2D 키포인트의 3차원 변환과 코사인 기반 기하 분석을 통한 인체 포즈 유사도 추정 및 비교

손명규, 이상헌, 김현덕, 김준광

대구경북과학기술원

smk@dgist.ac.kr, pobbylee@dgist.ac.kr, hyunduk00@dgist.ac.kr, kjk1208@dgist.ac.kr

3D Human Pose Similarity Estimation via Lifting from 2D Keypoints and Cosine-Based Geometric Analysis

Myoung-Kyu Sohn, Sang-Hoen Lee, Hyunduk Kim, Junkwang Kim

Daegu Gyeongbuk Institute of Science & Technology (DGIST)

요 약

본 연구는 2D 인체 키포인트를 3차원 공간으로 변환(Lifting)하여 인체 포즈 간의 유사도를 정량적으로 추정하는 방법을 제안한다. 기존 2D 기반 자세 비교는 시점 변화나 신체 크기, 위치 차이에 민감하다는 한계를 지니고 있으나, 제안한 방법은 3D 공간에서의 기하학적 관계를 반영함으로써 이러한 문제를 보완하였다. 2D 이미지로부터 추출된 키포인트를 딥러닝 기반 3D Lifting 모델을 통해 3차원 좌표로 변환하고, 위치·크기·시점에 대해 각각 이동 불변성, 신체 크기 불변성, 시점 불변성 정규화를 수행하여 비교 가능한 형태로 정규화하였다. 이후 정규화된 포즈 벡터 간의 방향적 일치도를 측정하기 위해 코사인 유사도(Cosine Similarity)를 적용하였다. 실험 결과, 2D 키포인트 기반 평균 유사도 0.7855에 비해 3D 기반 접근에서는 0.8901로 차이를 보였다. 본 연구는 인체 포즈 유사도 평가의 정확성을 높였으며, 향후에는 동영상 데이터에 적용하여 시간적 유사성 분석으로 확장할 예정이다.

Ⅰ. 서론

최근 비대면 환경의 확산과 함께 홈트레이닝, 재활 치료, 스포츠 코칭 등 다양한 분야에서 인체의 동작을 인식하고 분석하는 기술의 수요가 급격히 증가하고 있다. 특히 카메라를 이용한 인체 자세 분석 기술은 비용 효율성과 접근성 측면에서 주목받고 있으며, 기존의 3차원 센서 기반 시스템을 대체하거나 보완하는 방향으로 발전하고 있다. 이러한 기술의 핵심은 인체의 자세를 정량적으로 비교하고 유사도를 평가하는 것으로, 이는 동작 교정, 퍼포먼스 측정, 사용자 피드백 등에 직결된다. 기존 연구들은 주로 2차원 이미지 기반의 키포인트 추출 기술을 활용하여 인체의 자세를 비교해 왔으며, 코사인 유사도, 동적 시간 왜곡(Dynamic Time Warping), 가중치 기반 매칭 등 다양한 방법이 제안되었다. 그러나 2차원 공간에서의 비교는 카메라 시점(viewpoint) 변화나 인체의 크기, 위치에 민감하다는 한계가 존재한다. 이를 극복하기 위해 최근에는 2D 키포인트를 3차원 공간으로 변환하는 'Lifting' 기법을 적용하여 보다 구조적이고 의미 있는 유사도 분석을 시도하는 연구가 주목받고 있다[1,2].

본 연구에서는 2D 인체 키포인트를 3차원으로 lifting한 후, 공간상에서 코사인 유사도를 기반으로 한 기하학적 분석을 수행함으로써 인체 포즈 간의 유사도를 정량적으로 추정하는 방법을 제안한다. 이를 위해 먼저 2D 이미지로부터 2D 키포인트를 추출하고, 추출된 2D 키포인트로부터 3D 키포인트를 추정한다. 이후 이를 일련의 전처리 과정을 통해 정규화한다. 정규화된 포즈 벡터 간의 방향성과 구조적 유사성을 분석하기 위해 코사인 유사도를 적용한다. 본 방식은 기존 2D 기반 접근 방식보다 자세 간의 정밀한 차이를 더 잘 포착할 수 있으며, 다양한 시점 및 인체 조건에서도 높은 일관성을 유지할 수 있는 장점이 있다.

Ⅱ. 본론

인체 자세의 유사도를 정확하게 추정하기 위해서는 관절의 상대적 위치와 공간적 구조를 정밀하게 반영할 수 있는 3차원 표현이 필요하다. 그러나 일반적인 영상에서는 깊이 정보가 존재하지 않는다. 2차원 이미지에서 3차원 인체 키포인트를 추정하는 방법은 이미지에서 직접 3D 키포인트를 추정하는 방법이 있고, 최근에는 2차원 이미지에서 추출한 인체 키포인트를 3차원 공간으로 변환하는 lifting 방법이 좀 더 많이 쓰이고 있다[3]. 본 연구에서는 비전 트랜스포머(vision transformer)와 같은 딥러닝 기반의 포즈 추정 모델을 통해 2D 키포인트 (x,y) 좌표를 추출한 뒤, 3D lifting 딥러닝 모델을 통하여 2D 키포인트로부터 3D 좌표 (x,y,z)를 추정한다. 이 lifting 과정은 트랜스포머 기반 네트워크를 활용하여, 관절 간의 관계를 학습함으로써 2D 투영으로 손실된 깊이(z) 정보를 추정하여 최종 3D 좌표를 추출하는 것이다[3,4,5].

3D로 변환된 포즈 데이터는 포즈 간 비교의 일관성과 정밀도를 확보하기 위해 전처리(preprocessing) 단계를 거친다. 먼저, 인체의 위치 차이에 따른 영향을 제거하기 위해 이동 불변성(translation invariant) 처리를 수행한다. 이 단계에서는 인체의 중심점을 골반으로 기준점(origin) 설정한 뒤, 모든 키포인트 좌표에서 해당 중심점의 좌표를 빼는 방식으로 평행이동을 수행한다. 이를 통해 인체가 화면의 어느 위치에 있더라도 모든 포즈가 동일한 원점 기준에서 비교될 수 있다. 다음으로, 사람마다 신체 크기나 신장에 차이가 존재하므로 신체 크기 불변성(body size invariant) 정규화를 적용한다. 이를 위해 인체의 각 조인트 사이의 거리를 측정하고, 해당 값을 평균 크기로 정규화하여 모든 키포인트 좌표를 스케일링한다. 이렇게 하면 절대적인 신체 크기 차이는 제거되고, 관절 간의 상대적 비율과 형태 정보만 남게 되어 인체 구조의 일관성이 유지된다. 마지막으로,

시점 불변성(viewpoint invariant) 정규화를 수행하여 관측 시점에 따른 자세 왜곡을 줄인다. 이 과정에서는 신체의 주요 축(왼쪽 골반과 오른쪽 골반)을 기준으로 포즈를 회전시켜, 모든 포즈가 동일한 방향에 맞춰지도록 정렬한다. 이를 위해 3D 좌표계 상에서 회전 행렬(Rotation Matrix)을 계산하고, 각 키포인트에 이를 적용함으로써 시점 변화에 대한 영향을 최소화한다. 이러한 일련의 정규화 과정을 거치면, 각 포즈는 위치·크기·시점의 차이로부터 독립적인 상태로 표현되어, 이후의 유사도 계산에서 순수한 자세 형태의 차이만을 비교할 수 있게 된다.

전처리가 완료된 포즈는 각 키포인트를 연결한 벡터 형태로 표현되며, 이를 이용해 포즈 간 유사도를 계산한다. 본 연구에서는 코사인 유사도(Cosine Similarity) 를 사용하여 두 포즈 벡터의 방향적 일치도를 측정한다. 코사인 유사도는 두 벡터가 이루는 각도의 코사인을 계산함으로써, 자세의 절대적 크기보다는 형태적 유사성에 초점을 맞춘다. 수학적으로, 시범 포즈 P와 비교 포즈 Q 간의 유사도는 다음과 같이 정의된다:

$$S(P,Q) = \frac{P \cdot Q}{\| P \| \ \| Q \|}$$

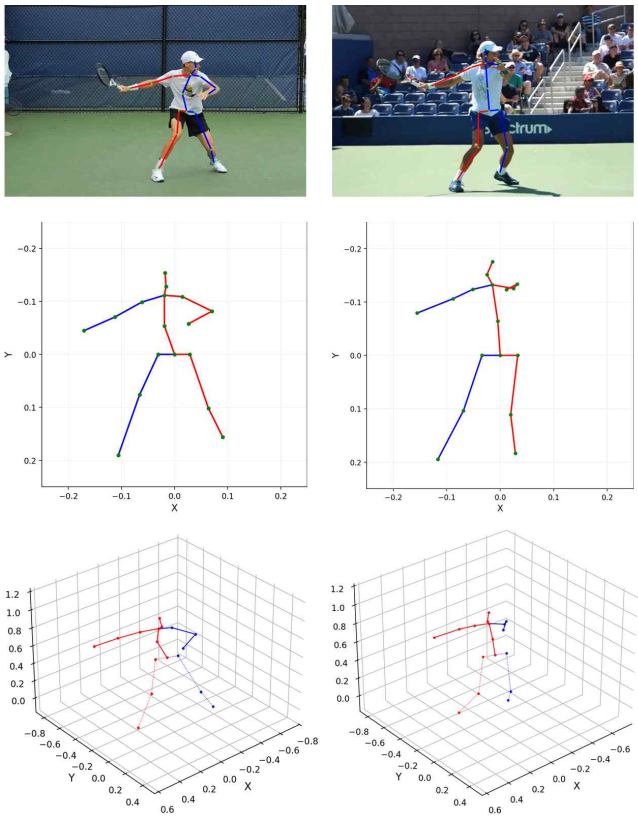

그림 1. 유사도 분석을 위한 키포인트 추정 (첫 번째 컬럼: 2d 입력이미지 + 추정된 2d 키포인트, 두 번째 컬럼: 정규화 된 2d 키포인트, 세 번째 컬럼: 2d 키포인트로부터 추정된 3d 키포인트의 정규화 된 키포인트)

유사도 측정을 위해서는 2D 키포인트 및 3D 키포인트 모두 각각 정규화된 키포인트를 사용하였다. 표1은 그림 1에서 보여지는 두 개 이미지의 유사도 측정 실험 결과를 보여준다. 실험 결과는 제안한 3차원 포즈 기반 유사도 측정 방법이 기존 2차원 키포인트 기반 접근과 서로 다른 유사도 값을 보여주고 있다. 2D 키포인트를 직접 이용하여 코사인 유사도를 계산한 경우 평균 유사도(Average similarity)는 0.7855로 나타났으며, 동일한 데이터에서 3D로 lifting한 후 동일한 방식으로 유사도를 계산한 결과 0.8901로 측정되었다. 이는 3차원 공간에서 포즈를 비교할 때 인체의 깊이 정보가 반영되어, 시점(viewpoint) 변화나 관절의 부분적 가림(occlusion)에 덜 민감하게 작용했기 때문으로 분석된다. 또한 3D 포즈 정규화 과정을 통해 크기와 위치 차이가 제거되어, 자세의 구조적 형태에 기반한 순수한 유사도 평가가 가능해졌다. 이러한 결과는 2D 기반 분석보다 3D 공간에서의 포즈 비교가 인체 동작의 실제적 유사성을 보다 정밀하게 반영할 수 있음을 보여주며, 제안한 방법의 유효성을 실험적으로 입증한다.

표 1. 평균 유사도 추정 결과

	2D keypoints	3D keypoints
Average similarity	0.7855	0.8901

Ⅲ. 결론

본 연구에서는 2D 인체 키포인트를 3차원 공간으로 lifting한 후, 코사인 기반 기하 분석을 통해 인체 포즈 간 유사도를 정량적으로 추정하는 방법을 제안하였다. 제안한 방법은 2D 기반 접근보다 깊이 정보를 반영함으로써 시점 변화나 가림 현상에 강인하며, 포즈의 구조적 형태를 보다 정밀하게 반영할 수 있음을 실험을 통해 확인하였다. 특히 평균 유사도가 각각 0.7855와 0.8901인 결과는 3D 공간에서의 비교와 2D 공간에서의 비교가 차이가 난다는 점을 분명히 한다. 향후 연구에서는 본 연구에서 제안한 정규화 및 유사도 측정 방법을 연속적인 동영상 데이터에 적용하여 시간에 따른 자세 변화나 동작의 유사성을 분석하고, 실시간 자세 교정 및 행동 인식과 같은 응용으로 확장하는 연구를 진행하고자 한다.

ACKNOWLEDGMENT

본 연구는 2025년 과학기술정보통신부의 재원으로 DGIST 기관고유사업(25-IT-01)과 연구개발특구진흥재단(과제번호 2710033150)의 지원을 받아 수행되었습니다.

참 고 문 헌

[1] Badiola-Bengoa, Aritz, and Amaia Mendez-Zorrilla. "A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise." Sensors 21.18 (2021): 5996.

[2] Guo, Yan, et al. "A Survey of the State of the Art in Monocular 3D Human Pose Estimation: Methods, Benchmarks, and Challenges." Sensors (Basel, Switzerland) 25.8 (2025): 2409.

[3] Liu, Yang, Changzhen Qiu, and Zhiyong Zhang. "Deep learning for 3D human pose estimation and mesh recovery: A survey." Neurocomputing 596 (2024): 128049.

[4] Xu, Yufei, et al. "ViTPose: Simple vision transformer baselines for human pose estimation." Advances in neural information processing systems 35 (2022): 38571-38584.

[5] Mehraban, Soroush, Vida Adeli, and Babak Taati. "MotionAGFormer: Enhancing 3d human pose estimation with a transformer-gcnformer network." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024.