The Virtual Tailor: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style
– Supplementary –

Chaitanya Patel* Zhouyingcheng Liao* Gerard Pons-Moll
Max Planck Institute for Informatics, Saarland Informatics Campus, Germany
{cpatel, zliao, gpons}@mpi-inf.mpg.de

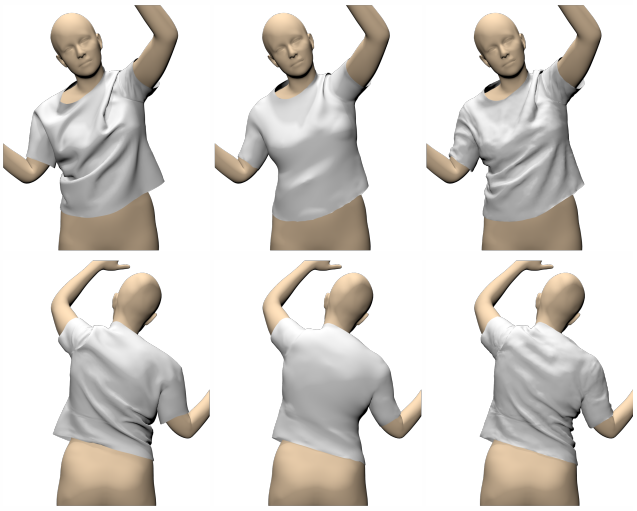

Figure 1. To illustrate that two methods can yield comparable quantitative error, but be uncomparable in realism and detail, we smooth out the output of our method (middle, error: 11.9mm) and compare it to our method (right, 10.9mm), and the ground truth (left). Clearly, by construction, our method (right) is significantly better than smoothed output(middle), despite being only 1mm better (10.9mm vs 11.9mm). Similarly, comparisons in Fig. 5 of the paper and video demonstrate our method is clearly superior to baselines, despite little quantitative improvements.

1. Improvement Over Baseline

TailorNet predictions are $\sim$1mm better than our baseline which may seem little. It is important to note that even our baseline is novel. It is trained from a dataset which covers the space of poses, shapes and styles, is far from trivial. A space of styles is learned on canonical pose, selected styles need to cover the space, and deformations need to be carefully un-posed.

Second, and more importantly, as demonstrated in Figure 1, low quantitative error is a necessary condition but not sufficient for a realistic result. As demonstrated in Section 6.2 in the paper and the video, TailorNet is only slightly better than our baseline quantitatively, but significantly better qualitatively. The fine details of TailorNet come from explicitly using mesh frequency decomposition. We note that finding the quantitative metric that reflects realism is an open research question.

2. Quasi-Static Effects versus Dynamic Effects

We want to add a short discussion on why we decided to focus on quasi-static effects versus dynamics. Quasi-static effects correspond to overall deformation and wrinkles that appear for a given pose, shape and style during a slow motion. We think a model of quasi-statics is very important as most of our daily activities involve slow motions. Consequently, images, videos and scans of people often exhibit these types of motions. Since we think the Virtual-Tailor could be useful for single image 3D reconstruction and analysis, we decided to learn a model of quasi-statics.
Analogously, the SMPL [3] model also represents quasi-static pose-deformation, and the dynamics (DMPL) are typically not used in image and video analysis. Hence, we decided to first focus on quasi-statics of clothing, similar to Garnet [2]. That being said, dynamic effects are also important; we are confident our method (decomposition into frequencies, mixture of shape-style prototypes) can be adapted to dynamics replacing MLPs by RNN type architectures– we leave this for future work. We note however, that our current model already produces compelling temporally coherent results for completely new (never seen during training) motion sequences.

3. Inter-penetration

While samples from the style subspace in a canonical pose do not have intersections, the pose dependent deformations produce intersections sometimes, similar to previous work [4, 1]. To resolve them, we push vertices out of the body surface as done in prior work [4, 1], which can be done in real time. This is however a limitation of our method that is shared with previous work [4, 1].

4. Network and Training Details

Our baseline and TailorNet use several MLPs - each of them has input layer, two hidden layers of 1024 neurons with ReLU activation, and output layer. The first hidden layer is followed by a dropout with $p = 0.2$. We set the learning rate $1e-4$, weight decay $1e-6$ and batch-size of 32. We arrive to these hyperparameters by tuning our baseline, and then keep them constant to train all other MLPs. The training converges after 160k iterations for baseline and $D^{LF}(\boldsymbol{\theta}, \phi)$, and 65k iterations for each $D^{HF}_{\phi,k}(\boldsymbol{\theta})$.

5. Further Simulation Details

Each style-shape pair (γ, β) is simulated in an independent session – style does not change during one simulation. Since the parametric model of style is learnt in canonical pose and shape (Section 4.2), a garment $G(\beta_0, \theta_0, \gamma)$ and a canonical body $M(\beta_0, \theta_0)$ are imported in the beginning of each session, after which the body slowly transitions to $M(\beta, \theta_0)$ and drives the garment to $G(\beta, \theta_0, \gamma)$. Then the body greedily traverse all poses while the style-shape is fixed as described in Section 5.3, which generates pose-variant garments $G(\beta, \theta_i, \gamma)$. It should be noted that the cloth material is kept fixed for all simulations. Modelling different materials is left for future work, see Conclusion.

References

[1] Peng Guan, Loretta Reiss, David A Hirshberg, Alexander Weiss, and Michael J Black. Drape: Dressing any person. *ACM Trans. Graph.*, 31(4):35–1, 2012. 1

[2] Erhan Gundogdu, Victor Constantin, Amrollah Seifoddini, Minh Dang, Mathieu Salzmann, and Pascal Fua. Garnet: A two-stream network for fast and accurate 3d cloth draping. *CoRR*, abs/1811.10983, 2018. 1

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

[4] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. Learning-based animation of clothing for virtual try-on. *Comput. Graph. Forum*, 38(2):355–366, 2019. 1