Video-based Accurate Human Kinematics Estimation during High-Speed Impact

Qiantailang Yuan, Svein Kleiven, Xiaogai Li

I. INTRODUCTION

Kinematics and posture estimation are crucial for injury reconstruction and enhancing safety features. Current markerless motion-capture systems, such as RGB-Depth cameras, have limitations in precision, accuracy and capture range. Existing deep learning pose estimation methods also face challenges, such as lacking depth information and limited generalisability across datasets. To address these limitations, this short communication presents a novel markerless multi-cam motion-capture system via deep learning pose estimation algorithms. The applicability and robustness of this system are demonstrated by analysing the pose configuration of fast and dynamic human movements in sports.

II. METHODS

Participants and Experimental Setup

Two healthy subjects, wearing protective gear, performed high-speed hockey tackles in a controlled environment (Fig. 1 (a)). Data were collected simultaneously using marker-based and markerless motion-capture systems. A total of 25 trials were recorded, with the marker-based system data serving as ground truth. Trials included tackle motions from different recording angles with different velocities to test the system's ability to track complex movements.

Data Collection and System Architecture

The markerless system utilised four GoPro HERO 10 cameras, with a 2704 x 1520 resolution and 240 fps sampling rate. The marker-based system used a 10-camera VICON V16 system capturing data at 240 Hz. An LED wand facilitated synchronisation between these two systems. Multi-view camera calibration was performed using the OpenCV Calibration module and a CharuCo calibration target.

Markerless Motion Capture System

The markerless system evaluated multi-cam, multi-person kinematics estimation using OpenPose [1], YOLOv4 [2], EasyMocap [3] and Skinned Multi-Person Linear (SMPL) model [4]. Monocular kinematics estimation employed SMPL model to estimate 3D postures from 2D joint coordinates and focal length estimation.

Data Analysis

To evaluate the accuracy of the markerless system, positional data from both marker-based and markerless systems were compared using several metrics. For multi-view reconstruction, Mean Per Joint Position Error (MPJPE) and Mean Absolute Error (MAE) were evaluated, while Mean Per Joint Velocity Error (MPJVE) and full body relative velocity error were evaluated for monocular reconstruction. In this evaluation, results obtained from the marker-based system are considered as ground truth.

III. INITIAL FINDINGS

Multi-view Kinematics Estimation Results

Figure 1 depicts examples of the multi-view, multi-person, 3D pose estimation results. Table I reports the detailed comparison of the two Mocap systems, with an MPJPE of around 55 mm from ground truth data. However, the MAE of the pelvis joint is larger than 50 mm, observed in all trials due to the difference in the pelvis marker position between the VICON and markerless systems. The system's performance remained robust and adaptable across different recording angles and velocities.

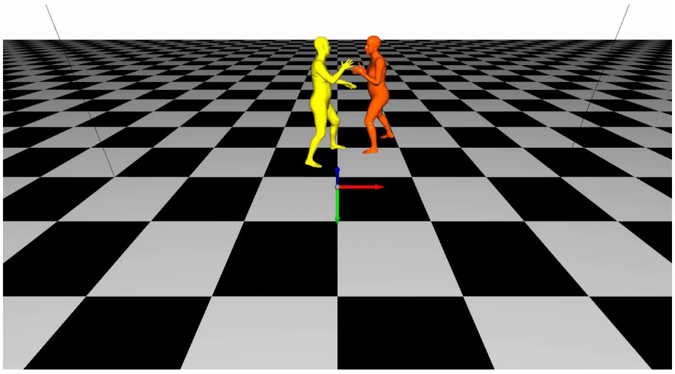
TABLE I

THE POSITIONAL ERROR EVALUATION FROM MULTI-VIEW RECONSTRUCTION; THE UNIT IS MILLIMETERS

	Neck	RShoulder	Relbow	RWrist	LShoulder	LElbow	LWrist	MidHip	Rhip	Rknee	Rankle	Lhip	Lknee	Lankle	LHeel	RHeel	MPJPE
Subject 1	28.7	53.6	56.5	47.7	46.7	45.5	50.0	97.8	110.9	78.4	64.8	105.5	59.6	49.1	64.1	51.6	55.7
Subject 2	25.2	38.7	48.9	32.8	42.1	51.5	41.5	83.8	101.3	60.1	45.3	105.8	67.3	48.6	52.4	50.5	55.3

Q. Yuan (e-mail: qyuan@kth.se; tel: +46 0760632468) is a doctoral student, S. Kleiven is a professor and X. Li is an associate professor, all at Neuronic Engineering, KTH Royal Institute of Technology, Stockholm, Sweden.

(a) Pose detection and tracking from each camera.



(b) Triangulation and SMPL fitting.

Fig. 1. Multi-view reconstruction results from markerless MoCap system.

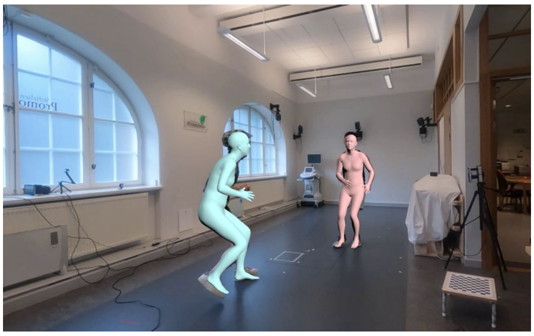
Monocular Kinematics Estimation Results

The results of the representative monocular kinematics estimation, as depicted in Table II and Fig. 2, exhibit an MPJVE lower than 1 m/s and full body relative velocity error around 10%. The MPJVE, which represents the average velocity error across adjacent frames, is derived from the first derivative of the MPJPE. A low MPJVE value is indicative of good temporal consistency in the tracked motion. The relative velocity error indicates the overall motion estimation is accurate and reliable.

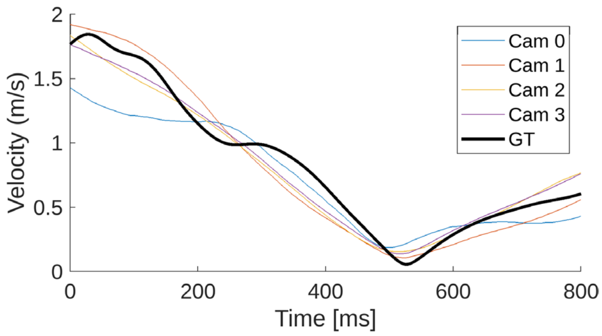
TABLE II

THE VELOCITY ERROR EVALUATION FROM MONOCULAR RECONSTRUCTION

MPJVE (m/s)	0.9	1.0	0.9	0.7	1.5	0.7	0.6	0.8	0.4	0.5	1.1	0.5	0.6	0.8	0.8
Error (%)	2.0	6.3	4.5	11.5	5.2	6.8	8.3	3.7	6.6	12.2	10.8	15.3	4.6	5.3	3.2



(a) 3D body model tracking from one camera.



(b) Representative kinematic time-histories for monocular estimation (colors) and ground truth (black).

Fig. 2. Monocular reconstruction results from markerless MoCap system.

IV. DISCUSSION

This study presents a markerless motion-capture system that accurately tracks complex human movements in sports scenarios, showing promising results in multi-view kinematics estimation. A key contribution is the implementation of a monocular 3D trajectory estimation approach, beneficial for situations where multi-camera configurations are not feasible or cost-effective. The system's overall accuracy is within acceptable ranges for biomechanical analysis in sports. Future work should focus on refining algorithms and models to improve performance across various recording angles and velocities. Testing the system's effectiveness in other sports movements and scenarios and integrating additional data sources to enhance depth information accuracy are other possible directions of future work.

V. REFERENCES

[1] Cao, Z., *et al*., CVPR, 2017.
[2] Bochkovskiy, *et al*., arXiv, 2020.
[3] EasyMocap, Github, 2021.
[4] Loper, *et al*., ACM, 2015.