Dress Anyone: A Method & Dataset for 3D Garment Retargeting

Anonymous 3DV submission

Paper ID 386

Figure 1. Our framework can retarget 3D (loose/multi-layer) garments on non-parametric meshes in arbitrary pose & shape.

Abstract

3D garment retargeting for digital characters & avatars involves non-rigid deformation of a 3D garment mesh to plausibly fit the target body mesh in a different pose. Existing neural methods for garment simulation/draping make assumption that the 3D garment is initially fitted over the 3D body, and generally require a canonicalized representation of garments, limiting them to parametric settings. In this paper, we present a novel approach to achieve 3D garment retargeting under non-parametric settings. We propose a novel isomap-based representation to first estimate robust correspondences between garment and body mesh to achieve an initial coarse retargeting, followed by a fast and efficient neural optimization, governed by Physics-based constraints. The proposed framework enables a fast inference pipeline and quick optimization for any 3D garment. We perform extensive experiments on publicly available datasets & our new dataset of 3D clothing and report superior quantitative and qualitative results in comparison to SOTA methods, while demonstrating new capabilities.

1. Introduction

3D garment modeling for digital characters & avatars finds several applications in fashion, e-commerce, gaming, movies, and AR/VR. One such useful application is 3D Virtual Tryon, i.e. retargeting 3D digital garments on various 3D characters/avatars. Given a 3D polygonal mesh representation of a garment and a *biped* target body, the objective is to repose and deform the garment mesh to *fit* the target body mesh in a new pose while inducing pose dependent high frequency geometrical details on garment surface, in a plausible manner. This task is challenging due to the articulated nature of the target body, topological variations in garments across different categories, and the complex non-rigid deformations caused by physical interactions between the garment and body (e.g., collisions) as well as external factors (e.g. gravity).

Traditionally, Physics-Based Simulations (PBS) is used to simulate 3D garments on a body undergoing non-rigid deformations [1, 20, 32, 33]. However, PBS assumes that the initial garment mesh is already *fitted* (in the same pose) to the underlying body before modeling the physical interactions between them. Additionally, PBS-based approaches often suffer from numerical instability, incur high computational costs, are difficult to parallelize, and require manual tuning of simulation parameters [5].

Advancement in human modelling and garment digitization has led to the emergence of several learning-based solutions for 3D garment simulation [3, 5, 13, 14, 21, 39, 44]. In particular, the introduction of Parametric body models, such as SMPL [29], offers a convenient way to deal with the articulation of the human body as well as garments,

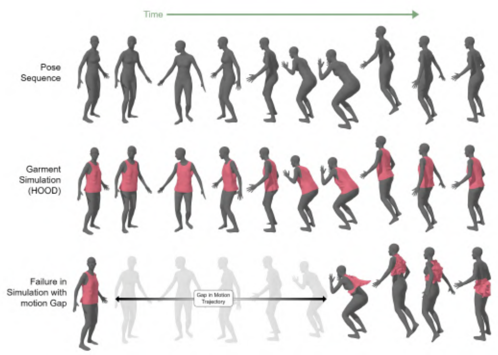

Figure 2. Neural simulation methods like HOOD[13] fail to directly retarget garments when there is gap in motion trajectory.

up to an extent. Recent developments in this direction have led to a plethora of self-supervised neural garment simulation approaches [6, 13, 14, 45]. These methods focus on modeling realistic garment deformation as the underlying body gradually changes the pose over an animated sequence. While these methods provide plausible modeling of pose-specific deformation and wrinkles by learning physics-based constraints, they require a continuous trajectory of the underlying body going from an initial pose to a final pose for training. The primary reason is that pose information from previous states of the underlying body along its trajectory is required to calculate simulation-specific parameters, such as velocity and acceleration, for the current pose. Consequently, when attempting to directly retarget the 3D garment from one arbitrary pose to another, these methods fail drastically due to the absence of motion or trajectory information between the initial garment pose and the target body pose. Figure 2 shows one such failure case of HOOD [13] in case of garment retargeting.

On the other hand, methods such as DIG [22] and DrapeNet [10] address aforementioned limitation by learning skinning weights to deform the *unposed* garment to an arbitrary pose in a self-supervised manner. However, to perform retargeting, the garment should be *unposed* (in a canonical T-pose), represented as a latent code of a large embedding space of garments (learned using supervision [10]). Recently proposed ISP [23] follows a similar approach to drape multi-layer garments, however, it assumes sewing-pattern representation of digitally created synthetic garments. Additionally, all of the aforementioned methods cannot support draping/retargeting the garment onto non-parametric human avatars or more general biped characters. Moreover, the intrinsic details of the garments (e.g. pocket, pleats, buttons etc.) are not directly preserved and are lost in the simulation.

In this work, we propose an optimization-based approach, bridging the aforementioned gaps for retargeting 3D parametric/non-parametric garments from any arbitrary pose over a target body model (parametric, non-parametric human avatar, biped characters) in a different pose, as shown in Figure 1. Given a 3D garment and a target human body as polygonal meshes, we first estimate the coarse correspondences between the two for the initial *fit*. Since existing correspondence matching methods[11, 12, 25, 28, 43] don't handle different non-rigidly deformed topologies (garment and body), we propose a novel *Isomap*-based representation, which builds upon SMPL to provide an initial non-rigid placement of the garment around the target body as a coarse retargeting initialization. We then perform a Laplacian-based detail transfer step [42] to preserve the high-fidelity geometric details (pleats, pockets, etc.) of the input garment and integrate it with the retargeted coarse garment. Finally, we employ a tiny-MLP to obtain refined pose-dependent garment deformations by efficiently optimizing for Physics-based constraints for the target pose, in a matter of seconds. Though there are several learning-based methods for generalized draping/simulation [6, 10, 22, 35, 39], all of them only work on a parametric body and need to be trained on a large collection of non-parametric target body meshes to support arbitrary out-of-distribution human avatars/scans. Our optimization-based approach provides a fast inference & quick optimization for any garments, while overcoming the aforementioned limitations. Moreover, once the tiny-MLP is optimized, it can be integrated into other differentiable pipelines (e.g. multi-view garment geometry optimization via differentiable rendering [27]). Unlike existing methods [7, 10, 22], our framework doesn't require skinning weights and, therefore, retargets any arbitrary non-parametric 3D garment on any parametric or non-parametric target body. Our key contributions are:

- We propose a novel framework for retargeting 3D garments in arbitrary pose onto parametric/non-parametric avatar, while handling loose and multilayer clothing.
- We propose a novel Isomap-based representation for estimating correspondences between 3D garment mesh and the target avatar mesh.
- We curate a new dataset "Real-3DVTON", comprising multiple 3D garments worn by different subjects in arbitrary poses, captured using a multi-view RGBD cameras. We plan a public release of the dataset & code.

2. Related Works & Background

3D Garment Simulation

Classical PBS based garment simulations [1, 20, 32, 33] yield good retargeting. However, they need a good initial mesh alignment and are typically computationally expensive and prone to numerical instability. Existing deep learning-based methods [4, 15, 35, 40] have made progress in this direction via supervised learning of skinning weights of the parametric garment for draping it onto a parametric human body. The accuracy of these methods is driven by the amount of ground truth data available for training. For mitigating this requirement, methods such as [5, 6, 37, 39, 50] adopted physics-inspired constraints for optimization, to learn a *per garment* model in a self-supervised fashion. However, skinning-based deformations fail to handle loose clothing, since they initialize skinning weights from underlying SMPL. A very recent work [7] addresses this limitation by employing an RBF-kernel over skinning weight-initialization, based on the distance of the garment from the underlying parametric body, however, it requires training for a fixed garment over a large dataset of parametric body animation sequences [31]. Other methods [13, 19, 21, 24, 45] aim towards a better generalization across different garment categories. HOOD [13] proposed a hierarchical graph-based approach extending [36] to learn skinning-free garment simulation over across different garment categories. One major criticism of the aforementioned method is the requirement of a perfect initial fitting of the 3D garment over the underlying body. Consequently, they are not suitable for retargeting garments from one arbitrary pose directly to another pose without going through intermediate body poses. Another major limitation of PBS inspired neural methods is that they only handle parametric body meshes and unlike classical PBS-based methods, do not support simulation over non-parametric meshes.

3D Garment Draping/Retargeting
The problem of 3D Garment Draping/Retargeting is different from simulation in the sense that it aims to retarget a given 3D garment in an initial static pose directly to a different final static body pose. Unlike simulation, this problem doesn't depend on the availability of intermediate dynamic pose trajectory between the initial and final pose. One naive approach to tackle this problem in parametric setting is to perform skinning of the garment using SMPL-based skinning weights [8], however, it is only applicable to extremely tight-fit clothing. Several methods have been proposed [10, 22] to address this problem by learning residual deformations over SMPL-based skinning. In particular, given a dataset of 3D garments simulated over a canonical SMPL body, DIG [22] follows an auto-decoding approach for learning the skinning weights, optimized using implicit-surface learning. Though the learned skinning weights can directly deform the garment to an arbitrary target pose, the deformations are purely statistical in nature and are not physically plausible. Drapenet [10] addresses this limitation by imposing physics-based losses while learning residual deformations over the initial SMPL-based

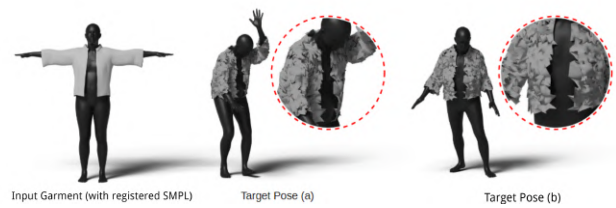

Figure 3. Coarse retargeting via nearest SMPL vertex yields noise.

skinning. For generalizing across different garment types, Drapenet [10] employs a supervised training scheme to learn a garment embedding space and then conditioning the deformation network with the garment latent vectors. However, in order to directly pose the garment to a target both the aforementioned methods require a 3D garment *unposed (in canonical T-pose)* garment perfectly fitted over a SMPL mesh. Moreover, their data-driven and parametric nature restricts them from handling arbitrary non-parametric 3D garments and target bodies. To the best of our knowledge, there is no support for 3D draping/retargeting of non-parametric 3D garments over non-parametric target human avatars/characters, from one arbitrary pose to another. Other novel view synthesis based approaches[46, 47] require multiview input data, typically captured using a sophisticated light-stage setup. In summary, there is a significant gap in literature for draping any arbitrary 3D garment from one person to another to enable 3D VTON use case.

Non-rigid Correspondence Estimation
3D garment retargeting from one pose to another can be seen as the problem of non-rigid shape deformation. More specifically, given a 3D garment and a target body, the objective is to deform the 3D garment in a plausible, non-rigid manner to fit a target body shape. In literature, methods have been proposed [25, 43] which attempt to solve this problem by first establishing a set of correspondences between topologically same source and target shapes, then using these correspondences to smoothly deform the source shape. However, in the context of 3D garment retargeting, the topology of the source shape (garment) differs significantly from the target shape (body). Another alternative is to use partial shape matching [11, 12, 28] to find correspondences across shapes of different topologies, but such methods are typically limited to partial regions of the same shape. We propose to address the non-rigid deformation between two topologically different shapes—specifically, the garment and the target body, by leveraging SMPL-based representation to establish the initial correspondences.

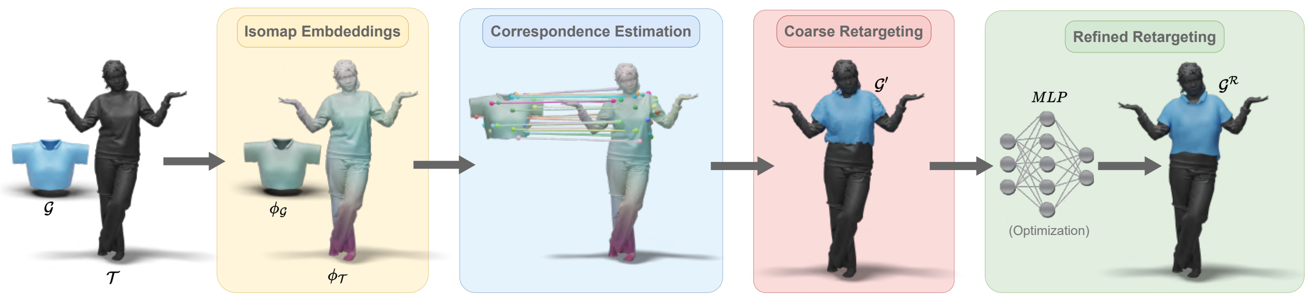

Figure 4. Outline of our proposed framework for 3D garment retargeting.

3. Methodology

Figure 4 illustrates the pipeline of the proposed approach. Given a garment mesh and a target body, we first estimate correspondences between the two using proposed isomap embeddings. These correspondences provide a crude idea of how the garment should be placed around the target. We then perform a coarse non-rigid deformation guided by these correspondences. We also perform a Laplacian-based detail preservation step to transfer the original details from the input garment to the deformed garment. Finally, we refine this coarse retargeting by optimizing the Physics-based simulation losses using a tiny MLP.

3.1. Correspondence-Guided Coarse Retargeting

This module aims to perform a coarse retargeting of the garment mesh over the target body by first establishing dense surface-level correspondences between the two. Utilizing these correspondences, we transform the garment mesh vertices to align with the target body mesh vertices. The key idea is to establish dense correspondences that can provide a *coarse* understanding of how the garment should be draped on the target body; e.g., sleeves going around the arms, the collar going around the neck etc. SMPL [29], being a parametric body model, is a natural choice for acting as a medium for establishing dense surface correspondences, as it can easily model variations in human shapes & poses. Therefore, we first perform dense non-rigid registration of both garment and target body mesh with the SMPL mesh, as shown in Figure 5. It is important to note that, unlike other methods [10, 22] which require initial garment mesh with perfectly registered SMPL mesh, our approach can deal with noise in SMPL registration as we use it only to achieve initial coarse retargeting of garments (see Figure 5(c)).

Let the garment mesh be $\mathcal{G}$, target body mesh be $\mathcal{T}$ and their corresponding SMPL meshes be $\mathcal{M}_{\mathcal{G}}$ & $\mathcal{M}_{\mathcal{T}}$, respectively. Establishing correspondences between $\mathcal{G}$ and $\mathcal{T}$ simply means for each vertex $v_i \in \mathbb{R}^3$ of $\mathcal{G}$, locating a 3D point $x_i \in \mathbb{R}^3$ on the surface of $\mathcal{T}$, where v_i should be placed. One can perform simple skinning of the garment by interpolation the skinning weights of the underlying SMPL mesh. However, that only allows re-posing the garment into various poses, not in retargeting to different subjects, and would also fail for loose garments. Alternatively, a naive way would be to find out the nearest SMPL vertex for the point on the garment and associate it with the corresponding nearest SMPL vertex to the human scan, but this approach produces a lot of local noise as an SMPL vertex can be associated to multiple garment/scan vertices (see Figure 3).

To mitigate the aforementioned issues and produce a locally smooth retargeting, we first define global features ϕ_i for each vertex q_i of the SMPL meshes $\mathcal{M}_{\mathcal{G}}$ & $\mathcal{M}_{\mathcal{T}}$. For a given SMPL mesh $\mathcal{M}$ with $\mathcal{V}_s$ number of vertices, the task is to estimate a feature vector $\phi_{smpl} = [\phi_1, \phi_2, ..., \phi_{\mathcal{V}_s}]$,$\phi_{smpl} \in \mathbb{R}^{\mathcal{V}_s \times d}$, where $\phi_i \in \mathbb{R}^d$. ϕ_{smpl} is the same for any SMPL mesh registered with any garment or body, i.e. $\phi_{smpl} = \phi_{\mathcal{M}_{\mathcal{G}}} = \phi_{\mathcal{M}_{\mathcal{T}}}$. The choice of appropriate ϕ_{smpl} must have the following essential properties. First, the feature embedding ϕ_{smpl} should incorporate both the local neighborhood information while maintaining global structural context. It should be agnostic to the position of SMPL vertices in 3D space, which means these features do not vary based on the pose or shape of SMPL. Moreover, ϕ_{smpl} should be continuous over the surface of SMPL mesh to ensure locally smooth encoding of neighborhood information. Finally, it should be concise yet representation-rich to uniquely characterize the associated surface, especially when extrapolating to the registered garment mesh or target body mesh. We experimented with existing representations such as CSE [34] and BodyMap [16] to serve the need for ϕ_{smpl}, as they promise to encode global structural information. However, we empirically found them to produce false matching due to the repetition of extrapolated features due to very low dimensionality (we provide a detailed study regarding this in the supplementary).

<u>**Isomap Embeddings**</u>
Keeping aforementioned issue in mind, we develop a new strategy to establish correspondence across different garments and human body via SMPL, leveraging the intrinsic geometry-based Isomap Embeddings [18]. We first encode

local neighborhood information by computing the pairwise geodesic distance matrix, $|\mathbb{D}_{geo}| = \mathcal{V}_s \times \mathcal{V}_s$, for all pairs of vertices (q_i, q_j) of the SMPL mesh; i.e.

$$\mathbb{D}_{geo}{}^{ij} = geodist(q_i, q_j) \tag{1}$$

To incorporate global information, we use isometric mapping to fit the vertices of SMPL mesh onto a d dimensional manifold by extending metric multi-dimensional scaling (MDS) based on $\mathbb{D}_{geo}$. This gives us a d-dimensional representation of each SMPL vertex q_i, i.e. ϕ_{smpl} Figure 5(a). We empirically found that setting d=128 ensures sufficient dimensionality to avoid repetitions while extrapolating on the target or registered mesh.

Once we have a global feature embedding ϕ_{smpl}, the feature embedding $\phi_{\mathcal{G}}$ for each vertex v_i of the garment mesh $\mathcal{G}$ is computed as follows:

$$\phi^i_{\mathcal{G}} = \frac{\sum_{j=1}^{k}[\phi^j_{\mathcal{M}_{\mathcal{G}}}/||v_i - q_j||^2]}{\sum_{j=1}^{k}[1/||v_i - q_j||^2]}; q_j \in \mathcal{N}^i \tag{2}$$

$$\mathcal{N}^i = [q_1, q_2, ..., q_k] \tag{3}$$

where, $||.||^2$ is the $\mathbb{L}_2$ distance, q_j is a vertex of the underlying SMPL mesh $\mathcal{M}_{\mathcal{G}}$ & j^{th} nearest neighbor of v_i in Euclidean space; and $|\mathcal{N}^i| = k = 32$ (set empirically). Similarly, we compute $\phi_{\mathcal{T}}$ by extrapolating $\phi_{\mathcal{M}_{\mathcal{T}}}$ based on k-nearest neighbor distance. We term these extrapolated features $\phi_{\mathcal{G}}$ and $\phi_{\mathcal{T}}$ as ***Isomap Embeddings***. These *Isomap Embedding* are common across garments and target bodies as shown in Figure 5(e) & (f).

For an arbitrary point on the garment, an initial target 3D point on the target is located via the estimated *Isomap Embedding vectors*. We first perform an initial retargeting to *coarsely* position the garment around the target body. In particular, for each vertex v_i of $\mathcal{G}$, the corresponding 3D target location x_i in the vicinity of $\mathcal{T}$ is estimated as follows:

$$x_i = \frac{\sum_{j=1}^{k}[u_j/||\phi^i_{\mathcal{G}} - \phi^j_{\mathcal{T}}||^2]}{\sum_{j=1}^{k}[1/||\phi^i_{\mathcal{G}}, \phi^j_{\mathcal{T}}||^2]}; \phi^j_{\mathcal{T}} \in \mathcal{N}^i \tag{4}$$

$$\mathcal{N}^i = [\phi^1_{\mathcal{T}}, \phi^2_{\mathcal{T}}, ..., \phi^k_{\mathcal{T}}]; \phi^j_{\mathcal{T}} \in \phi_{\mathcal{T}} \tag{5}$$

where, u_j is the vertex of target mesh $\mathcal{T}$ corresponding to $\phi^j_{\mathcal{T}}$, $\mathcal{N}^i$ the set of k-nearest neighbors of $\phi^i_{\mathcal{G}}$ in $\phi_{\mathcal{T}}$, and $|\mathcal{N}^i| = k = 32$. We replace the vertices v_i of $\mathcal{G}$ with corresponding x_i, coarsely retargeting the garment mesh around the target mesh $\mathcal{T}$.

<u>Garment Detail Preservation</u>

The coarse retargeted garments lack the original details like wrinkles, pleats, and collars. We take inspiration from [42], which relies on the Laplacian Matrix to encode the high-fidelity geometric details of the mesh. For given input garment mesh $\mathcal{G}$ with $V_{\mathcal{G}} = \{v_1, v_2, ...v_N\} \in \mathbb{R}^3$ where $\mathcal{N}$ is the total number of vertices, let $\mathcal{G}'$ be the coarsely retargeted garment mesh. For each vertex v_i let, $\mathcal{N}_i = \{j|(i,j) \in K\}$ be the neighborhood ring directly connected to v_i and degree d_i be the number of vertices in $\mathcal{N}_i$. The cotan Laplacian coordinate per vertex is given as:

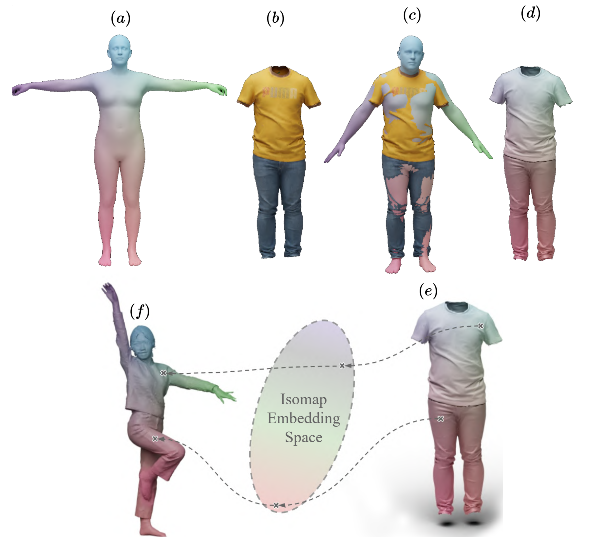

Figure 5. Isomap embedding estimation for arbitrary 3D scans: (a) SMPL mesh with per-vertex Isomap embeddings; (b) Input 3D garment(s); (c) SMPL registered with the input garment(s); (d) Isomap embeddings transferred to the input garment.

$$\delta_i(v_i) = v_i - \frac{1}{a_i}\sum_{j \in \mathcal{N}_k}(cot_{ij} + cot\beta_{ij})(v_i - v_j) \tag{6}$$

where a_i is the local area element, α and β are the opposite angles of the faces on either side of the edge ij.

In order to integrate the high-fidelity geometric details from the input garment onto retargeted garment, we first calculate the cotan Laplacian Matrix $L_{\mathcal{G}}$ and Laplacian coordinates $\delta_{\mathcal{G}}$ of the input mesh $\mathcal{G}$. For the coarsely retargeted mesh $\mathcal{G}'$, we sort the vertices based on their distance to the underlying SMPL mesh $\mathcal{M}_{\mathcal{T}}$ and choose the closest vertices as anchor points. The Laplacian matrix is recomputed as $\hat{L} = [L^T_{\mathcal{G}}, 1_i]^T$ where 1_i is the one hot encoding with i_{th} column value set to one. The Laplacian coordinates are recomputed as $\hat{\delta} = [\delta_{\mathcal{G}}, v_i]^T$ where v_i are the anchor points. By solving a linear system of equation $V^{\mathcal{G}'} = \hat{L}^{-1}\hat{\delta}$, we obtain the updated retargeted mesh $\mathcal{G}''$ with high fidelity details. Selecting only the close-body vertices as anchors, the loose garment details are also preserved from the original garment mesh (subsection 5.3).

3.2. Refined Retargeting via Optimization

The coarsely retargeted garment $\mathcal{G}''$ still lacks pose-specific deformations, e.g. the wrinkles and folds formed when a garment drops under the effect of gravity. These deformations can be obtained by Physically simulating the coarsely retargeted garment onto the static target body. While classic physics-based cloth simulations are more accurate, they are computationally expensive, difficult to parallelize and often prone to numerical instability. Neural cloth simulation methods like [6, 13] could be an alternative to classical simulations, however, they only handle SMPL. To avoid large-scale, resource-intensive training on diverse garment categories, and more importantly, due to the lack of any large, standard dataset of diverse non-parametric target meshes, we resort to an optimization-based approach for physics-guided deformation.

The coarsely retargeted garment mesh from previous step $\mathcal{G}^{''}$ with vertices $V_{\mathcal{G}^{''}}$ needs to be simulated on the static target body $\mathcal{T}$. We employ a *tiny* Multi-Layer Perceptron (MLP) network proposed in [41] to predict per-vertex deformation to simulate the garment. For each vertex $V_i^{G^{''}}$ of the refined retargeted garment $\mathcal{G}^R$, a $\Delta x_i \in \mathbb{R}^3$ is predicted by the MLP. The vertex position of the final simulated garment mesh $\mathcal{G}^{\mathcal{R}}$ is given as $v_i^{G^R} = v_i^{G^{''}} + \Delta x_i$. The predicted deformations are optimized via following constraints:

$$\begin{aligned} L_{total} = \lambda_1 L_{strain} + \lambda_2 L_{bend} + \lambda_3 L_{gravity} \\ + \lambda_4 L_{collision} + \lambda_5 L_{pin} \end{aligned} \tag{7}$$

where, L_{strain}, L_{bend} & $L_{gravity}$ are taken from [13] and we adopt Collision loss $L_{collision}$ from [26]. Pinning loss L_{pin} from [10] is used to avoid slipping of certain garment parts, e.g. straps, trouser-waist, etc., due to gravity.

4. Experimental Setup

Implementation Details: We use open-source frameworks, e.g.Trimesh[9] and Open3D[51] for implementing correspondense-guided coarse retargeting. The refined retargeting is implemented in PyTorch. We use Siren [41] as the tiny-MLP, with 3 hidden layers and 256 neurons per layer. For each garment and target body mesh pair, we optimize the refined retargeting module for $5k$ iterations, with a learning rate of $1e^{-5}$ using Adam optimizer. The optimization takes around 15-20 seconds for a garment mesh with $\sim 3.5k$ vertices on an NVIDIA RTX 4090 GPU. For Equation 7, the weights $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$ are empirically set to $\{1, 0.01, 1, 500, 1000\}$.

Public Datasets: For qualitative comparisons, we use the garments from popular datasets e.g. CLOTH3D[2] and VTO [38] datasets. For the parametric setting, we use SMPL meshes from AMASS [31] dataset. To quantitatively evaluate our approach, we take simulated 3D garments from CLOTH3D [2] as ground truth. For non-parametric setting, we use real human scans from THuman2.0 [49] and biped cartoon characters from 3DBiCar [30] dataset to demonstrate qualitative results.

Table 1. Benchmarking on Real-3DVTON dataset

Garment Type	CD ↓	P2S ↓	IR Ratio % ↓
Top	0.03660	0.14654	0.75695
Bottom	0.02368	0.11997	2.9165

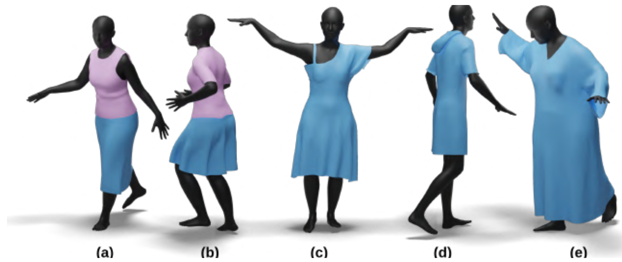

Figure 6. Our proposed framework can handle loose garments.

Real-3DVTON (Our Dataset): As stated in Sec. 1, there is a need for real-world dataset to evluate 3D garment retargeting, which contains a real 3D garment draped on real 3D human in different poses. To bridge this gap we captured real garments draped onto 15 human subjects with varied body shapes, with 44 unique garments, distributed across 255 data samples in total. For every sample, a subject is scanned in 5 different poses, wearing the same garment, using a static multi-view capture setup with 7 Azure Kinect RGBD cameras. To obtain final mesh reconstructions we employ multiview Kinect Fusion[17] on the captured RGBD data, which are then post-processed in Meshlab for noise-rectification to obtain clean, UV-parametrized garment meshes. Additionally, we perform SMPL registration for each mesh to approximate the pose & shape for future use. Our dataset captures realistic noise & topological deformations of real-world garments. We believe our dataset can prove to be extremely useful in the progress of the 3D-VTON domain. We benchmark this data with our proposed method in Table 1. Please refer to supplementary for images of our dataset and additional results of our garments retargeted to avatars from THuman 2.0 [49].

Evaluation Metrics: To quantitatively evaluate our proposed approach, we report widely used metrics like Chamfer Distance(CD), Interpenetration Ratio(IR) and Point-to-Surface Distance(P2S). Please refer to the supplementary material for more details about these metrics.

5. Results & Evaluation

5.1. Qualitative Evaluation

Qualitative Comparison: Figure 7 shows qualitative comparison of our method with Drapenet[10], where our method preserves original garment details (e.g. collar) while achieving better draping quality. Similarly, Figure 8

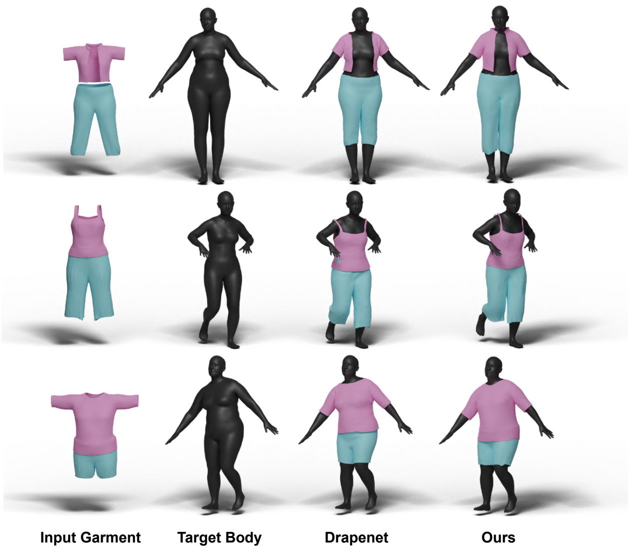

Figure 7. Qualitative comparison with DrapeNet[10]

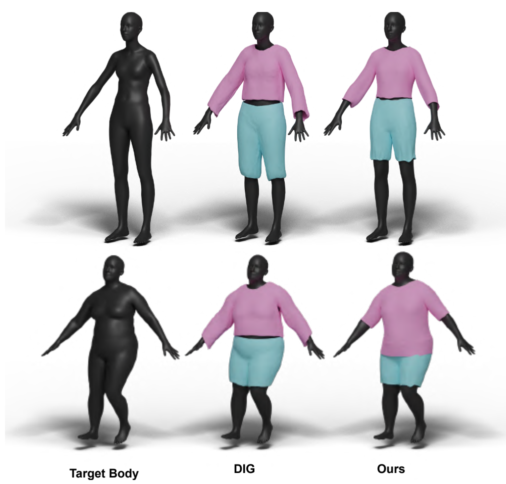

Figure 8. Qualitative comparison with DIG[22]

shows the comparison with DIG[22], where a lot of irregular deformations can be seen on the garment, while we achieve far superior retargeting. Please note that, since DIG authors do not provide inference code for arbitrary garments, we use the closest latent code to the input garment. Unlike DIG & Drapenet, we do not need to train on a large canonicalized garment dataset.

Non-Parametric 3D Garment Retargeting: Figure 1 highlights the capability of our method to retarget 3D garments onto any arbitrary parametric/non-parametric target mesh. Here, we also retarget 3D garment onto a 3D human mesh reconstructed from images (using [48, 52]) in a complex yoga pose. Figure 6 demonstrates that our method can effectively drape extremely loose garments from VTO [38] dataset really well. Figure 10 shows CLOTH3D garments

Figure 9. We are retargeting Cloth3D garment samples on 3DBiCar [30] dataset. Notice that our method handles varying shapes and variations in body proportions of the Biped characters.

Table 2. Quantitative Evaluation with Drapenet [10].

Module	Garment Type	CD ↓	P2S ↓	IR Ratio % ↓
DrapeNet	Top	0.2722	**0.0085**	**0.3752**
	Bottom	0.2897	0.0150	**1.1931**
Ours	Top	**0.00136**	0.01499	0.6857
	Bottom	**0.00054**	**0.0089**	1.7593

on real scans from THuman2.0 dataset. ***Please, refer to the supplementary for extended results***.

Multi-layered Clothing: Once we have a garment draped on a target mesh, we can treat both the target and draped garment as a single mesh, and re-compute the Isomap Embeddings following the steps in Figure 5. This allows us to retarget multilayered garments as shown in Figure 11.

Dressing Bipeds from 3DBiCar Dataset: As shown in [30], our Isomap embeddings can be adopted for retargeting garments on biped characters. We show results of Cloth3D garments draped on samples from dataset proposed in Figure 9. ***Please refer to supplementary regarding the Isomap embedding computation for 3DBiCar sample.***

5.2. Quantitative Evaluation

We report quantitative comparison with Drapenet[10] on CLOTH3D dataset in Table 2. We randomly sample 160 simulation sequences (80 topwear and 80 bottomwear). For each sequence, we randomly sample 5 timesteps (frames), resulting in 800 cloth-body paired meshes. Though P2S for

Figure 10. Results of Cloth3D garments draped on THuman 2.0 [49] human meshes in different poses and shapes.

Figure 11. Retargeting multi-layer garments on a single target.

both our method and drapenet is comparable, Chamfer Distance (CD) for Drapenet is significantly large due to its susceptibility towards outliers. We observe that for Drapenet, good initial skinning of canonical garment to the target pose is important, and any noise in skinning results in outlier vertices which contribute towards larger values of CD. Interpenetration ratio (IR) for Drapenet is lower because it computes residual deformations over initial skinning deformations (usually pointed away from the body). However, skinning suffers from the aforementioned issue and also limits applicability to loose and non-parametric garments. On the other hand, we try to empirically balance the trade-off between collision loss and plausible deformations to support non-parametric meshes. ***Please refer to supplementary for further discussion.***

5.3. Ablation Study

We discuss ablative analysis of all key component/stages in our proposed pipeline. In Table 3, we report CD, P2S and IR under the same experimental settings as quantitative evaluation. While the detail preservation step yields lower CD and P2S, it has a very high IR values compared to coarse retargeting as a side-effect of retaining original garment details. The refined retargeting achieves skinning-free, physically plausible deformations at the cost of slightly higher CD & P2S. For Laplacian-based detail transfer, Figure 12 shows the effect of using different % of garment vertices closest to the body as anchors. We use the top 20% of the closest vertices in the case of loose garments like skirts and 40% for other relatively tighter clothing.

Table 3. Ablative analysis of our pipeline.

Stages	Garment Type	CD x 10^{-3} ↓	P2S x 10^{-3} ↓	IR % ↓
Coarse	Top	1.071	16.74	1.046
	Bottom	0.574	12.67	2.110
Detail Transfer	Top	**0.902**	**13.53**	3.515
	Bottom	**0.495**	**8.950**	4.373
Refined	Top	1.360	14.99	**0.685**
	Bottom	0.547	8.992	**1.759**

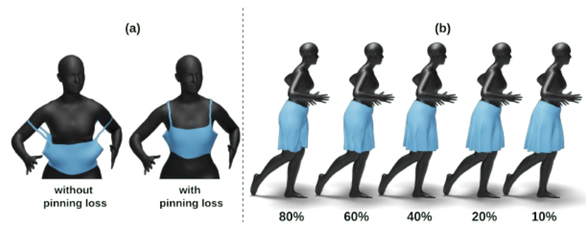

Figure 12. (a) Importance of pinning loss to avoid slipping. (b) The % of garment vertices chosen as anchors for detail transfer.

6. Conclusion

We present a novel non-parametric 3D garment retargeting method that transfers any 3D garment mesh to any target body using Isomap embeddings and SMPL for correspondence, enabling support for non-parametric meshes. Our tiny-MLP-based optimization yields physically plausible pose-specific deformations, while being fast and efficient. Though our approach is highly robust to SMPL registration noise, we wish to completely remove any dependence on parametric models in future. Secondly, we wish to improve the collission loss to mitigate the small interpenetrations due to the soft-constraint nature of the loss. We also wish to further speed up the optimization process to allow realtime, video-driven garment retargeting. We believe the proposed method acts as a crucial step towards non-parametric 3DVTON applications.

References

[1] David Baraff and Andrew Witkin. Large steps in cloth simulation. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, page 43–54, New York, NY, USA, 1998. Association for Computing Machinery. 1, 2

[2] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. CLOTH3D: Clothed 3d humans. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 6

[3] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Pbns: Physically based neural simulation for unsupervised garment pose space deformation. *ACM Trans. Graph.*, 40(6), 2021. 1

[4] Hugo Bertiche, Meysam Madadi, Emilio Tylson, and Sergio Escalera. Deepsd: Automatic deep skinning and pose space deformation for 3d garment animation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5471–5480, 2021. 3

[5] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Neural cloth simulation. *ACM Trans. Graph.*, 41(6), 2022. 1, 3

[6] Ruochen Chen, Liming Chen, and Shaifali Parashar. Gaps: Geometry-aware, physics-based, self-supervised neural garment draping. *2024 International Conference on 3D Vision (3DV)*, pages 116–125, 2023. 2, 3, 6

[7] Ruochen Chen, Liming Chen, and Shaifali Parashar. Gaps: Geometry-aware, physics-based, self-supervised neural garment draping, 2024. 2, 3

[8] Enric Corona, Albert Pumarola, Guillem Alenyà, Gerard Pons-Moll, and Francesc Moreno-Noguer. Smplicit: Topology-aware generative model for clothed people, 2021. 3

[9] Dawson-Haggerty et al. Trimesh library, 2019. 6

[10] Luca De Luigi, Ren Li, Benoît Guillard, Mathieu Salzmann, and Pascal Fua. Drapenet: Generating garments and draping them with self-supervision, 2022. 2, 3, 4, 6, 7

[11] Viktoria Ehm, Paul Roetzer, Marvin Eisenberger, Maolin Gao, Florian Bernard, and Daniel Cremers. Geometrically consistent partial shape matching, 2023. 2, 3

[12] Viktoria Ehm, Maolin Gao, Paul Roetzer, Marvin Eisenberger, Daniel Cremers, and Florian Bernard. Partial-to-partial shape matching with geometric consistency, 2024. 2, 3

[13] Artur Grigorev, Bernhard Thomaszewski, Michael J. Black, and Otmar Hilliges. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 2, 3, 6

[14] Artur Grigorev, Giorgio Becherini, Michael Black, Otmar Hilliges, and Bernhard Thomaszewski. ContourCraft: Learning to resolve intersections in neural multi-garment simulations. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. 1, 2

[15] Erhan Gundogdu, Victor Constantin, Shaifali Parashar, Amrollah Seifoddini, Minh Dang, Mathieu Salzmann, and Pascal Fua. Garnet++: Improving fast and accurate static 3d cloth draping by curvature loss. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(1):181–195, 2022. 3

[16] Anastasia Ianina, Nikolaos Sarafianos, Yuanlu Xu, Ignacio Rocco, and Tony Tung. Bodymap: Learning full-body dense correspondence map. In *CVPR*, 2022. 4

[17] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *UIST '11 Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011. 6

[18] Varun Jampani, Martin Kiefel, and Peter V. Gehler. Learning sparse high dimensional filters: Image filtering, dense crfs and bilateral neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 4

[19] Dohae Lee, Hyun Kang, and In-Kwon Lee. Clothcombo: Modeling inter-cloth interaction for draping multi-layered clothes. *ACM Transactions on Graphics (TOG)*, 42:1 – 13, 2023. 3

[20] Jie Li, Gilles Daviet, Rahul Narain, Florence Bertails-Descoubes, Matthew Overby, George E. Brown, and Laurence Boissieux. An implicit frictional contact solver for adaptive cloth simulation. *ACM Trans. Graph.*, 37(4), 2018. 1, 2

[21] Peizhuo Li, Tuanfeng Y. Wang, Timur Levent Kesdogan, Duygu Ceylan, and Olga Sorkine-Hornung. Neural garment dynamics via manifold-aware transformers. *Computer Graphics Forum (Proceedings of EUROGRAPHICS 2024)*, 43(2), 2024. 1, 3

[22] Ren Li, Benoît Guillard, Edoardo Remelli, and Pascal Fua. Dig: Draping implicit garment over the human body, 2022. 2, 3, 4, 7

[23] Ren Li, Benoît Guillard, and Pascal Fua. Isp: Multi-layered garment draping with implicit sewing patterns, 2023. 2

[24] Tianxing Li, Rui Shi, Qing Zhu, and Takashi Kanai. Swingar: Spectrum-inspired neural dynamic deformation for free-swinging garments. *IEEE transactions on visualization and computer graphics*, PP, 2023. 3

[25] Yang Li and Tatsuya Harada. Lepard: Learning partial point cloud matching in rigid and deformable scenes, 2022. 2, 3

[26] Yudi Li, Min Tang, Yun Yang, Zi Huang, Ruofeng Tong, Shuangcai Yang, Yao Li, and Dinesh Manocha. N-cloth: Predicting 3d cloth deformation with mesh-based networks, 2022. 6

[27] Yifei Li, Hsiao-yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, and Tuur Stuyck. DiffAvatar: Simulation-ready garment optimization with differentiable simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[28] O. Litany, E. Rodolà, A. M. Bronstein, and M. M. Bronstein. Fully spectral partial shape matching. *Computer Graphics Forum*, 36(2):247–258, 2017. 2, 3

[29] 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, 2015. 1, 4

[30] Zhongjin Luo, Shengcai Cai, Jinguo Dong, Ruibo Ming, Liangdong Qiu, Xiaohang Zhan, and Xiaoguang Han. Rabit: Parametric modeling of 3d biped cartoon characters with a topological-consistent dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 6, 7

[31] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pages 5442–5451, 2019. 3, 6

[32] Rahul Narain, Armin Samii, and James F O'brien. Adaptive anisotropic remeshing for cloth simulation. *ACM transactions on graphics (TOG)*, 31(6):1–10, 2012. 1, 2

[33] Andrew Nealen, Matthias Müller, Richard Keiser, Eddy Boxerman, and Mark Carlson. Physically based deformable models in computer graphics. In *Computer graphics forum*, pages 809–836. Wiley Online Library, 2006. 1, 2

[34] Natalia Neverova, David Novotný, Vasil Khalidov, Marc Szafraniec, Patrick Labatut, and Andrea Vedaldi. Continuous surface embeddings. *ArXiv*, abs/2011.12438, 2020. 4

[35] Chaitanya Patel, Zhouyingcheng Liao, and Gerard Pons-Moll. TailorNet: Predicting clothing in 3D as a function of human pose, shape and garment style. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 3

[36] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W. Battaglia. Learning mesh-based simulation with graph networks, 2021. 3

[37] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. Learning-based animation of clothing for virtual try-on. *Computer Graphics Forum*, 38, 2019. 3

[38] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. Learning-Based Animation of Clothing for Virtual Try-On. *Computer Graphics Forum (Proc. Eurographics)*, 2019. 6, 7

[39] Igor Santesteban, Miguel A Otaduy, and Dan Casas. SNUG: Self-Supervised Neural Dynamic Garments. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2, 3

[40] Yidi Shao, Chen Change Loy, and Bo Dai. Towards multi-layered 3d garments animation. In *ICCV*, 2023. 3

[41] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *arXiv*, 2020. 6

[42] Olga Sorkine, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and H-P Seidel. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 175–184, 2004. 2, 5

[43] Ramana Sundararaman, Gautam Pai, and Maks Ovsjanikov. Implicit field supervision fornbsp;robust non-rigid shape matching. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, page 344–362, Berlin, Heidelberg, 2022. Springer-Verlag. 2, 3

[44] Lokender Tiwari and Brojeshwar Bhowmick. Garsim: Particle based neural garment simulator. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4472–4481, 2023. 1

[45] Lokender Tiwari, Brojeshwar Bhowmick, and Sanjana Sinha. Gensim: Unsupervised generic garment simulator. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4169–4178, 2023. 2, 3

[46] Donglai Xiang, Fabian Prada, Timur Bagautdinov, Weipeng Xu, Yuan Dong, He Wen, Jessica Hodgins, and Chenglei Wu. Modeling clothing as a separate layer for an animatable human avatar. *ACM Transactions on Graphics*, 40(6): 1–15, 2021. 3

[47] Donglai Xiang, Timur Bagautdinov, Tuur Stuyck, Fabian Prada, Javier Romero, Weipeng Xu, Shunsuke Saito, Jingfan Guo, Breannan Smith, Takaaki Shiratori, Yaser Sheikh, Jessica Hodgins, and Chenglei Wu. Dressing avatars: Deep photorealistic appearance for physically simulated clothing. *ACM Transactions on Graphics*, 41(6):1–15, 2022. 3

[48] Yuliang Xiu, Jinlong Yang, Xu Cao, Dimitrios Tzionas, and Michael J. Black. ECON: Explicit Clothed humans Obtained from Normals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 7

[49] Tao Yu, Zerong Zheng, Kaiwen Guo, Pengpeng Liu, Qionghai Dai, and Yebin Liu. Function4d: Real-time human volumetric capture from very sparse consumer rgbd sensors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2021)*, 2021. 6, 8

[50] Meng Zhang, Duygu Ceylan, and Niloy J. Mitra. Motion guided deep dynamic 3d garments. *ACM Trans. Graph.*, 41 (6), 2022. 3

[51] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018. 6

[52] Heming Zhu, Lingteng Qiu, Yuda Qiu, and Xiaoguang Han. Registering explicit to implicit: Towards high-fidelity garment mesh reconstruction from single images, 2022. 7