DOI: 10.1111/cgf.14939
Pacific Graphics 2023
R. Chaine, Z. Deng, and M. H. Kim
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 7

Combating Spurious Correlations in Loose-fitting Garment Animation Through Joint-Specific Feature Learning

Junqi Diao, Jun Xiao†, Yihong He, and Haiyong Jiang†

School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.

† Joint Corresponding author.

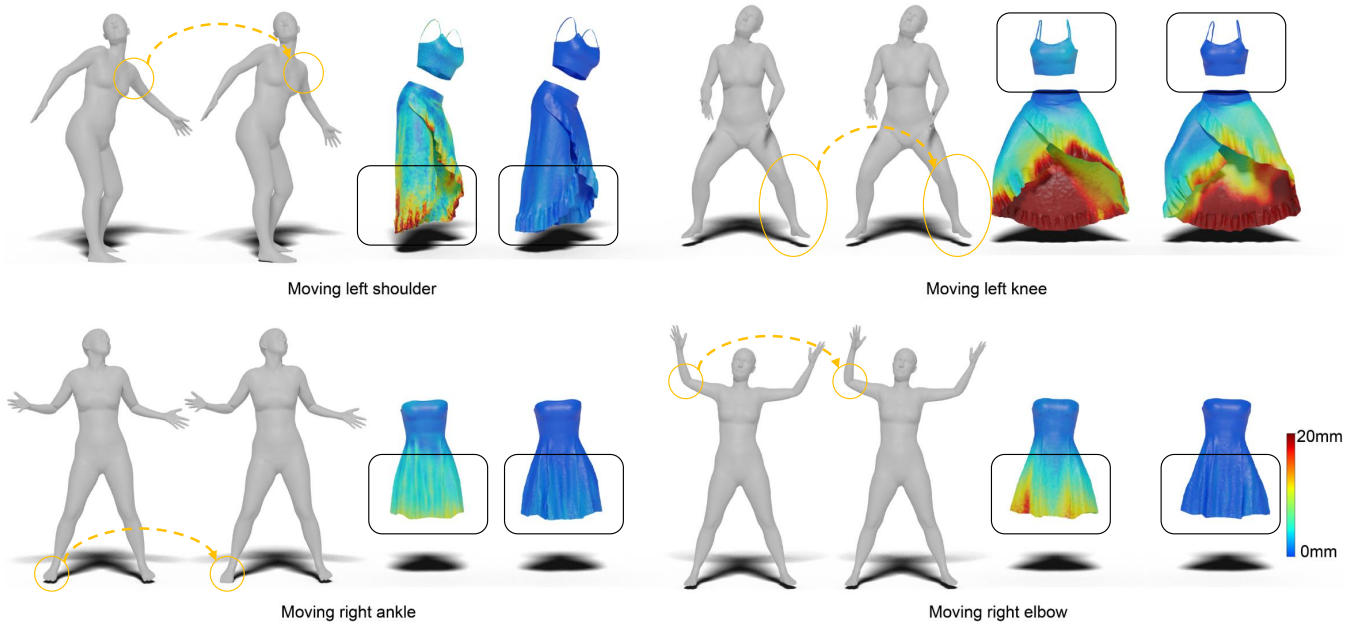

Figure 1: *Spurious correlations between body poses and garment dynamics. From left to right of each motion group: the original pose, the perturbed pose, the relative garment changes of Pan et al. [PMJ*22], and that of ours. For each motion group, we perturb only one joint (highlighted in yellow) and visualize the changes in animated garments. Note that garment parts inside the black boxes are not correlated with the joint motion and should not have too much influence on animated garments.*

Abstract

We address the 3D animation of loose-fitting garments from a sequence of body motions. State-of-the-art approaches treat all body joints as a whole to encode motion features, which usually gives rise to learned spurious correlations between garment vertices and irrelevant joints as shown in Fig. 1. To cope with the issue, we encode temporal motion features in a joint-wise manner and learn an association matrix to map human joints only to most related garment regions by encouraging its sparsity. In this way, spurious correlations are mitigated and better performance is achieved. Furthermore, we devise the joint-specific pose space deformation (PSD) to decompose the high-dimensional displacements as the combination of dynamic details caused by individual joint poses. Extensive experiments show that our method outperforms previous works in most indicators. Moreover, garment animations are not interfered with by artifacts caused by spurious correlations, which further validates the effectiveness of our approach. The code is available at https://github.com/qiji77/JointNet.

CCS Concepts

• ***Computing methodologies*** → ***Procedural animation;***

1. Introduction

> Dress for success. Image is very important.
>
> *Brian Tracy*

Garments are essential in our daily life. They not only keep us warm but also reflect our regional beliefs, individuality, and personality. Therefore, realistic garment simulation is in great demand in numerous applications involving humans, e.g., video games, virtual reality, the fashion industry, and virtual try-ons, just to name a few. A vivid garment is individualized by both its specific tailoring and animation style. The existing workflows usually employ physics-based simulation for high-quality garment dynamics. Nevertheless, this scheme is computationally burdensome and requires professional and tedious parameter tuning, which makes it difficult to scale to real-time applications for processing various kinds of garments.

Recently, data-driven paradigms have attracted extensive attention for improved efficiency. Pioneer work [GRH*12] learns garment deformations from simulated data and can be applied to different body shapes and poses. Motivated by promising achievements of deep learning methods on other vision tasks, there has been a growing interest in designing neural networks for highly non-linear garment deformations [BMTE21, PLPM20, PMJ*22, ZWCM21]. For example, TailorNet [PLPM20] learns garment deformations and wrinkles in the canonical space using a network and drives clothing to deform with body poses through fixed skinning weights. DeePSD [BMTE21] improves garment shapes by learning a set of blend weights and blend shapes, i.e., PSD, for each point on the garment template. However, both TailorNet and DeePSD posit that garments and the human body share similar skinning patterns, therefore often fail to deal with loose-fitting garments. Those garments are not completely consistent with body motions and have more flexible deformation or more wrinkles.

Some recent works [PMJ*22, ZWCM21] have made attempts at loose-fitting garment simulation. These methods generally consider temporal body motions as important clues since the whole body motion rather than a single joint pose determines garment deformations. In particular, Zhang et al. [ZWCM21] first learn the 3D shape and further enhance projected garment images by a rendering network. Therefore, it cannot cater to those tasks requiring fine 3D garments. Another interesting work [PMJ*22] can produce 3D garments by incorporating virtual bone-based skinning for low-frequency shapes and an additional network for high-frequency details. Despite amazing results, these methods rely on global contexts on the whole body pose to determine the skinned garment shape, which leads to spurious correlations between human poses and garment shapes. As exemplified in Fig. 1, moving the left shoulder leads to an unlikely swing of the skirt, which violates the intuition that garment dynamics are only affected by a small subset of related body joints. Moreover, these methods still have difficulty in producing fine garment wrinkles, especially for skirts.

In this work, we address the problem of loose-fitting garment animation from body motions. A good solution should (i) accurately associate motions of each joint to garment deformations so that they behave as consistent as expected; (ii) produce high-quality wrinkle details and garment animation; (iii) require as less training data as possible. To this end, we present a novel garment animation method, which renders sparse correlations between body joints as well as brings high-quality 3D garment wrinkles. Specifically, we learn joint-wise temporal features from a sequence of 3D body poses and then map joint-wise features to those of the relevant garment's virtual bones with a learnable association matrix. We encourage the association matrix to be sparse and non-negative to eliminate spurious correlations between body joints and irrelevant garment vertices. Subsequently, we enhance features of relevant virtual garment bones with garment-specific features that encode the intrinsic garment animation style. Afterward, we predict 3D transformations for each garment's virtual bone with enhanced features, which can be further skinned to produce the high-dimensional garment shape. As the skinned garment may lose track of fine garment wrinkles, we improve the result by predicting displacements for each vertex with a set of joint-specific PSDs in the canonical space. This design can further improve garment quality and make the method easier to learn.

Our contributions are summarized as follows.

- Joint-wise motion feature encoding to avoid spurious correlations between different body joints.
- An association mapping between human joints and a garment's virtual bones, which greatly reduces spurious correlations in garment animation.
- A joint-specific PSD in the canonical space learns more realistic garment details specific to each joint.

Extensive experiments show that our method achieves significant improvements in all indicators. Moreover, the qualitative results are not affected by false correlations between body joints and garment deformations, and better visual quality is obtained.

2. Related Works

3D garment animation can be achieved by two main paradigms including physically-based simulation (PBS) and data-driven garment synthesis. We review these related works in this section.

2.1. PBS

PBS is a well-studied methodology to generate high-quality garment animation by modeling the interactions between real-world forces and garments according to physical laws [BW98, Pro97, SSIF08, TPBF87]. However, PBS-based methods are usually computation-hungry and sensitive to low-quality garment typologies because of complex collisions and interactions among garment, body, and physical forces. Therefore these methods often struggle to generate high-resolution garment animation efficiently. Some works attempt to reduce the computational cost [VSC01] by designing more efficient physical energy objective functions [GHF*07, JGT17, LBK17] and position-based simulation [MHHR07]. Another line of work leverages CUDA acceleration and parallel algorithm design to efficiently simulate realistic garment animation [WWW22, Zel05]. For example, Tang et al. [TWL*18] use spatial hashing for continuous and incremental tracking of deformed garment vertices and resolve penetrations by a non-linear GPU-based impact zone solver. Wu et al. [WWYW20]

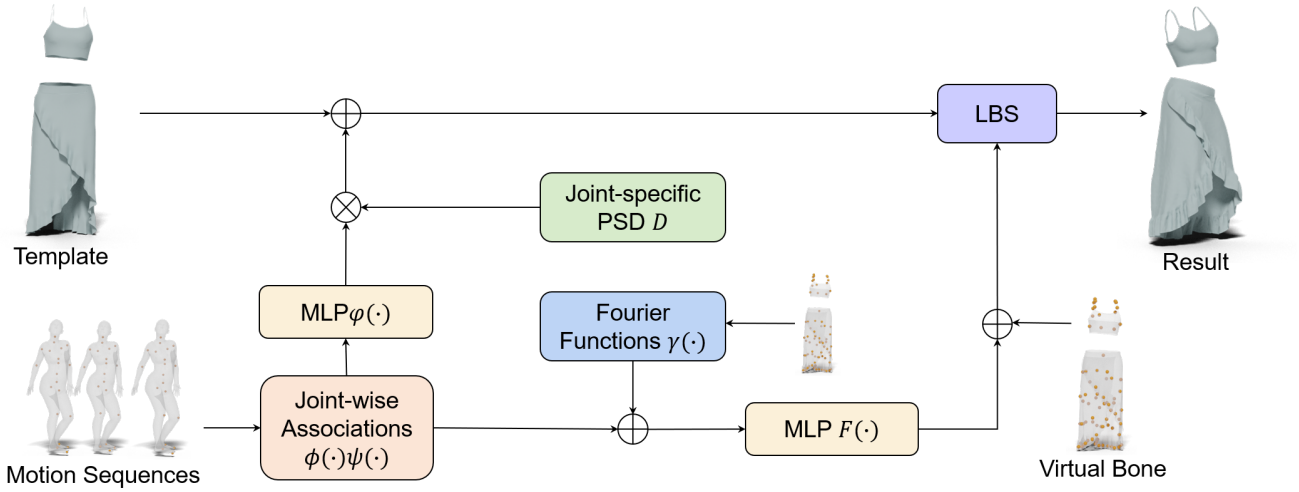

Figure 2: *Overview of our method. Given 3D body motions, we feed them into the joint-wise association module to map joint-wise features encoded by GRU to different virtual garment bones with a sparse association function $\phi(\cdot)\psi(\cdot)$. These features are then concatenated with garment-specific features encoded by a Fourier function $\gamma(\cdot)$ to predict rigid transformations of each virtual garment bone with MLP $F(\cdot)$. At the same time, joint-wise motion features are also sent to MLP $\varphi(\cdot)$ to predict a set of joint-specific blending weights, which are multiplied with joint-specific PSD D for garment displacement in the canonical space. Finally, the garment shape is skinned with LBS to produce the final result.*

cope with collisions by decomposing the problem into soft constraint and strict constraint enforcement. Li et al. [LTT*20] make it possible to perform cloth simulation on multiple GPUs. Another interesting work [Wan21] enables sub-millimeter level cloth simulation with millions of vertices, which is computationally expensive and cannot cater to real-time applications. In general, PBS-based methods can produce realistic garment simulations at the cost of significantly high computation complexity. Moreover, PBS relies on many physical parameters, which can take hours to tune even for experts.

2.2. Data-driven Simulation

Data-driven simulation methods learn to generate realistic garment deformations from a set of collected garments. Compared to PBS, this category of works usually delivers faster speed and requires fewer computation resources as highlighted in many papers [CGY*21, GRH*12, GCP*20, GCS*19, PLPM20, WCC*21]. These methods learn to deform garments from high-quality 3D data compiled by PBS methods off-line [BMTE21, BME20, JZGF20, KKN*13, PLPM20, SOC19, TB21]. The basic workflow is to first learn PSD [LCF00] for garments and then use linear blend skinning (LBS) to animate the mesh in the rest pose [VSGC20]. For example, TailorNet [PLPM20] decomposes 3D animated garments into pose-driven low-frequency shapes and shape-and-style-dependent high-frequency details. DeePSD [BMTE21] improves garment shapes by learning PSD on the garment template. DeepDraper [TB21] additionally considers the impact of measurements on clothing and proposes perceptual constraints to improve the representation of high-frequency details. Santesteban et al. [SOC19] further improve high-frequency wrinkles by introducing GRU units. These methods only demonstrate results on tight-fitting clothes and cannot be expected to work on loose ones. Because loose clothes usually have much larger deformation freedom and do not follow body motions as closely as tight-fitting clothes. In addition, some methods try to directly reconstruct clothed human bodies from digital scanning [SYMB21], images [PSRC*19] or videos [AMX*18]. Despite amazing results, these methods assume similar skinning weights between clothes and 3D human bodies, which makes it infeasible to animate loose clothing.

To address the above problem, some methods incorporate human kinematics and physics to drive cloth deformation [BME21, HLB*23, SRPMN21, SOC22, STOC21, ZWCM21, WSFM19, GBH23]. However, these methods may not be robust for large motions and do not work well on high-frequency deformations. Another line of work tackles the problem by designing novel frameworks and incorporating garment priors [TB23]. DNG [ZWCM21] first learns the deformation of the coarse model and then generates pixel-level frame renderings of complex target garments with an additional network. Zhang et al. [ZCM22] learn features for each garment point on the UV map based on body motion sequences and a history of how the garment deforms. The method subsequently predicts per-vertex skinning weights and displacements in the canonical space of the garment. Pan et al. [PMJ*22] suggest using virtual bones [LD12] for clothing deformation, bringing a new idea to loose-fitting clothing animation. Compared to those methods on tight-fitting clothing animation, Pan et al. [PMJ*22] learn virtual bone motions from a global code encoding a set of body sequences and finally produce clothing

wrinkles through virtual-bone skinned shapes and clothing motions achieving state-of-the-art performance on loose-fitting garment animation. AnchorDEF [ZLH*23] utilizes human motions as input to learn a set of anchor points for the garment, along with their corresponding rigid transformation matrices and the offsets of each point in the canonical space. However, these methods learned from the global code can be easily affected by irrelevant body joints and yield unexpected influences on loose-fitting clothes. In this work, we investigate how to make the influence of body motions on virtual bones more local to alleviate undesirable interference.

3. Method

Given a sequence of 3D body poses $\mathcal{P}_B^{1\cdots t}$ and a source 3D garment G_S at the rest pose, our goal is to predict the target 3D garment G_T^t for each body pose $\mathcal{P}_B^t$ so that a garment can be animated with a set of the body pose inputs. We represent human pose $\mathcal{P}_B^t$ with the axis-angle rotations of $J = 20$ body joints w.r.t. their parent joints $R_B^t \in \mathbb{R}^{J\times 3}$ and the position of the root joint $T_B^t \in \mathbb{R}^{1\times 3}$ following [LMR*15], where hand joints are not considered for their irrelevances with garment deformations. This problem is challenging in several aspects. First, garment deformations, especially loose-fitting ones, have a high degree of freedom and usually do not follow the posed body shapes closely. Second, correlations between body joints and garment deformations are complex but localized, where each joint is expected to affect only a subset of garment vertices. Third, the wrinkles of the garment are intricate and exhibit large variations, making them difficult to learn.

In this work, we present a novel method that ensures the localized influence of body joints on the garment and the high fidelity of garment wrinkles. The overall pipeline is shown in Fig. 2. At first, we introduce virtual bones that act as an intermediate representation between the garment and body poses to alleviate the gaps between high-dimensional garment motions and body motions (see Sec. 3.1). Afterward, we learn sparse correlations between body joints and virtual bones and encode garment-specific features for coarse garment shape skinning in Sec. 3.2. Last but not least, fine-level winkles of the garment are accounted for by learning displacements in the canonical space with a set of joint-specific PSDs so that the learned winkles are pose-independent and specific to different joints in Sec. 3.3.

3.1. Virtual Bones for Garment Animation

In 3D digital designs, artists usually rely on rigged garments for garment animation. However, garment rigging requires professional and tedious labor and cannot scale up. Following [PMJ*22], we first use Marvelous Designer to generate a sequence of mesh models. Next, we employ SSDR [LD12] to solve for a rest pose mesh U, the skinning weight W, and a series of virtual bone poses $\mathcal{P}_V = \{R_V^v|T_V^v\}^{|V|}$, where $R_V^v \in \mathbb{R}^{3\times 3}$ denotes the rotation matrix of the v-th virtual bone, $T_V^v \in \mathbb{R}^3$ is the v-th virtual bone's translation, and $|V| = 80$ is the number of virtual bones. Skinning weight W encodes the influences of each virtual bone on the final deformation of each garment vertex and is represented by a sparse, non-negative $N \times |V|$ matrix, N is the number of garment vertices, and $|V|$ is the number of virtual bones. The sum of elements in any row of W is equal to 1. A garment and its virtual bones are shown in the top left and bottom right of Fig. 2. Therefore, high-dimensional 3D garment deformations can be narrowed down to a small set of learned 3D rigid transformations defined on virtual bones. Though fine-scale wrinkle motions may be lost, we can recover most coarse garment deformations through skinned virtual bones.

Given virtual bones $\mathcal{P}_V$ and skinning weights $W \in \mathbb{R}^{N\times|V|}$, the deformed garment G_T can be attained according to virtual bone-based skinning. We can calculate the i-th deformed garment vertex $G_{T,i}$ by transforming a vertex $G_{S,i}$ at the rest pose with blended 3D rigid transformations of virtual bones as follows:

$$\begin{aligned} G_{T,i} &= LBS(\mathcal{P}_V, W, G_{S,i}) \\ &= \sum_{v=1}^{|V|} w^{i,v}(R_V^v G_{S,i} + T_V^v), \end{aligned} \tag{1}$$

where $w^{i,v}$ indicates the skin weight for vertex $G_{S,i}$ regarding to virtual bone v.

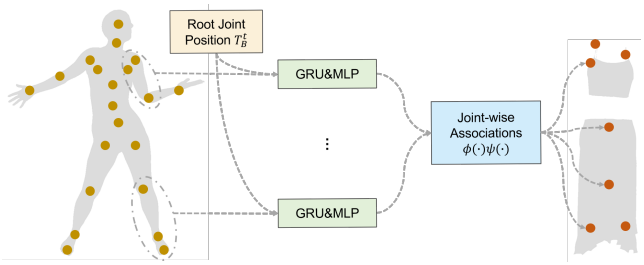

Figure 3: *Joint-wise associations. Both axis-angle vectors of neighboring joints and the position of the root joint are concatenated and fed into a joint-wise GRU layer to learn temporal features for each joint. Thereafter, we use a sparse association matrix to map joint-wise features to related virtual bones.*

3.2. Associating Body Joints and Virtual Bones

After we have virtual bones and corresponding skinning weights W, 3D garment animation boils down to estimating 3D transformations of virtual bones $\mathcal{P}_V^t$ from a sequence of t-frame poses. Previous works [PMJ*22] consider body joints as a whole and encode the concatenation of all body joints for virtual bones prediction, therefore virtual bones are influenced by all body joints. Although this scheme can capture contexts between different joints, it can easily lead to spurious correlations between garments and unrelated body joints. For example, arm poses should not affect the swing of the bottom of a skirt as shown in Fig. 1.

In this work, we cope with the above problem by requiring each body joint to only affect a subset of garment vertices as illustrated in Fig. 3. Given the body poses $\mathcal{P}_B^{1\cdots t}$, we encode temporal motion features with a function $\psi(\cdot) : \mathbb{R}^{T\times J\times C} \rightarrow \mathbb{R}^{J\times K}$, where K is the dimension of virtual bone features and C denotes the feature dimension of input joint poses. For each joint, we concatenate the root joint position T_B^t, its joint rotation R_B^t, and that of its two neighbors on the human skeleton as joint features. The temporal feature

encoder $\psi(\cdot)$ consists of a joint-wise single GRU layer (J GRUs with input size C and hidden size 600) following a single layer perceptron with 480 output channels, where joint-wise GRUs alleviate feature interferences between different joints.

Inspired by [OBB20], we devise a learnable association matrix $A \in \mathbb{R}^{|V|\times J}$ so that only essential joint poses are used for virtual bone estimation so that spurious correlations are mitigated. As the association matrix should be non-negative, we use $\phi(A) = ReLU(A)$ to threshold negative weights and restrain spurious correlations. Then, the rigid transformations of virtual bones can be determined by:

$$\mathcal{P}_V^t = F(\phi(A)\psi(\mathcal{P}_B^{1\cdots t})), \tag{2}$$

where $F(\cdot) \in \mathbb{R}^{|V|\times K} \rightarrow \mathbb{R}^{|V|\times 6}$ is a channel-wise function to predict the three-dimensional Euler angles and the three-dimensional offsets for each virtual bone. $F(\cdot)$ is implemented as a four-layer perceptron $(K, 1024, 2048, 1024, 6)$

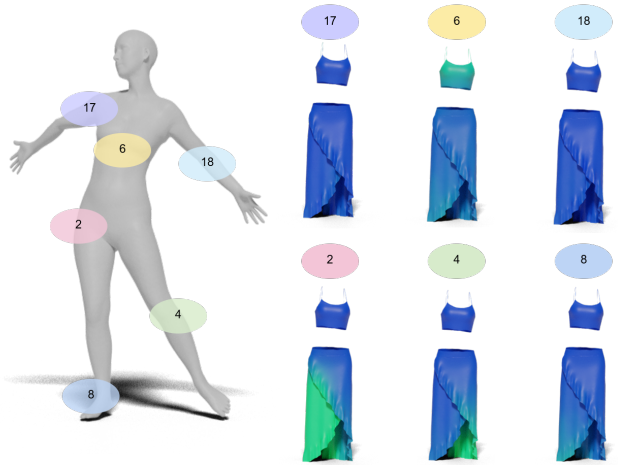

Figure 4: *Examples of learnt association function $\phi(A)$. On the left, each circle denotes a body joint and its corresponding index id. On the right, the influences of each joint are shown with the green color highlighting impacted garment regions.*

During training, body joints with zero activation with respect to virtual bones will have no corrective effect on the virtual bone predictions. We further encourage the association matrix $\phi(A)$ to be sparse so that unnecessary correlations can be eliminated. In Fig. 4, we show the learned association function $\phi(A)$, where only related garment regions are affected by body joints. However, this does not imply the loss of long-range effects of body joints on the garment. As shown in index 2, the motion of hips not only affects garment vertices near body joints but also has an impact on the distal part, such as the skirt hem.

Except for the body poses, each 3D garment usually exhibits specific geometric patterns, such as camisole and waistband. Therefore we augment the body pose-based features with a virtual bone-based garment prior. We feed 3D positions of each virtual bone T_V^i at the rest pose to a Fourier function [TSM*20] so that garment-specific feature f_V^i can be learned:

$$\begin{aligned} f_V^i = \gamma(T_V^i) = [&\alpha_1 cos(2\pi \mathbf{c_1} \cdot T_V^i), \alpha_1 sin(2\pi \mathbf{c_1} \cdot T_V^i), \\ &\cdots, \alpha_m cos(2\pi \mathbf{c_m} \cdot T_V^i), \alpha_m sin(2\pi \mathbf{c_m} \cdot T_V^i)], \end{aligned} \tag{3}$$

where α_m is the randomly initialized Fourier series coefficient, $\mathbf{c_m}$ is the Fourier basic frequency. We then concatenate f_V^i as garment-specific features f_V. Basically, these features encode garment-specific animation styles. The combined features of the body pose-based features and f_V are fed into $F(.)$ to estimate 3D rigid transformations for each virtual bone as follows:

$$\mathcal{P}_V^t = F(\phi(A)\psi(\mathcal{P}_B^{1\cdots t}) + f_V). \tag{4}$$

3.3. Joint-specific PSD Displacements

Notwithstanding virtual bone-based skinning can recover coarse garment shapes, it usually loses track of high-frequency wrinkles. Previous works [PMJ*22] learn displacements based on the skinned garment shape to enrich details. We instead model garment wrinkles in the canonical space. The advantage is that different posed garments are aligned in the same space and many of the local geometric details become invariant to large body articulations. Recent works on human reconstruction and neural radiance field [DLJ*20, SYMB21, CJS*22, CZB*21] suggest that learning shape details in the canonical space can improve the results.

However, learning these details still requires memorizing every pose-dependent detail, which hinders its generalization and learning efficiency. We observe that each joint only affects a small region of garment wrinkles, which can be further decomposed into a combination of basic blending shapes as PSD [LCF00]. Therefore, we devise joint-specific PSDs to learn joint-specific wrinkle variations. In this way, complex wrinkles are reduced to a very small set of joint-specific weights. The basic network structure is illustrated in Fig. 5. Then we can formulate the joint-specific PSD as follows:

$$\hat{G}_S^t = \sum_{j=1}^{J} \varphi(\psi(\mathcal{P}_B^{1\cdots t,j}))D^j + G_S, P \tag{5}$$

where $\psi(\cdot)$ is explained in Eq. (2), $\varphi(.) : \mathbb{R}^K \rightarrow \mathbb{R}^{|X|}$ is a three-layer perceptron $(K, 480, 128, |X|)$ taking the output features $\psi(\cdot)$ to predict blending weights for each joint-specific PSD, and $D^j \in \mathbb{R}^{|X|\times N\times 3}$ is the PSD matrix for j-th body joint. Then, $\hat{G}_S^t$ is deformed by virtual bone-based LBS to generate garment G_T^t with skinning weight W. The equation for the skinning process is shown in Eq. (1).

3.4. The Objective

In this subsection, we describe the loss terms for network optimization. The overall objective function is defined as:

$$L = L_{cloth} + \lambda_{norm} \cdot L_{norm} + \lambda_s \cdot L_{sparse}, \tag{6}$$

where hyper-parameter λ_{norm} and λ_s balance different loss terms. L_{cloth} calculates the difference between the deformed garment G_T^t and the corresponding ground truth G_T^t* as follows:

$$L_{cloth} = \frac{1}{T}\sum_{t=1}^{T} \left\| G_T^t - G_T^t* \right\|_1. \tag{7}$$

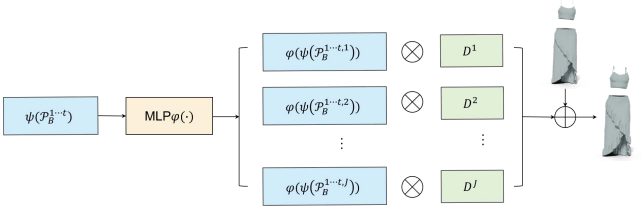

Figure 5: *Details of joint-specific PSD. We devise a set of learnable PSD D^j for each human joint and then multiply it with blending weights $\varphi(\psi(\mathcal{P}_B^{1\cdots t,j}))$ predicted from the motion features of each body joint. Then we add predicted offsets to the garment template as the rest-posed garment for skinning.*

Similarly, normal loss L_{norm} is enforced on normal between the deformed garment and the ground truth to improve local geometry. Sparsity loss L_{sparse} encourages the sparsity on correlation weights between body joints and virtual bones with $L1$ norm:

$$L_{sparse} = \|\phi(A)\|_1 . \tag{8}$$

4. Experments

4.1. Datasets and Evaluation Metrics

In our experiments, we adopt the garment animation dataset proposed in [PMJ*22]. The dataset is generated by garment simulation on two different types of loose-fitting garments driven by a digital avatar, denoting ($D1$,$D2$). The driven human motions are collected from the Internet and have a diversity of complex human poses. The dataset is divided into 35000 frames for training and 5000 frames for testing. For each garment, we obtain the virtual bones and corresponding skinning weights as [PMJ*22]. To validate the effectiveness on a wider range of garment types, we selected three different types of loose-fitting garments in Cloth3D [BME20], denoting ($D3$,$D4$,$D5$). We applied the same set of poses as in [PMJ*22] and generated the ground truth garment mesh sequences using Marvelous Designer. Then, we computed the virtual bone for the garments using SSDR [LD12]. We set the number of bones to 80 to maintain the consistency with [PMJ*22].

For evaluation, we embrace two metrics: RMSE (Root Mean Squared Error), and Hausdorff distance. Basically, RMSE and Hausdorff distance reflect the vertex distance between predicted garments and their ground truth.

4.2. Implementation Details

The method is implemented based on the released code of [PMJ*22] and is trained on an NVIDIA Tesla V100 GPU with 32GB memory. we adopt truncated back propagation through time [WP90] with 50-time steps to reduce the memory consumption. The network is optimized using RMSProp optimizer and we set the batch size to 8 and the initial learning rate to 10^{-3}, which decays by 70% every 30 epoch. In our experiments, we split the training process into two stages. In the first stage, we optimize the network without the garment wrinkle part in Sec. 3.3. Afterward, we tune the whole network with all modules. Both stages are trained with all loss terms. The code will be released for research purposes.

Table 1: *Quantative comparisons with previous works.*

	Method	Skinned		Final	
		RMSE↓	Hausdorff↓	RMSE↓	Hausdorff↓
D1	Tailornet [PLPM20]	-	-	41.59	133.26
	DNG [ZWCM21]	-	-	49.20	144.45
	Pan et al. [PMJ*22]	28.40	110.49	26.90	107.50
	Ours	**24.98**	**98.95**	**24.18**	**98.61**
D2	Tailornet [PLPM20]	-	-	36.71	91.87
	DNG [ZWCM21]	-	-	36.42	106.65
	Pan et al. [PMJ*22]	21.23	71.30	21.04	71.71
	Ours	**19.16**	**63.62**	**18.74**	**63.50**
D3	DNG [ZWCM21]	-	-	33.72	83.33
	Pan et al. [PMJ*22]	16.28	53.25	15.52	51.15
	Ours	**14.30**	**46.05**	**14.14**	**46.51**
D4	DNG [ZWCM21]	-	-	55.47	137.76
	Pan et al. [PMJ*22]	38.07	106.11	37.63	105.09
	Ours	**33.88**	**98.76**	**33.52**	**98.36**
D5	DNG [ZWCM21]	-	-	30.78	75.57
	Pan et al. [PMJ*22]	16.93	55.34	16.51	54.56
	Ours	**14.11**	**48.11**	**14.18**	**47.83**
D6	DNG [ZWCM21]	-	-	50.59	137.31
	Pan et al. [PMJ*22]	25.29	85.48	24.76	89.57
	Ours	**21.80**	**76.09**	**21.59**	**76.24**

4.3. Comparison with Previous Work

We compare our method with the state-of-the-art loose-fitting garment deformation methods, including Pan et al. [PMJ*22], DNG [ZWCM21], and TailorNet [PLPM20]. We retrain the network [PMJ*22,ZWCM21] with their released code for comparison. For TailorNet, we adopt the results reported in [PMJ*22], which has the same configuration as ours.

Tab. 1 shows the quantitative comparison. Compared to [PMJ*22] on the skinned results, our method achieves about $+2.75mm$ gains in RMSE, $+9.61mm$ in Hausdorff distances on dress (D1 and D2), and about $+2.75mm$ gains in RMSE, $+9.61mm$ in Hausdorff distances on average on dress (D3, D4, and D5). This suggests our method overwhelms [PMJ*22] on both geometry and local edge structures. A major difference between our method and [PMJ*22] is that Pan et al. [PMJ*22] encode motion features of all body joints with a GRU layer, while ours extracts motion features in a joint-wise manner and augments them with garment-specific features f_V. Therefore our joint-wise design can prohibit interferences in the garment shapes from irrelevant body joints as well as improve the animation quality. For final results, our method surpasses previous methods for at least $2.7mm$ on D1 and $2.3mm$ on D2 on RMSE. The results confirm our advantage over previous

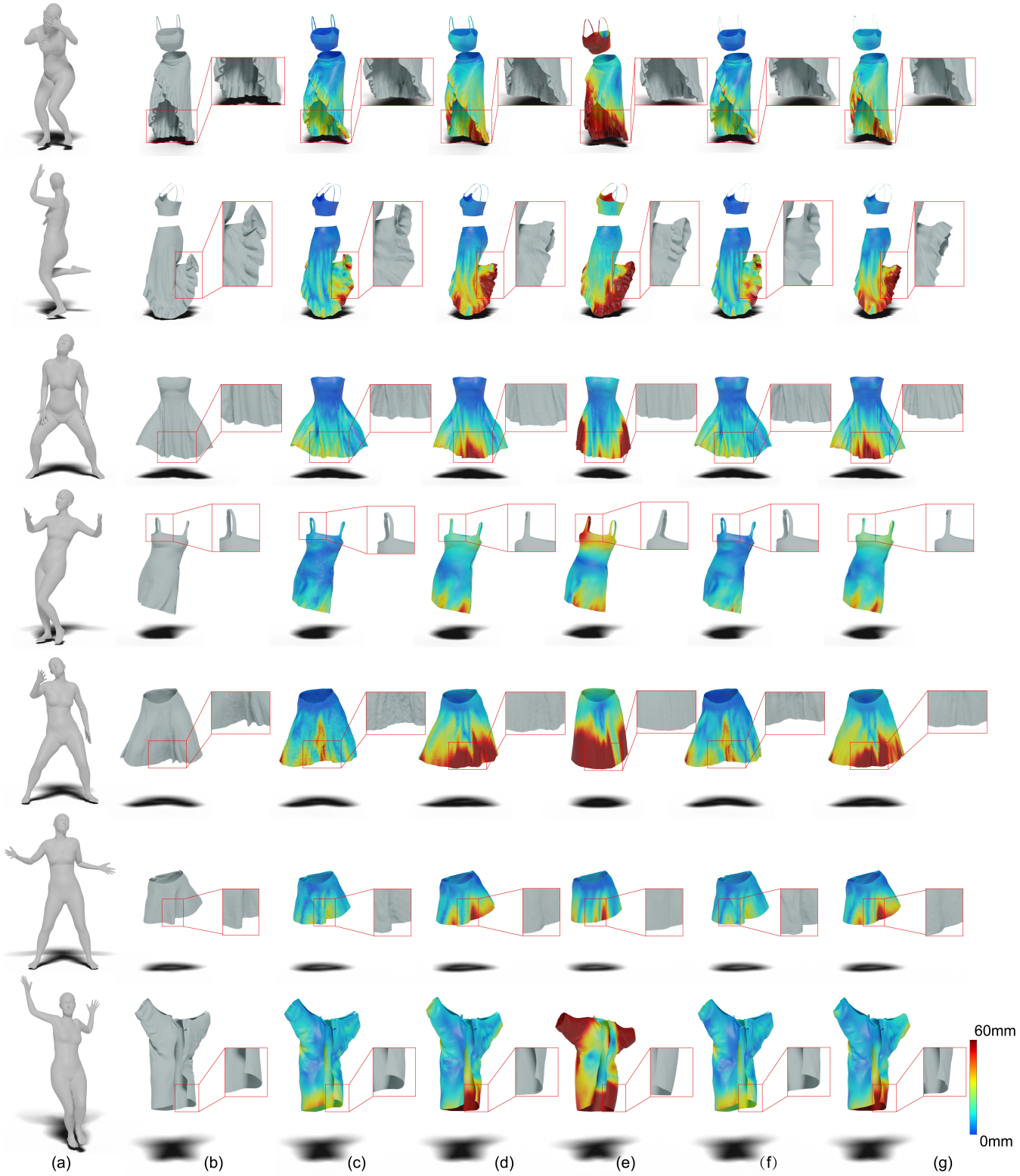

Figure 6: *Qualitative comparisons with previous methods. The first two rows are for dress1, and the last five rows are for dress2, dress3, dress4, dress5, and dress6. From left to right: (a) 3D body poses, (b) the ground truth garments, (c) ours, (d) Pan et al. [PMJ*22], (e) DNG [ZWCM21], (f) skinned results by ours, (g) skinned results by Pan et al. [PMJ*22].*

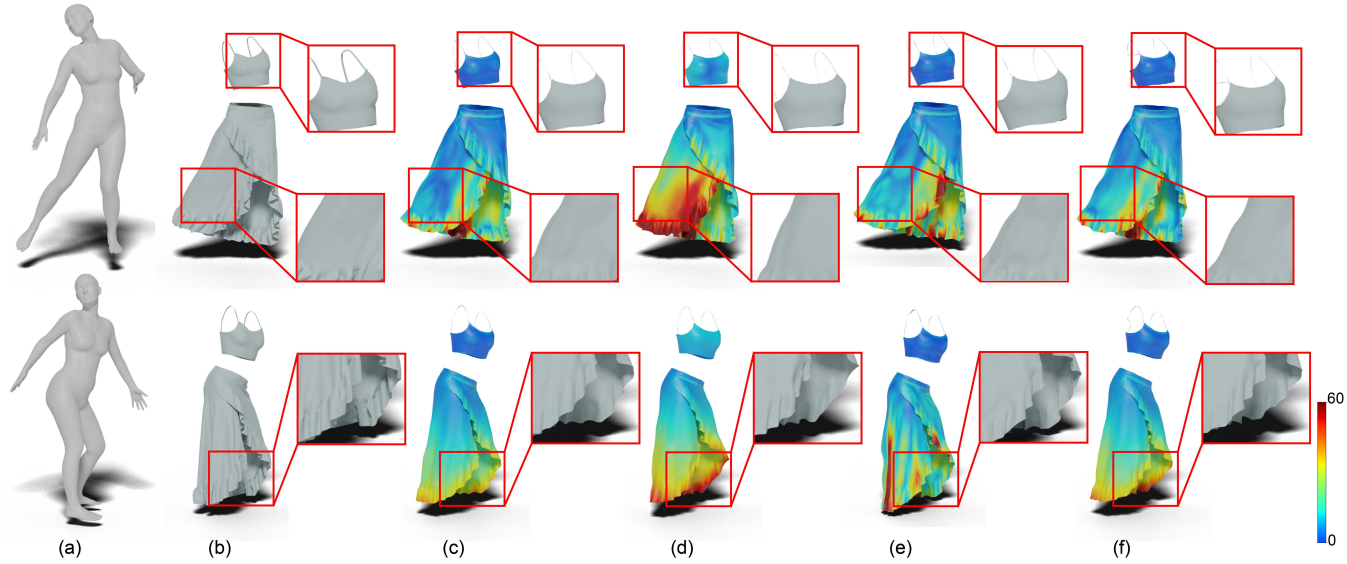

Figure 7: *Ablation studies of skinned results. From left to right: (a) body poses, (b) ground-truth garments, (c) our results, (d) results of ours without $\phi(\cdot)\psi(\cdot)$, (e) results of KNN associations, (f) results of ours without f_V.*

work on loose-fitting garment animation. Our method surpasses previous methods for at least 4.1*mm* on dress D4 on RMSE. Basically, our method can attain better results on looser-fitting garments. For example, dress D4 has fewer fixes than the other ones. Fig. 6 demonstrates the qualitative results of different methods. The results reveal that our method can produce more realistic garments, especially the loose parts and garment wrinkles.

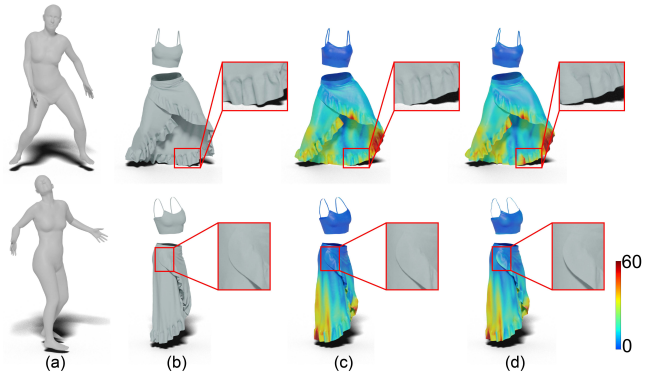

Figure 8: *Ablation studies of final results. From left to right: (a) body poses, (b) ground-truth garments, (c) our results, (d) results of ours without PSD. See the zoom-in regions for geometry differences.*

4.4. Ablation Experiments

In this section, we evaluate the influences of key module designs. We conduct all experiments on dress1.

Joint-specific Associations. Because the association matrix $\phi(A)$ and joint-wise motion features $\psi(\mathcal{P}_B^{1\cdots t})$ depends on each other, we replace them with a single GRU layer rather than removing them one by one (denoting without $\phi(A)\psi(\mathcal{P}_B^{1\cdots t})$). On the first row of Tab. 2, we find this change results in a 6.0*mm* drop in RMSE and an 11.5*m* drop in Hausdorff distance. To further validate how well a prior-based association matrix can work, we directly selected the nine nearest human joint nodes for each virtual bone based on their distance (denoted as KNN association), where the best performance is attained when $K = 9$. On the second row of

Table 2: *Ablation studies on skinned results and final results. For ablated methods,* ours without $\phi(\cdot)\psi(\cdot)$ *replaces the association module with one GRU layer,* KNN association *means directly selecting KNN human joint nodes,* ours without f_V *removes garment-specific features* f_V, ours without PSD *replaces the joint-specific PSD module with a high-frequency module on the baseline method [PMJ*22].*

Method		RMSE↓	Hausdorff↓
Skinned	without $\phi(\cdot)\psi(\cdot)$	30.94	110.50
	KNN association	27.48	107.18
	without f_V	26.52	103.01
	Ours	**24.98**	**98.95**
Final	without PSD	24.69	98.71
	Ours	**24.18**	**98.61**

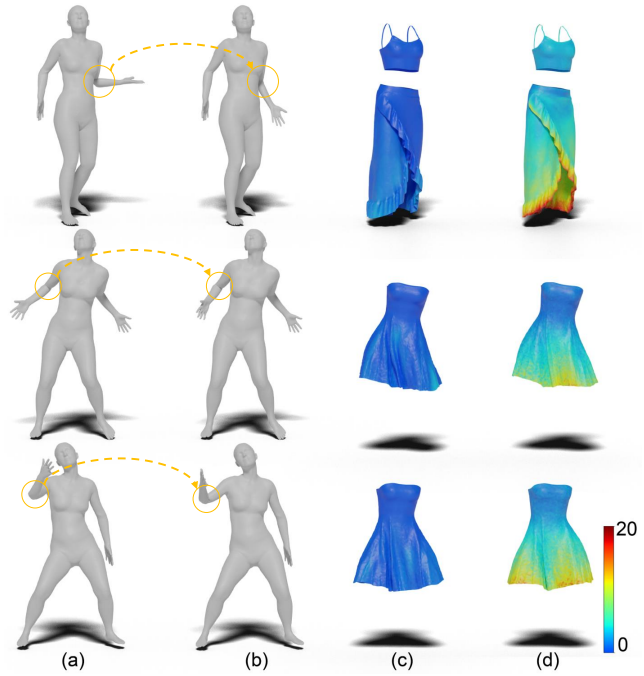

Figure 9: *Results on perturbed joint poses. From left to right: (a) original body pose, (b) the body pose with perturbed joints, (c) garment changes of ours, (d) garment changes of [PMJ*22]. We move only one joint in each row. Note that our method has fewer uncorrelated garment movements.*

Tab. 2, the metrics of RMSE and Hausdorff drop by 5.26*mm* and 22.65*mm*. Removing garment-specific features f_V on the third row of Tab. 2 leads to a 1.5*mm* drop in RMSE and a 4.0*mm* drop in Hausdorff distance. The baseline method [PMJ*22] can be treated as removing both joint-wise motion features, association matrix, and garment-specific features. By comparing the first row with results of [PMJ*22], we can see garment-specific features will have less influence if features of all body joints are joined for motion feature learning. Fig. 7 gives visual results of the experiments.

Notice $\phi(\cdot)\psi(\cdot)$ and f_V are important to high-quality garment results. Comparing (c) with (e), we can find that opting for the direct selection of joints yields inferior outcomes and fewer details. This outcome can be attributed to the fact that directly selecting a fixed number of joints might inadvertently neglect certain pivotal joints, consequently compromising the overall quality of the results.

Joint-specific PSD. We explored the impact of joint-specific PSD on the results. Removing this module yields a 0.5*mm* decrease in RMSE and a 0.1*mm* drop in the Hausdorff distance. Therefore the influence of joint-specific PSD is positive but very slight. Although the boost is small, the results of the PSD are closer to the details of the GT surface, as shown in Fig. 8

Robustness to Irrelavent Joints. We select several joints that usually affect the garment shape and perturb them with some random poses. After the perturbation, garment shapes should change as little as possible w.r.t. the original. In Fig. 9 and Tab. 3, we report the results. Notice that our method is more robust to perturbations on unrelated local joints.

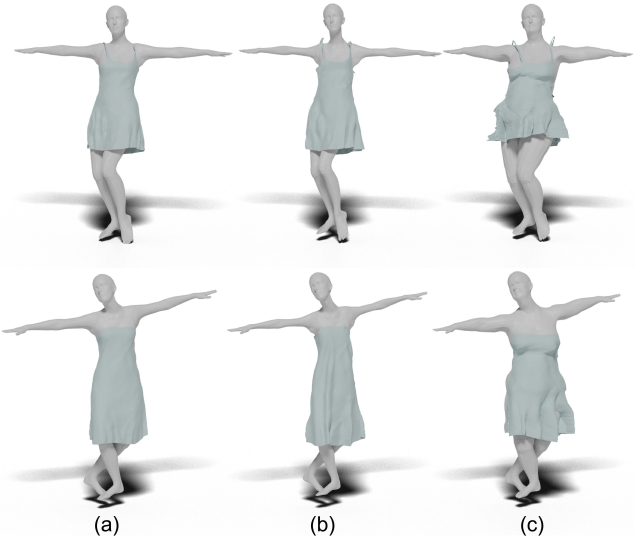

Figure 10: *Extension to Different Body Shapes. The first row shows different human bodies, and the second and third rows show animated garments in different poses.*

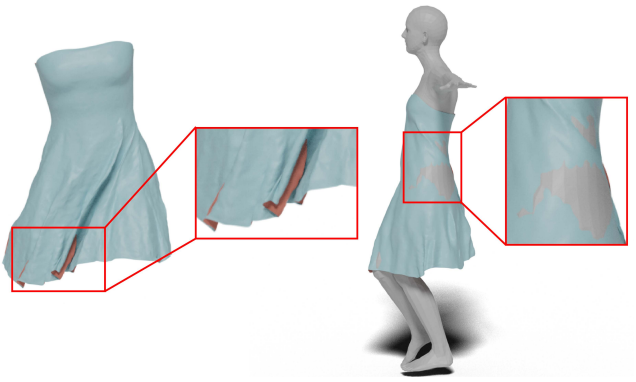

Figure 11: *Failure cases. Some parts of garments may have self-collisions or body-garments collisions as we do not model collisions.*

The Impact of Training Data Size. During experiments, we also found that our method is more data-efficient. In Tab.4, we show the results trained with different amounts of data (50% and 30%, respectively). Basically, our method can achieve similar results as [PMJ*22] even with only half of the training data.

4.5. Extension to Different Body Shapes

We demonstrate the generalization capabilities across different body types, as depicted in Fig 10. The network is trained on a

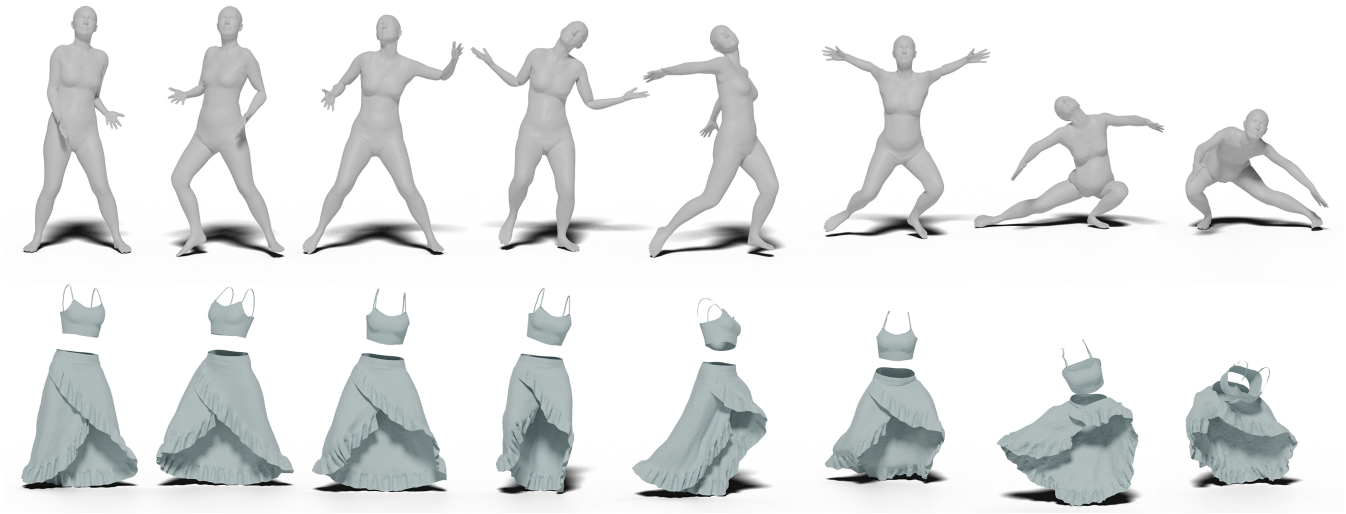

Figure 12: *Garment animations for a sequence of human motions. The top row shows human motions, and the second row denotes animated garments.*

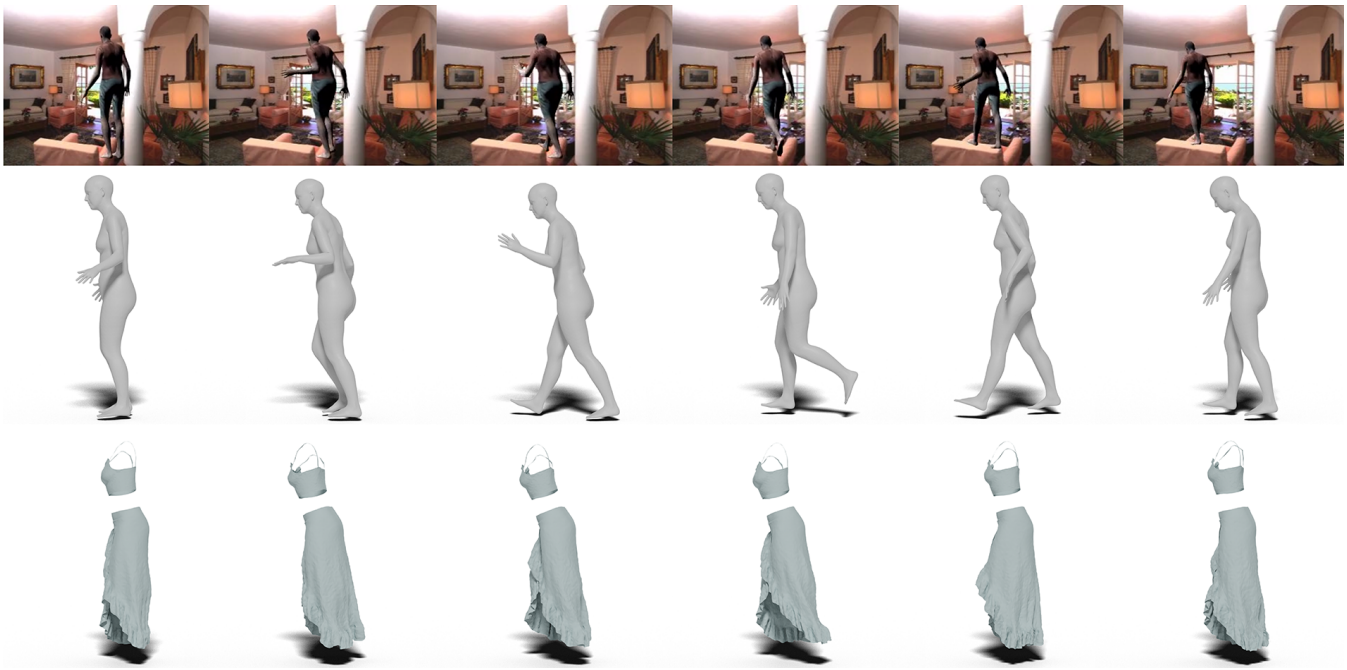

Figure 13: *Garment animations with human motions from video input. The first row gives a sequence of video frames, the second row is the corresponding human motion, and the last row demonstrates our animated results.*

medium body shape (a) and subsequently tested on lean and fat body shapes (b and c). During testing, we adjusted the garment template using body shape priors. The experimental results demonstrated that our method can be extended to different body shapes and produce satisfactory results by moving virtual bones based on the differences in body shapes.

4.6. Applications

Body motion-based garment animations enjoy wide applications, such as virtual try-ons and garment designs. In Fig. 12, we demonstrate an example of how a garment can be animated by a sequence of body motions. An interesting application of the method

Table 3: *RMSE results on four randomly perturbed joints.*

Method	Joint-18	Joint-16	Joint-17	Joint-19
Pan et al. [PMJ*22]	6.9	7.0	4.5	3.8
Ours	**0.5**	**1.2**	**1.0**	**0.6**

Table 4: *Results on different amounts of training data.*

Training data	Method	Skinned	Final
30%	Pan et al. [PMJ*22]	36.54	36.41
	Ours	32.27	31.55
50%	Pan et al. [PMJ*22]	32.10	31.88
	Ours	28.02	27.37
all data	Pan et al. [PMJ*22]	28.41	26.90
	Ours	**24.98**	**24.18**

is to animate garment dynamics based on a natural image or video input. In this part, we select a video sequence of the surreal dataset [VRM*17] and then use the corresponding human motions to animate garments. Results are shown in Fig.13.

4.7. Limitations

Our method still has some limitations. For instance, the trained network can only be used for the animation of a specific kind of garment. In addition, the present approach does not model the complex collisions between garment parts or between garments and body, therefore self-penetration and collisions may occur sometimes as shown in Fig. 11. We leave these two limitations to future works.

5. Conclusions

This work presents a novel approach to body motion-based garment animation. Our method first learns to map joint-wise motion features to features of related virtual garment bones with a sparse association matrix. Then the mapped features of virtual bones along with garment priors are concatenated to learn 3D transformations of virtual bones. Besides, we devise joint-wise PSDs to learn garment wrinkles in the canonical space. Experiments show that our method eliminates counterfeit garment deformations caused by unrelated body joints, improves the quality of garment deformations, and eases the learning process.

6. Acknowledgments

This work is supported by the National Natural Science Foundation of China (U2003109, U21A20515, 62102393, 62206263, 62271467), China Postdoctoral Science Foundation (2022T150639, 2021M703162), the State Key Laboratory of Robotics and Systems (HIT) (SKLRS-2022-KF-11), and the Fundamental Research Funds for the Central Universities.

References

[AMX*18] ALLDIECK T., MAGNOR M., XU W., THEOBALT C., PONS-MOLL G.: Video based reconstruction of 3d people models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 8387–8397. 3

[BME20] BERTICHE H., MADADI M., ESCALERA S.: Cloth3d: clothed 3d humans. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16* (2020), Springer, pp. 344–359. 3, 6

[BME21] BERTICHE H., MADADI M., ESCALERA S.: Pbns: Physically based neural simulation for unsupervised garment pose space deformation. *ACM Transactions on Graphics 40*, 6 (2021), 198. 3

[BMTE21] BERTICHE H., MADADI M., TYLSON E., ESCALERA S.: Deepsd: Automatic deep skinning and pose space deformation for 3d garment animation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 5471–5480. 2, 3

[BW98] BARAFF D., WITKIN A.: Large steps in cloth simulation. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), pp. 43–54. 2

[CGY*21] CHEN L., GAO L., YANG J., XU S., YE J., ZHANG X., LAI Y.-K.: Deep deformation detail synthesis for thin shell models. *arXiv preprint arXiv:2102.11541* (2021). 3

[CJS*22] CHEN X., JIANG T., SONG J., YANG J., BLACK M. J., GEIGER A., HILLIGES O.: gdna: Towards generative detailed neural avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 20427–20437. 5

[CZB*21] CHEN X., ZHENG Y., BLACK M. J., HILLIGES O., GEIGER A.: Snarf: Differentiable forward skinning for animating non-rigid neural implicit shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 11594–11604. 5

[DLJ*20] DENG B., LEWIS J. P., JERUZALSKI T., PONS-MOLL G., HINTON G., NOROUZI M., TAGLIASACCHI A.: Nasa neural articulated shape approximation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16* (2020), Springer, pp. 612–628. 5

[GBH23] GRIGOREV A., BLACK M. J., HILLIGES O.: Hood: Hierarchical graphs for generalized modelling of clothing dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 16965–16974. 3

[GCP*20] GUNDOGDU E., CONSTANTIN V., PARASHAR S., SEIFODDINI A., DANG M., SALZMANN M., FUA P.: Garnet++: Improving fast and accurate static 3d cloth draping by curvature loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence 44*, 1 (2020), 181–195. 3

[GCS*19] GUNDOGDU E., CONSTANTIN V., SEIFODDINI A., DANG M., SALZMANN M., FUA P.: Garnet: A two-stream network for fast and accurate 3d cloth draping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 8739–8748. 3

[GHF*07] GOLDENTHAL R., HARMON D., FATTAL R., BERCOVIER M., GRINSPUN E.: Efficient simulation of inextensible cloth. In *ACM SIGGRAPH 2007 papers*. 2007, pp. 49–es. 2

[GRH*12] GUAN P., REISS L., HIRSHBERG D. A., WEISS A., BLACK M. J.: Drape: Dressing any person. *ACM Transactions on Graphics (ToG) 31*, 4 (2012), 1–10. 2, 3

[HLB*23] HALIMI O., LARIONOV E., BARZELAY Z., HERHOLZ P., STUYCK T.: Physgraph: Physics-based integration using graph neural networks. *arXiv preprint arXiv:2301.11841* (2023). 3

[JGT17] JIANG C., GAST T., TERAN J.: Anisotropic elastoplasticity for cloth, knit and hair frictional contact. *ACM Transactions on Graphics (TOG) 36*, 4 (2017), 1–14. 2

[JZGF20] JIN N., ZHU Y., GENG Z., FEDKIW R.: A pixel-based framework for data-driven clothing. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 135–144. 3

[KKN*13] KIM D., KOH W., NARAIN R., FATAHALIAN K., TREUILLE A., O'BRIEN J. F.: Near-exhaustive precomputation of secondary cloth effects. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 1–8. 3

[LBK17] LIU T., BOUAZIZ S., KAVAN L.: Quasi-newton methods for real-time simulation of hyperelastic materials. *Acm Transactions on Graphics (TOG) 36*, 3 (2017), 1–16. 2

[LCF00] LEWIS J. P., CORDNER M., FONG N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), pp. 165–172. 3, 5

[LD12] LE B. H., DENG Z.: Smooth skinning decomposition with rigid bones. *ACM Transactions on Graphics (TOG) 31*, 6 (2012), 1–10. 3, 4, 6

[LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 34*, 6 (Oct. 2015), 248:1–248:16. 4

[LTT*20] LI C., TANG M., TONG R., CAI M., ZHAO J., MANOCHA D.: P-cloth: interactive complex cloth simulation on multi-gpu systems using dynamic matrix assembly and pipelined implicit integrators. *ACM Transactions on Graphics (TOG) 39*, 6 (2020), 1–15. 3

[MHHR07] MÜLLER M., HEIDELBERGER B., HENNIX M., RATCLIFF J.: Position based dynamics. *Journal of Visual Communication and Image Representation 18*, 2 (2007), 109–118. 2

[OBB20] OSMAN A. A., BOLKART T., BLACK M. J.: Star: Sparse trained articulated human body regressor. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16* (2020), Springer, pp. 598–613. 5

[PLPM20] PATEL C., LIAO Z., PONS-MOLL G.: Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 7365–7375. 2, 3, 6

[PMJ*22] PAN X., MAI J., JIANG X., TANG D., LI J., SHAO T., ZHOU K., JIN X., MANOCHA D.: Predicting loose-fitting garment deformations using bone-driven motion networks. In *ACM SIGGRAPH 2022 Conference Proceedings* (2022), pp. 1–10. 1, 2, 3, 4, 5, 6, 7, 8, 9, 11

[Pro97] PROVOT X.: Collision and self-collision handling in cloth model dedicated to design garments. In *Computer Animation and Simulation'97: Proceedings of the Eurographics Workshop in Budapest, Hungary, September 2–3, 1997* (1997), Springer, pp. 177–189. 2

[PSRC*19] PUMAROLA A., SANCHEZ-RIERA J., CHOI G., SANFELIU A., MORENO-NOGUER F.: 3dpeople: Modeling the geometry of dressed humans. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 2242–2251. 3

[SOC19] SANTESTEBAN I., OTADUY M. A., CASAS D.: Learning-based animation of clothing for virtual try-on. In *Computer Graphics Forum* (2019), vol. 38, Wiley Online Library, pp. 355–366. 3

[SOC22] SANTESTEBAN I., OTADUY M. A., CASAS D.: Snug: Self-supervised neural dynamic garments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 8140–8150. 3

[SRPMN21] SANCHEZ-RIERA J., PUMAROLA A., MORENO-NOGUER F.: Physxnet: A customizable approach for learning cloth dynamics on dressed people. In *2021 International Conference on 3D Vision (3DV)* (2021), IEEE, pp. 879–888. 3

[SSIF08] SELLE A., SU J., IRVING G., FEDKIW R.: Robust high-resolution cloth using parallelism, history-based collisions, and accurate friction. *IEEE transactions on visualization and computer graphics 15*, 2 (2008), 339–350. 2

[STOC21] SANTESTEBAN I., THUEREY N., OTADUY M. A., CASAS D.: Self-supervised collision handling via generative 3d garment models for virtual try-on. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 11763–11773. 3

[SYMB21] SAITO S., YANG J., MA Q., BLACK M. J.: Scanimate: Weakly supervised learning of skinned clothed avatar networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 2886–2897. 3, 5

[TB21] TIWARI L., BHOWMICK B.: Deepdraper: Fast and accurate 3d garment draping over a 3d human body. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 1416–1426. 3

[TB23] TIWARI L., BHOWMICK B.: Garsim: Particle based neural garment simulator. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2023), pp. 4472–4481. 3

[TPBF87] TERZOPOULOS D., PLATT J., BARR A., FLEISCHER K.: Elastically deformable models. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* (1987), pp. 205–214. 2

[TSM*20] TANCIK M., SRINIVASAN P., MILDENHALL B., FRIDOVICH-KEIL S., RAGHAVAN N., SINGHAL U., RAMAMOORTHI R., BARRON J., NG R.: Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems 33* (2020), 7537–7547. 5

[TWL*18] TANG M., WANG T., LIU Z., TONG R., MANOCHA D.: I-cloth: Incremental collision handling for gpu-based interactive cloth simulation. *ACM Transactions on Graphics (TOG) 37*, 6 (2018), 1–10. 2

[VRM*17] VAROL G., ROMERO J., MARTIN X., MAHMOOD N., BLACK M. J., LAPTEV I., SCHMID C.: Learning from synthetic humans. In *CVPR* (2017). 11

[VSC01] VASSILEV T., SPANLANG B., CHRYSANTHOU Y.: Fast cloth animation on walking avatars. In *Computer Graphics Forum* (2001), vol. 20, Wiley Online Library, pp. 260–267. 2

[VSGC20] VIDAURRE R., SANTESTEBAN I., GARCES E., CASAS D.: Fully convolutional graph neural networks for parametric virtual try-on. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 145–156. 3

[Wan21] WANG H.: Gpu-based simulation of cloth wrinkles at submillimeter levels. *ACM Transactions on Graphics (TOG) 40*, 4 (2021), 1–14. 3

[WCC*21] WU N., CHAO Q., CHEN Y., XU W., LIU C., MANOCHA D., SUN W., HAN Y., YAO X., JIN X.: Agentdress: Realtime clothing synthesis for virtual agents using plausible deformations. *IEEE Transactions on Visualization and Computer Graphics 27*, 11 (2021), 4107–4118. 3

[WP90] WILLIAMS R. J., PENG J.: An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural computation 2*, 4 (1990), 490–501. 6

[WSFM19] WANG T. Y., SHAO T., FU K., MITRA N. J.: Learning an intrinsic garment space for interactive authoring of garment animation. *ACM Transactions on Graphics (TOG) 38*, 6 (2019), 1–12. 3

[WWW22] WU B., WANG Z., WANG H.: A gpu-based multilevel additive schwarz preconditioner for cloth and deformable body simulation. *ACM Transactions on Graphics (TOG) 41*, 4 (2022), 1–14. 2

[WWYW20] WU L., WU B., YANG Y., WANG H.: A safe and fast repulsion method for gpu-based cloth self collisions. *ACM Transactions on Graphics (TOG) 40*, 1 (2020), 1–18. 2

[ZCM22] ZHANG M., CEYLAN D., MITRA N. J.: Motion guided deep dynamic 3d garments. *ACM Transactions on Graphics (TOG) 41*, 6 (2022), 1–12. 3

[Zel05] ZELLER C.: Cloth simulation on the gpu. In *ACM SIGGRAPH 2005 Sketches*. 2005, pp. 39–es. 2

[ZLH*23] ZHAO F., LI Z., HUANG S., WENG J., ZHOU T., XIE G.-S., WANG J., SHAN Y.: Learning anchor transformations for 3d garment animation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 491–500. 4

[ZWCM21] ZHANG M., WANG T. Y., CEYLAN D., MITRA N. J.: Dynamic neural garments. *ACM Transactions on Graphics (TOG) 40*, 6 (2021), 1–15. 2, 3, 6, 7