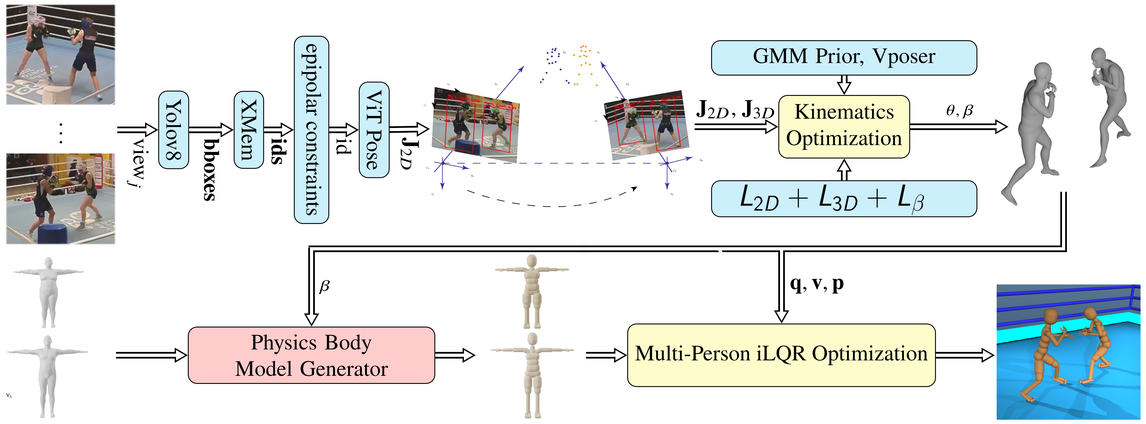
Multi-person Physics-based Pose Estimation for Combat Sports

Hossein Feiz[1,*] David Labbé[1] Sheldon Andrews[1]
[1]École de technologie supérieure, Montréal, Québec, Canada

Introduction

Motion capture from only RGB cameras in a typical sports scene (e.g., boxing), with the presence of coaches, viewers, and close interactions between athletes, brings many challenges, such as heavy occlusion, background crowding, and fast, complicated movements. To address this problematic, a multi-view configuration was used. Our system only required the visibility of an athlete from two cameras. We also created a pipeline using machine and deep learning techniques that automated the process of creating animations.

Pipeline



Physics Body Model Generator

This is the future...

Theorem

This is a theorem

Weighted Triangulation and Filtering

The weighted triangulation estimates the 3D positions of the keypoints from their corresponding 2D keypoints from all N cameras. The 3D positions of each joint are thus determined by solving the linear system:

Weighted Triangulation

$$\begin{bmatrix} \mu_1(\mathbf{P}_{11} - u_1\mathbf{P}_{31}) \\ \mu_1(\mathbf{P}_{21} - v_1\mathbf{P}_{31}) \\ \vdots \\ \mu_N(\mathbf{P}_{1N} - u_N\mathbf{P}_{3N}) \\ \mu_N(\mathbf{P}_{2N} - v_N\mathbf{P}_{3N}) \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (1)$$

$\mathbf{P}_{ij}$ extracts the i-th row of the projection matrix $\mathbf{P}_j$. μ_j is the average confidence of a 2D point for the current frame in camera j. Higher confidence values give more weight to certain cameras. We use SVD to solve Eq. 1.

After triangulation, outliers are handled by interpolating and smoothing using cubic spline for each joint's trajectory. If there is no solution for triangulation, we utilize an extended Kalman filter for dynamic state estimation based on velocity, acceleration of the keypoint, and position constraint to estimate the 3D keypoints. This results in robust filtering and smoothing while preserving trajectory integrity, enabling dependable 3D reconstructions from sparse and noisy 2D keypoints.

Multi-frame Multi-view Tracking IDs

- 1st item on the right
- 2nd item on the right
- 3rd item on the right

An Inner Block

Say something important here

$$E = mc^2 \quad (2)$$

$$a \in \mathbb{R}^n \quad (3)$$