Developing a Dynamic Human Model for Monocular Video-based Motion Capture by a Formulation of Generalized Coordinates for Lagrange's Equations

Meiyin Liu[1], SangHyun Lee[2], Veeru Talreja[3], Julia Penfield[3], Rick Barker[3]
[1]Department of Civil and Environmental Engineering, Rutgers University, New Brunswick, USA
[2]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, USA
[3]VelocityEHS, 222 Merchandise Mart Plaza, Chicago, USA
Email: meiyin.liu@rutgers.edu

Summary

The rapid development of vision-based human motion capture technologies implies potential in end-to-end automated biomechanical analysis from a monocular video. However, certain characteristics (e.g., loss in degree of freedom) of vision-based motion data require a specifically designed human model, especially for otherwise prohibitive applications such as biomechanical postural control. This research addresses this issue by proposing a dynamic human modeling approach with a formulation of generalized coordinates to use Lagrange's equations. Preliminary tests demonstrated a promising result in linear acceleration estimation, with a less than 1 m/s^2 discrepancy from marker-based mocap system (i.e., Vicon).

Introduction

Biomechanical modeling and postural control rely on motion data capture. Vision-based whole-body motion capture provides a potentially transformative way of streamlining the comprehensive process in a conveniently automated manner. Some works only focus on calculating kinematic data (e.g., body velocity and accelerations), limiting their applications to biomechanical modeling, which uses inverse dynamics without risks of indeterministic solutions of forces [1, 2]. Some works directly predict external forces and/or joint torques from motion data in Cartesian coordinates using AI models without explicitly estimating the intermediate kinematic data, limiting their applications from postural control with Lagrange's equations that produce deterministic solutions [3].

Methods

Given a monocular video, a 3D human pose estimation model is applied to capture frame-wise skeletons. All joint locations are represented by Cartesian coordinates. They are converted to the proposed dynamic human model via representing each body part by rigid bodies instead of vector-form links and via converting joint location representation from Cartesian coordinates to generalized coordinates with the proposed formulation (Figure 1). Specifically, the generalized coordinates are the three Euler angles that construct the rotation matrix between two connecting body parts' local Cartesian coordinate systems. With every body part represented by three Euler angles, Lagrange's equations are expressed with the generalized coordinates. When external forces are applied (e.g., hand forces, ground reaction forces), a similar process is applicable.

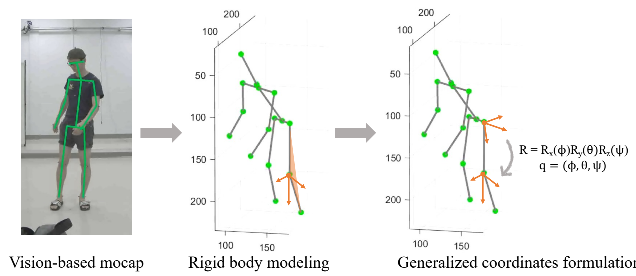

Figure 1: Major steps of dynamic human model development.

Results and Discussion

To test the feasibility of the proposed approach, we compare the estimated acceleration (linear and angular) with that from a marker-based motion capture system (i.e., Vicon). The preliminary result shows a mean absolute difference of 0.74 m/s^2 (0.96 m/s^2 for the task in Figure 2) between the estimated and ground truth linear accelerations for 5 tasks, demonstrating that the modeling does not distort the values of kinematic data after conversion to generalized coordinates and recovery back to Cartesian coordinates.

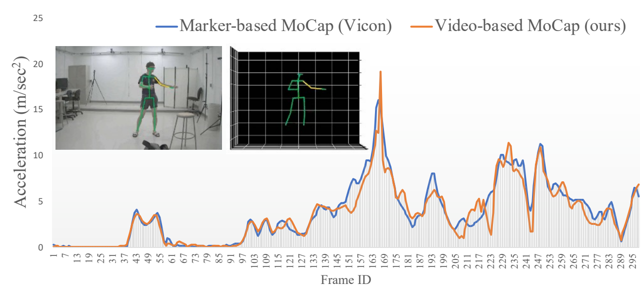

Figure 2: Example comparison result: estimated linear acceleration for the right lower arm.

Conclusions

The proposed formulation of generalized coordinates for Lagrange's equations can properly estimate human kinematic data compared to a marker-based motion capture system (i.e., Vicon). It suggests an impactful potential of streamlining dynamic biomechanical analysis for a video, allowing not only biomechanical modeling but also postural controlling.

References

[1] Messmore et al. (2023). U.S. Patent 11,783,495.
[2] Kösesoy and Öz. (2021). *European Journal of Science and Technology*, **32**: 92-100.
[3] Louis et al. (2022). *International Conference on Multimedia*, 3540-3548.