open-body-fit: Open-source pipeline for understanding sign language gestures from video

Aline Normoyle
Bryn Mawr College

Edward Zhang
University of Pennsylvania

Norman I. Badler
University of Pennsylvania

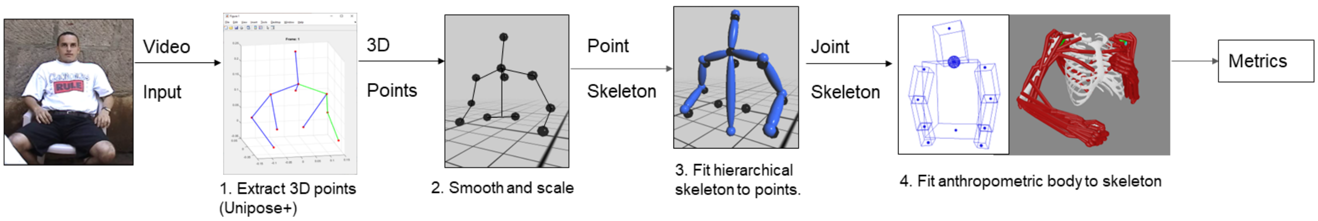

Figure 1: Pipeline implemented by open-body-fit for computing a 3D skeleton from a video. The system takes 3D positions from video and a skeleton specification as input. From the 3D positions, open-body-fit computes a 3D hierarchical skeletal motion using inverse kinematics. From the skeleton, we can estimate a biomechanic skeleton using an anthropometrics model of the body and a physics simulator. We demonstrate the system using Dart [Lee et al. 2018] and OpenSim [Delp et al. 2007]. The system outputs metrics, such as kinetic energy or symmetry.

ABSTRACT

We describe our open-source system, open-body-fit, for extracting articulatory effort metrics for sign language from standard video. The goal of this system is to provide insights into how factors related to physical movements, such as principles of least effort and ease of articulation, might affect how sign languages evolve or differ across individuals. Our pipeline extracts poses from standard video using a computer vision model and retargets these poses onto a 3D physically-based skeleton. Our system computes metrics for each frame of motion, such as kinematic energy and symmetry, and includes scripts for analyzing them in R.

KEYWORDS

motion capture, sign language, kinematics, physics-based character animation, muscle simulation

ACM Reference Format:
Aline Normoyle, Edward Zhang, and Norman I. Badler. 2022. open-body-fit: Open-source pipeline for understanding sign language gestures from video. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

We describe our open-source pipeline for extracting biomechanically-motivated kinematic metrics from standard video. Although extracting such metrics from videos has applications outside of linguistics, our goal is to use these metrics to better understand sign languages. Particularly, we are interested in how factors related to physical movements, such as principles of least effort and ease of articulation, might affect how sign languages evolve time. For example, signers might also reduce the number of repetitions in a sign, or cut short the articulation of a sign, or replace translational symmetry with reflective symmetry [Napoli et al. 2014]. Such shortcuts are analogous to co-articulations, contractions, and acronyms used in spoken English. Understanding the connection between articulatory effort and sign language helps us understand the extent that signs change to become energy efficient as opposed to other factors, such as whether signs change to make them easier to see while focusing on the face.

Our approach is straightforward. We take 3D poses of the upper-body extracted from video using a deep-learning model, Unipose+ [Artacho and Savakis 2021], and then fit a physically-based, hierarchical skeleton to it (Figure 1). Unipose+ estimates 3D positions for the upper body. These positions are in a unit-less coordinate system that can vary between video clips. For example, the same subject may have different median-sized limbs dependent on their distance from the camera. Thus, rather than use these coordinates directly, we use them to drive the motion of a predetermined skeletal model, whose size and joint hierarchy are defined apriori. The use of a skeletal model also allows us to define axes of rotation, maintain constant limb lengths, or define rotation limits.

We smooth the points from Unipose+ using a Gaussian filter and then scale the points to fit our skeletal model. We solve for the skeletal poses each frame using a standard inverse kinematics approach which minimizes the distances between the skeleton's joints with the 3D points from Unipose+. To validate the fit, we extract 2D poses from the video and then solve for a matrix P that projects our body model back onto the original video. The matrix P corresponds to the model-view-projection matrix that best projects

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

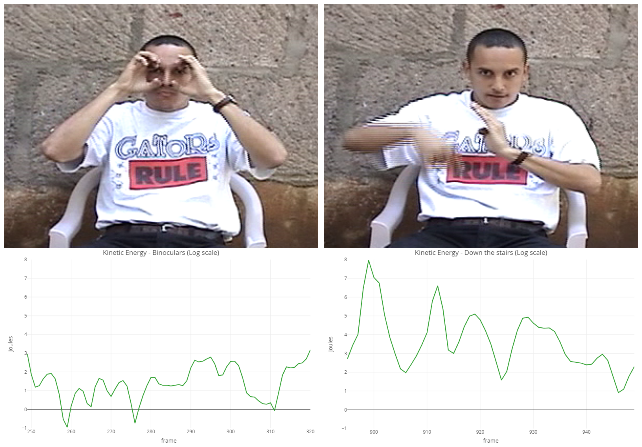

Figure 2: Kinematic energy. Left column shows results for the sign "look with binoculars". Right column shows results for the sign "go down the stairs". The bottom row shows time series curves for each sign (in Joules, shown with a log scale).

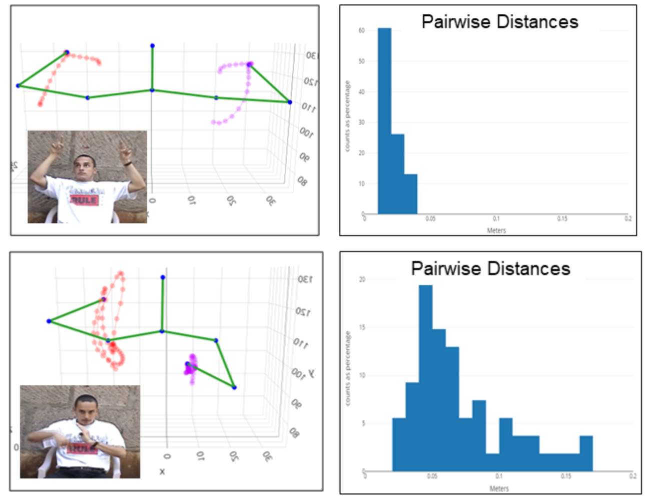

Figure 3: Symmetry for the signs "window" and "down the stairs". Window is a symmetry sign and has small pairwise distances compared to the Down the Stairs.

our skeleton to the image plane. We can then load the skeletal model into a physical simulator, such as Dart or OpenSim.

Dart [Lee et al. 2018] is an open-source physics simulator toolkit implemented in C++. In Dart, we define rigid bodies (as cuboids) for each limb. The masses of each rigid body are estimated using an anthropometrics model based on [Winter 2009] which outlines how weight and center of mass are typically distributed across the body, given a subject's height and weight. Our system estimates velocities and accelerations using 5-point differencing and directly integrates Dart to output forces and torques using inverse dynamics.

OpenSim [Delp et al. 2007] is an open-source physics simulator focused on body simulation and with built-in support for muscle simulation. In this work, we load use MoBL-ARMS Dynamic Upper Limb model [Saul et al. 2014] which defines muscles for the arms. To integrate with OpenSim, our system defines a marker set file that corresponds to our skeleton which OpenSim can load. The system also outputs a .trc (Track Row Column) file containing the motion of the joints. Using these, we can reproduce the motion from the video and use OpenSim's features to estimate forces, torques, and muscle activations.

2 METRICS

In this work, we analyze a dataset consisting of speakers of Nicaraguan sign language, recorded using standard video cameras. Here, we use two example metrics to demonstrate the data we obtain using open-body-fit.

2.1 Estimating Energy

We demonstrate our system by comparing the kinetic energy of two signs: "look with binoculars" and "go down the stairs". The kinetic energy is a function of the velocities and masses of each limb. The kinetic energy is high when large limbs are moving quickly and zero when none of the limbs are moving at all, as when the signer is in a resting position.

2.2 Symmetry

Our system can also be used to estimate other useful metrics for gesture analysis. Symmetry quantifies how similar the movements of the left and right wrists are. Symmetry can refer to both in tandem and mirrored movements. To compute it, we align the trajectories of the left and right wrists using the Kabsch algorithm and then compute the pairwise distances between them. To demonstrate the approach, we consider two very different signs: "window" and "go down the stairs" (Figure 3).

3 DISCUSSION

The ability to extract kinematic and dynamics metrics from a video has many potential benefits for both understanding human movement as well as for developing new models for recognition and procedural animation of movement. Our on-going work on this project is validating the data extracted using our computer vision model as well as the metrics we estimate using physics.

REFERENCES

Bruno Artacho and Andreas Savakis. 2021. UniPose+: A unified framework for 2D and 3D human pose estimation in images and videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

Scott L Delp, Frank C Anderson, Allison S Arnold, Peter Loan, Ayman Habib, Chand T John, Eran Guendelman, and Darryl G Thelen. 2007. OpenSim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering* 54, 11 (2007), 1940–1950.

Jeongseok Lee, Michael X Grey, Sehoon Ha, Tobias Kunz, Sumit Jain, Yuting Ye, Siddhartha S Srinivasa, Mike Stilman, and C Karen Liu. 2018. Dart: Dynamic animation and robotics toolkit. *Journal of Open Source Software* 3, 22 (2018), 500.

Donna Jo Napoli, Nathan Sanders, and Rebecca Wright. 2014. On the linguistic effects of articulatory ease, with a focus on sign languages. *Language* (2014), 424–456.

Katherine Saul, Xiao Hu, Craig Goehler, Meghan Vidt, Melissa Daly, Anca Velisar, and Wendy Murray. 2014. Benchmarking of dynamic simulation predictions in two software platforms using an upper limb musculoskeletal model. *Computer methods in biomechanics and biomedical engineering* 18 (07 2014), 1–14. https://doi.org/10.1080/10255842.2014.916698

David A. Winter. 2009. *Biomechanics and Motor Control of Human Movement, Fourth Edition*. John Wiley & Sons, Inc.