Inverse Kinematic, Dynamic, and Muscle Simulation from Video Files: Toward Markerless 3D Human Pose, Torque, and Muscle Estimation

Ali Nasr and John McPhee

May 17, 2024

The 7th International Conference on Multibody System Dynamics
June 9-13, 2024, Madison, WI, USA

Inverse Kinematic, Dynamic, and Muscle Simulation from Video Files: Toward Markerless 3D Human Pose, Torque, and Muscle Estimation

Ali Nasr, John McPhee

Systems Design Engineering, University of Waterloo, Canada
[a.nasr, mcphee]@uwaterloo.ca

Abstract

Optical skeletal motion capture technology is extensively employed across various domains, including entertainment, sports biomechanical analysis, healthcare, gaming, augmented reality, and human-computer interaction [1]. Conventional commercial systems mandate the usage of marker suits resulting in usability limitations. In reaction, scholars have devised markerless motion capture techniques enabling motion estimation in a broader range of settings through multi-view video analysis. Current leading pose estimation methods are heavily relying on deep learning techniques [3]. These machine learning methods depend on pre-processed databases, which are typically generated through visual detection and manual annotation processes. The manual processes introduce inaccuracies, violation of biomechanical limits, variable body segment lengths, or leading to non-existent degrees of freedom (DoFs). This study presents a method for correcting the manual annotations by integrating a human musculoskeletal (MSK) model with International Society of Biomechanics (ISB) standard.

Model

A scalable MSK model was employed to represent the human multibody system dynamics, which was constructed with MapleSim (Maplesoft Inc, Canada) and verified by experimental dynamometer data [4]. This model comprises two sub-models of skeletal dynamics and a biomechanical joint torque, which are used for dynamic simulations of movement. The full body model consists of 15 body segments with 40 DoFs. The skeletal system sub-model is created in accordance with the ISB standard and anthropometric data, taking into account anatomical landmarks location, frame definition, conventional DoFs, body segment geometric, body segment inertial parameters, and centers of mass (COMs).

The biomechanical joint torque is modeled using muscle torque generators (MTGs), each of which has a unique set of characteristics, including an active component that generates torque limited by the joint's angle and angular velocity, and a passive component that mimics the joint's viscosity or friction and constrains its range of motion (ROM). The MTGs comprise five sub-functions that are scalable with sex, age, body mass, height, dominant side, physical activity, and skin temperature, estimating these functions at the level of one joint. The five sub-functions are named as (1) passive torque, (2) peak joint isometric strength, (3) active-torque-angle scaling, (4) active-torque-angular-velocity scaling, and (5) excitation-to-activation signal functions. These functions used the mean maximum ROM, mean maximum joint velocity, peak isometric torque, and peak isokinetic torque separately for females and males from literature data.

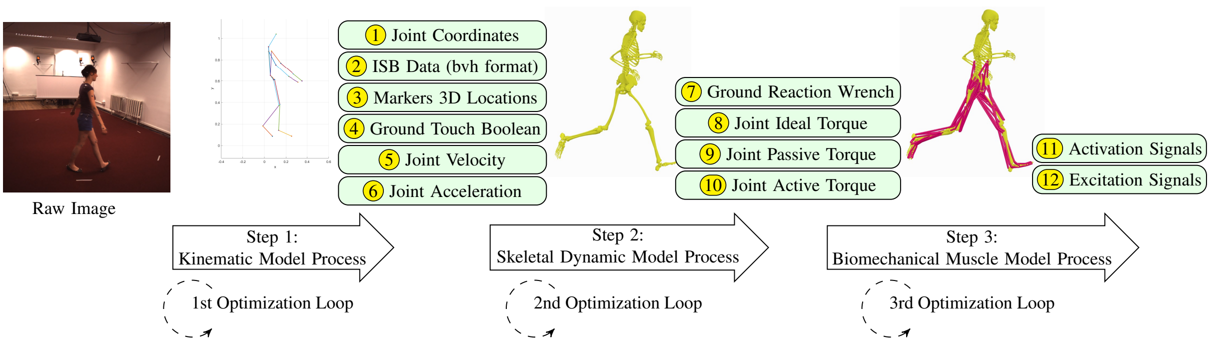

Figure 1: The schematic of a 3-step data generation pipeline consists of kinematic, skeletal dynamic, and MTG model processes using the multibody MSK model.

Simulation

The simulation input and output are illustrated in Figure 1. The input consists of videos with raw or manual marker annotations, and the output comprises 12 synchronized data points for each image frame.

The input data is processed by the aforementioned multibody system model, which performs an inverse simulation with three optimization loops. The first loop adjusts the ISB standard joint coordinates to align the anatomical landmarks with the input raw markers, while satisfying the biomechanical ROM constraints. The second loop computes the ground reaction wrench that minimizes the wrench related to the pelvic coordinates, within the feasible range of ground reaction wrenches. The third loop derives the excitation signals that produce the activation signals, with the constraint that both signals are between 0 and 1.

The simulation output includes: 1) standard joint coordinates, 2) BVH (Biovision Hierarchy) file for animation, 3) 3D marker locations, 4) ground contact detection, 5) joint velocities, 6) joint accelerations, 7) ground reaction wrenches, 8) joint ideal torques, 9) joint passive torques, 10) joint active torques, 11) activation signals, and 12) excitation signals. These data are generated for default average female and male subjects, as well as in a dimensionless format. For the dimensionless format, the positions are normalized by the subject's height, the forces are normalized by the subject's weight, and the torques are normalized by the product of the subject's weight and height.

Results and Discussion

The ISB marker locations obtained from the simulation output have a mean-squared error (normalized to body height) of 7% compared to the raw marker locations. This error can be attributed to the misalignment of the raw markers with the actual joint centers, which leads to variable human segment lengths, or some DoF that exceed the possible ROM.

The aim of this paper was to prepare 3D human joint positions, angle, torque, MTG activation, and excitation synchronized with single-camera images, a task known as monocular 3D human pose and dynamic estimation. Our physics-based pipeline estimators use a multibody MSK model [4] to refine the predicted motions and make them more consistent with physical principles [2]. Scientists can train a deep learning model that takes the raw images as input and our processed synchronized data as output, which can be applied to various domains such as motion analysis, sports biomechanics, healthcare, entertainment, augmented reality, and human-computer interaction [3].

Pure kinematic methods do not account for the physical laws that govern human motion, which may result in unrealistic poses such as floating, sliding, and leaning without support. Physics-based 3D human pose estimation methods, on the other hand, incorporate physics constraints to enforce contact consistency and avoid motion artifacts such as ground penetration, jitter, and imbalance. As future work, we propose to develop a simulation framework that can handle a wide range of motions and activities, incorporating more diverse constraints and possibly using central nervous system (CNS) or optimal control to generate the motions.

Acknowledgments

This research is funded by the Canada Research Chairs program and the Natural Sciences and Engineering Research Council of Canada.

References

[1] Febrer-Nafria et al.: Predictive multibody dynamic simulation of human neuromusculoskeletal systems: A review. Multibody System Dynamics: Springer, pp. 1–41, 2022.

[2] McPhee, J.; Nasr, A.: title=Multibody System Dynamics: A Fundamental Tool for Biomechatronic System Design, In P. ECCOMAS Thematic Conference on Multibody Dynamics, Lisbon, Portugal, 2023.

[3] McNally et al.: EvoPose2D: Pushing the boundaries of 2d human pose estimation using accelerated neuroevolution with weight transfer. IEEE Access, vol. 9, pp. 139403–139414, 2021.

[4] Nasr et al.: Scalable musculoskeletal model for dynamic simulations of upper body movement. Computer Methods in Biomechanics and Biomedical Engineering: Taylor & Francis, pp. 1–32, 2023.