

## Outcome-Based Comparison of Representative Pitching Motions Using Deep Learning-Based 3D Skeleton Sequences

Jae-Seung Kim\*, Hyung-Woo Moon\*\*, Yong-Tae Woo\*\*\*

\*Student, Dept. of Computer Engineering, Changwon National University, Gyeongsangnamdo, Korea

\*\*Chief Technical Officer, HiBall Co., Gyeongsangnamdo, Korea

\*\*\*Professor, Dept. of Computer Engineering, Changwon National University, Gyeongsangnamdo, Korea

### [Abstract]

In this paper, we propose a method to define representative pitching motions centered on pitching results in baseball games and proposes a method to analyze the influence of pitch type or pitching results on pitching motions. To this end, the videos taken from the front of the pitcher were classified into four types, and the pitching videos were analyzed by type. The pitching videos were divided into frames, and a 2D skeleton was extracted using HRNet, and then converted into a time-series 3D skeleton using PoseFormerV2. The relative coordinates based on the pelvis were normalized to correct the shooting distortion. DTW Barycenter Averaging (DBA) was applied to generate representative pitching motions by type, and the similarity between representative pitching motions was compared using the Dynamic Time Warping technique. The experimental results showed that the difference in pitching motions according to the pitching results was more significant, which means that individual pitchers have result-centered pitching motions.

▶ **Key words:** Deep Learning, HRNet, PoseFormerV2, Pitching motion analysis, DTW

### [요 약]

본 연구는 야구 경기에서 투구 결과 중심의 대표 투구 동작을 정의하고, 구종이나 투구 결과가 투구 동작에 미치는 영향을 분석하기 위한 방법을 제안한다. 이를 위하여 투수 정면에서 촬영된 영상을 네 가지 유형으로 분류하고, 유형별로 투구 패턴을 분석하였다. 투구 영상은 프레임 단위로 분할하고, HRNet을 이용하여 2D 스켈레톤을 추출한 뒤, PoseFormerV2를 통해 시계열 3D 스켈레톤으로 변환하였다. 촬영 왜곡을 보정하기 위하여 골반 기준의 상대 좌표로 정규화하였다. DTW Barycenter Averaging(DBA)을 적용하여 유형별로 대표 투구 동작을 생성하고, Dynamic Time Warping 기법을 이용하여 대표 투구 동작간의 유사도를 비교 실험하였다. 실험 결과, 투구 결과에 따른 투구 동작 차이가 더 유의미하였고, 이는 투수 개인별로 결과 중심의 투구 동작이 존재함을 의미한다.

▶ **주제어:** 딥러닝, HRNet, PoseFormerV2, 투구 동작 분석, 투수 평가, DTW(Dynamic Time Warping)

- First Author: Jae-Seung Kim, Corresponding Author: Yong-Tae Woo
- \*Jae-Seung Kim (jskim@hiball.net), Dept. of Computer Engineering, Changwon National University
- \*\*Hyung-Woo Moon (hwmoon@hiball.net), HiBall Co.
- \*\*\*Yong-Tae Woo (ytwoo@changwon.ac.kr), Dept. of Computer Engineering, Changwon National University
- Received: 2025. 05. 02, Revised: 2025. 05. 13, Accepted: 2025. 05. 21.

## I. Introduction

인공지능에 기반한 영상처리 기법을 이용하여 야구 경기 영상에서 투수의 투구 동작을 분석하여, 전략 수립이나 경기력 향상에 적용하기 위한 연구가 활발하게 진행되고 있다[1-5]. 투구 동작에서 어깨의 위치나 팔꿈치 각도의 차이로 인해 타격 결과가 안타가 되거나 헛스윙이 되는 경우가 빈번하게 발생한다[6,7]. 따라서 투수 개인별 신체 조건을 반영한 체계적인 투구 동작 분석은 투수의 기량을 향상시키는 데 중요한 역할을 할 수 있다. 요즘은 웨어러블 센서를 이용하여 투구 동작을 분석하는 방법도 활발하게 연구되고 있다[8-10].

최근에는 경기 중에 촬영한 영상에서 추출한 관절 정보를 이용하여 투구 동작을 체계적으로 분석하기 위한 연구가 활발하게 진행되고 있다[11-13]. J. Bright 등은 야구 중계 영상에서 TCN(Temporal Convolution Network)과 Transformer를 기반으로 투구 동작을 분석하여 릴리스 포인트, 구속, 투수 익스텐션 등과 같은 다양한 정보를 추출하기 위한 PitcherNet을 제안하였다[11]. N. Umezu 등은 3D 스켈레톤과 투구 동작 비교 기법을 통해 프로 선수의 투구 동작과 아마추어 선수들의 투구 동작을 비교 분석하기 위한 시각적 피드백 시스템을 제안하였다[12].

임준석 등은 2D 관절 좌표에서 DTW(Dynamic Time Warping) 알고리즘을 적용하여 투구 동작의 이상 여부를 판단하는 기법을 제안하였다[14]. 박준영 등은 CNN(Convolution Neural Network) 모델을 이용하여 경기 영상에서 투구 동작이 출현하는 장면을 분류하고 기록과 연계하여 장면을 검색하는 방법을 제안하였다[15]. 하지만 이러한 기법들은 2D 관절 좌표나 투구 동작 이미지만을 이용하는 관계로 투구 동작을 입체적으로 정밀하게 분석하기 어렵다. 따라서, 투구 동작에 영향을 미치는 요인을 찾아내고 이를 통해 투수 개인별로 최적화된 투구 동작을 분석하여 투수의 경기력을 향상시킬 수 있는 기법이 필요하다.

본 연구에서는 구종과 투구 결과별로 대표 투구 동작 간의 차이를 정량적으로 분석하기 위한 방법을 제안한다. 먼저, 인공지능 기반의 시계열 스켈레톤 분석 기법을 이용하여 구종(직구/변화구)과 투구 결과(스트라이크/볼)별로 대표 투구 동작을 정의한다. 그리고 구종과 투구 결과별로 분류된 영상은 딥러닝 기반의 스켈레톤 추출 및 변환 기법을 통해 관절 기반 데이터로 구성한다. 시계열 분석 기법을 이용하여 구종과 투구 결과별로 대표 투구 동작을 생성한다. 마지막으로 대표 투구 동작 간의 유사도를 정량적

로 비교하여 투수 개인별로 투구 결과 중심의 유의미한 투구 동작이 존재함을 검증하였다.

본 연구에서 제안한 기법의 유효성을 검증하기 위하여 구종과 투구 결과별로 60개씩, 투수 한 명당 총 240개의 투구 영상을 수집하여 실험을 진행하였다. 실험 결과, 투구 결과에 따른 투구 동작 차이가 구종보다 더 뚜렷하게 나타나는 경향을 검증하였다. 향후, 본 연구는 투수의 기량 향상을 위한 실시간 피드백 훈련 시스템, 투구 동작 자동 교정 시스템, 개인화된 퍼포먼스 트래킹 시스템 등에 응용할 수 있으리라 기대된다.

## II. Preliminaries

### 1. Related works

#### 1.1 A Study on Existing Deep Learning-Based Motion Analysis Methods

최근에 스포츠 분야에서는 경기 전략 수립이나 경기력 향상을 위하여 인공지능 기술을 활용하여 선수들의 경기 영상을 시각적으로 분석하여 정량화하기 위한 연구가 활발하게 진행되고 있다[16,17]. 기존의 주된 연구 방법은 2D 스켈레톤 기반의 시각적 분석이나 정적인 프레임 단위의 분석 방법이 대부분이다. 하지만 이러한 방식은 경기 영상에서 입체적인 깊이 정보를 파악하여 분석하기 어렵다. 또한 선수 개인의 특성이나 반복 동작 내에서 세부 패턴을 찾아내기 어렵다. 이에 따라, 3D 스켈레톤으로 구성된 시계열 데이터를 활용한 입체적인 분석과 개인 맞춤형 동작 평가 방법에 대한 연구가 필요하다.

#### 1.2 HRNet Model

인간 자세 추정(Human Pose Estimation, HPE)을 위한 딥러닝 모델 중에서 HRNet(High-Resolution Network)은 다양한 해상도의 특징 정보를 동시에 활용하기 위한 모델이다[18]. 이 모델은 고해상도를 유지하는 네트워크를 중심으로 다양한 해상도의 특징 맵을 병렬로 처리하고, 이를 반복적으로 상호 보완하는 방식으로 학습한다. 이 모델에서 고해상도 맵은 정밀한 위치 정보를 제공하고, 저해상도 맵은 전체적인 구조 정보를 제공한다. 또한 두 맵을 결합하여 보다 정확한 관절 위치를 예측할 수 있다. 전체적인 영상에 대한 구조적인 정보와 세부적인 관절 위치 정보를 동시에 고려하여 보다 세밀한 2D 스켈레톤 추출이 가능하다. 본 연구에서는 투구 영상에 대한 2D 스켈레톤을 추출하기 위한 목적으로 HRNet을 활용하였다.

### 1.3 PoseFormerV2 Model

PoseFormerV2 모델은 시계열 2D 스켈레톤을 입력하여 3D 스켈레톤을 예측하는 Transformer 기반 모델이다 [19]. 이 모델은 시계열 2D 스켈레톤을 저주파 중심 패턴 (Low-Frequency Pattern)으로 학습함으로써, 일시적인 이상값이나 잡음(High-Frequency Components)의 영향을 최소화하고 전반적인 동작 흐름을 보다 안정적으로 구성할 수 있는 장점이 있다. 본 연구에서는 기존의 2D 스켈레톤으로 분석이 어려운 입체적인 깊이 정보와 관절 간의 상대적 구조를 보완하기 위하여 PoseFormerV2 모델을 이용하여 3D 관절 좌표를 추정하였다.

### 1.4 DTW Algorithm

DTW 알고리즘은 시계열 데이터 간의 길이나 속도의 차이가 존재하더라도 두 데이터를 유연하게 정렬하여 유사도를 비교할 수 있는 알고리즘이다[20]. 영상에서 동작은 시점에 따라 구간 길이나 속도, 리듬 등이 달라질 수 있다. DTW 알고리즘은 이러한 종류의 시계열 데이터를 효과적으로 비교할 수 있다. 이 알고리즘은 두 개의 시계열 데이터를 비선형적으로 정렬하여 가장 유사한 시점 간의 매칭 경로를 계산한다. 이 경로를 기반으로 거리 값을 산출하여 유사도를 정량적으로 평가할 수 있다. 본 연구에서는 구종-결과 조합별로 생성된 대표 투구 동작 간의 차이를 정량적으로 분석하기 위하여 DTW 알고리즘을 적용하였다.

### 1.5 DTW Barycenter Averaging (DBA)

DBA 기법은 DTW 알고리즘을 기반으로 여러 개의 시계열 데이터를 평균화하여 대표 시계열을 생성하기 위한 기법이다[21]. 이 기법은 각 시계열 데이터를 정렬한 뒤 유사한 시점의 값을 평균하는 과정을 반복 수행하여 최종적으로 수렴된 대표 동작을 도출한다. 이를 통해 데이터의 정렬 기준에 대한 일관성을 제공하고, 개인별로 최적의 투구 동작을 효과적으로 정의할 수 있다. 본 연구에서는 DBA 기법을 통해 투수 개인별 구종-결과 조합별로 구성된 시계열 3D 스켈레톤으로부터 대표 투구 동작을 추출하였다. 그리고 추출된 대표 투구 동작을 기준으로 다른 조건의 대표 투구 동작과 유사도를 비교하는 목적으로 사용하였다.

## III. The Proposed Scheme

본 연구에서는 투수 개인별 투구 동작을 구종과 결과에 따라 조건별로 분류하였다. 그리고 각 조건에 해당하는 투

구에서 대표 투구 동작을 추출한 후, 이들 간의 투구 동작 차이를 정량적으로 비교·분석하기 위한 새로운 기법을 제안한다. 먼저, 투구 영상으로부터 HRNet 모델을 이용하여 2D 스켈레톤을 추출하고 투수를 식별하였다. 그리고 PoseFormerV2 모델을 이용하여 3D 스켈레톤으로 변환하고 정규화하였다. DBA를 이용하여 조건별 대표 투구 동작을 추출하였다. 마지막으로 DTW 알고리즘을 활용하여 조건별 대표 투구 동작 간의 유사도를 비교·분석하였다.

그림 1은 본 연구에서 제안한 딥러닝 기반의 3D 스켈레톤 시계열 분석 기법과 투구 결과 중심의 대표 투구 동작 추출 및 유사도 비교 과정을 개념적으로 나타낸 그림이다.

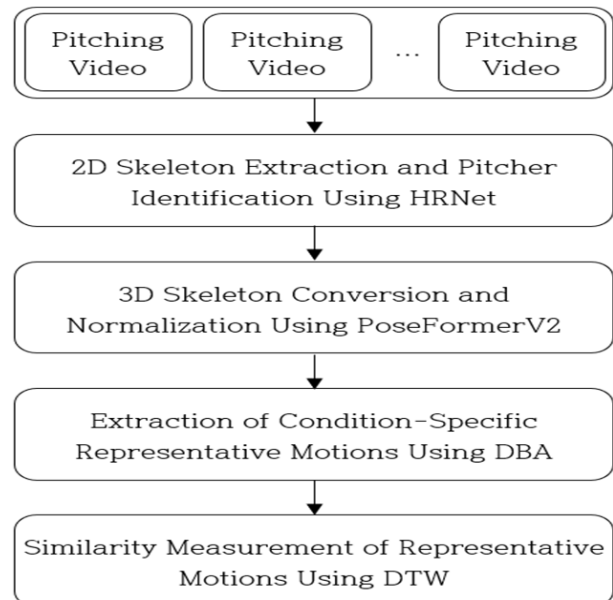


Fig. 1. Concept Diagram for Deep Learning-Based Analysis of Result-Specific Representative Pitching Motions

### 1. 2D Skeleton Extraction and Format Conversion

본 과정은 HRNet 모델을 이용하여 경기 영상에서 투수를 효과적으로 식별하기 위한 방법과 식별된 투수의 투구 동작으로부터 2D 스켈레톤을 추출하는 방법으로 구성된다. 또한 이 과정에서는 포맷 정제 과정을 통하여 투수 개인별 투구 동작에 대한 시계열 2D 스켈레톤을 구성한다.

#### 1.1 Filtering and Identification of the Pitcher's Skeleton

HRNet 모델은 경기 영상 프레임에 포함된 모든 인물을 인식하여 2D 스켈레톤을 추출하기 위한 모델이다. 본 연구에서는 투수의 투구 동작을 효과적으로 분석하기 위하여 영상 프레임에서 스켈레톤의 길이가 가장 긴 인물을 투수로 식별한다. 투수의 스켈레톤 길이는 영상에서 인식된

사람 중에서 머리 위치의  $y$ 축과 양 발중 가장 낮은 발 위치의  $y$ 축간의 차이에 의해 계산한다. 이 방법을 통하여 경기 영상에서 투수외에 불필요한 인물을 효과적으로 제거하여 투수만으로 구성된 2D 스켈레톤을 구성할 수 있다.

### 1.2 Structuring of Sequential 2D Skeletons for 3D Pose Estimation

경기 영상에서 추출된 2D 스켈레톤은 프레임 단위의 시계열 구조로 구성되는 관계로 PoseFormerV2 모델의 입력 형식에 맞게 재구성한다. 각 프레임은 17개 관절에 대한  $x$ ,  $y$  좌표 정보를 포함하고 있다. 시간 순서대로 정렬된 데이터는 하나의 numpy 배열로 구성된다. 또한 모델 학습과 일관된 비교 분석을 위하여 .npz 포맷으로 저장한다. 이러한 데이터 구조화는 구종-결과 조합에 따른 다양한 투구 동작을 일관된 형식으로 관리하고, 3D 변환과 시계열 기반 분석을 위한 데이터를 구성하는데 활용된다.

## 2. 3D Skeleton Conversion and Pelvis-Centered Normalization Using PoseFormerV2

본 과정은 PoseFormerV2 모델을 이용하여 2D 스켈레톤을 3D 시계열 데이터로 변환하는 과정이다. 이 과정에서는 영상의 촬영 각도나 투수 위치 변화로 발생하는 왜곡 현상을 보정하기 위하여 골반 중심 기준으로 정규화를 수행한다. 이 과정을 통해 모든 투구 동작은 동일한 기준으로 비교 가능한 형태로 정제된다.

### 2.1 3D Skeleton Prediction Using PoseFormerV2

이 과정은 PoseFormerV2를 이용하여 구종-결과 조건별로 분류된 투구 영상을 시계열 3D 스켈레톤으로 변환하는 과정이다. 일반적으로 투구 동작은 좌우 움직임뿐만 아니라 앞뒤 움직임에 대한 분석도 중요하다. 하지만, 2D 스켈레톤은 정면 방향으로 움직이는 투구 동작은 분석하기 어렵다. 따라서, 입체적인 깊이 정보를 추출할 수 있는 3D 스켈레톤으로 변환이 필요하다. 변환된 3D 스켈레톤을 통하여 투구 동작 간의 일관성과 패턴의 차이를 정량적으로 비교하여 투구 동작의 구조적인 특성을 파악할 수 있다.

### 2.2 Pelvis-Centered Normalization and Data Structuring

3D로 변환된 스켈레톤에서는 촬영 각도나 투구의 위치 변화로 인해 시계열 간의 위치와 방향에 미세한 왜곡이 발생할 수 있다. 이러한 왜곡은 유사도 비교 시 오차를 증가시킬 수 있다. 본 연구에서는 위치에 대한 왜곡을 최소화

하기 위하여 좌우 엉덩이 관절의 평균 위치를 기준으로 골반 중심을 정규화하였다. 정규화는 해당 기준점을 원점 (0,0,0)으로 설정한 뒤, 나머지 관절 좌표를 상대 좌표계로 재정렬하는 과정이다. 이 과정을 통하여 정규화된 관절 좌표 데이터는 투구 동작을 분석하기 위하여 DTW 기반의 유사도 분석 과정에 사용된다.

## 3. Outcome-Based Comparison of Condition-Specific Representative Motions Generated via DBA Using DTW

본 과정은 투구 동작을 구종과 결과에 따라 분류하고, 투구 조건별로 대표 투구 동작을 추출하기 위한 과정이다. 또한 추출된 대표 투구 동작 간의 유사도를 비교하기 위한 방법을 제안한다. 제안 방법을 통하여 구종에 따른 투구 동작의 차이뿐만 아니라, 투구 결과에 따른 투구 동작의 차이를 정량적으로 분석하였다.

### 3.1 Representative Motion Generation Using DBA

투수 개인별로 대표 투구 동작을 추출하기 위하여 DTW 기반의 평균화 기법인 DBA를 사용하였다. DBA는 여러 종류의 시계열 3D 스켈레톤을 시간축을 기준으로 정렬한다. 그리고 각 프레임에서 유사한 시점의 스켈레톤 내 관절 좌표를 반복적으로 평균화 및 수렴하여 대표 시계열을 생성하는 방법이다. 이 방법은 단순 시간 평균보다 각 동작의 패턴을 보존하면서 평균화가 가능한 장점이 있다. 본 연구에서는 직구-변화구와 같은 구종 조합이나 스트라이크-볼 같은 결과 조합에 해당하는 투구 시계열을 분류한 뒤, DBA를 적용하여 조건별 대표 투구 동작을 도출하였다.

### 3.2 Similarity Analysis Between Representative Motions Using DTW

이 과정은 투구 조건별로 생성된 대표 투구 동작에 대하여 DTW 알고리즘을 이용하여 상호 간의 유사도를 비교하는 과정이다. 제안 방법은 3D 스켈레톤내에서 프레임 단위로 비교하여, 각 관절 좌표의 차이를 기준으로 최적의 정렬 경로를 계산한다. 이 과정을 통해 프레임 간의 시간이나 속도 차이로 발생할 수 있는 왜곡을 보정한다. 다음은 3D 스켈레톤내의 대응되는 관절 위치의 거리 차이를 계산하여 대표 투구 동작 간의 유사도를 계산한다. 이 방법은 전체적인 투구 동작 흐름뿐만 아니라 릴리스 타이밍과 같은 세부 구간에서도 구조적인 차이를 효과적으로 비교할 수 있다.

## IV. Experimental Results

본 연구에서 제안한 기법의 타당성을 입증하기 위하여 2024년 한국프로야구 경기에서 볼넷 비율이 낮은 투수와 높은 투수를 대상으로한 비교 실험을 진행하였다. 일반적으로 볼넷 비율이 낮은 투수는 투구 동작이 안정적인 투수일 가능성이 높고, 볼넷 비율이 높은 투수는 투구 동작이 불안정한 투수일 가능성이 높다. 본 실험에서는 투수 정면에서 촬영된 영상을 구종(직구, 변화구)과 투구 결과(스트라이크, 볼)에 따라 4종류로 분류하였다. 실험용 투구 영상은 구종과 투구 결과별로 60개씩, 투수 한 명당 총 240개의 영상을 수집하여 대표 투구 동작 추출을 위한 시계열 데이터를 확보하였다.

### 1. Experimental Process of 2D Skeleton Extraction and Format Conversion

본 과정에서는 투구 영상에서 2D 스켈레톤을 추출하고, 이를 시계열 데이터로 정제하여 3D 변환을 위한 포맷을 구성하기 위한 실험을 진행하였다. 실험은 HRNet을 이용한 2D 스켈레톤 추출과 투수 식별 및 3D 포즈 예측을 위하여 2D 스켈레톤 시계열을 구성하는 실험을 진행하였다.

#### 1.1 Experimental Identification of Pitcher Skeletons Using HRNet

투구 영상에서 HRNet 모델을 적용하여 각 프레임별로 출현하는 인물에 대한 2D 스켈레톤을 추출하였다. 그러나 투구 영상에는 투수 외에도 타자, 포수, 심판 등의 인물이 포함되는 경우가 있다. 본 실험에서는 각 프레임에서 머리부터 발끝까지의 y축 거리 차이가 가장 큰 사람을 투수로 식별하는 방법을 적용하였다. 이 방법을 통하여 투구 영상으로부터 투수만 추출하여 2D 스켈레톤을 효과적으로 구성할 수 있었다.

그림 2는 투구 영상에서 투수를 식별하여 투수 영상만 추출한 스켈레톤의 예이다. 그림의 왼쪽 부분은 경기 영상에서 투수, 심판, 타자가 함께 포함된 경우이다. 오른쪽 부분은 y축 좌표 차가 가장 큰 객체를 투수로 식별하고, 타자나 심판 같은 다른 인물 영상을 제거한 결과 영상이다.

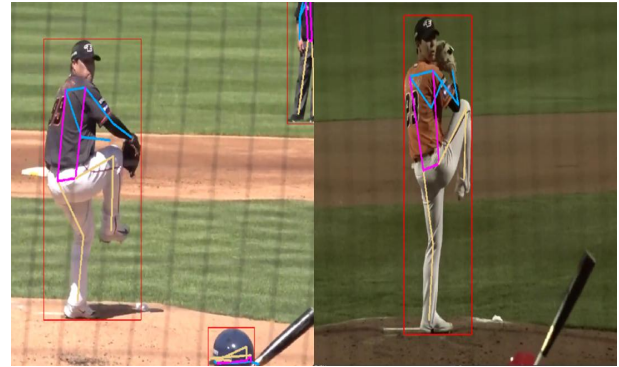


Fig. 2. Example of Extracted Skeleton After Pitcher Identification from Pitching Video

#### 1.2 Experimental Structuring of 2D Skeleton Sequences for 3D Conversion

본 실험은 식별된 투수 2D 스켈레톤을 3D 스켈레톤으로 변환하기 위하여 입력 형식으로 재구성하는 과정이다. 정렬된 데이터는 하나의 numpy 배열로 변환한 후, PoseFormerV2 모델의 입력 형식에 따라 .npz 포맷으로 저장하였다. 저장된 데이터는 프레임 수의 불일치로 인한 불균형을 방지하기 위하여 일정 프레임 수 이상 확보한 투구 동작만을 선별하여 구성하였다. 이를 통해 다양한 구종 및 결과 조합에 따른 투구 동작을 통일된 시계열 데이터로 관리할 수 있었고, 3D 변환과 시계열 기반의 유사도 분석을 위한 신뢰성 있는 데이터셋을 확보할 수 있었다.

그림 3은 HRNet 모델을 사용하여 투구 영상에서 투수만 별도로 추출한 2D 스켈레톤 구성 결과이다. 그림 3에서 빨간색으로 표시된 좌표 값은 2D 스켈레톤 내에서 17개의 관절이 위치한 부분에 대한 x, y 좌표를 의미한다. 그리고 파란색으로 표시된 부분은 17개의 관절을 연결한 투구 스켈레톤을 의미한다.

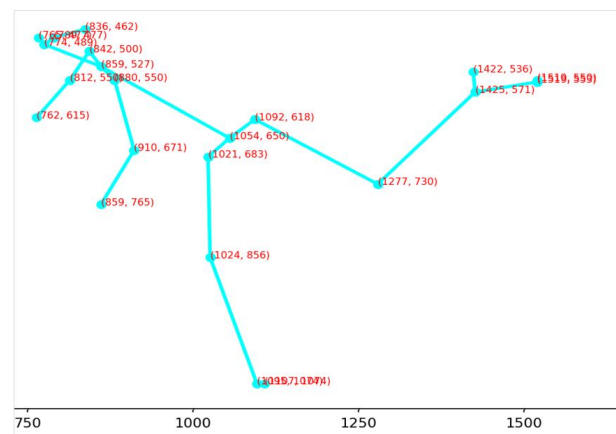


Fig. 3. 2D Skeleton Visualization for Pitcher Using HRNet

## 2. Experimental Conversion and Normalization of 3D Skeletons Using PoseFormerV2

본 실험에서는 2D 스켈레톤을 3D 스켈레톤으로 변환하고, 변환 과정에서 발생할 수 있는 위치 왜곡을 정규화하여 정밀한 시계열 데이터를 구축하는 과정을 수행하였다. 이 실험은 크게 PoseFormerV2를 활용한 3D 관절 좌표 예측과 골반 중심의 정규화를 통한 데이터 구조화의 두 단계로 구성된다.

### 2.1 Experimental 3D Skeleton Prediction Using PoseFormerV2

본 과정에서는 시계열 2D 스켈레톤을 Human3.6M 포맷으로 변환한 후, PoseFormerV2 모델에 입력하여 각 프레임의 3차원 관절 좌표를 예측하는 실험을 진행하였다. PoseFormerV2는 시간 흐름에 따른 관절 간의 관계를 학습하여, 단일 프레임 기반 분석의 한계를 극복하고 투구 동작의 입체적인 움직임 패턴을 정밀하게 추정할 수 있다. 이 과정에서 구성된 3D 스켈레톤은 각 관절에 대한 x, y, z 좌표를 포함한다. 그리고 시계열 흐름을 반영한 3차원 투구 동작 데이터로 구성된다.

### 2.2 Normalization and Data Structuring for Experimental 3D Skeleton Analysis

3D 스켈레톤 데이터는 촬영 각도나 투수 위치 변화로 인해 시계열 간에 공간적인 왜곡 현상이 발생할 수 있다. 이러한 왜곡은 투구 동작 간의 유사도 비교 과정의 정확도를 떨어뜨릴 수 있다. 본 연구에서는 좌우 엉덩이 관절의 평균 위치를 골반 중심으로 설정하고 모든 관절 좌표를 이 기준에 대해 상대적으로 재정렬하는 정규화 과정을 수행하였다. 정규화된 3D 스켈레톤 데이터는 .npz 포맷으로 저장하여, 이후 대표 투구 동작 생성(DBA)과 DTW 기반 유사도 분석 과정의 입력 데이터로 활용하였다.

그림 4는 2D 스켈레톤을 PoseFormerV2를 통해 정규화하여 완성된 시계열 3D 스켈레톤 결과를 시각화한 실험 결과 그림이다. 그림 4에서 파란색 선은 투수의 오른팔과 오른발을 의미하고, 빨간색 선은 왼팔과 왼발을 의미한다. 그리고 척추 라인은 해당 투수가 왼손 투수인 경우 빨간색으로 표시하고, 오른손 투수인 경우 파란색으로 표시한다. 그림 4는 왼손 투수가 오른발을 들고 투구하는 순간의 3D 스켈레톤을 보여주는 그림이다. 또한 3D 스켈레톤 변환 과정에서 두 팔이 겹치는 순간에도 앞뒤 프레임을 확인하여 정상적으로 변환되고 있는지를 판단할 수 있다.

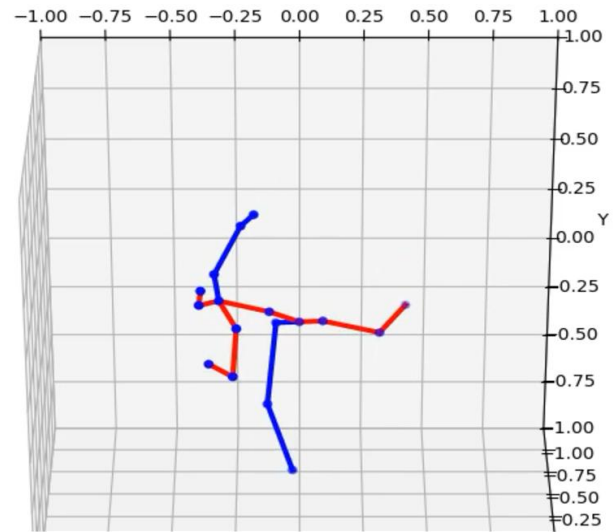


Fig. 4. 3D Coordinate Conversion Results Using PoseFormerV2 Model

## 3. Experimental Comparison of Representative Motions by Outcome Using DBA and DTW

본 과정에서는 시계열 3D 스켈레톤 데이터를 기반으로 구종(직구/변화구)과 투구 결과(스트라이크/볼)에 따라 조건별로 대표 투구 동작을 생성하는 실험을 진행하였다. 그리고 이들 간의 동작 차이를 정량적으로 비교하기 위한 실험도 함께 수행하였다. 실험은 DBA 기법을 이용하여 조건별로 대표 동작을 추출하는 과정과 DTW 기법을 이용하여 대표 동작 간의 유사도를 비교하는 과정으로 구성된다.

### 3.1 Experimental Generation of Representative Motions Using DBA

본 실험에서는 각 조건 그룹에 대해 DBA 기법을 적용하여 투수별로 대표 투구 동작을 생성하였다. DBA 기법은 반복 투구 동작 간의 시간축을 유연하게 정렬하고, 관절별로 움직임을 평균화하여 대표성을 갖는 하나의 투구 시퀀스를 생성하는 기법이다. 이 기법은 투구 시퀀스 간의 프레임 수나 리듬 차이를 보정하여 일관된 패턴을 도출할 수 있다. 이 실험을 통하여 각 조건별로 안정적이고 일관된 투구 동작을 도출할 수 있었다. 또한 이후 동작 간의 유사도 비교를 위한 기준 데이터를 확보하였다.

### 3.2 Experimental Similarity Analysis of Representative Motions Using DTW

투수 간의 투구 동작 차이를 비교 실험하기 위하여 한국 프로야구 투수 중에서 볼넷 비율이 높은 투수와 낮은 투수를 선정하여 조건별로 대표 투구 동작을 생성하였다. 생성된 대표 투구 동작 간의 유사도는 DTW 알고리즘을 적용

하여 계산하였다. DTW distance는 시계열의 길이나 정규화 방식, 관절 좌표의 분포 특성 등에 따라 절대적인 기준을 설정하기 어렵다. 이에 따라 실험 대상 간의 상대적인 DTW distance 분포에 기반한 해석 방법을 적용하였다.

실험에서는 구종-결과 조합별 대표 투구 동작 간의 유사도 비교를 통해 DTW distance 값을 수집하였다. DTW distance 분포에서 최소값 0.27236과 최대값 1.24708를 기준으로 유사도 해석 구간을 정의하였다. 전체 DTW distance의 범위(약 0.27 ~ 1.24)를 기준으로 약 0.35 간격으로 4단계 기준을 설정하였다. 표 1은 본 실험에서 정한 DTW distance에 따라 동작성 유사성을 판별하기 위한 기준이다. 표 1에서 유사도의 해석 범위를 4단계로 설정한 이유는 조건 간의 대표 투구 동작을 간략하게 비교하기 위해서이다.

Table 1. Interpretation Criteria for Motion Similarity Based on DTW Distance

Distance Range	Interpretation Criteria
0.00 ~ 0.35	High Structural Similarity
0.36 ~ 0.70	Moderate Structural Similarity
0.71 ~ 1.00	Low Structural Similarity
1.01 이상	Structural Divergence

표 2는 볼넷 비율이 낮은 투수와 볼넷 비율이 높은 투수에 대하여 4개 조건 그룹별로 유사도를 비교한 결과이다. 표 2에서처럼 두 투수 모두가 같은 구종이라도 투구 결과에 따라 투구 동작의 차이가 크게 나타났다. 예를들어, Fastball(Strike)-Fastball(Ball)의 DTW distance가 A 투수는 0.93725, B 투수는 1.10365로 일부 차이가 존재하거나 구조적인 차이가 뚜렷하게 나타났다. 반면에 서로 다른 구종이라도 투구 결과가 같은 경우에는 유사한 투구 동작이 나타났다. 표 2에서 A 투수는 Fastball(Ball)과 Curveball(Ball) 간의 DTW distance가 0.27236으로 매우 유사한 결과를 보였다. B 투수는 Fastball(Strike)과 Curveball(Strike) 간의 DTW distance가 0.32642로 비교적 유사한 결과를 보였다.

이러한 실험 결과는 구종보다는 투구 결과가 투구 동작에 더 큰 영향을 줄 수 있음을 의미한다. 즉, 투구 결과가 동일한 경우에는 릴리스 타이밍이나 체중 이동 등의 패턴도 유사하게 나타날 가능성이 높다. 또한 실험 결과에서처럼 투수별로 투구 결과 중심의 일관된 투구 동작이 존재할 수 있음을 의미한다.

Table 2. DTW-Based Similarity Across Four Pitching Conditions for Pitchers A (Low BB%) and B (High BB%)

Pitcher	Group	DTW distance
A	Fastball(S) - Fastball(B)	0.93725
	Fastball(S) - Curve(S)	0.32258
	Fastball(S) - Curve(B)	0.78792
	Fastball(B) - Curve(S)	0.82285
	Fastball(B) - Curve(B)	0.27236
	Curve(S) - Curve(B)	0.67414
B	Fastball(S) - Fastball(B)	1.10365
	Fastball(S) - Curve(S)	0.32642
	Fastball(S) - Curve(B)	1.24708
	Fastball(B) - Curve(S)	0.84855
	Fastball(B) - Curve(B)	0.41920
	Curve(S) - Curve(B)	0.93320

## V. Conclusions

본 연구에서는 야구 영상에서 투수 개인별로 투구 동작에 대한 시계열 3D 스켈레톤 변환과 유사도 분석을 통해 투구 동작에 영향을 미치는 주요 요인을 정량적으로 비교하기 위한 기법을 제안하였다. 그리고 딥러닝 기반의 시계열 스켈레톤 분석 기법을 이용하여 직구나 변화구 같은 구종이나 스트라이크나 볼과 같은 투구 결과의 조합별로 대표 투구 동작을 추출하였다. 이를 위하여, HRNet 모델을 활용하여 프레임 단위의 2D 스켈레톤을 추출하였다. PoseFormerV2 모델을 통해 3D 스켈레톤으로 변환한 후, 골반 중심 정렬을 통해 스켈레톤을 정규화하였다. 마지막으로 3D 스켈레톤에 대하여 DBA를 적용하여 각 조건에 대한 대표 투구 동작을 생성하였다.

제안 기법의 타당성을 입증하기 위하여 DTW 알고리즘을 이용하여 구종-투구 결과별로 대표 투구 동작 간의 유사성을 비교하기 위한 실험을 진행하였다. 실험 과정에서는 구종과 투구 결과별로 60개씩, 투수 한 명당 총 240개의 투구 영상을 수집하여 실제 투구와 대표 투구 동작 간의 유사도를 비교하였다. 실험 결과, 구종보다는 투구 결과에 따른 투구 동작 차이가 더 유의미한 결과를 보였다. 또한 투수 개인별로 투구 결과 중심의 일관된 투구 동작이 존재할 수 있음을 입증하였다.

본 연구 결과는 투수의 기량을 향상하기 위한 투수 맞춤형 훈련 시스템이나 개인화된 투구 동작 교정 시스템에 응용될 수 있으리라 기대된다. 또한 인공지능 기반의 실시간 코칭 시스템 등과 같이 실제 투수 훈련 환경에 적용 가능

한 실용적 기술로 응용할 수 있으리라 기대한다. 향후, 다양한 유형의 투수 영상을 확보하여 통계적인 타당성을 보완하고, 시계열 학습 모델인 TLMS(Time-series Learning Motion Similarity) 기반의 학습형 유사도 판단 모델로 확장할 계획이다.

## ACKNOWLEDGEMENT

This research was supported by Changwon National University in 2025~2026.

## REFERENCES

- [1] T. Kim, S. Lim, J. Koh, and J. Lee, "A Study on the Prediction of Win-Loss in Korean Professional Baseball Using Artificial Intelligence Models," *Journal of the Korea Big Data Society*, vol. 5, no. 2, pp. 77-84, 2020.
- [2] J. Lee, "Consistency Analysis of an AI-Based Automatic Ball-Strike System (ABS) in Korean Professional Baseball (KBO)," *Korean Journal of Physical Education and Sports Science*, vol. 33, no. 1, pp. 123-135, 2024.
- [3] K. Van Areem, "Prediction of Future Values of Player Performance KPIs in Football," *Mathematics of Data Science*, TU Delft (EEMCS), 2024.
- [4] J. Lee, J. Lee, S. Moon, A. Oh, D. Nam, and W. Yoo, "A Study on Video-Based Sports Player Tracking Using Deep Learning," *Proceedings of the Summer Conference of the Institute of Electronics Engineers of Korea (IEEK)*, pp. 1889-1891, 2020.
- [5] "Sports Video Athlete Detection Based on Deep Learning," *Artificial Intelligence Technologies in Sports and Art Data Applications*, vol. 35, pp. 4201-4210, 2023.
- [6] M. Shinya, Y. Kuroyanagi, Y. Kanosue, and S. Fujii, "Pitching form determines probabilistic structure of errors in pitch location," *Journal of Sports Sciences*, vol. 35, no. 21, pp. 2142-2147, 2017.
- [7] W. A. Laughlin, K. W. Fortenbaugh, R. L. Fleisig, and G. S. D'Lima, "The effects of baseball bat mass properties on swing mechanics, ground reaction forces, and swing timing," *Sports Biomechanics*, vol. 15, no. 1, pp. 36-47, 2016.
- [8] C. L. Camp, S. Loushin, S. Nezelek, A. P. Fiegen, D. Christoffer, and K. Kaufman, "Are wearable sensors valid and reliable for studying the baseball pitching motion? An independent comparison with marker-based motion capture," *American Journal of Sports Medicine*, vol. 49, pp. 3094-3101, 2021.
- [9] T. Yoshikawa, A. Inui, Y. Mifune, H. Nishimoto, K. Yamaura, and S. Mukohara, "Factors associated with medial elbow torque measured using a wearable sensor in junior high school baseball pitchers," *Applied Sciences*, vol. 13, no. 19, pp. 10573-10579, 2023.
- [10] A. Lizzio, G. Cross, W. Guo, and E. Makhni, "Using wearable technology to evaluate the kinetics and kinematics of the overhead throwing motion in baseball players," *Arthroscopy Techniques*, vol. 9, no. 9, pp. 429-431, 2020.
- [11] J. Bright, B. Balaji, Y. Chen, D. A. Clausi, and J. S. Zelek, "PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3420-3429, 2024.
- [12] N. Umezu and S. Akiyama, "Visualizing difference between bodily movements of athletes and users learning to play baseball," *Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 252-254, 2022.
- [13] M. Singh, M. U. Ghani, S. Kumar, and T. Baker, "Review on wearable technology in sports: Concepts, challenges and opportunities," *Computers in Biology and Medicine*, vol. 139, pp. 104854, 2021.
- [14] J. Lim and J. Kim, "A Quantitative Method for Measuring Human Motion Similarity Using Dynamic Time Warping (DTW) of Keypoint Data Extracted from Video," *Journal of the Institute of Electronics and Information Engineers*, vol. 24, no. 4, pp. 1109-1116, 2020.
- [15] J. Park, J. Kim, and Y. Woo, "A Motion Classification and Retrieval System for Baseball Game Videos Using a CNN (Convolution Neural Network) Model," *Journal of the Korea Society of Computer and Information*, vol. 26, no. 8, pp. 31-37, 2021.
- [16] M. Fraser, M. McCormack, and J. Mei, "Baseball Injury Prevention and Pitching Biomechanics," *Worcester Polytechnic Institute, Major Qualifying Project Report*, pp. 1-55, 2022.
- [17] M. Fraser, M. McCormack, and J. Mei, "Baseball Injury Prevention and Pitching Biomechanics," *Worcester Polytechnic Institute, Major Qualifying Project Report*, pp. 1-55, 2022.
- [18] 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 (CVPR)*, pp. 5693-5703, 2019.
- [19] Q. Zhao, C. Zheng, M. Liu, P. Wang, and C. Chen, "PoseFormerV2: Exploring frequency domain for efficient and robust 3D human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18658-18668, 2023.
- [20] P. Senin, "Dynamic Time Warping algorithm review," *Information and Computer Science Department, University of Hawaii, Tech. Rep. CSTR-07-89*, 2008.
- [21] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678-693, 2011

## Authors



Jae-Seung Kim received the B.S. and M.S. degrees in Computer Engineering from Changwon National University, Korea, in 2007 and 2009, respectively. Mr. Kim is a Developer & Architecture & CEO in HiBall Co.

He is interested in Software Architecture & Micro Service Architecture.



Hyung-Woo Moon received the M.S. and Ph.D. degrees in Computer Engineering from Changwon National University, Korea, in 2007 and 2014, respectively. Mr. Moon is a CTO in HiBall Co.

He is interested in Analysis of sports Bigdata & Data Architecture.



Yong-Tae Woo received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Kyungpook National University, Korea, in 1982, 1984 and 1995, respectively.

Dr. Woo is a Professor in the Department of Computer Engineering, Changwon National University since 1987. He is also CEO of Hibrain.net Co. He is interested in Data Modeling, Internet Business, and Big Data Analysis areas.