착용 모델 생성 결과에 대한 평균 정량적 오차 비교 결과

Table 2. Comparison results of mean quantitative errors on the CAPE data set^[21] (refer Fig. 21) measured by creating a detailed clothed model using a parameter-based 3D human model generated from five SOTA techniques

SOTA Technique Name	Chamfer	P2M
ROMP	17.761	19.655
PARE	17.050	18.740
PKCN	18.699	20.137
PyMAF-x	19.022	20.765
4D Humans HMR2.0	16.313	17.232

면 모델을 생성한 후, 생성된 모델들을 각각 ECON 신경망으로 입력하여 세부적 의복 착용 모델을 생성하고, 세부적 의복 착용 모델의 정량적 평가에 많이 사용되는 척도인 Chamfer distance와 P2M distance를 측정하여 비교하였다. 그 결과가 표 2에 나타나 있다.

먼저 표 1의 결과를 보면 4D Humans 기법이 MPJPE (Mean Per Joint Position Error) 값 측면에서 가장 오차가 적은 좋은 결과를 보였으며, PyMAF-x, PKCN, PARE 기법이 다음으로 거의 유사한 오차를 보였다. 마지막으로 ROMP 기법이 가장 부정확한 결과를 보였다. 한편, PA-MPJPE 값 측면에서는 PKCN과 4D Humans 기법이 가장 좋은 결과를 보였고, PARE, PyMAF-x, ROMP 기법 순으로 오차가 적은 결과를 보였다.

다음으로, 표 2의 세부적 의복 착용 모델 생성 결과를 보면, Chamfer distance 측면에서는 4D Humans 기법에 생성한 파라미터 기반 모델을 이용한 결과가 오차가 가장 적은 결과를 보였으며, PARE, ROMP, PKCN, PyMAF-x 순으로 오차가 적게 나타났다. 한편 P2M distance 측면에서도 동일한 정확도 순서를 보였다.

따라서, 정량적 수치 측면에서 파라미터 기반 3차원 휴먼 모델의 정확도가 세부적 의복 착용 모델 생성 결과에 미치는 영향은 기법 간 정확도 차이가 큰 경우 일부 존재한다고 할 수 있으나, 그 정확도가 차이가 크지 않은 경우, 세부적 의복 착용 모델 생성 기법에 의해 정확도가 개선되어 파라미터 기반 3차원 휴먼 모델의 정확도 영향이 상대적으로 적었다고 볼 수 있다.

IV. 결론 및 추후 과제

본 논문에서는 대표적인 3차원 휴먼 모델 생성 기법 중 파라미터 기반 휴먼 모델 생성 기법과 내재함수 기반 휴먼 모델 생성 융합 기법 중의 하나인 ECON 기법을 기반으로 사전 정보로 사용된 파라미터 기반 3차원 휴먼 모델(SMPL 메쉬 모델)을 5종의 SOTA 기법들로부터 생성한 후 이들의 정확도가 최종 세부적 의복 착용 3차원 휴먼 모델의 정확도에 미치는 영향에 대해 비교실험을 통해 알아보았다.

총 8종의 다양한 테스트 영상에 대한 비교실험 수행 결과, 먼저 정성적으로 파라미터 기반 3차원 휴먼 모델 생성시 대부분 영상에서 가장 좋은 성능을 보인 기법은 PyMAF-x 기법으로 생성한 결과였으며, 대체로 PKCN, PARE 기법의 결과가 거의 비슷한 정확도를 보였다. 또한 상대적으로 ROMP 기법의 결과가 가장 좋지 않은 성능을 보여주었다. 이러한 파라미터 기반 휴먼 모델들을 세부적 의복 착용 모델 생성 기법인 ECON 기법의 입력으로 사용하여 세부적 의복 착용 3차원 휴먼 모델을 생성한 결과에서는 ROMP 기법이나 PARE 기법 결과에서 정확도가 파라미터 기반 모델 생성 결과보다 정확도가 향상됨을 많이 볼 수 있었으며, 나머지 기법 결과와 거의 유사한 결과를 보여 기법 간의 정확도 차이가 크게 나타나지 않았고, 옷 주름 등 일부 미세한 차이 정도만 나타났다. 그리고, 파라미터 기반 휴먼 모델의 정량적 정확도 측정 결과에서는 4D Humans 기법이 가장 좋은 결과를 보였고, PyMAF-x, PKCN, PARE 기법이 거의 유사한 결과를 보였고, ROMP 기법의 정확도가 가장 낮아 정성적 결과와 유사한 결과를 보였다. 한편, 세부적 의복 모델 생성 결과의 정량적 비교에서 4D Humans 기법이 가장 높은 정확도를 보인 결과는 파라미터 모델 생성 결과와 유사한 경향을 보였으나, ROMP 기법이나 PARE 기법의 정확도가 많이 향상되어 나머지 기법들과 거의 유사한 정확도 보였다. 따라서, 파라미터 기반 모델의 정확도를 ECON 기법에 의해 생성된 최종 세부적 의복 착용 3차원 휴먼 모델의 정확도가 반드시 따라가진 않음을 알 수 있었다.

추후 과제로는 세부적 의복 착용 3차원 휴먼 모델 생성시에 파라미터 기반 모델의 정보를 좀 더 효율적으로 융합하는 방법에 관한 연구가 필요하다.

참 고 문 헌 (References)

- [1] 7080 Old Pop Legend, ABBA also metaverse!, <https://brunch.co.kr/@poppol12/14>, (accessed Jan. 13, 2024).
- [2] SPRi, *Newcommers in Mataverse, Digital Human*, Issue Report IS-135, 2022, https://spri.kr/posts/view/23372?code=data_all&study_type=&board_type=issue_reports, (accessed Jan. 13, 2024).
- [3] M. Park, J. Kang, and J. Yeon, "Human modeling technology trends for metaverse services," *Special issues on 3D object and space modeling and visualization technology, Broadcast and Media Magazine*, vol. 26, no.4, pp. 61-71, Oct. 2021. <https://www.kibme.org/resources/journal/20220617111442058.pdf>
- [4] Intelrealsense, <https://www.intelrealsense.com/>, (accessed Jan. 13, 2024).
- [5] Azure Kinect DK, <https://azure.microsoft.com/ko-kr/services/kinect-dk/>, (accessed Jan. 13, 2024).
- [6] Z. Zheng, T. Yu, Y. Liu, and Q. Dai, "Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 3170 - 3184, Jun. 2022. doi: <https://doi.org/10.1109/TPAMI.2021.3050505>
- [7] W. Nicolaa, M. Yvana, C. Frederic, "A survey of 3D human body reconstruction from single and multiple camera views," *Computers & Graphics*, Jan. 2023. <http://dx.doi.org/10.2139/ssrn.4335838>, Sept. 2023.
- [8] H. A. Correia and J. H. Brito, "3D reconstruction of human bodies from single-view and multi-view images: A systematic review," *Computer Methods and Programs in Biomedicine*, vol. 239, no. 107620, Sep. 2023. doi: <https://doi.org/10.1016/j.cmpb.2023.107620>
- [9] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," *ACM Transactions on Graphics*, vol. 34, no. 6, pp. 248:1 - 248:16, Oct. 2015. doi: <https://doi.org/10.1145/2816795.2818013>
- [10] Y. Sun, Q. Bao, W. Liu, Y. Fu, M. J. Black, and T. Mei, "Monocular, One-stage, Regression of Multiple 3D People," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11179-11188, Oct. 2021, (accessed Jan. 13, 2024). doi: <https://doi.org/10.48550/arXiv.2008.12272>
- [11] M. Kocabas, C. P. Huang, O. Hilliges, and M. J. Black, "PARE: Part Attention Regressor for 3D Human Body Estimation," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11127-11137, Oct. 2021, (accessed Jan. 13, 2024). doi: <https://doi.org/10.48550/arXiv.2104.08527>
- [12] H. Zhang, Y. Tian, Y. Zhang, M. Li, L. An, Z. Sun, and Y. Liu, "Pymaf-x: Towards well-aligned full-body model regression from monocular images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 10, pp. 12 287 - 12 303, Oct. 2023. doi: <https://doi.org/10.1109/TPAMI.2023.3271691>
- [13] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and H. Li, "PIFu: Pixel-aligned implicit function for high resolution clothed human digitization," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2304-2314, Oct. 2019. doi: <https://doi.org/10.48550/arXiv.2304.09502>
- [14] S. Saito, T. Simon, J. Saragih, and H. Joo, "Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 84 - 93, Jun. 2020. doi: <https://doi.org/10.48550/arXiv.2004.00452>
- [15] B. Bhatnagar, C. Sminchisescu, C. Theobalt, and G. Pons-Moll, "Combining Implicit Function Learning and Parametric Models for 3D Human Reconstruction," *Proceedings of European Conference on Computer Vision*, pp. 311-329, Aug. 2020. doi: <https://doi.org/10.48550/arXiv.2007.11432>
- [16] Xiu, J. Yang, D. Tzionas, and M. J. Black, "ICON: Implicit Clothed humans Obtained from Normals," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13286-13296, Jun. 2022. doi: <https://doi.org/10.48550/arXiv.2112.09127>
- [17] Y. Xiu, J. Yang, X. Cao, D. Tzionas, and M. J. Black, "ECON: Explicit Clothed Humans Optimized via Normal Integration," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 512-523, Jun. 2023. doi: <https://doi.org/10.48550/arXiv.2212.07422>
- [18] J. Kim, G.-M. Um, J. Seo, and W. Kim, "Part-attentive kinematic chain-based regressor for 3D human modeling," *Journal of Visual Communications and Image Representation*, vol. 95, pp. 103881 - 103889, Sep. 2023. doi: <https://doi.org/10.1016/j.jvcir.2023.103881>
- [19] S. Goel, G. Pavlakos, J. Rajasegaran, A. Kanazawa, and J. Malik, "Humans in 4D : Reconstructing and Tracking Humans with Transformers," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14783-14794, Oct. 2023. doi: <https://doi.org/10.48550/arXiv.2305.20091>
- [20] H. Zhang, Y. Tian, X. Zhou, W. Ouyang, Y. Liu, L. Wang, and Z. Sun, "Pymaf: 3d human pose and shape regression with pyramidal mesh alignment feedback loop," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11446 - 11456, Oct. 2021. doi: <https://doi.org/10.48550/arXiv.2103.16507>
- [21] Q. Ma, J. Yang, A. Ranjan, S. Pujades, G. Pons-Moll, S. Tang, M. J. Black, "Learning to Dress 3D People in Generative Clothing," *Proceedings of CVPR2020*, June, 2020. doi: <https://doi.org/10.48550/arXiv.1907.13615>

저 자 소 개



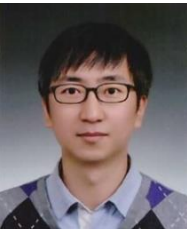
엄 기 문

- 1993년 2월 : 서강대학교 전자공학과(공학박사)
- 2000년 4월 ~ 현재 : 한국전자통신연구원 실감미디어연구실 책임연구원
- 2001년 6월 ~ 2002년 5월 : 캐나다 CRC(Communications Research Centre Canada) 방문연구원
- ORCID : <https://orcid.org/0000-0002-0929-0695>
- 주관심분야 : 컴퓨터 비전, 휴먼 3차원 모델 복원



송 찬 들

- 2023년 2월 : 건국대학교 학사
- 2023년 3월 ~ 현재 : 건국대학교 석사
- ORCID : <https://orcid.org/0009-0003-3071-7512>
- 주관심분야 : 컴퓨터 비전, 영상처리, 심층학습



김 원 준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문연구원
- 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
- ORCID : <https://orcid.org/0000-0001-5121-5931>
- 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습



이 희 경

- 2022년 1월 : 한국정보통신대학교(ICU) 공학석사
- 2002년 1월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0002-1502-561X>
- 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습, VCM, 메타버스, 360VR, 시선추적, 메타데이터



양 승 준

- 2001년 2월 : 전남대학교 전산학과 석사
- 2001년 3월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0002-8425-9620>
- 주관심분야 : 디지털방송, 컴퓨터 비전, 영상처리, 신경망학습

저 자 소 개



정 원 식

- 1992년 2월 : 경북대학교 전자공학과(공학사)
- 1994년 2월 : 경북대학교 대학원 전자공학과(공학석사)
- 2000년 2월 : 경북대학교 대학원 전자공학과(공학박사)
- 2000년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0001-5430-2969>
- 주관심분야 : 이머시브 미디어 기술, 기계를 위한 영상 부호화, 딥러닝기반 신호처리, 멀티미디어 표준화



추 현 곤

- 1998년 2월 : 한양대학교 전자공학과(공학사)
- 2000년 2월 : 한양대학교 전자공학과(공학석사)
- 2005년 2월 : 한양대학교 전자통신전파공학과(공학박사)
- 2005년 2월 ~ 현재 : 한국전자통신연구원 책임연구원
- 2015년 1월 ~ 2017년 1월 : 한국전자통신연구원 디지털홀로그래피연구실실장
- 2023년 2월 ~ 현재 : 한국전자통신연구원 실감미디어연구실실장
- 2017년 9월 ~ 2018년 8월 : Warsaw University of Technology, Poland 방문연구원
- ORCID : <https://orcid.org/0000-0002-0742-5429>
- 주관심분야 : Computer vision, 3D imaging and holography, Immersive media, 3D broadcasting system