

단안 영상 기반 3차원 인체 자세 추정 신뢰도 측정 및 평가 방법

하재현¹, 김휘강^{2*}, 이상범², 김광용²

¹ 경북대학교, ² 한국전자통신연구원

mini0950@knu.ac.kr, {hwigangkim*, sblee230, kimky}@etri.re.kr

Monocular Image-based 3D Human Pose Estimation Reliability Measurement and Evaluation Method

Jaehyeon Ha¹, Hwi-Gang Kim^{2*}, Sangbeom Lee², Kwang-Yong Kim²

¹ School of Electronics Engineering, Kyungpook National University,

² Electronics and Telecommunications Research Institute

요약

본 논문은 단일 시점 영상으로부터 추출한 3차원 인체 자세 추정 기술의 신뢰도를 측정하기 위한 방법과 관련한 것으로, 하나의 카메라로 촬영한 영상 내 사용자의 3차원 자세를 추정할 때 인체 관절 사이의 상대 거리 및 특정 움직임에 따른 각도 등을 활용하여 3차원 실측 자료 없이 자세 추정 결과에 대한 신뢰도를 측정하고 평가할 수 있는 간단하고 효과적인 방법을 소개한다. 이와 관련하여 본 논문에서는 자세 추정 결과의 신뢰도 측정을 위해 일관성이란 이름의 지표를 제시하며, 이는 실측 자료를 포함한 학습 영상과는 별개의 직접 획득한 영상을 이용하여 모델을 평가함으로써 보다 능동적인 어플리케이션 및 서비스 개발이 가능할 것으로 보인다.

I. 서론

영상을 기반으로 한 인간의 동작을 인식하고 이해하는 방법은 오랜 기간동안 계속해서 연구되어 온 인공지능의 핵심 과제이다. 이를 위해 영상 내 사람 신체의 키포인트(Keypoints)들을 찾고 이를 기반으로 한 인체 자세 추정 연구가 활발하게 진행되고 있으며, 인공지능 발달로 이는 상당한 진전을 이루었지만, 단일 시점에서의 자세 추정의 경우 인체 내 일부 관절이 보이지 않거나 다른 물체나 사용자에게 의해 가려짐 등으로 인한 많은 어려움이 있다. 이러한 문제의 해결을 위해 여러 대의 카메라를 이용한 다중 시점 기반의 3차원 인체 자세 추정 기술들이 최근 많이 소개되고 있으나, 이 역시 학습 및 테스트를 위한 실측값(Ground-truth)를 획득하는 것이 쉽지 않고 학습 환경과 동일한 카메라 환경 구성 및 캘리브레이션 정보 등 보다 정확한 3차원 자세 추정 결과를 얻기 위한 데이터가 부족하다는 문제가 있다 [1].

이와 관련하여 본 연구에서는 인체 내 고유한 특성을 갖는 신체 부위 및 이를 활용한 기하학적 분석이 가능한 특정 동작을 제시하여 단일 시점 영상만으로 3차원 자세 추정 결과의 신뢰도를 측정 및 평가할 수 있는 방법을 소개하였으며, 이를 통하여 실측값을 포함한 데이터셋을 활용한 테스트 및 성능 평가의 필요성을 완화할 수 있을 것으로 기대된다. 또한 본 논문에서는 3차원 자세 추정 분야 내 실제 SOTA 모델[2]들과의 순위 비교를 통하여 제안한 신뢰도 측정 방법이 유효함을 입증하였다.

II. 본론

본 논문에서는 3차원 자세 추정 모델의 신뢰도 평가를 위해 앞서 설명한 것과 같이 신체 내 특정 부위 및 특정 동작을 척도로 제시하고자 한다. 예를 들어 인체의 상완, 전완, 종아리, 허벅지 등의 길이는 어느 각도에서 보아도 항상 일정해야 하며, 동일하게 팔을 뻗었을 때 팔꿈치의 각도는 항상 180도로, 팔을 앞으로 뻗고 있을 때 몸통과 팔(겨드랑이)의 각도는 항상 90도에 가깝게 일정하여야 한다. 따라서 아래 그림 1과 같이, 한 팔은 앞으로 곧게 뻗고 다른 팔은 위로 곧게 뻗은 상태로 하나의 카메라를 기준으로 360도로 돌며 영상을 획득한다고 가정했을 때,

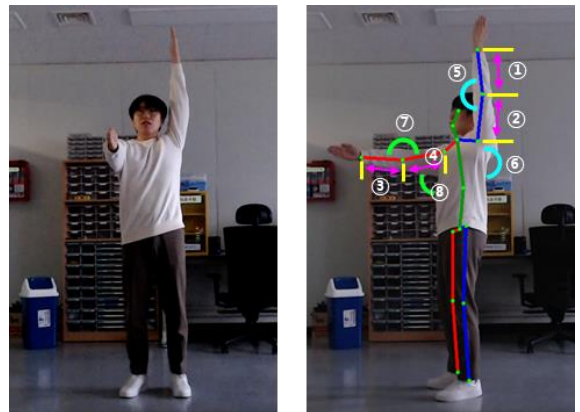


그림 1. 자세 추정 신뢰도 평가 척도의 예 (총 8개, 길이 4부위 (①~④), 전완 및 상완), 각도 4곳 (⑤~⑧, 팔꿈치 및 겨드랑이))



그림 2. 자세 추정 신뢰도 평가 그래프 (총 8 개, 길이 4 부위 (①~④, 전완 및 상완), 각도 4 곳 (⑤~⑧, 팔꿈치 및 겨드랑이))

이 영상에 대한 3 차원 자세 추정 결과로부터 산출한 두 팔의 상완 및 전완의 길이와 두 팔꿈치의 각도 (180 도) 및 두 겨드랑이 각도(90 도, 180 도)는 영상의 시작부터 끝까지 항상 일정하여야 한다. 따라서 본 연구에서는 위 제시한 특정 동작을 촬영한 단안 영상으로부터 신체 내 특정 부위를 척도로 활용해 그 길이나 각도가 일정하게 유지되는 지를 확인하고자 하며, 이를 보기 위해 위 그림 2 와 같이 각 척도에 대한 길이와 각도를 그래프로 그려 확인할 수 있었으며, 그래프의 변화 정도를 아래 수식 1 과 같이 계산하였다. 본 논문에서는 이와 같이 측정된 값을 '일관성'이란 이름의 지표로 정의하고 일관성 값이 작을수록 3 차원 자세 추정 결과 신뢰도가 높은 모델이라 평가할 수 있게 된다.

$$Consistency(x) = \frac{\sum_{i=1}^f |x_i - \sigma|}{\alpha_{scale} \times N}, \quad \begin{cases} \alpha_{scale}: scale\ factor \\ N: number\ of\ frames \end{cases} \quad (1)$$

$$where\ \sigma = \frac{M + m}{2}, \quad \begin{cases} M = \max(X) \\ m = \min(X) \end{cases}$$

위 수식에서 일관성 값은 영상 모든 프레임에서 특정 부위의 길이 및 각도 값의 집합을 X 라 할 때, 그 집합의 최대값과 최소값의 반을 σ 로 하여 각 프레임에서의 길이 및 각도 값과 σ 의 차의 절대값을 모두 더한 값으로, 이를 각 일관성 값의 스케일링을 위해 스케일링 팩터 α_{scale} 및 N 으로 나뉜다. 이 때, 스케일 팩터 α_{scale} 은 일관성 값이 길이일 때는 신체 부위 중 가장 긴 부위의 길이, 각도일 때는 신체 각도 중 가장 큰 값인 180 을 사용하고, N 은 영상의 전체 프레임 수를 의미한다.

III. 시험 결과

본 논문에서는 여러 3 차원 자세 추정 모델을 이용하여 제안한 방법 및 일관성 지표의 유효성을 확인하였으며, 이를 위해 사용한 모델은 총 3 가지로 Julieta Martinez 외 연구진들이 제안한 베이스라인 모델인 SIM[3], Bastian Wandt 외 연구진들이 제안한 self-supervised 방법의 CanonPose[4], 및 Wenhao Li 외 연구진들이 제안한 트랜스포머 모델인 MHFormer[5]이다. 이 3 개의 모델을 사용하여 위 단락에서 제시한 동작을 수행하며 촬영한 영상으로부터 3 차원 자세 추정 결과를 획득한 뒤, 이를 바탕으로 앞서 선택한 8 부위에 대한 일관성 그래프는 위 그림 2 와 같고, 각 척도에 대한 일관성 값은 다음 표 1 과 같다. 그림 2 의 그래프에서 볼 수 있는 것과 같이 그래프의 모양이 떨림없이 일정하게 유지되고 그 변화의 정도가 가장 적은 MHFormer 모델이 셋 중 가장 우수한 모델이라 할 수 있으며, 표 1 과 같이 위 수식 1 을 통해 일관성 값을 측정함으로써 각 모델의 정량적 신뢰도 수치 및 모델 간 비교 평가 결과를 확인할 수 있다.

표 1. 자세 추정 신뢰도 평가 결과 (일관성)

	①	②	③	④	⑤	⑥	⑦	⑧	평균
SIM	0.087	0.101	0.072	0.052	0.068	0.064	0.105	0.059	0.070
Canon	0.088	0.059	0.101	0.105	0.046	0.034	0.182	0.100	0.083
MHF	0.024	0.074	0.063	0.083	0.064	0.043	0.088	0.033	0.059

또한 위 표의 신뢰도 측정을 통한 모델 간 비교 평가 결과 순서가 SOTA[2]의 모델들 간 비교 결과 순서와 (Human3.6M 데이터셋 / 지표 Average MPJPE 기준: MHFormer (Rank 35, 30.5 mm), SIM (Rank 117, 45.5 mm, CanonPose (Rank 287, 74.3 mm) 동일함을 볼 수 있어서 본 논문에서 제안한 신뢰도 지표가 유효한 것도 확인할 수 있다.

IV. 결론

본 논문에서는 단일 시점 영상을 활용한 3 차원 인체 자세 추정 모델의 신뢰도 측정 방법을 소개하였으며, 이 때 인체 내의 일부 신체 부위 및 특정 동작을 척도로 한 일관성 지표를 제안하고 이에 따른 결과들을 확인하였다. 이는 제한된 환경에서 3 차원 자세 추정 기술을 활용한 어플리케이션 개발 시 쉽고 간단한 방법으로 모델들의 비교 평가를 가능하게 하며, 이를 통하여 보다 능동적인 관련 서비스의 개발이 가능할 것으로 기대된다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [24ZD1140, 대경권 지역산업 기반 ICT 융합 기술 고도화 지원(의료)]

참고 문헌

- [1] 락민지 외, "3 차원 합성 데이터와 2 차원 실제 데이터를 활용한 현실 환경에서의 3 차원 사람 자세 추정," 대한전자공학회 하계학술대회 논문집, pp. 2919-2922, 2023.
- [2] "3D Human Pose Estimation on Human3.6M," (<https://paperswithcode.com/sota/3d-human-pose-estimation-on-human36m>)
- [3] J. Martinez et al., "A simple yet effective baseline for 3d human pose estimation," *Proceedings of the IEEE International Conference on Computer Vision*, pp.2640-2649, 2017.
- [4] B. Wandt et al., "CanonPose: Self-Supervised Monocular 3D Human Pose Estimation in the Wild," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 13294-13304, 2021.
- [5] W. Li et al., "MHFormer: Multi-Hypothesis Transformer for 3D Human Pose Estimation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 13137-13146, 2022.