EUROGRAPHICS 2024 / A. Bermano and E. Kalogerakis
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 43 (2024), Number 2

Supplementary material: Neural Garment Dynamics via Manifold-Aware Transformers

Peizhuo Li[1], Tuanfeng Y. Wang[2], Timur Levent Kesdogan[1], Duygu Ceylan[2], Olga Sorkine-Hornung[1]

[1]ETH Zurich, Switzerland [2]Adobe Research, United Kingdom

1. Deformation Gradients

In this section, we review the preliminaries for the deformation gradients and Poisson equation.

1.1. Deformation gradients field

Given a discretized garment mesh, assumed to be a 2-manifold triangular mesh with vertices position $\mathbf{V}^{\text{rest}}$ at rest status and triangulation $\mathbf{T}$. For a triangle i associated with vertices v_j, v_k, v_l (counter-clockwise), its local coordinate Q_i as a 3×3 matrix is defined as $\left[v_k - v_j, v_l - v_j, \mathbf{n}_i\right]$ and $\mathbf{n}_i$ is the unit outward normal of triangle i. The deformation gradient of triangle i at frame t with vertices position $\mathbf{V}^t$ w.r.t. to the rest status can then be defined as $\Phi_i^t = Q_i^t Q_i^{-\text{rest}}$, where $Q_i^{-\text{rest}}$ refers to matrix inversion of Q_i^{rest}, the local coordinate of triangle j at rest status. We denote the deformation gradient field at frame t as $\Phi^t := \{\Phi_i^t\}_{i\in\mathbf{T}}$. We choose deformation gradients because they fully capture the local deformation of the garment and are invariant to the global translation and rotation.

1.2. Relative deformation gradient

Although the deformation gradients field provides a triangulation-agnostic representation, directly predicting the deformation gradient of frame i based on features of previous frames is not the most efficient way. Instead, the residual of the deformation gradient field between two frames, namely the relative deformation gradients $\Psi^t = \Phi^t\Phi^{-(t-1)}$, where $\Phi^{-(t-1)}$ is the element-wise matrix inverse of Φ^{t-1}, is more suitable for learning the dynamics of the garment deformation and frees the network from the effort of remembering the absolute deformation from previous frames.

1.3. Poisson equation

Given an arbitrary deformation gradient field Φ^t, we can reconstruct the deformed mesh by solving a Poisson equation:

$$\mathbf{V}^* = \arg\min_{\mathbf{V}} \sum_{i\in\mathbf{T}} s_i \left\|\Phi_i(\mathbf{V}) - \Phi_i^t\right\|_F^2, \quad (1)$$

where s_i is the area of triangle i and $\Phi_i(\mathbf{V})$ is the deformation gradient of triangle i given vertex positions $\mathbf{V}$. It is a sparse linear system w.r.t. to $\mathbf{V}$ and can be efficiently solved with the Laplace-Beltrami operator [SP04; SCL*04].

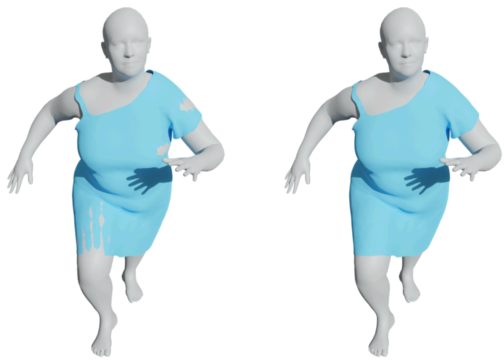


Figure 1: *Collision refinement. Our post-process refines the collision between garment and body in raw prediction by minimizing the proposed energy.*

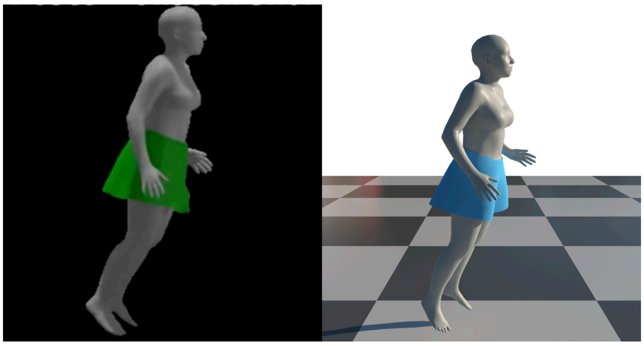
2. Collision refinement

We use a post-process adopted from DRAPE [GRH*12] for collision refinement. Assuming the predicted cloth's vertex position is $\tilde{\mathbf{V}}$ and the body's vertex position is $\mathbf{U}$, we solve for a new vertex position $\mathbf{V}$ to minimize the following energy functions:

$$E_{\text{collision}} = \sum_{(i,j)\in C} \|\varepsilon + \vec{n}_j \cdot (\mathbf{V}_i - \mathbf{U}_j)\|_2^2, \quad (2)$$

where C is the set containing the paired indices of the cloth vertex i in collision with the body and its nearest body vertex j, and $\vec{n}_j$ is the outward normal of the body vertex j. This energy pushes the vertices inside the body away from the body surface. Besides, we also want to keep the local geometry of the cloth unchanged. We thus also include the following Laplacian term:

$$E_{\text{lap}} = \|\Delta\mathbf{V} - \Delta\tilde{\mathbf{V}}\|_2^2. \quad (3)$$



GarSim [TB23] Ours

Figure 2: *Qualitative comparison to GarSim.*

To make this system determined, we add the following regularization term:

$$E_{\text{reg}} = \|\mathbf{V} - \tilde{\mathbf{V}}\|_2^2. \tag{4}$$

The overall energy function is thus:

$$E = E_{\text{collision}} + \lambda_{\text{lap}} E_{\text{lap}} + \lambda_{\text{reg}} E_{\text{reg}}, \tag{5}$$

and we use $\lambda_{\text{lap}} = 0.5$ and $\lambda_{\text{reg}} = 1 \times 10^{-3}$ in our experiments. It is a sparse least-square problem and can be efficiently solved with Cholesky decomposition. We show an example of collision refinement in Figure 1.

3. Qualitative Comparisons

In this section, we qualitatively compare our results to the results of GarSim [TB23]. As can be seen in Figure 2, our model is able to create more dynamics thanks to the global awareness introduced by our manifold-aware transformer architecture.

4. Network architecture

In this section, we describe the detailed network architecture of our framework.

We use an encoder-only transformer [VSP*17] architecture. Our model contains 8 layers of transformer block. Each transformer block contains a multi-head self-attention layer and a feed-forward layer. The embedding dimension and the feed-forward layer dimension are set to 512. The number of heads is set to 8. We use a dropout rate of 0.1.

For the input of the network, we gather the feature extracted from the past 10 frames, namely $n_{\text{hist}} = 10$. The range of manifold-aware self-attention is controlled by $p_{\text{geo}} = 20$. We choose to use $n_{\text{conn}} = 2$ manifold-aware self-attention heads out of 8 heads. During training time, we split the input geometry into $n_s = 4$ disjoint subsets in the first 100 epochs for efficiency consideration. We then stop the splitting (i.e. $n_s = 1$) for the remaining epochs to enable the network to learn fine details of garments. We train the network with a batch size of 16.

We use hyper-parameters λ_{sv} and λ_{vel} to balance our global loss term. We find that an equal weighting of the singular value loss $\mathcal{L}_{\text{sv}}$ and a higher weighting of the global velocity loss $\mathcal{L}_{\text{vel}}$ delivers optimal results. Thus, we use $\lambda_{\text{sv}} = 1$ and $\lambda_{\text{vel}} = 3$ for all our experiments.

References

[GRH*12] Guan, Peng, Reiss, Loretta, Hirshberg, David A, et al. "Drape: Dressing any person". *ACM Transactions on Graphics (ToG)* 31.4 (2012), 1–10 1.

[SCL*04] Sorkine, Olga, Cohen-Or, Daniel, Lipman, Yaron, et al. "Laplacian surface editing". *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*. 2004, 175–184 1.

[SP04] Sumner, Robert W and Popović, Jovan. "Deformation transfer for triangle meshes". *ACM Transactions on graphics (TOG)* 23.3 (2004), 399–405 1.

[TB23] Tiwari, Lokender and Bhowmick, Brojeshwar. "GarSim: Particle Based Neural Garment Simulator". *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023, 4472–4481 2.

[VSP*17] Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, et al. "Attention is all you need". *Advances in neural information processing systems* 30 (2017) 2.