

특집논문 (Special Paper)

방송공학회논문지 제31권 제5호, 2026년 9월 (JBE Vol.31, No.5, September 2026)

<https://doi.org/10.5909/JBE.2026.31.5.1006>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

단일 시점 영상 기반 골퍼 간 스윙 자세 유사도 분석

설 동 규^{a)}, 최 유 주^{a)†}

Inter-Golfer Swing Pose Similarity Analysis from Single-View Videos

Dong-Kyu Seol^{a)} and Yoo-Joo Choi^{a)†}

요 약

본 논문은 정면에서 촬영한 두 인물의 단일 시점 골프 스윙 영상을 입력받아 스윙 세그먼트 단위로 자세 유사도를 분석하는 딥러닝 및 규칙 기반 유사도 분석 방법을 제안한다. 기존의 인체 및 골프 클럽 특징점 추정 모델과 스윙 이벤트 추정 모델은 촬영 환경과 골퍼의 자세 변화에 따라 예측값에 결측치와 오류가 발생할 수 있다. 이를 보완하기 위하여 제안 방법은 사전 학습 모델로부터 추정된 인체 및 골프 클럽 특징점에 대해 결측치와 이상치를 보정하고, 운동학적 규칙을 적용하여 스윙 이벤트 시점을 보정한다. 또한, 동적 시간 워핑을 이용하여 두 스윙 영상의 프레임 시퀀스를 정렬한 후, 스윙 세그먼트별 인체 자세 거리와 스윙 이벤트별 코킹 각도 차이를 산출한다. 두 지표를 정규화한 후 가중 합산하여 최종 스윙 유사도 값을 구한다. 프로 골퍼들의 드라이버 스윙 영상을 이용한 실험 결과, 제안 보정 방법을 통하여 특징점 및 이벤트 예측 정확도가 향상되었으며, 코킹 각도를 고려한 유사도 분석이 인체 특징점만을 이용한 방법보다 골퍼 간 자세 차이를 효과적으로 구분함을 확인하였다.

Abstract

This paper proposes a deep learning- and rule-based similarity analysis method that takes two single-view golf swing videos captured from the front and analyzes pose similarity at the swing-segment level. Existing models for estimating human and golf club keypoints and swing events may produce missing or erroneous predictions depending on the shooting environment and variations in golfer's postures. To address these limitations, the proposed method corrects missing values and outliers in the human and golf club keypoints estimated by pretrained models and refines the timing of swing events by applying kinematic rules. In addition, the frame sequences of the two swing videos are aligned using Dynamic Time Warping (DTW), after which the human pose distance for each swing segment and the cocking angle difference for each swing event are calculated. The two metrics are normalized and combined using a weighted sum to obtain the final swing similarity score. Experimental results using driver swing videos of professional golfers showed that the proposed correction methods improved the accuracy of keypoints and swing event predictions. Furthermore, similarity analysis incorporating the cocking angle was found to distinguish differences in golfer's swing poses more effectively than methods based solely on human keypoints.

Keyword : Pose Estimation, Golf Swing Analysis, Similarity Metric, Temporal Segmentation, Rule-based Refinement

1. 서론

인간 자세 추정(Human Pose Estimation, HPE) 기술은 딥러닝과 컴퓨터 비전 분야의 발전과 더불어 큰 성장을 이루었다. 기존에는 마커(Marker)를 부착하거나 특수한 센서에 의존하여 제한된 환경에서 자세 분석을 수행하였다면^[1], 최근에는 단일 카메라로 촬영된 2D 영상만을 기반으로 인체 특징점(Keypoint)을 높은 정확도로 추론한다^[2]. HPE 기술이 발전함에 따라 스포츠 동작 분석 분야로 HPE 적용이 확장되고 있는데, 대표적인 HPE 적용 스포츠 분야로 골프를 들 수 있다. 골프 분야에서 HPE 기술은 골퍼들의 스윙 영상을 기반으로 스윙 자세 정확도를 분석하기 위한 목적으로 활용된다.

기존의 HPE 기반 골프 스윙 분석 연구 및 상용 시스템들은 주로 실내의 모션 캡처 스튜디오 등 특수 환경에서 인체 특징점을 획득하거나, 단일 시점 스윙 영상을 입력으로 받아 인체 특징점을 추정한 후, 인체 특징점 간의 상대적 위치 또는 거리를 사전에 설정된 임계값(Threshold)과 비교하여 스윙 자세에 대한 점수를 계산하는 등 규칙 기반 방식에 의존하였다.

딥러닝 기술이 발전함에 따라 골프 스윙을 분석하기 위한 다양한 딥러닝 모델들이 발표되었는데, 대표적인 모델로 단일 프레임 시퀀스에서 자세가 크게 변화하는 시점인 스윙 이벤트를 예측하는 SwingNet^[3] 모델, 골프 클럽의 특징점을 추정하는 GolfPose^[4] 모델, 동적 시간 워핑(Dynamic Time Warping) 알고리즘으로 두 스윙 자세 시퀀스를 정렬한 후 인체 특징점 간 거리 계산을 통하여 서로 다른 스윙 자세를 비교 분석하는 방법^[5] 등이 제안되었다.

이러한 기존 모델들을 이용하여 유튜브 등을 통해 수집된 골프 스윙 영상을 적용하여 스윙 이벤트 예측, 골프 클럽 특징점 예측을 수행하는 경우, 예측 정확도의 개선이 필요

함을 확인할 수 있다. 이는 제한된 실험 환경에서 획득된 영상 데이터를 기반으로 학습된 것에 기인할 수 있다. 예측 정확도를 높이기 위해서 보다 다양한 환경에서 획득된 학습 데이터를 구축하여 파인 튜닝을 수행하는 방법을 적용할 수 있으나 데이터 구축에 많은 시간과 비용이 요구된다.

이에 본 논문에서는 두 골퍼의 스윙 자세를 구간별로 구분하여 상세하게 유사도를 분석하기 위하여, 사전 학습 모델로부터 추정된 인체 및 골프 클럽 특징점에 대해 결측치와 이상치를 보정하고, 운동학적 규칙을 적용하여 스윙 이벤트 시점을 보정하는 방법을 제안한다. 또한, 제안 방법에서는 스윙 이벤트 추정, 인체 및 골프 클럽 탐지, 동적 시간 워핑 알고리즘을 단일 구조로 통합하고, 왼쪽 손목의 코킹 각도(Cocking Angle)에 대한 차이값을 계산하여 인체 특징점뿐만 아니라 손목 코킹 각도가 포함된 스윙 유사도 측정 방법을 적용한다. 이때, 골프 스윙 구성의 특징을 기반으로, 스윙 이벤트, 인체 및 골프 클럽 추정 정확도를 높이기 위한 규칙 기반 알고리즘을 제안한다. 제안 방법은 다음의 세부 과정을 수행한다.

첫째, 골프 스윙 영상을 사전 학습(Pre-train)된 모델에 입력하여 그 프레임 시퀀스를 7개의 스윙 세그먼트로 분할하고, 인체 및 골프 클럽의 특징점을 추출한다.

둘째, 모션 블러(Motion Blur)와 자기 폐색(Self Occlusion)으로 인하여 발생하는 골프 클럽 특징점 및 스윙 이벤트 예측 오류를 완화하기 위하여, 각 스윙 세그먼트별 클럽 샤프트 길이의 제3사분위 수를 계산하고 이를 바탕으로 클럽 헤드 특징점의 이상치 및 결측치를 탐지한 다음, 극좌표계 기반 3차 스플라인을 적용하여 모든 이상치와 결측치를 보간한다. 이렇게 보정된 클럽 특징점 좌표 시퀀스를 활용하여 기존 모델의 부정확했던 스윙 이벤트 추정 결과를 규칙 기반으로 재조정한다.

셋째, 보정된 특징점 좌표 시퀀스를 기반으로 동적 시간 워핑 알고리즘을 통하여 두 영상의 스윙 자세 시퀀스를 정렬함으로써 최적의 프레임 대응 관계를 찾는다. 골프 스윙의 운동학적 특성을 반영한 자세 비교 지표를 생성하기 위하여, 왼팔과 골프 클럽 샤프트가 이루는 각도를 코킹 각도로 정의하고 두 영상 간의 코킹 각도 차이와 인체 특징점 간 유클리드 거리(Euclidean Distance)를 계산한다. 두 지표에 대하여 정규화 및 가중 평균을 수행하여 유사도 점수를

a) 서울미디어대학원대학교 인공지능융합소프트웨어학과(Dept. of AI Software Engineering, Seoul Media Institute of Technology)

✉ Corresponding Author : 최유주(Yoo-Joo Choi)

E-mail: yjchoi@smit.ac.kr

Tel: +82-2-6953-3154

ORCID: <https://orcid.org/0000-0001-7520-097X>

· Manuscript received August 26, 2026; Revised September 9, 2026;

Accepted September 9, 2026.

계산한다. 스윙 영상 간의 유사도 점수를 전체 평균 및 스윙 세그먼트별로 도출하여, 스윙을 구성하는 세부 자세에서의 유사도를 분석한다.

제안 방법에 대한 성능 검증은 서로 다른 프로 골퍼 5명의 스윙 영상을 이용하여 동일 선수 간 및 다른 선수 간 스윙 자세 유사도 결과를 구하는 실험을 통하여 수행한다. 실험 결과로 제안 방법 적용 후, 스윙 이벤트 예측, 인체 및 클럽 특징점 추정 정확도 향상을 입증하고, 인체 특징점만을 기반한 스윙 자세 비교에 비해 인체 특징점 간 거리와 골프 클럽 각도 차이를 가중합하는 방식이 골프 스윙의 정량적 비교 분석에 유효하게 작용함을 증명한다.

본 논문은 총 5개의 장으로 구성되며, 본 장 이후 내용은 다음과 같다. 먼저 2장에서는 연구의 이론적 배경이 되는 스포츠 분야, 특히 골프 스윙 자동 분석 선행 연구들의 특징을 살펴본다. 3장에서는 본 연구에서 제안하는 방법의 처리 절차를 설명하고, 입력 영상 전처리, 스윙 세그멘테이션, 특징점 추정, 규칙 기반 후처리, 유사도 계산 기법 구조를 상세하게 기술한다. 4장에서는 시스템 성능 검증에 관한 실험 구성과 결과를 서술한다. 마지막으로 5장에서는 연구의 기여점 및 한계점과 향후 연구 방향을 제시한다.

II. 관련 연구

1. 스포츠 분야에서의 인간 자세 추정

인간 자세 추정(Human Pose Estimation, HPE)은 이미지 프레임에서 인체 특징점의 위치를 예측하는 기술이다. 초기의 HPE 기술은 주로 수동 설계한 특징(Feature) 기반 인체 구조 모델링에 의존하였는데, 조명 등 외부 환경의 변화에 영향을 크게 받고 특이한 자세에 대한 제한적인 강건성을 보여 정확도 향상이 필요했다^[6]. 이후 GPU 등 하드웨어 성능이 발전하였고, 나아가 COCO^[7] 및 Human3.6M^[8]과 같은 대규모 데이터셋이 등장하며 입력 데이터에서 특징을 자동 학습하는 딥러닝 모델의 발전을 촉진하였다.

HPE는 인간-컴퓨터 상호작용(HCI) 분야뿐만 아니라 연속적인 동작에 대한 분석이 필요한 스포츠 분야에서도 널리 활용되고 있다. 스포츠는 일상적인 인간 자세와 달리 빠

르고 복잡한 동작으로 구성되므로 범용 HPE 모델을 적용할 경우 자세 추정 정확도가 감소한다. 이러한 문제를 해결하기 위하여 각 스포츠 종목의 특수성을 반영한 HPE 연구가 진행되고 있다. 이에 대한 예시로, 운동 훈련 보조를 위한 딥러닝 기반 시스템인 AI Coach^[9]는 스포츠 영상 속 인물의 동작에 대한 실시간 피드백 제공을 목표로 한다. 이 시스템은 자체 HPE 모델로 특징점을 예측하고 교정 필요 유무에 따라 분류한 다음, 교정이 필요한 경우 표준 자세의 예시 영상을 사용자에게 보여주는 구조다. 정교한 라켓 움직임을 파악함이 중요한 탁구에서는 손목, 팔꿈치, 양쪽 어깨로 구성된 특징점 좌표를 예측하고 이를 스매시, 드라이브 등 11가지의 패턴으로 분류하는 연구^[10]가 제시되었다. 테니스 분야의 경우는 인물의 특징점 좌표를 추출하고 서브, 포핸드, 백핸드 등 동작 구간을 BoW(Bag of Words) 방식으로 분류함으로써 샷의 성공 확률을 분석하는 방법론이 제시되었다^[11]. 단일 깊이 이미지로 3D 골프 스윙 자세를 추정하는 연구^[12]에서는 랜덤 포레스트 방식으로 특징점 위치를 예측하고, 해당 예측값의 신뢰도 값을 검증 포레스트로 평가하였다. 현재 골프 스윙 분석을 주제로 하는 연구는 인체 특징점을 추정하고 이를 기반으로 자세를 분석하는 연구들과 클럽의 특징점을 추정하고 특징점 데이터의 분포 및 시간에 따른 움직임을 기반으로 자세를 분석하는 연구들로 구분해 볼 수 있다.

2. 인체 특징점 추정 기반 골프 스윙 자세 분석

초기의 인체 특징점 기반 골프 스윙 분석 연구는 키넥트(Kinect) 등 Time-of-Flight(ToF) 센서가 있는 특수 장비로 수집한 좌표 데이터를 주로 기반하였다. J. Park 등^[1]은 좌표 데이터에 허프 변환(Hough Transform)을 적용하여 직선으로 표현된 골프 클럽을 검출하고, 클럽 헤드의 위치, 속도, 이동 방향을 바탕으로 골프 스윙을 7개의 구간으로 분류한 연구를 수행하였다. H. Kim 등^[13]은 양쪽 손의 순간 속도와 골반, 어깨, 손의 상대 좌표를 수집해 골프 스윙을 4개의 구간으로 분류한 연구를 수행하였다. L. Zhang 등^[14]은 인체 특징점 데이터를 SC-GMM(Serial Correlation Gaussian Mixture Model)에 입력하여 각 스윙 구간별 자세 점수를 계산한 후 전문가 스윙의 분포를 모델링한

GMM(Gaussian Mixture Model)으로 로그-우도 점수를 계산하고, 이를 SVM(Support Vector Machine)에 입력하여 스윙 등급을 분류한 연구를 수행하였다.

딥러닝 모델을 적용한 골프 스윙 분석 연구로, 족압 센서로 측정된 스윙 시의 압력 중심(Center of Pressure) 데이터 시퀀스를 이미지로 변환하고 CNN(Convolutional Neural Network)에 입력하여 초보자과 숙련자로 분류하는 연구^[15]가 진행되었다. K. Ko 등^[2]은 단일 카메라로 촬영한 2차원 이미지 시퀀스를 CNN에 입력하여 스윙 구간을 분할하고 자세 특징을 추출한 다음, 이를 Bi-LSTM 모델에 입력하여 몸통의 좌우 움직임(Sway), 머리의 상하 움직임(Head-Up), 스윙의 정점에서 어깨선과 골반 선이 회전한 각도의 차이(X-factor)와 같은 3차원의 동작 지표를 예측하는 연구를 발표하였다. HPE 모델의 특징점 추정 결과에 대한 보정은 이상치(Outlier)를 Bi-LSTM 모델로 탐지하고 선형 보간함으로써 개선하는 방식이 제시되었다^[16]. 또한, DEKR^[17]이나 HRNet^[18]과 같은 HPE 모델을 골프 데이터를 이용하여 파인튜닝(Fine-tuning)한 GolfMate^[19] 모델이 제시되었다. AI Golf^[20]에서는 MobileNetV2^[21]와 Bi-LSTM을 결합한 SwingNet^[3]으로 스윙 이벤트를 예측한 후, MMPose로 각 스윙 이벤트별 인체 특징점을 추정하고, 해당 특징점 이미지 8장을 수평 결합한 단일 이미지를 ResNet-50 모델에 입력하여 스윙 수준을 이진 분류하였다.

이와 같은 선행 연구들은 주로 스윙 영상의 프레임 시퀀스 간 정렬 방법론 또는 인체 특징점 간의 위치 차이 분석 방법론에 집중되어 있다. 실제 골프 스윙에서 자세의 수준을 가늠하는 척도인 코킹 각도 등은 반영하지 않은 한계가 존재한다.

3. 골프 클럽 특징점 추정 기반 골프 스윙 자세 분석

정확한 골프 스윙 분석을 위해서는 인체뿐만 아니라 골프 스윙을 구성하는 요소인 골프 클럽의 특징점을 추정하는 것이 필요하다. 이때 전통적인 알고리즘 기반 방식은 샤프트나 클럽 헤드를 추정함에 있어 모션 블러나 복잡한 배경을 이유로 추정 정확도가 떨어지는 문제가 존재한다.

골프 클럽을 추정하는 연구 초기에는 이러한 한계를 극

복하기 위하여, 인체 특징점과 함께 1-2개의 골프 클럽 특징점을 수작업으로 추가한 데이터셋을 구축하여 기존 HPE 모델을 추가 학습하는 것을 목표로 하였다. D. Lee 등^[22]은 자세 추정을 위한 모션 캡처 스튜디오 환경의 접근성을 지적하며, GolfDB 데이터셋으로 사전 학습된 Faster R-CNN과 HRNet으로 인체 특징점 17개와 클럽 헤드 특징점 1개를 추정하여 데이터셋을 구축하는 방법을 제시하였다. J. Kang 등^[23]은 골프 스윙 영상에서 HRNet으로 인체 특징점 17개를 추정하고, 클럽 그림과 클럽 헤드 특징점을 수동 추가한 데이터셋으로 HigherHRNet^[24]을 학습하여 클럽 특징점 추정 시스템을 구축하였다.

이후 GolfPose^[4]은 샤프트와 헤드를 포함한 5개의 골프 클럽 3D 특징점을 추정하였다. GolfPose는 240fps로 촬영된 골프 스윙 영상과 2D 및 3D 특징점 좌표로 구성되는 GolfSwing 데이터셋을 구축하여 Faster R-CNN과 YOLOX-s^[25] 및 2D HPE 모델인 HRNet, ViTPose, DEKR, 그리고 2D 특징점 좌표를 3D로 변환하는 3D 리프팅 모델인 MixSTE^[26]를 각각 파인튜닝하였다. 이를 통하여 GolfPose 모델은 인물(Golfer), 골프 클럽(Club), 그리고 두 객체를 한 클래스로 합친 “Golfer-with-club”으로 바운딩 박스 추정 대상을 다르게 할 수 있으며 2D 특징점 추정 단계에서도 객체를 사람(Person), 골프 클럽(Club) 또는 사람과 골프 클럽(Golfer)으로 나눠 추정하도록 하였다.

본 논문에서는 서로 다른 골퍼의 스윙 자세 간 유사도를 정량적으로 분석하기 위해, 인체 특징점과 골프 클럽의 특징을 함께 고려하는 “골프 스윙 유사도 측정 시스템”을 제안한다.

III. 제안 시스템 설계

1. 시스템 개요

제안 시스템은 [그림 1]과 같이 2명의 골퍼에 대한 단일 시점 영상 2개를 입력받아, 영상별로 프레임 단위로 분할하여 하나의 시퀀스를 구성한 후 처리된다. 주요 수행 단계는 데이터 전처리(Frames Pre-processing) 단계, 딥러닝 모델 기반 스윙 이벤트 추정(Swing Events Detection) 단계, 딥러

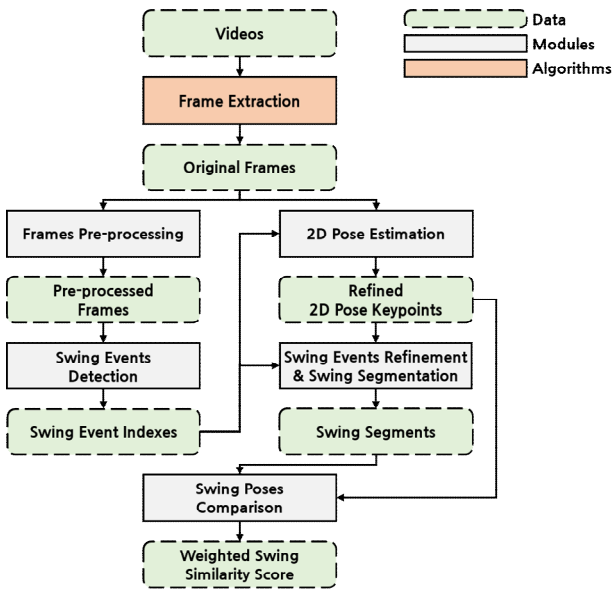


그림 1. 제안 시스템의 전체 구조도
Fig. 1. Overall architecture of the proposed system

닝 모델 기반 인체 및 골프 클럽 특징점 추정(2D Pose Estimation) 단계, 규칙 기반 스윙 이벤트 보정(Swing Events Refinement & Swing Segmentation) 단계, 인체 특징점 간 유클리드 거리 및 코킹 각도 차이 기반 스윙 유사도 측정(Swing Poses Comparison) 단계로 구분된다. 3절의 세부 섹션에서는 주요 수행 단계에 대하여 설명한다.

2. 데이터 전처리

제안 시스템의 첫 단계는 [그림 2]와 같이, 영상의 이미지 프레임으로 구성된 시퀀스를 규격화하는 데이터 전처리 과정이다. 제안 시스템은 SwingNet의 백본(Backbone)인 MobileNetV2^[21]가 스윙 동작의 특징을 안정적으로 포착하도록, GolfSwing^[4]으로 미세 조정(Fine-tuning)된 YOLOX-s^[25] 모델로 인물에 대한 바운딩 박스를 탐지한 후, 프레임 내 인물 영역을 분리하는 크롭핑(Cropping), 영상 비율에 맞추어 프레임 크기를 조정하는 리사이징(Resizing), 나머지 영역을 채우는 패딩(Padding), 픽셀값을 정규화(Normalization)하는 과정을 수행한다. 각 단계별 내용은 다음과 같다.

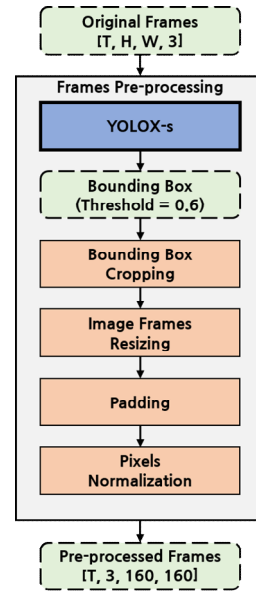


그림 2. 제안 시스템의 데이터 전처리 절차
Fig. 2. Data pre-processing procedure of the proposed system

먼저 YOLOX-s 모델로 인물에 대한 바운딩 박스를 탐지할 때는, 신뢰도가 0.6 미만인 경우 필터링하며, 한 프레임에 여러 인물이 탐지되면 가장 높은 신뢰도의 바운딩 박스를 해당 프레임의 인물 영역으로 결정한다.

크롭핑 단계에서는 YOLOX-s 모델이 예측한 바운딩 박스 좌표를 기준으로 프레임마다 인물 영역을 자른다. 리사이징 시, 이중 선형 보간(Bilinear interpolation) 방식으로 원본 프레임의 해상도 비율을 유지하면서 긴 쪽을 160픽셀에 맞춰 크기를 조정한다. 패딩 단계는 SwingNet 모델의 학습 과정과 유사하게, 이미지의 빈 영역에 ImageNet^[27] 데이터셋의 평균 BGR 색상 값을 채워 넣는다. 정규화 시, 각 프레임은 RGB 순으로 전환 후 [0.0, 1.0] 범위의 텐서로 변환되며, ImageNet의 평균 및 표준편차를 사용하여 픽셀값을 정규화한다. 이로써 개별 프레임은 (160, 160, 3) 해상도의 이미지가 된다.

3. 딥러닝 모델 기반 스윙 이벤트 추정

다음으로 스윙 이벤트 추정 단계는 “스윙 세그멘테이션”을 수행한다. 스윙 세그멘테이션은 단일 골프 스윙의 전체

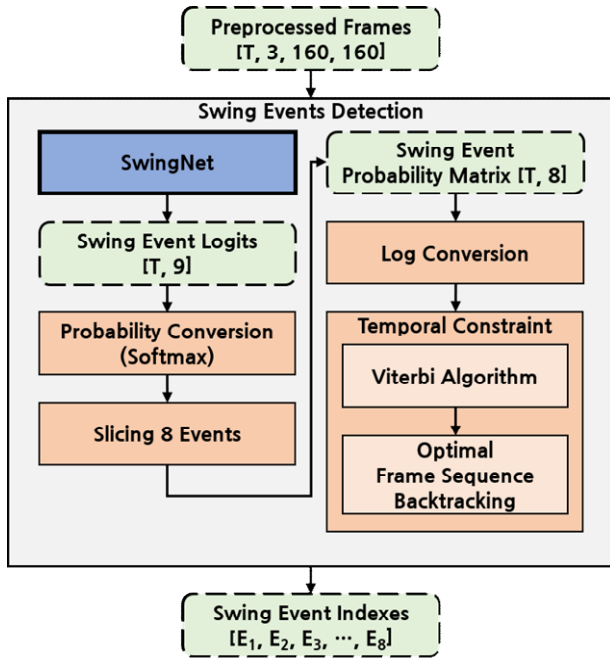


그림 3. 제안 시스템의 스윙 이벤트 추정 절차
Fig. 3. Swing events detection procedure of the proposed system

동작 시퀀스를 물리적 자세 변화에 따른 특정한 기준을 바탕으로 분할한 구간인 “스윙 세그먼트”를 구성하는 과정이다. 본 연구에서는 SwingNet 모델로 골프 스윙 자세의 변화 시점인 “스윙 이벤트”의 프레임 인덱스 8개를 추정하고, 인접한 스윙 이벤트 사이의 연속하는 프레임을 묶어 하나의

스윙 세그먼트로 정의하였다.

SwingNet의 기본 구조는 다음과 같다. 먼저 정규화된 프레임 시퀀스 텐서가 입력되면 MobileNetV2의 컨볼루션 필터로 각 텐서의 공간적 특징 벡터를 추출한다. 다음으로 이를 시퀀스의 순방향 및 역방향 정보가 학습된 Bi-LSTM에 입력하여 프레임 간 시간적 문맥을 모델링한다. Bi-LSTM을 통과한 특징 벡터 시퀀스는 완전 연결 층을 거쳐 Softmax 함수로 전달되어, 각 프레임은 Address, Toe-up, Mid-back-swing, Top, Mid-downswing, Impact, Mid-follow-through, Finish로 구성되는 스윙 이벤트 8개와 어느 스윙 이벤트에도 해당하지 않음을 의미하는 “No-event”를 포함하여 9개 클래스로 구분한다. SwingNet은 9개 클래스에 해당할 확률을 담은 [전체 프레임 수×9] 형태의 행렬을 출력한다. 제안 시스템에서는 [그림 3]과 같이, 사전 학습된 SwingNet 모델을 통하여 각 프레임별 개별 스윙 클래스에 속할 확률값을 획득하고, 이를 기반으로 스윙 이벤트 프레임 인덱스를 추출한다. 이때 기존 SwingNet 모델의 연구에서는 각 이벤트일 확률이 가장 높은 프레임을 순차적으로 개별 선택하는데, 골프 스윙 순서상 이후에 발생해야 할 이벤트가 그 전 이벤트보다 먼저 예측되는 오류가 발생할 수 있다. 따라서 본 시스템은 비터비 (Viterbi) 알고리즘의 형식대로 “각 이벤트는 반드시 이전 이벤트보다 시간적으로 뒤에 위치해야 한다”는 제약을 설정하고, 첫 번째 프레임부터 마지막 프레임까지 로그(Log) 확률의 누적합을 계산하여, 누적합이 최대가 되는 시퀀스 경로를

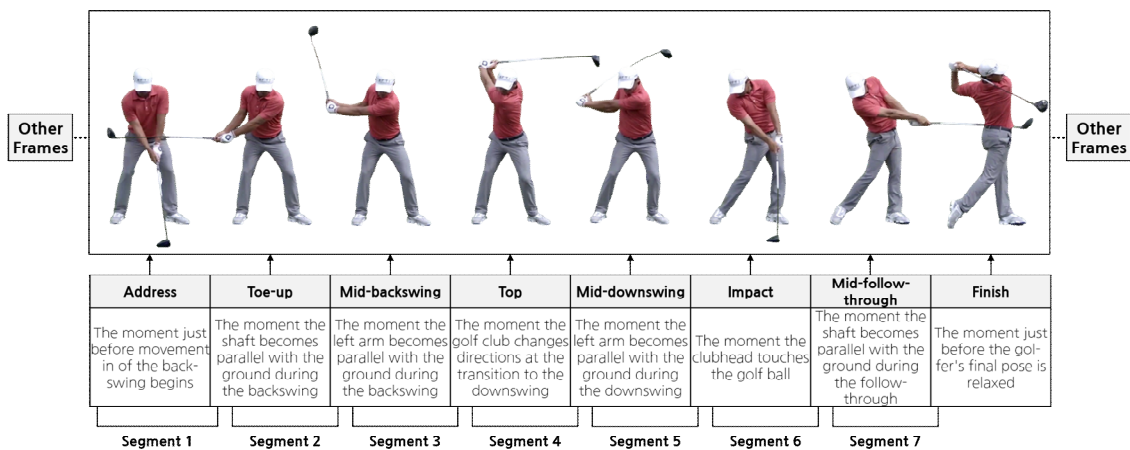


그림 4. 스윙 이벤트와 스윙 세그먼트의 정의
Fig. 4. Definitions of swing events and swing segments

역추적함으로써 스윙 이벤트가 실제 동작 순으로 올바르게 나열된 프레임 인덱스 배열을 출력한다.

출력된 스윙 이벤트에 대한 인덱스들은 골프 스윙 영상을 7개의 스윙 세그먼트로 분할하는 기준이 된다. 각 스윙 세그먼트의 범위는 한 스윙 이벤트의 인덱스부터 다음 스윙 이벤트의 직전 인덱스까지로 정의한다. 이에 따라 Segment 1은 “Address” 이벤트부터 “Toe-up”의 직전 프레임까지며, Segment 7은 “Mid-follow-through”부터 “Finish”까지의 구간으로 설정된다. “Address” 이벤트의 이전 프레임과 “Finish” 이벤트의 이후 프레임들은 스윙 세그먼트이션 대상에서 제외하였다. 정의된 스윙 이벤트와 스윙 세그먼트는 [그림 4]와 같다.

4. 딥러닝 모델 기반 인체 및 골프 클럽 특징점 추정

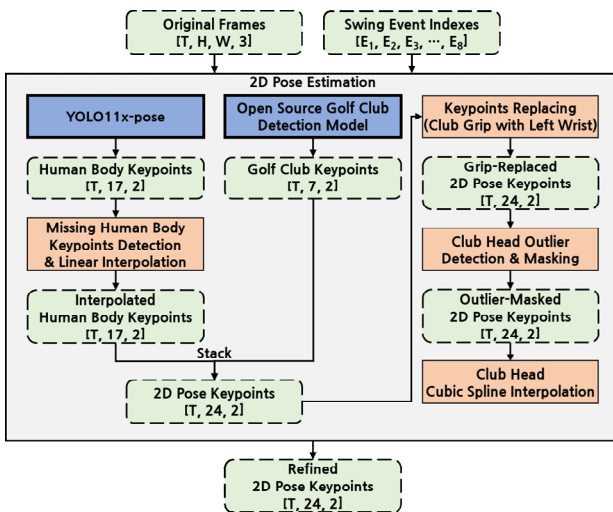


그림 5. 제안 시스템의 인체 및 골프 클럽 특징점 추정 절차

Fig. 5. 2D pose estimation procedure of the proposed system

스윙 이벤트 인덱스가 예측되면, 제안 시스템은 [그림 5]와 같이 인체와 골프 클럽을 검출함에 각각 특화된 딥러닝 모델로 원본 프레임 시퀀스에서 2D 특징점을 예측한다. 이때 과정은 크게 스윙 자세 추정 단계와 인체 및 골프 클럽 특징점 결측치 보간 단계로 구성된다.

자세 추정 단계에서는 먼저 YOLO11x-pose^[28] 모델과 오픈소스 골프 클럽 탐지 모델^[29]에 스윙 영상의 원본 프레임을 각각 입력한다. YOLO11x-pose 모델은 프레임마다 인

체의 바운딩 박스를 탐지하고, 해당 영역 내에서 2D 인체 특징점 17개를 추정한다. 오픈소스 골프 클럽 탐지 모델은 골프 클럽의 손잡이를 의미하는 그립(Grip)부터 본체 부분인 헤드(Head)까지로 구성되는 2D 골프 클럽 특징점 7개를 추정한다. 이후 수직 스택(Vertical Stack) 연산으로 인체와 골프 클럽 특징점 시퀀스를 병합하여, 프레임당 24개의 2D 특징점 좌표로 구성되는 시퀀스를 형성한다.

인체 및 골프 클럽 특징점 결측치 보간 단계는 빠른 스윙으로 발생하는 모션 블러와 신체 회전에 따른 폐색 문제를 해결하기 위한 목적을 가진다. 특히 인체 특징점 추정 시 회전 동작에 의하여 오른쪽 상체에 해당하는 특징점이 가려져 결측되거나, 골프 클럽 특징점 추정 시 모델 성능에 한계가 있어 실제와 다른 위치에 특징점이 추정되었다. 더불어 이러한 추정 오류를 보정함에 있어 규칙 없이 선형 보간 등 수학적인 방법만을 적용했을 때, 보간된 특징점의 흐름이 실제 스윙 궤적과 반대되는 현상을 확인하였다. 따라서 제안 시스템에서는 다음의 순서대로 인체 및 골프 클럽 특징점 추정 모델 각각의 결과 데이터를 보정하도록 동작 구조를 설계하였다.

첫째, 결측된 인체 특징점 좌표를 탐색하고, 양쪽 골반 중앙 지점을 인체의 회전 중심점으로 설정한 다음, 직교좌표계를 극좌표계로 변환한다. 이후 SwingNet 모델이 추정한 “Top” 이벤트 인덱스를 기준으로, 전체 프레임 시퀀스를 “백스윙(Backswing)”과 “다운스윙(Downswing)” 구간으로 구분하여 각도 변화량에 따라 연속된 특징점들이 백스윙 구간에서는 시계 방향으로, 다운스윙 구간에서는 반시계 방향으로 회전하도록 제약을 둔다. 방향이 교정된 각도(Radian)와 반지름 값에 대하여 선형 보간(Linear Interpolation)을 수행하고 이를 다시 직교좌표계로 복원하여 완전한 인체 궤적을 구성한다. 마지막으로 인체에 비하여 추정 성능이 불안정한 클럽 그립 특징점의 좌표를 왼쪽 손목 좌표로 모든 프레임에서 대체한다.

둘째, 영상에서 스윙 영역별로 보여지는 클럽 샤프트의 평균 길이를 계산하고 이를 기반으로 추정 클럽 특징점의 이상치를 탐지하기 위하여, 먼저 비디오의 0번째 프레임부터 “Finish” 이벤트 인덱스까지의 시퀀스를 [표 1]과 같이 6개의 섹션(Section)으로 분할한다. 각 섹션을 구성하는 프레임마다 “왼쪽 손목부터 클럽 헤드까지의 유클리드 거리”

표 1. 클럽 헤드 이상치 판별을 위한 섹션의 범위
Table 1. Section ranges for club head outlier detection

Section	Start Frame	End Frame
Section 1	Frame 0	Address index
Section 2	Address index + 1	Mid-backswing index
Section 3	Mid-backswing index + 1	Top index
Section 4	Top index + 1	Impact index
Section 5	Impact index + 1	Mid-follow-through index
Section 6	Mid-follow-through index + 1	Finish index

를 “샤프트 길이”로 상정하여 계산한다. 이때 섹션별 샤프트 길이의 제3사분위 수(Q3)를 기준으로 특정 프레임의 샤프트 길이가 Q3의 0.7배 미만이거나 1.3배를 초과할 경우 이상치로 특정하고 해당 프레임에서의 클럽 헤드 특징점 좌표를 [0, 0] 형식의 결측치로 변환한다.

셋째, 결측 처리된 프레임과 기존에 결측된 프레임에 보간을 적용한다. 이는 결측 프레임 전후의 정상 추정된 특징점 벡터를 극좌표계의 각도로 변환하고, 인체 보간과 동일한 방식으로 백스윙과 다운스윙 시의 회전 방향 규칙을 설정한 다음, 시간 축에 따른 각도 변화량을 3차 스플라인 곡선 함수에 대입하여 결측 프레임의 회전 각도를 추론하며, 이 각도에 섹션별 샤프트 길이의 Q3 값을 곱하여 직교좌표계 상의 클럽 헤드 좌표를 복원하는 방식이다.

5. 규칙 기반 스윙 이벤트 보정

규칙 기반 스윙 이벤트 보정 단계에서는 [그림 6]의 구조

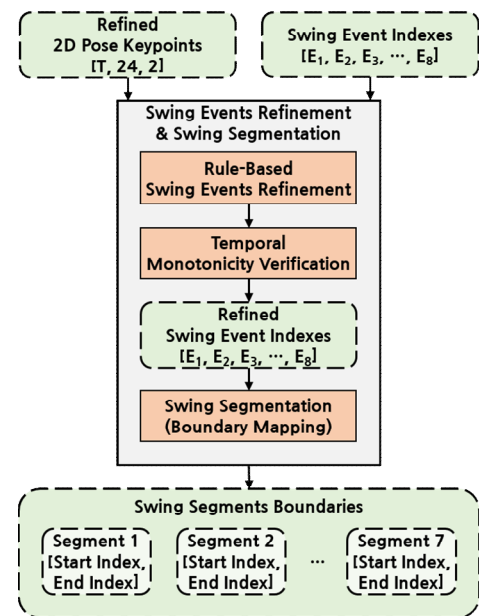


그림 6. 제안 시스템의 스윙 이벤트 보정 절차
Fig. 6. Swing events refinement procedure of the proposed system

표 2. 스윙 이벤트별 인덱스 보정 규칙
Table 2. Index refinement rules for swing events

Order	Swing event	Refinement rules
1	Address	The frame immediately preceding the point where the inter-frame displacement of the club head, calculated using Euclidean distance, exceeds 3 pixels
2	Toe-up	The frame in which the difference between the club head's Y-coordinate and the average Y-coordinate of both wrist points is minimized during the backswing
3	Mid-backswing	The frame where the difference between the Y-coordinates of the left elbow and the left wrist is minimized
4	Top	The frame where the Y-coordinates of the grip, club head, and both wrists reach their peaks and simultaneously begin to descend
5	Mid-downswing	The frame during the club's descent phase where the difference between the Y-coordinates of the left elbow and the left wrist is minimized
6	Impact	The frame exhibiting the maximum Y-coordinate of the club head within the interval where it is positioned below both knees
7	Mid-follow-through	The frame where the Y-coordinate difference between the club head and the grip is minimized, while the center points of both shoulders, elbows, wrists, and the club head are sequentially oriented toward the right side of the screen
8	Finish	The frame where the club head descends after reaching its peak, minimizing the difference between its Y-coordinate and the Y-coordinate of the center point between both shoulders

대로, 결측치 보간을 통하여 완성된 인체 및 골프 클럽 특징 점 시퀀스와 [표 2]와 같이 설정한 규칙을 기반으로 SwingNet 모델이 추정된 스윙 이벤트 인덱스를 보정한다. 그 보정 결과가 스윙 이벤트 발생 순서에 맞는지 검증한 후, 스윙 세그먼트별 시작 지점과 종료 지점의 인덱스를 설정하여 스윙 세그먼트이션을 수행한다.

[표 2] 규칙은 [그림 4]와 같이 SwingNet 연구에서 정의한 스윙 이벤트별 자세를 바탕으로 설계한 보정 규칙이다. 제안 시스템에서는 보정 규칙에 부합하는 프레임을 탐색한 다음 각 이벤트의 인덱스를 갱신하여, 시퀀스 분할의 기준으로써 총 7개의 스윙 세그먼트를 설정한다. 이와 같은 규칙 기반 스윙 이벤트 보정 절차를 통하여, 기존에 SwingNet 모델이 슬로우 모션(Slow-Motion)으로 촬영되지 않았거나 배경이 복잡한 골프 스윙 영상을 입력받았을 때, [그림 4]의 정의에 정합하지 않는 프레임을 스윙 이벤트로 추정하여, 두 스윙을 비교할 때 동일한 스윙 이벤트일지라도 실제로는 서로 다른 스윙 자세를 비교하게 되어 부정확한 스윙 유사도 평가가 되는 문제를 보완하고자 하였다. 각 스윙 이벤트별 프레임 인덱스 보정 구조는 다음과 같다.

$$D_{ClubHead}^t = \| P_{ClubHead}^t - P_{ClubHead}^{t-1} \| \quad (1)$$

$$\hat{t}_{Address} = t - 1 \text{ (if } D_{ClubHead}^t > D_{threshold} \text{)} \quad (2)$$

정지 상태에서 백스윙이 시작되는 “Address” 이벤트는 영상의 0번째 프레임부터 SwingNet이 추정된 “Toe-up” 이벤트 인덱스 직전까지의 구간에서 클럽 헤드가 속도 임계값 이상으로 움직이는 첫 프레임을 찾아 그 직전 프레임으로 인덱스를 보정한다. 현재 프레임 t 와 직전 프레임 $t-1$ 에서의 클럽 헤드 좌표를 각각 $P_{ClubHead}^t, P_{ClubHead}^{t-1}$ 라 할 때, 인접 프레임 간 클럽 헤드 위치 변화량인 $D_{ClubHead}^t$ 를 식 (1)과 같이 계산한다. 이후, $D_{ClubHead}^t$ 가 임계값 $D_{threshold}=3.0$ 픽셀을 최초로 초과할 때 식 (2)와 같이 해당 프레임의 직전을 새로운 인덱스 $\hat{t}_{Address}$ 로 설정한다. 이후의 모든 스윙 이벤트 보정 과정에서는 직전 이벤트로 확정된 프레임의 다음 인덱스부터 탐색하여 시간상 역방향으로 스윙 이벤트가 결정되는 문제를 방지한다.

$$\hat{t}_{ToeUp} = \operatorname{argmin}_t |P_{ClubHead,y}^t - P_{grip,y}^t| \quad (3)$$

이후, 백스윙 구간에서 골프 클럽의 샤프트가 지면과 수평이 되는 “Toe-up”과 골퍼의 왼팔이 지면과 수평이 되는 “Mid-backswing”을 식 (3)과 같이 순차적으로 보정한다. “Toe-up”의 경우 클럽 헤드의 Y좌표 $P_{ClubHead,y}^t$ 와, 왼쪽 손목으로 대체한 상태인 클럽 그림의 Y좌표 $P_{grip,y}^t$ 간 값의 차이가 가장 작은 프레임 $\hat{t}_{ToeUp}$ 으로 보정한다. 이때 함수를 최솟값으로 만드는 프레임 인덱스 t 를 반환하는 argmin_t 를 적용한다. 이와 동일하게, “Mid-backswing”은 왼쪽 팔꿈치의 Y좌표와 왼쪽 손목의 Y좌표 값 차이가 최소화되는 프레임을 찾은 후 해당 인덱스로 스윙 이벤트 인덱스를 보정한다.

$$\hat{t}_{Top} = \min \left\{ t \mid \begin{array}{l} P_{grip,y}^t < P_{grip,y}^{t+1}, \\ P_{ClubHead,y}^t < P_{ClubHead,y}^{t+1}, \\ P_{wrist,u}^t < P_{wrist,u}^{t+1} \end{array} \right\} \quad (4)$$

“Top” 이벤트 인덱스는 식 (4)처럼 화면상 양쪽 손목이 머리보다 높은 곳에 위치하는 조건 $P_{wrist,y}^t < P_{face,y}^t$ 를 만족하는 프레임 구간 t 내에서, 클럽 궤적이 다운스윙 방향으로 변화하는 정점 이후 시점 중 클럽 그림과 클럽 헤드의 Y좌표인 $P_{grip,y}^t$ 와 $P_{ClubHead,y}^t$, 양쪽 손목의 평균 Y좌표 $P_{wrist,y}^t$ 값이 모두 동시에 증가하는 첫 프레임 $\hat{t}_{Top}$ 으로 보정한다.

$$\hat{t}_{Impact} = \operatorname{argmax}_t P_{ClubHead,y}^t \quad (5)$$

그 후, 다운스윙 시 왼팔이 지면과 평행을 이루는 이벤트 “Mid-downswing”은 화면상 클럽 헤드가 머리보다 높고, 인체의 우측에 있으며, 양쪽 손목이 모두 왼쪽 발목 및 왼쪽 팔꿈치보다 화면상 좌측에 위치할 때, 왼쪽 팔꿈치와 왼쪽 손목의 Y좌표 차이가 최소화되는 지점을 찾는다. 클럽 헤드가 골프공에 접촉하는 순간인 “Impact” 이벤트는 클럽 헤드가 양쪽 무릎보다 아래에 있는 프레임 중, 식 (5)와 같이 클럽 헤드의 Y좌표가 최댓값을 가지는 프레임 $\hat{t}_{Impact}$

를 새로운 스윙 이벤트 인덱스로 선택한다. “Mid-follow-through” 이벤트는 “Impact” 이후의 프레임 시퀀스 중 클럽 샤프트가 지면과 평행을 이루는 시점이어야 하므로, 양쪽 어깨, 양쪽 팔꿈치, 양쪽 손목의 중앙 지점, 클럽 헤드가 수평 정렬된 상태에서 클럽 헤드와 그림의 Y좌표 차이가 최소인 프레임으로 보정한다. 이어 골퍼의 최종 자세가 이완되기 직전인 “Finish”는 스윙 정점 이후 클럽 헤드가 양쪽 어깨의 평균 Y좌표 값보다 높아지는 시점부터, 양쪽 어깨의 평균 Y좌표와 클럽 헤드의 Y좌표 간 차이가 최소화되는 지점을 발견한다.

제안 시스템은 위와 같은 스윙 이벤트별 규칙 기반 보정 절차를 통하여 SwingNet이 추정하였던 스윙 이벤트 인덱스들을 실제 골퍼 스윙 동작의 변화 시점에 부합하도록 스윙 이벤트 인덱스 시퀀스를 보정하고, 기존의 원본 SwingNet 모델의 확률적 예측 구조를 개선하였다.

6. 유클리드 거리 및 코킹 각도 기반 스윙 유사도 측정

마지막으로 스윙 유사도 측정 절차는 [그림 7]과 같이 전체 시퀀스에서의 인체 자세 간 유클리드 거리를 계산하고, 보정된 스윙 이벤트 프레임에서 영상 간 “코킹 각도”의 차이를 계산한다. 유클리드 거리 값과 코킹 각도 차이값을 각각 0-1 사이의 값으로 정규화한 다음, 두 값을 더하여 각 스윙 세그먼트별 스윙 유사도와 전체 평균 스윙 유사도를 계산한다. 이때 제안 시스템에서 상정한 코킹 각도는 [그림 8]과 같이 두 손목의 중앙 지점을 기준으로 왼쪽 팔꿈치와 클럽 헤드로 각각 향하는 두 벡터 간 내적으로 계산하며,

스윙 간 유사도 계산 시 180°로 나눠 정규화된다. 이 두 지표에 개별적인 가중치를 부여 및 합산하고 비선형 활성화 함수를 적용하여 1에 가까울수록 스윙 자세가 일치함을 나타내는 유사도 점수(Weighted Swing Similarity Score)로 만든다.

골퍼 간의 스윙 동작을 정밀하게 비교하기에 앞서, 각 영상의 스윙 구간별 길이를 분석한 결과는 [표 3]과 같다. 골퍼마다 스윙 속도와 스윙 소요 시간이 다르므로 백스윙 중 가장 자세의 차이가 뚜렷하게 나타나는 “Segment 3” 구간의 프레임 수를 관찰하였다. 관찰 결과, “찰스 하웰 3세”의 첫 번째 영상이 47프레임, “안병훈”의 2번째 영상이 6프레임의 길이를 보였다. 이처럼 영상마다 세그먼트별 프레임 수가 서로 다르므로, 두 스윙 간의 직접적인 시계열 비교가 불가능한 문제가 발생한다. 본 연구에서는 이와 같은 영상 간 스윙 속도 차이로 인한 프레임 수 불일치 문제를 해결하기 위하여, 유사도를 계산하기 전에 먼저 프레임 당 인체 특징점 좌표 17개로 구성되는 시퀀스 P_{human} 을 구성하고 동적 시간 워핑(Dynamic Time Warping) 알고리즘 방식으로 두 시퀀스 간 누적 거리 비용이 최소가 되는 매핑 경로, 즉 프레임 인덱스 쌍 (a, b)의 집합인 $Path$ 를 탐색한다.

$$Path_{seg} = \left\{ (a, b) \in Path \mid \begin{aligned} &\hat{t}_{A,j} \leq a \leq (\hat{t}_{A,j+1} - 1), \\ &\hat{t}_{B,j} \leq b \leq (\hat{t}_{B,j+1} - 1) \end{aligned} \right\} \quad (6)$$

보정된 스윙 이벤트 인덱스 $\hat{t}$ 를 기준으로 경로를 분할하여 j 번째 스윙 세그먼트에 속하는 경로의 부분 집합 $Path_{seg}$ 를 식 (6)과 같이 정의한다. 이때 $\hat{t}_{A,j+1}$, $\hat{t}_{B,j}$, $\hat{t}_{B,j+1}$ 은 영상 A와 영상 B에서 각 스윙 세그먼트

표 3. 영상별 3번째 스윙 세그먼트의 프레임 수 비교
Table 3. Comparison of frame counts for the third swing segment by video

Data name	Number of frames in Segment 3	Stitched frame images of Segment 3
Byeonghun An-1	28	
Byeonghun An-2	6	
Charles Howell III-1	47	
Charles Howell III-2	7	
Tiger Woods-1	14	
Tiger Woods-2	24	
Julietta Granada-1	9	
Julietta Granada-2	10	

의 시작 지점 및 종료 지점에 해당하는 프레임의 인덱스다.

$$D_{pose} = \frac{1}{|Path_{seg}|} \sum_{(a,b) \in Path_{seg}} \| P_{A, human}^a - P_{B, human}^b \| \quad (7)$$

$Path_{seg}$ 내에 속한 모든 프레임 쌍 (a, b)에 대하여, 식 (7)과 같이 스윙 세그먼트별 인체 특징점 간 평균 유클리드 거리 D_{pose} 를 계산한다. 이때 $|Path_{seg}|$ 는 해당 세그먼트 내에서 짝지어진 프레임 쌍의 수이며, $P_{A, human}^a$ 과 $P_{B, human}^b$ 는 각각 영상 A의 a번째 프레임과 영상 B의 b번째 프레임에 해당하는 17개의 2D 인체 특징점 좌표 벡터이다.

$$P_{Pivot} = \frac{P_{Left Wrist} + P_{Right Wrist}}{2} \quad (8)$$

각 스윙 이벤트 프레임에서 코킹 각도의 차이값을 계산하기 위하여, 먼저 식 (8)과 같이 왼쪽 손목 좌표 $P_{Left Wrist}$ 와 오른쪽 손목 좌표 $P_{Right Wrist}$ 의 중간 지점을 P_{Pivot} 으로 정의하고 위치를 계산한다.

$$v_{arm} = P_{Left Elbow} - P_{Pivot} \quad (9)$$

$$v_{club} = P_{ClubHead} - P_{Pivot} \quad (10)$$

다음으로 P_{Pivot} 를 기준으로 식 (9)에 따라 왼쪽 팔꿈치 특징점 $P_{Left Elbow}$ 에서 P_{Pivot} 을 잇는 벡터 v_{arm} 를 생성하고, 식 (10)과 같이 클럽 헤드 최상단 $P_{ClubHead}$ 에서 P_{Pivot} 을 잇는 샤프트 벡터 v_{club} 를 생성한다.

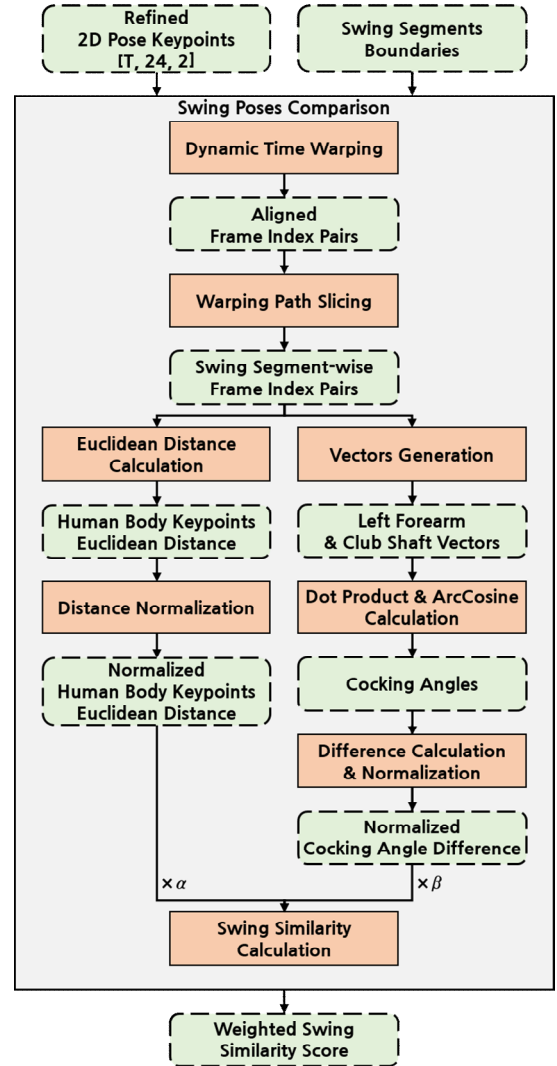


그림 7. 제안 시스템의 스윙 유사도 측정 절차

Fig. 7. Swing similarity calculation procedure of the proposed system

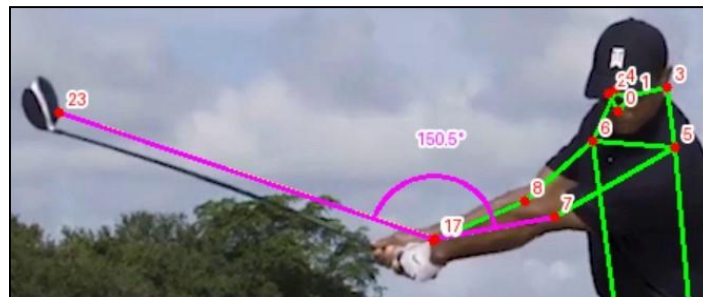


그림 8. 코킹 각도의 구성 예시

Fig. 8. Example of cocking angle configuration

$$\theta = \arccos\left(\frac{v_{arm} \cdot v_{club}}{\|v_{arm}\| \cdot \|v_{club}\|}\right) \quad (11)$$

$$\Delta\theta = |\theta_A - \theta_B| \quad (12)$$

$$\overline{\Delta\theta} = \frac{\Delta\theta}{180} \quad (13)$$

이후 식 (11)처럼 생성된 두 벡터 v_{arm} 과 v_{club} 사이의 내적을 두 벡터의 크기의 곱으로 나누고 아크코사인 ($\arccos$)을 취하여 코킹 각도 θ 를 도출하고, 식 (12)에 따라 두 영상 A와 B의 동일한 스윙 이벤트 시점에서 계산한 코킹 각도 θ_A 와 θ_B 의 절댓값 차이 $\Delta\theta$ 를 계산한 다음, 식 (13)과 같이 왼팔의 가동 범위인 180° 를 기준으로 나눠 정규화한 “코킹 각도 차이” 지표 $\overline{\Delta\theta}$ 를 구한다.

$$D_{max} = \sqrt{W^2 + H^2} \quad (14)$$

$$\overline{D}_{pose} = \frac{D_{pose}}{D_{max}} \quad (15)$$

“인체 특징점 간 유클리드 거리”와 “코킹 각도 차이”로 단위가 서로 상이한 D_{pose} 와 $\Delta\theta$ 를 단일 유사도 점수로 결합하기 위하여, 거리 지표 D_{pose} 를 0부터 1 사이의 값으로 정규화한다. 이는 식 (14)와 같이 프레임의 너비(W)와 높이(H)로 최대 픽셀 대각선 길이 D_{max} 를 구하고, 이를 분모로 식 (15)에 따라 정규화된 인체 간 거리 $\overline{D}_{pose}$ 를 도출하는 방식이다.

$$SwingDistance = \alpha \cdot \overline{D}_{pose} + \beta \cdot \overline{\Delta\theta} \quad (16)$$

마지막으로, 정규화된 $\overline{D}_{pose}$ 와 $\overline{\Delta\theta}$ 에 식 (16)과 같이 각 가중치 α 와 β 를 곱하여 합산함으로써, 두 영상 간의 스윙 차이 $SwingDistance$ 를 계산한다. 이때 가중치는 동일 인물의 스윙과 다른 인물의 스윙을 구분함에 코킹 각도 차이 지표가 인체 특징점 간 유클리드 거리 지표보다 더 높은 변별력을 보여, α 보다 β 에 더 큰 가중치를 부여하였다.

$$S = \frac{1}{1 + SwingDistance} \quad (17)$$

최종적으로 계산하는 영상 내 인물 간 자세 유사도 점수 S 는 식 (17)처럼 $SwingDistance$ 가 0일 때 최댓값이 1.0이 되며, 반대로 $SwingDistance$ 값이 커질수록 0에 수렴하는 비선형 활성화 함수 형태로 정의하였다.

IV. 실험 및 결과

1. 시스템 성능 평가를 위한 검증용 데이터셋 구축

본 장에서는 제안 시스템의 성능 검증 방법 및 검증용 데이터셋 구축 과정을 다룬다. 제안 시스템에 대한 성능 검증은 특징점 탐지 모델의 추정 결과와 제안 시스템의 결측치 보간 결과를 비교하고, $SwingNet^{[3]}$ 모델의 스윙 이벤트 추정 결과와 제안 시스템의 보정 결과를 대조한 후, 제안 시스템에서 설계한 “코킹 각도 차이” 지표의 유효성을 검증하며, 동일 인물 및 다른 인물 간 스윙 유사도를 관찰하여 스윙 자세에 대한 구분 성능을 평가한다.

이와 같은 실험의 목적은 다음과 같다. 첫째, 제안 시스템이 수행한 인체 및 골프 클럽 특징점 보정 작업이 실제 인체 및 골프 클럽에 부합하는 위치로 교정되었는지 증명한다. 둘째, 딥러닝 모델이 확률적으로 예측한 초기 추정치를 제안 시스템에서 활용하되 명시적인 “스윙 이벤트별 정답 자세”의 기준을 갖고 스윙 이벤트를 보정하는지 확인한다. 셋째, 제안 시스템에서 제시한 “코킹 각도 차이” 지표가 인물 간 스윙 자세 차이를 변별하는 데 유효한 영향을 미치는지 검증한다. 이와 같은 실험을 위하여, 야외 골프장 환경에서 촬영된 프로 골퍼 5명의 드라이버 스윙 영상을 인물당 2개씩 수집하여 검증용 데이터셋을 구축하였다. 한 인물의 영상은 “1번째 영상”, “2번째 영상”으로 구분되며 각 영상은 모두 1280×720 픽셀 크기로 통일하였다. 영상들은 “동일 인물 간 스윙”과 “다른 인물 간 스윙”끼리 중복 없이 조합되어 각각의 스윙 유사도를 계산한다. 일관된 조건에서 실험하기 위하여 각 영상마다 골퍼가 스윙을 시작하는 프레임으로부터 1프레임에서 5프레임 이전 시점을 영상의 시작

지점으로, “Finish” 자세가 해제된 이후로부터 1프레임에서 5프레임 이후 시점을 영상의 종료 지점으로 일괄 트리밍(Trimming)하였다.

2. 실험 환경 및 시스템 성능 평가 척도

본 절에서는 제안 시스템의 성능 검증 실험 환경과 성능 평가 척도를 기술한다. 실험은 NVIDIA GeForce RTX 4090 GPU와 128GB RAM이 장착된 데스크탑 컴퓨터로 하드웨어 환경을 구성하였고, Windows 운영체제 내 Ubuntu 18.04 환경에 구축된 Anaconda를 기반으로 Python 3.8과 PyTorch 프레임워크를 활용하여 일관된 조건에서 실험하였다.

실험 과정에서 프레임 시퀀스를 SwingNet 모델에 입력할 때는 메모리 한계를 고려하여 64프레임 길이로 분할하였다. 또한, 인체 특징점을 예측할 때는 각 특징점 좌표의 신뢰도 임계값(Threshold)이 0.75 이상인 경우에만 유효한 좌표로 수용하고 이에 미달하는 좌표는 [9999, 9999]로 기록하여 결측치로 설정한다.

3. 인체 및 골프 클럽 특징점 추정 결과 비교 실험

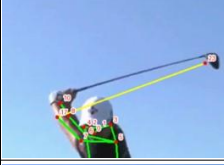
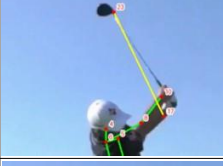
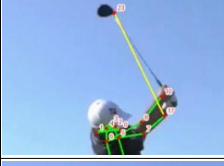
본 절에서는 YOLO11x-pose^[28] 모델의 인체 특징점 추정 결과와 제안 시스템의 보간 결과를 비교 분석하고, 오픈소스 골프 클럽 탐지 모델^[29]의 골프 클럽 특징점 추정 결과와 제안 시스템의 이상치 탐지 및 결측치 보간 결과를 대조한다. 이로써 인체와 클럽 헤드 특징점이 일부 프레임에서 결측되어 불완전한 특징점 시퀀스를 형성하는 문제를 해결할 수 있는지 검증한다.

인체 특징점 결측치 보간 전후 결과는 [표 4]와 같다. 표에서 “보간 전”은 YOLO11x-pose 모델의 추정 결과이며, “보간 후”는 해당 결과를 기반으로 제안 시스템에서 특징점을 보간한 결과다. 이때 두 시스템에 입력한 데이터는 모두 “안병훈” 골퍼의 2번째 영상이다. 프레임에 표시된 빨간색 점은 특징점의 위치를 표현하고, 노란색 선은 골프 클럽 그립과 클럽 헤드를 이은 클럽 샤프트를 나타내며, 초록색 선은 인체 특징점 간의 구성을 의미한다.

추정 결과를 보면, 보간 전에는 백스윙 동작 중 자기 폐색

표 4. 인체 특징점 결측치 보간 전후 결과 비교

Table 4. Comparison of missing human body keypoints interpolation results

Index	Before missing human body keypoints interpolation	After missing human body keypoints interpolation
33		
46		
50		

으로 인하여 영상의 전체 시퀀스 중 33번째, 46번째, 50번째 프레임에서 상체 특징점 중 7번에 해당하는 왼쪽 팔꿈치가 결측되었으나, 제안 시스템에서 제약 기반 선형 보간을 적용함으로써 해당 특징점들이 실제 인체 관절에 근접하는 위치로 복원되었음을 확인하였다.

[표 5]는 제안 시스템의 인체 특징점 결측치 보간에 대한 성능 평가 결과이다. 이때 평가를 위하여 각 영상마다 0번부터 16번까지의 인체 특징점이 가장 많이 결측된 스윙 세그먼트를 대상으로 정답(Ground Truth) 좌표를 수동 제작하였다. 표에서 ‘보간 전’은 YOLO11x-pose가 추정한 특징점 좌표와 정답 좌표 간 유클리드 거리의 평균(Mean)을 의미하고, ‘보간 후’는 제안 시스템이 보정한 특징점 좌표와 정답 좌표 간 유클리드 거리의 평균(Mean)을 나타낸다.

측정 결과, 특징점이 가장 많이 결측된 스윙 세그먼트는 Segment 7(Mid-follow-through부터 Finish까지의 구간)인 영상이 6개, Segment 3(Mid-backswing부터 Top까지의 구간)인 영상이 2개로 확인되었다. 또한 보간 전 결과에서는 임계값을 초과하지 못하여 결측된 특징점의 좌표가 포함되어, 최저 1693.57 픽셀에서 최대 2688.47 픽셀의 오차를 보였다. 반면, 보간 후 결과에서는 제안 시스템의 보간 과정

표 5. 인체 특징점 보간 전후 정답 데이터 대비 평균 유클리드 거리 비교

Table 5. Comparison of mean euclidean distance against GT before and after human keypoint interpolation

Data name	Swing segment with the most missing frames	Number of frames with missing keypoints	Distance before interpolation	Distance after interpolation
Byeonghun An-1	Segment 7	32	2097.94 px	9.29 px
Byeonghun An-2	Segment 7	21	2348.72 px	8.49 px
Charles Howell III-1	Segment 3	47	1952.09 px	1.98 px
Charles Howell III-2	Segment 7	8	2884.34 px	6.18 px
Tiger Woods-1	Segment 7	31	1693.57 px	7.89 px
Tiger Woods-2	Segment 3	24	2194.46 px	2.83 px
Julieta Granada-1	Segment 7	14	2688.47 px	7.41 px
Julieta Granada-2	Segment 7	15	2025.78 px	5.03 px

을 통하여 최저 1.98 픽셀에서 최대 9.29 픽셀의 한 자릿수 단위로 오차가 급감하였다. 이로써 제안 시스템의 인체 특징점 선형 보간 절차가 좌표를 실제 인체 관절 위치와 유사한 수준으로 복원시키는 성능을 확인하였다.

다음으로, 골프 클럽 헤드 특징점 결측치 보간 전후를 비교한 예시는 [표 6]과 같다. 해당 표는 타이거 우즈의 1번째 영상의 일부를 보여준다. 보간 전 결과를 보면, 180장의 소규모 이미지 데이터셋으로 학습된 오픈소스 골프 클럽 탐

표 6. 골프 클럽 헤드 특징점 결측치 보간 전후 결과 비교

Table 6. Comparison of missing golf club head keypoints interpolation results

Index	Before missing golf club head keypoints interpolation	After missing golf club head keypoints interpolation
69		
70		
71		

지 모델의 성능 한계로 인하여 70번째 프레임에서 골프 클럽 그립과 클럽 헤드의 특징점이 결측되었다. 이러한 결측치는 두 손목의 중앙 지점과 클럽 헤드를 잇는 벡터를 형성할 수 없게 만들며, 스윙 유사도 계산 요소 중 하나인 “코킹 각도”의 차이값을 계산할 수 없게 된다. 이는 스윙 궤적의 연속성을 훼손하여 스윙 비교 신뢰도를 감소시킬 수 있다. 반면 보간 후 결과에서는, 먼저 골프 클럽의 그립 부분에 해당하는 특징점의 좌표를 왼쪽 손목 좌표로 대체한다. 이후 정상 예측된 전후 인접 프레임의 클럽 헤드 좌표를 활용하여 3차 스플라인 보간을 적용함으로써, “Impact” 순간의 결측치가 실제 골프 클럽 위치와 방향에 부합하는 좌표로 복원되었다. 결과적으로 실제 스윙과 유사한 연속적인 궤적을 완성하였다.

[표 7]은 검증용 데이터셋을 대상으로 이와 같은 골프 클럽 특징점 좌표 보정에 대한 성능 검증을 수행한 결과를 보여준다. 보간 전 결과를 보면, 실험에 사용된 8개의 영상 모두 오픈소스 골프 클럽 탐지 모델의 특징점 추정 성능 한계로 최소 2프레임, 최대 28프레임에서 클럽 헤드 특징점이 결측되었다. 세부적으로는 전체 105프레임 중 28프레임이 결측된 타이거 우즈의 1번째 영상이나, 355프레임 중 16프레임이 결측된 찰스 하웰 3세의 1번째 영상의 사례와 같이, 원본 데이터의 프레임 시퀀스 길이와 무관하게 스윙 내에서 클럽 헤드의 탐지 실패가 모두 나타났다. 이러한 결측은 영상 간 스윙 자세를 비교할 때 유사도에 대한 신뢰도를 하락시킬 수 있다.

표 7. 골프 클럽 헤드 특징점 보간 전 결측 프레임 수

Table 7. The number of missing club head frames before interpolation

Data name	Total Number of frames	Number of missing club head frames before interpolation
Byeonghun An-1	164	10
Byeonghun An-2	64	2
Charles Howell III-1	355	16
Charles Howell III-2	59	4
Tiger Woods-1	105	28
Tiger Woods-2	140	7
Julieta Granada-1	56	4
Julieta Granada-2	68	17

4. 스윙 이벤트 추정 결과 비교 실험

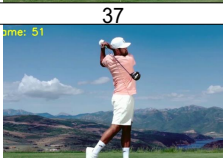
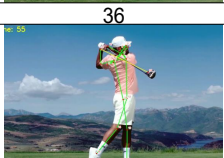
본 절에서는 SwingNet 모델이 추정한 스윙 이벤트 인덱스와, 제안 시스템이 결측치가 보간된 특징점을 활용하여 보정한 스윙 이벤트 인덱스를 비교한다. 비교 시 III-3절에서 정의된 스윙 이벤트별 자세를 기준으로 두 시스템이 추정한 결과 중 어떤 프레임이 더 해당 기준에 부합하는지 평가한다. 이로써 SwingNet 모델의 스윙 이벤트 추정 성능 한계가 완화되었는지 검증한다.

실험 결과는 [표 8]과 같으며, 이때 입력한 데이터는 프로 골퍼 “토니 피나우”의 1번째 영상이다. 스윙 이벤트별로 분석해 보면, “Address” 이벤트 인덱스는 스윙을 이미 시작한 프레임으로, “Toe-up”은 골프 클럽 샤프트가, “Mid-back-swing”은 팔이 지면과 평행하지 않은 프레임으로 추정하였다. 또한 백스윙에서 다운스윙으로 전환되기 시작한 이후 시점으로 “Top”을 추정했고, 팔이 완전히 평행하지 않은 시점으로 “Mid-downswing”을 추정했으며, Impact는 클럽이 골프공을 타격하기 직전 시점으로 추정되지 않았고, Mid-follow-through는 샤프트가 더 평행하게 표현된 프레임이 있음에도 해당 프레임으로 추정되지 않았으며, Finish는 스윙 동작이 종료되지 않은 시점으로 추정되었다.

이와 같이 SwingNet 모델의 불안정한 확률적 추론 결과에 기반하여 골프 스윙 간 유사도를 계산할 경우, 동일한 이벤트임에도 서로 다른 스윙 시점을 비교하게 되어 객관적인 비교가 되지 못할 수 있다. 따라서 딥러닝 모델이 추정한 클럽 및 인체 특징점에 대하여 좌표 보정 및 결측치 보간을 수행하고, 완성된 특징점 좌표 시퀀스를 바탕으로 딥러닝 모델이 추정한 스윙 이벤트 인덱스를 갱신하여 객관적인 스윙 세그멘테이션을 수행하는 후처리 작업이 필요하다.

이에 따라 설계한 제안 시스템의 규칙 기반 스윙 이벤트 보정 결과를 관찰했을 때, “Address”부터 “Finish”까지의 각 스윙 이벤트가 백스윙이 시작되기 직전 시점, 샤프트 또는 팔이 지면과 평행하는 시점, 다운스윙을 시작하며 골프 클럽 방향이 바뀌는 시점, 클럽 헤드가 골프공에 닿는 시점, 최종 자세가 풀리기 직전 시점으로 보정되었다. 이로써 제안 시스템이 SwingNet 모델의 스윙 이벤트 추정 결과를 개선하여, SwingNet 연구에서 정의한 스윙 이벤트별 자세를 기준으로 할 때 상대적으로 더 기준에 부합하는 프레임으로 보정하였음을 확인할 수 있다.

표 8. SwingNet의 스윙 이벤트 추정 결과와 제안 시스템의 보정 결과 비교
Table 8. Comparison of swing events estimation and refinement results

Swing events	Swing events Estimation results	Swing events Refinement results
Address		
Index	6	4
Toe-up		
Index	10	12
Mid-backswing		
Index	20	17
Top		
Index	26	25
Mid-downswing		
Index	29	30
Impact		
Index	35	34
Mid-follow-through		
Index	37	36
Finish		
Index	51	55

5. 제안 시스템의 스윙 유사도 측정 결과 분석

본 절에서는 스윙 유사도 계산 시 “코킹 각도”의 차이값을 적용함이 스윙 구분 성능에 주는 영향을 확인하고, 동일한 인물 또는 서로 다른 인물로 구성된 한 쌍의 골프 스윙 영상에 대하여 “스윙 유사도” 점수를 계산함으로써, 제안 시스템이 스윙의 같고 다름을 구분하는 성능을 확인한다. 또한 특정한 스윙 세그먼트에서 발생하는 자세의 차이를 정량적인 스윙 유사도 점수로 제시하여, 사용자가 목표로 하는 스윙과 본인의 스윙 간 차이를 파악할 수 있는지 검증한다. 이때 스윙 유사도는 정규화된 스윙 세그먼트별 인체 특징점 간 유클리드 거리와 정규화된 스윙 이벤트별 코킹 각도 차이에 각각 가중치 α 와 β 를 곱하여 합산한 다음, 정규화하여 스윙 세그먼트별 점수와 전체 프레임 시퀀스 간 평균 점수를 각각 계산한다.

본 연구에서는 가중치의 비율에 따른 제안 시스템의 스윙 구분 성능 변화를 확인하기 위하여, 유사한 스윙 자세를 보이는 골퍼 간 비교와 상이한 스윙 자세를 보이는 골퍼 간 비교로 구분하여 가중치 변화에 따른 유사도 값을 분석하였다. [표 9]는 유사한 스윙 자세를 보이는 골퍼 간 비교 2건과 상이한 스윙 자세를 보이는 골퍼 간 비교 2건을 대상으로, 두 가중치의 변화에 따른 스윙 유사도 값을 제시한 것이다. 유사한 스윙 자세의 경우에는 유사도 수치가 가장 높게 나타나는 가중치 조합을 확인하였으며, 상이한 스윙 자세의 경우에는 유사도 값이 가장 낮게 나타나는 가중치 조합을 확인하였다. 분석 결과, 서로 다른 골퍼의 스윙 자세를 비교한 모든 사례에서 $\alpha=0.2$, $\beta=0.8$ 일 때, 유사한 자세와 상이한 자세 간의 차이가 상대적으로 명확하게 나타났다. 이에 따라 본 연구에서는 두 유형의 스윙 자세를 효과적으로 구분하기 위한 가중치로 $\alpha=0.2$, $\beta=0.8$ 을 적용하였다.

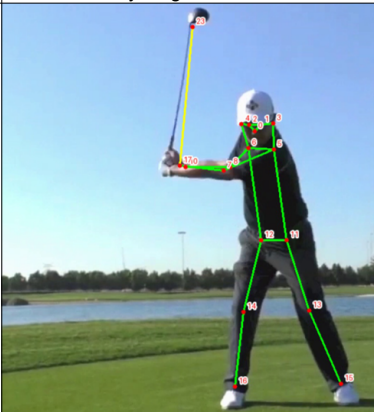
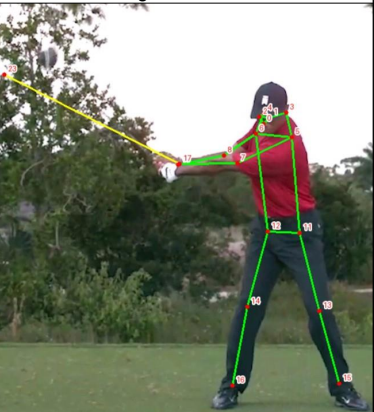
표 9. 다양한 가중치 조합에 따른 스윙 유사도 비교
Table 9. Comparison of swing similarity for different weight combinations
(a) Similar swing pose comparison - Case 1

Data name	Charles Howell III-1		Charles Howell III-2
Frame image			
Weights	α	β	Swing pose similarity
	0.2	0.8	0.9345
	0.3	0.7	0.9191
	0.4	0.6	0.9043
	0.5	0.5	0.8899
	0.6	0.4	0.8761
	0.7	0.3	0.8626
	0.8	0.2	0.8496

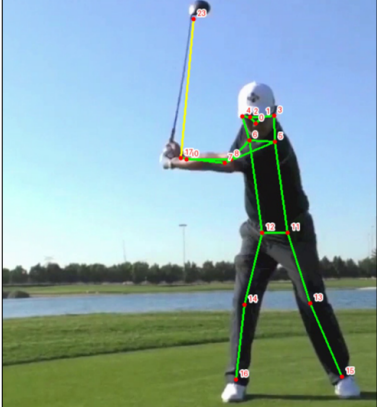
(b) Similar swing pose comparison - Case 2

Data name	Charles Howell III-2		Julieta Granada-1
Frame image			
Weights	α	β	Swing pose similarity
	0.2	0.8	0.9114
	0.3	0.7	0.8977
	0.4	0.6	0.8844
	0.5	0.5	0.8716
	0.6	0.4	0.8591
	0.7	0.3	0.8470
	0.8	0.2	0.8352

(c) Dissimilar swing pose comparison - Case 3

Data name	Byeonghun An-2		Tiger Woods-1
Frame image			
Weights	α	β	Swing pose similarity
	0.2	0.8	0.9055
	0.3	0.7	0.9101
	0.4	0.6	0.9148
	0.5	0.5	0.9197
	0.6	0.4	0.9248
	0.7	0.3	0.9300
	0.8	0.2	0.9354

(d) Dissimilar swing pose comparison - Case 3

Data name	Byeonghun An-2		Tiger Woods-2
Frame image			
Weights	α	β	Swing similarity
	0.2	0.8	0.9255
	0.3	0.7	0.9285
	0.4	0.6	0.9316
	0.5	0.5	0.9348
	0.6	0.4	0.9382
	0.7	0.3	0.9417
	0.8	0.2	0.9453

다음으로, 두 골프 스윙 영상의 자세 유사도 계산 시 적용되는 “스윙 세그먼트별 인체 특징점 간 유클리드 거리”, “스윙 이벤트 프레임별 코킹 각도 차이”와 제안 시스템의 “규칙 기반 스윙 이벤트 보정”과 “동적 시간 워핑”의 각 지표에 대한 기여도를 분리하여 실험하였다. 본 실험은 제안 시스템의 구조를 기준으로 하여, 평가 대상인 각 단일 지표만을 시스템에서 소거(Ablation)하였을 때 스윙 유사도의 변화 정도를 관찰하는 방식으로 진행하였다. 동일 골퍼의 반복 스윙은 서로 다른 골퍼 간의 스윙보다 상대적으로 높은 자세 유사성을 보인다는 가정하에, 동일 골퍼의 서로 다른 스윙 4건과 서로 다른 골퍼 간 스윙 24건을 대상으로

평균 유사도를 산출하고, 두 그룹 간 평균 유사도의 차이를 separation gap으로 정의하여 비교하였다. [표 10]은 실험 결과를 보여주고 있다. 전체 제안 시스템의 separation gap은 0.0258로 나타났다. 코킹 각도 차이를 제거한 경우 separation gap은 0.0085로 감소하였으며, 규칙 기반 스윙 보정을 제거한 경우에는 0.0011로 크게 감소하였다. 이는 코킹 각도 정보와 스윙 이벤트 보정 과정이 동일 골퍼와 서로 다른 골퍼의 스윙 자세를 구분하는데 중요한 역할을 함을 보여준다. 또한, 동적 시간 워핑(DTW)을 제거한 경우, separation gap은 0.0254로 전체 시스템과 유사한 수준을 보였다. 한편, 인체 특징점 차이를 제거한 경우 separation gap

표 10. 제안된 스윙 유사도 측정 방법의 Ablation Study
Table 10. Ablation study of the proposed swing similarity method

Method	Within-Golfer Similarity ↑	Between-Golfer Similarity ↓	Separation Gap ↑
w/o Human Keypoint Distance	0.9519	0.9214	0.0305
w/o Cocking Angle Difference	0.8756	0.8671	0.0085
w/o Rule-based Swing Event Refinement	0.9128	0.9117	0.0011
w/o Dynamic Time Warping	0.9324	0.9070	0.0254
Proposed System (Full)	0.9345	0.9087	0.0258

이 0.0305로 증가하여, 해당 요소의 기여도에 대해서는 추가적인 분석이 필요한 것으로 나타났다.

골프 스윙 자세 분석과 관련된 선행 연구들을 살펴보면, [14]는 Kinect로 스윙 영상을 촬영하여 프로 골퍼의 스윙이 모델링된 GMM을 기반으로 SVM에서 사용자의 스윙 등급을 판별하였다. AI Golf^[20] 연구는 SwingNet^[3] 이용해 스윙 이벤트를 예측하고, 각 이벤트별로 추정된 인체 특징점 이미지를 활용하여 스윙 수준을 이진 분류하였다. GolfMate^[19]는 인체 특징점의 시계열 변화량을 Bi-LSTM으로 보정하고, 관절 각도 정보를 반영한 설명 가능한 임베딩(Embedding)을 구축하여 자세 유사성을 분석하였다. 또한 GolfPose^[4]는 연구실 환경에서 촬영된 이미지 데이터셋을 구축하고, 딥러닝 모델을 파인튜닝하여 사람과 골프 클럽에 대한 바운딩 박스 및 특징점을 추정하는 방법을 제안하였다. 이와 같은 선행 연구들은 사전에 학습된 데이터 분포나 정규화된 임베딩 모델을 기반으로 사용자의 골프 스윙 수준을 평가하거나 등급을 분류하는데 초점을 두고 있다. 반면 본 연구에서 제안하는 시스템은 고가의 전문 촬영 장비 없이 스마트폰 등 휴대용 전자기기로 촬영한 단일 시점 영상을 입력으로 사용할 수 있다. 또한 인체 특징점뿐만 아니라 골프 클럽 특징점을 함께 추정하여 스윙 유사도 산출에 반영하고, 스윙 전체를 단일 점수로 평가하는 대신 각 스윙 동작 단계별로 유사도를 계산한다. 나아가 사용자가 비교를 원하는 특정 프로 골퍼의 스윙 영상을 직접 입력하여 자신의 스윙과 비교하거나, 서로 다른 두 프로 골퍼의 스윙 자세를 비교할 수 있도록 구성하였다.

V. 결 론

본 논문에서는 스윙 이벤트 추정, 인체 관절점 및 골프 클럽 특징점 추정 정확도를 높이기 위하여, 딥러닝 및 규칙 기반 골프 스윙 포즈 유사도 분석 시스템을 제안하였다. 제안 시스템에서는 먼저 사전 학습된 딥러닝 모델을 기반으로 인체 및 골프 클럽의 특징점을 추정하고, 스윙 단계별 구간을 의미하는 “스윙 세그먼트”의 시작 지점 “스윙 이벤트”에 대한 프레임 인덱스를 추정하였다. 주요 특징점 및 스윙 이벤트의 추정 정확도를 높이기 위하여 다양한 조건

에서 촬영된 영상을 기반한 학습 데이터셋을 구축하는 것 대신에, 특징점에 대한 이상치 정의 및 보정 알고리즘을 제안하였고, 골프 스윙 자세의 명확한 정의를 기반한 스윙 이벤트 보정 규칙을 제안하였다.

본 연구의 핵심 기여는 다음과 같다. 첫째, 기존에 제안되었던 YOLO11x-pose^[28] 모델은 골퍼의 상체 특징점에 대한 자기 폐색(Self Occlusion) 문제가 발생하였으며, 오픈소스 골프 클럽 탐지 모델^[29]은 소규모 데이터로 학습되어 불안정한 특징점 추정 성능을 보였다. 반면 제안 시스템에서는 결측된 인체 및 골프 클럽 특징점을 각각 선형 및 스플라인 방식으로 보간하여 실제 스윙 자세와 유사한 형태로 좌표를 복원하였다.

둘째, 선행 연구인 SwingNet^[3] 모델이 확률적으로 스윙 이벤트를 예측하는 것에서 발전하여, 해당 연구에서 제안된 각 스윙 이벤트별 스윙 자세에 따라 규칙을 설정 후 이를 기반으로 스윙 이벤트 인덱스를 보정함으로써 기존 방법론에서 제안된 기준을 기반으로 기존 모델보다 더 정확한 스윙 이벤트 추정 성능을 달성하였다.

셋째, 기존 연구에서는 골프 스윙을 전체 동작이나 넓은 범위의 구간 단위로 비교하였으나, 제안 시스템은 단계적 과정을 수행하는 골프 스윙의 특성을 고려하여 각 스윙 세그먼트별로 자세 차이를 분석하였다. 이로써 프로 골퍼 영상을 대상으로 한 사례 분석을 통하여 제안 방법이 스윙 구간별 차이를 식별할 수 있는 기술적 가능성을 확인하였다. 또한 향후 사용자의 스윙과 프로 골퍼의 스윙을 비교하여 차이가 크게 나타나는 구간을 안내하는 코칭 보조 도구 및 골프 방송에서 선수 간 스윙을 정량 비교하는 시각화 자료로 활용될 가능성을 보였다.

현재 제안 시스템은 스윙 시 골프 클럽이 탄성으로 인하여 형태가 왜곡되는 현상을 반영하지 않았으나, 향후 연구에서는 이러한 변형을 반영한 골프 클럽 포즈 추정이 요구된다. 또한 초보자 대상 코칭 효과 및 실제 방송 환경에서의 활용성을 직접 검증하지 않았으며, 현재 시스템은 제한적인 프로 골퍼 영상만을 사용하여 실험하였으므로 다양한 숙련도, 체형 및 스윙 유형을 가진 골퍼에게 연구 결과를 일반화하는 데 한계가 있다. 나아가 3차 스플라인 곡선 위에 골프 클럽 특징점 결측치가 보간되는 구조를 발전시켜 골프 클럽이 골퍼의 손목을 중심으로 회전하는 점을 고려

한 보정 처리 연구가 필요하다.

본 논문을 통해 골프 스윙 자세 분석에 있어서 인체 관절 점 추정뿐만 아니라 골프 클럽 특징점 추정의 정확도가 크게 영향을 미치게 됨을 확인하였다. 이에 다양한 골프 클럽의 특징점 추정이 가능한 모델 구축을 위하여 다양한 골프 클럽 특징점 추정에 대한 추가 데이터셋을 구축하고 골프 클럽 탐지에 유리한 모델 설계 및 구축이 요구된다.

참 고 문 헌 (References)

- [1] J. Park, and S. Kwak, "Golf Swing Classification Using Fuzzy System," *Journal of Broadcast Engineering*, Vol.18, No.3, pp.380-392, May 2013.
doi: <https://doi.org/10.5909/JBE.2013.18.3.380>
- [2] 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, October 2019.
doi: <https://doi.org/10.5573/ieie.2019.56.10.49>
- [3] 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*, Long Beach, CA, USA, pp. 2553-2562, 2019.
doi: <https://doi.org/10.1109/CVPRW.2019.00311>
- [4] M. Lee, Y. Zhang, K. Wu, and Y. Tseng, "GolfPose: From Regular Posture to Golf Swing Posture," *International Conference on Pattern Recognition*, Kolkata, India, pp. 387-402, 2024.
doi: https://doi.org/10.1007/978-3-031-78305-0_25
- [5] J. Park, "The Methodology of the Golf Swing Similarity Measurement Using Deep Learning-Based 2D Pose Estimation," *Journal of The Korea Society of Computer and Information*, Vol.28, No.1, pp.39-47, January 2023.
doi: <https://doi.org/10.9708/jksci.2023.28.01.039>
- [6] R. Sun, Z. Lin, S. Leng, A. Wang, and L. Zhao, "An In-Depth Analysis of 2D and 3D Pose Estimation Techniques in Deep Learning: Methodologies and Advances," *Electronics*, Vol.14, pp.1307, 2025. (accessed Jul. 23, 2026)
doi: <https://doi.org/10.3390/electronics14071307>
- [7] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Zitnick, "Microsoft COCO: Common Objects in Context," *European Conference on Computer Vision*, Zurich, Switzerland, pp. 740-755, 2014.
doi: https://doi.org/10.1007/978-3-319-10602-1_48
- [8] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.36, No.7, pp.1325-1339, December 2014.
doi: <https://doi.org/10.1109/TPAMI.2013.248>
- [9] J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu, "AI Coach: Deep Human Pose Estimation and Analysis for Personalized Athletic Training Assistance," *Proceeding of ACM International Conference on Multimedia*, Nice, France, pp. 2228 - 2230, 2019.
doi: <https://doi.org/10.1145/3343031.3350910>
- [10] K. Kulkarni and S. Shenoy, "Table Tennis Stroke Recognition Using Two-Dimensional Human Pose Estimation," *Proceeding of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, pp. 4571 - 4579, 2021.
doi: <https://doi.org/10.1109/CVPRW53098.2021.00515>
- [11] R. Kurose, M. Hayashi, T. Ishii, and Y. Aoki, "Player pose analysis in tennis video based on pose estimation," *International Workshop on Advanced Image Technology*, Chiang Mai, Thailand, pp. 1-4, 2018.
doi: <https://doi.org/10.1109/IWAIT.2018.8369762>
- [12] S. Park, J. Chang, H. Jeong, J. Lee, and J. Park, "Accurate and Efficient 3D Human Pose Estimation Algorithm Using Single Depth Images for Pose Analysis in Golf," *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, pp. 105-113, 2017.
doi: <https://doi.org/10.1109/CVPRW.2017.19>
- [13] H. Kim, S. Jang, H. Yoo, J. Lee, Y. Jeon, D. Kim, and S. Kim, "Analysis of Golf Swing by Using Kinect," *Proceedings of KIIT Conference*, Gwangju, Korea, pp. 511-512, 2018.
- [14] L. Zhang, J. Hsieh, T. Ting, Y. Huang, Y. Ho, and L. Ku, "A Kinect based Golf Swing Score and Grade System using GMM and SVM," *International Congress on Image and Signal Processing*, Chongqing, China, pp. 711 - 715, 2012.
doi: <https://doi.org/10.1109/CISP.2012.6469827>
- [15] J. Lee, S. Park, and C. Kim, "Golf Swing Analysis using COP Trajectory Image Learning," *Proceedings of KSPE Autumn Conference*, Gunsan, Korea, pp.494-495, 2018.
- [16] C. Ju, J. Park, and D. Lee, "Deep Learning-Based Outlier Detection and Correction for 3D Pose Estimation," *Korea Information Processing Society Transactions: Software and Data Engineering*, Vol.11, No.10, pp.419-426, October 2022.
doi: <https://doi.org/10.3745/KTSDE.2022.11.10.419>
- [17] Z. Geng, K. Sun, B. Xiao, Z. Zhang, and J. Wang, "Bottom-Up Human Pose Estimation Via Disentangled Keypoint Regression," *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, pp. 14676-14686, 2021.
doi: <https://doi.org/10.1109/CVPR46437.2021.01444>
- [18] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, pp. 5693-5703, 2019.
doi: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2019.00584>
- [19] C. Ju, J. Kim, and D. Lee, "GolfMate: Enhanced Golf Swing Analysis Tool through Pose Refinement Network and Explainable Golf Swing Embedding for Self-Training," *Applied Sciences*, Vol.13, No.20, pp.11227, 2023. (accessed Jul. 23, 2026)
doi: <https://doi.org/10.3390/app132011227>
- [20] AI Golf: Golf Swing Discrepancy Detector, <https://github.com/>

- harryyizihan/ai_golf_swing (accessed Jul. 23, 2026)
- [21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 4510-4520, 2018.
doi: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00474>
- [22] D. Lee, and J. Nang, "A Dataset and Keypoint Detection for Club Head Trajectory Estimation in 2D Golf Swing Video," *Proceeding of Korean Institute of Information Scientists and Engineers: Korea Software Congress*, Pyeongchang, Korea, pp. 523-525, 2020.
- [23] J. Kang, J. Kim, and B. Park, "A Study on Pose Estimation Technique using external keypoints for golf swing pose correction," *Conference of Korean Institute of Communications and Information Sciences*, pp.332-333, 2020.
- [24] B. Cheng, B. Xiao, J. Wang, H. Shi, T. Huang, and L. Zhang, "HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation," *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 5385-5394, 2020.
doi: <https://doi.org/10.1109/CVPR42600.2020.00543>
- [25] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," *arXiv Preprint arXiv:2107.08430*, 2021. (accessed Jul. 23, 2026)
doi: <https://doi.org/10.48550/arXiv.2107.08430>
- [26] J. Zhang, Z. Tu, J. Yang, Y. Chen, and J. Yuan, "MixSTE: Seq2seq Mixed Spatio-Temporal Encoder for 3D Human Pose Estimation in Video," *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, pp. 13222-13232, 2022.
doi: <https://doi.org/10.1109/CVPR52688.2022.01288>
- [27] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248-255, 2009.
doi: <https://doi.org/10.1109/CVPR.2009.5206848>
- [28] R. Khanam, and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements," *arXiv Preprint arXiv:2410.17725*, 2024. (accessed Jul. 23, 2026)
doi: <https://doi.org/10.48550/arXiv.2410.17725>
- [29] Golf Club 7 Keypoints Computer Vision Model, <https://universe.roboflow.com/testproject-caupq/golf-club-7-keypoints-kjgnr> (accessed Jul. 23, 2026)

저 자 소 개

설 동 규



- 2022년 2월 : 동아방송예술대학교 뉴미디어콘텐츠과 전문학사
- 2024년 2월 : 한국예술원 방송영상디지털콘텐츠예술학과 학사
- 2026년 8월 : 서울미디어대학원대학교 인공지능응용소프트웨어학과 석사
- ORCID : <https://orcid.org/0009-0004-9480-265X>
- 주관심분야 : 컴퓨터 비전, 인간 자세 추정, 에이전틱 AI, AI 표준화, 디지털 트윈, 메타버스, 프로젝트 맵핑

최 유 주



- 1989년 2월 : 이화여자대학교 컴퓨터공학과 학사
- 1991년 2월 : 이화여자대학교 컴퓨터공학과 석사
- 2005년 2월 : 이화여자대학교 컴퓨터공학과 박사
- 2005년 9월 ~ 2010년 2월 : 서울벤처대학원대학교 교수
- 2010년 3월 ~ 현재 : 서울미디어대학원대학교 교수
- ORCID : <https://orcid.org/0000-0001-7520-097X>
- 주관심분야 : 컴퓨터 그래픽스, 이미지 프로세싱, 비주얼 컴퓨팅, HCI, 가상현실