

## 3 차원 모션을 통한 아바타 생성 기술

박소현<sup>1</sup>, 전유채<sup>2</sup>, 고재은<sup>2</sup>, 강지우<sup>3</sup>

<sup>1</sup>숙명여자대학교 IT 공학전공 학부생

<sup>2</sup>숙명여자대학교 컴퓨터과학과 학부생

<sup>3</sup>숙명여자대학교 인공지능공학부 교수

{so209, wjsdbco, rhwodms1223, jwkang}@sookmyung.ac.kr

### Avatar Generation from 3D Motion

So-Hyun Park<sup>1</sup>, U-Chae Jun<sup>2</sup>, Jae-Eun Ko<sup>2</sup>, Ji-Woo Kang<sup>3</sup>

<sup>1</sup>Department of IT Engineering, Sookmyung Women's University

<sup>2</sup>Department of Computer Science, Sookmyung Women's University

<sup>3</sup>Division of Artificial Intelligence Engineering, Sookmyung Women's University

#### 요 약

버추얼 유튜버로서 자신의 동작을 3D 가상 캐릭터로 나타내고, SNS 에서 춤을 공유하는 경우가 많아졌다. 본 논문에서는 2D 영상에서 MediaPipe BlazePose 모델로 추정된 사람 포즈를 3D 인체 모델인 SMPL 에 피팅하여 사용자 정의 3D 모델을 생성하는 방법을 제안한다. 이를 통해 자신의 춤 영상으로 3D 모델을 생성하여 공유하거나, 기존의 춤 동영상으로 3D 모델을 생성하여 댄스 게임에 사용할 수 있다. 이처럼 본 기술은 예술 및 엔터테인먼트 분야에서 다양하게 활용될 수 있다.

#### 1. 서론

버추얼 (virtual) 유튜버의 등장으로 자신의 동작을 3D 가상 캐릭터로 표현하는 경우와, SNS (Social Network Service) 플랫폼에서 댄스 챌린지의 인기로 자신의 춤을 공유하는 사람이 많아졌다. 만일 자신의 동작 또는 춤을 3D 모델에 피팅하여 공유할 수 있다면, 이 기능을 다양한 예술 및 엔터테인먼트 분야에서 활용할 수 있을 것이다.

사람의 3D 포즈 (pose)를 추정하는 여러 모델 중 MediaPipe BlazePose [1] 모델은 2D 이미지에서 3D 포즈를 실시간으로 추정할 수 있으며 보이지 않는 부분 또한 추정이 가능하다는 장점이 있다. 하지만 추정 결과가 랜드마크의 3D 좌표이고, 스켈레톤 모델이 다른 휴머노이드 모델과 달리 척추가 없는 구조이기 때문에 3D 인체 모델에 피팅하는 것이 어렵다는 문제가 있다. 따라서 본 논문에서는 BlazePose 모델의 포즈 추정 결과인 랜드마크의 3D 좌표를 SMPL (Skinned Multi-Person Linear model) [2] 포즈 파라미터로 변환하는 방법을 제안한다. 결과적으로 임의의 2D 영상에서 추정한 사람 포즈를 3D 모델에 피팅하여 사용자 정의 3D 모델을 생성할 수 있다.

이러한 기술은 댄스 게임과 같은 엔터테인먼트 에 플리케이션에 적용할 수 있을 뿐만 아니라 사용자가 자신의 동작을 가상공간에서 공유하고 상호작용할 수 있는 환경 또한 제공할 수 있다.

#### 2. SMPL

SMPL 은 6,890 개의 정점 (vertex)과 24 개의 관절 (joint)로 이루어진 3 차원 인체 선형 모델로, 파라미터  $\beta$  로 체형 (shape)을, 파라미터  $\theta$  로 포즈를 제어할 수 있다.

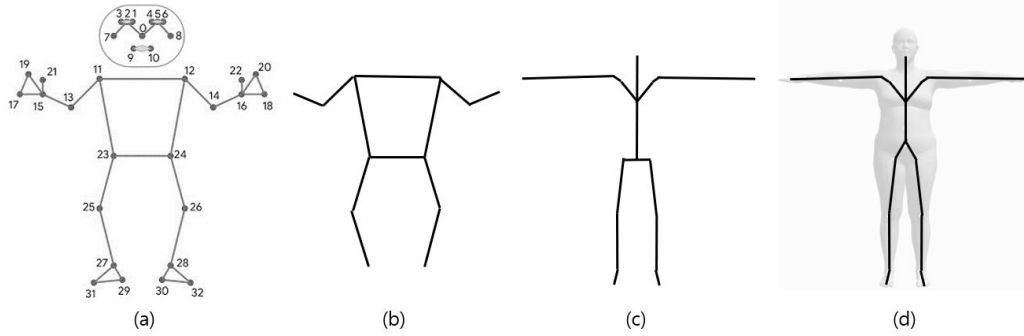
포즈는 관절을 움직인 양으로 취할 수 있고, 각 관절의 회전은 스켈레톤의 키네마틱 트리 (kinematic tree) 상하 구조에 따른 상대 각도의 축 회전 (axis angle) 표현으로 나타내며, 이를 다음과 같은 Rodrigues 공식을 통해  $3 \times 3$  회전행렬  $R$ 로 변환한다.

$$R = I + K \sin \theta + K^2(1 - \cos \theta)$$

여기서  $I \in \mathbb{R}^{3 \times 3}$  는 항등 행렬,  $\theta$  는 회전 각도,  $K \in \mathbb{R}^{3 \times 3}$  는 회전축의 방향 벡터로 이루어진 반대칭 행렬이다. 관절 위치에 회전행렬  $R$ 을 곱하면  $K$ 를 회전축으로  $\theta$ 만큼 회전한 관절 위치를 구할 수 있다.

#### 3. MediaPipe BlazePose 모델

2D 이미지로부터 3D 포즈를 추정하는 BlazePose 모델은 프레임에서 사람이 존재하는 영역을 감지하는 감지기 (detector)와 키포인트 좌표를 예측하고 사람의 위치를 추적하는 추적기 (tracker)로 두 단계의 파이프라인으로 이루어진다. 학습 과정에서 히트맵과 오픈셋 손실을 사용하고, 추론 과정에서는 트래킹 기반으로 추정하여 모델의 무게를 줄이고 속도를 높인다.



(그림 1) 초기 포즈 (rest pose)의 스켈레톤 모델. (a) 33 개의 랜드마크로 이루어진 MediaPipe BlazePose 모델, (b) BlazePose 모델에서 12 개 랜드마크를 선택한 결과, (c) 12 개 키포인트 위치를 상대적으로 계산하여 추가한 결과, (d) SMPL의 스켈레톤.

새로운 이미지가 입력되면 자동으로 이미지를 회전하고 크기를 조절하여 사람을 따라가고, 학습 데이터 이미지를 랜덤으로 가려 학습되어 보이지 않는 부분에 대해서도 추정이 가능하다. 최종적으로 BlazePose 모델의 33 개 랜드마크의 3D 좌표와 이미지에서 각 랜드마크가 보일 가능성 (visibility)을 알 수 있다.

4. 방법

BlazePose 모델로 추정된 33 개의 랜드마크 3D 좌표를 225 차원 SMPL 포즈 파라미터로 변환하는 것이 목표이다. 구하고자 하는 SMPL 포즈 파라미터는 루트 (root)를 포함한 24 개 각 관절의 9 차원 회전 행렬과 전역 위치 좌표, 6 개의 패딩 값으로 이루어진다 ( $24 \times 9 + 3 + 6 = 225$ ).

먼저 (그림 1)의 (a)에서 (c)와 같이 SMPL 스켈레톤과 유사하도록 키포인트의 3D 위치를 구한다. BlazePose 모델의 33 개 랜드마크 3D 좌표 중 (그림 1)의 (b)와 같이 얼굴, 손, 발을 제외한 12 개를 가져오고, 이를 이용해 목, 머리 등 12 개 키포인트의 위치를 상대적으로 계산하여 추가한다. 총 24 개의 키포인트를 SMPL 스켈레톤의 키네마틱 트리와 동일한 계층 구조를 갖도록 하고, 포즈를 변경하지 않은 초기 포즈 (rest pose)를 정의한다.

다음으로 키포인트의 위치로부터 키포인트의 회전 축과 각도를 구한다. 각 키포인트의 상위 키포인트로부터의 상대적인 회전을 누적하여 총 회전 행렬과 오일러 각을 계산한다. 이를 쿼터니언 회전으로 변환하고, 축 회전 표현으로 나타낸다.

마지막으로 SMPL 과 동일하게 Rodrigues 공식을 사용하여  $3 \times 3$  회전행렬로 변환하고, 루트 키포인트 위치를 전역 위치 좌표로 사용하여 225 차원 행렬을 얻는다.



(그림 2) 오일러 각으로 변환한 키포인트 회전

5. 결과

BlazePose 랜드마크에서 24 개의 키포인트 위치를 계산하고, 각 키포인트의 회전을 ZXY 오일러 각으로 변환한 결과는 (그림 2)와 같다. 쿼터니언 회전, 축 회전, Rodrigues 공식을 통해 SMPL 포즈 파라미터로 변환 가능하며, 이를 SMPL 에 피팅할 수 있다.

6. 결론

본 논문에서는 BlazePose 모델을 통해 추정된 포즈를 SMPL 에 피팅하였다. 피팅 결과는 유니티와 같은 게임 엔진을 통해 다른 휴머노이드 모델에 쉽게 피팅할 수 있다. BlazePose 모델로 추정되는 동시에 SMPL 포즈 파라미터로 변환되도록 하면 실시간 아바타 생성 또한 가능하다. 이는 추후에 구현하고, 유사한 기술인 FrankMocap [3]과 비교하고자 한다.

본 논문에서 제안한 방법을 통해 사용자가 자신의 춤을 촬영하여 3D 모델을 생성하여 공유하거나, 기존의 춤 영상으로 3D 모델을 생성하여 댄스 게임에 사용하는 등 가상 공간에서 창의적인 활용이 가능하다.

ACKNOWLEDGMENT

이 논문은 2023 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2023-00229451, 이종 플랫폼간 상호호환이 가능한 디지털휴먼(아바타) 연동 기술)

참고문헌

[1] Bazarevsky, Valentin, et al. Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204, 2020.  
 [2] Loper, Matthew, et al. "SMPL: A Skinned Multi-person Linear Model." ACM Transactions on Graphics. 2015. p. 34.6: 1-16.  
 [3] Rong, Yu, et al. "Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 1749-1759.