

딥러닝 기반 포즈 인식 및 교정을 통한 효율적인 홈 트레이닝 시스템 개발

Development of an Efficient Home Training System through Deep
Learning-based Pose Recognition and Correction

이용준, 김태영¹⁾

Yong-jun Lee, Tae-Young Kim

(02713) 서울특별시 성북구 서경로 124 서경대학교 컴퓨터공학과
jjjuny7712@gmail.com, tykim@skuniv.ac.kr

요 약

COVID-19 확산으로 인해 헬스장이나 야외보다는 집에서 운동하는 것을 선호하는 사람들이 늘어나고 있다. 홈 트레이닝을 할 때 잘못된 방법으로 운동을 하게 되면 다칠 수도 있고 운동의 효과가 절감될 수 있는 문제가 있다. 본 논문은 학습에 기반한 포즈 분류 모델을 통해 실시간으로 사용자의 포즈를 분류하고, 사람의 관절 점을 기반으로 포즈를 분석하여 잘못된 포즈를 교정할 수 있도록 도와주는 홈 트레이닝 시스템을 제안한다. 트레이너의 비디오 영상과 웹캠을 통하여 직접 입력받은 운동 영상을 MediaPipe Pose API를 사용하여 포즈의 관절 좌표를 추출하고, 이를 데이터 세트로 학습하여 포즈 분류 모델을 구성한다. 사용자는 실시간으로 인식된 본인의 포즈를 확인하면서 운동할 수 있고 잘못된 포즈에 대해서는 올바른 어드바이스를 제공받을 수 있다. 또한 사용자가 올바른 포즈로 운동을 한 경우 운동 횟수가 증가되어 보다 효율적인 운동이 가능하도록 한다. 맨몸 운동의 대표적 운동인 스쿼트와 푸쉬업에 대해 본 시스템의 성능을 실험한 결과 스쿼트는 96%이고, 푸쉬업은 95%의 정확도를 보였고 포즈 추론 속도는 평균 31ms가 소요되어 실시간으로 활용 가능함을 알 수 있었다.

Abstract

Due to the spread of COVID-19, more and more people prefer to exercise at home rather than at gyms or outdoors. If you exercise in the wrong way during home training, you may get hurt and the effectiveness of the exercise may be reduced. This paper proposes a home training system that helps correct wrong posture by classifying user posture in real time through a deep learning-based posture classification model and analyzing posture based on human joint points. The joint coordinates of the pose are extracted using the MediaPipe Pose API for the video images of the trainer and the exercise input images obtained directly from the webcam, and the pose classification model is constructed by learning them as a dataset. The user can exercise while checking his or her pose recognized in real

1) 교신저자

time, and can be provided with the right advice for the wrong pose. In addition, when the user exercises in the right pose, the number of exercises increases, enabling more efficient exercise. As a result of testing the performance of this system for squats and push-ups, which are representative exercises of bare-body exercise, it was found that squats were 96%, push-ups showed 95% accuracy, and pose inference speed took an average of 31 ms, making it available in real-time.

키워드: 딥러닝, 포즈 인식, 홈 트레이닝, 미디어파이프

Keyword: Deep-Learning, Pose Recognition, Home Training, MediaPipe

1. 서론

글로벌 헬스케어 기업인 뉴스킨 파마넥스가 최근에 실시한 ‘건강한 삶과 운동에 대한 한국인 인식 조사’ 결과에 따르면 국내 성인 10명 중 9명은 현재 운동을 하고 있다고 답할 정도로 국내에 헬스 열풍이 불고 있다[1]. 이와 더불어 COVID-19 확산으로 인해 헬스장이나 야외에서 하는 운동 대신 집에서 운동을 하는 홈 트레이닝을 선호하는 사람들이 늘어나고 있는 실정이다.

기존 출시된 홈 트레이닝 시스템 중 카카오 VX의 스마트 홈트[2]는 기본적으로 트레이너의 비디오를 제공하며, 사용자가 해당 비디오를 따라 하면서 운동을 수행할 수 있도록 한다. AI 기능이 있는 일부 운동들은 포즈 인식 기술을 통해 사용자의 관절 정보와 트레이너의 관절 정보를 비교하여 트레이너의 포즈와 비슷해지도록 운동 포즈를 실시간으로 피드백해준다. 그 밖의 관련 연구로 포즈 트레이너[3]는 자신이 운동한 영상을 녹화 후 제출하면 포즈 트레이너가 해당 영상을 분석하여 잘못된 포즈를 감지하고 올바른 포즈로 유도하는 내용을 담아서 사용자에게 전달해준다. 요가 포즈를 인식하는 AI 요가 트레이너[4]는 요가 포즈 영상을 데이터 세트로 하여 학습한 포즈 분류 모델을 통해 요가 포즈를 분류한다. 이를 위해 MobileNetV2 기반 포즈 추정을 통해 관절을 감지하고, 해당 모델의 입력값으로 사용하여 6가지의 요가 포즈를 분류한다. 또한 스마트 헬스케어 운동관리를 위한 애플리케이션[5]은 PoseNet을 사용하여 17개의 관절 좌표를

추출하고, 해당 관절 좌표를 분석하여 손뱁치기 수행을 감지하며, 그 결과값을 신체 활동 에너지 비용을 평가하는 척도인 신진대사 해당치로 환산하여 보여준다. 비대면 온라인 홈 트레이닝 어플리케이션 연구[6]는 운동 강도를 나타내는 METs와 사용자의 체중을 토대로 운동별 소모 칼로리를 계산한다. 또한 MoveNet을 통해 17개의 관절 좌표를 추출하고, 트레이너 영상의 관절 좌표와 사용자의 관절 좌표를 프레임별로 비교하여 정확한 자세인지 판별하도록 한다. 판별 과정에서 사용자의 관절 좌표가 트레이너와 비교하여 반경 3픽셀 이내이면 정확한 운동 자세로 판단하여 GOOD 메시지를 출력하고 아니면 BAD 메시지를 출력한다.

홈 트레이닝을 수행할 때 잘못된 방법으로 운동을 하게 되면 다칠 수도 있고 운동의 효과가 절감될 수 있으므로 정확한 방법으로 운동하는 것이 매우 중요하다. 기존에 출시된 홈 트레이닝 시스템의 경우 트레이너의 지시에 따라 운동을 수정하도록 하는 경우가 대부분이고 사용자의 포즈를 트레이너의 포즈와 비교하므로 객관적인 포즈 교정이 안될 수 있다. 따라서 본 논문은 PC와 모바일 환경에서 학습에 기반한 포즈 분류 모델을 통해 실시간으로 사용자의 포즈를 분류하고 관절 점을 기반으로 잘못된 포즈를 인식하여 교정할 수 있도록 도와주는 홈 트레이닝 시스템을 제안하고자 한다.

본 논문은 트레이너 비디오 영상과 웹캠으로 직접 입력받은 운동 포즈 영상에 대하여 MediaPipe Pose API를 사용하여 관절 좌표를 추출하고, 이를 데이터 세트로 학습하여 포즈 분류 모델을 구성한

다. 사용자는 실시간으로 인식된 본인의 포즈를 보면서 운동할 수 있고 잘못된 포즈에 대해서는 올바른 어드바이스를 제공받을 수 있다. 또한 사용자가 올바른 포즈를 한 경우에 대해서만 운동 횟수를 세어주어 효율적인 운동을 할 수 있도록 한다.

본 논문에서 지원하는 운동은 맨몸 운동의 대표적인 운동인 스쿼트와 푸쉬업 2가지이고 다른 포즈의 운동도 쉽게 확장 가능하다. 포즈 교정의 기준으로 스쿼트는 발의 각도와 위치를, 푸쉬업은 손의 각도와 엉덩이의 위치를 사용한다. 또한 운동 횟수 세기에 사용된 정보로 스쿼트는 다리의 각도를 사용하고 푸쉬업은 팔의 각도를 사용한다. 트레이너 비디오를 분석하여 추출된 정보들을 바탕으로 유효 범위 또는 기준치를 두어 포즈 교정 및 운동 횟수 세기를 수행하게 되며 포즈 분류와 포즈 교정 및 운동 횟수 세기 등의 결과를 사용자에게 전달하기 위해 TTS(Text-To-Speech) 안내 음성을 출력한다.

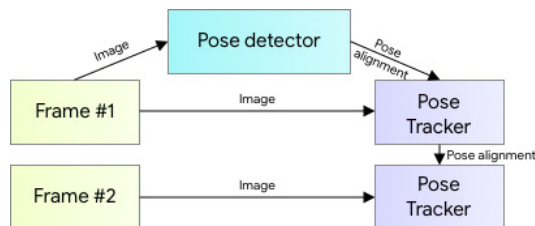
개발된 시스템의 성능을 실험한 결과 포즈 분류 모델의 스쿼트는 약 96%, 푸쉬업은 약 95%의 실전 정확도를 보이고 포즈 분류 모델 추론 속도는 평균 31ms가 소요되어 실시간으로 활용 가능함을 알 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서 MediaPipe 기반 포즈 인식 방법을 소개하고, 3장에서는 본 논문의 딥러닝 기반 홈 트레이닝 시스템에 대한 상세한 내용을 설명한다. 4장에서는 본 논문의 실험을 위하여 구성된 훈련 데이터 세트와 포즈 분류 모델의 실험 결과를 설명한 다음 5장에서 결론을 맺는다.

2. MediaPipe 기반 포즈 인식

최근 딥러닝 기술의 발달로 이를 응용한 연구사례가 급증하고 있다[7~8]. 다양한 사례 중 사람의 포즈를 인식하는 것은 사진이나 영상 속에서 사람의 신체 관절이 어떻게 구성되어 있는지 위치를 추정하는 문제이다. 포즈 인식은 사람의 행동 및 상태, 의도 파악을 위한 핵심 요소가 되기 때문에 상당히 중요한 연구 주제이다[9].

MediaPipe는 구글에서 제공하는 AI 프레임워크로서 비디오 형식 데이터를 이용한 다양한 비전 AI 기능을 파이프라인 형태로 손쉽게 사용할 수 있도록 도와준다. 해당 솔루션 중 하나인 MediaPipe Pose는 BlazePose[10]를 기반으로 구현된 것으로 사람의 포즈 인식에 유용하게 사용할 수 있다. BlazePose는 각 신체 관절을 추정하고, 사람의 신체 관절을 그룹화하는 상향식 접근방법을 사용하는 OpenPose[11]를 기반으로 실시간 포즈 인식을 위해 (그림 1)의 2단계의 포즈 탐지기(Pose Detector)와 포즈 추적기(Pose Tracker)로 구성된 머신러닝 파이프라인을 사용한다. 포즈 탐지기는 관절 좌표, 현재 프레임의 사람 존재 여부, 영상 상에서 관심 영역(ROI) 등을 예측한다. 포즈 탐지기가 사람이 없다고 판단하면 다음 프레임에서 탐지기 네트워크를 다시 실행한다. 포즈 탐지기는 많은 객체 탐지에서 사용하는 NMS(Non Maximum Suppression) 알고리즘을 사용하는데 이 알고리즘은 단단한 물체는 잘 탐지(detect)하지만 사람의 포즈를 탐지하는 상황에서는 문제가 발생한다. 이러한 한계를 극복하기 위해 사람의 얼굴과 같이 상대적으로 단단한 신체 부위의 바운딩 박스(Bounding Box)를 탐지하는데 목표를 둔다. 즉, 사람의 얼굴을 빠르게 탐지하여 이를 토대로 몸통의 위치를 예측한다. 이후 포즈 추적기는 ROI에서 33개의 포즈 키포인트를 모두 예측한다.



(그림 1) 사람의 포즈 인식 파이프라인 개요[10]

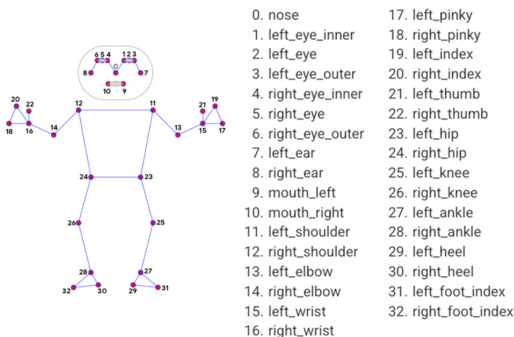
<표 1>은 BlazePose와 OpenPose 간의 성능을 비교한 결과이다. 성능 평가 지표로 관절 점의 추정 좌표와 정답 좌표의 거리가 임곗값보다 작다면 그

관절 점이 옳다고 판단하는 PCK(Percentage of Correct Key points)와 실시간 처리 성능을 보여주는 FPS를 사용한다. BlazePose의 경우 OpenPose에 비해 PCK 수치가 큰 차이를 보이지 않으면서 FPS가 더 높게 나타나 실시간 처리 성능이 더 좋다는 것을 확인할 수 있다

〈표 1〉 BlazePose vs OpenPose

Model	FPS	AR Dataset, PCK@0.2	Yoga Dataset, PCK@0.2
OpenPose (body only)	0.4	87.8	83.4
BlazePose Full	10	84.1	84.5

(그림 2)는 MediaPipe Pose Landmark Model에 대한 그림이다. 총 33개의 관절로 구성되어 있으며 각 관절에 대한 이름과 위치를 보여준다.



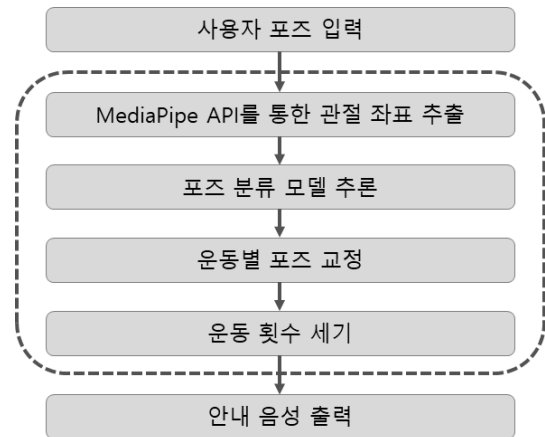
(그림 2) MediaPipe Pose Landmark Model[10]

본 논문에서는 MediaPipe Pose를 통해 추출된 관절 좌표와 관절 선, 관절 좌표의 신뢰도 값을 기반으로 사용자가 운동을 효율적으로 수행할 수 있도록 하는 딥러닝 기반 홈 트레이닝 시스템을 제안한다.

3. 딥러닝 기반 포즈 인식 및 교정을 통한 효율적인 홈 트레이닝 시스템

본 논문에서 제안하는 딥러닝 기반 홈 트레이닝

시스템의 처리 과정은 (그림 3)과 같다. PC나 모바일 기기에 연결된 웹캠을 통하여 사용자의 포즈를 입력받으면 MediaPipe API의 포즈 인식을 통하여 관절 좌표를 추출한다. 이는 포즈 분류 모델의 입력으로 사용하여 포즈 분류 모델 추론을 통해 UP, DOWN, NOTHING 3가지 포즈로 분류된다. 각 운동별 포즈 교정 알고리즘을 통하여 운동에 올바른 포즈인지 판단한 후 올바른 포즈일 때 음성 출력과 함께 운동 횟수를 세어주고, 올바르지 못한 포즈라고 판단되면 안내 음성이 출력되어 사용자가 효율적으로 운동을 수행할 수 있도록 도와준다.



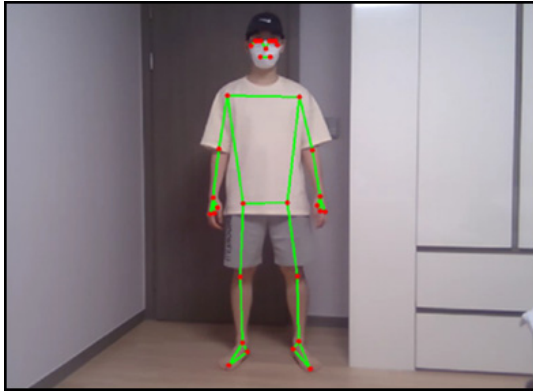
(그림 3) 홈 트레이닝 시스템 처리 과정

3.1 포즈 분류 모델 추론

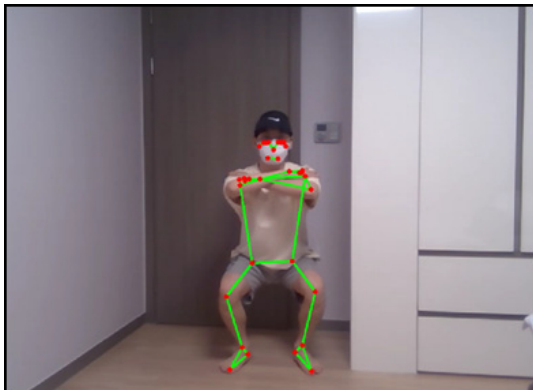
학습 데이터는 사전에 수집된 관절 좌표 정보들로 구성된다. 스쿼트의 포즈 분류 모델은 카메라 정면을 바라보고 운동하는 것을 기준으로 UP, DOWN, NOTHING 3가지로 포즈를 분류한다. 푸쉬업의 포즈 분류 모델은 카메라가 측면을 바라보고 운동하는 것을 기준으로 분류하기 때문에 푸쉬업 하나의 운동에 좌우가 상반된 2가지 모델이 존재한다. 푸쉬업도 스쿼트와 마찬가지로 두 모델 각각 UP, DOWN, NOTHING 3가지로 포즈를 분류한다.

(그림 4~7)은 포즈 분류 모델의 추론을 통한 사용자의 포즈 분류 예시를 보여준다. (그림 8~9)는 정확도를 높이기 위해 구성한 NOTHING에 대한 데

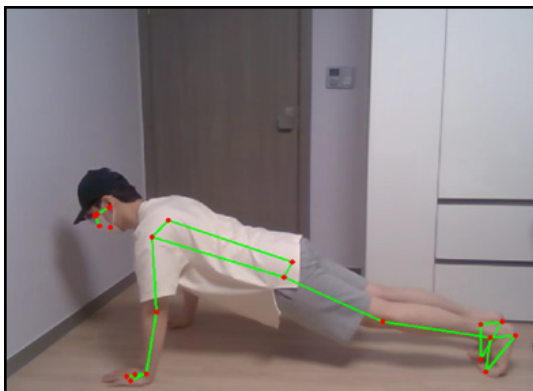
이터의 예시이다.



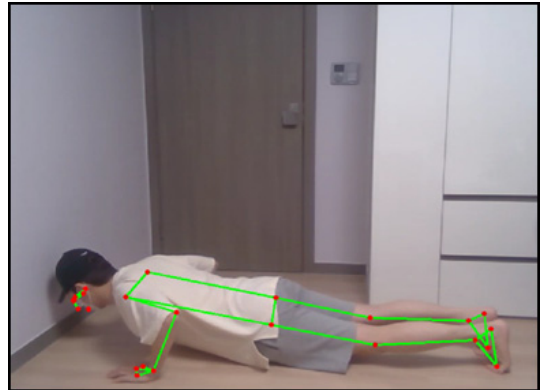
(그림 4) 스쿼트 포즈 분류 - UP



(그림 5) 스쿼트 포즈 분류 - DOWN



(그림 6) 푸쉬업 포즈 분류 - UP



(그림 7) 푸쉬업 포즈 분류 - DOWN



(그림 8) 스쿼트 포즈 분류 - NOTHING



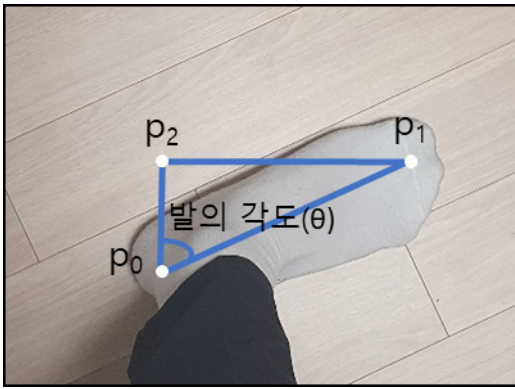
(그림 9) 푸쉬업 포즈 분류 - NOTHING

3.2 운동별 포즈 교정 방법

• 스쿼트 포즈 교정

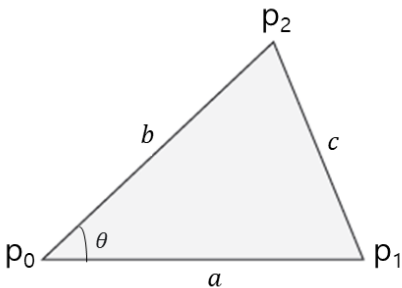
사용자가 스쿼트를 수행할 때 올바른 포즈라고 판단하는 기준은 발의 각도와 발의 위치이다. 발의

각도는 발뒤꿈치만 땅에 붙인 채로 발을 좌우로 움직일 때 변하는 각도이다. 발의 각도를 구하기 위해서 사용하는 관절 좌표는 (그림 2)를 기준으로 왼발은 29, 31의 관절 좌표를 오른발은 30, 32의 관절 좌표를 사용한다. 각도를 구하는 방법은 먼저, (그림 10)와 같이 발뒤꿈치 관절 점(P0), 발가락 관절 점(P1), (P0의 x 좌표, P1의 y 좌표)에 해당하는 관절 점(P2)을 이은 삼각형을 생성한다.



(그림 10) 발의 각도 예시

그 다음 단계로 삼각형의 각 관절 점 사이의 거리를 구한 다음 식 (1)을 적용하여 해당 관절 점의 각도를 구한다 (그림 11). 이때, 발의 각도가 기준 범위 안에서 움직이면 올바른 포즈라고 판단한다.



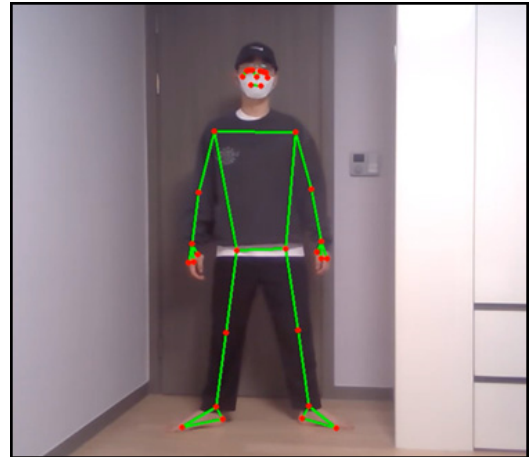
$$P_0 \text{각도}(\theta) = \arccos\left(\frac{a^2 + b^2 - c^2}{2ab}\right) \cdot \frac{180}{\pi} \quad \text{식(1)}$$

(그림 11) 삼각형 변의 길이와 각도의 관계

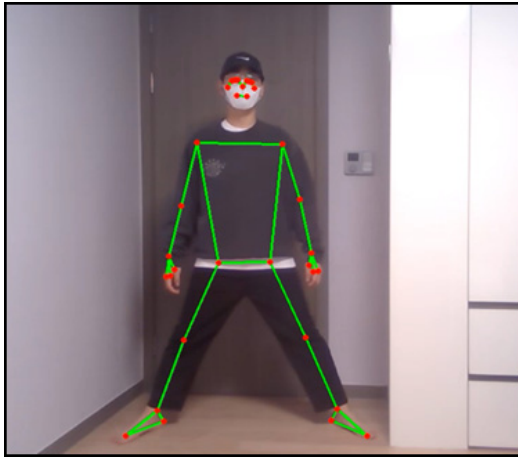
발의 위치는 사용자의 어깨를 기준으로 발과 어깨 사이의 거리를 구하여 발의 벌려진 정도를 인식한다. 마찬가지로 기준 범위 내에서 움직일 때 올바른 포즈로 판단한다. 발의 상대적인 위치는 식 (2)를 통해 구한다.

$$\text{발의 상대적인 위치} = |\text{발목의 좌표 } x - \text{어깨의 좌표 } x| \quad \text{식(2)}$$

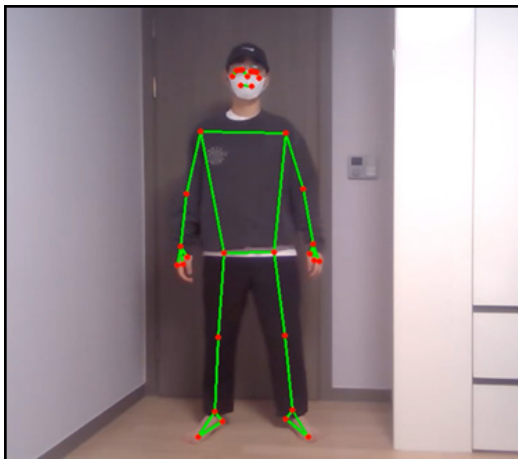
(그림 12~14)는 스쿼트의 포즈 교정 예시를 보여준다. 그림의 경우 발의 각도가 기준치보다 크거나(그림 12), 발의 위치가 어깨너비보다 많이 벌어져 있어(그림 13) 잘못된 포즈로 판단된 경우이다. 포즈 교정을 위한 기준치는 트레이너 비디오의 관절 좌표를 분석하여 발의 각도와 발의 위치의 임계값을 결정하였다.



(그림 12) 스쿼트 포즈 교정 - 잘못된 발의 각도



(그림 13) 스쿼트 포즈 교정 - 잘못된 발의 위치

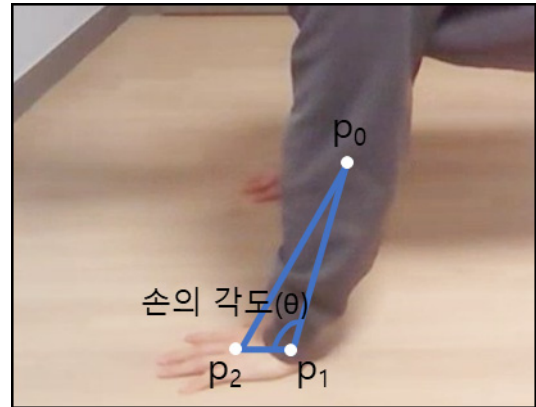


(그림 14) 스쿼트 포즈 교정 - 올바른 자세

• 푸쉬업 포즈 교정

푸쉬업을 수행할 때 올바른 포즈라고 판단하는 기준은 손의 각도와 엉덩이의 위치이다. 손의 각도는 푸쉬업 준비 포즈에서 손이 바라보는 방향을 움직이면 변하는 각도를 말한다. 손의 각도를 구하기 위해서 사용하는 관절 좌표는 (그림 2)를 기준으로 왼손은 13, 15, 19의 관절 좌표를, 오른손은 14, 16, 20의 관절 좌표를 사용한다. 각도를 구하는 방법은 팔꿈치 관절 점(P0), 손목 관절 점(P1), 손가락(중지) 관절 점(P2)을 이은 삼각형을 생성한 다음 스쿼트의 경우와 같은 방법으로 손목 관절 점

(P1)의 각도를 구한다(그림 15).



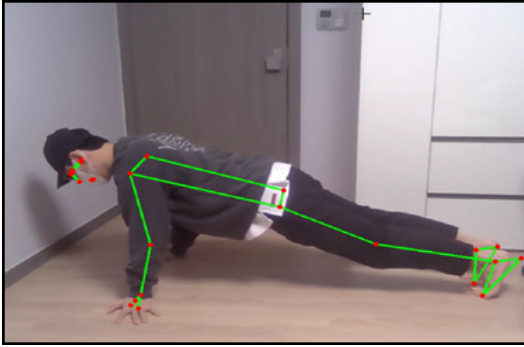
(그림 15) 손의 각도 예시

유효한 손의 각도 범위를 정해놓고 해당 범위 안에서만 사용자의 손의 각도가 움직일 때와 손가락에 해당하는 3개의 관절 좌표가 몸을 기준으로 손목의 관절 좌표보다 앞쪽에 있을 때만 올바른 포즈라고 판단한다.

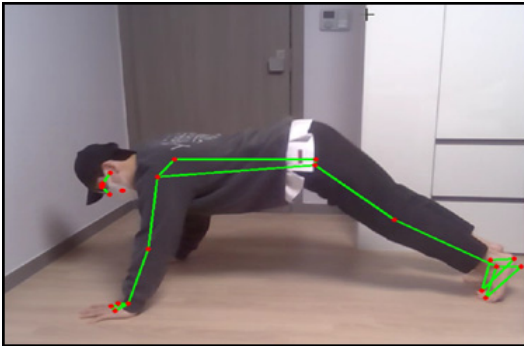
엉덩이의 위치는 식 (3)을 통해 엉덩이의 위치를 구하고 해당 수치의 유효한 범위를 기준으로 두어 엉덩이가 너무 높거나 낮은 잘못된 포즈를 교정할 수 있도록 한다. 즉, 어깨보다 엉덩이가 아래쪽에 있을 때만 유효한 범위인지 확인하고 그렇지 않으면 잘못된 포즈라고 판단한다.

엉덩이의 상대적인 위치
 $= |\text{엉덩이의 좌표 } y - \text{어깨의 좌표 } y|$ - 식(3)

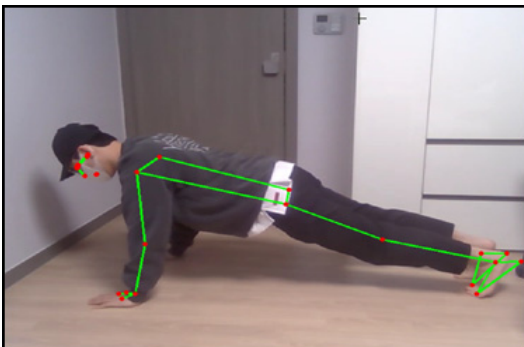
(그림 16~18)은 푸쉬업 포즈 교정 예시를 보여준다. 그림에서 잘못된 예는 손의 방향이 정면을 향하지 않고 측면을 향하는 포즈인 경우(그림 16)과 엉덩이의 위치가 너무 올라가 있는 경우(그림 17)이다.



(그림 16) 푸쉬업 포즈 교정 - 잘못된 손의 각도



(그림 17) 푸쉬업 포즈 교정 - 잘못된 엉덩이의 위치



(그림 18) 푸쉬업 포즈 교정 - 올바른 자세







이와 같이 포즈 분류 모델 추론 결과와 각 운동별 포즈 교정 기준에 해당하는 정보들을 실시간으로 추적하여 유효한 기준 범위를 벗어나게 되면 TTS (Text To Speech) 안내 음성을 출력하여 사용자가 올바른 포즈로 교정할 수 있도록 한다.

3.3 운동 횟수 세기

운동 횟수를 세기 위한 정보는 다음과 같다. 스쿼트 운동은 (골반, 무릎, 발목) 관절 점을 잇는 다리의 각도를 구하고, 푸쉬업 운동은 (어깨, 팔꿈치, 손목) 관절 점을 잇는 팔의 각도를 구한다. 트레이너 비디오를 분석하여 UP, DOWN을 위한 최대, 최소 각도의 평균을 추출하고 이를 토대로 각도 기준치를 둔다. 포즈 분류 모델을 통해 DOWN 포즈가 되고 스쿼트는 다리의 각도, 푸쉬업은 팔의 각도가 정해진 각도 기준치를 만족하면 세기 준비가 완료된다. 이 상태에서 UP 포즈로 변하고 핵심 부위의 각도가 각도 기준치를 만족하면 횟수가 증가 된다.

<표 2>는 운동 횟수 세기에 대한 예시이다. 초기 포즈 분류는 UP이고(표 2의 왼쪽 그림의 예) 이후 포즈 분류가 DOWN으로 변하고, 운동별 DOWN 각도 기준치를 만족하면 카운트 준비가 True로 변한다(표 2의 가운데 그림의 예). 이 상태에서 포즈 분류가 UP으로 변하고, UP 각도 기준치를 만족하면 운동 횟수가 1만큼 증가한다(표 2의 오른쪽 그림의 예).

〈표 2〉 운동 횟수 세기 예시

스쿼트			
푸쉬업			
포즈 분류	UP	DOWN	UP
각도 기준치		110°	150°
카운트 준비	False	True	False
운동 횟수	0	0	1

4. 실험

포즈 분류 모델을 구현하기 위해 웹캠과 트레이너 비디오를 통해 신체의 관절 좌표 정보를 수집하였고, 실험 환경 및 성능 분석 결과는 다음과 같다.

4.1 학습 데이터 세트

필요한 특정 관절 좌표를 수집하여 운동별 포즈 분류 모델의 학습 데이터를 생성한다. 학습 데이터는 트레이너 비디오 데이터를 통해 관절 좌표를 생성하여 전체적인 포즈 분류 모델의 틀을 잡고 더 세밀하게 포즈를 분류하기 위해 직접 해당 포즈를 취해 학습 데이터를 생성하였다. 각 운동별 학습에 필요한 관절 좌표를 각각 정의하고 MediaPipe Pose API를 통해 관절 좌표 정보를 가져오는데 잘못된 관절 좌표가 저장되는 문제를 방지하기 위해 운동별 포즈 분류 모델에 사용된 모든 특징점을 비교하여 신뢰도 값이 80% 이상인 좌표만 가져와 학습 데이터를 구성하였고, 이를 통해 포즈 분류 모델의 정확도를 개선하였다.

〈표 3〉은 각 운동별 학습 데이터 세트와 특징점 수를 나타낸다. 스쿼트의 특징점은 (그림 2)를 기준으로 (0, 11, 12, 13, 14, 15, 16, 23, 24, 25, 26, 27, 28)로 정의하였다. 스쿼트 포즈 분류 모델에 사용된 데이터 세트는 정면을 바라보고 스쿼트를 수행하는 5명의 트레이너 비디오와 웹캠을 통해 직접 스쿼트 포즈를 취하여 총 4,564개의 인스턴스(instances)를 수집하였다. 푸쉬업은 측면을 바라보고 수행하기 때문에 좌우 2개의 모델을 각각 생성하였는데 웹캠으로부터 가까운 팔의 위치를 기준으로 왼팔이 보이는 경우는 푸쉬업(L) 오른팔이 보이는 경우는 푸쉬업(R)로 정의한다. 푸쉬업(L)의 특징점은 (그림 2)를 기준으로 (0, 11, 12, 13, 15, 23, 24, 25, 27)로 정의하였고, 푸쉬업(R)의 특징점은 (0, 11, 12, 14, 16, 23, 24, 26, 28)로 정의 하였다. 푸쉬업(L)과 푸쉬업(R)의 특징점이 상이한 이유는 푸쉬업은 특정 신체가 가려질 수 있기 때문에 데이터 수집 기준에 해당하는 신뢰도 값을 잘 만족하지 못하고, 가려진 신체의 관절 좌표는 잘못된 좌표를 가질 확률이 높기 때문에 포즈 분류 모델의 정확도 개선을 위해 특정한 특징점을 제외하였다. 푸쉬업은 4명의 트레이너 비디오와 웹캠을 통해 직접 푸쉬업 포즈를 취하여 푸쉬업(L)은 6,755개, 푸쉬업(R)은 6,761개의 인스턴스를 각각 수집하였다.

〈표 3〉 학습 데이터 세트 및 특징점 개수

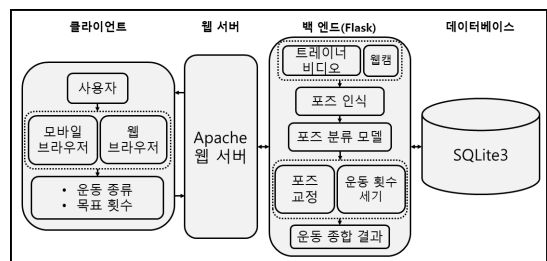
운동종류	UP	DOWN	NOTHING	특징점(x,y)
스쿼트	2,653	1,111	800	26
푸쉬업(L)	3,108	1,747	1,900	18
푸쉬업(R)	3,173	1,888	1,700	18

4.2 실험 환경

본 실험은 클라이언트로 Windows 10 Pro OS와 프로세서 AMD Ryzen 5 3500U, 그래픽 카드 Radeon Vega 8 Graphics, 12GB RAM으로 구성된 노트북을 사용하고, 웹 서버로 프로세서 Intel Core i5 6400, 그래픽 카드 GeForce GTX 1050, 8GB RAM으로 구성된 PC를 사용하여 Python 3.7, Tensorflow 2.3.0 버전에서 실험되었다.

(그림 19)는 웹 기반으로 구현된 홈 트레이닝 시스템 구성도이다. 사용자가 모바일 또는 PC로 웹 브라우저에 접속하고 운동 종류와 목표 횟수를 선택하면 선택한 운동과 횟수에 맞는 트레이너 비디오를 제공하고 사용자가 트레이너의 포즈를 실시간으로 따라 하면서 포즈 교정과 운동 횟수 세기 등의 안내 음성을 통해 효율적으로 운동을 수행할 수 있도록 도와준다.

운동이 종료되면 사용자에게 날짜, 운동 종류, 달성 횟수 및 목표 횟수, 수행 시간 등으로 구성된 종합 결과를 보여주고, 결과를 사용자가 저장하면 SQLite3[12] 데이터베이스에 저장되어 과거에 수행된 운동 결과들을 확인할 수 있도록 한다.



(그림 19) 홈 트레이닝 시스템 구성도

4.3 성능 분석

본 실험에서 사용된 스쿼트, 푸쉬업(L), 푸쉬업(R) 학습 데이터는 10%의 비율로 테스트데이터 세트를 구성하여 운동별 포즈 분류 모델의 성능을 분석하였다. 실험결과 <표 4>와 같이 스쿼트는 99.78%, 푸쉬업(L)은 98.96%, 푸쉬업(R)은 98.96%의 정확도를 보였다.

<표 4> 테스트데이터 포즈 분류 정확도

운동종류	UP	DOWN	NOTHING	총합	정확도
스쿼트	265	111	80	456	99.78%
푸쉬업(L)	310	174	190	674	98.96%
푸쉬업(R)	317	188	170	675	98.96%

또한 학습 데이터와 관련이 없는 임의의 여러 트레이너 비디오를 수집하여 UP, DOWN 데이터 세트를 수집하였고 해당 데이터 세트를 통해 포즈 분류 모델의 추론 정확도를 산출하였다. <표 5>와 같이 스쿼트는 96.03%, 푸쉬업(L)은 95.12%, 푸쉬업(R)은 95.62%의 정확도를 보였다.

<표 5> 비디오데이터 포즈 분류 정확도

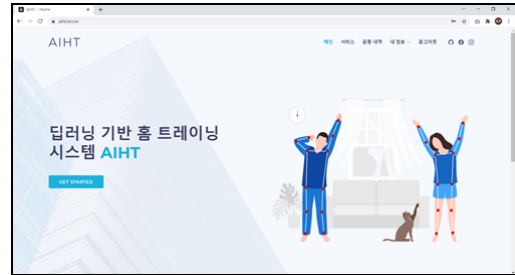
운동종류	UP	DOWN	총합	정확도
스쿼트	940	1,230	2,170	96.03%
푸쉬업(L)	1,447	872	2,319	95.12%
푸쉬업(R)	1,487	983	2,470	95.62%

포즈 분류 모델의 추론 속도는 사용자의 포즈를 실시간으로 분석하기 위한 중요한 지표이다. 본 실험에 사용된 PC로 측정된 결과 스쿼트는 평균적으로 31.5ms가 소요되고 푸쉬업은 평균적으로 31.7ms가 소요되어 실시간으로 활용 가능성을 알 수 있었다.

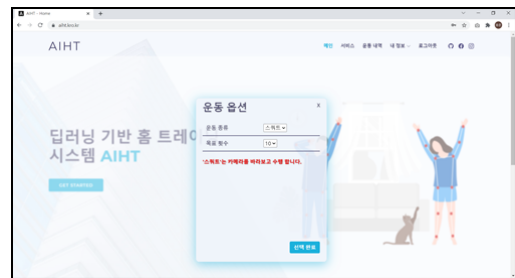
다양한 신체를 가진 5명의 20대 학생을 대상으로 본 논문에서 제시된 홈 트레이닝 시스템을 테스트한 결과 전반적으로 사용이 편리하고 운동 결과

가 만족스럽다는 평가와 운동에 효과적일 것 같다는 답변이 있었다.

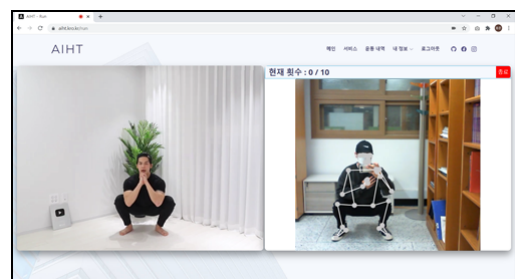
(그림 20~22)는 본 시스템의 실행화면을 보여준다(<https://www.aiht.kro.kr>).



(그림 20) 홈 트레이닝 시스템 - 메인 화면



(그림 21) 홈 트레이닝 시스템 - 운동 선택 화면



(그림 22) 홈 트레이닝 시스템 - 운동 수행 화면

5. 결론

본 논문은 학습에 기반한 포즈 분류 모델을 통해 사용자의 포즈를 실시간으로 분류하고, 사람의 관절 점을 기반으로 포즈를 분석하여 사용자가 효율적으로 운동을 할 수 있도록 도와주는 홈 트레이닝

시스템을 제안한다. 본 시스템은 사용자가 PC나 모바일 기기에서 보이는 트레이너의 자세를 따라 하면서 운동을 할 수 있고 잘못된 포즈에 대한 교정을 받거나 올바른 포즈에 대해서 운동 횟수를 세어주는 등의 기능을 제공한다. 운동별 포즈 분류 모델의 추론 시간을 측정한 결과 평균 31ms의 속도와 평균 95% 이상의 정확도를 보임을 알 수 있었다. 향후 연구로 2D 포즈 추정이 아닌 3D 포즈 추정 기법을 적용하여 보다 정밀한 홈 트레이닝 시스템을 구현할 예정이다.

■ 감사의 글

본 연구는 2020학년도 서경대학교 교내연구비 지원에 의하여 이루어졌음.

■ 참고문헌

- [1] <https://www.consumernews.co.kr/news/articleView.html?idxno=608514>
- [2] <https://www.kakaoht.com/>
- [3] S. Chen, R. Yang, "Pose Trainer: Correcting Exercise Posture using Pose Estimation," arXiv preprint arXiv:2006.11718, 2020.
- [4] 최지운, 남현우, 정진호, "OpenVINO와 라즈베리파이 기반의 Yoga Pose Estimation 구현," 융복합지식학회 논문지, pp.85-93, 2021.
- [5] 하태용, 이후진, "인공지능 기반의 스마트 헬스케어 운동관리를 위한 애플리케이션 구현," 전자공학회 논문지, pp.44-51, 2020.
- [6] 신윤지, 이현주, 김준희, 권다영, 이선애, 추운진, 박지혜, 정자현, 이형석, 김준호, "딥러닝 기반 영상처리 기법 및 표준 운동 프로그램을 활용한 비대면 온라인 홈트레이닝 어플리케이션 연구," 국제문화기술진흥원 논문지, pp.577-582, 2021.
- [7] 김태일, 고영진, 김태영, "모바일 기반 Air Writing을 위한 객체 탐지 및 광학 문자 인식 방법," 한국차세대컴퓨팅학회 논문지, pp.53-63, 2019.

- [8] 오동한, 이병희, 김태영, "외부 환경에 강인한 딥러닝 기반 손 제스처 인식," 한국차세대컴퓨팅학회 논문지, pp.31-39, 2018.
- [9] 조정찬, "단일 이미지에 기반을 둔 사람의 포즈 추정에 대한 연구 동향," 한국차세대컴퓨팅학회 논문지, pp.31-42, 2019.
- [10] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, M. Grundmann, "BlazePose: On-device Real-time Body Pose tracking," arXiv preprint, arXiv:2006.10204v1, 2020.
- [11] Z. Cao, G. Martinez Hidalgo, T. Simon, SE. Wei, YA. Sheikh, "OpenPose: Realtime multi-person 2d pose estimation using part affinity fields," IEEE transactions on pattern analysis and machine intelligence, pp.172-186, 2019.
- [12] <https://www.sqlite.org/version3.html>

■ 저자소개

◆ 이용준



- 2016년 3월~현재 서경대학교 컴퓨터 공학과 학사 재학
- 관심 분야: 머신러닝, 컴퓨터 비전, 웹 백엔드

◆ 김태영



- 1991년 2월 이화여자대학교 전자계산학과 학사
- 1993년 2월 이화여자대학교 전자계산학과 석사
- 1993년 3월~2002년 2월 한국통신 멀티미디어연구소 선임연구원
- 2001년 8월 서울대학교 전기컴퓨터공학부 박사
- 2002년 3월~현재 서경대학교 컴퓨터 공학과 교수
- 관심 분야: 딥러닝, 증강현실, 영상처리, 모바일 3D