

시니어 홈트레이닝을 위한 BiLSTM 기반 동작 분류 및 자세 유사도 분석에 관한 연구 - MediaPipe Pose 기반 접근법

김하진. 하인우. 임재영. 김현기. 윤수연*

Windsor-Essex Catholic District School, 국민대학교, *국민대학교

hajin.kim@mytools2go.ca, herb@kookmin.ac.kr, dlawowo01@kookmin.ac.kr, avalonia@kookmin.ac.kr,
*1104py@kookmin.ac.kr

A Study on BiLSTM-Based Motion Classification and Pose Similarity Analysis for Senior Home Training: A MediaPipe Pose-Based Approach

Hajin Kim, InWoo Ha, Jae Young Im, Kim Hyun Ki, Soo-Yeon Yoon*

Windsor-Essex Catholic District School, Kookmin University, *Kookmin University

요약

본 논문은 시니어 대상 홈트레이닝 환경에서 운동 수행의 안전성과 효과성을 제고하기 위해, 인공지능 기반 코칭 시스템을 위한 자세 분류 및 정밀 분석 프레임워크를 제안한다. MediaPipe Pose를 활용하여 실시간으로 사용자의 관절 키포인트를 추출하고, Bidirectional Long Short-Term Memory(이하 BiLSTM) 기반 시계열 분류모델을 통해 8종의 운동 동작을 자동분류하였다. 또한, Dynamic Time Warping(이하 DTW), Cosine Similarity, Object Key Points Similarity(이하 OKS)를 결합한 복합 유사도 지표를 적용하여 시범 동작과 사용자 동작 간의 자세 유사도를 정량적으로 평가하였다. 실험결과, 제안한 포즈 분류모델은 90% 이상의 정확도를 보였으며, 다양한 실제 운동 영상에서도 안정적인 성능을 확인하였다. 본 연구는 시니어 홈트레이닝을 위한 AI 코칭 시스템의 실용적 적용 가능성과 확장성을 제시한다.

I. 서론

고령화 시대, 시니어 인구 비중이 증가하면서 건강하고 활동적인 노후를 추구하는 액티브 시니어 계층 역시 늘어나고 있다. 이에 따라 이들을 대상으로 한 AI 기반의 디지털 헬스케어 기술이 빠른 속도로 발전하는 중이다.

시니어 계층은 운동을 위해 피트니스 클럽을 찾기에는 이동이 불편한 경우가 많고, 피트니스 클럽에서도 기구 운동을 하기에는 근력이 부족하여 적합한 운동을 하지 못하는 경우가 상당하다. 이러한 상황은 홈트레이닝을 기반으로 한 시니어 대상 헬스케어 산업이 성장할 수 있는 가능성을 보여준다.

단, 홈트레이닝 시 사용자의 현재 상태, 목표, 자세 등을 토대로 개인별 맞춤형으로 코칭할 필요가 있다. 그러지 않을 경우 피트니스 프로그램의 주 수요층인 젊은 층을 대상으로 설계된 프로그램과 콘텐츠로 인해 시니어들의 부상발생위험이 증가하고 운동 가능한 범위도 제한될 수밖에 없다.

본 연구는 시니어들의 안전하고 효과적인 홈트레이닝을 지원하기 위해 AI 기반을 통한 개인적 맞춤형 분석 및 피드백 시스템을 제안한다. 이를 위해 딥러닝 기반 동작 분류 및 자세 유사도 분석 시스템을 제안한다.

II. 관련 연구

2.1. Pose Estimation

Pose estimation은 영상 데이터로부터 인체의 주요 관절 키포인트를 검출하여 사람의 관절 위치를 추정하여 실시간 자세 분석, 운동 모니터링, 행동 인식 등의 이러한 특성으로 인해 여러 응용 환경에서 폭넓게 활용되고 있다. 딥러닝 기반 Pose estimation 기술은 이미지나 영상입력으로부터 인체 관절 키포인트의 위치를 정확하게 예측함으로써 사용자의 행동과 수행 상태를 수치화해 분석할 수 있게 하며, 특히 실시간 피드백이 필요한 피트니스 및 운동 분석환경에서 적용 가능성이 높아 지고 있다[1].

대표적인 Pose estimation 모델로는 OpenPose, HRNet 등 고성능 모델과 MoveNet·BlazePose·MediaPipe Pose 등 경량 실시간 처리 모델이 있다.

OpenPose는 높은 정확도를 보이나 계산량이 많아 실시간 처리 및 GPU 의존성이 높은 반면, MoveNet/BlazePose류는 경량화와 실시간 처리에 강점이 있다.[2]. 특히 MediaPipe Pose는 CPU 환경에서도 실시간 처리 가능하며 33개의 관절 랜드마크를 안정적으로 추출할 수 있어 다양한 플랫폼(Python, Web, Mobile)에서 활용 가능하다[3]. 따라서 본 연구에서는 실시간성과 확장성을 고려하여 MediaPipe Pose를 최종 선택하였다.

2.2. Pose Classification

Pose classification은 추정된 관절 좌표를 바탕으로 사람의 자세 또는 행동을 자동으로 분류하는 기술이다. 이는 운동 동작 인식, 행동 모니터링, 이상 행동 탐지 등 여러 응용 영역에 활용되고 있다.

Pose classification을 구현하기 위한 딥러닝 기반 접근법으로는 CNN, RNN 계열 모델, Transformer 등이 있으며, 각 모델은 시퀀스 형태의 관절 데이터를 처리하는 방식과 장·단점이 다르다. 특히 신체 관절 시퀀스를 입력하여 사용하는 1차원/Temporal CNN은 국소적인 시간 패턴 학습하는 데에는 효과적이지만 전반에 걸친 장기 의존성 모델링에는 한계를 보인다. 이는 동작 전체의 흐름을 반영하지 못한다는 문제가 보고되었다[4]. Transformer 기반 모델은 전역 문맥(global context) 학습에 강점을 가지나 중소규모 데이터셋에서는 과적합 위험이 있으며 계산 비용이 높아 실시간 처리에 제약이 있다[5].

양방향 시계열 구조를 갖는 BiLSTM은 시퀀스 내 동작 전이 구간을 효과적으로 포착할 수 있다. 이러한 특성은 특히 Skeleton 기반 데이터와 결합했을 때 조명이나 배경 변화에 강인하며 비교적 소규모 데이터셋에서도 강한 성능을 보인다[6].

이에 본 연구에서는 시니어의 운동 동작을 정밀하게 인식하기 위한 핵심 모델로 BiLSTM을 채택하였다.

2.3 DTW (Dynamic Time Warping)

한편, 사용자의 동작 정확도를 평가하기 위해서는 전문가의 표준 동작과 사용자 동작 간의 유사도를 정량적으로 측정해야 한다. 그러나 시니어 계층은 신체적 능력의 저하로 인해 표준 동작 대비 수행 속도가 느리거나 일정하지 않은 경향이 있다. 단순한 유클리디안 거리 기반의 비교는 이러한 시간적 불일치를 고려하지 못해 정확한 평가가 어렵다는 한계가 있다..

이에 비해 DTW의 경우 두 시계열 데이터의 길이가 다르거나 속도 차이가 발생하더라도, 시점 간의 비선형적 매칭을 통해 최적의 정렬 경로를 탐색할 수 있다. 이러한 특성 덕분에 DTW는 동작의 속도가 변하더라도 자세의 형상적 유사성을 측정할 수 있는 기법으로 널리 활용되고 있다. 따라서 본 연구에서는 시니어의 불규칙한 운동 속도를 보정하고 정확한 유사도를 측정하기 위한 알고리즘으로 DTW를 적용하였다.

III. 실험 설계

3.1. 시스템 아키텍처

본 AI 분석시스템은 Pose Estimation, Pose Classification, Pose Similarity의 총 3 단계로 구성된다.

먼저, 포즈 추정(Pose Estimation) 단계에서는 MediaPipe Pose 모델을 이용하여 실시간으로 사용자의 운동 영상과 시범 영상에서 33개 관절의 키포인트를 프레임 단위로 추출한다. 추출된 각 관절 포인트들은 x, y, z 좌표 및 Visibility 정보를 포함한다.

포즈 분류(Pose Classification) 단계에서는 BiLSTM 모델을 채택하였으며, 추출된 랜드마크 시퀀스를 바탕으로 총 132 차원(33개 랜드마크 × 4차원 특성)의 입력 시퀀스를 구성하여 8개의 운동 동작을 분류하도록 설계하였다. 또한 Stratified Split과 Pack Padded Sequence를 적용하여 가변 길이 시퀀스를 효과적으로 학습하고자 하였다.

마지막으로 포즈 유사도 측정(Pose Similarity) 단계에서는 시범 영상과 사용자 운동 영상 간 프레임을 DTW 알고리즘으로 정렬하고, Cosine Similarity, L2 Distance, OKS를 기반으로 자세의 유사도를 정량적으로 측정한다.

이렇게 구성된 시스템 아키텍처는 그림 1과 같이 표현될 수 있다.

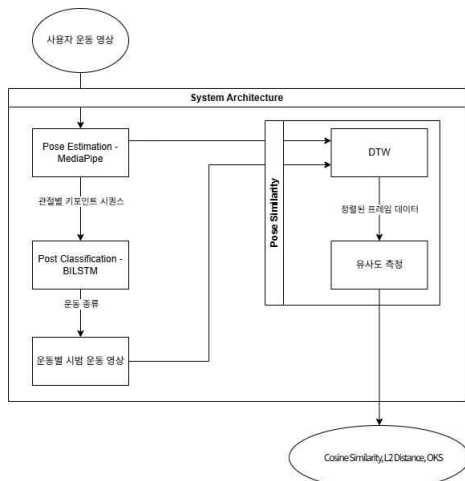


그림 3. 시스템 아키텍처

유사도 측정지표의 경우, Cosine Similarity는 관절 쌍방향의 형태가 일치하는 정도를 평가하고, OKS는 COCO(Common Object in Context)에서 제안된 표준 유사도 지표로서 관절별 위치 오차를 고려한다. L2 Distance는 공간적 위치 오차를 측정하고, 일정 임계값을 초과할 경우 자세 불일치(NG)로 분류하게 된다.

이를 기반으로 최종 유사도 점수는 Cosine Similarity와 OKS의 곱으로 산출되며, 위치와 형태정보를 동시에 반영하도록 하였다.

3.2. 데이터셋 구성

데이터는 유튜브를 통해 총 145개의 홈트레이닝 영상을 수집하여 구성하였다. 수집한 영상은 합계 2시간가량의 분량으로 총 1,203회의 운동 반복 동작이 포함되어 있다. 구축된 데이터셋의 파라미터는 다음 표 1과 같다.

항목	내용
모델구조	2-layer BiLSTM, Hidden Size 64, Dropout 0.3
입력 차원	132 (33개 관절 X 4개 속성)
출력	8개 운동항목 스쿼트, 런지, 섯업, 하이니즈, 브릿지, 코브라, 사이드런지, 점핑잭
데이터분할	학습데이터 60%, 검증데이터 20%, 테스트데이터(Test)20% Stratified Split
Optimizer	Adam (Adapted Moment Estimation)
Loss Function	Cross Entropy Loss
Epoch	80
Batch Size	4

표 1. 데이터셋 파라미터

IV. 실험 결과

4.1. 포즈 추정의 결과

MediaPipe Pose 모델을 사용하여 조명, 배경, 신체 등 다양한 조건에서 안정적으로 33개의 관절 키포인트를 추출할 수 있었다. 그 결과 시니어 사용자의 실제 동작에서도 높은 정확도를 유지하였다.

4.2. 포즈 분류의 결과

BiLSTM 기반 포즈 분류 모델을 통해 8개의 운동 동작을 분류한 결과 전체 정확도는 90% 이상으로 나타났다. Confusion Matrix 분석 결과, 섯업, 점핑잭 등 일부 동작은 95% 이상의 높은 인식률을 보였다. Skeleton 기반의 입력방식은 RGB 기반 방식 대비 배경과 조명변화에도 인식오류가 적은 특징을 나타냈다. 그 결과를 요약하면 다음 그림 2와 같다.

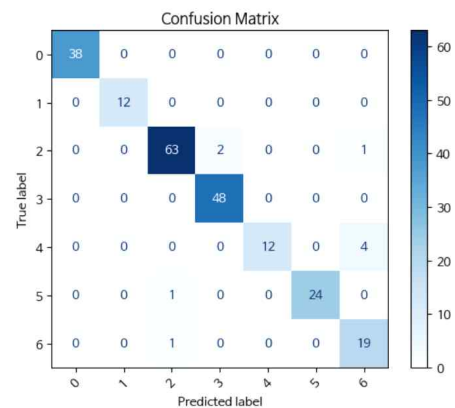


그림 2. BiLSTM 기반 포즈 분류 모델의 Confusion Matrix
위 결과는 본 연구에서 사용한 Skeleton 기반 포즈 분류 모델이 프레임 단위 동작 인식 환경에서도 안정적인 분류 성능을 유지함을 의미하며, 앞서 제시한 정량 분석 결과의 신뢰성을 뒷받침한다.

4.3. 포즈 유사도 측정 결과

시범 영상과 사용자 운동 영상의 포즈 유사도는 DTW 정렬 결과, 유사도 측정지표, 최종 유사도 계산의 방법으로 측정하였다.

DTW 정렬 결과의 경우, 같은 사람의 서로 다른 시퀀스를 비교하였을 때 DTW Distance는 8.35로 낮은 값을 보였으나, 다른 사람의 시퀀스를 비교하였을 때 DTW Distance는 57.68로 차이가 컸다. 이는 DTW 알고리즘이 시계열 정렬 및 비교에 효과적임을 의미한다.

V. 결론

본 연구 결과 BiLSTM 기반의 자세 분류는 고정된 환경뿐만 아니라 다양한 실제 운동 영상에서도 높은 정확도를 보였고, 복합 유사도 지표(DTW + Cosine + OKS)를 적용한 자세 비교시 형태적·공간적 정확도를 동시에 평가할 수 있음을 확인하였다.

이는 광범위하게 구축된 데이터셋을 바탕으로 포즈 추정 - 포즈 분류 포즈 유사도 측정의 단계를 통해 정량적으로 분석한 결과를 사용자에게 피드백하는 시스템이 시니어 대상의 홈트레이닝 프로그램에 기술적으로 적용할 수 있음을 의미한다.

이를 더 발전시키기 위해서는 시니어들의 동작분석 데이터를 기반으로 시니어들에게 적합한 시범 영상과 운동모델을 데이터베이스화하는 것, 시니어들의 접근성과 이해도를 높일 수 있는 장비와 피드백 언어의 대중화가 요구되며, 시니어들의 운동 효과를 검증하기 위해 추가적인 연구가 필요하다.

ACKNOWLEDGEMENT

본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음 (2022-0-00964).

참 고 문 헌

- [1] J. Cho, "Recent Trends in Human Pose Estimation Based on a Single Image," *The Journal of Korean Institute of Next Generation Computing*, vol. 15, (5), pp. 31-42, 2019. .
- [2] Y. & K. Serdaliyev G., "Serdaliyev, Yerlan & Kazbekova, G. (2025). REAL-TIME HUMAN POSE ESTIMATION TECHNIQUES: A COMPARATIVE REVIEW OF MODERN DEEP LEARNING APPROACHES. 10.5281/zenodo.17880888." vol. 7, (7), pp. 17-27, 2025. . DOI: 10.5281/zenodo.17880888.
- [3] J. Kim *et al*, "Human pose estimation using mediapipe pose and optimization method based on a humanoid model," *Applied Sciences*, vol. 13, (4), pp. 2700, 2023. .
- [4] W. Li, J. Tan and Y. Fan, "Human Pose Classification Based on Pose Long Short-Term Memory," *Mod Intell Times*, vol. 2, (3), 2024. .
- [5] A. Martín-Cirera *et al*, "Comparison of Transformers with LSTM for classification of the behavioural time budget in horses based on video data," *Biosystems Engineering*, vol. 242, pp. 154-168, 2024.
- [6] A. A. Ridha *et al*, "Human activity recognition by BiLSTM recurrent neural networks and support vector machine," in *2023 6th International Conference on Engineering Technology and its Applications (IICETA)*, 2023, .