Carnegie Mellon University
Electrical & Computer ENGINEERING

Virtual Draping

Aditti Ramsisaria[1], Caroline Pang[2], Claire Lin[3]
[1] ECE & Robotics [2] ECE & HCI [3] ECE, Carnegie Mellon University, Pittsburgh, PA

Introduction

Virtually modelling how 3D garments drape on the human body has widespread applications in the domains of AR/VR content generation, e-commerce, virtual try-on, gaming, and more.

3D reconstruction of garments with accurate deformations (such as folds and wrinkles) on a custom, virtual body can help a person infer how a garment might look on their own body. There are several previous works which employ supervised techniques to learn how clothing deforms as a function of shape and pose, garment style, and sizing of garments. There are also self-supervised learning approaches which leverage optimization-based schemes to formulate a set of physics-based loss terms to train neural networks [3].

The goal of this project is to build an end-to-end garment draping pipeline which leverages self-supervised techniques for redressing garments in 3D on bodies reconstructed from 2D images, that adapts robustly to changes in pose, shape, garment style and material specifications.

Methods

We propose a pipeline that consists of three main steps in the following order:
(1) 2D image preprocessing with OpenPose [1] to get 2D joints for posing
(2) 2D to 3D reconstruction with SMPLify-X [2], and SMPL body parameter generation
(3) 3D garment draping derived from SNUG [3].

OpenPose [1]

- A multi-stage, feedforward CNN which extracts feature maps then uses a greedy bipartite matching algorithm to get poses of each person in an image
- Takes a color, 2D image as an input
- Outputs the 2D locations of the anatomical joints of every person in the input image
- **Our efforts: set up library, pass in self-collected body frames images as input**

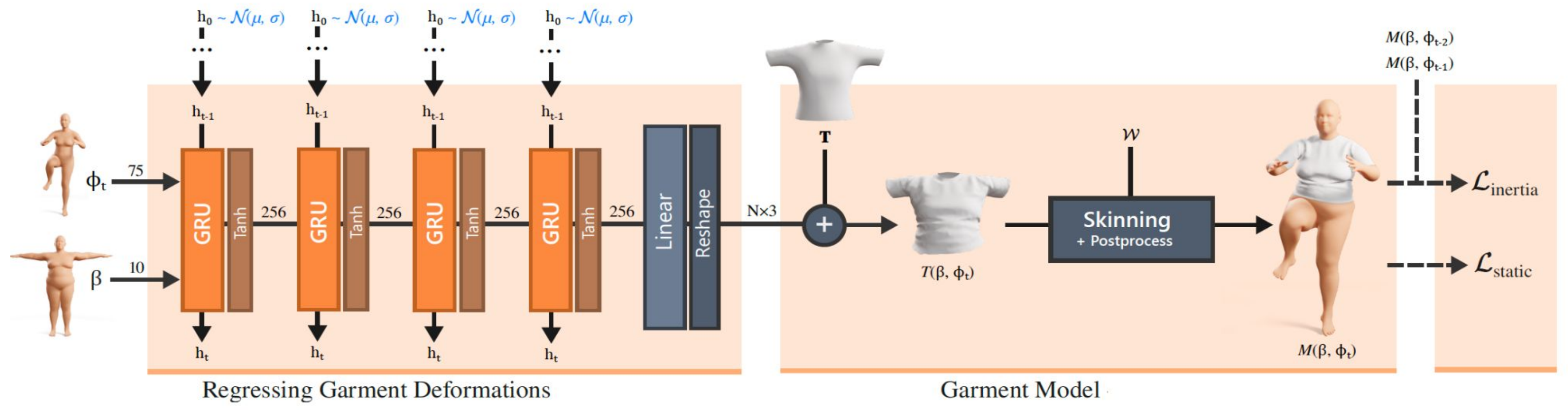
SMPLify-X [2]

- A CNN to extract 3D human body pose from a single 2D image
- Takes 2D joints and a 2D image as input
- Outputs reconstructed 3D SMPL body pose, shape, translation and camera orientation parameters
- **Our efforts: Experimented and selected a subset of SMPL pose parameters that is most ideal for SNUG input. Modified the output of SMPLify-X to be compatible with SNUG.**

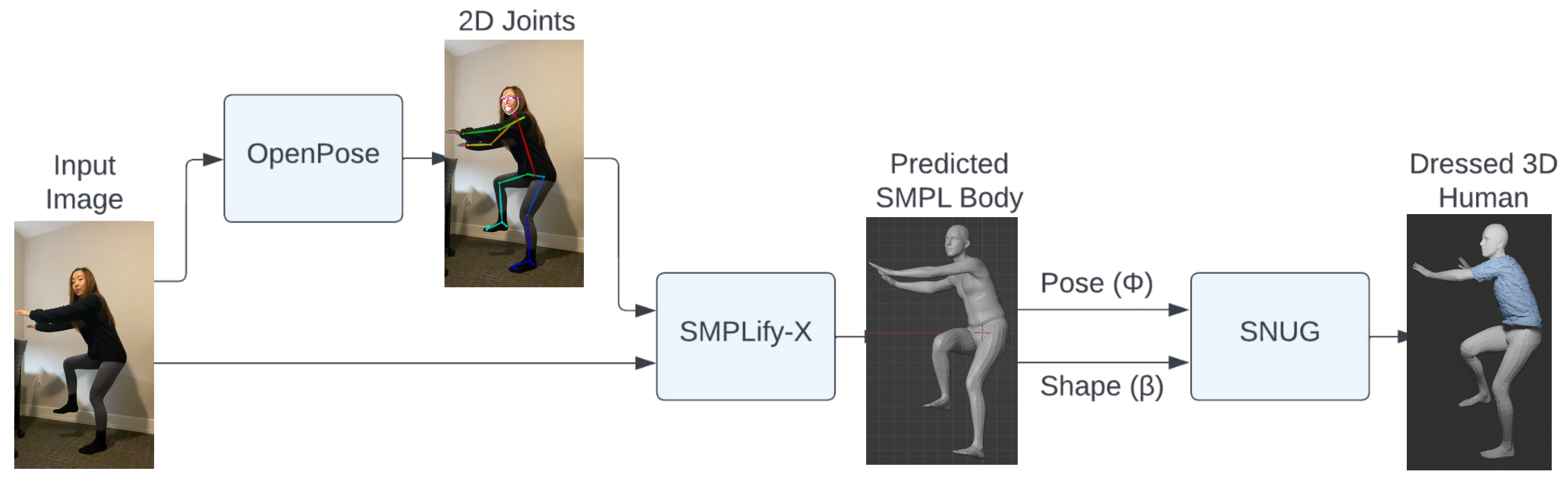
SNUG [3]

- A self-supervised model that computes physics-based deformations in the garment template based on material specifications and body/pose sequences.
- Takes in a subset of output parameters from SMPLify-X
- **Our efforts: Implemented the training module from scratch. Trained a new model with garment model.**
- Training Data: We randomly selected 4000 frames of various human body poses from the AMASS dataset

$$\mathcal{L}_{\text{static}} = \mathcal{L}_{\text{strain}} + \mathcal{L}_{\text{bending}} + \mathcal{L}_{\text{gravity}} + \mathcal{L}_{\text{collision}}$$
$$\mathcal{L} = \mathcal{L}_{\text{inertia}} + \mathcal{L}_{\text{static}}$$



SNUG Network Architecture

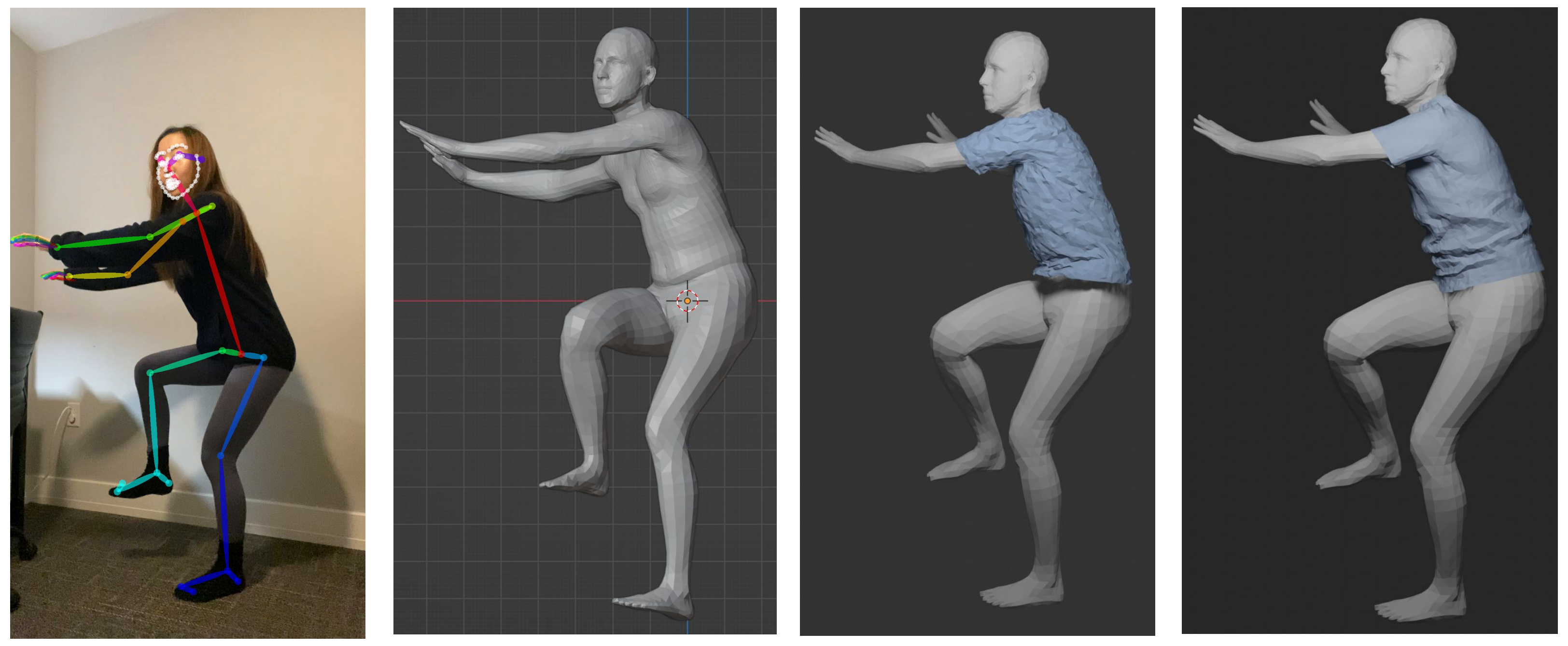


Overview of Pipeline

The garment model is built on the SMPL format where M() represents the human body as a function of pose(θ), shape(β), global translation(t). SMPL applies a series of linear displacements to base template T, and applies linear blend skinning W(). Bp(·) models pose-dependent deformations of a skeleton J, and Bs(·) models the shape dependent deformations. W represents the blend weights.

$$M(\beta, \theta) = W(T(\beta, \theta), J(\beta), \theta, W)$$
$$T(\beta, \theta) = T + Bs(\beta) + Bp(\theta)$$

Results and Discussion

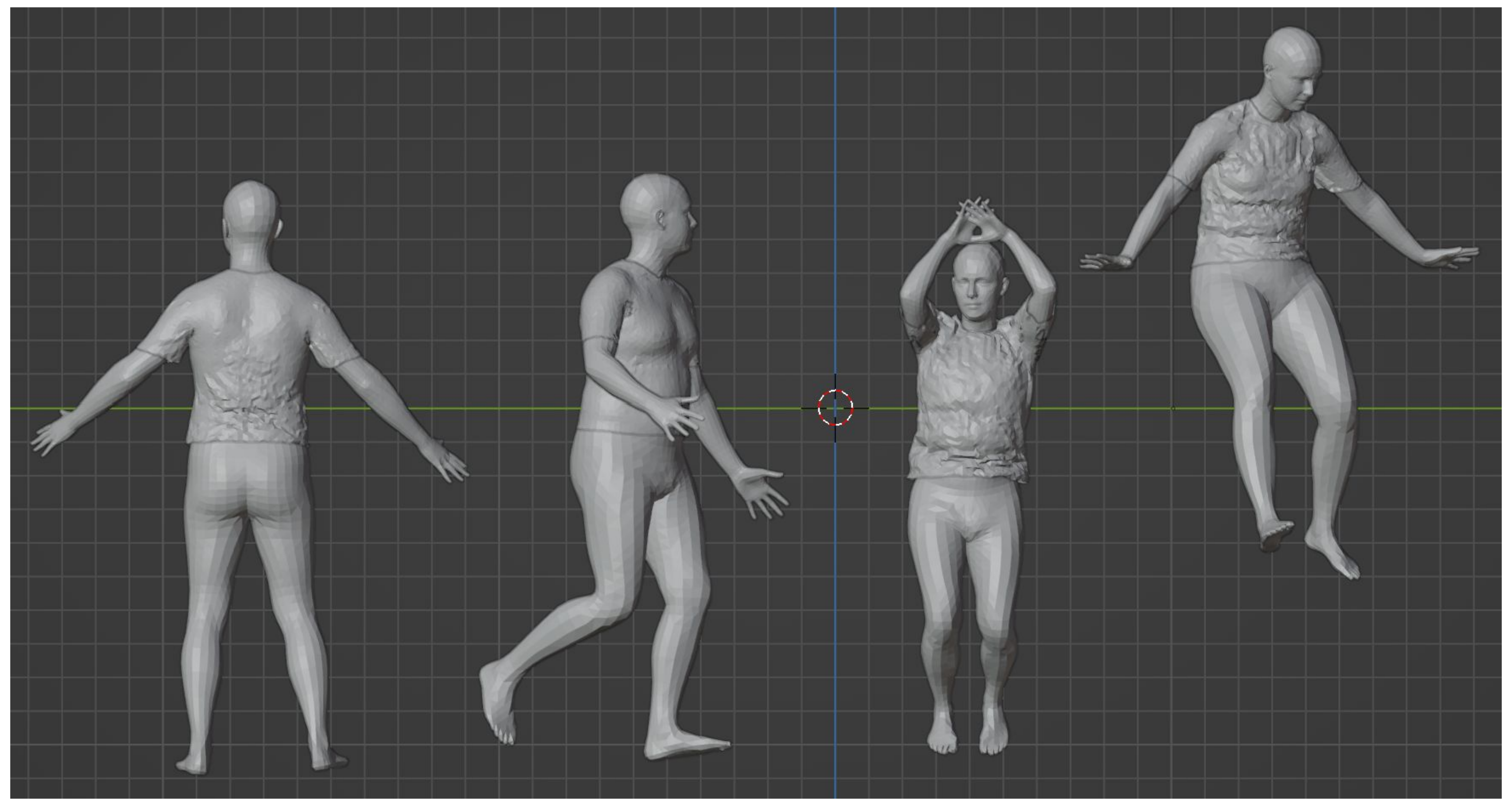


(a) OpenPose Joints visualized on Input Image
(b) SMPLify-X Body Reconstruction
(c) Our SNUG Garment Reconstruction (100% Cotton)
(d) Pre-trained SNUG Garment Reconstruction (100% Cotton)

In the first part of the pipeline, we were able to successfully use OpenPose to obtain accurate 2D pose estimations via landmarking from video frames. In the second part of the pipeline, we were also able to successfully utilize the SMPLify-X model to transform these 2D joints and images into 3D reconstructions in the SMPL format, through which we derive the body pose, shape, and translation parameters. We qualitatively observe that the 3D reconstructions accurately capture the pose and body shape of the figures from the 2D images.

In our implementation of the SNUG training module, we were able to successfully train the model to learn garment deformations and adapt to body poses and materials. We trained using a batch size of 8 for 5 epochs with a learning rate of 0.001 on a 16GB RAM system with a CPU. Our model also produced folds in realistic positions across a variety of dynamic poses.

However, we do think there is room for improvement in terms of adapting to different body shapes and creating realistic folds and wrinkles. We notice that our model tends to add many excess small deformations due to lack of convergence in the membrane strain energy loss. We hypothesize that this may be due to a lack of compute resources when training, since the original SNUG implementation used a batch size of 16 on a 32GB system, with close to 6900 frames for 10 epochs. Additionally, when testing with different body shapes, we noticed that there were interpenetrations between the garment and body vertices during inference. This is likely due to the model not generalizing to different body shape parameters in our implementation of linear blend skinning, causing posed vertices to overlap with each other.



Output of SNUG - 3D predictions of Garment Draping on Dynamic Poses

Conclusions

We were able to successfully combine three models, OpenPose, SMPLify-X, and SNUG and build a pipeline that takes in a single 2D image/video frame of human and produces the 3D reconstruction for the same body with a garment of choice draped on the body. A lot of effort was put into fixing deprecated versions of code, bridging the different models, and implementing the SNUG training module from scratch.

With the possibility to visualize clothing fit in 3D by simply inputting a single 2D image, our project can be adapted to e-commerce platforms and online clothing stores to help increase accuracy and confidence when shoppers select clothing sizes and styles.

Regarding future work, we would like to extend the predictions to different garments. Currently, our model is trained and tested on 100% cotton material. Additionally, we would like to incorporate lighting and color accuracy into the pipeline, forming a more holistic online clothing shopping experience for the customers. We also plan to use better post-processing techniques to account for collision penalties. Lastly, we would like to use the full functionality of the SIMPLify-X to predict well-articulated hands and facial expressions.

References

[1] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[2] Pavlakos, Georgios and Choutas, Vasileios and Ghorbani, Nima and Bolkart, Timo and Osman, Ahmed A. A. and Tzionas, Dimitrios and Black, Michael J. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019*

[3] Igor Santesteban, Miguel A. Otaduy, Dan Casas. SNUG: Self-Supervised Neural Dynamic Garments. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022 (Oral).

Acknowledgements

We would like to thank Professor Marios Savvides and teaching assistants Fangyi Chen, Han Zhang, and Sai Mitheran Jagadesh Kumar for their support throughout the project. We would also like to thank SNUG author Dan Casas for his guidance over the course of the semester.

Fall 2022
18-794 Pattern Recognition Theory