

# 관절 각도 기반 인체 자세 유사도 평가 기법

정시은\*<sup>1</sup>, 강지희\*<sup>2</sup>, 서현\*\*

## A Human Pose Similarity Evaluation using Joint Angles

SiEun Jung\*<sup>1</sup>, JiHee Kang\*<sup>2</sup>, and Hyeon Seo\*\*

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2023-00280241)

### 요약

자세 추정의 기존 픽셀 기반 평가는 외부 환경 변화에 민감해 실제 자세를 반영하지 못하며, HRNet과 같은 대규모 모델은 높은 정확도를 제공하지만, 자원이 제한된 환경에서 실시간 처리가 어렵다. 본 연구는 MediaPipe 기반 관절 좌표에서 포즈 벡터를 구성해 코사인 유사도를 적용하는 평가 방법을 제안한다. 실험 결과, 각도 기반 평가는 픽셀 기반보다 구분력이 높았고, 가장 높은 FPS를 기록해 제한된 연산 환경에서 적합성을 보였다. 이는 제안한 방법이 제한된 연산 환경에서 실시간 동적 자세를 분석하는 데 유효함을 보여준다.

### Abstract

Pixel-based pose evaluation is sensitive to environmental changes and fails to reflect actual poses, while large models like HRNet offer high accuracy but unsuitable for real-time use in resource-limited setting. This study builds pose vectors using MediaPipe joints and cosine similarity. Experiments show that angle-based evaluation offer higher discriminative power than pixel-based methods and achieves the highest FPS, demonstrating suitability for low-resource setting. Thus, the proposed method supports real-time dynamic pose analysis in constrained environments.

### Keywords

pose estimation, human body postures, similarity, real-time

### 1. 서론

인체 자세 추정 기술은 운동 분석, 인간-컴퓨터 상호작용, 가상현실 등 다양한 응용 분야의 핵심 기반을 제공한다. 이러한 환경에서는 즉각적인 추정과 피드백 제공이 요구되므로[1] 실시간 처리 성능이

중요한 조건으로 제시된다. 더불어 실제 환경에서는 연산 자원이 제한되기 때문에 제약 속에서 안정적인 동작할 수 있는 경량 모델의 설계가 중요하다. 한편, 기존 인체 자세 추정 연구들은 주로 HRNet과 같은 대규모 딥러닝 모델을 활용해왔다[2]. 이러한 구조는 높은 정확도를 확보하는 데 효과적이거나, 연산량이 방대하여 실시간 환경에는 적합하지 않다.

\* 경상국립대학교 컴퓨터공학부 학사과정  
- ORCID<sup>1</sup>: <https://orcid.org/0009-0006-3874-965X>

- ORCID<sup>2</sup>: <https://orcid.org/0009-0009-1947-0153>

\*\* 경상국립대학교 IT공과대학 컴퓨터공학부 교수(교신저자)  
- ORCID: <https://orcid.org/0000-0001-8488-3733>

• Received: Oct. 13, 2025, Revised: Dec. 07, 2025, Accepted: Dec. 10, 2025

• Corresponding Author: Hyeon Seo

Dept. of Computer Science, Gyeongsang National University, Jinju, Korea

Tel.: +82-55-772-1385, Email: [hseo0612@gnu.ac.kr](mailto:hseo0612@gnu.ac.kr)

자세 유사도 평가 연구에서는 주로 정적 이미지를 대상으로 픽셀 또는 특징 맵 기반 거리 계산 방식을 사용해왔으나, 배경·조명 변화 및 카메라 시점 변화에 민감해 실제 자세의 구조적 차이를 반영하기 어렵다[3][4]. 이런 한계를 보완하기 위해 본 연구에서는 경량 인체 자세 추정 모델인 MediaPipe[5]를 활용해 영상으로부터 관절 좌표를 추출하고, 기하 구조기반 포즈 벡터와 코사인 유사도를 적용한 자세 유사도 분석 방법을 제안한다.

## II. 데이터셋

본 연구에서 사용한 데이터셋은 총 10개의 샘플로 구성되어 있으며, 각 샘플은 관절 시각화 영상(mp4), 음악 파일(mp3), 포즈 좌표 데이터 파일(txt)로 이루어져 있다. 음악 파일은 동작 타이밍을 맞추기 위한 기준 신호로 사용되었으며, 피실험자는 박자에 맞춰 동작을 수행함으로써 리듬 단위별 일관된 자세 비교가 가능하였다. 데이터셋은 재생 시간을 기준으로 60초 이하의 short 버전 7개와 60초를 초과하는 long 버전 3개로 구분하였다. 각 샘플의 재생 시간(Duration), FPS(Frames Per Second) 등 상세 정보는 표 1과 같다. 모든 데이터는 원본 영상의 재생 시간을 기준으로 균등 간격 샘플링하여, 각 영상당 총 100프레임으로 통일하였다.

표 1. 실험 데이터 세트  
Table 1. Experimental data sets

ID	Type (Short/Long)	Duration(sec)	FPS
0	Short	26	26.13
1	Long	113	113.20
2	Short	35	34.93
3	Short	21	20.73
4	Long	78	78.38
5	Short	12	11.60
6	Short	29	28.83
7	Short	40	39.87
8	Long	87	86.81
9	Short	27	86.81

각 포즈 좌표 데이터 파일에는 프레임 단위로 추출된 33개 관절의 좌표가 기록되며 (그림 1), 본 연구에서는 이를 이용해 포즈 벡터를 구성하고 유사도 계산에 활용하였다.

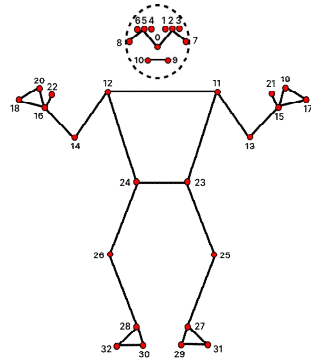


그림 1. 33개의 신체 랜드마크 위치  
Fig. 1. 33 body landmark locations

## III. 방법론

### 3.1 관절 각도 산출

관절 좌표는 MediaPipe Pose Landmarker를 사용해 프레임별로 33개 키포인트의 2차원 좌표를 추출하고, 좌·우 골반(landmark 23, 24)의 평균을 원점으로 중심 정규화를 수행하였다. 여러 인물이 검출된 경우, 머리 키포인트(landmark 0)를 기준으로 좌→우 순서로 정렬해 일관된 인덱싱을 유지하였다. 관절 각도는 사전 정의된 10개의 관절 트리플에 대해 계산하였으며, 구성 정보는 표 2에 제시하였다. 각 트리플  $(i_1, i_2, i_3)$ 에서 중심 관절( $i_2$ )을 기준으로 두 벡터  $u = p_{i_1} - p_{i_2}, v = p_{i_3} - p_{i_2}$ 를 구성하고, 이들 사이의 각도  $\theta$ 는 코사인 관계식(식 (1))에 따라 계산하였다.

표 2. 관절 트리플  
Table 2. Joint triples

Target angle	$i_1$	$i_2$	$i_3$
Left elbow	11	13	15
Right elbow	12	14	16
Left ankle	25	27	29
Right ankle	26	28	30
Left hip	11	23	25
Right hip	12	24	26
Left shoulder	13	11	23
Right shoulder	14	12	24
Pelvis-R.leg	23	24	26
Pelvis-L.leg	24	23	25

$$\cos\theta = \frac{u \cdot v}{\|u\| \|v\|} \quad (1)$$

코사인값을 각도로 변환하고, 범위를 [-1, 1]로 보정하여 계산의 안정성을 확보한 뒤 이를 각 프레임의 10개 관절 트리플에 적용해 10차원 포즈 벡터를 생성하였다.

### 3.2 포즈 유사도 산출

원본 안무와 실시간 사용자 자세의 시계열 데이터로부터 각각 10차원 벡터를 생성하고, 두 벡터 간 코사인 유사도를 식 (2)와 같이 계산한다.

$$s = \frac{A \cdot B}{\|A\| \|B\| + \epsilon} \quad (2)$$

수치적 안정성을 위해 분모에 작은 상수  $\epsilon$ 를 추가하였으며, 연산 효율을 고려해 동시 처리 인원을 최대 4명으로 제한하였다. 벡터 연산은 배치 처리로 수행해 검출된 모든 인물의 유사도를 병렬 계산하였다. 코사인 유사도 임계값은 실험적으로 0.6으로 설정하였으며, 이는 두 자세의 구조적 차이가 명확히 구분되는 지점으로 확인된 값이다. 식 (3)와 같이 설정된 임계값 이상인 구간은 0~100% 범위로 선형 변환하고, 그 이하 구간은 0으로 클리핑하였다.

$$s_{percent} = \begin{cases} 100 & \text{if } s \geq 1.0 \\ \frac{s - 0.6}{0.4} \times 100 & \text{if } 0.6 \leq s \leq 1.0 \\ 0 & \text{if } s < 0.6 \end{cases} \quad (3)$$

## IV. 결 과

본 연구에서는 픽셀 기반과 각도 기반, 두 가지 접근법으로 자세 유사도를 비교 분석하였다. 그림 2와 그림 3은 각 방법의 적용 예시를 보여준다.

표 3은 10개의 데이터셋에 대한 자세 유사도 평가 결과를 나타낸다. 픽셀 기반 방식의 평균 유사도는 동일 자세 92.61%, 다른 자세 92.38%로 차이가 0.23%에 불과해 실제 자세 차이를 구별하기 어렵

다. 반면, 각도 기반 방식은 동일 자세 평균 93.98%, 다른 자세 80.68%로 유사도 차이가 13.30%에 나타나, 자세 구분에 더 효과적이며 차이가 뚜렷하게 나타남을 확인할 수 있었다.

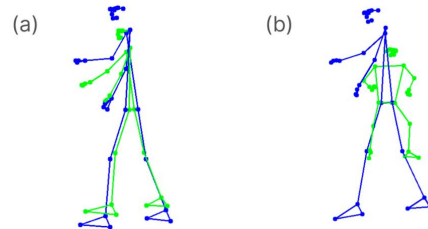


그림 2. 픽셀 기반 자세 추정 (a) 동일 자세, (b) 다른 자세 (파란색은 정답 자세, 초록색은 추정 자세를 나타냄)  
Fig. 2. Pixel-based pose estimation (a) Same pose, (b) Different pose (Blue: ground truth; Green: estimated pose)

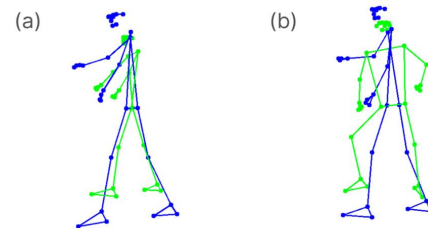


그림 3. 각도 기반 자세 추정 (a) 동일 자세, (b) 다른 자세 (파란색은 정답 자세, 초록색은 추정 자세를 나타냄)  
Fig. 3. Angle-based pose estimation  
(a) Same pose, (b) Different pose (Blue: ground truth; Green: estimated pose)

표 3. 자세 유사도 결과  
Table 3. Pose similarity results

ID	Pixel-based		Angle-based	
	Correct pose similarity (%)	Incorrect pose similarity (%)	Correct pose similarity (%)	Incorrect pose similarity (%)
0	92.4	92.3	94.8	81.2
1	92.6	92.8	94.6	78.8
2	93.3	92.4	94.2	79.8
3	92.1	92	90.6	73.6
4	93.4	92.6	87.8	82.8
5	92.8	92.1	93.1	86.5
6	92.7	92.1	93.9	80.4
7	91.7	91.6	92.7	78.8
8	93.7	92.6	93.2	83.3
9	92.7	93.0	93.0	82.2

표 1에서 구분된 Short와 Long 영상의 영향을 분석하였다. Long 구간의 경우 FPS와 프레임 수가 많아 프레임 간 관절 각도 변화 폭이 작게 유지된다. 이에 따라 각도 기반 유사도는 구간 전체에서 안정적이고 일관된 값을 보였다. 반면, Short 영상에서는 급격한 자세 변동 구간이 많아 구간 내 빠른 동작 변화로 인해 유사도 값의 변동이 다소 컸지만, 픽셀 기반 대비 높은 분별력을 유지하였다.

제안 방식의 실시간 적용 가능성을 확인하기 위해, 고정밀 포즈 추정 모델인 HRNet[2]을 동일한 환경에서 평가하였다. 실험은 CPU 기반 M2 MacBook Air(16GB RAM)에서 수행하였으며, 초당 처리 가능한 프레임 수(FPS)를 측정하였다.

표 4는 세 가지 방식의 FPS 결과를 나타낸다. MediaPipe 각도 기반 방식은 91 FPS, 픽셀 기반 방식은 68 FPS, HRNet 기반 방식은 6 FPS로 확인되었다. 저-중간 사양의 CPU 환경에 MediaPipe 기반 방식은 높은 처리 속도를 보인 반면, HRNet 기반 방식은 낮은 FPS로 인해 실시간 적용이 어려웠다.

표 4. 포즈 추정 방식별 FPS 결과

Table 4. Results of FPS among pose estimation methods

Methods	FPS(Frame Per Second)
Angle-based	91
Pixel-based	68
HRNet (DeepLearning)	6

## V. 결 론

본 연구에서는 인체 관절 각도 기반 포즈 벡터와 코사인 유사도를 이용하여 자세 유사도를 정량적으로 평가하는 방법을 제안하였다. 제안 방식은 기존 픽셀 기반 방식보다 자세의 구조적 차이를 정확히 반영하며, 다양한 환경에서도 일관된 결과를 보였다. 또한, 각도 기반, 픽셀 기반, HRNet 모델을 동일 환경에서 비교한 결과, 제안 방식이 가장 높은 FPS를 보였다. 이에 본 방법이 다양한 분야에서 즉각적인 피드백이 요구되는 응용에 활용 가능성을 확인하였다. 향후 연구에서는 관절 수 확장, 3차원 데이터 적용, 동작 유형 다양화 및 대규모 데이터셋을 활용한 실험을 통해 제안 방식의 일반화 성능과 강건성을 더욱 고도화할 예정이다.

## References

- [1] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 6, pp. 7157-7173, Jun. 2023. <https://doi.org/10.1109/TPAMI.2022.3222784>.
- [2] M. Wang, J. Tighe, and D. Modolo, "Combining Detection and Tracking for Human Pose Estimation in Videos", *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 11085-11093, Jun. 2020. <https://doi.org/10.1109/CVPR42600.2020.01110>.
- [3] X. Sun, M. Adamu, R. Zhang, X. Guan, and Q. Li, "Pixel-Coordinate-Induced Human Pose High-Precision Estimation Method", *Electronics*, Vol. 12, No. 7, p. 1648, Mar. 2023. <https://doi.org/10.3390/electronics12071648>.
- [4] W. J. Kim, "A Posture Similarity Analysis Scheme based on the Confidence Score of Human Body Keypoints", *Journal of Korean Institute of Information Technology*, Vol. 21, No. 12, pp. 171-177, Dec. 2023. <https://doi.org/10.14801/jkiit.2023.21.12.171>.
- [5] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A Framework for Building Perception Pipelines", *arXiv preprint arXiv:1906.08172*, pp. 1-17, Jun. 2019. <https://doi.org/10.48550/arXiv.1906.08172>.