

Deep Learning-Based Outlier Detection and Correction for 3D Pose Estimation

Chan-Yang Ju[†] · Ji-Sung Park^{††} · Dong-Ho Lee^{†††}

ABSTRACT

In this paper, we propose a method to improve the accuracy of 3D human pose estimation model in various move motions. Existing human pose estimation models have some problems of jitter, inversion, swap, miss that cause miss coordinates when estimating human poses. These problems cause low accuracy of pose estimation models to detect exact coordinates of human poses. We propose a method that consists of detection and correction methods to handle with these problems. Deep learning-based outlier detection method detects outlier of human pose coordinates in move motion effectively and rule-based correction method corrects the outlier according to a simple rule. We have shown that the proposed method is effective in various motions with the experiments using 2D golf swing motion data and have shown the possibility of expansion from 2D to 3D coordinates.

Keywords : Human Pose Estimation, Pose Refinement, Deep Learning

3차원 자세 추정을 위한 딥러닝 기반 이상치 검출 및 보정 기법

주 찬 양[†] · 박 지 성^{††} · 이 동 호^{†††}

요 약

본 논문에서는 다양한 운동 모션에서 3차원 사람 자세 추정 모델의 정확도를 향상하는 방법을 제안한다. 기존의 사람 자세 추정 모델은 사람의 자세를 추정할 때 좌표 오차를 유발하는 흔들림, 반전, 교환, 오검출 등의 문제가 발생한다. 이러한 문제는 사람 자세 추정 모델의 정확한 자세 추정을 어렵게 한다. 이를 해결하기 위해 본 논문에서는 딥러닝 기반 이상치 검출 및 보정 방법을 제안한다. 딥러닝 기반의 이상치 검출 방법은 여러 모션에서 좌표의 이상치를 효과적으로 검출하고, 모션의 특징을 활용한 규칙 기반 보정 방법을 통해 이상치를 보정한다. 다양한 실험과 분석을 통하여 제안하는 방법이 골프 스윙 모션과 다양한 운동 모션에서도 사람의 자세를 정확히 추정할 수 있고, 3차원 좌표 데이터에서도 확장 가능성을 보인다.

키워드 : 사람 자세 추정, 좌표 보정, 딥러닝

1. 서 론

단일 영상 속 등장하는 인물의 관절 좌표를 추정하는 사람 자세 추정(Human Pose Estimation) 기술의 성장이 가속화되면서 다양한 산업 분야에서 해당 기술을 활용한 사례가 증가하고 있다. 축구, 야구, 농구 등 다양한 운동에서 개인 맞춤형 코칭 서비스를 제공하기 위해 사람 자세 추정 기술을

활용하는 시도들이 나오고 있다. 대표적으로는 골프 산업에서 사람 자세 추정 기술을 통해 골프 코칭 서비스를 제공하려는 시도들도 나오고 있다. 사람 자세 추정 기술의 최근 연구 결과들을 보면 상당히 높은 수준으로 사람의 자세를 검출하는 것을 알 수 있다.

하지만 운동 모션에서 기존에 연구된 자세 추정 모델을 적용했을 때는 여러 문제점이 존재한다. [1]에서는 일반적으로 사람 자세 추정 모델에서 발견되는 문제점을 흔들림(Jitter), 반전(Inversion), 교환(Swap), 오검출(Miss)로 분류하고 이런 문제들이 모델 성능에 미치는 영향에 대해 연구가 진행되었다.

[2]에서는 골프 스윙 모션에서 사람의 자세를 추정할 때, [1]에서 언급한 오검출, 반전, 흔들림, 미검출의 문제를 해결하는 연구를 진행했다. 해당 논문에서는 사람 자세 추정 시 발생하는 이상치(Outlier) 데이터들을 Bi-LSTM(Bidirectional Long Short-Term Memory) 네트워크로 학습하여 이상치를 효과적으로 검출하고 골프 스윙 모션의 특징을 활용하여

※ 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2020-0-01343, 인공지능융합연구센터지원(한양대학교 ERICA).

※ 이 논문은 2021년 한국정보처리학회 ACK 2021의 우수논문으로 "골프 스윙 모션 추정에서 Bi-LSTM 기반의 효율적인 이상치 검출 및 보정 기법"의 제목으로 발표된 논문을 확장한 것임.

† 초 회 원 : 한양대학교 인공지능융합학과 바이오인공지능융합전공 석·박사통합과정

†† 비 회 원 : 한양대학교 인공지능융합학과 바이오인공지능융합전공 박사과정

††† 비 회 원 : 한양대학교 인공지능융합학과 바이오인공지능융합전공 정교수

Manuscript Received : December 24, 2021

First Revision : March 22, 2022

Accepted : March 24, 2022

* Corresponding Author : Dong-Ho Lee(dhlee72@hanyang.ac.kr)

규칙 기반으로 좌표의 이상치를 검출하는 방법을 제안했다.

본 논문에서는 [2]에서 제안한 골프 스윙 모션에서의 좌표 정확도 개선 실험을 확장하여 다양한 운동 모션 데이터에서 이상치 검출 모델을 학습하고 운동 모션 데이터에서 효과적으로 이상치를 검출할 수 있음을 보인다. 또 제안된 방법이 2차원 좌표뿐 아니라 3차원 좌표에도 적용할 수 있음을 보인다. 그리고 검출된 이상치를 제거하고 간단한 선형 보간법(Linear Interpolation)을 통해 결측치를 보간하여 연속된 자세에서 좌표의 정확도를 효과적으로 개선하는 방법을 제안한다.

본 논문의 2장에서는 사람 자세 추정 모델과 관련된 연구를 살펴보고, 3장에서는 제안하는 이상치 검출 및 보정 기법을, 4장에서는 다양한 운동 모션과 3차원 좌표에 제안한 방법을 적용한 결과를 보고, 5장에서 결론 및 향후 연구 방향에 대해 제시한다.

2. 관련 연구

[3]에서는 2차원의 사람 자세 추정 연구 분야를 단일 사람 추정(Single Person Pose Estimation) 모델과 여러 사람 추정(Multi-Person Pose Estimation) 모델로 분류하였다. 또 그 각각을 회귀(Regression) 방식과 히트맵(Heatmap) 방식 그리고 상향식(Bottom-up)과 하향식(Top-down) 검출 방식으로 분류했다[3].

단일 사람 추정 연구의 회귀 방식은 이미지의 특징 맵(Feature Map)에서 직접 관절 좌표를 회귀하는 방식으로 관절을 검출한다. 이 방식은 빠르고 직접적이며 종단 간(End-to-end) 방식으로 훈련된다. 이런 이유로 별도 변화 없이 3차원 관절 추정에 이용될 수 있지만, 관절의 위치를 학습시키는 것이 어렵고, 여러 사람의 자세 추정에는 적용할 수 없다는 단점이 있다. BlazePose[4]는 회귀 방식을 사용한 모델로 추적(Tracking) 방식을 사용하여 이전 좌표 정보를 다음 좌표 정보 예측에 사용한다. 또한 3차원 좌표 추정까지 확장한 모델로 x, y, z의 좌표를 얻을 수 있다. 운동 모션을 입력으로 사용했을 때 화질이 떨어질수록 모든 좌표에서 미검출 문제가 발생하였다.

단일 사람 추정 연구의 히트맵 방식은 관절의 위치로 예상되는 부분의 히트맵 예측을 통해 관절 좌표를 추론하는 방식이다. 이는 시각화를 하기 쉽고, 복잡한 경우에 적용이 가능한 장점이 있다. 하지만 히트맵을 얻기 위해 많은 메모리 소비가 필요하고, 3차원 좌표 추정으로 확장이 어려운 단점이 있다. HRNet[5]은 히트맵 방식을 사용한 모델로 학습에서 다양한 해상도를 병렬적으로 적용하여 전체적인 맥락과 국소적인 맥락을 받아들이는 방법으로 학습된 모델이다. 해당 모델은 실험에 사용한 다른 모델들보다 높은 성능으로 사람의 관절 좌표를 추출했지만, 오검출과 좌표 반전 문제가 여전히 존재하였다.

여러 사람 추정 연구의 상향식 좌표 검출 방식은 영상 속 인물들의 관절 좌표를 추정한 후 사람을 구별하는 방식이다.

좌표를 먼저 찾고 개별 사람을 검출하기 때문에 빠르게 좌표를 검출할 수 있지만, 성능이 떨어진다는 단점이 있다. 상향식 좌표 검출 모델인 OpenPose[6]는 신체 부위의 특징을 통해 사람을 구별해내는 방법을 사용하여 정확도를 향상했다. 골프 스윙 모션 데이터를 사용했을 때 좌표의 미검출 문제가 발생하였다.

여러 사람 추정 연구의 하향식 좌표 검출 방식은 사람을 먼저 검출하고 반복적으로 단일 사람의 관절 좌표를 추정하는 방법이다. 사람을 검출하고 그 후에 반복적으로 사람의 자세를 추정하여 속도는 느리지만 더 정확한 좌표를 검출한다는 장점이 있다. 좌표 추정의 성능을 높이기 위해서는 사람을 탐지하는 부분과 탐지된 사람의 자세를 검출하는 두 부분의 성능이 개선되어야 한다. [7]은 최근 연구된 모델로 STN(Spatial Transformation Network)과 SDTN(Spatial De-transformer Network)를 통해 기하학적, 공간적 변형 과정을 거친다. 이를 통해 고품질의 사람 후보 프레임을 추출하고 인식 성능을 향상하는 특징을 보인다.

사람 자세 추정 모델에서 좌표의 정확도를 개선하는 연구는 종단 간 방법과 후처리(Post-processing) 방법으로 분류된다[8]. 해당 연구는 사람 자세 추정 모델이 학습한 좌표 추정의 정확도를 높이는 연구로, 본 논문에서 증점적으로 다루는 연구 분야이다.

[9-15]는 종단 간 학습 방식으로 연구된 모델이다. 모델별로 구현 방식은 다르지만, 포즈 추정과 개선이 모델 내부에서 함께 진행된다는 특징은 동일하다. [13]의 경우 반복적인 오류 피드백(Error Feedback)을 통해 모델 안에서 단계별로 오류를 전달하여 모델의 추정 결과를 보정하는 과정을 거치며 점진적으로 포즈를 개선해가는 방법을 사용한다. 다른 모델들과 마찬가지로 단계적인 자세 개선 과정들은 모델 내부에 자세 추정 과정과 함께 구현되어 있다. [14]은 내부에서 추정된 자세를 개선하는 PRM(Pose Refine Machine)을 통해 포즈 추정 과정 마지막 단계에서 포즈 개선을 진행한다. 해당 연구에서 PRM은 고수준(High-level)의 판별 의미 정보와 저수준(Low-level)의 공간 정보를 통해 더 정밀한 자세 추정을 가능하게 한다. [15]는 2단계의 과정을 통해 사람의 자세를 추정하는 구조로 되어 있다. 두 번째 단계의 GPR(Graph Pose Refinement) 모듈을 통해 개선된 자세를 얻는 방법을 사용한다. GPR 모듈은 관절 간의 관계를 고려하여 그래프 구조의 개선 모듈로 설계되었다. 이와 같은 종단 간 방식의 개선 모듈은 자세 추정 모델의 추정 결과에 의존하여 출력값을 내고 이는 개선 모듈이 모델 내부에서 성공적으로 작동한다고 보장할 수 없는 결과를 가져오게 된다.

[8, 16, 17]는 성능 개선 연구 중 후처리 방식을 사용한 모델로 사람 자세 추정 모델이 출력한 좌표를 입력 받아서 좌표를 보정하는 네트워크 구조에 대한 연구를 수행했다. [16]와 [8]의 연구는 비슷한 구현 방식을 보이는데, 좀 더 높은 성능을 보인 [8]에서 제안된 방법은 입력 이미지와 좌표를 CNN(Convolutional Neural Network) 네트워크의 Backbone

+ Upsampler 구조를 통과시켜 교정된 좌표를 얻는 방식을 사용한다. [17]에서는 추출된 관절을 PRM(Pose Refinement Module)을 통해 좌표를 보정한 후 이를 활용하여 동작을 인식하는 연구를 수행했다. 해당 모델에서 사용된 PRM은 GCL(Graph Convolution Layer)과 TCL(Temporal Convolution Layer)를 통과하여 보정된 자세를 얻는 구조로 설계되었다. 이러한 후처리 방식은 모델 내부에서 작동되는 종단간 방식과 달리 모델이 좌표를 추정한 후에 적용된다는 점에서 모델 의존도를 낮추었지만, 추가적인 좌표 보정 과정으로 인한 연산량 증가의 문제가 있다.

본 논문에서는 후처리 방식을 사용하여 기존의 연구 [8, 16]에서 제안된 방법보다 더 적은 계산 비용으로 좌표의 정확도를 개선할 수 있는 좌표 이상치 검출 네트워크 및 보간 알고리즘에 대한 연구를 수행한다.

3. Bi-LSTM 네트워크를 활용한 운동 모션에서 관절 좌표 추정의 정확도 개선

본 논문에서는 1개 층의 Bi-LSTM을 활용하여 적은 연산 비용으로 효과적으로 이상치를 제거하는 네트워크를 학습시키는 방법을 제안한다. 이상치를 제거하는 Bi-LSTM 모델은 자세 추정 모델이 자세를 추정하는 과정에서 발생하는 이상치 좌표를 모방한 데이터로 학습하여 각 관절의 프레임별 이상치 발생 유무를 출력한다. 이를 통해 이상치가 발생한 프레임의 관절 좌표를 제거한 후 보간법을 통해 결측된 좌표를 보간한다. 3장에서는 이상치를 검출하는 Bi-LSTM 모델을 학습하는 과정과 결측치 보간 방법을 보이고, 4장에서 2차원 골프 모션 데이터에서의 실험 결과와 다양한 2차원 운동 모션 및 3차원 모션 데이터에 적용한 실험을 통해 제안하는 방법이 관절 좌표 추정의 정확도를 개선하여 자세 추정 모델의 성능을 향상할 수 있음을 보인다.

3.1 이상치 데이터 생성

Fig. 1은 [2]의 골프 스윙 모션에 대한 좌표 정확도 개선 과정을 시각화 한 것이다. 알고리즘은 추정된 좌표의 이상치를 단계별로 제거하고, 보간하는 과정을 포함한다. 본 논문에서는 Fig. 1의 (b)네트워크만을 사용하여 이상치를 검출 및 제거하고 보간 알고리즘(e)을 통해 결측 좌표를 보간한다. (b)네트워크는 Bi-LSTM 모델을 통한 프레임별 이상치 검출 모델로, 자세 추정 모델이 추정된 좌표에서 이상치 데이터의 발생 여부를 출력해준다. 이상치 데이터 수집을 위해 정답 관절 좌표 라벨(Label)이 포함된 데이터를 기반으로 이상치를 모방하는 데이터 생성 기법을 사용하였다. 기존 자세 추정 모델에서 발생하는 이상치는 실제 정답 좌표에서 임의의 x , y 거리만큼 떨어진 위치를 관절 좌표로 추정하는 형태로 발생한다. 이를 모방하기 위해 실제 정답 관절 좌표에서 임의의 x , y 값에 증폭 혹은 감쇠 값을 주어 실제 이상치를 모방한 데이터를 생성한다. 좌푯값의 증폭과 감쇠는 랜덤 변수를 통해 선택된 프레임에 적용하여 모델이 연속된 프레임에서의 변화를 파악할 수 있도록 한다.

3.2 이상치 검출 및 보정 알고리즘

1) Bi-LSTM 기반의 이상 프레임 검출

이상치 검출 모델은 1개 층의 Bi-LSTM 네트워크를 사용하고, 이상치가 발견되면 해당 프레임의 출력값으로 알려주는 역할을 한다. Fig. 2는 4번 관절(오른쪽 손목)의 변화량을 그래프로 시각화 한 것이다. 변화량을 분석한 결과 이상치가 발생할 때는 좌표의 변화량이 이전 프레임들보다 비정상적으로 크게 변하는 것을 확인할 수 있었다. Fig. 3은 골프 스윙 동작에서 부정확한 좌표 추정을 유발하는 동작을 구분 동작으로 분리하여 시각화한 것이다. 동작의 속도가 빠르지 않은 동작의 경우 일정 수준 정확하게 자세를 추정하지만, 동작이 빨라지는 경우 부정확한 좌표 추정 비율이 늘어나는 것을 확

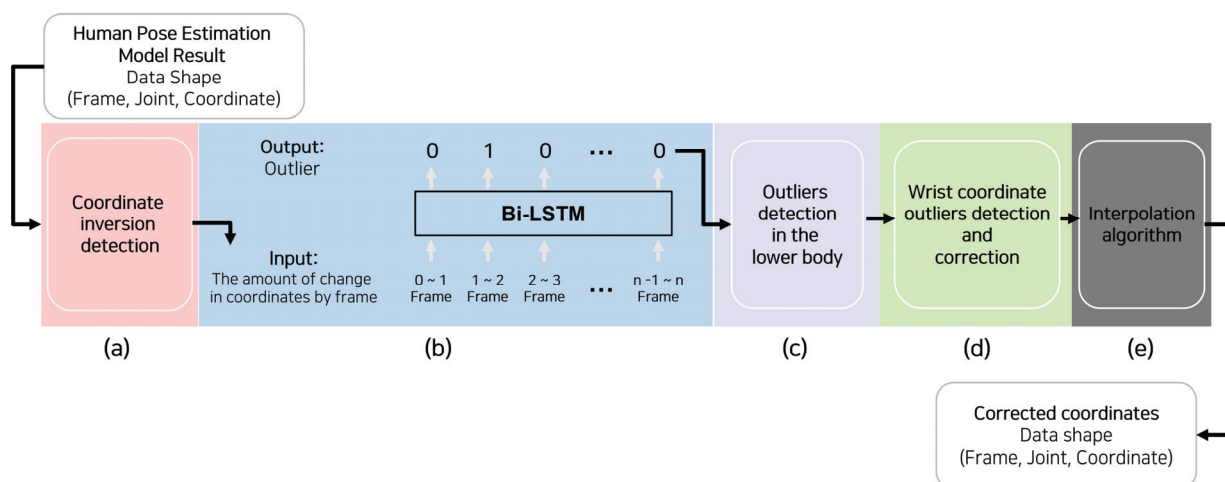


Fig. 1. The Structure of Outlier Detection and Interpolation Process for Golf Swing Motion Estimation (a)~(d) The Process of Outlier Detection and Removing, (e) The Process of Missing Value Interpolation

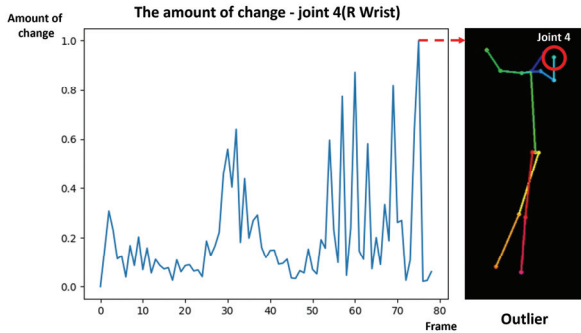


Fig. 2. A Graph of the Amount of Change in Joint 4 when Outliers Occur

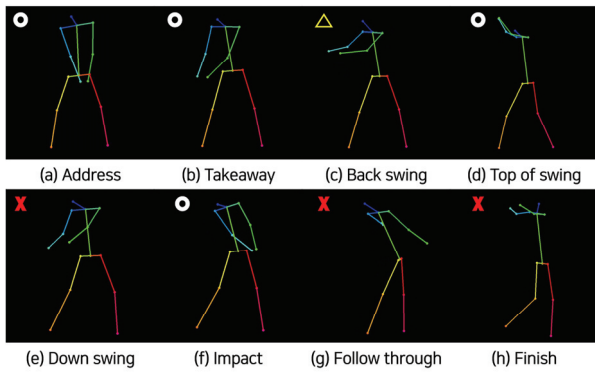


Fig. 3. The Motion Causing Inaccurate Coordinates in Golf Swing Motion

인할 수 있었다. Bi-LSTM 네트워크에 입력으로 들어가는 데이터는 프레임별 좌표 변화량이며, 이때 좌표의 변화량은 유클리드 거리(Euclidean Distance) 계산법으로 프레임 간의 관절별 좌표 거리 변화량을 계산한다. 한 개의 데이터에는 하나의 관절에 대한 변화량이 기록되어 있다. 변화량의 경우 영상 크기에 따라 값의 범위가 달라지므로 각 데이터의 최댓값으로 나눠주어 0과 1 사이의 값으로 정규화(Normalization)한다. 또한 네트워크의 최대 입력 길이를 설정하고 더 짧은 경우에는 0으로 패딩(Padding)값을 준다. 입력 데이터는 프레임별 [좌표 변화량, 해당 관절의 변화량 평균]을 기본 데이터로 실험했고, [관절 정보] 데이터를 추가로 입력하여 정보량에 따른 성능 변화 실험을 진행했다.

2) 결측 좌표 보간

Fig. 1의 (e)는 이상치 검출 모델을 통해 이상치를 제거한 후 결측된 좌표를 보간하는 알고리즘이다. Linear, Next, Previous, Nearest 등의 보간 방법이 있으며 각 방법의 성능 비교를 통해 Linear 방법이 높은 성능을 보이는 것을 알 수 있었다. Linear Interpolation은 두 점을 봤을 때 그 사이의 값인 x에 대한 f(x)의 값을 추정하는 데 사용되는 방법이고 계산식은 Equation (1)과 같다.

$$f(x) = \frac{x - x_1}{x_2 - x_1} (f(x_2) - f(x_1)) + f(x_1) \quad (1)$$

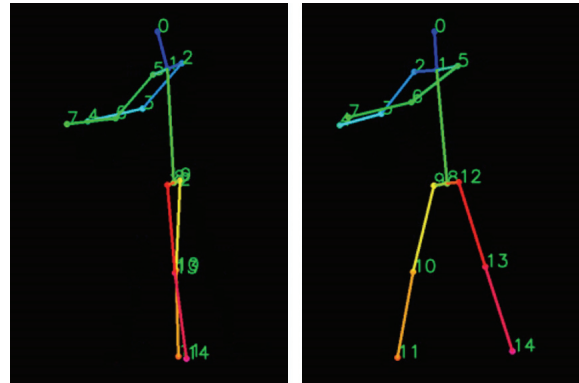


Fig. 4. (Left) Before the Coordinates Correction (Right) After the Coordinates Correction

Fig. 4는 이상치를 제거한 후 보간 알고리즘을 통해 결측치를 보간한 후의 결과를 보여주고 있다. 개선 과정을 통해 반전, 흔들림, 오검출 등의 문제가 개선되었고, 추정된 좌표가 안정적으로 보정된 것을 확인할 수 있다.

4. 실험 및 분석

4.1 골프 스윙 모션에서의 좌표 추정의 정확도 개선 실험

1) 실험 환경

2차원 골프 스윙 좌표 데이터에서의 성능 평가를 위해 총 5개, 483프레임의 관절 라벨이 포함된 골프 스윙 영상에 대해 실험을 진행했다. 실험은 사람 자세 추정 모델이 추정된 좌표 데이터에 논문에서 제안한 방법을 적용한 후 성능 변화를 분석한다. 성능 평가를 위한 지표로는 MPJPE(Mean Per Joint Position Error) 계산식을 응용하여 사용했으며 계산식은 Equation (2)와 같다.

$$MPJPE = \frac{1}{T} \frac{1}{N} \sum_{t=1}^T \sum_{i=1}^N \| (J_i^{(t)} - J_{root}^{(t)}) - (\hat{J}_i^{(t)} - \hat{J}_{root}^{(t)}) \|_2 \quad (2)$$

T는 총 프레임의 수, N은 측정에 사용한 관절의 수이다. MPJPE는 정답 데이터의 중심 관절($J_{root}^{(t)}$)과 측정하려는 관절($J_i^{(t)}$)의 거리와 예측 데이터에서 중심 관절($\hat{J}_{root}^{(t)}$)과 측정하려는 관절($\hat{J}_i^{(t)}$)의 거리의 차이를 계산하여 오류를 측정한다.

2) Bi-LSTM을 이용한 이상 프레임 검출 성능

Table 1은 2차원 골프 모션 데이터에서 실험한 결과이다. Bi-LSTM은 Fig. 1의 (b)네트워크이며, Rule based bias correction은 골프 동작의 규칙을 활용한 이상치 검출 방법이다. 골프 동작의 규칙의 경우 아래 '이상치 검출 규칙에 따른 성능 변화' 항목에 상세 내용을 기술하였다. 비교를 위한 기준 모델은 BlazePose[4]를 사용했고, 각 결과는 Linear Interpolation을 사용하여 보간한 후에 성능을 측정했다.

Table 1. MPJPE Results of Bias Correction Model[2] Each Experimental Method in Golf Swing Motion

Models	MPJPE
	Mean
Base model (BlazePose[4])	22.29
Bi-LSTM	16.81
Bi-LSTM + Rule based bias correction	16.43

실험 결과 기준 모델만 사용하여 좌표를 추정할 때보다 기준모델에 논문에서 제안한 Bi-LSTM 네트워크를 추가로 적용했을 때 좌표의 정확도가 개선되는 것을 알 수 있었다. Bi-LSTM 네트워크 학습의 경우 [2]에서 높은 성능을 보인 실험 환경을 활용하여 10-40px 사이의 좌표값의 증폭 혹은 감쇠 값을 주었고, 1~3프레임 사이의 연속된 프레임에 이상치를 생성했다.

3) 이상치 검출 규칙에 따른 성능 변화

이상치 검출 규칙은 [2]에서 제안한 방법이며, Fig. 1의 (a), (c), (d)에 해당한다. (a)는 좌우 반전 검출 규칙으로 골프 스윙 모션에서 11번 관절(오른쪽 발목)은 12번 관절(왼쪽 발목)보다 작은 x 좌표를 갖는다는 규칙을 활용한 방법이다. 11번 관절이 12번 관절보다 큰 x값을 가지면 좌우를 반전시킨다. (b)는 하체 이상치 검출 방법으로 양쪽 발목 좌표가 일정 거리 이상 좁혀지면 이상치로 판단하고 제거한다. (c)는 손목 이상치 검출 규칙으로 골프 스윙 중 두 손목의 거리가 일정 거리 이상 떨어지면 이상치로 판단하여 제거한다.

Bi-LSTM 모델만 사용했을 때는 16.81의 오차를 보였고, 해당 모델에 골프 동작의 특징을 이용한 규칙을 추가로 적용했을 때는 오차가 16.43까지 개선이 되었다. 최종적으로는 골프 스윙 동작의 특징을 이용한 규칙 모두를 적용했을 때 높은 성능을 보였으며 기준 모델보다 5.86의 오차 감소를 보여주었다.

4) 다른 모델에 적용한 결과

Table 2는 다른 자세 추정 모델에 제안한 방법을 적용한 결과이다. 성능 비교에 사용한 'Bias correction model'은

Table 2. MPJPE Results of the Bias Correction Model[2] on [4-6] in Golf Swing Motion

Models	MPJPE
	Mean
BlazePose[4]	22.29
BlazePose[3] + Bias correction model[2]	16.43
Open Pose[6]	53.60
Open Pose + Bias correction model[2]	14.60
HRNet[5]	12.28
HRNet + Bias correction model[2]	11.98

Table 1에서 가장 높은 성능을 보인 모델이다. 실험 결과를 통해 자세 추정 모델 [4-6]에 제안된 방법을 적용하면 전체적으로 오차가 감소하는 것을 확인할 수 있었다.

4.2 운동 모션 추정에서의 좌표 추정의 정확도 개선 실험

1) 실험 환경

Table 1의 실험 결과를 보면 골프 스윙 모션의 특성을 이용한 규칙을 적용하지 않고, 이상치 검출 모델만으로도 MPJPE가 낮아진 것을 알 수 있었다. 이를 바탕으로 Bi-LSTM 기반 이상치 제거 모델을 추가로 학습시켜 다른 운동 모션에도 적용해 보는 실험을 진행하였다.

실험을 하는 조건은 4.1의 실험에서 성능이 좋았던 학습 방식을 사용했으며, 골프 모션의 특징을 이용한 규칙 (a), (c), (d)는 사용하지 않았다. 이상치를 제거한 후 결측값을 채울 때는 Linear Interpolation을 사용하였다. 모델을 학습하기 위해 Open 3D Human Pose Data Set인 3DPW[18]에서 제공하는 2차원 데이터인 '2d pose' 데이터를 사용했다. 또한 추가적으로 Baseball, Basketball, Dance, Gait, Pilates, Squash 모션을 수집하여 각 관절에 라벨이 달린 데이터를 각각 학습 및 테스트 데이터로 사용했다. 학습에 사용한 데이터는 Open Data Set과 수집한 데이터를 포함하여 21개의 모션, 약 13700프레임이다. 3.1의 데이터 생성 기법을 통해 데이터를 증가시켜서 실험을 진행했다. 본 실험을 통해 위에서 언급한 6개의 운동 모션, 약 1000프레임에 대한 MPJPE를 계산해 보고 다양한 운동 모션에도 제안하는 방법이 효과적임을 보인다.

2) Bi-LSTM을 이용한 이상 프레임 검출 성능

Table 3의 모델 이름은 [학습에 사용한 모션 수(생성한 데이터 개수)], 데이터의 길이 - Bi-LSTM 모델이다. Bi-LSTM 모델은 좀 더 긴 모션에 적용할 수 있게 최대 프레임 길이를 200으로 설정하였으며, 길이가 부족한 경우 0으로 패딩 값을 주었다. 학습에 사용한 데이터는 각 프레임 별로 [좌표 변화량, 변화량의 평균]을 주었다. 골프 스윙에서는 관절 번호를 주는 것이 성능이 더 좋았지만, 일반적인 운동 모션에서는 관절의 움직임의 정도가 각각 다르기 때문에 관절 번호를 주는 것은 모델 학습에 좋은 영향을 주지 못했다. 또한 학습에 사용한 모션 수를 늘려가며 실험한 결과 예측하려는 운동 모션과 관계없는 모션이라도 다양한 모션을 통해 모델을 학습시켰을 때 더 좋은 성능이 나왔다. 최종적으로는 골프 모션이 아닌 다른 운동 모션에서도 좌표 추정의 정확도가 개선되는 것을 확인할 수 있었고, 논문에서 제안된 방법을 사용하기 전에는 평균 50.14의 오차를 보인 모델이 평균 39.09까지 오차 수치가 감소한 것을 확인할 수 있었다.

3) 보간 방법 비교

본 실험에서 시도한 보간 방법은 Linear, Cubic, Nearest, Previous, Next이다. Table 3에서 가장 높은 성능을 보이는

Table 3. MPJPE results of each experimental method in sports motion

Models	MPJPE						
	Baseball	Basketball	Dance	Gait	Pilates	Squash	Mean
Base model (BlazePose[4])	34.24	71.12	89.81	35.85	42.01	27.78	50.14
[5 Motion(102400)], 3 - Model	34.25	55.97	65.70	35.71	26.58	28.04	41.04
[5 Motion(102400)], 2 - Model	34.52	54.80	64.39	35.11	27.29	27.73	40.64
[15 Motion(102400)], 2 - Model	32.37	53.02	64.11	34.85	27.38	27.66	39.90
[15 Motion(204800)], 2 - Model	31.27	52.43	65.25	34.75	25.99	27.67	39.56
[21 Motion(204800)], 2 - Model	30.24	52.48	61.22	35.96	26.73	27.91	39.09

Table 4. MPJPE Results of Each Interpolation Method in Sports Motion

Methods	MPJPE
	Mean
Linear	39.09
Cubic	95.17
Nearest	40.04
Previous	39.87
Next	40.26

모델을 기준으로 실험을 진행했다. Table 4를 보면 가장 높은 성능을 보인 방법은 선형 보간법이다. Cubic 방법의 경우 3차 다항식을 사용하여 값을 보간하는데, 이를 통해 중간 값을 채울 경우 정답과는 전혀 다른 곡선 형태의 함수가 생성되고 오히려 오차를 크게 만들었다. 다른 방법들은 결과가 비슷했지만, Previous와 Next는 단순히 값을 복사한다는 점에서 본 연구의 목적과는 다른 결과를 가져왔고, Nearest 방법은 일정 수준의 성능을 보였으나 Linear Interpolation 방법보다 성능이 떨어졌다.

4.3 3차원 모션에서 좌표 추정의 정확도 개선 실험

1) 실험 환경

4.2의 실험에서 높은 성능을 보인 모델과 유사하게 실험을 진행했다. 학습 데이터는 3DPW[18]에서 제공하는 3차원 좌표 데이터를 사용했으며, 이중 단일 사람이 등장한 14개의 모션을 선별하였고, 10개의 모션은 학습에, 4개의 모션은 테스트에 사용하였다. 데이터 생성 방법은 동일하며 좌표 값의 증폭 혹은 감쇠의 범위는 4.1의 실험에서 성능이 좋았던 10px에서 40px 값을 사용하였고, x, y, z 좌표에 각각 적용했다. 좌표 보간 방법도 Table 4에서 가장 성능이 좋았던 Linear Interpolation을 사용하였다. 성능 테스트에 사용한 모델은 BlazePose[4]로 단일 사람 검출 및 3차원 좌표에 대한 추정이 가능하기 때문에 비교 모델로 선정했다.

2) 3차원 좌표 데이터에서 좌표 추정의 정확도 개선 효과

Table 5는 3차원 좌표 추정의 정확도 개선 실험의 결과이다. 4개의 실험 데이터에 대해 BlazePose[4]가 예측한 관절

Table 5. Results of 3d coordinates using the proposed method

Models	MPJPE
	Mean
BlazePose[4]	159.91
BlazePose[4] + Ours	154.57

의 MPJPE는 159.91이고, 제안된 방법을 적용하면 154.57까지 오차가 감소하여 본 논문에서 제안한 방법이 3차원 데이터에 대해서도 성능 향상에 기여할 수 있다는 것을 보였다.

이상치를 찾는 핵심 방법은 각 관절 좌표의 프레임 별 변화량을 기록하여 이전 기록에 비해 비정상적으로 변화량이 치솟는 좌표를 이상치로 판단하여 제거하는 것이었다. Fig. 5는 3차원 데이터에서 이상치 데이터를 제거하기 전후의 프레임 별 관절 좌표의 변화량을 비교한 그래프이다. 첫 번째 그래프가 0번 관절의 그래프이고, 제일 마지막 그래프가 14번 관절이며, 각 관절 번호에 대한 라벨은 Fig. 6에 나타난 관절 부위와 같다. 각 그래프의 x축은 프레임, y축은 변화량이다. 이상치를 제거한 후 프레임 별 좌표의 변화량이 안정적으로 변한 것을 확인할 수 있었으며, 해당 그래프를 그리는데 사용된 데이터는 MPJPE가 기존 201.74에서 보정 후 183.61까지 낮아진 것을 확인할 수 있었다. 이를 통해 실제 변화량을 안정화 시키는 작업이 좌표 추정의 정확도 향상에 기여할 수 있음을 알 수 있었다.

5. 결론

본 논문에서는 기존의 사람 자세를 추정 모델의 성능을 개선하는 이상치 검출 네트워크 및 보간 알고리즘을 제안하였다. 실험을 통해 사람 자세 추정 모델에 제안된 방법을 적용한 결과 관절 예측의 성능 향상이 이루어짐을 알 수 있었다. 또한 본 논문에서 제안된 이상치 검출 모델이 다양한 2차원 운동 모션 및 3차원 모션 데이터에서 좌표 추정의 정확도를 개선할 수 있다는 것을 확인했다. 또 프레임 별 관절 좌표의 변화량 그래프를 제안한 방법을 적용하기 전후로 비교하여 좌표의 변화량이 좌표 추정의 정확도에 영향을 미침을 보였다. 기존의 사람 자세 추정 모델의 추정 결과를 개선하는 연

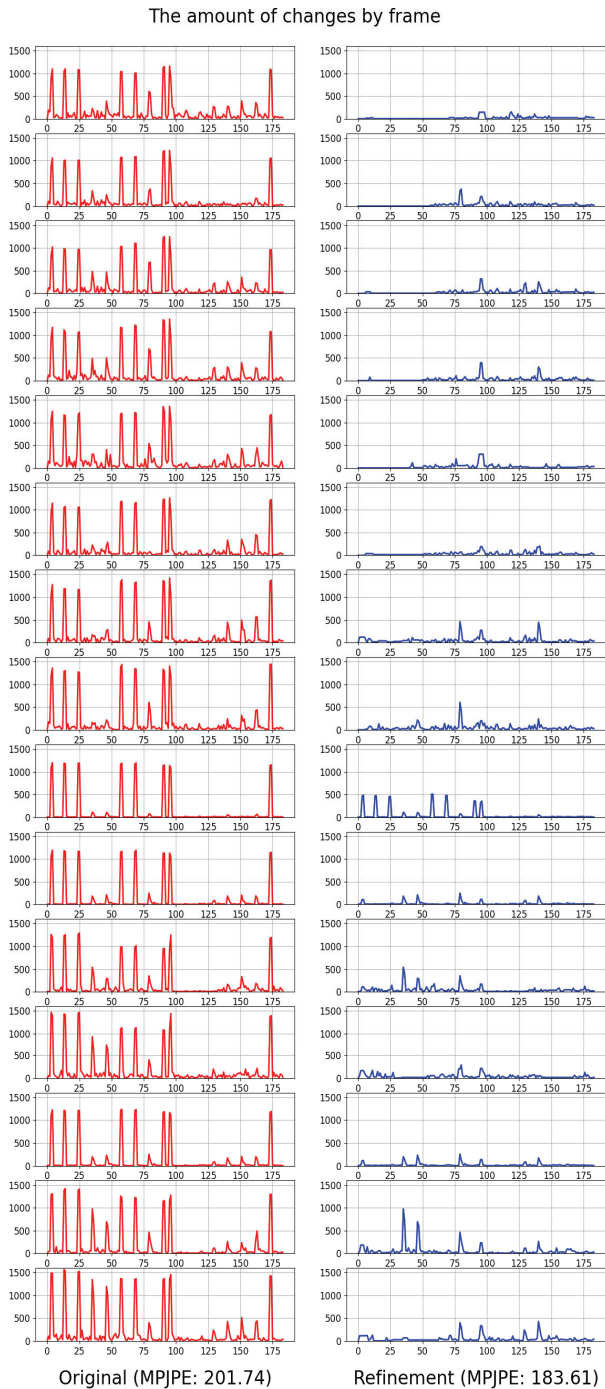


Fig. 5. Graphs Comparing the Amounts of Changes by Frame (Left) Original Results (Right) Refinement Results

구는 이미지를 통한 CNN 기반 좌표 보정 방법으로 3차원 좌표 개선으로 활용하기에는 어려움이 있다. 하지만 본 논문에서 제안하는 방법은 관절 좌표 데이터를 통해 개선이 이루어지므로 3차원 좌표 데이터에 대한 확장이 용이하다. 이를 통해 향후 가상 현실(Metaverse), 골프 레슨, 재활 치료 보조 시스템, 산업 안전 관리 시스템 등 다양한 분야에 적용될 수 있을 것으로 기대한다. 다만 본 논문에서 제안된 모델은

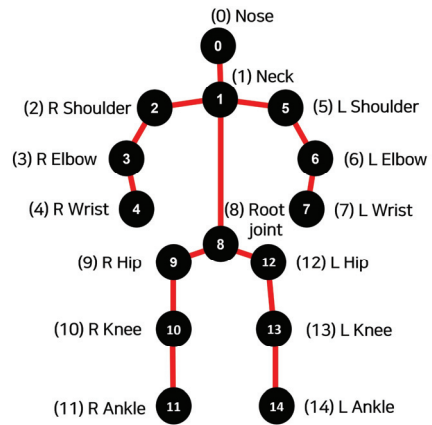


Fig. 6. The Human Skeleton Model Used for Learning

Bi-LSTM 모델의 한계로 사전에 정해진 길이를 초과하는 영상에는 대응하지 못하여 실시간으로 자세 교정이 되지 않는다는 문제점이 있다. 이를 보완하기 위해 향후 연구는 시계열 정보를 실시간으로 분석하여 가변적인 길이에도 대응할 수 있는 모델을 개발하고자 한다.

References

- [1] M. R. Ronchi and P. Perona “Matteo: PERONA, Pietro. Benchmarking and error diagnosis in multi-instance pose estimation,” In *Proceedings of the IEEE International Conference on Computer Vision*, pp.369-378, 2017.
- [2] C. Y. Ju, J. S. Park, G. S. Oh, H. J. Choi, and D. H. Lee. “An efficient Bi-LSTM based method for outlier detection and correction in golf swing motion estimation.” *Proceedings of the Korea Information Processing Society Conference*, pp.787-790, 2021.
- [3] Q. Dang, J. Yin, B. Wang, and W. Zheng, “Deep learning based 2d human pose estimation: A survey,” *Tsinghua Science and Technology*, Vol.24, No.6, pp.663-676, 2019.
- [4] V. Bazarevsky, L. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “BlazePose: On-device real-time body pose tracking,” *arXiv preprint arXiv:2006.10204*, 2020.
- [5] K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.5693-5703, 2019.
- [6] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.43, No.1, pp.172-186, 2019.
- [7] Y. Li, D. Yang, Y. Chen, C. Peng, Z. Sun, and L. Jiao, “A lightweight top-down multi-person pose estimation method based on symmetric transformation and global matching,” *IEEE Access*, Vol.10, pp.22112-22122, 2022.

[8] G. Moon, J. Y. Chang, and K. M. Lee, "Posefix: Model-agnostic general human pose refinement network," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.7773-7781, 2019.

[9] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," In *European Conference on Computer Vision*, Springer, Cham, pp.483-499, 2016.

[10] A. Bulat and G. Tzimiropoulos, "Human pose estimation via convolutional part heatmap regression," In *European Conference on Computer Vision*, Springer, Cham, pp.717-732, 2016.

[11] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp.4724-4732, 2016.

[12] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded pyramid network for multi-person pose estimation," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.7103-7112, 2018.

[13] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.4733-4742, 2016.

[14] Y. Cai et al. "Learning delicate local representations for multi-person pose estimation," In *European Conference on Computer Vision*, pp.455-472, 2020.

[15] J. Wang, X. Long, Y. Gao, E. Ding, and S. Wen, "Graph-pcnn: Two stage human pose estimation with graph pose refinement," In *European Conference on Computer Vision*, pp.492-508, 2020.

[16] M. Fieraru, A. Khoreva, L. Pishchulin, and B. Schiele, "Learning to refine human pose estimation," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.205-214, 2018.

[17] S. Li, J. Yi, Y. A. Farha, and J. Gall, "Pose refinement graph convolutional network for skeleton-based action recognition," *IEEE Robotics and Automation Letters*, Vol.6, No.2, pp.1028-1035, 2021.

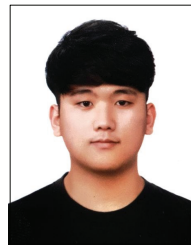
[18] T. V. Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll, "Recovering accurate 3d human pose in the wild using imus and a moving camera," In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp.601-617, 2018.



주 찬 양

<https://orcid.org/0000-0002-5064-802X>
 e-mail : karunogi@hanyang.ac.kr
 2021년 한양대학교 문화콘텐츠학과(학사)
 2021년~현 재 한양대학교 인공지능융합학과
 바이오인공지능융합전공
 석·박사통합과정

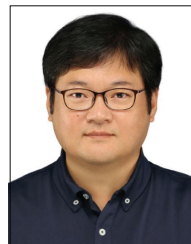
관심분야 : Artificial Intelligence(AI), Explainable AI, Digital Health Care



박 지 성

<https://orcid.org/0000-0001-8947-6387>
 e-mail : jsdms316@gmail.com
 2017년 한양대학교 컴퓨터공학과(학사)
 2020년 한양대학교 컴퓨터공학과(석사)
 2020년~현 재 한양대학교 인공지능융합학과
 바이오인공지능융합전공 박사과정

관심분야 : Natural Language Process, Information Leakage Prevention, Artificial Intelligence(AI)



이 동 호

<https://orcid.org/0000-0003-0305-9182>
 e-mail : dhlee72@hanyang.ac.kr
 1995년 홍익대학교 컴퓨터공학과(학사)
 1997년 서울대학교 전기·컴퓨터공학과
 (석사)
 2001년 서울대학교 전기·컴퓨터공학과(박사)

2004년~현 재 한양대학교 인공지능융합학과 바이오인공지능융합전공
 정교수

관심분야 : Semantic Multimedia, Flash-based System Software, NoSQL and Big Data Processing