

## The Methodology of the Golf Swing Similarity Measurement Using Deep Learning-Based 2D Pose Estimation

Jonghyuk Park\*

\*Assistant Professor, Dept. of AI, Big Data & Management, Kookmin University, Seoul, Korea

### [Abstract]

In this paper, we propose a method to measure the similarity between golf swings in videos. As it is known that deep learning-based artificial intelligence technology is effective in the field of computer vision, attempts to utilize artificial intelligence in video-based sports data analysis are increasing. In this study, the joint coordinates of a person in a golf swing video were obtained using a deep learning-based pose estimation model, and based on this, the similarity of each swing segment was measured. For the evaluation of the proposed method, driver swing videos from the GolfDB dataset were used. As a result of measuring swing similarity by pairing swing videos of a total of 36 players, 26 players evaluated that their other swing sequence was the most similar, and the average ranking of similarity was confirmed to be about 5th. This ensured that the similarity could be measured in detail even when the motion was performed similarly.

▶ **Key words:** Motion similarity, Motion analysis, Golf, Sports, Deep learning, Artificial intelligence

### [요 약]

본 논문에서는 골프 동영상 속 스윙 자세 사이의 유사도를 측정할 수 있는 방법을 제안한다. 딥러닝 기반 인공지능 기술이 컴퓨터 비전 분야에 효과적인 것이 알려지면서 동영상을 기반으로 한 스포츠 데이터 분석에 인공지능을 활용하기 위한 시도가 증가하고 있다. 본 연구에서는 딥러닝 기반의 자세 추정 모델을 사용하여 골프 스윙 동영상 속 사람의 관절 좌표를 획득하였고, 이를 바탕으로 각 스윙 구간별 유사도를 측정하였다. 제안한 방법의 평가를 위해 GolfDB 데이터셋의 Driver 스윙 동영상을 활용하였다. 총 36명의 선수에 대해 스윙 동영상들을 두 개씩 짝지어 스윙 유사도를 측정한 결과, 본인의 또 다른 스윙이 가장 유사하다고 평가한 경우가 26명이었으며, 이때의 유사도 평균 순위는 약 5위로 확인되었다. 이로부터 비슷한 동작을 수행하고 있는 경우에도 면밀히 유사도를 측정하는 것이 가능함을 확인할 수 있었다.

▶ **주제어:** 동작 유사도, 동작 분석, 골프, 스포츠, 딥러닝, 인공지능

- 
- First Author: Jonghyuk Park, Corresponding Author: Jonghyuk Park
  - \*Jonghyuk Park (jonghyuk@kookmin.ac.kr), Dept. of AI, Big Data & Management, Kookmin University
  - Received: 2022. 12. 22, Revised: 2023. 01. 10, Accepted: 2023. 01. 12.

## I. Introduction

딥러닝 기반의 인공지능 모델이 컴퓨터 비전 분야에서 기존 모델 대비 우수한 성능을 보여주면서, 이를 여러 분야에 응용하기 위한 시도가 증가하고 있다[1-3]. 사람의 자세를 추정하기 위한 연구[4-8] 또한 컴퓨터 비전 분야의 대표적인 응용 연구 중 하나로, 이미지 속 사람의 여러 신체 부위에 대한 관절 좌표를 정확하게 추정하는 것이 연구의 목표이다. Fig. 1은 사람의 자세 추정 결과를 시각화한 그림으로, 사전에 정의된 관절들의 픽셀 좌표를 구하여 파란색 점으로 표현한 것을 확인할 수 있다.

이렇게 획득한 이미지 속 사람의 각 신체 부위에 대한 관절 좌표는 동영상 단위로 확장되어, 사람의 동작을 인식하고 분석하기 위해 사용될 수 있다. 동영상에서 수행되고 있는 사람의 동작이 어떤 동작인지 인식할 수도 있고[8, 9], Closed-circuit television(CCTV) 속 사람의 이상행동을 탐지하기 위해 동작을 분석할 수도 있다[10].

스포츠 분야에서도 선수의 동작을 인식, 평가하고 이를 바탕으로 선수의 경기력을 향상시키기 위해 인공지능을 활용하는 사례가 증가하고 있다. 딥러닝 기반의 자세 추정 모델을 통해 이미지 또는 동영상 속 사람의 관절 좌표를 획득하여, 이를 바탕으로 사람의 자세를 측정하고 스포츠 동작을 분석하는 것이 대표적인 활용 사례라 할 수 있다 [11, 12].

여러 스포츠 중, 골프는 최근 들어 시장 규모가 급격하게 성장하고 있는 종목이다[13]. 골프를 즐기는 연령대가 20대 및 30대까지 확장되며 의류, 장비, 스크린 골프 등 관련 산업까지 함께 시장 규모가 커지고 있다[14]. 데이터를 분석하고 인공지능을 학습시키는 측면에 있어, 골프는 스윙 동작의 시작 시점과 종료 시점을 명확하게 알 수 있기 때문에 인공지능을 활용하기에 적합한 종목이라 할 수 있다. 또한 같은 자리에서 이동하지 않고 스윙 동작을 수행하기 때문에 촬영하는 카메라가 고정되어 있다는 점 또한 보다 정확한 분석을 가능하게 하는 요인이다.

이에, 본 연구에서는 골프 스윙에 대해 동작 유사도를 측정할 수 있는 방법론을 제안한다. 제안 방법으로 골프 스윙의 동작 유사도를 측정하기 위하여, 먼저 자세 추정 인공지능 모델을 통해 골프 스윙을 수행하고 있는 영상 속 사람의 자세, 즉 2차원 신체 관절 좌표를 획득한다. 그리고 유사도 측정 대상인 두 사람의 2차원 신체 관절 좌표를 Dynamic time warping(DTW)[15]을 사용하여 동작 수행 시간에 대해 정렬시킨다. 사람마다 스윙의 수행 속도에 차이가 있기 때문에 같은 동작을 수행하는 자세, 즉 관절 좌

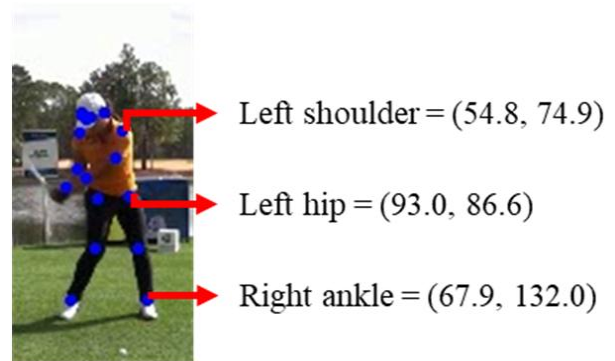


Fig. 1. Example Image of Pose Estimation Result

표끼리 비교하기 위해서는 이를 정렬시켜야 한다. 정렬된 자세 속 관절 좌표에 대해서는 Euclidean 거리를 사용하여 거리를 측정하였고, 이를 모든 관절 및 동작을 구성하는 모든 자세에 대해 평균 내어 최종 유사도를 계산하였다.

골프 스윙 동작의 유사도를 측정하기 위해서, GolfDB[16] 데이터셋의 Driver 스윙 영상을 활용하였다. GolfDB는 프로 골프 선수들의 스윙 영상을 제공하는 데이터셋이다. 같은 촬영 각도에서 스윙하는 여러 선수들의 영상이 존재하기 때문에 유사도를 측정하는데 있어 적합한 데이터셋이다. 또한, Address, Top, Impact, Finish 같은 스윙 동작의 구간에 대한 정보를 함께 제공하기 때문에 Backswing 동작(Address에서 Top까지의 동작), Downswing 동작(Top에서 Impact까지의 동작), Finish 동작(Impact에서 Finish까지의 동작)과 같이 전체 스윙을 세분화하여 유사도를 측정하는 것이 가능하다.

제안 방법을 평가하기 위해서, 같은 선수의 스윙 동작 쌍을 다른 선수와의 스윙 동작 쌍 대비 높은 유사도로 측정할 수 있는지 확인하고자 하였다. 이를 위해, GolfDB 데이터셋에서 같은 카메라 시점에 대해 두 개 이상의 골프 스윙이 있는 선수만 추출하여 유사도를 측정하였다. 유사도 측정 결과, 본인의 스윙 쌍에 대해 가장 유사하다고 측정된 경우가 36명 중 26명, 72%에 해당되었다. 이러한 결과로부터, 동작 수행 과정이 서로 유사한 골프 스윙에 대해서도 면밀하게 유사도를 측정하는 것이 가능한 것을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 우선, 제 2장에서는 관련 연구를 소개한다. 제 3장에서는 제안 방법에 대해 설명하고, 제 4장에서는 제안 방법의 검증을 위해 사용한 데이터셋, 실험 설정 및 결과에 대해 기술한다. 마지막으로 제 5장에서는 결론 및 향후 연구에 대해 서술하며 논문을 마무리한다.

## II. Related Works

제 1장에서 언급한 것처럼, 딥러닝 기반의 인공지능 모델이 이미지와 동영상 데이터를 처리하는데 효과적인 것이 드러나면서, 스포츠 경기 동영상을 분석하는데 인공지능을 활용하기 위한 시도가 증가하고 있다. 본 장에서는 먼저 골프 스윙 분석에 딥러닝 기반 인공지능 모델이 쓰인 사례에 대해 소개하고, 제안 방법에 활용되어진 자세 추정 방법론과 DTW에 대해 설명한다.

### 1. Golf swing analysis using deep learning

자세 추정 모델의 성능 발전과 함께 사람의 관절 위치를 보다 정확하게 추론하는 것이 가능해지며, 이를 골프 스윙 동작 분석에 활용하기 위한 연구가 수행되고 있다. [16]의 연구에서는 프로 골프 선수들의 스윙 동영상을 수집하여 공개함과 동시에 골프 스윙 동작을 8개의 이벤트로 세분화할 수 있도록 데이터셋에 주석을 달고 직접 세분화 모델을 학습시켰다. [17]의 저자들은 사전 학습된 High-resolution net(HRNet)[5]을 활용하여, 17개의 신체 관절 좌표를 추정하였고, 여기에 클럽 헤드 특징점 좌표를 추가해 HRNet을 추가 학습시킴으로써 클럽 헤드의 궤적을 추정할 수 있는 시스템을 제안하였다. 박정욱 외 4인의 연구 또한 HRNet을 사용하여, 실내 골프 연습장에서 수집한 골프 스윙 데이터를 바탕으로 모델을 학습시켜 골프 스윙 자세를 추론할 수 있는 모델을 제시하였다[18]. 고경리와 반성범은 합성곱 신경망[19]을 사용하여 스윙 구간 정보를 갖는 시계열을 생성하였고, 이를 양방향 장단기 메모리[20]에 투입하여 바디-스웨이, 헤드-업, X-factor에 대한 분석을 수행하였다[21]. 한편, 프로 골프 선수의 영상을 분석하여 표준 자세를 정의하고, 이를 골퍼를 배우고 싶은 사람의 자세와 비교하기 위한 시스템을 제안한 연구도 존재한다[22]. 이때 저자들은 OpenPose[4]를 활용하여 스윙 동작의 관절 좌표를 추론하였다.

위 연구들은 자세 추정 모델을 활용하여 스윙을 분석하였지만, 스윙 간 직접적인 비교를 수행하지 못했다. 본 논문에서는 HRNet을 발전시킨 자세 추정 모델인 Disentangled keypoint regression(DEKR)[7]을 바탕으로, 두 골프 스윙의 동작 유사도를 계산하여 스윙을 직접 비교할 수 있는 방법을 제시하고자 한다.

### 2. 2D human pose estimation

2차원 사람 자세 추정은 이미지나 동영상 같은 시각 정보에 나타나 있는 사람의 각 관절 위치, 즉 2차원 좌표를

추론하기 위한 과업으로, 이 문제를 해결하기 위한 대표적인 모델로는 OpenPose[4], HRNet[5] 등이 있다. OpenPose는 이미지에 나타나 있는 모든 신체 관절을 먼저 찾아내고, 이후에 신체 관절들을 어떻게 연결할 수 있는지 추론해내는 Bottom-up 방식을 채택하고 있는 모델이다. 따라서 여러 사람이 이미지에 등장할 경우 각 신체 관절 부위가 어느 사람에 속하는지 알아내야 한다. 이를 위해 Part Affinity Field를 구성하여 신체 관절 부위를 연결하게 된다. HRNet은 사람 객체의 경계 상자를 먼저 찾은 다음에, 경계 상자 안에서 자세를 추정하는 Top-down 방식을 채택하고 있는 모델이다. 기존 모델들과 달리 고해상도 정보와 저해상도 정보를 지속적으로 혼합(Fusion)하는 방식을 통해 자세 추정 문제에서 높은 성능을 달성하였다.

HRNet의 우수한 성능은 이를 기반으로 한 여러 변형 모델들을 만들어지게 하였고, 대표적인 모델로는 HigherHRNet[6]과 DEKR[7]이 있다. 기존의 Bottom-up 방식은 빠른 속도로 사람의 자세를 추론 가능하다는 장점이 있지만, 이미지 상에서 작게 표현된 사람의 자세를 추정하는데 낮은 성능을 보인다는 약점이 있었다. HigherHRNet은 이러한 Bottom-up 방식의 한계를 HRNet 기반의 고해상도 Feature pyramid 모델 구조를 통해 극복하고자 하였다. DEKR은 HRNet Backbone 구조에 Adaptive convolution을 사용하여 각 관절 특징점 영역에 있는 픽셀을 활성화시켰다. 그리고, 활성화된 픽셀의 표현을 학습하게 하는 방식으로 해당 관절 특징점 영역에 집중할 수 있도록 유도하여 Bottom-up 방식에서 우수한 성능을 달성하였다.

본 연구에서는 DEKR을 사용하여 골프 스윙 동작의 관절 좌표를 추론하였다. DEKR은 Bottom-up 방식으로 관절 좌표를 추론하기 때문에 높은 추론 성능과 함께 빠른 속도로 관절 위치를 추정할 수 있다는 장점이 있다.

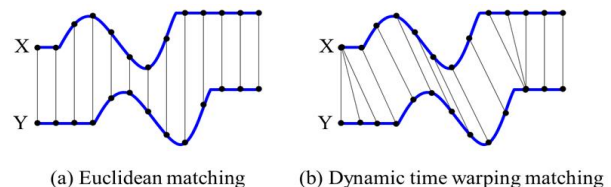


Fig. 2. Comparison of Euclidean Matching and DTW Matching

### 3. Dynamic time warping

DTW는 길이와 속도가 다른 두 시계열  $X = (x_1, x_2, \dots, x_N)$ ,  $Y = (y_1, y_2, \dots, y_M)$ ,  $N, M \in \mathbb{N}$  사이의 유사도 기반 최적의 정렬을 찾기 위해

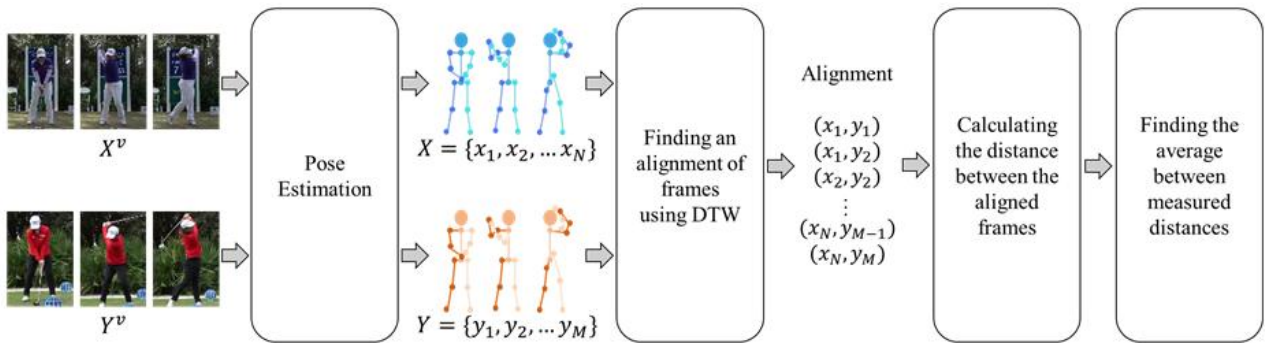


Fig. 3. Proposed Method for the Measuring Similarity between Two Golf Swing Motions

사용되는 알고리즘이다. 비용 행렬  $C \in \mathbb{R}^{N \times M}$  :  $C_{nm} = \|x_n, y_m\|$ ,  $n \in [1 : N]$ ,  $m \in [1 : M]$  를 계산하고,  $C_{11}$ 에서  $C_{NM}$ 까지 비용의 합이 최소화 되는 경로를 찾아 두 시계열의 각 지점을 정렬하게 된다. Fig. 2은 Euclidean 기반 정렬과 DTW 기반 정렬을 비교한 그림이다. 시간에 따라 정렬한 Euclidean 기반 정렬 방법과 달리 DTW 기반 정렬은 각 지점에서 시계열 구성 요소의 유사도에 따라 각 지점을 짝지은 것을 확인할 수 있다.

본 연구에서는 DTW를 통해 두 자세의 시계열, 즉 동작을 정렬하였고 정렬된 자세끼리 유사도를 측정된 뒤 이를 모든 지점에 평균내어 두 동작의 최종 유사도를 산출하였다.

### III. The Proposed Method

Fig. 3는 본 논문에서 제안하는 두 스윙 사이의 유사도를 측정하기 위한 방법을 도식화한 그림이다. 두 스윙 사이의 유사도를 측정하기 위해서, 먼저 골프 스윙을 포함하고 있는 두 동영상  $X^v$ ,  $Y^v$ 의 프레임 단위 자세 추정을 수행한다. 그 결과 생성된  $X^v$ 의 각 프레임에 대한 자세 추정 시계열을  $X = (x_1, x_2, \dots, x_n, \dots, x_N)$ ,  $Y^v$ 의 각 프레임에 대한 자세 추정 시계열을  $Y = (y_1, y_2, \dots, y_m, \dots, y_M)$ 이라고 하자. 여기서  $x_n$ ,  $y_m$ 은 각각  $X^v$ 의  $n$ 번째 프레임과  $Y^v$ 의  $m$ 번째 프레임에 대한 자세 추정 결과를 의미하며,  $x_n, y_m \in \mathbb{R}^{2 \times K}$ 이다. 또한  $K$ 는 사전에 정의된 신체 관절의 개수를 의미한다.

본 연구에서는 앞에서 언급한 것처럼, 자세 추정 모델로 DEKR을 사용하였다. DEKR의 저자들이 공개한 Github 공식 저장소의 프로그램과, MS COCO[24] 데이터셋을 사용하여 모델을 학습하였고, 그 결과 COCO 데이터셋이 정의하고 있는 17개의 관절 좌표를 생성할 수 있는 모델을 만들었다. 이렇게 만들어진 COCO 데이터셋의 관절 중 오

른쪽 눈, 왼쪽 눈, 오른쪽 귀, 왼쪽 귀는 스윙 자세 유사도를 측정하는데 불필요하였기 때문에 제거하였다. 또한, 추정된 관절 좌표를 신체 중심을 기준으로 상대 좌표로 변환하기 위해 가운데 엉덩이(Middle hip)를 관절 구조에 추가하여 총 15개의 관절 좌표를 획득하였다. 이는 Table. 1에 정리되어 있다.

Table 1. Human Joints Used in the Proposed Method

Index	Joint
0	Nose
1	Neck
2	Right shoulder
3	Right elbow
4	Right wrist
5	Left shoulder
6	Left elbow
7	Left wrist
8	Middle hip
9	Right hip
10	Right knee
11	Right ankle
12	Left hip
13	Left knee
14	Left ankle

이후, 2.3절에서 설명한 DTW를 통해  $X$ 와  $Y$ 의 프레임을 정렬시킨다. 골프 스윙의 속도는 그 스윙 동작을 수행하는 사람마다 다를 수 있다. 본 연구에서는 스윙 자세의 비교에 초점을 맞춘 유사도 측정 방법을 제안하는 것이 목표이기 때문에, 사람 간 스윙 속도의 차이는 고려하지 않았다. 따라서, 두 스윙의 각 프레임, 즉 자세에서 대응되는 신체 관절 사이의 거리로 DTW의 비용 행렬  $C$ 를 계산하고, 계산된 비용의 총합이 최소가 되는 프레임 쌍을 구하게 된다. 이러한 프레임 쌍의 집합을  $A = \{a_1, \dots, a_l, \dots, a_L\}$ 라고 하자. 여기서 집합  $A$ 의 원소  $a_l$ 은  $X$ 와  $Y$  각각의 프레임에서 자세를 추정한 결과로 구성된 쌍  $(x_n, y_m)$ 이라 하자. 이렇게 짝지어진 두 프레임 속 사람의 자세는 서로 유사하다고 할 수 있기 때문에



Fig. 4. Example Images of Three Views of the GolfDB Dataset

스윙 속도를 고려하지 않고, 비슷한 스윙 단계를 수행하고 있는 프레임끼리 비교할 수 있게 된다.

다음으로, 짝지어진 두 프레임의 대응 관절 간 거리를 측정하고, 이를 모든 관절  $K$ 개에 대해 평균 내어 프레임 쌍에서의 평균 관절 거리를 측정한다. 구체적으로, 본 연구에서는 코부터 왼쪽 발목까지 관절을 15개 사용했기 때문에  $x_n$ 의 코 좌표와  $y_m$ 의 코 좌표 사이의 거리를 계산하고, 같은 방식으로  $x_n$ 과  $y_m$ 의 왼쪽 발목 좌표 사이의 거리까지, 총 15개 대응 관절 사이의 거리를 구하여 계산한 모든 거리의 평균을 구하는 방식이다. 그리고, 이를 모든 프레임 쌍에 대해 평균내어 두 비교 동작에서의 최종 평균 거리를 구한다.  $l$ 번째 프레임 쌍  $a_l$ 의  $k$ 번째 관절에 대해 측정된 거리를  $d_l^k$ 라고 했을 때, 최종 평균 거리는 다음 수식과 같이 쓸 수 있다.

$$AverageDistance = \frac{1}{L} \frac{1}{K} \sum_{l=1}^L \sum_{k=1}^K d_l^k$$

(1)

본 논문의 제안 방법에서는 거리 측정 함수로 Euclidean 거리를 사용하였다. 이와 같은 과정을 통해 구해진 최종 평균 거리로부터 두 시계열  $X$ 와  $Y$  사이의 최종 유사도를 산출하기 위해 다음 수식을 활용하여 유사도로 전환한다.

$$Similarity = \frac{\alpha}{AverageDistance + \epsilon}$$

(2)

본 연구의 최종 모델에서는  $\alpha = 3$ ,  $\epsilon = 0.00001$  값을 사용하여 유사도로 변환하였다.

Table 2. Swing Events Defined by GolfDB Dataset

Event	Description
Address	The moment just before the takeaway begins
Toe-up	Shaft parallel with the ground during the backswing
Mid-backswing	Arm parallel with the ground during the backswing
Top	The moment the golf club changes directions at the transition from backswing to downswing
Mid-downswing	Arm parallel with the ground during the downswing
Impact	The moment the clubhead touches the golf ball
Mid-follow-through	Shaft parallel with the ground during the follow-through
Finish	The moment just before the golfer's final pose is relaxed

## IV. Experiments

### 1. Dataset

제안 방법의 평가를 위한 데이터셋으로, GolfDB[16]을 사용하였다. GolfDB는 YouTube로부터 수집한 총 246명의 골프 프로 선수의 스윙 동영상으로 구성되어 있다. Driver, Fairway, Hybrid, Iron, Wedge Club 각각에 대하여 206명, 57명, 19명, 60명, 11명의 스윙 동영상을 포함하고 있으며, 카메라 시점 정보는 Fig. 4에 나타나 있는 세 가지(Face-on, Down-the-line, Other) 시점으로 분류된다. 또한 여덟 가지의 스윙 Event가 발생한 시점에 대한 프레임 정보를 제공하기 때문에 스윙 Event의 발생 시점을 탐지하는 모델을 학습시킬 수 있다. Table 2는 GolfDB에서 정의한 스윙 Event를 스윙 순서대로 제시한 표로, 각 스윙 Event에 대한 설명을 확인할 수 있다.

본 연구에서는 동일한 선수의 서로 다른 스윙이 다른 선수와의 스윙과 비교하였을 때 낮은 최종 평균 거리를 산출할 수 있는지, 즉 높은 유사도를 기록할 수 있는지 확인하고자 하였다. 따라서 두 카메라 시점, Face-on과 Down-the-line 각각에 대해 스윙 동영상이 2개 이상 존재하는 36명의 선수를 선별하였고, 스윙 종류는 Driver 스윙으로 한정하였다. Table 3은 각 시점별 선별된 선수의 선수 경력을 요약한 표로, 프로 선수들 중에서도 우수한 경력은 지닌 선수들이 선별되었음을 알 수 있다. 이로부터 각 카메라 시점당 각각 4,560개와 4,753개의 비교 동영상 쌍을 만들었고, 이때 같은 스윙 동영상으로 구성된 쌍은 삭제하여 다른 스윙 동영상 사이에서만 비교가 수행될 수 있도록 하였다.

Table 3. Career Stats of the Selected Players on the PGA/LPGA

Players in Face-on view videos	The number of Victories	The number of Top 10s	Players in Down-the-line view videos	The number of Victories	The number of Top 10s
AI MIYAZATO	9	60	AI MIYAZATO	9	60
ALISON LEE	-	16	ANNA NORDQVIST	9	77
ANNA NORDQVIST	9	77	AYAKO UEHARA	-	8
BERNHARD LANGER	44	210	BEATRIZ RECARI	3	26
BROOKE HENDERSON	12	70	BILLY HORSCHER	7	55
CHARLES HOWELL	3	98	BRITTANY LINCICOME	8	65
CHARLEY HULL	2	28	BROOKE HENDERSON	12	70
COLIN MONTGOMERIE	7	63	BUD CAULEY	-	23
CRISTIE KERR	20	182	CHARLES HOWELL	3	98
GARY WOODLAND	4	53	CHARLEY HULL	2	28
GRAEME MCDOWELL	4	38	CINDY LACROSSE	-	-
GREG NORMAN	20	128	CRISTIE KERR	20	182
HARRIS ENGLISH	4	39	DUSTIN JOHNSON	24	115
HYO JOO KIM	5	44	GARY WOODLAND	4	53
INBEE PARK	21	118	GERINA PILLER	-	40
IN-GEE CHUN	4	40	HEE YOUNG PARK	3	40
JASON DAY	12	84	IAN POULTER	3	52
JASON DUFNER	5	47	IK KIM	7	79
JENNY SHIN	1	37	INBEE PARK	21	118
JERRY KELLY	11	62	JOHN DALY	5	35
JUSTIN THOMAS	15	71	LEE WESTWOOD	2	48
LEXI THOMPSON	11	84	LEXI THOMPSON	11	84
LYDIA KO	19	101	LYDIA KO	19	101
MATT KUCHAR	9	111	MARK WILSON	-	3
MICHELLE WIE	5	49	MICHELLE WIE	5	49
MINJEE LEE	8	65	MIKE WEIR	1	14
PAULA CREAMER	10	104	MORGAN PRESSEL	2	66
RETIEF GOOSEN	2	32	PADRAIG HARRINGTON	4	13
RYAN PALMER	4	68	PAULA CREAMER	10	104
SANDRA GAL	1	28	RETIEF GOOSEN	2	32
SEAN OHAIR	4	45	RICKIE FOWLER	5	76
SEI YOUNG KIM	12	63	RORY MCILROY	23	106
SO YEON RYU	6	96	RORY SABBATINI	6	80
STACY LEWIS	13	115	SANDRA GAL	1	28
TOM LEHMAN	12	88	STACY LEWIS	13	115
YANI TSENG	15	67	TIM CLARK	2	54

## 2. Swing segment configuration

본 연구에서는 GolfDB 데이터셋에서 제공하는 Event 정보를 활용하여, 전체 스윙을 세부 구간으로 나누어 각 구간별 유사도를 측정하고자 시도하였다. 이에, 전체 Event를 하나의 스윙 구간으로 간주하는 방식 이외에 추가로 두 가지 방식의 스윙 구간을 정의하였다. 첫 번째 방식은 전체 스윙을 두 개의 세부 스윙 구간으로 나누는 방식으로, Address 자세에서 Top 자세까지를 첫 번째 구간으로, Top 자세에서 Finish 자세까지를 두 번째 구간으로 한다. 두 번째 방식은 세 개의 세부 스윙 구간으로 전체 스윙을 나눈 방식이다. 구체적으로, Address 자세에서 Top 자세까지를 첫 번째 스윙 구간으로, Top 자세에서 Impact 자세까지를 두 번째 스윙 구간으로, 마지막으로 Impact 자세에서 Finish 자세까지를 세 번째 스윙 구간으로 정의하였다. 이는 Table 4에 표로 정리되어 있다.

Table 4. Swing Segment Configuration

The number of segments	Swing segment configuration
1	1. Address → Toe-up → Mid-backswing → Top → Mid-downswing → Impact → Mid-follow-through → Finish
2	1. Address → Toe-up → Mid-backswing → Top 2. Top → Mid-downswing → Impact → Mid-follow-through → Finish
3	1. Address → Toe-up → Mid-backswing → Top 2. Top → Mid-downswing → Impact 3. Impact → Mid-follow-through → Finish

## 3. Results

제안 방법의 유사도 측정 방식을 평가하기 위하여, 먼저 골프 선수마다 고유의 스윙 자세가 있어 동일한 선수의 스

윙을 비교했을 때, 다른 선수의 스윙과 비교했을 때보다 더 높은 유사도를 기록할 것이라는 가정을 설정했다. 이 가정 하에 동일한 선수의 스윙 쌍으로부터 측정된 유사도가 모든 선수와의 유사도 중 몇 위를 기록하는지 확인하였고 이러한 순위를 36명의 모든 선수에 대해 평균내어 제안 방법의 효과를 입증하고자 하였다. 또한, 가장 효과적인 유사도 측정 방식의 탐색을 위해 4.2절에서 설명한 추가적인 두 가지 방식의 스윙 구간 정의를 사용하여, 각각의 스윙 구간에서 유사도를 측정한 뒤 이를 평균내어 최종 유사도를 측정하였다. 또한 식 (1)의 거리 측정을 Euclidean 거리 이외에 Cosine 거리로도 측정하여 최적의 방법을 탐색하고자 하였다.

Table 5는 Face-on과 Down-the-line 각각의 카메라 시점에 대한 유사도 순위 측정 결과를 나타내는 표이다. 세 개의 스윙 세부 구간과 Euclidean 거리를 사용했을 때 동일한 선수 스윙의 유사도 평균 순위가 각각의 카메라 시점에서 6.83등, 5.17등으로 우수한 성능을 기록하였다. 특히, Down-the-line 카메라 시점에서 동일 선수의 스윙 유사도가 Face-on 보다 더 높은 순위를 기록한 것을 확인할 수 있다. 이는 Down-the-line 시점에서 스윙의 차이가 더 명확하게 드러나기 때문이다. Fig. 5는 INBEE PARK과 LYDIA KO의 Impact 직후의 자세 추정 결과를 시각화한

Table 5. Average ranking for pairs of the same player

The number of segments	Distance metric	Average rank for pairs of the same player	
		Face-on	Down-the-line
1	Euclidean	7.64th	5.19th
	Cosine	9.31th	8.19th
2	Euclidean	7.36th	5.19th
	Cosine	8.36th	7.92th
3	Euclidean	<b>6.83th</b>	<b>5.17th</b>
	Cosine	7.75th	6.92th

그림으로, INBEE PARK과 LYDIA KO의 경우 Down-the-line 시점에서는 본인 스윙의 유사도가 1위였으나, Face-on 시점에서는 24위와 9위로 각각 떨어졌다. Down-the-line 시점의 경우 두 선수의 어깨 기울기 차이가 명확하게 표현되는 것을 확인할 수 있는 것에 반해, Face-on 시점에서는 이러한 점이 잘 드러나지 않는 것을 알 수 있다. 이렇게, Down-the-line 시점에서는 팔과 몸통의 자세가 잘 드러나기 때문에 이러한 점이 유사도 측정에 반영되어 본인의 스윙을 비교할 때 높은 유사도를 기록하였다. 반면, Face-on 시점에서는 본인 스윙의 고유한 자세가 동영상 상에서 잘 표현되지 않고 다른 선수와 비슷해 보이기 때문에 상대적으로 낮은 유사도 순위를 기록하였다.



Fig. 5. Pose Comparison between Down-the-line and Face-on

Fig. 6은 1위부터 5위까지 해당 등수 내 동일 선수의 스윙 유사도 순위가 포함되는 인원 수를 기록한 그래프이다. Down-the-line 시점의 경우, 본인 스윙의 유사도 순위가 1위인 선수가 26명으로 20명인 Face-on 시점 대비 6명이나 많았다. 또한, 본인 스윙의 유사도 순위가 5위 안에 포함되는 선수가 Down-the-line 시점의 경우 29명인 것에 비해, Face-on 시점의 경우에는 그보다 낮은 26명을 기록하였다. 이러한 사실로부터 골프 스윙 유사도를 측정하는데 있어 Down-the-line 시점이 Face-on 시점보다 효과적이며 스윙 자세를 비교하기에 충분한 정보를 제공하는 것을 알 수 있다.

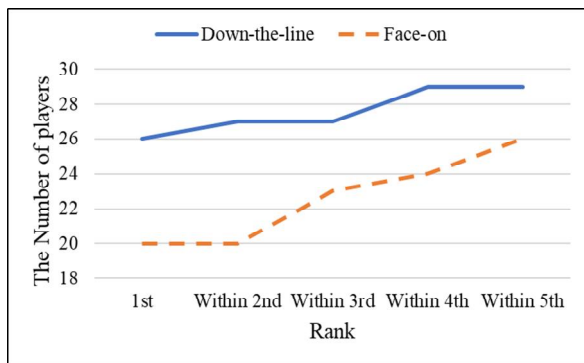


Fig. 6. The Number of Golfers Whose Swing Similarity Is in the Each Rank

## V. Conclusions

딥러닝 기반 인공지능의 활용도가 점차 커짐에 따라 스포츠 분석에도 인공지능을 활용하려는 시도가 늘고 있다. 본 연구에서는 특히 최근 들어 높은 시장 성장을 기록하고 있는 골프 종목에 인공지능을 접목하여, 스윙 동작을 비교하고 유사도를 측정할 수 있는 방법론을 제시하였다. 이를 위해 자세 추정 모델을 사용하여 스윙 동작의 각 프레임별 사람 관절 좌표를 추정하였고, 이를 DTW를 활용하여 정렬시킨 다음 각 관절별 거리를 측정하여 최종 유사도를 산출하였다. 골프 스윙 동영상을 포함하고 있는 GolfDB 데이터셋으로 실험한 결과, 선수의 동일한 선수의 스윙을 가장 유사하다고 평가한 경우가 약 전체 36명 중 26명을 기록하였다. 또한, 동일한 선수의 스윙 유사도가 5위 이내인 경우는 전체 36명 중 29명이었다. 이러한 결과로부터 제안 방법은 골프 스윙이라는 서로 유사한 동작에 대해서도 면밀히 유사도를 측정하는 것이 가능한 것을 확인할 수 있었다. 더불어, Down-the-line 카메라 시점에서 촬영한 골프 스윙 동영상을 사용해 유사도를 측정하는 것이

Face-on 시점보다 더 정확한 유사도를 측정 가능하게 한다는 것 또한 알 수 있었다.

이와 같은 유사도 측정 방식은 자세 추정 결과가 정확하지 않으면 면밀한 유사도 측정이 불가능하게 된다. 따라서 동작이 크고, 빠른 스포츠 동작에서도 높은 정확도를 보이는 자세 추정 모델을 개발하는 것은 본 연구의 추후 연구라고 할 수 있다. 또한, 골프 스윙 이외에 동작의 수행 자세가 중요한 야구나 농구 같은 종목에도 제안한 방법을 적용해볼 수 있을 것이다. 마지막으로, 자세 교정 피드백을 제공할 수 있는 시각화 모듈을 개발하는 것 또한 의미가 있을 것이다.

## ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2022-00166634).

## REFERENCES

- [1] J. Lee, M. Shin, J. Park, and N. Moon, "Deep Learning-Based Companion Animal Abnormal Behavior Detection Service Using Image and Sensor Data," *Journal of The Korea Society of Computer and Information*, Vol. 27, No. 10, pp. 1-9, Oct. 2022. DOI: 10.9708/jksci.2022.27.10.001
- [2] H. Kim, and J. Cho, "A Study on Intelligent Skin Image Identification From Social Media Big Data," *Journal of The Korea Society of Computer and Information*, Vol. 27, No. 9, pp. 191-203, Sep. 2022. DOI: 10.9708/jksci.2022.27.09.191.
- [3] J. Im, and D. Kim, "Corneal Ulcer Region Detection With Semantic Segmentation Using Deep Learning," *Journal of The Korea Society of Computer and Information*, Vol. 27, No. 9, pp. 1-12, Sep. 2022. DOI: 10.9708/jksci.2022.27.09.001.
- [4] Z. Cao, G. H. Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 1, pp. 172-186, Jan. 2021. DOI: 10.1109/TPAMI.2019.2929257.
- [5] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao, "Deep High-Resolution Representation Learning for Visual Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 10, pp. 3349-3364, Oct. 2021. DOI: 10.1109/TPAMI.2020.2983686.



- [6] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, "HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5385-5394, 2020. DOI: 10.1109/CVPR42600.20.00543.
- [7] Z. Geng, K. Sun, B. Xiao, Z. Zhang, and J. Wang, "Bottom-Up Human Pose Estimation Via Disentangled Keypoint Regression," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14676-14686, 2021. DOI: 10.1109/CVPR46437.2021.01444.
- [8] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "AdaSGN: Adapting Joint Number and Model Size for Efficient Skeleton-Based Action Recognition," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 13393-13402, 2021. DOI: 10.1109/ICCV48922.2021.01316.
- [9] X. Zhang, C. Xu, and D. Tao, "Context Aware Graph Convolution for Skeleton-Based Action Recognition," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14321-14330, 2020. DOI: 10.1109/CVPR42600.2020.01434.
- [10] Y. Yang, F. Angelini, and S. M. Naqvi, "Pose-Driven Human Activity Anomaly Detection in a CCTV-like Environment," IET Image Process, pp. 1-13, 2022. DOI: 10.1049/ipr2.12664.
- [11] P. Parmar, A. Gharat, and H. Rhodin, "Domain Knowledge-Informed Self-supervised Representations for Workout Form Assessment," Proceedings of the 17th European Conference on Computer Vision, pp. 105-123, 2022. DOI: 10.1007/978-3-031-19839-7\_7.
- [12] H. Li, Q. Lei, H. Zhang, J. Du, and S. Gao, "Skeleton-Based Deep Pose Feature Learning for Action Quality Assessment on Figure Skating Videos," Journal of Visual Communication and Image Representation, Vol. 89, 2022. DOI: 10.1016/j.jvcir.2022.103625.
- [13] J. Shin, G. Kim, and S. Kim, "A Study on the Current Status and Prospect of the Korean Golf Industry," Journal of Golf Studies, Vol. 16 No. 1, pp. 215-228, Mar. 2022. DOI: 10.34283/ksgs.2022.16.1.18.
- [14] S. Lee, S. Kim, and I. Park, "Analysis of the Empirical Relationship between Golf Wear Selection Attributes and Customer Behaviors of MZ Generation," The Society of Convergence Knowledge Transactions, Vol. 10 No. 3, pp. 59-71, Sep. 2022. DOI: 10.22716/sckt.2022.10.3.025.
- [15] R. Bellman, and R. Kalaba, "On Adaptive Control Processes," IRE Trans. on Automatic Control, Vol. 4, No. 2, pp. 1-9, Nov. 1959. DOI: 10.1109/TAC.1959.1104847.
- [16] W. McNally, K. Vats, T. Pinto, C. Dulhanty, J. McPhee, and A. Wong, "GolfDB: A Video Database for Golf Swing Sequencing," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 2553-2562, 2019. DOI: 10.1109/CVPRW.2019.00311.
- [17] D. Lee, and J. Nang, "A Dataset and Keypoint Detection for Club Head Trajectory Estimation in 2D Golf Swing Video," Proceedings of the Korean Information Science Society Conference, pp. 523-525, 2020.
- [18] J. Park, G. Baek, J. Lee, J. Kang, and J. Kim, "Design of Human Pose Estimation Model using HRNet for Golf Swing Motion Correction," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, pp. 991-992, 2021.
- [19] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," Proceedings of the Advances in neural information processing systems, pp. 396-404, 1990. DOI:
- [20] M. Schuster, and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. on Signal Processing, Vol. 45, No. 11, pp. 2673-2681, Nov. 1997. DOI: 10.1109/78.650093.
- [21] K. Ko, and S. Pan, "Swing Sequence Images based 3D Golf Swing Analysis using Deep Learning," Journal of the Institute of Electronics and Information Engineers, Vol. 56 No. 10, pp. 49-55, Oct. 2019. DOI: 10.5573/ieie.2019.56.10.49
- [22] W. Byeon, Y. Shim, H. You, and S. Kang, "Implementation of Computer Vision and Deep Learning-Based Golfer Pose-Estimation System And Coaching System," Proceedings of the Korea Information Processing Society Conference, pp. 1040-1043, 2020.
- [23] T. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár "Microsoft COCO: Common Objects in Context," arXiv preprint arXiv:1405.0312, May. 2015. DOI:10.48550/arXiv.1405.0312.

## Authors



Jonghyuk Park received the B.S. and Ph.D. degrees in Industrial Engineering from Seoul University, Korea, in 2015 and 2021, respectively. In 2015, Dr. Park was a Product Management Engineer with Samsung

Electronics. From 2020 to 2021, he was a Research Engineer in ai.m, Seoul, Korea. Since 2021, he has been an Assistant Professor at the Department of AI, Big Data & Management, Kookmin University, Seoul, Korea. His current research interests include computer vision, and machine learning applications.