This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

PGC: Physics-Based Gaussian Cloth from a Single Pose

Michelle Guo*[1] Matt Jen-Yuan Chiang[2] Igor Santesteban[2] Nikolaos Sarafianos[2]
Hsiao-yu Chen[2] Oshri Halimi[2] Aljaž Božič[2] Shunsuke Saito[2] Jiajun Wu[1] C. Karen Liu[1]
Tuur Stuyck[2] Egor Larionov[2]
[1]Stanford University [2]Meta Reality Labs



Figure 1. We present a method to recover photorealistic, simulation-ready garments from a multi-view capture of a *single* static pose. The recovered garments consist of simulatable geometry and fine-detail appearance. Moreover, our results generalize to novel motion as they can be simulated on human motion sequences, and our garments are relightable.

Abstract

We introduce a novel approach to reconstruct simulation-ready garments with intricate appearance. Despite recent advancements, existing methods often struggle to balance the need for accurate garment reconstruction with the ability to generalize to new poses and body shapes or require large amounts of data to achieve this. In contrast, our method only requires a multi-view capture of a single static frame. We represent garments as hybrid mesh-embedded 3D Gaussian splats, where the Gaussians capture near-field shading and high-frequency details, while the mesh encodes far-field albedo and optimized reflectance parameters. We achieve novel pose generalization by exploiting the mesh from our hybrid approach, enabling physics-based simulation and surface rendering techniques, while also capturing fine details with Gaussians that accurately reconstruct garment details. Our optimized garments can be used for simulating garments on novel poses, and garment relighting. Project page: phys-gaussian-cloth.github.io.

1. Introduction

Clothing plays a vital role in how we present ourselves to the world, and it is essential to accurately recreate and represent a person's attire when creating virtual avatars. The automatic reconstruction of animatable garments is a vital research area, and it's no wonder that the field has witnessed a surge in interest in garment reconstruction techniques. Over the years, many advances have been made to automatically recreate them from scans [23, 29, 71], text-guidance [16, 48], monocular videos [27, 42], RGB-D sensors [63], or even images [25, 26, 45, 62].

Traditionally, garment reconstruction methods have focused on mesh-based representations [2, 29, 58] due to their natural fit for simulation purposes [49]. Physics-based simulation models allow one to generate natural and physically plausible motions under a variety of body movements. However, mesh-based representations are limited to the mesh resolution and often struggle to capture geometric details such as belts, pockets, and zippers or fabrics with significant thickness and fuzziness such as knits or furs. On the other hand, recent development of 3D Gaussian splat (3DGS) reconstruction [21] applied to clothed avatars [30, 31, 44] have shown that they are capable of capturing such detailed features resulting in an appearance that is true to the real clothing. Despite significant progress in point-based simulation models for point-based representations of volumetric objects [8, 19, 59, 70], point-based cloth simulation is considered computationally expensive [64] but largely under-explored. In addition, existing works tend to require multi-frame tracking [44, 71]. A concurrent

*This work was conducted during an internship at Meta Reality Labs

work called Gaussian Garments [44] also proposes a hybrid representation. Gaussian Garments achieves this by leveraging video sequences of multi-view captures. They use StyleUNet to predict color shifts and lighting effects. In order to train this module, they require tracked sequences of garments in motion. However, tracking highly deformable and dynamic cloth surfaces is computationally expensive and remains a challenging problem. Inaccuracy can lead to blurring of the appearance in the final result. In contrast, we propose a novel appearance model that reconstructs the appearance from a single time instance only, eliminating the need for temporally tracking deformable surfaces, leading to a less cumbersome and more efficient method.

The goal of this work is to reconstruct the appearance of a garment from a multi-view capture of a *single frame*. A natural approach to this problem is to leverage the recent work on 3D Gaussian splat (3DGS) scene reconstruction [21], which uses a set of 3D Gaussian splats rendered from multiple views to fit a set of captures. Each 3D Gaussian splat is essentially a point with a local coordinate frame and associated view-dependent appearance properties rendered with a soft falloff modelled using a Gaussian kernel. With this approach we capture the fine volumetric details on the surface of fabrics like pilling and fly-aways that give many materials their characteristic look.

A limitation of the original approach [21] is that splats are not constrained to specific locations in space, which can lead to severe visually jarring artifacts when their positions or rotations are perturbed. To alleviate this issue, recent works employed a mesh-embedded splat representation [14, 33, 40, 44] that anchors splats on a triangle mesh to better control their position and scale, which allows them to follow the deforming mesh more faithfully. We adopt this approach, recognizing the advantage of this representation as it allows for physics-based deformations of the underlying mesh using a simulation pipeline.

Another difficulty with Gaussian splatting is that splats overfit to the scene appearance, baking in shading effects like lighting conditions, specular highlights, and self-shadowing. This limits their use beyond mere reconstruction. We propose an approach to address this limitation in garment assets generated from multi-view single pose garment captures. Instead of relying solely on Gaussian splat optimization to reconstruct the static scene, we exploit the strengths of traditional physically-based rendering (PBR) techniques to reproduce far-field pose-dependent shading, and use 3DGS reconstruction to fill in the missing pose-independent details such that the final result remains faithful in appearance to the original garment. This decomposition of responsibilities based on representation fits well with frequency-based decomposition of signals.

To generate garments in novel poses, we employ a physics-based cloth simulator to generate a deformed garment mesh. Using the deformed mesh, our method combines the high frequency information of the reconstructed mesh-embedded 3DGS with the low frequency information obtained from the mesh-based representation, which is obtained through physics-based rendering of the mesh with its reconstructed texture map and reflectance parameters. To summarize, we propose the following contributions:

- The first method to reconstruct high quality simulation-ready garments leveraging only a *single frame* of a multi-view capture.
- A hybrid garment rendering method, which leverages both the PBR-based shading for far-field effects and Gaussian splats for reproducing near-field fine details on the surface of fabrics.
- We evaluate the sheen's impact on far-field garment appearance and achieve better reconstruction results with a cloth-specific PBR model.

2. Related Work

Mesh-based Garment Reconstruction: Garment reconstruction methods have tackled this problem using specialized patterns to track cloth geometry accurately [7, 15, 46, 56], but the reconstruction is limited to garments manufactured with a specific pattern thus limiting the ability to capture appearance. Recent work [67] demonstrated the possibility of synthetically augmenting appearance for captured patterned clothing. A plethora of mesh-based works have been introduced recently that approach the problem from a deformation [45, 58], differentiable cloth simulation [13, 28, 29, 50], and inverse rendering [54, 71] perspectives. However, mesh-based works are unable to represent high-frequency details on the surface of the garment that are challenging to model with a mesh (*e.g.* fur, knits, pockets). In contrast, we augment our mesh garments with Gaussians to model the appearance of near-field geometry.

Clothed Gaussian Avatars: Several methods [31, 36, 41, 44, 54] have been proposed to animate cloth avatars using 3D Gaussian splats, including learning deformations based on the underlying avatar movements. D3GA [73] embeds 3D Gaussians onto a tetrahedral cage driven by joints and keypoints, while Animatable Gaussians (AG) [30] predict pose-dependent Gaussian offsets on 2D position maps and use linear blend skinning to animate the Gaussians. However, these approaches require multi-frame input and often struggle with tracking garments due to the lack of distinctive texture cues and self-occlusion in clothing. This can lead to a degradation in appearance. In contrast, our method overcomes these limitations by eliminating the need for tracking and requiring only a single static frame.

Hybrid Approaches: The key motivation behind leveraging hybrid approaches for clothed human applications is that they combine the best of both worlds. Mesh-based approaches are fast and easy to render but fail to capture loose

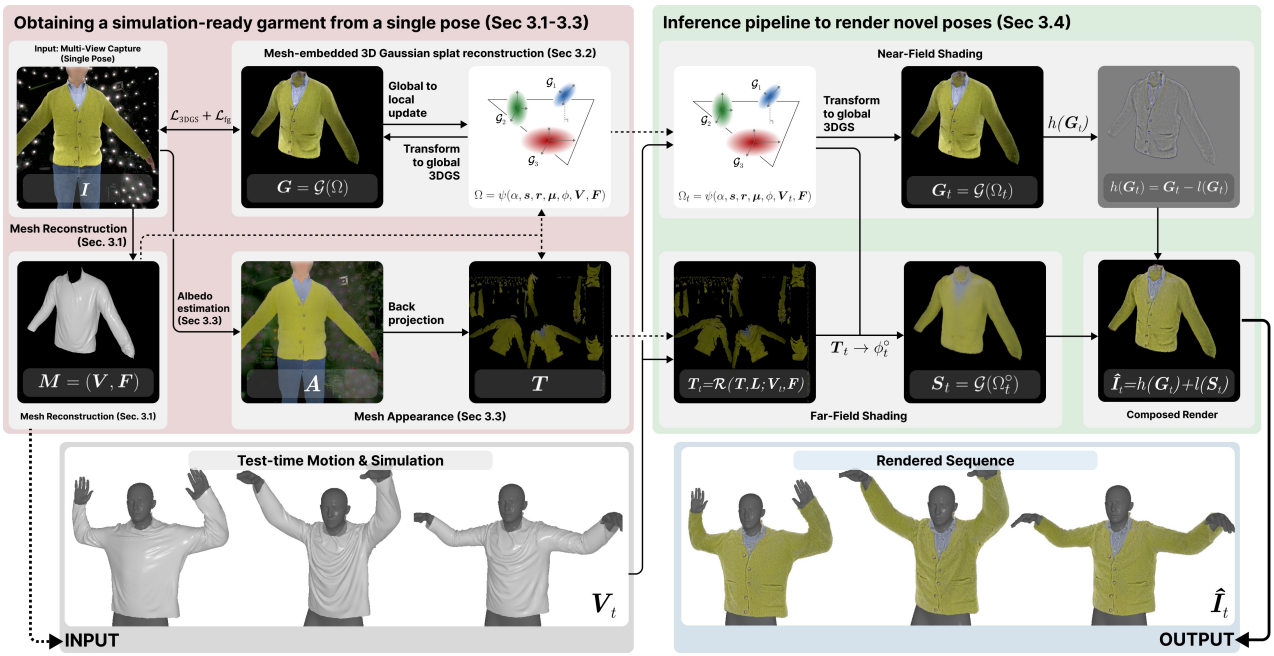


Figure 2. **Method Overview**. Given a multi-view capture of a clothed human in a single pose, we first extract the garment mesh and fit a 3DGS representation with mesh-embedded Gaussian splats. Simultaneously, we fit an albedo map of the ground truth image and back-project onto the mesh to generate a textured mesh. At inference, the mesh is shaded with a physically-based shading model and the resulting mesh colors are then transferred to zero-order spherical harmonics on the pre-optimized splats. Finally, we combine the high-pass of the original Gaussian splat reconstruction with the low-pass of the traditionally shaded result to produce the final render.

clothing and fine-level details whereas Gaussian splat-based approaches generate hiqh-quality results but their ability to be animated and driven remains an active research problem. One line of work [24, 53, 60, 61] has focused on Gaussian heads driven by underlying mesh-based representations. In addition, a wide-range of works [11, 33, 35, 38, 51, 52, 65] has introduced hybrid approaches for modeling the full clothed body. Unlike such approaches that rely on monocular video inputs, we focus on extracting the fine details from a single-pose multi-view setup. Furthermore, our main idea is to leverage Gaussians to fit only the high-frequency details of fabrics, while leaving low-frequency details to the physics-based mesh rendering approach.

Gaussians from Single Pose: Several works [43, 59, 70] have investigated reconstructing objects given a single multi-view observation for a variety of applications. SplatSim [43] leverages Gaussian splat reconstruction for rigid body objects to improve robot manipulation tasks while PhysDreamer [70] optimizes Gaussians from a multi-view capture of a scene. However, they have baked-in appearance and their method is not applicable to garments. Unlike existing works, we reconstruct the Gaussians modeling appearance details as well as the underlying surface mesh from a single pose, enabling realistic garment simulation.

3. Method

We propose a novel method for reconstructing photorealistic garment appearance from multi-view *single-pose* capture. Our system (Fig. 2) consists of three key components: 1. Mesh-embedded 3DGS reconstruction (Sec. 3.2) — building on previous work, we create a mesh-embedded 3DGS representation from our capture. It leverages the strength of 3DGS in reconstructing fine volumetric fabric details while anchoring them on a mesh to enable more faithful deformations; 2. Physics-based rendering (PBR) [39] appearance reconstruction (Sec. 3.3) — we propose a novel cloth-specific physically-based shading model to reconstruct PBR appearance from real world captures. This accurately captures shadowing and indirect illumination effects under novel pose and lighting conditions; 3. Gaussian-PBR hybrid rendering (Sec. 3.4) — we combine the strength of physically-based rendering for far-field pose-dependent shading, and use 3DGS reconstruction to fill in the missing near-field pose-independent details.

3.1. Preliminaries

We represent garments with a hybrid representation that integrates a mesh with Gaussian splats. The splats' positions are defined as offsets from the mesh. This allows the mesh to aid in simulating new dynamics and predicting shading, while the Gaussian splats are used to capture and predict

high-frequency volumetric details in the appearance of the simulated surface. In this section, we discuss geometric optimization for mesh-embedded Gaussian splatting and review physically-based rendering used to produce the far-field shading effects.

Mesh Reconstruction: Given a set of multi-view images $\{\boldsymbol{I}_k\}_{k=1}^{C}$ from C camera views of a clothed human in a static pose, we reconstruct a triangle mesh $(\boldsymbol{V}, \boldsymbol{F})$ with vertex positions $\boldsymbol{V}$ and triangles $\boldsymbol{F}$. We select a near A-pose frame such that the garment's surface is visible, and has minimal wrinkles. The images are then processed to produce a noisy point-cloud via stereo matching [4]. We further reconstruct the surface [20] using the generated point-cloud, and apply remeshing [47] to improve the triangle quality making the mesh suitable for simulation. From the reconstructed mesh of the whole body, we automatically segment out the garment mesh using image-based segmentation [12] and project the maps back to the mesh.

Mesh-embedded Gaussian Splat Representation: To couple the Gaussians with the mesh, each Gaussian is mapped to a triangle on the mesh via a triangle-local coordinate frame centered at the centroid of each triangle [40]. The orientation of this frame is determined by the triangle's normal, one of the edges normalized, and their cross product. Each 3D Gaussian is parameterized by its position $\boldsymbol{\mu} \in \mathbb{R}^3$, rotation quaternion $\boldsymbol{r} \in \mathbb{H}$ and scale $\boldsymbol{s} \in \mathbb{R}^3_+$ relative to this frame as well as the coefficients of three degrees of spherical harmonics $\phi \in [0,1]^{16\times 3}$ and opacity $\alpha \in \mathbb{R}$. The triangle-local splat positions and rotations are transformed into world space (Sec. 3.2) giving a set of global splat parameters $\Omega = \psi(\alpha, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{\mu}, \phi; \boldsymbol{V}, \boldsymbol{F})$, before the splats are finally rendered producing the final image $\boldsymbol{G} = \mathcal{G}(\Omega)$, where $\mathcal{G}$ is the Gaussian rasterizer [21]. The view parameters are implicit in $\mathcal{G}$.

3.2. Mesh-Embedded 3DGS Reconstruction

Each garment is rendered using a mesh-embedded 3D Gaussian splat representation. In addition to prior work, we further exploit the underlying mesh representation to model far-field lighting and self-shadowing effects using PBR techniques. The pre-computed asset consists of a 3D Gaussian splat reconstruction of the garment and a corresponding mesh with an albedo texture estimated from real world captures. These two representations are later combined to render the fine result as described in Sec. 3.4.

Gaussian Appearance: We sample 1 million Gaussians on the surface of the mesh via triangle point picking [55], which defines the local position $\boldsymbol{\mu}$ of each Gaussian. We initialize the local rotation $\boldsymbol{r}$, scale $\boldsymbol{s}$ and opacity α to identity, 1 and 0.1, respectively. Spherical harmonics coefficients ϕ are initialized randomly. The parameters are optimized through minimizing the sum of two losses. First, the following loss [21] between the reference and rendered images:

$$\mathcal{L}_{\text{3DGS}} = \lambda \left\| \boldsymbol{I}_k - \boldsymbol{G}_k \right\|_1 + (1-\lambda)\text{SSIM}(\boldsymbol{I}_k, \boldsymbol{G}_k), \quad (1)$$

where $\boldsymbol{G}_k$ is the rendered image, $\boldsymbol{I}_k$ is the reference image for camera k and SSIM is the structural similarity index [72]. Following [21], we use $\lambda = 0.8$ for all our experiments. To discourage outliers (*i.e.* Gaussian splats away from the surface that may fit details not related to the fabric), we adapt a regularization loss from NeRF [3] to penalize splats far away from the surface:

$$\mathcal{L}_{\text{fg}} = \lambda_{\text{fg}} \left\| \boldsymbol{M}_k - \boldsymbol{\alpha}_k \right\|_1, \quad (2)$$

where $\boldsymbol{M}_k$ is the foreground mask of the ground truth image $\boldsymbol{I}_k$ obtained using Sapiens [22] and $\boldsymbol{\alpha}_k$ is the alpha channel of the rendered Gaussians. This encourages foreground and background Gaussians to have accumulated opacity close to 1 and 0 respectively. We use $\lambda_{\text{fg}} = 0.1$ in our optimization. Unlike [44], which trains on segmented garment images, we train Gaussians on entire images. To do so, we use $\boldsymbol{\alpha}_k$ to blend the rendered garment with images of the empty capture dome. This allows us to learn fuzzy details on the garment boundaries instead of being limited by the segmentation model accuracy. Note that the alpha regularization loss is also sensitive to the segmentation accuracy and can negatively affect the reconstruction of fuzzy details. We alleviate this issue by not enforcing the loss at the contours of the garment, which we automatically identify by eroding and dilating the foreground mask $\boldsymbol{M}_k$, and masking the pixels at the exclusion. After optimizing Gaussians on the full body, we only keep the Gaussians that are bound to mesh triangles belonging to the segmented garment.

3.3. PBR Appearance Reconstruction

While the optimized 3D Gaussians (Sec. 3.2) are capable of reconstructing the garment appearance well, especially for the near-field high-frequency fabric details, they contain baked-in lighting from the training frame, and thus novel deformations (*e.g.*, new wrinkles) would not be shaded correctly as demonstrated in Fig. 5(b). Previous works typically rely on video sequences to reproduce Gaussian parameters for different mesh configurations [36, 44, 71] and involve multi-frame tracking. Instead, we leverage physically-based rendering, which naturally accounts for lighting effects such as shadowing and indirect illumination under mesh deformation. This considerably simplifies the pipeline and compute load. In this section, we detail how we reconstruct a PBR representation from our capture. We first introduce our cloth-specific PBR shading model that is suitable for both cloth appearance modeling and reconstruction. Then, we estimate the garment color using a learning-based approach, removing baked-in lighting from the input images. The other shading parameters of our model are estimated with physically-based differentiable rendering. Finally, together with the reconstructed

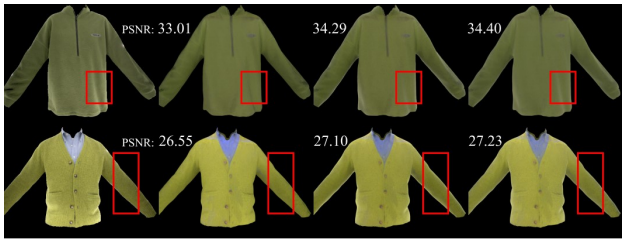


Figure 3. **PBR model evaluation**. Each mesh is shaded with 3 different shading models and compared with ground truth (a). Notice the sheen around the cardigan sleeve and the wrinkle near the fleece abdomen denoted by the red square. The Lambertian model (b) shows no sheen, Disney BRDF (c) over-estimates the forward scattered sheen, whereas our model (d) produces the best match to ground truth sheen. PSNR values are computed to quantitatively demonstrate the improvement of our model over prior methods.

mesh (Sec. 3.2), such PBR-based representation can be rendered under novel pose, view, and lighting conditions. The relighting capability is shown in Fig. 8.

Cloth Appearance Model: The Lambertian model has been commonly used in previous work [71]. However, as shown in Fig. 3(b), this naive view-independent approach misses several key fabric appearance properties and does not fit the actual data well. In contrast, we base our model on the Disney Bidirectional Reflectance Distribution Function [5] (BRDF) f_d , a versatile PBR model that is capable of creating realistic appearance for general hard surfaces and beyond. Garments exhibit strong backward and forward scattering at grazing angles, due to their flyaway fibers. This is apparent for fabrics constructed from loose staple yarns such as the fleece and cardigan in Fig. 3(a). This effect is commonly known as *sheen*. We observed that the sheen component of [5] is unable to match the capture and instead adopt the sheen BRDF f_s proposed by [66]. It is modeled based on optical simulation and takes the important fiber multiple scattering into account. Combining two BRDFs, our final appearance model f results in:

$$f(\boldsymbol{x}, \boldsymbol{i}, \boldsymbol{o}) = \sigma_s f_s(\boldsymbol{x}, \boldsymbol{i}, \boldsymbol{o}) + H_s(\boldsymbol{o})\sigma_d(\boldsymbol{x}) f_d(\boldsymbol{x}, \boldsymbol{i}, \boldsymbol{o}), \quad (3)$$

where $\boldsymbol{x}$ is the position and $(\boldsymbol{i}, \boldsymbol{o})$ are the lighting and viewing directions. σ_s is the constant sheen color and $\sigma_d(\boldsymbol{x})$ is the spatially-varying base color. $H_s(\boldsymbol{o})$ is the view-dependent sheen transmission [66], which ensures we combine two BRDFs in an energy conserving way. We found that the render using our model matches the input image better compared to both Lambertian model (completely lack of sheen) and Disney BRDF (only having forward scattering component of sheen and lack of roughness control) as demonstrated in Fig. 3.

Base Color Reconstruction: To reconstruct the base color σ_d of our model for PBR shading, our goal is to extract a spatially-varying lighting-less albedo map from input views of the training frame. To remove baked-in lighting, we use a pre-trained intrinsic image decomposition network [6] to predict the albedo for each view. Intrinsic image decomposition is inherently under-constrained and a common problem is multi-view inconsistency. To address this, we calculate a color correction matrix per view that minimizes the difference between the input image $\boldsymbol{I}_k$ and the predicted albedo within the segmentation mask $\boldsymbol{M}_k$. We found that this normalizes the predicted albedo across views effectively. We estimate and normalize the albedo from 10 input views, then back-project and merge them in texture space.

Shading Reconstruction: Given the estimated base color and the carefully calibrated capture conditions, we use Mitsuba [18], a physically-based differentiable renderer, to obtain the other stationary parameters of our model from our training frame. The parameters optimized are *roughness*, *sheen color* and *sheen roughness*. To better constrain our model in the cloth appearance space, we set all other parameters as the configuration of a rough dielectric material, including setting *metallic* and *clearcoat* to 0. We optimize the parameters jointly across all input views of the training frame. We observed that together with the reconstructed base color, our PBR-based method produces a faithful far-field reconstruction for both training and test frames.

3.4. Gaussian—PBR Hybrid Rendering

We propose a novel approach that combines the strengths of 3DGS (Sec. 3.2) and Physically-Based Rendering (PBR) representations (Sec. 3.3). The former excels at capturing fine volumetric details on fabric surfaces, but struggles with pose-dependent effects, while the latter accurately renders far-field lighting effects with novel poses, but is limited by the underlying mesh quality. By decomposing an image into high-frequency and low-frequency components, we approximate the former using Gaussian-based representation and the latter using PBR-based shading. Our key observation is that pose-dependent effects like shadowing and indirect illumination are mostly in the low-frequency signals of a render and can be well-approximated by PBR-based shading, while high-frequency signals corresponding to fine fabric details remain largely pose-independent and can be effectively captured by a Gaussian-based representation. To be specific, a $W \times H$ RGB image $\boldsymbol{I} \in [0, 1]^{W \times H \times 3}$ can be naturally decomposed into low- and high- frequency components as

$$\boldsymbol{I} = h(\boldsymbol{I}) + l(\boldsymbol{I}), \quad (4)$$

where $l(\boldsymbol{I})$ and $h(\boldsymbol{I})$ are a low-pass and high-pass filter on image $\boldsymbol{I}$ respectively. Our hybrid approach will approximate $h(\boldsymbol{I})$ with our 3D-Gaussian representation (we call *near-field shading*) and $l(\boldsymbol{I})$ with our PBR model (we call *far-field shading*) and then combine them into a final render. With this, we generalize our garment rendering to novel

deformation using single-pose reconstruction. We are able to reproduce both high-frequency fabric details and pose-dependent lighting effects. We demonstrate our results in Sec. 4. We discuss implementation details below.

Far-field Shading: To create a deformed mesh, we use a physics simulator [32] to generate a set of deformed vertices at time t, denoted by $\boldsymbol{V}_t$. It is natural to render the mesh with a PBR model directly into screen space [18]. However, mesh rasterization requires a substantially different rendering pipeline than Gaussian splatting. To combine the two representations at run-time, we choose to render into texture space. To be specific, we render the mesh using our PBR workflow (introduced in Sec. 3.3) denoted by $\mathcal{R}$ with reconstructed parameters $\boldsymbol{T}$ (including base color and other shading parameters) and a given lighting configuration $\boldsymbol{L}$. This produces a new shaded texture: $\boldsymbol{T}_t = \mathcal{R}(\boldsymbol{T}, \boldsymbol{L}; \boldsymbol{V}_t, \boldsymbol{F})$. Our cloth-specific PBR shading model Eq. (3) and viewing information are implicit in $\mathcal{R}$. For each Gaussian splat, we find the color on $\boldsymbol{T}_t$ using the pre-computed correspondence (introduced in Sec. 3.1) to obtain the zero-order spherical harmonics $\phi_t^\circ \in \mathbb{R}^3$. Thus, we can use the existing 3DGS renderer $\mathcal{G}$ to produce a final image using the new colors, but with all other parameters unchanged from the original 3DGS optimization as follows:

$$\begin{aligned} \boldsymbol{S}_t &= \mathcal{G}(\Omega_t^\circ), \\ \Omega_t^\circ &= \psi(\alpha, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{\mu}, \phi_t^\circ; \boldsymbol{V}_t, \boldsymbol{F}). \end{aligned} \tag{5}$$

In contrast to directly rendering Gaussians, this allows us to reconstruct wrinkle details under novel deformations.

Near-field Shading: As mentioned earlier, PBR shading lacks many high-frequency details of textured fabric like flyaways. To recover these details, we first render the full 3DGS reconstruction at the new configuration

$$\begin{aligned} \boldsymbol{G}_t &= \mathcal{G}(\Omega_t) \\ \Omega_t &= \psi(\alpha, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{\mu}, \phi; \boldsymbol{V}_t, \boldsymbol{F}). \end{aligned} \tag{6}$$

We then extract the near-field shading and appearance from the optimized 3D Gaussians by applying a high-pass filter h to the Gaussian rendering: $h(\boldsymbol{G}_t) = \boldsymbol{G}_t - l(\boldsymbol{G}_t)$, where $l(\boldsymbol{I})$ is an $\boldsymbol{\alpha}$-scaled low-pass filter on image $\boldsymbol{I}$ where $\boldsymbol{\alpha}$ the alpha channel of $\boldsymbol{I}$. The low-pass is computed with an alpha-aware Gaussian blur, so that the background does not contaminate the garment color at the boundary. To achieve this, we renormalize the kernel after scaling it by alpha as $K(x, y)\boldsymbol{\alpha}(x, y)$ for each kernel pixel at position (x, y). The alpha weighting is applied before kernel normalization. We use the rasterized Gaussian opacity as our alpha mask.

Composed Render: Our final render $\hat{\boldsymbol{I}}$ combines the high-pass near-field reconstruction of Eq. (6) with a low-pass of the corresponding far-field reconstruction from Eq. (5):

$$\hat{\boldsymbol{I}}_t = h(\boldsymbol{G}_t) + l(\boldsymbol{S}_t). \tag{7}$$

Fig. 6 demonstrates the composed render on 4 different garments. Together with Eq. (4) we identify the reconstruction error $|\boldsymbol{I} - \hat{\boldsymbol{I}}_t|$ of our hybrid approach for the training frame as shown in Fig. 4 and empirically find the size of the blur kernel that minimizes the reconstruction error. Please refer to the supplemental material for more discussion.



Figure 4. **Reconstruction.** The fleece garment $\boldsymbol{I}$ is reconstructed from the sum of the far-field approximation $l(\boldsymbol{S})$ and near-field approximation $h(\boldsymbol{G})$ to produce the final reconstruction $\hat{\boldsymbol{I}}$. The image on the right shows the reconstruction error.

4. Experiments

Data: We use four garments: a loose t-shirt, a loose dress, a fleece quarter-zip, and a knit cardigan to demonstrate our ability to handle loose garments and fuzzy fabrics. The clothed subjects are recorded with a multi-view system consisting of 170 cameras arranged uniformly in a half-dome configuration above the ground. Images are captured at 4096×2668 resolution. Each subject followed a predefined script that includes a variety of poses to ensure diverse data capture. However, despite capturing temporal sequences, we utilize only a single frame near the A-pose for each subject to reconstruct the garments. The rest of the ground truth captures represent novel poses, which we use to evaluate our method in all figures except Fig. 3 and Fig. 4.

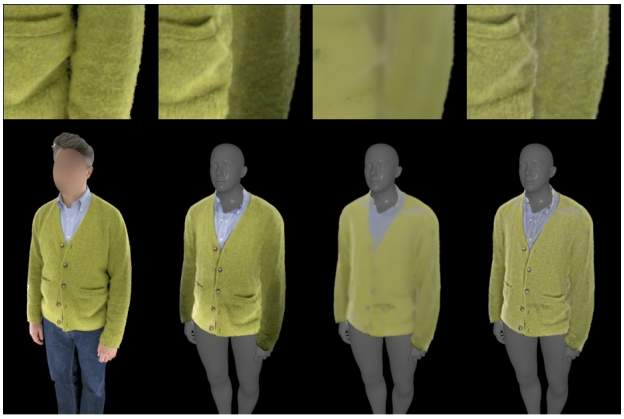
Metrics: We validate our method using two metrics, learned perceptual similarity (LPIPS) [69] and feature similarity (FSIM) [68], which compares salient low-level features to better match the human visual system. Note that pixel-based metrics such as PSNR are less appropriate for our work at test time since we use a simulated result with manually defined parameters that may not align with ground truth as can be seen in Fig. 7 (see Sec. 4.4 for proposed improvements). The goal of our approach is not to achieve perfect pixel-to-pixel correspondence at test time. Instead, we show that new wrinkles are generated under new deformations and their aesthetic is preserved, which we demonstrate qualitatively in Fig. 5 and Fig. 7. Quantitative metrics were computed on four video sequences containing 150 unseen frames each, including a goal-post pose, spooky pose and upside-down U pose as shown in Fig. 1.

4.1. Ablation Studies

We compare our method qualitatively and quantitatively with a series of ablations where the importance of key components of our method are evaluated. We provide qualitative comparisons in Fig. 5 as well as quantitative comparisons in Tab. 1 (bottom). Specifically, we explore the performance

Table 1. **Quantitative Comparisons**: Our proposed approach improves significantly upon top-performing past work (top). In addition we conduct ablation studies (bottom) on novel frames with unseen poses. Our full method outperforms related work and ablated results of our hybrid representation.

Method	FSIM↑	LPIPS ($\times 10^{-2}$) ↓
SCARF [10]	0.764	5.00
Animatable Gaussians [30]	0.827	3.39
3DGS-Only	0.825	3.41
PBR-Only	0.809	4.67
Ours	**0.834**	**3.38**



(a) GT (b) 3DGS-Only (c) PBR-Only (d) Ours

Figure 5. **Ablation**. We compare our approach to different ablations of our method and ground truth (a) of a novel pose. 3DGS-Only (b) has baked in shading from the training pose, resulting in unrealistic appearance when deformed to novel poses. PBR-Only (c) has reshading, but lacks detailed shading, and can only model flat 2D texture colors. Ours (d) has the best of both worlds.

of: a) 3DGS-Only that has baked in shading from the training pose, which makes the appearance unrealistic when deformed to novel poses, b) PBR-Only that has reshading, but lacks the detailed shading and volumetric flyaways meaning that it can only render 2D surfaces, and c) our approach that combines the far-field relightable BRDF render with the near-field shading and fine volumetric details of the mesh-embedded 3DGS.

Relighting: We demonstrate that our garments can be rendered under a variety of lighting conditions in Fig. 8. Note that compared to pure 3DGS-based methods that require learning pose-dependent intrinsic maps from multi-view multi-pose data, we rely on a single pose only. This figure also demonstrates a clear sheen effect in novel lighting optimized against real world captures.

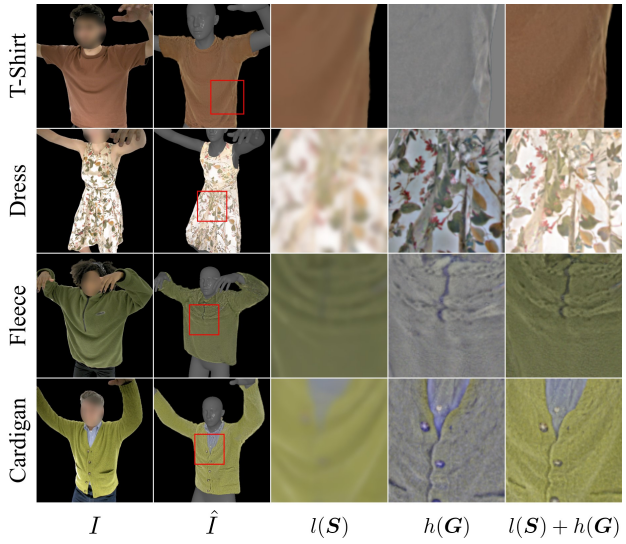


Figure 6. **Decomposition**. Our method composes the lowpass shaded Gaussians $l(\boldsymbol{S})$ with the highpass Gaussians $h(\boldsymbol{G})$ to render garments under novel deformations. Note that $l(\boldsymbol{S})$ contributes the far-field shading (*e.g.*, mesh wrinkles) while $h(\boldsymbol{G})$ contributes the near-field, high-frequency details.

4.2. Body Animation & Cloth Simulation

Although any simulation method would be effective, we simulate the garment on body mesh sequences using an implementation of the eXtended Position Based Dynamics (XPBD) [32] simulation method, chosen for its excellent performance characteristics. We use a SMPL-like parameterized statistical body model. Given the tracked skeleton motion from the video sequence, we use linear blend skinning (LBS) to animate the human body. The garment is simulated with the body mesh as the collider. The garment meshes are directly used for rendering the novel frames.

4.3. Comparisons with State of the Art

We compare with two top-performing recent approaches: Animatable Gaussians (AG) [30] and SCARF [10] and provide our qualitative and quantitative comparisons in Fig. 7 and Tab. 1. We refer the reader to the supplementary material for additional details for each baseline method. We observe that both past works struggle on loose clothing like the dress. AG has difficulty modeling loose garments since the underlying skinning module cannot represent dynamic motions, so it produces blurriness at the loose part of the clothing. Since neither baselines use physics simulation to drive the dress motion, the resulting dynamical movements are unrealistic in comparison to ours. However, despite producing higher quality motion, our approach does not optimize for garment rest shape and instead uses the reconstructed geometry as the rest shape for simulation. This is a well-known phenomenon (*i.e.* sagging [17]), which causes



(a) GT (b) SCARF (c) AG (d) Ours

Figure 7. **Qualitative Comparisons**. We compare our proposed approach (d) with the state-of-the-art methods SCARF [10] (b) Animatable Gaussians [30] (c) in a novel pose setting. We observe that both baselines struggle with loose clothing and at the same time they are not lighting-aware while our method can faithfully represent the dress and render it under novel lighting.

the dress to sag under gravity and hence the patterns on the dress (although more crisp) appear lower in the frame than in the reference. We believe this is the main reason for AG to produce slightly higher PSNR values and we note that other state-of-the-art methods [44] excluded quantitative comparisons to AG altogether. Additionally, our shading model also provides better generalization.

Although we cannot explicitly compare results against Gaussian Garments [44], since their code is not available, we note that not unlike their method, we also outperform SCARF in all metrics, without the need for multi-frame tracking. As a result, our method produces more crisp details since we can completely avoid motion-blur effects in video captures and averaging artifacts during multi-frame tracking. In terms of performance, Gaussian Garments requires an additional 24.5 hours for multi-view registration, which our method bypasses entirely.

4.4. Limitations and Future Work

Pose-dependent Details and Relightability. Our approach ignores the effects of deformation and novel lighting on high-frequency appearance. It is also limited by the capacity of 3DGS of generalizing to under-observed areas (*e.g.* armpit area). While this can be partially inferred on a full garment by tracking multi-view video capture per garment as in [44], another approach would be to capture fine details using close-up captures of fabrics and with varying lighting conditions for a fully relightable model.

Albedo Extraction and Back-projection. Our method uses an off-the-shelf albedo extraction technique [6], which we apply per camera view, that generates inconsistent colors when back-projected onto the mesh. This causes some artifacts in the far-field render. Addressing this issue would enable us to make use of a smaller blur kernel and rely more heavily on the generalizable far-field PBR component.

Range of Fabrics. We target common fabrics with fuzzy surfaces that are difficult to reproduce with pure PBR methods. Long fur or transparent fabrics would require mesh-reconstruction techniques and accurate tracking [44].

Geometry Misalignment. To better match ground truth geometry, we can leverage differentiable simulation [29, 50] to recover more accurate rest shape and material parameters. This would enable editability, resizing [9], and sag-free rest shape estimation [17] to obtain simulated garments that remain consistent in size with the reconstructed geometries. This would improve feature alignment at test time, potentially improving similarity metrics to ground truth. Finally, our method assumes shell geometry along the entire garment, which would not accurately model the deformation of pockets, multi-layer garments, or garment openings.

Runtime: Once reconstructed, our garments can be leveraged in real-time applications such as games or virtual telepresence. Our method relies exclusively on techniques that run in real-time: a) PBR techniques are commonly used in real-time games [1], and b) numerous recent works [34, 37, 57] demonstrate real-time Gaussian splat rendering. This, in combination with real-time simulation [32], enable our garments to be rendered and simulated in real-time, making them highly versatile for various applications.



Figure 8. **Relighting**. The PBR component of the composed render enables our assets to correctly respond to different lighting.

5. Conclusion

We introduced a new method that generates real-time capable virtual garment assets from a single frame multi-view capture. Our approach extends 3D Gaussian splatting applied to garment captures with the ability to generalize to new deformations such as those produced by cloth simulation tools and new lighting conditions. Our results outperform both qualitatively and quantitatively recent state-of-the-art methods while remaining simple and requiring only captures of a single pose. This eliminates the need for complex surface tracking and prevents blurring of fine details due to averaging and motion blur effects of multi-frame captures. We believe that this work brings us one step closer to more realistic clothed avatars with garment assets suitable for virtual try-on and telepresence applications.

Acknowledgements

We thank Sébastien Speierer for providing advice on inverse rendering and Tomas Simon for help with processing captures. This research was supported by Meta Reality Labs. Michelle Guo is also supported by the NSF GRFP, NSF RI #2211258 and #2338203, and the Stanford Institute for Human Centered AI (HAI).

References

[1] Tomas Akenine-Moller, Eric Haines, and Naty Hoffman. *Real-time rendering*. AK Peters/crc Press, 2019. 8

[2] Seungbae Bang, Maria Korosteleva, and Sung-Hee Lee. Estimating garment patterns from static scan data. In *Computer Graphics Forum*, pages 273–287. Wiley Online Library, 2021. 1

[3] Sai Bi, Zexiang Xu, Pratul P. Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Milos Hasan, Yannick Hold-Geoffroy, David J. Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *CoRR*, abs/2008.03824, 2020. 4

[4] Michael Bleyer, Christoph Rhemann, and Carsten Rother. Patchmatch stereo-stereo matching with slanted support windows. In *Bmvc*, pages 1–11, 2011. 4

[5] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *Acm Siggraph*, pages 1–7. vol. 2012, 2012. 5

[6] Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM TOG*, 43(1), 2023. 5, 8

[7] He Chen, Hyojoon Park, Kutay Macit, and Ladislav Kavan. Capturing detailed deformations of moving human bodies. *ACM Transactions on Graphics (TOG)*, 40(4):1–18, 2021. 2

[8] Hsiao-yu Chen, Edith Tretschk, Tuur Stuyck, Petr Kadlecek, Ladislav Kavan, Etienne Vouga, and Christoph Lassner. Virtual elastic objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15827–15837, 2022. 1

[9] Hsiao-yu Chen, Egor Larionov, Ladislav Kavan, Gene Lin, Doug Roble, Olga Sorkine-Hornung, and Tuur Stuyck. Dress anyone: Automatic physically-based garment pattern refitting. *arXiv preprint arXiv:2405.19148*, 2024. 8

[10] Yao Feng, Jinlong Yang, Marc Pollefeys, Michael J. Black, and Timo Bolkart. Capturing and animation of body and clothing from monocular video. In *SIGGRAPH Asia 2022 Conference Papers*, 2022. 7, 8

[11] Yao Feng, Jinlong Yang, Marc Pollefeys, Michael J Black, and Timo Bolkart. Capturing and animation of body and clothing from monocular video. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 3

[12] Cheng-Yang Fu, Tamara L Berg, and Alexander C Berg. Imp: Instance mask projection for high accuracy semantic segmentation of things. In *ICCV*, 2019. 4

[13] Artur Grigorev, Michael J Black, and Otmar Hilliges. Hood: Hierarchical graphs for generalized modelling of clothing dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16965–16974, 2023. 2

[14] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. *CVPR*, 2024. 2

[15] Oshri Halimi, Tuur Stuyck, Donglai Xiang, Timur M Bagautdinov, He Wen, Ron Kimmel, Takaaki Shiratori, Chenglei Wu, Yaser Sheikh, and Fabian Prada. Pattern-based cloth registration and sparse-view animation. *ACM Trans. Graph.*, 41(6):196–1, 2022. 2

[16] Kai He, Kaixin Yao, Qixuan Zhang, Jingyi Yu, Lingjie Liu, and Lan Xu. Dresscode: Autoregressively sewing and generating garments from text guidance. *ACM Transactions on Graphics (TOG)*, 43(4):1–13, 2024. 1

[17] Jerry Hsu, Nghia Truong, Cem Yuksel, and Kui Wu. A general two-stage initialization for sag-free deformable simulations. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2022)*, 41(4):64:1–64:13, 2022. 7, 8

[18] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer, 2022. https://mitsuba-renderer.org. 5, 6

[19] Ying Jiang, Chang Yu, Tianyi Xie, Xuan Li, Yutao Feng, Huamin Wang, Minchen Li, Henry Lau, Feng Gao, Yin Yang, et al. Vr-gs: A physical dynamics-aware interactive gaussian splatting system in virtual reality. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–1, 2024. 1

[20] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013. 4

[21] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4), 2023. 1, 2, 4

[22] Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and Shunsuke Saito. Sapiens: Foundation for human vision models. In *Computer Vision – ECCV 2024*, pages 206–228, Cham, 2025. Springer Nature Switzerland. 4

[23] Taeksoo Kim, Byungjun Kim, Shunsuke Saito, and Hanbyul Joo. Gala: Generating animatable layered assets from a single scan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1535–1545, 2024. 1

[24] Tobias Kirschstein, Simon Giebenhain, Jiapeng Tang, Markos Georgopoulos, and Matthias Nießner. Gghead: Fast and generalizable 3d gaussian heads. *arXiv preprint arXiv:2406.09377*, 2024. 3

[25] Boqian Li, Xuan Li, Ying Jiang, Tianyi Xie, Feng Gao, Huamin Wang, Yin Yang, and Chenfanfu Jiang. Garmentdreamer: 3dgs guided garment synthesis with diverse geometry and texture details. *arXiv preprint arXiv:2405.12420*, 2024. 1

[26] Ren Li, Corentin Dumery, Benoit Guillard, and Pascal Fua. Garment Recovery with Shape and Deformation Priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 1

[27] Yue Li, Marc Habermann, Bernhard Thomaszewski, Stelian Coros, Thabo Beeler, and Christian Theobalt. Deep physics-aware inference of cloth deformation for monocular human performance capture. In *2021 International Conference on 3D Vision (3DV)*, pages 373–384. IEEE, 2021. 1

[28] Yifei Li, Tao Du, Kui Wu, Jie Xu, and Wojciech Matusik. Diffcloth: Differentiable cloth simulation with dry frictional contact. *ACM Transactions on Graphics (TOG)*, 42(1):1–20, 2022. 2

[29] Yifei Li, Hsiao-yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, and Tuur Stuyck. DiffAvatar: Simulation-ready garment optimization with differentiable simulation. In *CVPR*, 2024. 1, 2, 8

[30] Zhe Li, Zerong Zheng, Lizhen Wang, and Yebin Liu. Animatable gaussians: Learning pose-dependent gaussian maps for high-fidelity human avatar modeling. In *CVPR*, 2024. 1, 2, 7, 8

[31] Siyou Lin, Zhe Li, Zhaoqi Su, Zerong Zheng, Hongwen Zhang, and Yebin Liu. Layga: Layered gaussian avatars for animatable clothing transfer. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 1, 2

[32] Miles Macklin, Matthias Müller, and Nuttapong Chentanez. Xpbd: position-based simulation of compliant constrained dynamics. In *Proceedings of the 9th International Conference on Motion in Games*, pages 49–54, 2016. 6, 7, 8

[33] Gyeongsik Moon, Takaaki Shiratori, and Shunsuke Saito. Expressive whole-body 3d gaussian avatar. *arXiv preprint arXiv:2407.21686*, 2024. 2, 3

[34] Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, and Federico Tombari. Radsplat: Radiance field-informed gaussian splatting for robust real-time rendering with 900+ fps. *arXiv preprint arXiv:2403.13806*, 2024. 8

[35] Panwang Pan, Zhuo Su, Chenguo Lin, Zhen Fan, Yongjie Zhang, Zeming Li, Tingting Shen, Yadong Mu, and Yebin Liu. Humansplat: Generalizable single-image human gaussian splatting with structure priors. *arXiv preprint arXiv:2406.12459*, 2024. 3

[36] Bo Peng, Yunfan Tao, Haoyu Zhan, Yudong Guo, and Juyong Zhang. Pica: Physics-integrated clothed avatar. *arXiv preprint arXiv:2407.05324*, 2024. 2, 4

[37] Zhexi Peng, Tianjia Shao, Yong Liu, Jingke Zhou, Yin Yang, Jingdong Wang, and Kun Zhou. Rtg-slam: Real-time 3d reconstruction at scale using gaussian splatting. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 8

[38] Marco Pesavento, Yuanlu Xu, Nikolaos Sarafianos, Robert Maier, Ziyan Wang, Chun-Han Yao, Marco Volino, Edmond Boyer, Adrian Hilton, and Tony Tung. Anim: Accurate neural implicit model for human reconstruction from a single rgb-d image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5448–5458, 2024. 3

[39] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, 2023. 3

[40] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner. Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20299–20309, 2024. 2, 4

[41] Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, and Siyu Tang. 3dgs-avatar: Animatable avatars via deformable 3d gaussian splatting. In *CVPR*, 2024. 2

[42] Lingteng Qiu, Guanying Chen, Jiapeng Zhou, Mutian Xu, Junle Wang, and Xiaoguang Han. Rec-mv: Reconstructing 3d dynamic cloth from monocular videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4637–4646, 2023. 1

[43] Mohammad Nomaan Qureshi, Sparsh Garg, Francisco Yandun, David Held, George Kantor, and Abhishesh Silwal. Splatsim: Zero-shot sim2real transfer of rgb manipulation policies using gaussian splatting. *arXiv preprint arXiv:2409.10161*, 2024. 3

[44] Boxiang Rong, Artur Grigorev, Wenbo Wang, Michael J. Black, Bernhard Thomaszewski, Christina Tsalicoglou, and Otmar Hilliges. Gaussian garments: Reconstructing simulation-ready clothing with photorealistic appearance from multi-view video, 2024. 1, 2, 4, 8

[45] Nikolaos Sarafianos, Tuur Stuyck, Xiaoyu Xiang, Yilei Li, Jovan Popovic, and Rakesh Ranjan. Garment3dgen: 3d garment stylization and texture generation. *2025 International Conference on 3D Vision (3DV)*, 2025. 1, 2

[46] Volker Scholz, Timo Stich, Marcus Magnor, Michael Keckeisen, and Markus Wacker. Garment motion capture using color-coded patterns. In *ACM SIGGRAPH 2005 Sketches*, pages 38–es, 2005. 2

[47] SideFX. Houdini, 2024. 4

[48] Astitva Srivastava, Pranav Manu, Amit Raj, Varun Jampani, and Avinash Sharma. Wordrobe: Text-guided generation of textured 3d garments. In *European Conference on Computer Vision*, pages 458–475. Springer, 2025. 1

[49] Tuur Stuyck. *Cloth simulation for computer graphics*. Springer Nature, 2022. 1

[50] Tuur Stuyck and Hsiao-yu Chen. Diffxpbd: Differentiable position-based simulation of compliant constraint dynamics. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 6(3):1–14, 2023. 2, 8

[51] David Svitov, Pietro Morerio, Lourdes Agapito, and Alessio Del Bue. Haha: Highly articulated gaussian human avatars with textured mesh prior. *arXiv preprint arXiv:2404.01053*, 2024. 3

[52] Angtian Wang, Yuanlu Xu, Nikolaos Sarafianos, Robert Maier, Edmond Boyer, Alan Yuille, and Tony Tung. Hisr: Hybrid implicit surface representation for photorealistic 3d human reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5298–5308, 2024. 3

[53] Cong Wang, Di Kang, He-Yi Sun, Shen-Han Qian, Zi-Xuan Wang, Linchao Bao, and Song-Hai Zhang. Mega: Hybrid mesh-gaussian head avatar for high-fidelity rendering and head editing. *arXiv preprint arXiv:2404.19026*, 2024. 3

[54] Shaofei Wang, Božidar Antić, Andreas Geiger, and Siyu Tang. Intrinsicavatar: Physically based inverse rendering

of dynamic humans from monocular videos via explicit ray tracing. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[55] Eric W. Weisstein. Triangle point picking, 2024. Visited on Nov. 6. 4

[56] Ryan White, Keenan Crane, and David A Forsyth. Capturing and animating occluded cloth. *ACM Transactions on Graphics (TOG)*, 26(3):34–es, 2007. 2

[57] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. 8

[58] Donglai Xiang, Timur Bagautdinov, Tuur Stuyck, Fabian Prada, Javier Romero, Weipeng Xu, Shunsuke Saito, Jingfan Guo, Breannan Smith, Takaaki Shiratori, et al. Dressing avatars: Deep photorealistic appearance for physically simulated clothing. *ACM TOG*, 41(6):1–15, 2022. 1, 2

[59] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4389–4398, 2024. 1, 3

[60] Yuelang Xu, Benwang Chen, Zhe Li, Hongwen Zhang, Lizhen Wang, Zerong Zheng, and Yebin Liu. Gaussian head avatar: Ultra high-fidelity head avatar via dynamic gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1931–1941, 2024. 3

[61] Yuelang Xu, Lizhen Wang, Zerong Zheng, Zhaoqi Su, and Yebin Liu. 3d gaussian parametric head model. In *European Conference on Computer Vision*, pages 129–147. Springer, 2025. 3

[62] Shan Yang, Zherong Pan, Tanya Amert, Ke Wang, Licheng Yu, Tamara Berg, and Ming C. Lin. Physics-inspired garment recovery from a single-view image. *ACM Trans. Graph.*, 37(5), 2018. 1

[63] Tao Yu, Zerong Zheng, Yuan Zhong, Jianhui Zhao, Qionghai Dai, Gerard Pons-Moll, and Yebin Liu. Simulcap: Single-view human performance capture with cloth simulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5504–5514, 2019. 1

[64] Weiran Yuan, Yujun Chen, and André Gagalowicz. Meshless virtual cloth. In *Computer Vision/Computer Graphics CollaborationTechniques: 4th International Conference, MIRAGE 2009, Rocquencourt, France, May 4-6, 2009. Proceedings 4*, pages 308–320. Springer, 2009. 1

[65] Ye Yuan, Xueting Li, Yangyi Huang, Shalini De Mello, Koki Nagano, Jan Kautz, and Umar Iqbal. Gavatar: Animatable 3d gaussian avatars with implicit mesh learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 896–905, 2024. 3

[66] Tizian Zeltner, Brent Burley, and Matt Jen-Yuan Chiang. Practical multiple-scattering sheen using linearly transformed cosines. In *ACM SIGGRAPH 2022 Talks*, pages 1–2, 2022. 5

[67] Youyi Zhan, Tuanfeng Y Wang, Tianjia Shao, and Kun Zhou. Pattern guided uv recovery for realistic video garment texturing. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[68] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. 6

[69] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 6

[70] Tianyuan Zhang, Hong-Xing Yu, Rundi Wu, Brandon Y Feng, Changxi Zheng, Noah Snavely, Jiajun Wu, and William T Freeman. Physdreamer: Physics-based interaction with 3d objects via video generation. In *European Conference on Computer Vision*, pages 388–406. Springer, 2025. 1, 3

[71] Yang Zheng, Qingqing Zhao, Guandao Yang, Wang Yifan, Donglai Xiang, Florian Dubost, Dmitry Lagun, Thabo Beeler, Federico Tombari, Leonidas Guibas, and Gordon Wetzstein. Physavatar: Learning the physics of dressed 3d avatars from visual observations. In *ECCV*, 2024. 1, 2, 4, 5

[72] Wang Zhou. Image quality assessment: from error measurement to structural similarity. *IEEE TIP*, 13:600–613, 2004. 4

[73] Wojciech Zielonka, Timur Bagautdinov, Shunsuke Saito, Michael Zollhöfer, Justus Thies, and Javier Romero. Drivable 3d gaussian avatars. In *International Conference on 3D Vision (3DV)*, 2025. 2