골프 스윙 모션 추정에서 Bi-LSTM 기반의 효율적인 이상치 검출 및 보정 기법

주찬양, 박지성, 오경수, 최현준, 이동호
한양대학교 인공지능융합학과 바이오인공지능융합전공
karunogi@hanyang.ac.kr, jsdms316@gmail.com, dhrudtn7187@naver.com, chlqudfud12@naver.com, dhlee72@hanyang.ac.kr

An efficient Bi-LSTM based method for outlier detection and correction in golf swing motion estimation

Chan-Yang Ju, Ji-Sung Park, Gyeong-Su Oh, Hyun-Jun Choi, Dong-Ho Lee
Dept. of Applied Artificial Intelligence, Hanyang University, Major In Bio Artificial Intelligence

요 약

본 논문에서는 최신 모션 인식 기술을 활용하여 골프 스윙 비디오에서 사람의 자세를 추정한 후 다양한 원인으로 오검출된 좌표들을 보정하여 자세 추정의 정확도를 높이는 방법을 제안한다. 기존의 사람 자세 추정 모델은 골프 스윙 데이터에서 오검출, 반전, 불안정성, 미검출의 문제를 보여 정확한 자세 추정을 어렵게 했다. 이를 해결하기 위하여 본 연구에서는 자세 추정시 발생하는 이상치 데이터들을 Bi-LSTM 으로 학습하고 골프 스윙의 특징을 고려한 간단한 규칙을 통하여 이상치 데이터를 효과적으로 검출하고 이를 보정하는 방법을 제안한다. 또한 다양한 실험과 분석을 통하여 제안하는 방법이 골프 스윙 모션에서 사람의 자세를 정확히 추정할 수 있음을 보인다.

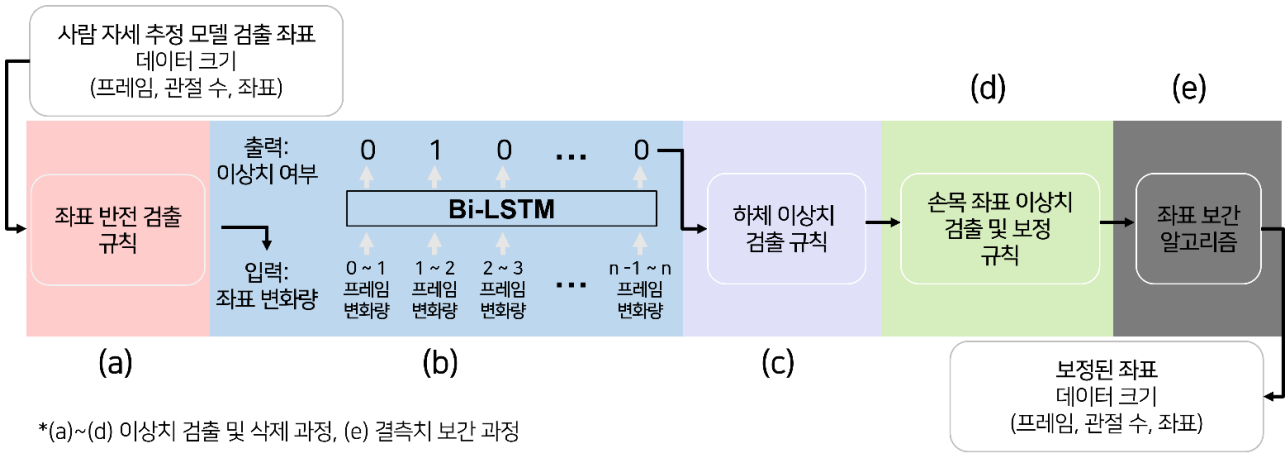
1. 서론

최근 단일 영상에서 사람 자세 추정 기술 성장이 가속화되면서 해당 기술을 다양한 산업 분야에 적용하는 시도가 이어지고 있다. 특히 골프 산업에서 자세 추정 기술을 활용하여 저렴한 비용으로 레슨 서비스를 제공하고자 하는 시도들이 많이 나오고 있다. 사람 자세 추정 기술에 대한 최근 연구 결과들을 살펴보면 높은 성능으로 사람의 자세를 검출하는 것을 확인할 수 있다. 하지만 골프 스윙 영상에서 기존의 자세 추정 모델을 적용했을 때는 여러 문제점이 발생하는 것을 알 수 있었다. [1]에서는 일반적인 사람 자세 추정 모델에서 나타나는 문제점을 불안정, 반전, 교환, 오검출로 분류하고 이런 문제점들이 성능에 미치는 영향에 대해 연구했다. 본 연구에서는 골프 스윙 모션에서 사람의 자세를 추정할 때, [1]에서 언급한 오검출, 반전, 불안정성, 미검출의 문제를 해결하기 위하여 자세 추정시 발생하는 이상치 데이터들을 Bi-LSTM 으로 학습하여 효과적으로 이상치를 검출하는 방법과 골프 스윙 모션만의 독특한 특징을 고려한 간단한 규칙(예, 골프 스윙 중에는 양쪽 손목의 거리가 일정한 특징을 가짐)을 활용하여 이상치를 검출하는 방법을 제안한다. 그리고 이러한 이상치를 제거하고 간단한 선형 보간법을 통하여 연속된 자세의 좌표를 효과적으로 추정하는 방법을 제안한다.

2. 관련 연구

[2]에서는 2 차원의 사람 자세 추정 연구 분야를 단일 사람 추정 모델과 여러 사람 추정 모델로 나눴으며 또 그 각각을 회귀 방식과 히트맵 방식 그리고 상향식과 하향식 검출 방식으로 분류했다.

단일 사람 검출 연구의 회귀 방식은 이미지의 특징맵에서 직접 관절 좌표를 회귀하는 방식으로, 빠르지만 관절의 위치를 학습하는게 어려운 특징이 있다. BlazePose[3]는 추적 방식을 사용한 회귀 모델로 이전 좌표 정보를 다음 좌표 정보 예측에 사용하는 특징을 보이며, 골프 스윙 데이터에 적용했을 때는 미검출 문제가 발생했다. 히트맵 방식은 관절 예상 위치의 히트맵 예측을 통해 관절 좌표를 추론하는 방식이고, 시각화를 하기 쉽지만, 히트맵을 얻기 위해 많은 메모리 소비가 필요하다는 단점이 있다. HRNet[4]은 히트맵 방식을 사용한 모델로 학습에서 다양한 해상도의 맥락을 받아들이는 방법으로 학습된 모델이다. 실험에서 사용한 모델 중에 가장 높은 성능을 보였으나, 좌표 반전이나 오검출 문제가 존재했다.



(그림 1) 골프 스윙 모션 추정을 위한 이상치 검출 및 보간 과정의 전체 구조

여러 사람 검출 연구의 상향식 좌표 검출 방식은 좌표 데이터를 먼저 추정한 뒤 사람을 검출하는 방식이다. 상향식 검출 방법은 빠르게 좌표를 검출할 수 있지만, 성능이 떨어진다는 단점이 있다. OpenPose[5]는 상향식 좌표 검출 모델이며 신체 부위가 어느 사람에게 속하는지 유추할 수 있는 특징을 이용하여 정확도를 높였고, 골프 스윙 데이터를 입력으로 사용했을 때 일부 관절 미검출 문제가 나타났다. 하향식 좌표 검출 방식은 사람을 먼저 검출하고 반복적으로 관절 좌표를 추정하는 점에서 느리지만 좀 더 정확하게 좌표를 검출할 수 있다는 장점이 있다.

사람 자세 추정 모델에서 추정된 좌표를 개선하는 연구는 종단간 학습 방법과 후처리 방법으로 나눠진다.[6]

[7]은 종단간 학습 방식으로 연구된 모델이며 모델마다 모델 안에서 단계별로 모델의 추정 결과를 보정하는 과정을 거치고 최종적으로 합쳐서 좌표를 출력하는 형식이다. 이 방식을 사용하게 되면 개선 모듈은 추정 결과에 의존하여 출력을 내고 이는 모델 개선이 성공적으로 작동한다고 보장할 수 없는 결과를 가져오게 된다.[6]

[6, 8]은 후처리 방식의 성능 개선 모델로 사람 포즈 추정 모델이 출력한 결과를 다시 받아서 보정하는 네트워크 구조를 보인다. 높은 성능을 보인 [6]의 구조를 살펴보면 이미지와 좌표를 CNN 네트워크의 Backbone + Upsampler 구조를 통과시켜 교정된 자세를 얻는 모델이다. 이러한 방식은 모델 추정 결과와 별개로 후 작업을 한다는 점에서 모델 의존도를 낮추고 다양한 방식의 모델에 적용할 수 있는 장점이 있지만, 연산 과정이 늘어나면서 사람 자세 추정 과정의 연산량이 늘어난다는 단점이 있다.

본 논문에서 제안하는 방식은 후처리 방식으로 기존의 연구 [6, 8]에서 사용된 네트워크보다 적은 계산 비용으로 성능을 개선할 수 있는 이상치 검출 네트워크 및 좌표 보정 알고리즘을 구현했다.

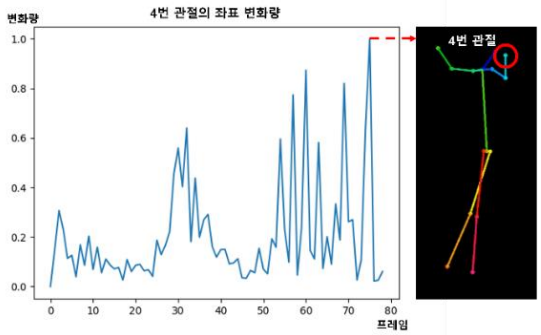
3. Bi-LSTM 및 골프 스윙 모션의 특성을 활용한 관절 좌표 추정의 정확도 개선

3.1 이상치 데이터 생성

그림 1 의 (b)네트워크를 학습시키기 위해서 이상치에 대한 데이터가 필요하지만 학습에 사용할 수 있는 데이터는 4 개의 주석이 달린 골프 스윙 영상, 총 308 프레임, 60 개 관절에 좌표 정보가 있다. Bi-LSTM 네트워크를 학습시키기 위해서는 더 많은 데이터가 필요하므로 이상치 데이터 생성 기법을 통해 총 102,400 개의 관절 데이터를 생성하였다. 데이터 생성은 랜덤한 프레임 수와 선택할 프레임에 대한 랜덤 변수 그리고 연속적인 이상치 생성에 대한 랜덤 변수를 통해 선택된 프레임의 x, y 좌푯값에 임의의 감쇠 혹은 증폭을 주어 이상치 데이터를 생성했다. 이는 실제 골프 스윙 데이터에서 발생하는 오검출 사례를 재현한 것으로 실제 예측 데이터에서 발생하는 이상치를 가깝게 따르는 감쇠 및 증폭의 양을 찾기 위해 실험을 진행했다.

3.2 이상치 검출 및 보정 알고리즘

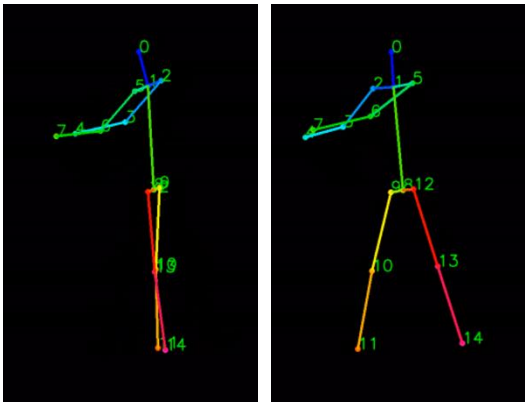
Step1: Bi-LSTM 기반의 이상 프레임 검출



(그림 2) 4 번 관절(오른쪽 손목)의 변화량 그래프 및 이상치에서의 예측 관절 시각화

이상치 검출 모델은 그림 1 의 (b)모델로 1 개 층의 Bi-LSTM 네트워크를 사용하여 이상치가 발견된 프레임을 출력값으로 알려주는 역할을 한다. 그림 2 는 특정 관절의 변화량을 그래프로 시각화 한 것이다. 데이터의 프레임 간 좌표 변화량을 분석한 결과 오검출 사례에서는 좌표의 변화량이 이전 프레임들보다 비정상적인 모습을 보이며 변화량이 크게 변하는 것을 확인할 수 있었다. 네트워크의 입력 데이터는 관절 좌푯값의 프레임 별 변화량이며 유클리드 거리 계산법을 통해 n-1 번째 프레임과 n 번째 프레임 간의 변화량을 계산했다. 데이터 한 개는 하나의 관절에 대한 프레임별 움직인 거리가 기록되어 있다. 영상 크기 별로 좌푯값의 범위가 다르므로 각 변화량 데이터는 각각의 최댓값으로 나눠주어 값의 범위를 0 과 1 사이의 값으로 정규화시켰다. 네트워크의 최대 입력 길이는 150 프레임이며 더 짧은 프레임의 경우에는 0 으로 패딩값을 준다. 입력으로 들어가는 데이터는 프레임별로 [좌표 변화량, 해당 관절의 변화량 평균]을 기본으로 했고, [관절 번호] 데이터를 추가로 입력하여 성능 변화 실험을 진행했다.

Step2: 골프 모션의 특징을 고려한 이상 좌표 검출



(그림 3) 좌표 반전 문제와 하체 이상치 문제 해결 전후 관절 이미지. (좌) 해결 전, (우) 해결 후

그림 1 의 (a)는 관절 좌표의 반전 문제를 검출 및 보정하는 방법이다. 그림 3 을 보면 골프 자세에서 11 번 관절(오른쪽 발목)은 14 번 관절(왼쪽 발목)보다 항상 작은 x 값의 좌표를 가진다. 이 규칙을 이용하여 11 번 관절의 x 좌표가 14 번 관절의 x 좌표보다 작으면 관절을 좌우 반전을 시키는 간단한 규칙을 정의했다.

그림 1 의 (c)는 하체 이상치 검출 방법으로 양쪽 발목이 골프 스윙 모션 중에 일정 거리 이상 좁혀지지 않는 특징을 이용한 규칙이다. 프레임별 양쪽 발목의 거리를 유클리드 거리 계산법을 통해 계산한 후 0 과 1 사이로 정규화한다. 그 후, 임의의 수치 이하로 내려간 좌표는 이상치로 판단하고 제거한다.

그림 1 의 (d)는 손목 이상치 검증 규칙으로 두 손목은 골프 스윙 중 일정하다는 것을 이용하여 (c)의 방법과 마찬가지로 일정 이상 거리가 벌어지면 이상치로 판단하고 제거한 후 반대 손의 좌표를 이용하여 보정하는 방식이다.

그림 3 은 좌표 반전 문제와 발목 이상치 문제가 발생했을 때 (a)와 (c)의 규칙을 통해 이상치를 제거하고, 보간한 결과이다.

Step3: 결측 좌표 보간(Interpolation)

그림 1 의 (e)는 이상치를 제거한 후 결측 좌표를 보간하는 알고리즘이다. 보간 방법에는 Linear, Nearest, Previous, Next 등이 있으며 각 방법의 성능을 비교해 본 결과 Linear 한 방법이 높은 성능을 보였다. Linear Interpolation 은 두 점$(x_1, f(x_1)), (x_2, f(x_2))$가 주어졌을 때 그 사이 값인 x에 대한 $f(x)$의 값을 추정하는데 사용되는 방법이고 수식은 아래와 같다.

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

4. 실험 및 분석

<표 1> 실험 방법에 따른 MPJPE 변화

모델	MPJPE					
	S1	S2	S3	S4	S5	평균
기준모델 BlazePose[3]	14.18	16.52	16.77	46.14	17.86	22.29
2 - Model(10,40)[D]	14.16	16.38	17.02	20.01	17.64	17.04
3 - Model(10,40)[D]	14.01	15.57	16.81	20.11	17.55	16.81
3 - Model(10,40)[J]	14.21	**14.93**	17.92	20.07	19.02	17.23
3 - Model(35,70)[D]	14.03	15.27	17.10	20.37	17.99	16.95
3 - Model(60,150)[D]	14.03	15.98	17.40	21.29	17.82	17.30
3 - Model(10,40)[D] + (a)	14.01	15.57	16.43	20.11	17.55	16.73
3 - Model(10,40)[D] + (a) + (c)	14.01	15.58	16.28	20.11	17.55	16.71
3 - Model(10,40)[D] + (a) + (d)	**13.16**	15.43	16.31	**19.83**	**17.54**	16.45
3 - Model(10,40)[D] + (a) + (c) + (d)	**13.16**	15.44	**16.16**	**19.83**	**17.54**	**16.43**

총 5 개, 483 프레임의 관절 주석이 달린 골프 스윙 영상에 대해 실험을 진행했다. 각 실험 영상은 S1 부터 S5 로 작성하였다. 성능 측정은 MPJPE(Mean Per Joint Position Error) 방법을 응용하여 사용했으며 계산식은 아래와 같다.

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

T는 총 프레임의 수, N은 측정에 사용한 관절의 수이다. MPJPE 는 정답 데이터에서 중심 관절과 측정하려는 관절의 거리와 예측데이터에서 중심 관절과 측정하려는 관절의 거리의 차를 통해 오류를 계산한다.

표 1 의 모델 이름은 '학습데이터 길이'-'Bi-LSTM 모델'('좌표 이상치 범위')['이상치 생성 방법'] 이며 각 (a), (c), (d)는 그림 1 에 나온 규칙을 의미한다. 기준 모델은 BlazePose[3]를 사용했으며, 각각의 모델을 통해 제거된 이상치를 보간한 후에 MPJPE 측정을 진행했다.

4.1 Bi-LSTM 방법을 이용한 이상 프레임 검출 성능

실험 결과 Bi-LSTM 모델을 학습할 때 데이터를 [좌표 변화량, 변화량의 평균]으로 주는 것보다 [관절

번호]를 추가하여 3 개의 정보를 주는 것이 더 높은 성능을 보였다. 이는 각 관절의 움직임의 정도가 달라서 일괄적인 적용보다 관절 별로 다른 기준을 적용하는 것이 더 높은 성능을 보임을 알 수 있었다. 이상치의 범위는 10px 에서 40px 사이의 좌표의 감쇠 혹은 증폭을 줄 경우에 성능이 높았다. 이상치 생성 방법의 경우 [D]는 1~3 사이의 연속한 프레임에 이상치를 생성했고, [J]의 경우에는 3 프레임 이상의 연속한 이상치를 생성했는데, [D]를 사용한 결과가 더 높은 성능을 보였다. [J]를 사용할 경우 모델이 이상치 여부를 학습하는데 어려움을 보였다.

4.2 이상치 검출 규칙에 따른 성능 변화

좌표 반전 검출 규칙(a)을 사용한 결과 평균 0.08 의 오차 감소가 있었고, 하체 이상치 검출 규칙(c)만 추가로 사용하였을 때는 평균에서 0.02 의 오차 감소가 생겼다. 손목 이상치 검출 규칙(d)을 사용했을 때는 사용하지 않을 때 보다 0.28 의 오차 감소가 있었다. 최종적으로는 (a), (b), (c)를 모두 사용하였을 때 제일 좋은 성능을 보였으며 기준 모델보다 5.86 의 오차 감소를 보여주었다.

4.3 다른 모델에 적용한 결과

<표 2> [3,4,5]에 본 연구에서 제안한 방법을 적용한 후 MPJPE 변화

모델	MPJPE					
	S1	S2	S3	S4	S5	평균
BlazePose[3]	14.18	16.52	16.77	46.14	17.86	22.29
BlazePose[3] + Ours	**13.16**	**15.44**	**16.16**	**19.83**	**17.54**	**16.43**
Open Pose[5]	14.31	70.75	91.64	57.48	33.83	53.60
Open Pose + Ours	**11.83**	**14.81**	**17.04**	**12.04**	**17.27**	**14.60**
HRNet[4]	**12.65**	10.66	14.44	9.15	**14.52**	12.28
HRNet + Ours	12.93	**10.20**	**13.39**	**8.84**	14.55	**11.98**

표 2 결과를 통해 자세 추정 모델 [3, 4, 5]에 본 논문에서 제안한 보정 과정을 추가로 적용하면 전체적으로 MPJPE 가 낮아진 것을 확인할 수 있었다. 또한 좌표 데이터의 손실이 많은 경우(S3, S4)에 큰 성능 향상이 있었는데, 이는 미검출된 좌표를 앞뒤 프레임의 좌표정보를 가지고 보간 했기 때문에 상당한 성능 개선이 이루어졌다고 볼 수 있다.

5. 결론 및 향후연구

본 논문에서는 기존의 사람 자세를 추정하는 모델의 성능을 개선하는 이상치 검출 및 보간 알고리즘을 제안하였다. 다양한 모델에 알고리즘을 적용한 결과 관절 예측의 성능 향상이 이루어 짐을 알 수 있었다. 기존의 연구는 CNN 을 이용한 좌표 보정 방법으로 3D 좌표 개선으로 활용하기에는 어려움이 있다. 하지만 본 논문에서 제안하는 알고리즘은 좌표 데이터를 통해 보정이 이루어지므로 3D 좌표 데이터에 대한 확장이 용이하다. 본 연구는 향후 가상 현실, 골프 게임, 골프 레슨 등 다양한 분야에 적용될 수 있을 것으로 기대한다.

사사표기

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2020-0-01343,인공지능융합연구센터지원(한양대학교 ERICA))

이 논문은 2021 년 과학기술정통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 연구결과로 수행되었음(2018-0-00192)

이 논문은 산업통상자원부 ‘산업혁신인재성장지원사업’의 재원으로 한국산업기술진흥원(KIAT)의 지원을 받아 수행된 연구임. (2021 년 산업 융합형 웨어러블 스마트 디바이스 전문인력 양성사업, 과제번호 : P0002397)

참고문헌

[1] RUGGERO RONCHI, Matteo; PERONA, Pietro. Benchmarking and error diagnosis in multi-instance pose estimation. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 369-378.

[2] DANG, Qi, et al. Deep learning based 2d human pose estimation: A survey. Tsinghua Science and Technology, 24.6: 663-676, 2019

[3] Bazarevsky, Valentin, et al. "BlazePose: On-device Real-time Body Pose tracking." arXiv preprint arXiv:2006.10204 (2020).

[4] SUN, Ke, et al. Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 5693-5703.

[5] CAO, Zhe, et al. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. IEEE transactions on pattern analysis and machine intelligence, 43.1: 172-186, 2019

[6] MOON, Gyeongsik; CHANG, Ju Yong; LEE, Kyoung Mu. Posefix: Model-agnostic general human pose refinement network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 7773-7781.

[7] CARREIRA, Joao, et al. Human pose estimation with iterative error feedback. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 4733-4742.

[8] FIERARU, Mihai, et al. Learning to refine human pose estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018. p. 205-214.