PhysAvatar: Learning the Physics of Dressed 3D Avatars from Visual Observations

Yang Zheng[1⋆], Qingqing Zhao[1⋆], Guandao Yang[1], Wang Yifan[1], Donglai Xiang[2], Florian Dubost[3], Dmitry Lagun[3], Thabo Beeler[3], Federico Tombari[3,4], Leonidas Guibas[1], and Gordon Wetzstein[1]

[1] Stanford University
[2] Carnegie Mellon University
[3] Google
[4] Technical University of Munich

Abstract. Modeling and rendering photorealistic avatars is of crucial importance in many applications. Existing methods that build a 3D avatar from visual observations, however, struggle to reconstruct clothed humans. We introduce PhysAvatar, a novel framework that combines inverse rendering with inverse physics to automatically estimate the shape and appearance of a human from multi-view video data along with the physical parameters of the fabric of their clothes. For this purpose, we adopt a mesh-aligned 4D Gaussian technique for spatio-temporal mesh tracking as well as a physically based inverse renderer to estimate the intrinsic material properties. PhysAvatar integrates a physics simulator to estimate the physical parameters of the garments using gradient-based optimization in a principled manner. These novel capabilities enable PhysAvatar to create high-quality novel-view renderings of avatars dressed in loose-fitting clothes under motions and lighting conditions not seen in the training data. This marks a significant advancement towards modeling photorealistic digital humans using physically based inverse rendering with physics in the loop. Our project website is at: `https://qingqing-zhao.github.io/PhysAvatar`.

Keywords: neural rendering · physics · dynamic modeling · 3D avatar

1 Introduction

Digital avatars are a vital component in numerous applications, ranging from virtual reality and gaming to telepresence and e-commerce [26,43,101]. A realistic 3D avatar of a person can be readily obtained from visual observations, such as multi-view image [116] and video [6] data. The task of rendering animated 3D avatars from novel viewpoints or when performing unseen motions, however, presents considerable challenges, particularly when the avatar wears loose-fitting garments. Accurately rendering the dynamics of garments in conditions that are not observed in the training data necessitates a holistic approach that not only

⋆ Equal Contribution

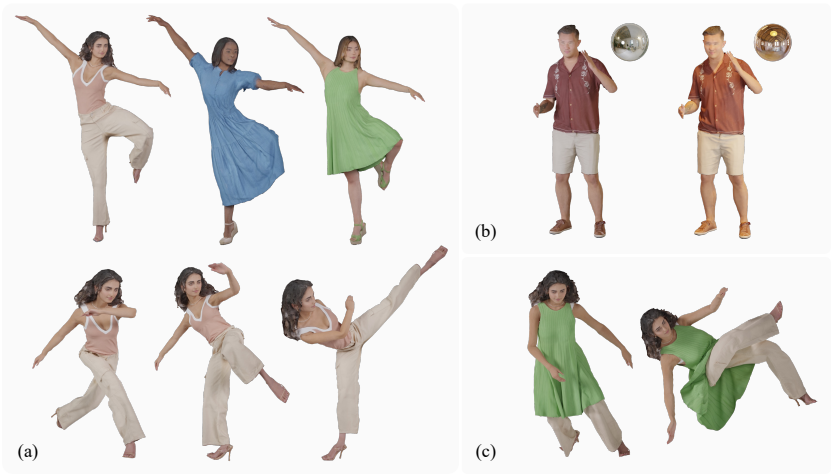

Fig. 1: PhysAvatar is a novel framework that captures the physics of dressed 3D avatars from visual observations, enabling a wide spectrum of applications, such as (a) animation, (b) relighting, and (c) redressing, with high-fidelity rendering results.

models the shape and appearance of the person but also the physical behavior of their clothes, including friction and collision.

Learning 3D scene representations from visual data is the core problem addressed by inverse rendering—a field that has recently shown remarkable progress in estimating the geometry and appearance of static and dynamic scenes from multi-view images and videos [105, 106]. In the context of reconstructing 3D avatars, the ability to explicitly control the motions of the avatar in post processing becomes imperative. Most existing methods for reconstructing animatable avatars drive the animation through an underlying skeleton using linear blending skinning (LBS) [64], which adequately models the dynamics of humans dressed in tight-fitting clothes [1,7,23,80,99,119]. With this approach, garments are treated as a rigidly attached part of the body adhering to piece-wise linear transformations, which results in motions that appear rigid and unconvincing in many cases [79, 120, 132]. Several successful attempts to introduce non-rigid deformations through pose-dependent geometry adjustments, such as wrinkles, have recently been made [18,57,59,60,114], although these methods are prone to overfitting to the body poses and motion sequences observed during training. A key problem is that most existing 3D avatar reconstruction approaches neglect to model the dynamics of loose garments in a physically accurate manner, leading to unrealistic fabric behavior and issues like self-penetration. To our knowledge, only the work of Xiang *et al*. [118] includes a physics-based approach to inverse rendering of digital humans, but their approach requires a tedious manual search process to find the parameters that model the dynamic behavior of their cloth fabrics with reasonable accuracy.

Here, we introduce PhysAvatar, a novel approach to 3D avatar reconstruction from multi-view video data. PhysAvatar combines inverse rendering with

"inverse physics" in a principled manner to not only estimate the shape and appearance of the avatar but also physical parameters, including density and stiffness, of the fabrics modeling their (loose) clothes. Our approach includes several parts: Given the reconstructed 3D mesh of the garment in one frame, we first leverage mesh-aligned 4D Gaussians [117] to track surface points of the garment across all frames of the input video sequence. This process establishes dense correspondences on the 3D garment. These data are used as supervision in an inverse physics step, where the physical parameters of the fabric are optimized using a finite-difference approach [50,123]. Finally, we employ a physically based inverse renderer [78] to jointly estimate ambient lighting and the surface material. By leveraging the refined geometry from our simulation step, the inverse renderer can effectively factor in pose-dependent effects, such as shadows caused by self-occlusion, resulting in accurate appearance reconstruction that enables the avatar to be rendered in novel lighting conditions.

PhysAvatar offers a comprehensive solution for reconstructing clothed human avatars from multi-view video data to perform novel view and motion synthesis with state-of-the-art realism. The key contributions of our work include:

1. The introduction of a new inverse rendering paradigm for avatars created from real-world captures that incorporates the physics of loose garments in a principled manner;
2. A pipeline that includes accurate and efficient mesh reconstruction and tracking using 4D Gaussians; automatic optimization of the garments' physical material properties; and accurate appearance estimation using physically based inverse rendering.

PhysAvatar demonstrates a novel path to learning physical scene properties from visual observations by incorporating inverse rendering with physics constraints. Code will be available on the project website.

2 Related Work

2.1 Scene Reconstruction from Visual Observations

Reconstructing 3D scenes from visual observations is grounded in classic approaches, such as Structure from Motion (SfM) [77, 93, 96]. Recently, this field has witnessed a paradigm shift towards inverse rendering techniques that estimate the shape and appearance of a scene from visual input [105, 106]. Among the many representations developed in this area are neural (radiance) fields [75, 95], neural volumes, surfaces, and signed distance fields [63, 73, 83, 94, 113, 122] as well as differentiable rasterization techniques for meshes [39, 61] and point clouds [40, 41, 48, 124]. Recent inverse rendering approaches are capable of modeling dynamic scenes [25, 84, 89, 108] and enable sophisticated post-capture deformations [36, 126].

2.2 Animatable Avatars

One central requirement for generating full-body digital humans is to enable animations driven by skeleton-level control [42, 71]. To this end, early works on

avatar reconstruction primarily relied on skinning techniques, which successfully model the dynamics of bodies with minimal or tight clothing [3, 64]. Clothing is an integral part of everyday human appearance, captured by the saying "clothes make the (wo)man". Reconstructing their dynamic behavior from visual data, however, remains a challenge. Several different types of representations have been explored for clothed avatars, including meshes [68] with dynamic textures [4, 33, 130], neural surface [16, 90, 107] and radiance fields [14, 18, 24, 27, 44–46, 86, 87, 99, 131], point sets [66, 67, 69], and 3D Gaussians [34, 58, 82, 132]. Many of these works condition the deformation of clothing on the body pose and some predict future dynamics based on a small number of frames [32, 33, 82].

We now discuss recent research most closely related to ours. TAVA [53] learns a deformable neural radiance field of humans or animals using a differentiable skinning method like SNARF [17]. ARAH [115] models 3D avatars using SDF representations and proposes an SDF-based volume rendering method to reconstruct the avatar in canonical space. GS-Avatar [34] achieves real-time animation by leveraging a learned 3D Gaussian predictor. Although these methods show reasonable effectiveness on humans wearing tight garments, they struggle with loose garments because their LBS-based deformation module cannot accommodate the complex dynamics of loose clothing. To address this limitation, Xiang *et al.* [118] incorporate a cloth simulator based on the eXtended Position Based Dynamics (XPBD) formulation [70] into a deep appearance model to reconstruct realistic dynamics and appearance. However, they require tedious manual adjustments of physical parameters to produce reasonable garment motion, and the code is not publicly available. Concurrent with our work, AniDress [14], while using physics-based simulation, only uses the simulation to produce a garment rigging model. This can cause inaccuracy in cloth dynamics modeling and unrealistic motion artifacts. In contrast, we estimate the physical parameters through a principled gradient-based inverse physics approach and accurately recover the shape and (relightable) appearance using a physically based inverse renderer, achieving high-fidelity results on novel motions.

2.3 Physics-Based Simulation

Physics-based simulation of cloth [5, 62, 88, 103, 109, 111, 125, 127] has been extensively studied and is particularly useful for modeling complicated dynamic effects in interaction with human bodies, such as large deformations, wrinkling, and contact. In this work, we adopt the Codimensional Incremental Potential Contact (C-IPC) solver [52] for its robustness in handling complicated body–cloth collision by the log-barrier penalty function [51]. We refer readers to dedicated surveys [37, 54] for a comprehensive review, and focus on works that solve the inverse parameter estimation problem. One line of work uses specialized devices to simultaneously measure forces and deformation in a controlled setting [47, 74, 112, 128]. Our work falls into the category that estimates physical parameters directly from imagery of garments worn on human subjects [12, 31, 97]. For this purpose, some differentiable simulators have been developed and applied to the optimization problem [55], but only for specific types of solvers,

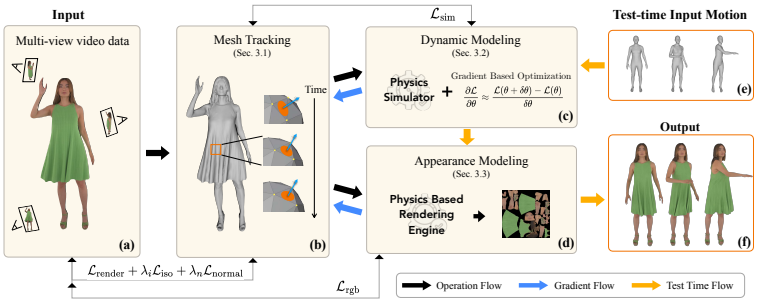

Fig. 2: Method Overview: (a) PhysAvatar takes multi-view videos and an initial mesh as input. We first perform (b) dynamic mesh tracking (Sec. 3.1). The tacked mesh sequences are then used for (c) garment physics estimation with a physics simulator combined with gradient-based optimization (Sec. 3.2); (d) and appearance estimation through physics-based differentiable rendering (Sec. 3.3). At test time, (e) given a sequence of body poses (f), we simulate garment dynamics with the learned physics parameters and employ physics-based rendering to produce the final images.

such as Projective Dynamics [56] and XPBD [98]. The heavy runtime cost of high-quality physics-based simulation motivates research in *neural cloth simulation*. Early works create datasets of garments simulated on human bodies [8,30], and then train the network to predict garment deformation in a supervised manner [19,81,91,110]. Recently, more attention has been paid to the self-supervised formulation that applies the elasticity energy defined in traditional simulation to predict garments, thereby training the network to act like a simulation solver, including both the (quasi-) static [9,11,22,49,92] and the dynamic [10,28] cases. Most of these works focus on making forward predictions that resemble traditional simulators without solving inverse parameter estimation on real-world garment data, except CaPhy [100]. CaPhy estimates the Saint Venant-Kirchhoff (StVK) elasticity parameters of clothing from 4D scans but only predicts pose-dependent clothing deformation in a quasi-static manner. In addition, all the works mentioned above treat simulation as a separate problem without modeling the photorealistic appearance. By comparison, our work builds avatars with physically based appearance and dynamics that are optimized for faithfulness to the real human capture in a holistic manner.

3 Method

Given multi-view videos as input, our goal is to reconstruct a 3D avatar that can be realistically animated and rendered under novel motions and novel views. The core pipeline (Fig. 2) consists of three primary components: 1. Mesh tracking (Sec. 3.1)—given multi-view videos and a reconstructed mesh at the initial time step as input, we track the deformation of the geometry, which provides the

geometry reference for the ensuing physics simulation; 2. Physics-based dynamic modeling (Sec. 3.2)—with the tracked meshes as reference, we estimate the garments' physical properties using a physics simulator; the simulator, together with the estimated physical parameters are used to generate novel motion; 3. Physics-based appearance refinement (Sec. 3.3)—given the tracked geometry, we further enhance the appearance estimation through a physics-based differentiable renderer, which considers self-occlusion and thus eliminate artifacts such as baked-in shadows. In the following, we delve into the details of each component.

3.1 Mesh Tracking

Given multiview videos of T frames and C views, $\{\mathbf{I}_{1:T,1:C}\}$, and an initial mesh $(\mathbf{V}_1, \mathbf{F})$, with vertices $\mathbf{V}_1$ and faces $\mathbf{F}$, that can be obtained via any existing static scene reconstruction methods, our goal is to accurately track the mesh deformation through the video sequence $\mathbf{V}_{1:T}$, which will be used as reference in the subsequent physics simulation. However accurately tracking mesh deformation from visual observation is very challenging. Inspired by the robust and real-time tracking performance from dynamic 3D Gaussians [65], we use 3D Gaussians as a surrogate representation for mesh tracking.

We adapt 3D Gaussians to align with the reconstructed mesh surface. Following the definition of dynamic 3D Gaussians, at frame t, each Gaussian is parameterized with position $\mathbf{p}_t \in \mathbb{R}^3$, rotation quaternion $\mathbf{q}_t \in \mathbb{R}^4$, color $\mathbf{c}_t \in \mathbb{R}^3$, scale $\mathbf{s}_t \in \mathbb{R}^3$ and opacity $o_t \in \mathbb{R}_+$. To couple the Gaussians with the mesh, we attach one Gaussian at the barycenter of each face $\mathbf{f} \in \mathbf{F}$, and determine the Gaussians' quaternions from the attached triangle. Specifically, the rotation of each Gaussian is computed from the rotation of its local frame, whose x- and z-axis are defined using the longest triangle edge at $t = 1$ and the face normal respectively. Furthermore, we set the last scaling factor, which corresponds to the scaling in face normal direction, to be a small constant value, ensuring the Gaussian lies on the mesh surface. The Gaussians can then be rendered in a differentiable manner [40] and we optimize their parameters by minimizing the loss [65] between the rendered images and the reference images:

$$\mathcal{L}_{\text{render}} = \lambda \left\| \mathbf{I}_{i,t} - \hat{\mathbf{I}}_{i,t} \right\|_1 + (1 - \lambda) \cdot \text{SSIM} \left(\mathbf{I}_{i,t}, \hat{\mathbf{I}}_{i,t} \right), \tag{1}$$

where $\hat{\mathbf{I}}_{i,t}$ denotes the image rendered from the perspective of the i-th camera at time t. Note that since the Gaussians' position and rotation are bound to the mesh, the optimization of the mesh vertices will implicitly optimize the parameters of the Gaussians. At the first time step, we optimize color, scale, and opacity. In the following frames $1 < t \leq T$, we fix the opacity and scale, and optimize Gaussian colors $\mathbf{c}_t$ and mesh vertices $\mathbf{V}_t$.

In addition to the photometric loss, we apply the following regularization terms to preserve the local geometric features: 1) an isometry loss $\mathcal{L}_{\text{iso}}$ [65] that preserves edge length through the training sequence, 2) a normal loss $\mathcal{L}_{\text{normal}}$ that encourages smooth mesh surfaces, enhancing mesh deformation accuracy

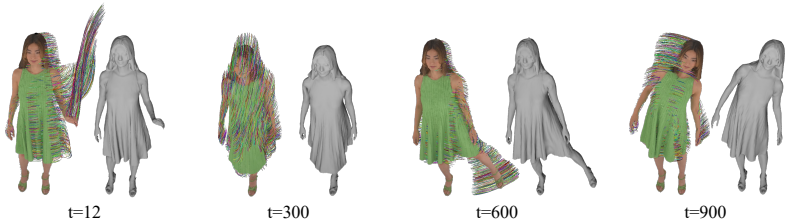

Fig. 3: Our method can robustly track a dynamic mesh from input images, providing accurate long-term correspondences. Here we show the rendered images overlaid with the Gaussian trajectories from the previous 12 frames and the optimized meshes.

and stability. Denoting the vertex and the triangle of the mesh at frame t as $\mathbf{v}_t$ and $\mathbf{f}_t$, the normal of a face as $\mathbf{n}(\cdot)$, and the 1-ring neighbors $N(\cdot)$, the losses are formally defined as:

$$\mathcal{L}_{\text{iso}} = \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \sum_{\mathbf{v}' \in N(\mathbf{v}_1)} \frac{(\|\mathbf{v}_1 - \mathbf{v}_1'\|_2 - \|\mathbf{v}_t - \mathbf{v}_t'\|_2)}{|N(\mathbf{v})|}, \tag{2}$$

$$\mathcal{L}_{\text{normal}} = \frac{1}{|\mathbf{F}|} \sum_{\mathbf{f}_t \in \mathbf{F}} \sum_{\mathbf{f}_t' \in N(\mathbf{f}_t)} (1 - \mathbf{n}(\mathbf{f}_t) \cdot \mathbf{n}(\mathbf{f}_t')). \tag{3}$$

The full loss used for mesh tracking is defined as:

$$\mathcal{L}_{\text{mesh}} = \mathcal{L}_{\text{render}} + \lambda_i \mathcal{L}_{\text{iso}} + \lambda_n \mathcal{L}_{\text{normal}}, \tag{4}$$

where we set $\lambda_i = 10$ and $\lambda_n = 0.1$ during our training. With the introduced losses, our method can reconstruct a smooth mesh sequence with accurate correspondences, as shown in Fig. 3.

3.2 Physics Based Dynamic Modeling

Given the tracked meshes, $(\mathbf{V}_{1:T}, \mathbf{F})$, we estimate the physical properties of the garments such that the simulation of their dynamics matches the reference tracked meshes. Once learned, these parameters can be used to simulate the garment under novel motions and novel views.

Garment Material Model and Simulation. We model the elastodynamics of the garment by equipping each simulation mesh with the discrete shell hinge bending energy [29,102] and isotropic StVK elastic potential [15,20]. Specifically, we estimate density ρ, membrane stiffness κ_s, and bending stiffness κ_b of the garment given the visual observations. For our generic use cases, it is reasonable to assume homogeneous physical properties for each garment piece, i.e., the density, membrane stiffness, and bending stiffness can be parameterized using scalar values, denoted as ρ, κ_s, and κ_b. In choosing the simulators, we consider differentiability and quality. Unfortunately, the development of differentiable simulators suitable for our application remains an open research question. Current implementations, such as those documented in [56], cannot be directly applied due to limitations in modeling complex body colliders. Therefore, we propose to use

the state-of-the-art robust and stable cloth simulator C-IPC [52] that guarantees no interpenetration with frictional contact. While C-IPC is differentiable, it lacks an implementation of accurate analytical gradients; hence, we use finite difference methods to estimate the gradients for parameter updates. Specifically, we employ an open-source implementation of C-IPC [52] as our cloth simulator, which is a state-of-the-art simulator achieving high-fidelity garment simulation with stability and convergence (see [52] for a comprehensive review of the base algorithm and implementation).

Garment Material Estimation. From the reconstructed mesh, we segment out the garment vertices to obtain ground truth garment mesh sequences $(\widehat{\mathbf{V}}^g_{1:T}, \widehat{\mathbf{F}}^g)$. The cloth simulator, $f(\cdot)$, takes as input the current garment state—position and velocity $(\mathbf{V}^g_t,\ \dot{\mathbf{V}}^g_t)$—the garment's physical properties $(\rho, \kappa_s, \kappa_b)$, the body collider $\{\mathbf{V}^C_t, \mathbf{F}^C\}$, and garment boundary vertices $\mathbf{V}^b_{t+1}$, and step size Δt. The boundary vertices of the garment are points of the garment that are in permanent contact with the human body, such as the shoulder region for draped garments. We employ the SMPL-X body model [64] as the body collider $\{\mathbf{V}^C_{t+1}, \mathbf{F}^C\}$, extracting SMPL-X body shapes and poses from the video data during the preprocessing (see Sec. 4). The garment and the boundary vertices are manually annotated once on the initial mesh $\mathbf{V}_1$ for the entire simulation. The simulation function predicting the garment shape in the next frame can be written as below:

$$\mathbf{V}^g_{t+1} = f(\mathbf{V}^g_t, \dot{\mathbf{V}}^g_t, \mathbf{V}^b_{t+1}, \mathbf{V}^C_{t+1}, \rho, \kappa_s, \kappa_b, \Delta t). \tag{5}$$

We find the optimal physical parameters by minimizing the difference between the simulated and the reconstructed garment meshes. Formally, the optimization objective can be expressed as:

$$\min_{\{\rho,\kappa_s,\kappa_b\}} \mathcal{L}_{\text{sim}}(\rho, \kappa_s, \kappa_b) = \sum_{t=0}^{T} \|\mathbf{V}^g_{t+1} - \widehat{\mathbf{V}}^g_{t+1}\|_2^2 \quad , \tag{6}$$

where $\mathbf{V}^g_1 = \widehat{\mathbf{V}}^g_1$ and $\mathbf{V}^g_{t+1}$ is computed by Eq. (5). We estimate gradients via a numerical approach using finite differences [50]. This method is particularly feasible in our context due to the low-dimensional nature of the optimization problem, which involves only three scalar parameters. We employ the forward difference formula for gradient estimation:

$$\frac{\partial \mathcal{L}_{\text{sim}}}{\partial \rho} \approx \left(\mathcal{L}_{\text{sim}}(\rho + \delta_\rho, \kappa_s, \kappa_b) - \mathcal{L}_{\text{sim}}(\rho, \kappa_s, \kappa_b)\right)/\delta_\rho. \tag{7}$$

Where δ_ρ is a hyperparameter for forward difference gradient estimation. The gradients for other parameters are estimated using the same approach, and the parameters are subsequently updated using the Adam optimizer.

Animation. At test time, we have the initial mesh $\{\mathbf{V}_1, \mathbf{F}_1\}$ and we are given a sequence of novel human body poses, e.g., provided by skeleton parameters such as SMPL [64]. Using standard LBS, we obtain the boundary locations $\mathbf{V}^b_{t+1}$. We assume the initial garment is at rest, i.e., $\dot{\mathbf{V}}_0 = 0$. We use the same garment

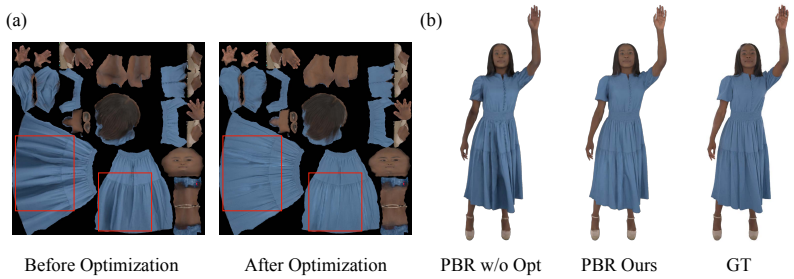

Fig. 4: Ablation study on appearance estimation: (a) Initial texture map $\mathbf{T}$ extracted from Gaussian splatting (left) has baked-in shadows highlighted in the red boxes; post-optimization (right), the baked-in shadows are substantially removed. (b) Rendering comparisons demonstrate that our method with the optimized texture map more closely aligns with the ground truth.

segmentation, boundary points, and SMPL body collider points as during training, and use Eq. (5) to simulate the deformation of the garment. This is then combined with the body, driven by LBS, to create the holistic animation.

3.3 Physics Based Appearance Modeling

While tracking with dynamic Gaussians jointly estimates the garments appearance, it does not account for light–surface interactions, consequently the reconstructed appearance contains baked-in shadows, causing uncanny visual artifacts when rendered in novel views and motions (see Fig. 4). In this step, given the mesh sequence $(\mathbf{V}_{1:T}, \mathbf{F})$ and the per-frame mesh colors $\mathbf{c}_{1:T}$ reconstructed from Section 3.1, we utilize Mitsuba3 [78]—a physics-based differentiable Monte Carlo renderer—to enhance the appearance reconstruction. For this purpose, we use a time-invariant diffuse texture map $\mathbf{T}$ to model the appearance. The value $\mathbf{T}$ is initialized from the average Gaussian colors from the tracking step (Section 3.1), $\sum_{1<t\leq T}(\mathbf{c}(\mathbf{x})_t)/T$. For the environment lighting, since the capture studio is typically equipped with uniform point light arrays, we can approximate the environment lighting as an ambient illumination, parameterized with a global $L_a \in \mathbb{R}^3$.

We optimize for $\mathbf{T}$ and L_a using the photometric loss defined in Eq. (8) for all training views and frames (i, t), where $\mathcal{R}$ denotes the rendering function and $c_{i,t}$ the camera pose:

$$\mathcal{L}_{\text{rgb}} = \min_{\{\mathbf{T}, L_a\}} \sum_{i,t} \|\mathbf{I}_{i,t} - \mathcal{R}\left(\mathbf{T}, L_a, \mathbf{V}_t, c_{i,t}\right)\|_2^2 . \tag{8}$$

At test time, we can swap the ambient light with more complex illuminations, and produce realistic rendering in arbitrary views and body poses (see Sec. 4.4).

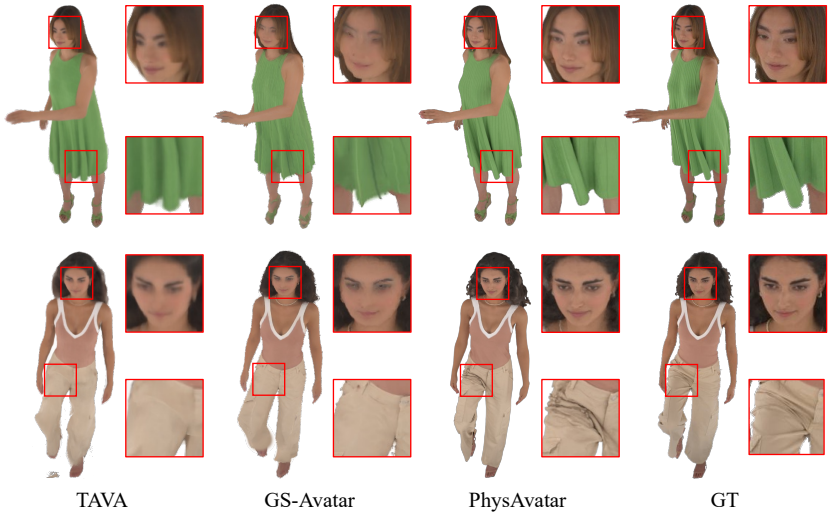

Fig. 5: Qualitative results on test poses from the ActorHQ [38] dataset. Our method PhysAvatar achieves state-of-the-art performance in terms of geometry detail and appearance modeling.

4 Experiments

In this section, we explain our experimental setup and results. We recommend watching the supplementary video for better visualization.

4.1 Experimental setup

Data Preparation. We choose four characters from the ActorHQ [35] dataset for our experiments, including two female characters with loose dresses, one female character wearing a tank top and loose-fitting pants, and one male character with a short-sleeved shirt and khaki shorts, as shown in Fig. 1. For each character, we use the ground truth mesh provided by the dataset and re-mesh it to obtain a simulation-ready initial mesh, i.e., one where the triangles have similar areas, to enhance the performance of garment simulation. We then segment the garment, define the boundary vertices of the garment for simulation purposes, and unwrap the whole mesh to get the UV map, all performed in Blender [13]. Since we observe that the human pose estimations from the dataset are quite noisy, we fit the SMPL-X [85] body mesh using multi-view videos by minimizing the distance between the projected SMPL-X joints and 2D keypoint detection results from DWPose [21, 121].

Implementation Details. For training, we utilize the entire set of 160 camera views from the ActorHQ [35] dataset. Although the dataset includes approximately 2000 frames for each character, we selectively use 24 frames with large movements for estimating garment material parameters and 200 frames for optimizing the texture map, which leads to good results while conserving compu-

Table 1: Quantitative comparison with baselines. PhysAvatar outperforms all baselines with respect to geometry accuracy. We achieve the best or competitive results in appearance metrics. See Fig. 5 and Fig. 6 for qualitative comparisons.

Method	Geometry		Appearance		
	CD ($\times 10^3$) (↓)	F-Score (↑)	LPIPS (↓)	PSNR (↑)	SSIM (↑)
ARAH [115]	1.12	86.1	0.055	28.6	<u>0.957</u>
TAVA [53]	<u>0.66</u>	<u>92.3</u>	0.051	29.6	**0.962**
GS-Avatar [34]	0.91	89.4	<u>0.044</u>	**30.6**	**0.962**
PhysAvatar (ours)	**0.55**	**92.9**	**0.035**	<u>30.2</u>	<u>0.957</u>

tational resources. For evaluation, we choose 200 frames for each character, with novel motions unseen during training. At test time, given a motion sequence defined by SMPL-X [85], we use an LBS weight inpainting algorithm [2] to obtain the skinning weights of the human body with respect to the SMPL-X skeleton. We animate the human body using LBS and simulate the garment, which is driven by the boundary vertices and utilizes the SMPL-X body mesh as the collider. More details are available at supplementary document.

Baselines. We benchmark our method and current open-sourced state-of-the-art methods modeling full-body avatars, including ARAH [115] which leverages SDF-based volume rendering to learn the canonical representation of the avatar, TAVA [53] which learns the canonical radiance field [75] and skinning weight of the character, and GS-Avatar [34] based on a pose-conditioned neural network which predicts the parameters of 3D Gaussians [40]. For all baselines, we use the official public codebase and extend them to our dense-view setting.

4.2 Comparison

We compare PhysAvatar against current state-of-the-art approaches in terms of geometry and appearance quality across test pose sequences, as shown in Table 1, Fig. 5, and Fig. 6.

Geometry Evaluation. For geometry evaluation, we employ the mean Chamfer distance [73] and the mean F-Score [104] at $\tau = 0.001$ as quantitative metrics. PhysAvatar outperforms all baseline according to these metrics, demonstrating a superior capability in accurately capturing the dynamic nature of garment geometry over time, as shown in Table 1. As shown in Fig. 5, PhysAvatar generates more accurate garment deformation and better captures fine wrinkle details, thanks to the integration of a physics-based simulator. This contrasts with all baseline methods [34, 53, 115], which do not explicitly model garment dynamics, but instead depend on learning quasi-static, pose-conditioned skinning weights or deformations. Additionally, we show results using motion data from the AMASS dataset [72]. PhysAvatar demonstrates its robustness by consistently synthesizing realistic garment deformation with fine details, while baseline methods sometimes fail to capture large deformations in loose garments (Fig. 6).

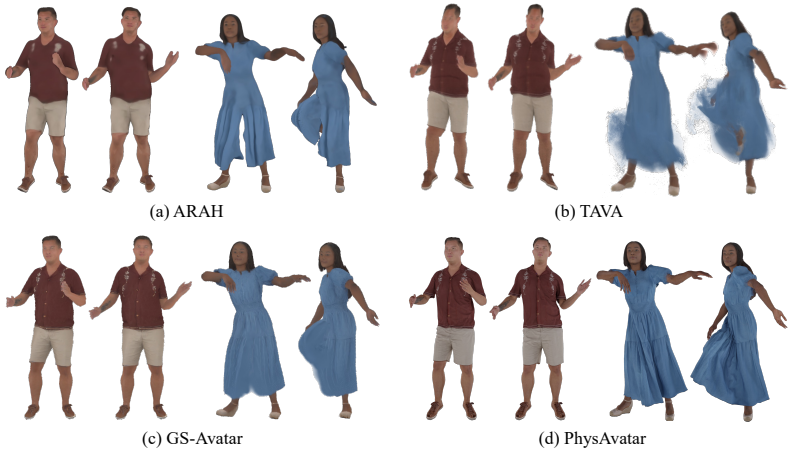

Fig. 6: Here we show animation results of current state-of-the-art methods including ARAH [115], TAVA [53] and GS-Avatar [34], and our results on test motions from AMASS [72] dataset. Images are rendered from novel views. Please zoom in for details.

Appearance Evaluation. For appearance evaluation, we utilize Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and the Learned Perceptual Image Patch Similarity (LPIPS) [129]. As detailed in Table 1, PhysAvatar surpasses all baselines in LPIPS, a metric shown to be better aligned with human perception [129]. While we achieve the second-best results in PSNR and SSIM, it is important to note that even minor misalignments in images can significantly impact these metrics, potentially not fully reflecting the quality of the renderings. As demonstrated in Fig. 5, PhysAvatar can capture more high-frequency details in facial features and garment textures, thereby creating images that are visually richer and more detailed, compared to the baseline approaches.

4.3 Ablation

We perform ablation studies to assess the impact of each component in our pipeline on the overall performance. For this, we focus on two female characters from the ActorHQ dataset, specifically chosen for their loose garments.

Garment physics estimation and modeling. We ablate our approach against two variants: one employing Linear Blend Skinning (LBS) for garment deformation and another utilizing a physics simulator for garment dynamics without optimized parameters, henceforth referred to as 'w/o physics', which employs random garment parameters. As shown in Table 2, both alternatives exhibit worse performance across all geometric metrics. Further, as shown in Fig. 7, LBS fails to produce plausible deformations for loose garments, such as dresses. While simulation with random parameters yields plausible garment dynamics, they do not accurately reflect real-world observations. In contrast, our method, with learned physics parameters, achieves deformations that closely align with the reference, demonstrating its effectiveness.

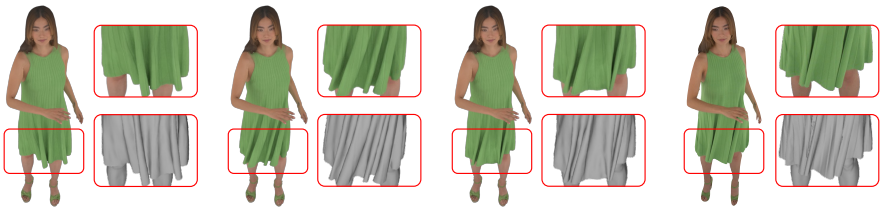

Fig. 7: Ablation Study on garment physics estimation: LBS-based garment deformation (LBS) fails to generate realistic garment deformation. Physics-based simulation with random parameters (w/o physics) achieves realistic deformation but falls short of matching the ground-truth reference (GT). In contrast, PhysAvatar accurately captures garment dynamics, closely aligning with the reference.

Method	Geometry		Appearance		
	CD ($\times 10^{-3}$) (↓)	F-Score (↑)	LPIPS (↓)	PSNR (↑)	SSIM (↑)
LBS	0.71	91.6	0.0354	<u>30.27</u>	**0.951**
w/o physics	<u>0.61</u>	<u>92.2</u>	<u>0.0347</u>	30.26	**0.951**
w/o appearance	—	—	0.0402	28.26	<u>0.947</u>
PhysAvatar (ours)	**0.56**	**93.0**	**0.0343**	**30.29**	**0.951**

Table 2: Ablation Study: PhysAvatar demonstrates improved geometric accuracy compared to both LBS-based animation (LBS) and simulations with random garment parameters (w/o physics), also shown in Fig. 7. PhysAvatar exhibits superior appearance metrics compared to the setting without physics-based appearance estimation (w/o appearance) (see Fig. 4).

Appearance estimation and modeling. We compare rendering outcomes with those obtained without appearance optimization through physics-based inverse rendering [78] (denoted as 'w/o appearance'), where the texture is directly extracted from the Gaussian framework, as detailed in section 3.3. Fig. 4 illustrates that, since the Gaussian framework does not model complex light-surface interactions, shadows are inherently baked into the texture. However, post-optimization, we note a significant removal of these baked shadows. Quantitatively, we see improvements across all appearance metrics post-optimization, shown in Table 2.

4.4 Application

PhysAvatar can facilitate a wide range of applications, such as animation, relighting, and redressing, as demonstrated in Fig. 1. Our method can also benefit from other motion databases in addition to motions defined under the regime of SMPL-X [85] or SMPL [64]. Fig. 1 shows our animation results on challenging motions from Mixamo [76], where we use its auto-rigging algorithm to drive the human body, boundary vertices of the garment, and SMPL-X mesh as the

(a) Our results are readily exportable (b) Results with artistic modifications

Fig. 8: Our results can be exported into a traditional computer graphics pipeline, e.g., Blender [13], which enables the avatars performing novel motions to be rendered in unseen environments (a) as well as editing lighting and texture of the garments (b).

collider. Current state-of-the-art approaches cannot handle such settings since they are mainly based on LBS weights from SMPL or SMPL-X. Moreover, while current methods relying on NeRF [75] or 3D Gaussians [40] representation are usually incompatible with legacy graphics pipelines, our results are readily exportable and can be seamlessly integrated with computer graphics software tools (e.g., Blender [13]), enabling post-processing artistic modifications (Fig. 8).

5 Limitations and Future Work

Our method currently depends on manual garment segmentation and mesh UV unwrapping. We are posed to evolve our pipeline towards a more automated system by incorporating neural networks to streamline these data preprocessing tasks. Another limitation of our method is the absence of a perfect underlining human body for garment simulation. Our current use of the SMPL-X [85] model as the body collider introduces potential inaccuracies in collision detection and contact information due to mismatches between the SMPL-X mesh and the actual human body shape. Our method would benefit from using more advanced parametric models that promise a closer approximation to real human body shape. Furthermore, our pipeline leverages dense view captures to model avatars, an approach that, while effective, has its limitations. Adapting our method to a sparse-view setting would enhance the versatility of its applications in environments where capturing dense views is impractical.

6 Conclusion

We introduce PhysAvatar, a comprehensive pipeline to model clothed 3D human avatars with state-of-the-art realism. By integrating a robust and efficient mesh tracking method, and a novel inverse rendering paradigm to capture the physics of loose garments and the appearance of the body, we believe our work represents a significant advancement in digital avatar modeling from visual observations, which also paves the way for innovative applications in entertainment, virtual reality, and digital fashion.

Acknowledgement

We would like to thank Jiayi Eris Zhang for the discussions. This material is based on work that is partially funded by an unrestricted gift from Google, Samsung, an SNF Postdoc Mobility fellowship, ARL grant W911NF-21-2-0104, and a Vannevar Bush Faculty Fellowship.

References

1. Abdal, R., Yifan, W., Shi, Z., Xu, Y., Po, R., Kuang, Z., Chen, Q., Yeung, D.Y., Wetzstein, G.: Gaussian shell maps for efficient 3d human generation. arXiv preprint arXiv:2311.17857 (2023)
2. Abdrashitov, R., Raichstat, K., Monsen, J., Hill, D.: Robust skin weights transfer via weight inpainting. In: SIGGRAPH Asia 2023 Technical Communications, pp. 1–4 (2023)
3. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. In: ACM SIGGRAPH 2005 Papers, pp. 408–416 (2005)
4. Bagautdinov, T., Wu, C., Simon, T., Prada, F., Shiratori, T., Wei, S.E., Xu, W., Sheikh, Y., Saragih, J.: Driving-signal aware full-body avatars. ACM Transactions on Graphics (TOG) **40**(4), 1–17 (2021)
5. Baraff, D., Witkin, A.: Large steps in cloth simulation. In: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 767–778 (2023)
6. Bashirov, R., Larionov, A., Ustinova, E., Sidorenko, M., Svitov, D., Zakharkin, I., Lempitsky, V.: Morf: Mobile realistic fullbody avatars from a monocular video. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3545–3555 (2024)
7. Bergman, A., Kellnhofer, P., Yifan, W., Chan, E., Lindell, D., Wetzstein, G.: Generative neural articulated radiance fields. Advances in Neural Information Processing Systems **35**, 19900–19916 (2022)
8. Bertiche, H., Madadi, M., Escalera, S.: Cloth3d: clothed 3d humans. In: European Conference on Computer Vision. pp. 344–359. Springer (2020)
9. Bertiche, H., Madadi, M., Escalera, S.: Pbns: physically based neural simulation for unsupervised garment pose space deformation. ACM Transactions on Graphics (TOG) **40**(6), 1–14 (2021)
10. Bertiche, H., Madadi, M., Escalera, S.: Neural cloth simulation. ACM Transactions on Graphics (TOG) **41**(6), 1–14 (2022)
11. 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. pp. 5471–5480 (2021)
12. Bhat, K.S., Twigg, C.D., Hodgins, J.K., Khosla, P.K., Popović, Z., Seitz, S.M.: Estimating cloth simulation parameters from video. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation. pp. 37–51 (2003)
13. Blender Online Coummunity: Blender, blender Foundation, https://www.blender.org/
14. Chen, B., Shen, Y., Shuai, Q., Zhou, X., Zhou, K., Zheng, Y.: Anidress: Animatable loose-dressed avatar from sparse views using garment rigging model. arXiv preprint arXiv:2401.15348 (2024)

15. Chen, H.Y., Sastry, A., van Rees, W.M., Vouga, E.: Physical simulation of environmentally induced thin shell deformation. ACM Transactions on Graphics (TOG) **37**(4), 1–13 (2018)
16. 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. pp. 20427–20437 (2022)
17. 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. pp. 11594–11604 (2021)
18. Chen, Y., Wang, X., Chen, X., Zhang, Q., Li, X., Guo, Y., Wang, J., Wang, F.: Uv volumes for real-time rendering of editable free-view human performance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16621–16631 (2023)
19. Chentanez, N., Macklin, M., Müller, M., Jeschke, S., Kim, T.Y.: Cloth and skin deformation with a triangle mesh based convolutional neural network. In: Computer Graphics Forum. vol. 39, pp. 123–134. Wiley Online Library (2020)
20. Clyde, D., Teran, J., Tamstorf, R.: Modeling and data-driven parameter estimation for woven fabrics. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. pp. 1–11 (2017)
21. Contributors, M.: Openmmlab pose estimation toolbox and benchmark. `https://github.com/open-mmlab/mmpose` (2020)
22. De Luigi, L., Li, R., Guillard, B., Salzmann, M., Fua, P.: Drapenet: Garment generation and self-supervised draping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1451–1460 (2023)
23. Dong, Z., Chen, X., Yang, J., Black, M.J., Hilliges, O., Geiger, A.: Ag3d: Learning to generate 3d avatars from 2d image collections. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14916–14927 (October 2023)
24. Feng, Y., Yang, J., Pollefeys, M., Black, M.J., Bolkart, T.: Capturing and animation of body and clothing from monocular video. In: SIGGRAPH Asia 2022 Conference Papers. pp. 1–9 (2022)
25. Fridovich-Keil, S., Meanti, G., Warburg, F.R., Recht, B., Kanazawa, A.: K-planes: Explicit radiance fields in space, time, and appearance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12479–12488 (2023)
26. Genay, A., Lécuyer, A., Hachet, M.: Being an avatar "for real": a survey on virtual embodiment in augmented reality. IEEE Transactions on Visualization and Computer Graphics **28**(12), 5071–5090 (2021)
27. Geng, C., Peng, S., Xu, Z., Bao, H., Zhou, X.: Learning neural volumetric representations of dynamic humans in minutes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8759–8770 (2023)
28. 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. pp. 16965–16974 (2023)
29. Grinspun, E., Hirani, A.N., Desbrun, M., Schröder, P.: Discrete shells. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation. pp. 62–67. Citeseer (2003)
30. 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. pp. 8739–8748 (2019)
31. Guo, J., Li, J., Narain, R., Park, H.S.: Inverse simulation: Reconstructing dynamic geometry of clothed humans via optimal control. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14698–14707 (2021)
32. Habermann, M., Liu, L., Xu, W., Pons-Moll, G., Zollhoefer, M., Theobalt, C.: Hdhumans: A hybrid approach for high-fidelity digital humans. Proceedings of the ACM on Computer Graphics and Interactive Techniques **6**(3), 1–23 (2023)
33. Habermann, M., Liu, L., Xu, W., Zollhoefer, M., Pons-Moll, G., Theobalt, C.: Real-time deep dynamic characters. ACM Transactions on Graphics (ToG) **40**(4), 1–16 (2021)
34. Hu, L., Zhang, H., Zhang, Y., Zhou, B., Liu, B., Zhang, S., Nie, L.: Gaussianavatar: Towards realistic human avatar modeling from a single video via animatable 3d gaussians. arXiv preprint arXiv:2312.02134 (2023)
35. Işık, M., Rünz, M., Georgopoulos, M., Khakhulin, T., Starck, J., Agapito, L., Nießner, M.: Humanrf: High-fidelity neural radiance fields for humans in motion. ACM Transactions on Graphics (TOG) **42**(4), 1–12 (2023). https://doi.org/10.1145/3592415, https://doi.org/10.1145/3592415
36. Jambon, C., Kerbl, B., Kopanas, G., Diolatzis, S., Leimkühler, T., Drettakis, G.: Nerfshop: Interactive editing of neural radiance fields. Proc. ACM Comput. Graph. Interact. Tech. **6**, 1:1–1:21 (2023). https://doi.org/10.1145/3585499
37. Jiang, Y., Wang, R., Liu, Z.: A survey of cloth simulation and applications. In: 2008 9th International Conference on Computer-Aided Industrial Design and Conceptual Design. pp. 765–769. IEEE (2008)
38. Jiang, Y., Shen, Z., Wang, P., Su, Z., Hong, Y., Zhang, Y., Yu, J., Xu, L.: Hifi4g: High-fidelity human performance rendering via compact gaussian splatting. arXiv preprint arXiv:2312.03461 (2023)
39. Kato, H., Ushiku, Y., Harada, T.: Neural 3d mesh renderer. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 3907–3916 (2017). https://doi.org/10.1109/CVPR.2018.00411
40. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (2023)
41. Keselman, L., Hebert, M.: Approximate differentiable rendering with algebraic surfaces. In: European Conference on Computer Vision. pp. 596–614. Springer (2022)
42. Komatsu, K.: Human skin model capable of natural shape variation. The visual computer **3**, 265–271 (1988)
43. Korban, M., Li, X.: A survey on applications of digital human avatars toward virtual co-presence. arXiv preprint arXiv:2201.04168 (2022)
44. Kwon, Y., Kim, D., Ceylan, D., Fuchs, H.: Neural human performer: Learning generalizable radiance fields for human performance rendering. Advances in Neural Information Processing Systems **34**, 24741–24752 (2021)
45. Kwon, Y., Kim, D., Ceylan, D., Fuchs, H.: Neural image-based avatars: Generalizable radiance fields for human avatar modeling. In: The Eleventh International Conference on Learning Representations (2023)
46. Kwon, Y., Liu, L., Fuchs, H., Habermann, M., Theobalt, C.: Deliffas: Deformable light fields for fast avatar synthesis. Advances in Neural Information Processing Systems **36** (2023)

47. Larionov, E., Eckert, M.L., Wolff, K., Stuyck, T.: Estimating cloth elasticity parameters using position-based simulation of compliant constrained dynamics. arXiv preprint arXiv:2212.08790 (2022)
48. Lassner, C., Zollhofer, M.: Pulsar: Efficient sphere-based neural rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1440–1449 (2021)
49. Lee, D., Kang, H., Lee, I.K.: Clothcombo: Modeling inter-cloth interaction for draping multi-layered clothes. ACM Transactions on Graphics (TOG) **42**(6), 1–13 (2023)
50. LeVeque, R.J.: Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. SIAM (2007)
51. Li, M., Ferguson, Z., Schneider, T., Langlois, T.R., Zorin, D., Panozzo, D., Jiang, C., Kaufman, D.M.: Incremental potential contact: intersection-and inversion-free, large-deformation dynamics. ACM Trans. Graph. **39**(4), 49 (2020)
52. Li, M., Kaufman, D.M., Jiang, C.: Codimensional incremental potential contact. ACM Transactions on Graphics (TOG) **40**(4), 1–24 (2021)
53. Li, R., Tanke, J., Vo, M., Zollhofer, M., Gall, J., Kanazawa, A., Lassner, C.: Tava: Template-free animatable volumetric actors (2022)
54. Li, X., Li, X.R., Li, Y., Feng, W.: Review of cloth modeling and simulation for virtual fitting. Textile Research Journal **93**(7-8), 1699–1711 (2023)
55. Li, Y., Chen, H.y., Larionov, E., Sarafianos, N., Matusik, W., Stuyck, T.: Diffavatar: Simulation-ready garment optimization with differentiable simulation. arXiv preprint arXiv:2311.12194 (2023)
56. Li, Y., Du, T., Wu, K., Xu, J., Matusik, W.: Diffcloth: Differentiable cloth simulation with dry frictional contact. ACM Transactions on Graphics (TOG) **42**(1), 1–20 (2022)
57. Li, Z., Zheng, Z., Liu, Y., Zhou, B., Liu, Y.: Posevocab: Learning joint-structured pose embeddings for human avatar modeling. arXiv preprint arXiv:2304.13006 (2023)
58. Li, Z., Zheng, Z., Wang, L., Liu, Y.: Animatable gaussians: Learning pose-dependent gaussian maps for high-fidelity human avatar modeling. arXiv preprint arXiv:2311.16096 (2023)
59. Liu, L., Habermann, M., Rudnev, V., Sarkar, K., Gu, J., Theobalt, C.: Neural actor: Neural free-view synthesis of human actors with pose control. ACM transactions on graphics (TOG) **40**(6), 1–16 (2021)
60. Liu, L., Xu, W., Zollhoefer, M., Kim, H., Bernard, F., Habermann, M., Wang, W., Theobalt, C.: Neural rendering and reenactment of human actor videos. ACM Transactions on Graphics (TOG) **38**(5), 1–14 (2019)
61. Liu, S., Li, T., Chen, W., Li, H.: Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7708–7717 (2019)
62. Liu, T., Bouaziz, S., Kavan, L.: Quasi-newton methods for real-time simulation of hyperelastic materials. Acm Transactions on Graphics (TOG) **36**(3), 1–16 (2017)
63. Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., Sheikh, Y.: Neural volumes: Learning dynamic renderable volumes from images. arXiv preprint arXiv:1906.07751 (2019)
64. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. ACM Transactions on Graphics **34**(6) (2015)
65. Luiten, J., Kopanas, G., Leibe, B., Ramanan, D.: Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. arXiv preprint arXiv:2308.09713 (2023)

66. Ma, Q., Saito, S., Yang, J., Tang, S., Black, M.J.: Scale: Modeling clothed humans with a surface codec of articulated local elements. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16082–16093 (2021)
67. Ma, Q., Yang, J., Black, M.J., Tang, S.: Neural point-based shape modeling of humans in challenging clothing. In: 2022 International Conference on 3D Vision (3DV). pp. 679–689. IEEE (2022)
68. Ma, Q., Yang, J., Ranjan, A., Pujades, S., Pons-Moll, G., Tang, S., Black, M.J.: Learning to dress 3d people in generative clothing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6469–6478 (2020)
69. Ma, Q., Yang, J., Tang, S., Black, M.J.: The power of points for modeling humans in clothing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10974–10984 (2021)
70. Macklin, M., Müller, M., Chentanez, N.: Xpbd: position-based simulation of compliant constrained dynamics. In: Proceedings of the 9th International Conference on Motion in Games. p. 49–54. MIG '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2994258.2994272, https://doi.org/10.1145/2994258.2994272
71. Magnenat, T., Laperrière, R., Thalmann, D.: Joint-dependent local deformations for hand animation and object grasping. Tech. rep., Canadian Inf. Process. Soc (1988)
72. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G., Black, M.J.: AMASS: Archive of motion capture as surface shapes. In: ICCV (2019)
73. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4460–4470 (2019)
74. Miguel, E., Bradley, D., Thomaszewski, B., Bickel, B., Matusik, W., Otaduy, M.A., Marschner, S.: Data-driven estimation of cloth simulation models. In: Computer Graphics Forum. vol. 31, pp. 519–528. Wiley Online Library (2012)
75. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
76. Mixamo, www.mixamo.com
77. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE international symposium on mixed and augmented reality. pp. 127–136. Ieee (2011)
78. Nimier-David, M., Vicini, D., Zeltner, T., Jakob, W.: Mitsuba 2: A retargetable forward and inverse renderer. ACM Transactions on Graphics (TOG) **38**(6), 1–17 (2019)
79. Noguchi, A., Sun, X., Lin, S., Harada, T.: Neural articulated radiance field. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5762–5772 (October 2021)
80. Noguchi, A., Sun, X., Lin, S., Harada, T.: Unsupervised learning of efficient geometry-aware neural articulated representations. In: European Conference on Computer Vision. pp. 597–614. Springer (2022)
81. 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. pp. 1–10 (2022)

82. Pang, H., Zhu, H., Kortylewski, A., Theobalt, C., Habermann, M.: Ash: Animatable gaussian splats for efficient and photoreal human rendering. arXiv preprint arXiv:2312.05941 (2023)
83. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 165–174 (2019)
84. Park, K., Sinha, U., Hedman, P., Barron, J.T., Bouaziz, S., Goldman, D.B., Martin-Brualla, R., Seitz, S.M.: Hypernerf: a higher-dimensional representation for topologically varying neural radiance fields. ACM Transactions on Graphics (TOG) **40**(6), 1–12 (2021)
85. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3D hands, face, and body from a single image. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 10975–10985 (2019)
86. Peng, S., Dong, J., Wang, Q., Zhang, S., Shuai, Q., Zhou, X., Bao, H.: Animatable neural radiance fields for modeling dynamic human bodies. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14314–14323 (2021)
87. Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., Zhou, X.: Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9054–9063 (2021)
88. Peng, T., Wu, W., Liu, J., Li, L., Miao, J., Hu, X., He, R., Li, L.: Pgn-cloth: Physics-based graph network model for 3d cloth animation. Displays **80**, 102534 (2023)
89. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-nerf: Neural radiance fields for dynamic scenes. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 10313–10322 (2020). https://doi.org/10.1109/CVPR46437.2021.01018
90. 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. pp. 2886–2897 (2021)
91. Santesteban, I., Otaduy, M.A., Casas, D.: Learning-based animation of clothing for virtual try-on. In: Computer Graphics Forum. vol. 38, pp. 355–366. Wiley Online Library (2019)
92. 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. pp. 8140–8150 (2022)
93. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016)
94. Sitzmann, V., Thies, J., Heide, F., Nießner, M., Wetzstein, G., Zollhöfer, M.: Deepvoxels: Learning persistent 3d feature embeddings. In: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE (2019)
95. Sitzmann, V., Zollhöfer, M., Wetzstein, G.: Scene representation networks: Continuous 3d-structure-aware neural scene representations. In: Advances in Neural Information Processing Systems (2019)
96. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. In: ACM siggraph 2006 papers, pp. 835–846 (2006)

97. Stoll, C., Gall, J., De Aguiar, E., Thrun, S., Theobalt, C.: Video-based reconstruction of animatable human characters. ACM Transactions on Graphics (TOG) **29**(6), 1–10 (2010)
98. Stuyck, T., Chen, H.y.: Diffxpbd: Differentiable position-based simulation of compliant constraint dynamics. Proceedings of the ACM on Computer Graphics and Interactive Techniques **6**(3), 1–14 (2023)
99. Su, S.Y., Yu, F., Zollhöfer, M., Rhodin, H.: A-nerf: Articulated neural radiance fields for learning human shape, appearance, and pose. Advances in Neural Information Processing Systems **34**, 12278–12291 (2021)
100. Su, Z., Hu, L., Lin, S., Zhang, H., Zhang, S., Thies, J., Liu, Y.: Caphy: Capturing physical properties for animatable human avatars. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14150–14160 (2023)
101. Sun, M., Yang, D., Kou, D., Jiang, Y., Shan, W., Yan, Z., Zhang, L.: Human 3d avatar modeling with implicit neural representation: A brief survey. In: 2022 14th International Conference on Signal Processing Systems (ICSPS). pp. 818–827. IEEE (2022)
102. Tamstorf, R., Grinspun, E.: Discrete bending forces and their jacobians. Graphical models **75**(6), 362–370 (2013)
103. Tang, M., Tong, R., Narain, R., Meng, C., Manocha, D.: A gpu-based streaming algorithm for high-resolution cloth simulation. In: Computer Graphics Forum. vol. 32, pp. 21–30. Wiley Online Library (2013)
104. Tatarchenko, M., Richter, S.R., Ranftl, R., Li, Z., Koltun, V., Brox, T.: What do single-view 3d reconstruction networks learn? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3405–3414 (2019)
105. Tewari, A., Fried, O., Thies, J., Sitzmann, V., Lombardi, S., Sunkavalli, K., Martin-Brualla, R., Simon, T., Saragih, J.M., Nießner, M., Pandey, R., Fanello, S., Wetzstein, G., Zhu, J.Y., Theobalt, C., Agrawala, M., Shechtman, E., Goldman, D.B., Zollhofer, M.: State of the art on neural rendering. Computer Graphics Forum **39** (2020). https://doi.org/10.1111/cgf.14022
106. Tewari, A., Thies, J., Mildenhall, B., Srinivasan, P., Tretschk, E., Yifan, W., Lassner, C., Sitzmann, V., Martin-Brualla, R., Lombardi, S., et al.: Advances in neural rendering. In: Computer Graphics Forum. vol. 41, pp. 703–735. Wiley Online Library (2022)
107. Tiwari, G., Sarafianos, N., Tung, T., Pons-Moll, G.: Neural-gif: Neural generalized implicit functions for animating people in clothing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11708–11718 (2021)
108. Tretschk, E., Tewari, A., Golyanik, V., Zollhöfer, M., Lassner, C., Theobalt, C.: Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 12939–12950 (2020). https://doi.org/10.1109/ICCV48922.2021.01272
109. Vassilev, T., Spanlang, B., Chrysanthou, Y.: Fast cloth animation on walking avatars. In: Computer Graphics Forum. vol. 20, pp. 260–267. Wiley Online Library (2001)
110. Vidaurre, R., Santesteban, I., Garces, E., Casas, D.: Fully convolutional graph neural networks for parametric virtual try-on. In: Computer Graphics Forum. vol. 39, pp. 145–156. Wiley Online Library (2020)
111. Wang, H.: Gpu-based simulation of cloth wrinkles at submillimeter levels. ACM Transactions on Graphics (TOG) **40**(4), 1–14 (2021)

112. Wang, H., O'Brien, J.F., Ramamoorthi, R.: Data-driven elastic models for cloth: modeling and measurement. ACM transactions on graphics (TOG) **30**(4), 1–12 (2011)
113. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. Advances in Neural Information Processing Systems **34**, 27171–27183 (2021)
114. Wang, S., Schwarz, K., Geiger, A., Tang, S.: Arah: Animatable volume rendering of articulated human sdfs. In: European conference on computer vision. pp. 1–19. Springer (2022)
115. Wang, S., Schwarz, K., Geiger, A., Tang, S.: Arah: Animatable volume rendering of articulated human sdfs. In: European Conference on Computer Vision (2022)
116. Wang, T., Zhang, B., Zhang, T., Gu, S., Bao, J., Baltrusaitis, T., Shen, J., Chen, D., Wen, F., Chen, Q., et al.: Rodin: A generative model for sculpting 3d digital avatars using diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4563–4573 (2023)
117. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4d gaussian splatting for real-time dynamic scene rendering. ArXiv **abs/2310.08528** (2023). https://doi.org/10.48550/arXiv.2310.08528
118. Xiang, D., Bagautdinov, T., Stuyck, T., Prada, F., Romero, J., Xu, W., Saito, S., Guo, J., Smith, B., Shiratori, T., et al.: Dressing avatars: Deep photorealistic appearance for physically simulated clothing. ACM Transactions on Graphics (TOG) **41**(6), 1–15 (2022)
119. Xu, Y., Yifan, W., Bergman, A.W., Chai, M., Zhou, B., Wetzstein, G.: Efficient 3d articulated human generation with layered surface volumes. arXiv preprint arXiv:2307.05462 (2023)
120. Xu, Y., Chen, B., Li, Z., Zhang, H., Wang, L., Zheng, Z., Liu, Y.: Gaussian head avatar: Ultra high-fidelity head avatar via dynamic gaussians. arXiv preprint arXiv:2312.03029 (2023)
121. Yang, Z., Zeng, A., Yuan, C., Li, Y.: Effective whole-body pose estimation with two-stages distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4210–4220 (2023)
122. Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. Advances in Neural Information Processing Systems **34**, 4805–4815 (2021)
123. Yee, K.S., Chen, J.S.: The finite-difference time-domain (fdtd) and the finite-volume time-domain (fvtd) methods in solving maxwell's equations. IEEE Transactions on Antennas and Propagation **45**(3), 354–363 (1997)
124. Yifan, W., Serena, F., Wu, S., Öztireli, C., Sorkine-Hornung, O.: Differentiable surface splatting for point-based geometry processing. ACM Transactions on Graphics (TOG) **38**(6), 1–14 (2019)
125. Yu, T., Zheng, Z., Zhong, Y., Zhao, J., Dai, Q., Pons-Moll, G., Liu, Y.: Simulcap: Single-view human performance capture with cloth simulation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5504–5514 (2019)
126. Yuan, Y.J., Sun, Y.T., Lai, Y.K., Ma, Y., Jia, R., Gao, L.: Nerf-editing: geometry editing of neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18353–18364 (2022)
127. Zeller, C.: Cloth simulation on the gpu. In: ACM SIGGRAPH 2005 Sketches, pp. 39–es (2005)
128. Zhang, J.X., Lin, G.W.C., Bode, L., Chen, H.y., Stuyck, T., Larionov, E.: Estimating cloth elasticity parameters from homogenized yarn-level models. arXiv preprint arXiv:2401.15169 (2024)

129. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
130. Zhao, F., Jiang, Y., Yao, K., Zhang, J., Wang, L., Dai, H., Zhong, Y., Zhang, Y., Wu, M., Xu, L., et al.: Human performance modeling and rendering via neural animated mesh. ACM Transactions on Graphics (TOG) **41**(6), 1–17 (2022)
131. Zheng, Z., Huang, H., Yu, T., Zhang, H., Guo, Y., Liu, Y.: Structured local radiance fields for human avatar modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15893–15903 (2022)
132. Zielonka, W., Bagautdinov, T., Saito, S., Zollhöfer, M., Thies, J., Romero, J.: Drivable 3d gaussian avatars. arXiv preprint arXiv:2311.08581 (2023)