Quantifying sign-language movement kinematics from video

Aline Normoyle
Bryn Mawr College

Samuel J. Rothstein
Swarthmore College

Norman I. Badler
University of Pennsylvania

ABSTRACT

We describe our system for extracting kinematic metrics for sign-language gestures from a video. The goals of these analyses are 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.

KEYWORDS

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

ACM Reference Format:

Aline Normoyle, Samuel J. Rothstein, and Norman I. Badler. 2021. Quantifying sign-language movement kinematics from video. In *Proceedings of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (i3D '21).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

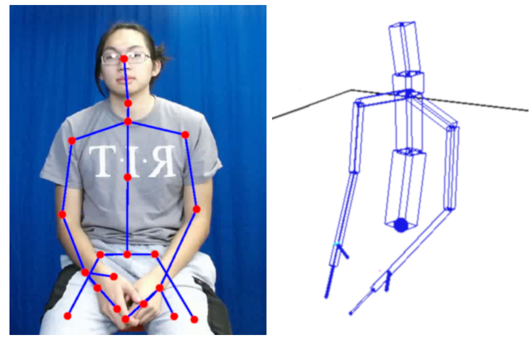

Figure 1: From image-based poses to physically-based skeleton. Left, a sign language speaker with pose extracted from RGB-D coordinates. Right, the same speaker represented with the physically-based skeleton (shown in a dart physics simulator [Lee et al. 2018]).

1 INTRODUCTION

We describe our system for extracting biomechanically-motivated kinematic metrics for sign-language gestures from a 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, differ across signers, or perhaps even vary for an individual.

Although 3D motion-capture-based data sets of sign-language exist [Lu and Huenerfauth 2014], it is much more common to have data sets based on videos [Imashev et al. 2020; Neidle et al. 2012; Özdemir et al. 2020].

Regardless of the source, fluent sign-languages speakers are needed to transcribe meanings and segment gestures into individual signs. However, new computer vision techniques for detecting people and extracting poses have made it possible to compute quantitative measurements which would be prohibitively difficult to do before.

For example, there is numerous evidence suggesting that psychophysical factors influence sign-languages [Napoli et al. 2014]. Signers might use a one-handed version of a sign in casual conversations, and a two-handed version in formal settings [Zimmer 1989]. 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. The ability to extract kinematic metrics automatically from videos gives us new capabilities for understanding how the human body might affect sign languages.

Our approach is straightforward. We take poses of the upper-body extracted from video [Artacho and Savakis 2020] and then fit a physically-based, hierarchical skeleton to it (Figure 1). The examples here use a data set based on RGB-D images recorded with the Kinect [Hassan et al. 2020]. From the image-based poses (that contain depth), we estimate 3D positions in meters and then fit a hierarchical skeleton that enforces constant limb lengths and joint rotation constraints. To estimate joint masses, we use 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. Using our physically-based skeleton, we compute a wide variety of metrics, including speeds, volumetric regions of movements, symmetry measures, kinetic energy, and torques.

1.1 Articulatory metrics

To goal of our metrics is to quantify different aspects of motion. In many cases, metrics are consistent with direct observation. For example, in Figure 2, we show how several metrics vary as someone introduces himself with American Sign Language (ASL).

1.1.1 Point distance. . Point distance measures the difference (in meters) between the current pose and the rest pose shown in Figure 1. This metric is lowest when both hands rest in the lap, moderate when one hand is raised, and highest when both hands are raised.

1.1.2 Symmetry. . Symmetry (meters) is the sum of distances between the left arm and the reflected positions of the right arm. Symmetry is high when both hands mirror each other in height and distance to the torso.

i3D '21, April 20–22, 2021,

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

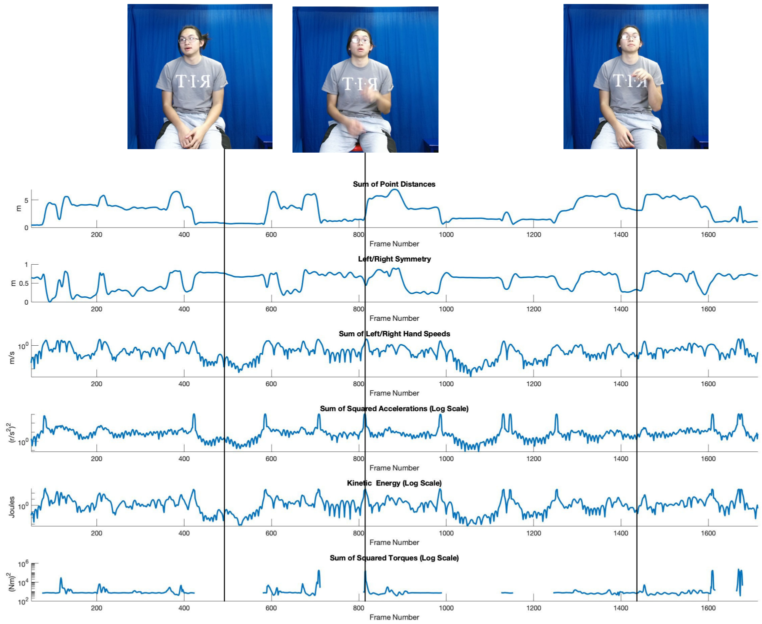

Figure 2: Time series showing all metrics. The x-axis shows the frame number. The y-axis corresponds to the units for each metric. Distance, hand speed, and symmetry are on a linear scale. Accelerations, kinetic energy, and torques are on a log scale. Most metrics have their lowest values when both hands are in the lap (frame 500). High values correspond to frames with large velocities (frames 815 and 1424).

1.1.3 Speed and acceleration. . Hand speed (m/s^2) is the sum of the speeds of both wrists. The sum of squared accelerations ($(radians/s^2)^2$) summarizes the angular accelerations at each joint. These metrics are highest when the arms move quickly.

1.1.4 Dynamics. . The kinetic energy (Joules) is the velocity of each joint scaled by its mass and inertia. The sum of squared torques ($(Nm)^2$) summarizes the forces at each joint. Higher kinetic energy implies greater movement, with larger joints having more influence than smaller ones. For torques, we omit frames where the hands rest on the subject's lap (*e.g.*, in contact with an object). In these poses, the hands are at rest and thus not applying forces to support the pose. Because accelerations, kinetic energy, and torques all reply on velocities, they all have similar curves. We compute angular velocities and accelerations using 5-point finite differences. We compute torques using a standard inverse dynamics algorithm [Featherstone 2014].

2 DISCUSSION

Several challenges need to be addressed to complete this system. The body model currently used to estimate mass and inertia does not take into account different body sizes or gender. Furthermore, the distribution of mass across the body requires the weight and height of the individual, which is difficult to discern from a video ([Hassan et al. 2020] includes height and weight information). Also, detailed motions of the hands cannot be reliably tracked and videos require a clear, unobstructed view of the speaker.

The ability to extract kinematic metrics from a video has many potential benefits for both understanding human movement as well as for character animation. For example, this approach could help us analyze the performances of different actors. Once computed, the physically-based skeleton can be used to edit the recorded motion or even generate new ones.

REFERENCES

Bruno Artacho and Andreas Savakis. 2020. UniPose: Unified human pose estimation in single images and videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7035–7044.

Roy Featherstone. 2014. *Rigid body dynamics algorithms.* Springer.

Saad Hassan, Larwan Berke, Elahe Vahdani, Longlong Jing, Yingli Tian, and Matt Huenerfauth. 2020. An Isolated-Signing RGBD Dataset of 100 American Sign Language Signs Produced by Fluent ASL Signers. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives.* 89–94.

Alfarabi Imashev, Medet Mukushev, Vadim Kimmelman, and Anara Sandygulova. 2020. A Dataset for Linguistic Understanding, Visual Evaluation, and Recognition of Sign Languages: The K-RSL. In *Proceedings of the 24th Conference on Computational Natural Language Learning*. 631–640.

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.

Pengfei Lu and Matt Huenerfauth. 2014. Collecting and evaluating the CUNY ASL corpus for research on American Sign Language animation. *Computer Speech & Language* 28, 3 (2014), 812–831.

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.

Carol Neidle, Ashwin Thangali, and Stan Sclaroff. 2012. Challenges in development of the american sign language lexicon video dataset (asllvd) corpus. In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC.* Citeseer.

Oğulcan Özdemir, Ahmet Alp Kındıroğlu, Necati Cihan Camgöz, and Lale Akarun. 2020. Bosphorussign22k sign language recognition dataset. *arXiv preprint arXiv:2004.01283* (2020).

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

June Zimmer. 1989. Toward a description of register variation in American Sign Language. In *The sociolinguistics of the Deaf community.* Elsevier, 253–272.