

일반논문 (Regular Paper)

방송공학회논문지 제29권 제5호, 2024년 9월 (JBE Vol.29, No.5, September 2024)

<https://doi.org/10.5909/JBE.2024.29.5.729>

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

## 모션 캡처 장비와 카메라의 시공간 매칭 시점 추천 시스템 제안

이승주<sup>a)</sup>, 박구만<sup>a)†</sup>, 김성제<sup>b)</sup>, 이수현<sup>c)</sup>

### Proposal of a Recommendation System for Time-space Matching of Motion Capture Device and Camera

Seungju Lee<sup>a)</sup>, Gooman Park<sup>a)†</sup>, Sungjei Kim<sup>b)</sup>, and Suhyeon Lee<sup>c)</sup>

#### 요약

본 논문에서는 모션 캡처 장비와 카메라 간의 시간적, 공간적 동기화를 위한 파라미터 추천 시스템을 제안한다. 제안된 시스템에서는 모션 캡처 데이터(3D)와 카메라 데이터(2D)를 이미지상으로 프로젝션하여 각각의 관절점(=keypoints) 데이터를 2D 형태로 생성 후, PDJ 평가 매트릭스를 통하여 관절점이 유사한 프레임 범위를 추천한다. 추천된 프레임은 두 장비의 프레임 차이를 나타낸다. 이 시스템을 이용하여 포즈 추정에서 필요한 관절점 labeling에 도움이 될 것으로 기대된다.

#### Abstract

This paper presents a parameter recommendation system for temporal and spatial synchronization between motion capture equipment and camera. The proposed system employs a two-step process. First, the motion capture data (3D) and camera data (2D) are projected onto the image to generate the respective joint point data (keypoints) in 2D form. Second, the PDJ evaluation matrix is used to recommend the frame range where the joint points are similar. The recommended frame represents the frame difference between the two devices. It is anticipated that this system will facilitate the labeling of the joint points required for pose estimation.

Keyword : Pose Estimation, Camera Intrinsic/Extrinsic parameters, Motion Capture, Sensor Synchronization

a) 서울과학기술대학교 일반대학원 스마트ICT융합공학과(Department of Smart ICT Convergence Engineering, Seoul National University of Science and Technology, Republic of Korea)

b) 한국기술교육대학교 컴퓨터공학부(Computer Science and Engineering, Korea University of Technology and Education)

c) 한국전자기술연구원(Korea Electronics Technology Institute, KETI)

† Corresponding Author : 박구만(Gooman Park)

E-mail: [gmpark@seoultech.ac.kr](mailto:gmpark@seoultech.ac.kr)

Tel: +82-2-970-6430

ORCID: <https://orcid.org/0000-0002-7055-5568>

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2023년도 문화기술 연구개발 사업으로 수행되었음(과제명: 메타버스 기반 스포츠 플레이 트위닝 및 실감화 기술 개발, 과제번호: R2022020052, 기여율: 100%)

· Manuscript August 22, 2024; Revised September 5, 2024; Accepted September 6, 2024.

Copyright © 2024 Korean Institute of Broadcast and Media Engineers. All rights reserved.

"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

## I. 서론

영상 내에서 관절점을 찾기 위해서는 Pose estimation을 위한 데이터 셋이 필요하다. 이미 MS COCO<sup>[1]</sup>, MPII Human Pose<sup>[2]</sup>, 3DPW<sup>[3]</sup>, CrowdPose<sup>[4]</sup> 등 2D 관절점 (keypoints) 뿐만 아니라 3D 관절점 데이터도 잘 구축되어 있다. 그러나 디지털 콘텐츠나 메타버스<sup>[5]</sup> 등의 상황에서 캐릭터가 인체와 다른 관절 구조를 가지고 있거나 관절의 위치가 다른 경우에는 사용하는 목적에 따라 여전히 데이터 셋의 구축이 필요하며, 데이터 구축 시 인적 물적 자원이 필요하다. 본 논문에서는 데이터 구축 시 소요되는 자원들을 줄이기 위하여, 모션 캡처 장비와 카메라의 시간축과 공간축을 동기화하는 파라미터를 추천하는 시스템을 제안한다. 제안된 시스템은 서로 다른 장비에서 취득한 관절 정보를 동일 시점으로 동기화하여, 시나리오별로 두 장비의 관절점을 선택적으로 사용할 수 있다. 예를들어, 시나리오마다 해당 시점에서 이미지로 추출한 관절점의 정확도가 높은 경우 이미지에서 추출한 관절점을 사용하고, 반대로 모션 캡처의 정확도가 높은 경우 3D 좌표를 직접 사용하거나 2D의 관절점으로 프로젝션하여 사용할 수 있다. 모션 캡처는 몸에 센서를 부착하거나, 적외선을 이용하는 등의 방법으로 인체의 움직임을 디지털 형태로 저장한다. 이러한 모션 데이터를 캐릭터에 입력하여 애니메이션을 제작<sup>[6]</sup>하거나, 의료 목적으로 자세 교정<sup>[7]</sup>하는 등에 사용된다. 본 시스템에서는 모션 캡처의 출력인 3D 관절점과 포즈 추정 출력인 2D 관절점을 매칭하여, 두 데이터가 유사한 Pose를 취하는 공간적 데이터인 카메라 외부 파라미터를 찾고 그에 해당하는 프레임 인덱스인 시간축을 찾아서 파라미터 추천한다. 제안된 시스템에서는 모션 캡처 장비 1대와 카메라 1대를 동기화하였지만, 2대 이상의 카메라로 확장이 가능하다.

## II. 관련연구

### 1. 포즈 추정

제안하는 시스템은 2D pose estimation 알고리즘을 사용하여, 영상에서의 관절점을 추정한다. 2D Pose estimation에서

는 2가지 접근 방법이 있다. 첫 번째, Top-down 방식<sup>[8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]</sup>은 영상에서 사람을 찾은 후 사람이 존재하는 영역(Bounding Box)에서 관절점을 추출 후 Pose 정보를 취득한다. 이 방식은 정확도가 높으나 사람을 인식하지 못하면 관절점을 추정할 수 없으며, 영상 내 사람수에 따라 연산량이 선형 증가한다. 두 번째, Bottom-up 방식<sup>[20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]</sup>은 영상에서 관절점을 찾은 후에 각 관절점끼리 연결하여 Pose 정보를 취득한다. 이 방식은 관절점을 매칭할 수 있는 조합의 수가 많기 때문에 매칭하는데 시간이 오래걸리지만, 사람을 찾는 과정이 없기 때문에 실시간 시스템에 적용이 가능하다. 본 시스템에서는 관절점의 추정 정확도가 실시간 처리보다 더 중요하기 때문에 Top-down 방식의 AlphaPose<sup>[19]</sup>를 채택하였으며, Halpe 데이터 셋으로 학습한 pre-trained model를 사용한다. 추가적으로 3D pose estimation 알고리즘은 단일 이미지에서 3D pose 재구성 방법<sup>[31, 32, 33]</sup>과 multi-view 또는 RGBD 데이터를 통한 3D pose 추정 방법<sup>[34, 35]</sup>이 있다. 해당 분야의 연구는 주로 객체 검출 문제로부터 자유롭고 데이터 셋 확보가 용이한 단일 이미지에서의 3D pose 재구성 방법이 활발하게 연구되고 있다.

### 2. 관절점 평가 지표

데이터 셋 또는 장비마다 관절점의 정의가 다르다. 그림 1은 COCO 17 keypoints와 Halpe 26 keypoints, 모션 캡처의 관절점을 나타낸다. COCO와 Halpe는 각각 17, 26개의 관절점으로 정의되며, 모션 캡처 장비는 Neuronmocap사의 perception neuron studio system 기준으로 총 59개의 관절점으로 구성되어 있으나, 손가락 관련 관절점을 제외하면 21개의 관절점으로 구성되어 있다. 결론적으로 각각의 데이터 셋의 관절점 위치와 관절점 수가 완전하게 같지 않다는 것을 확인할 수 있다.

다양한 관절점 평가 지표가 존재하지만, 대표적인 3가지에 대해서 정리하였다. 첫 번째, Percentage of Correct Parts (PCP)는 예측한 관절 값이 ground-truth segment 길이의 50% 내에 있는 경우 참으로 판별한다. 예를들어, 팔꿈치를 측정할 경우, 팔꿈치에서 반경(=팔의 길이의 절반을 반지름으로 하는 원)내에 예측 결과가 있으면 참으로 판별한다. 이 방식은 관절 사이의 거리가 가까우면 검출이 어려워

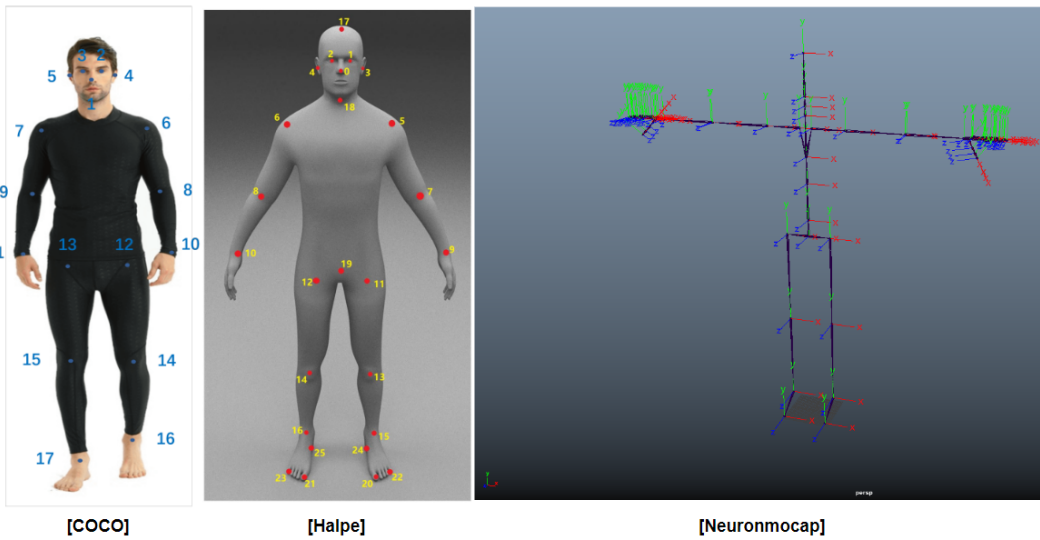


그림 1. COCO<sup>[1]</sup>, Halpe<sup>[19]</sup>, Motion capture 데이터 비교  
 Fig. 1. Comparison of COCO<sup>[1]</sup>, Halpe<sup>[19]</sup> and Motion capture data

지는 단점이 있다. 수식은  $PCP = (\text{전체 데이터 셋에 대한 올바른 관절}) / (\text{전체 데이터 셋에 대한 총 관절})$ 으로 표현된다. 두 번째, Percentage of Correct Keypoints (PCK)는 관절의 예측값과 정답값 사이 거리가 임계값보다 작은 경우를 참으로 판별한다. 여기서 임계값은 신체 부분의 길이를 사용한다. 예를 들어,  $PCKh @0.5$ 는 머리 사이즈의 50% 길이를 임계값으로 사용한다. 수식은  $PCKh @0.5 = 0.5 * \text{머리 직경 길이}$ 로 표현된다. 세 번째, Percentage of Detected Joints (PDJ)는 관절의 예측값과 정답값 사이 거리가 임계값보다 작은 경우를 참으로 판별한다. 임계값은 사람의 대각선 길이와 특정 임계값을 곱하여 사용한다. 예를 들어, 사람의 신체 길이가 대각선으로 500px 이고 특정 임계값으로 0.05를 사용하면, 각 관절은 정답값을 원의 중심으로하여 반경 25px 내에 예측값이 있으면 참으로 판별한다. 본 시스템에서는 해당 평가 지표를 사용한다. 수식

$$PDJ = \frac{\sum_{i=1}^n \text{bool}(d_i < 0.05 * \text{diagonal})}{n}$$

으로 표현된다.

### 3. 모션 동기화 시스템

CHATZITOFIS, Anargyros<sup>[36]</sup>은 MS k4A 3대를 이용하여 실시간 처리가 가능한 모션 동기화 시스템이다. 해당 논문

의 한계점은 복잡한 촬영 환경 설정과 고성능의 데스크탑이 필요하여 데이터 획득 과정 자체가 복잡하다. 우리가 제안하는 논문은 실시간 처리는 불가능하지만 촬영 환경 설정이 간편하다. RIECICKY, Adam<sup>[37]</sup>은 모션 캡처 장비에서 추출한 skeleton의 실루엣과 카메라에서 취득한 이미지 실루엣을 기반으로 매칭한다. 해당 논문의 한계점은 촬영 배경이 일정해야 하며, 실루엣을 매칭하는 search direction을 사전 정의해야 한다. 우리가 제안하는 방법은 촬영 배경과 상관없이 입력을 기반으로 모션 캡처 장비와 카메라를 end-to-end로 출력한다. ZHANG, Zhe<sup>[38]</sup>은 멀티 카메라를 이용하여 포즈 추정 (HigherHRNet<sup>[21]</sup>)을 수행하고 수행된 결과인 관절점을 이용하여 하나의 graph로 생성한다. 생성된 graph를 이용하여 energy distance function이 낮아지는 지점으로 동기화한다. 본 논문에서 제안하는 방법은 정밀 값을 추정하기 위해 모션 캡처 장비와 영상에서의 관절점의 위치를 동기화한다.

## III. 시공간 매칭 시점 추천 시스템

### 1. 전체 시스템

본 논문에서 제안하는 시스템은 그림 2와 같다. 첫 번째

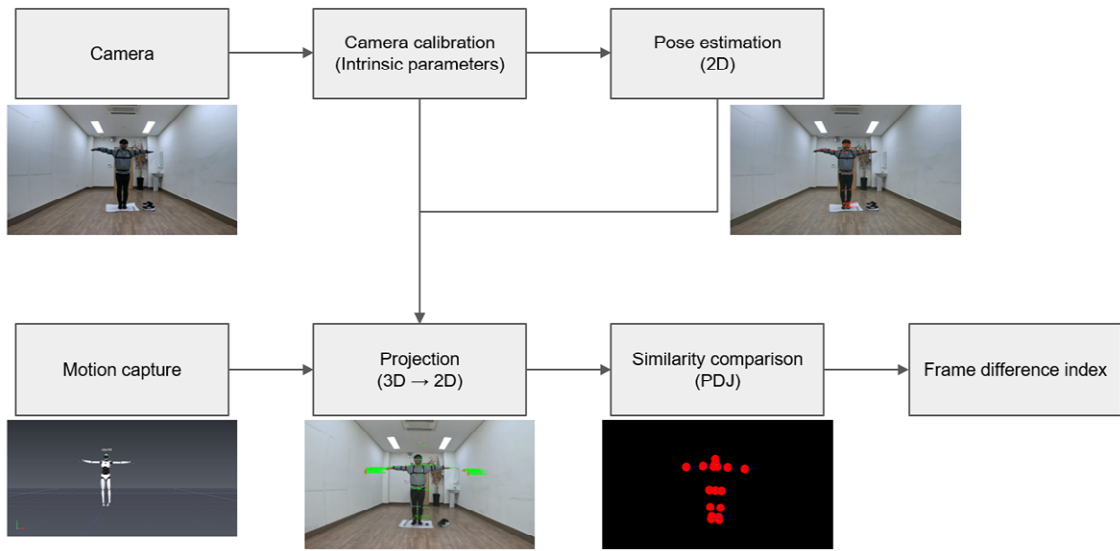


그림 2. 전체 시스템 구성도  
Fig. 2. Overall system configuration diagram

로 카메라에서 획득한 동영상과 모션 캡처 장비에서 각각 T-pose에 해당하는 프레임 인덱스를 찾는다. 두 번째로 동영상에서 Camera calibration을 수행하고, Pose estimation을 수행하여 동영상에서 관절점을 추출한다. 세 번째로 Camera calibration에서 얻은 내부 파라미터 값과 Pose estimation에서 얻은 관절점(2D point)을 사용하여 외부 파라미터 값을 추출한다. 추출된 외부 파라미터 값을 이용하여 모션 캡처 데이터(3D point)를 이미지상의 2D 좌표로 프로젝션하여 모션 캡처의 관절점을 추출한다. 마지막으로 각

각의 데이터에서 추출된 2D 상의 관절점을 PDJ 평가를 통하여 유사도가 제일 높은 프레임 인덱스의 차이를 출력한다. 최종적 출력되는 프레임 인덱스는 모션 캡처와 카메라의 시간 차이를 나타내는 수치이다.

## 2. 모듈

Camera calibration은 모션 캡처를 3D point에서 2D point로 프로젝션하기 위해 사용된다. 체커보드를 이용하

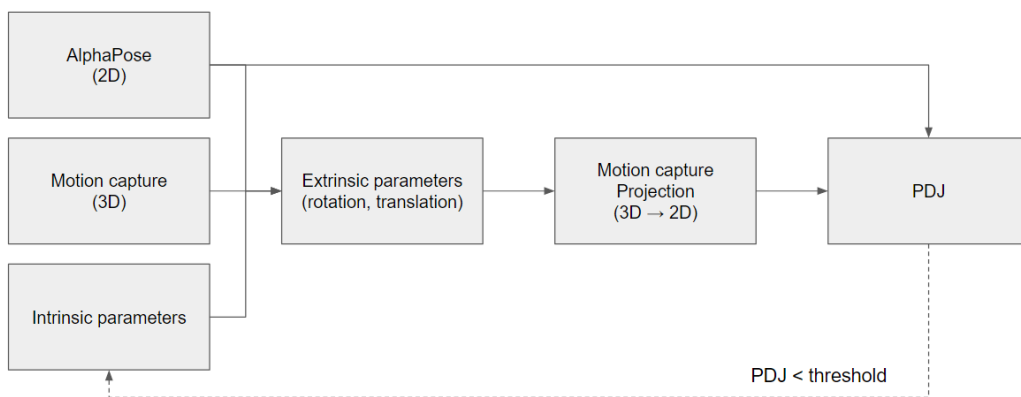


그림 3. Projection 모듈. 점선은 조건문이 참일 때 수행됨  
Fig. 3. Projection module. The dotted line is executed when the conditional statement is true

여, 카메라의 내부 파라미터를 추정한다. Pose estimation은 모션 캡처 데이터와 동영상 데이터의 관절점을 비교하기 위해서, 동영상의 관절점을 2D Pose estimation으로 추정한다. 사용된 2D pose estimation 알고리즘은 Alphapose이며, Halpe 데이터 셋으로 학습한 model을 사용한다. Projection은 그림 3처럼 작동한다. 자세히 설명하면, 모션 캡처의 3D point를 2D point로 정밀하게 프로젝션하기 위해, 다양한 카메라 외부 파라미터 값을 적용한 결과를 PDJ를 통하여 평가 후 최적의 값을 반환한다. 구체적으로 그림 3의 순서도를 보면, 첫 번째로 T-Pose 구간의 AlphaPose(동영상) 프레임과 모션 캡처 프레임 그리고 내부 파라미터를 입력받아 외부 파라미터(rotation, translation vector)를 추정한다. 추정된 외부 파라미터를 모션 캡처 데이터(3D Point)에 프로젝션하여 2D Point로 변환한다. 그리고 프로젝션된 모션 캡처 데이터(2D point)와 AlphaPose 데이터(2D point)를 PDJ 평가 매트릭스를 통하여 관절점의 유사도를 측정한다. 마지막으로 측정된 PDJ 수치가 임계값보다 작다면 해당 모듈의 전 과정을 다시 반복한다. Similarity comparison은 Projection 모듈에서의 모션 캡처 범위는 각 데이터의 T-pose 프레임에 대한 PDJ 평가이며, 본 모듈에서는 각각

의 데이터의 전체 프레임에 대한 관절점 유사도 평가이다. 빠른 평가를 위하여, 동영상 프레임의 절반과 모션 캡처 프레임의 절반을 사용하여 평균 PDJ 값과 프레임 인덱스 차이를 저장하였으며, 최종적으로 평균 PDJ값이 높은 프레임 인덱스 차이(Frame difference index) 값을 추천한다. 예를 들어, 동영상 데이터는 총 100 프레임이고 모션 캡처는 총 50 프레임으로 가정하면, 첫 번째 루프는 프레임 차이가 0인 범위를 탐색한다. 즉, 동영상 프레임이 1 ~ 50 프레임이고 모션 캡처 프레임이 1~ 50이다. 두 번째 루프는 프레임 차이가 1인 범위를 탐색한다. 동영상 프레임은 2~51, 모션 캡처 프레임은 1~50을 탐색한다. 이러한 방법으로 50 프레임 차이까지 반복한다.

#### IV. 실험

##### 1. 실험 환경

모션 캡처 장비는 Perception Neuron Studio Inertial System을 사용하였고, 카메라는 GoPro를 사용하였다. 데

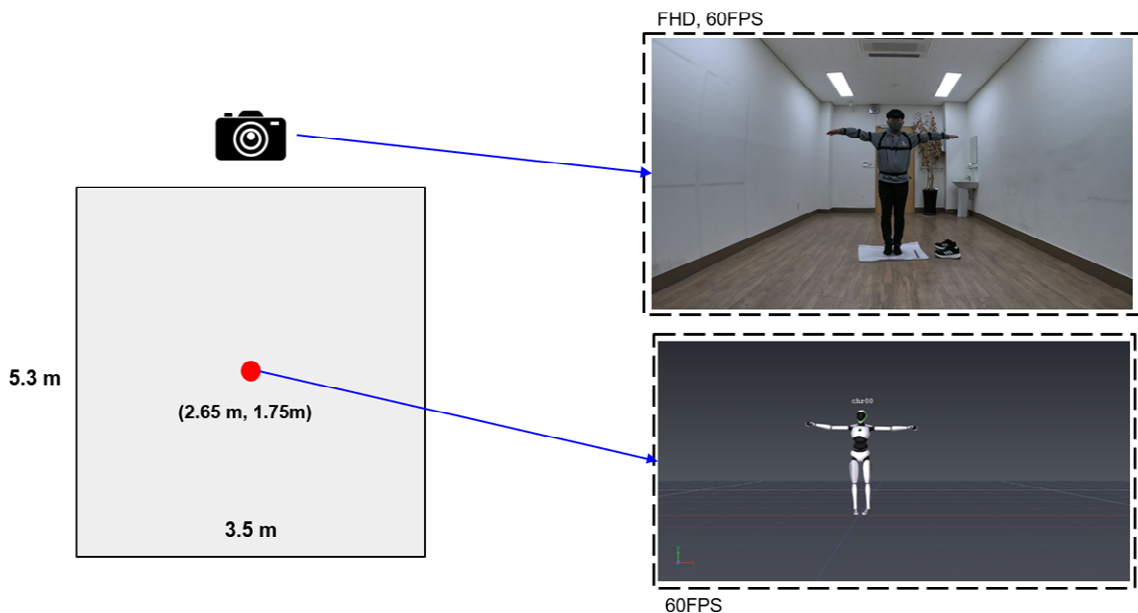


그림 4. 데이터 취득 환경

Fig. 4. Data acquisition environment

이터 취득 환경은 그림 4와 같다. 전체 공간은 가로 3.5m, 세로 5.3m이며, 카메라의 위치는 상단 중앙(가로 2.65m, 세로 0m)에 위치한다. 두 장비 모두 60FPS로 설정하고 카메라는 FHD 화질로 설정하였다. 촬영 방법은 카메라를 먼저 작동시키고 모션 캡처 장비를 나중에 작동시켰다. 즉, 카메라의 영상 프레임 길이가 모션 캡처의 프레임 길이보다 길다.

### 2. 관절점 재정의

그림 2를 보면 Alphapose의 관절점과 모션 캡처의 관절점은 위치와 수가 다르다. 따라서, 두 데이터에서 유사한 위치에 있는 관절점을 선별하여 총 17개의 관절점을 사용한다. 17개의 관절점은 표 1과 같다.

### 3. 결과 및 분석

그림 5는 모션 캡처 장비 1대와 카메라 장비 2대를 앞과 뒤에 배치하여 촬영하고 시간축으로 동기화한 결과이다. 왼쪽 그림은 동기화 전이고, 오른쪽은 동기화 후 결과이다. 동기화 후의 왼쪽 그림의 상단은 모션 캡처와 Alphapose의 유사한 17개의 관절점만을 출력한 결과이며, 왼쪽 그림의 하단은 모션 캡처의 관절점을 모두 출력한 결과이다. 나머지 동기화 후의 오른쪽 상단과 하단 그림은 스윙동작에서의 모션 캡처 관절점을 출력한 결과이다.

표 1. 모션 캡처와 Alphapose의 관절점 재정의

Table 1. Redefinition of keypoints of Motion capture and Alphapose

Index	Alphapose	Motion Capture
0	LShoulder	LeftShoulder
1	RShoulder	RightShoulder
2	LElbow	LeftArm
3	RElbow	RightArm
4	LWrist	LeftForeArm
5	RWrist	RightForeArm
6	LHip	LeftUpLeg
7	RHip	RightUpLeg
8	LKnee	LeftLeg
9	Rknee	RightLeg
10	Head	Head
11	Neck	Neck
12	Hip	Hips
13	LBigToe	LeftFoot_end
14	RBigToe	RightFoot_end
15	LHeel	LeftFoot
16	RHeel	RightFoot

그림 6 (a)는 임계값에 따른 프레임 차이와 빈도를 나타낸다. 프레임 차이는 모션 캡처와 카메라 간의 동일시점에 대한 프레임 인덱스를 의미하며, 임계값은 각 장비의 관절 데이터에 대한 유사도 측정(PDJ)시 사용되는 임계값을 의미한다. 빈도값은 두 장비의 관절 데이터에 대한 PDJ 값이 임계값 이상인 경우의 빈도를 나타낸다. 두 장비의 프레임

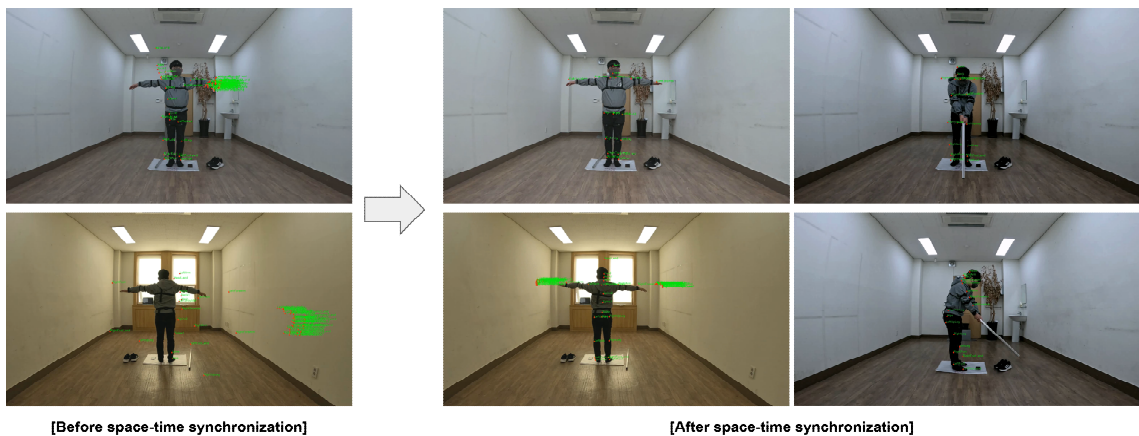


그림 5. 시간축과 공간축 동기화 결과. 오른쪽 상단은 표 1의 관절점만을 출력하였고 오른쪽 하단은 모션 캡처의 전체 관절점을 출력함  
Fig. 5. Time and space axis synchronization results. The top right outputs only the joint points in Table 1, and the bottom right outputs all joint points of motion capture

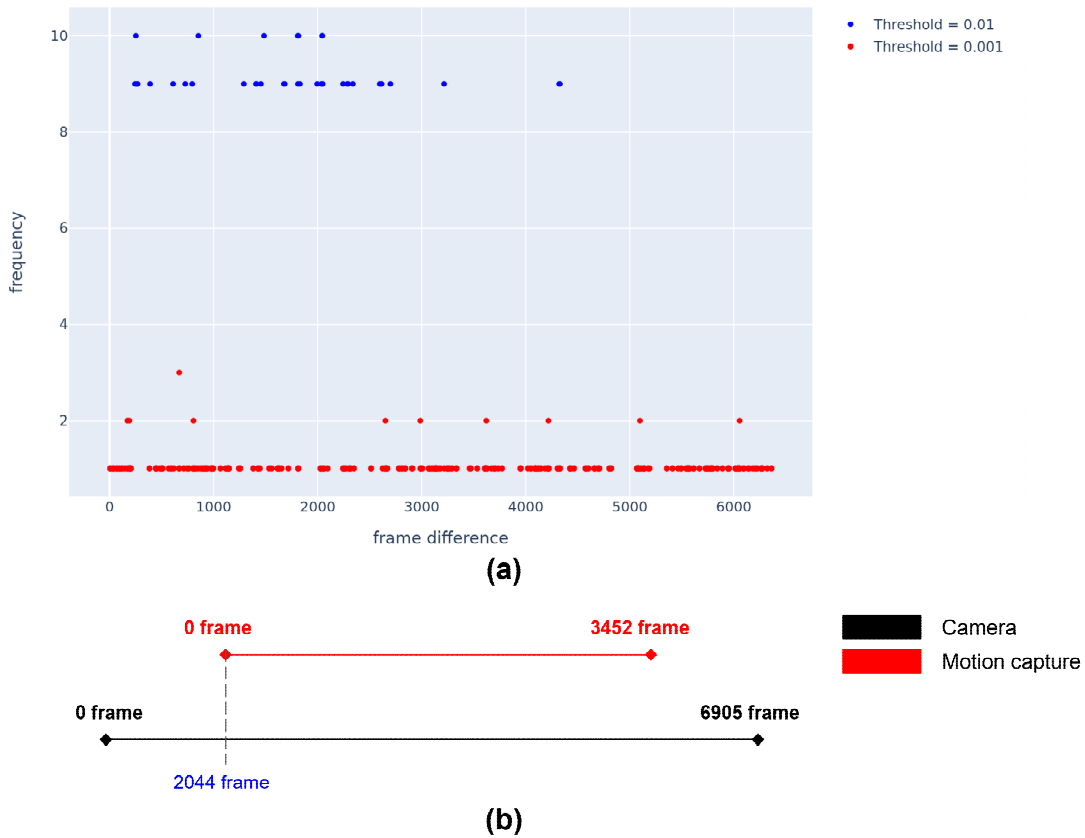


그림 6. (a) 임계값 별 프레임 차이. 프레임 차이는 모션 캡처의 프레임과 카메라의 프레임의 차이임. (b) 카메라와 모션 캡처의 동일시점 프레임 인덱스

Fig. 6. (a) Frame difference by threshold. Frame difference is the difference between the motion capture frame and the camera frame. (b) Same point frame index for camera and motion capture

차이는 2044이며, 임계값 0.01(파란색)에 빈도수 10에서 가장 오른쪽에 있는 점을 의미한다. 해당 내용을 자세하게 설명하기 위하여 그림 6 (b)를 첨부하였다. 각 선들은 카메라와 모션 캡처의 프레임 수를 나타내며, 카메라의 프레임 수가 모션 캡처의 프레임 수보다 길다. 이는 앞서 실험 환경에서 설명한 내용이다. 모션 캡처의 0 프레임은 카메라의 2044 프레임과 같은 시점이므로 임계값이 0.01일 때 정확한 시간축을 동기화하는 것을 확인할 수 있다. 그러나 그림 6 (a)의 그래프를 보면, PDJ 임계값에 따라 추천하는 프레임의 차이가 다른 문제가 있다. 이는 두 데이터 셋의 관절점 위치가 다르기 때문에 발생하는 문제로 확인된다. 즉, 두 데이터 셋의 관절점을 사용하여 카메라 외부 파라미터를 추정하는데, 관절점이 100% 같은 위치가 아니기 때문에 발

생되는 문제이다. 따라서, 현재의 시스템에서는 임계값 0.01에서는 잘 작동하지만, 다른 임계값에서는 문제가 발생할 수 있다.

## V. 결론

제안된 시스템에서 모션 캡처와 카메라 간의 시간축과 공간축을 동기화하는 파라미터를 추천 및 두 데이터가 동기화된 결과를 실험을 통해 확인하였다. 그러나 PDJ의 임계값에 따라 추천되는 프레임 차이의 인덱스가 달라지며, 정확한 동기화 시간이 아닌 여러 범위를 추천하는 한계가 있다. 이러한 문제는 Halpe 데이터 셋과 모션 캡처의 데이

터 셋의 관절점의 정의가 다르므로 인해 PDJ 값이 정확하게 평가되지 않는 것으로 실험결과를 통해 확인하였다. 따라서 다음 연구에서는 동일한 관절점으로 모델을 학습하여, 정확한 동기화 시간을 예측하는 시스템과 3D 관절점을 이용한 동기화 시스템을 제안할 예정이다.

### 참 고 문 헌 (References)

- [1] LIN, Tsung-Yi, et al. Microsoft coco: Common objects in context. In: Computer Vision - ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014. p. 740-755.  
doi: [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48)
- [2] ANDRILUKA, Mykhaylo, et al. 2d human pose estimation: New benchmark and state of the art analysis. In: Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. 2014. p. 3686-3693.
- [3] VON MARCARD, Timo, et al. Recovering accurate 3d human pose in the wild using imus and a moving camera. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 601-617.
- [4] LI, Jiefeng, et al. Crowdpose: Efficient crowded scenes pose estimation and a new benchmark. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 10863-10872.
- [5] JIANG, Ling, et al. SADNet: Generating immersive virtual reality avatars by real time monocular pose estimation. Computer Animation and Virtual Worlds, 2024, 35.3: e2233.  
doi: <https://doi.org/10.1002/cav.2233>
- [6] KUMARAPU, Laxman; MUKHERJEE, Prerana. Animepose: Multi-person 3d pose estimation and animation. Pattern Recognition Letters, 2021, 147: 16-24.  
doi: <https://doi.org/10.1016/j.patrec.2021.03.028>
- [7] KANASE, Rahul Ravikant, et al. Pose estimation and correcting exercise posture. In: ITM Web of Conferences. EDP Sciences, 2021. p. 03031.  
doi: <https://doi.org/10.1051/itmconf/20214003031>
- [8] CAI, Yuanhao, et al. Learning delicate local representations for multi-person pose estimation. In: Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part III 16. Springer International Publishing, 2020. p. 455-472.  
doi: [https://doi.org/10.1007/978-3-030-58580-8\\_27](https://doi.org/10.1007/978-3-030-58580-8_27)
- [9] HUANG, Junjie, et al. The devil is in the details: Delving into unbiased data processing for human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 5700-5709.
- [10] HUANG, Shaoli; GONG, Mingming; TAO, Dacheng. A coarse-fine network for keypoint localization. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 3028-3037.
- [11] LI, Wenbo, et al. Rethinking on multi-stage networks for human pose estimation. arXiv preprint arXiv:1901.00148, 2019.  
doi: <https://doi.org/10.48550/arXiv.1901.00148>
- [12] LIU, Huajun, et al. Polarized self-attention: Towards high-quality pixel-wise regression. arXiv preprint arXiv:2107.00782, 2021.  
doi: <https://doi.org/10.48550/arXiv.2107.00782>
- [13] MOON, Gyeongsik; CHANG, Ju Yong; LEE, Kyoung Mu. Posefix: Model-agnostic general human pose refinement network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019. p. 7773-7781.
- [14] PAPANDREOU, George, et al. Towards accurate multi-person pose estimation in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4903-4911.
- [15] SUN, Ke, et al. Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 5693-5703.
- [16] WANG, Jian, et al. Graph-pcnn: Two stage human pose estimation with graph pose refinement. In: Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part XI 16. Springer International Publishing, 2020. p. 492-508.  
doi: [https://doi.org/10.1007/978-3-030-58621-8\\_29](https://doi.org/10.1007/978-3-030-58621-8_29)
- [17] XIAO, Bin; WU, Haiping; WEI, Yichen. Simple baselines for human pose estimation and tracking. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 466-481.
- [18] ZHANG, Feng, et al. Distribution-aware coordinate representation for human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 7093-7102.
- [19] FANG, Hao-Shu, et al. Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45.6: 7157-7173.  
doi: <https://doi.org/10.1109/TPAMI.2022.3222784>
- [20] CAO, Zhe, et al. Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 7291-7299.  
doi: <https://doi.org/10.48550/arXiv.1611.08050>
- [21] CHENG, Bowen, et al. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 5386-5395.  
doi: <https://doi.org/10.48550/arXiv.1908.10357>
- [22] FIERARU, Mihai, et al. Learning to refine human pose estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018. p. 205-214.  
doi: <https://doi.org/10.48550/arXiv.1804.07909>
- [23] INSAFUTDINOV, Eldar, et al. Arttrack: Articulated multi-person tracking in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 6457-6465.  
doi: <https://doi.org/10.48550/arXiv.1612.01465>
- [24] JIN, Sheng, et al. Differentiable hierarchical graph grouping for multi-person pose estimation. In: Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23 - 28, 2020, Proceedings, Part VII 16. Springer International Publishing, 2020. p. 718-734.



- doi: [https://doi.org/10.1007/978-3-030-58571-6\\_42](https://doi.org/10.1007/978-3-030-58571-6_42)
- [25] KREISS, Sven; BERTONI, Lorenzo; ALAHI, Alexandre. Pifpaf: Composite fields for human pose estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 11977-11986.
- [26] NEWELL, Alejandro; HUANG, Zhiao; DENG, Jia. Associative embedding: End-to-end learning for joint detection and grouping. Advances in neural information processing systems, 2017, 30.
- [27] NIE, Xuecheng, et al. Single-stage multi-person pose machines. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 6951-6960.
- [28] TIAN, Z.; CHEN, H.; SHEN, C. Directpose: Direct end-to-end multi-person pose estimation. arXiv 2019. arXiv preprint arXiv:1911.07451, 449.  
doi: <https://doi.org/10.48550/arXiv.1911.07451>
- [29] WANG, Haixin, et al. Regularizing vector embedding in bottom-up human pose estimation. In: European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022. p. 107-122.  
doi: [https://doi.org/10.1007/978-3-031-20068-7\\_7](https://doi.org/10.1007/978-3-031-20068-7_7)
- [30] WANG, Yihan, et al. Lite pose: Efficient architecture design for 2d human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. p. 13126-13136.  
doi: <https://doi.org/10.48550/arXiv.2205.01271>
- [31] MEHTA, Dushyant, et al. Vnect: Real-time 3d human pose estimation with a single rgb camera. Acm transactions on graphics (tog), 2017, 36.4: 1-14.  
doi: <https://doi.org/10.1145/3072959.3073596>
- [32] ZHANG, Jinlu, et al. Mixste: Seq2seq mixed spatio-temporal encoder for 3d human pose estimation in video. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 13232-13242.  
doi: <https://doi.org/10.48550/arXiv.2203.00859>
- [33] ZHU, Wentao, et al. Motionbert: A unified perspective on learning human motion representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. p. 15085-15099.  
doi: <https://doi.org/10.48550/arXiv.2210.06551>
- [34] ZHANG, Jianfeng, et al. Direct multi-view multi-person 3d pose estimation. Advances in Neural Information Processing Systems, 2021, 34: 13153-13164.
- [35] WU, Size, et al. Graph-based 3d multi-person pose estimation using multi-view images. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021. p. 11148-11157.  
doi: <https://doi.org/10.48550/arXiv.2109.05885>
- [36] CHATZITOFIS, Anargyros, et al. A low-cost & real-time motion capture system. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. p. 21453-21458.
- [37] RIECICKÝ, Adam, et al. Optical-inertial Synchronization of MoCap Suit with Single Camera Setup for Reliable Position Tracking. In: VISIGRAPP (1: GRAPP). 2018. p. 40-47.  
doi: <https://doi.org/10.5220/0006531100400047>
- [38] ZHANG, Zhe; WANG, Chunyu; QIN, Wenhui. Semantically Synchronizing Multiple-Camera Systems with Human Pose Estimation. Sensors, 2021, 21.7: 2464.  
doi: <https://doi.org/10.3390/s21072464>

---

저 자 소 개

---



**이 승 주**

- 2016년 2월 : 상명대학교 컴퓨터소프트웨어공학과 공학사
- 2020년 8월 : 서울과학기술대학교 스마트ICT융합공학과 공학석사
- 2023년 9월 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 박사과정
- ORCID : <https://orcid.org/0000-0001-8969-4296>
- 주관심분야 : 컴퓨터비전, Video Analytics, Object Detection



**박 구 만**

- 1984년 2월 : 한국항공대학교 전자공학과
- 1986년 2월 : 연세대학교 전자공학과 공학석사
- 1991년 2월 : 연세대학교 전자공학과 공학박사
- 1991년 3월 ~ 1996년 9월 : 삼성전자 신호처리연구소 선임연구원
- 2016년 1월 ~ 2017년 12월 : 서울과학기술대학교 나노IT디자인융합대학원
- 1999년 8월 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 교수
- 2006년 1월 ~ 2007년 8월 : Georgia Institute of Technology Dept.of Electrical and Computer Engineering, Visiting Scholar
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터비전, 지능형실감미디어

---

저 자 소 개



김 성 제

- 2004년 2월 : 연세대학교 전기전자공학과 학사
- 2006년 8월 : 연세대학교 전기전자공학과 석사
- 2011년 8월 : 연세대학교 전기전자공학과 박사
- 2011년 3월 ~ 2015년 8월 : 삼성전자 S.LSI 책임연구원
- 2015년 9월 ~ 2024년 8월 : 한국전자기술연구원 책임연구원
- 2024년 9월 ~ 현재 : 한국기술교육대학교 컴퓨터공학부 조교수
- ORCID : <https://orcid.org/0009-0008-6466-649X>
- 주관심분야 : 영상/비디오 신호처리, 온-디바이스 AI



이 수 현

- 2023년 8월 : 연세대학교 박사
- 2023년 9월 ~ 현재 : 한국전자기술연구원 선임 연구원
- ORCID : <https://orcid.org/0000-0003-1989-9004>
- 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습