Temporally Consistent 3D Human Body Animation from RGB Video

Chair of Computer Graphics and Visualization
TUM School of Computation, Information and Technology
Technical University of Munich

Abstract

Modern applications in graphics, vision, AR/VR, and human-computer interaction require realistic 3D human body motion, but in practice, input data is often limited to monocular RGB video. Common approaches rely on 2D joint detectors and parametric body models, such as the Skinned Multi-Person Linear (SMPL) model, by fitting shape and pose parameters to each frame independently so that projected joints align with 2D detections. However, this per-frame method typically ignores temporal context, resulting in jittery and unnatural motion artifacts. We propose a pipeline that reconstructs full 3D human motion from RGB video by extending standard SMPL fitting with explicit temporal constraints. Our goal is to reduce jitter and obtain smooth and realistic 3D body animations. The proposed method will be systematically compared with a per-frame baseline in terms of both quantitative metrics and visual quality.

1. Technical Approach

We propose a pipeline designed to generate temporally consistent 3D body animation from monocular RGB video. Our approach builds upon the standard SMPLify method [2], extending it with explicit temporal constraints to reduce jitter and optimize processing performance for video sequences.

1.1. Baseline

We implement a baseline system inspired by the original SMPLify framework [2], as illustrated in Figure 1. The pipeline operates on a per-frame basis and consists of the following sequential steps:

1. Read video frame using OpenCV.
2. Detect 2D keypoints with OpenPose [3].
3. Load SMPL [4] model to map low-dimensional shape and pose parameters into a full 3D human mesh. We plan to adapt the forward kinematics of SMPL model for use within the Ceres Solver.
4. Optimize SMPL parameters [4] for the current frame. The problem is formulated as a non-linear least-squares optimization using the Ceres Solver [1]. For a given frame, the SMPL parameters are optimized to minimize an objective function which includes a reprojection term (alignment with 2D keypoints) and prior terms that enforce anatomically plausible estimates [2].
5. The 3D mesh is generated from estimated parameters and rendered over the current frame for visual inspection using GLFW / OpenGL.

1.2. Proposed Enhancements

To address the baseline's lack of temporal consistency and to optimize performance for video sequences, we will investigate the following complementary strategies:

- **Modify objective function:** We will augment the optimizer objective function by adding a temporal cost term (e.g., velocity or acceleration penalties) to ensure the model transitions smoothly between frames.
- **Warm initialization:** Rather than initializing the solver from the mean pose every frame, we can initialize it using the optimized pose from the previous frame. This provides a better starting point and is expected to lead to significantly faster convergence.
- **Parameter freezing:** Since body shape remains constant throughout a video, we can freeze the SMPL shape parameters after optimizing them for the initial frames. This makes the optimization problem simpler and enforces consistency across the sequence.

1.3. Evaluation

We will evaluate our approach against the baseline using a combination of error metrics and visual inspection. Our comparison will focus on accuracy (Mean Per-Joint Position Error and 2D Reprojection Error), temporal stability (Acceleration Error), and runtime performance. For the visual inspection, we will conduct a side-by-side comparison of the proposed method against the baseline.

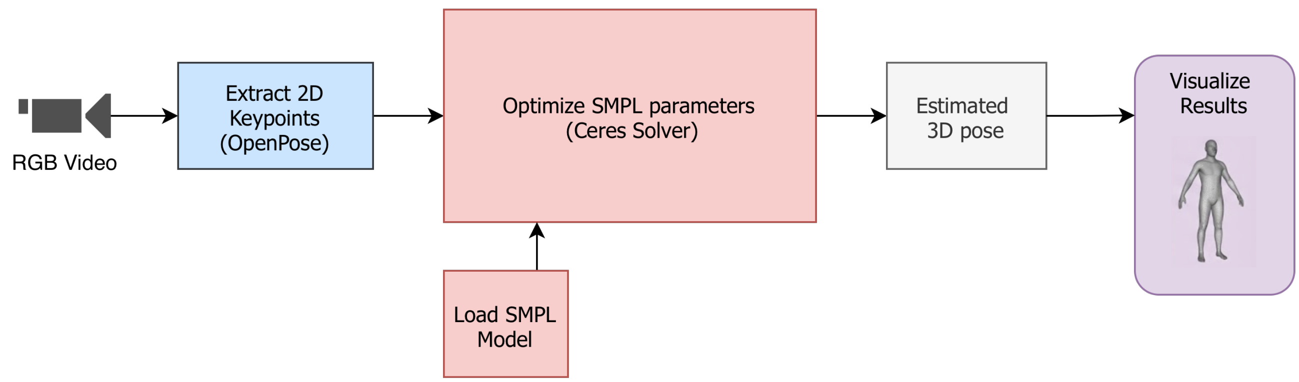

Figure 1. Pipeline for baseline.

2. Requirements

2.1. Datasets

We plan to use monocular RGB video sequences with visible human motion, focusing on widely used datasets for comparison with existing methods:

- **Human3.6M:** Controlled indoor recordings with consistent lighting and camera setups, suitable for validating the pipeline. `http://vision.imar.ro/human3.6m/description.php`
- **MPI-INF-3DHP:** Diverse motions, viewpoints, and backgrounds, to assess generalization to realistic scenarios. `https://vcai.mpi-inf.mpg.de/3dhp-dataset/`

2.2. Software Libraries & Frameworks

The proposed project relies on following software libraries and frameworks:

- **OpenCV:** For extracting and processing video frames. `https://opencv.org/`
- **OpenPose:** For robust 2D human joint detection from RGB images. `https://github.com/CMU-Perceptual-Computing-Lab/openpose`
- **SMPL:** Widely used in research for generating realistic 3D human body meshes from low-dimensional pose and shape parameters. `https://smpl.is.tue.mpg.de/`
- **Eigen:** For efficient matrix and vector operations. `https://eigen.tuxfamily.org/`
- **Ceres Solver:** Used for non-linear least squares optimization. `http://ceres-solver.org/`
- **GLFW / OpenGL:** For visualization of the estimated 3D model. `https://www.glfw.org/` `https://www.opengl.org/`

3. Milestones

Table 1. Project Schedule

Activity	15.12	22.12	12.01	19.01	26.01	02.02	09.02	17.02
A	•							
B	•	•						
C		•	•					
D			•					
E				•	•			
F					•			
G						•	•	
H							•	•

A: Configure repository, CMake and Docker.
B: Implement 2D keypoint detection with OpenPose.
C: Implement SMPL parameters optimization with Ceres.
D: Evaluate baseline on test sequences.
E: Implement proposed enhancements.
F: Evaluate proposed method on test sequences.
G: Write the final report.
H: Prepare the final presentation.

4. Team Members

Alexandru Ulesan	(03811739)
Ece Karasu	(03797980)
Lucas Pilla Pimentel	(03817782)
Ozden Gursoy	(03778621)

References

[1] Sameer Agarwal, Keir Mierle, and The Ceres Solver Team. Ceres solver. `https://github.com/ceres-solver/ceres-solver`, 2023. 1

[2] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J. Black. Keep it SMPL:

Automatic estimation of 3D human pose and shape from a single image. In *Computer Vision – ECCV 2016*. Springer International Publishing, 2016. 1

[3] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1

[4] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, 2015. 1