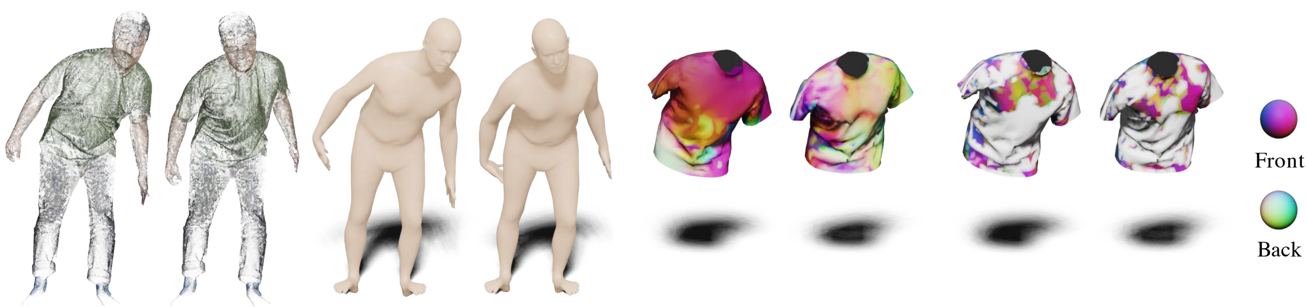
Inverse Simulation: Reconstructing Dynamic Geometry of Clothed Humans via Optimal Control

Jingfan Guo[1] Jie Li[1] Rahul Narain[2] Hyun Soo Park[1]
[1]University of Minnesota
[2]Indian Institute of Technology Delhi



(a) Input point cloud (b) Estimated body (c) Velocity direction (d) Force direction

Figure 1: We present an inverse simulation method to reconstruct time-varying geometry of the clothes and the underlying body. Given a sequence of point clouds (a) capturing a clothed human, we optimize the body states (shape and pose) (b) such that the generated cloth motion (c, d) via physics simulation is matched to the point clouds. Colors in (c) and (d) illustrate the instantaneous velocity and contact force directions on the cloth vertices. The vertices not in contact are colored white.

Abstract

This paper studies the problem of inverse cloth simulation—to estimate shape and time-varying poses of the underlying body that generates physically plausible cloth motion, which matches to the point cloud measurements on the clothed humans. A key innovation is to represent the dynamics of the cloth geometry using a dynamical system that is controlled by the body states (shape and pose). This allows us to express the cloth motion as a resultant of external (skin friction and gravity) and internal (elasticity) forces. Inspired by the theory of optimal control, we optimize the body states such that the simulated cloth motion is matched to the point cloud measurements, and the analytic gradient of the simulator is back-propagated to update the body states. We propose a cloth relaxation scheme to initialize the cloth state, which ensures the physical validity. Our method produces physically plausible and temporally smooth cloth and body movements that are faithful to the measurements, and shows superior performance compared to the existing methods. As a byproduct, the stress and strain that are applied to the body and clothes can be recovered.

1. Introduction

The clothing worn by a person generates complex time-varying geometry, e.g., wrinkles and folds, as the person moves. This is a form of secondary motion, arising in response to the body movement due to physical forces exerted by the contact with body surface, gravity, and the fabric's own material response. This behavior of cloth has been modeled in computer graphics [5, 10, 32] to produce physically plausible character animation. On the other hand, computer vision methods have enabled reconstructing the static geometry of the clothes [1,2,11,16,33,34] that is faithful to the visual measurements of appearance and shape. However, these approaches lack understanding of physics, resulting in non-interpretable and unrealistic cloth motion. This paper studies a problem of *inverse cloth simulation*—to estimate the underlying body shape and time-varying pose that generates physically plausible cloth motion that matches to the proxy measurements of cloth geometry, e.g., time series of point clouds.

Inverse cloth simulation is a fundamentally ill-posed problem because of the primary function of the clothes, i.e., it is designed to cover and protect the body, resulting in self-occlusion. There exist multiple shape and pose configurations that can match to the given cloth geometry, e.g., a skinny person wearing a loose shirt vs. a doughy person wearing a tight shirt. Further, a cloth's secondary motion is inherently time-varying whereas existing cloth modeling approaches based on a stationary relationship between body and cloth [1, 2, 16, 33] have limited expressibility on its dynamics.

We address this challenge by representing the time-varying cloth geometry as a result of physical interactions with the body movement. Specifically, we model the dynamics of the cloth using a dynamical system whose state is controlled by the movement history of the body states (shape and pose). Inspired by the theory of optimal control [4], we provide a solution for the inverse simulation problem, by matching the dynamic cloth geometry to the measurements as optimizing over the body states. Figure 1 illustrates the dynamics of the clothes that align with the point cloud measurements, which produces physically plausible cloth and body motion.

Our method takes as input a sequence of 3D point sets capturing a moving clothed human, and reconstructs time-varying geometry of the clothes and the underlying body. We initialize the cloth geometry by relaxing it to fit the shape and pose of the undressed body at the first time instant, parametrized by the SMPL model [23]. Given the equation of cloth motion that describes the rate of cloth velocity changes due to the external (skin friction and gravity) and internal (elasticity) forces, we use a differentiable physics simulator to solve the dynamical system, which generates the movement of the clothes in response to the body motion. We optimize the body state such that the distance between the resulting time-varying geometry of the clothes and the 3D point measurements is minimized. To enable the optimization, we derive the analytic gradient of the simulator, which allows us to back-propagate the errors to update the body states.

We demonstrate that our method produces physically plausible and temporally smooth cloth and body motion that match to the noisy point cloud measurements. As a byproduct, the stress and strain that were applied to both the clothes and body can be recovered (Figure 1(c) and (d)). Our method is agnostic to the type of the clothes, e.g., T-shirt, dress, and pants, and shows strong performance compared to the existing baselines that have lack of physics understanding.

The main contributions of this paper include (1) a new formulation of a dynamical system that models the dynamics of the clothes as a function of the shape and time-varying pose of the body; (2) a novel application of optimal control that uses inverse simulation to optimize the body states to fit the time-varying cloth geometry to the measurement (3) an analytic gradient of the simulator that allows us to update the body states; (4) a cloth initialization scheme using a steady state cloth simulation as transforming the body from the T-pose to the pose at the first time instant; (5) physically plausible reconstruction of the diverse types of the clothes faithful to the real-world measurements.

2. Related Work

We review recent works on clothed human body reconstruction in computer vision and cloth simulation in computer graphics, and discuss how they differ from the proposed method.

Clothed Human Reconstruction Reconstruction of the geometry of the clothed humans is a challenging task because it is a resultant of complex physical interactions between body and clothes. Existing works [1, 2, 11, 16, 33, 34] by large focus on fitting the cloth and body geometry to the measurements without considering their physical interactions. A few studies address this limitation by reconstructing the cloth and body as separate entities. For example, DoubleFusion [43] reconstructs undressed body and clothed body as two separate layers from RGB-D videos. IP-Net [7] learns to reconstruct implicit functions of double layered surfaces for undressed body and clothed body, respectively. The SMPL and SMPL+D models are registered to the implicit function to generate the reconstruction surfaces. ClothCap [31] segments clothes from a high resolution 4D scan of a clothed human body, and estimates the undressed body under the clothes. Multi-Garment Net [8] predicts static body and clothes from a short clip of video. While body and clothes are modeled separately, it still lacks awareness of physical interactions.

Physics based simulation produces physically plausible motion that is ideal for understanding the interactions between clothes and body. SimulCap [44] is a physics-based performance capture method that can capture details of the clothing deformation using a single depth camera. Because of the nature of the physics, temporal coherence of the cloth deformation can be enforced and missing data can be handled. However, its over-simplified simulator (e.g., mass-spring system) has limited expressibility to model complex cloth dynamics. To account for this limitation, SimulCap uses an artificial force to fit the simulated cloth to the depth.

Cloth Simulation The simulation of thin-shell has been of great interest to the graphics community for decades since the pioneering method allowing for stable cloth simulation with considerable time steps [5]. Depending on how the internal forces are formed, there are a variety of different methods being explored. One of the most common methods is finite element method (FEM) [5, 27], which is physically conforming as being based on continuum mechanics and widely used in both cloth and 3D solid deformable objects. The detailed introduction can be found at [40]. Mass-spring method [10, 22, 32] is an over-simplified model that only captures the forces within edges, which can be used for the occasions only requiring very rough cloth-like behaviors. There are also methods including position based dynamics (PBD) [25], projective dynamics [9, 24] and ADMM [26, 28] which are useful in real-time simu-

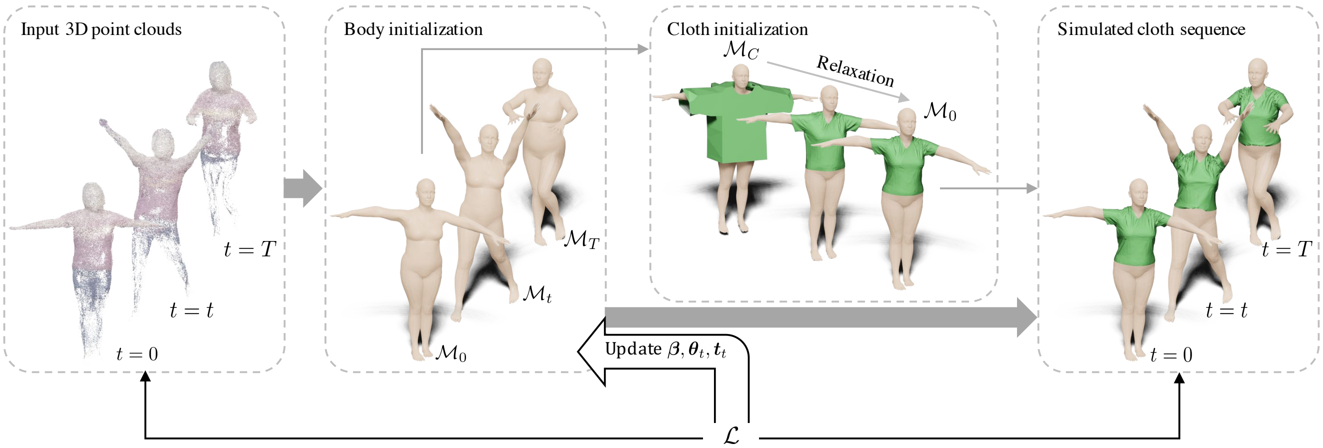


Figure 2: The pipeline of the proposed method. Given a time series of 3D point clouds capturing a clothed human, we estimate the body pose and shape parameters by fitting the SMPL body surfaces to the point clouds. Based on a canonical body and the estimated body at the first time instant, we initialize the cloth geometry using a cloth relaxation scheme. We simulate the cloth motion, and measure the error between the simulated time series of cloth geometry and the point clouds. We optimize the body states using the analytic gradients of the physics simulator. We iterate this process until the cloth motion is faithfully matched to the point clouds.

lation scenarios where the running time is reduced by sacrificing accuracy to some extent. Recently, material point method (MPM) [37, 39] has become popular in the simulation area which handles the coupling of different object types (rigid body, deformable body, water, sand etc.) very well, with some of these works [13, 15] focusing on cloth simulation. There are also methods going even further in the cloth model by a yarn-level simulation [17, 18].

Our method integrates cloth simulation into the clothed human reconstruction. We leverage a differentiable physics simulation [14, 21, 42] that can be optimized to fit to the measurement data. To our best knowledge, this is the first paper that applies the differentiable simulator to reconstruct the time-varying geometry of body and clothes.

3. Method

We present a novel method to reconstruct dynamic clothes and underlying human body shape and pose from a time series of 3D point sets on the clothes. Inspired by the theory of optimal control [4], we model the cloth motion/deformation using Newtonian dynamics that is modulated by the body movement.

3.1. Overview

Figure 2 illustrates the pipeline of the proposed method. We first fit the SMPL body to the input time series of 3D point clouds, and obtain a sequence of body pose and shape parameters as the initial body state. The cloth is then initialized using a cloth relaxation scheme, which transforms the cloth geometry from a canonical state to the state of the first time instant, and ensures the physical validity. We simulate the cloth motion with the body sequence, and measure the error between the simulated time series of cloth geometry and the point clouds. We compute the analytic gradients of the physics simulator, and use them to optimize the body states. By iterating this process, we match the cloth motion faithfully to the point clouds.

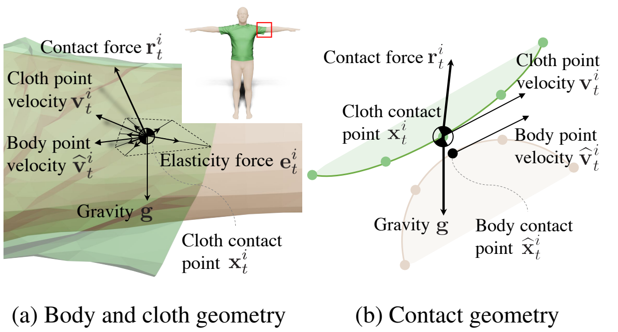
3.2. Body and Cloth Representation

We represent the geometry of the human body using a time-varying SMPL model [23], $\mathcal{M}_t \in \mathbb{R}^{3N_b}$, where N_b is the number of vertices of the SMPL model, and t is the time index. The model is parametrized by a time-varying pose $\boldsymbol{\theta}_t \in \mathbb{R}^{72}$ and global translation $\boldsymbol{t}_t \in \mathbb{R}^3$ and a time-invariant shape $\boldsymbol{\beta} \in \mathbb{R}^{10}$, i.e., $\mathcal{M}_t = \mathcal{M}(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{t}_t)$.

We represent the state of the cloth via its vertex positions $\mathbf{x}_t = \begin{bmatrix} {\mathbf{x}_t^1}^\intercal & \dots & {\mathbf{x}_t^{N_c}}^\intercal \end{bmatrix}^\intercal \in \mathbb{R}^{3N_c}$ and velocities $\mathbf{v}_t = \begin{bmatrix} {\mathbf{v}_t^1}^\intercal & \dots & {\mathbf{v}_t^{N_c}}^\intercal \end{bmatrix}^\intercal \in \mathbb{R}^{3N_c}$, where N_c is the number of cloth vertices. The dynamics of the cloth state is computed using a state-of-the-art cloth simulator [20]. Abstractly, this can be viewed as a discrete-time dynamical system,

$$(\mathbf{x}_t, \mathbf{v}_t) = f(\mathbf{x}_{t-1}, \mathbf{v}_{t-1}; \mathcal{M}_t), \tag{1}$$

since the current state of the cloth geometry is dependent only on the current body state $\mathcal{M}_t$ and the previous cloth state ($\mathbf{x}_{t-1}$ and $\mathbf{v}_{t-1}$). This representation of the cloth dynamics explicitly parametrized by the body state (shape and pose) is the key innovation, which allows us to jointly estimate the cloth and body states that follow the laws of physics.



(a) Body and cloth geometry (b) Contact geometry

Figure 3: The geometry of the body and clothes with physical properties. (a) The external (contact force and gravity) and internal (elasticity) forces are applied at each cloth vertex that drives the movement of the clothes (velocity). (b) A cloth vertex $\mathbf{x}$ is in contact with a point on the body mesh face (not with the body vertex), $\widehat{\mathbf{x}}$. The contact force that is a function of the velocity of the body is computed by weighted average of the body vertices on the contact face.

3.3. Equation of Cloth Motion

To derive the gradients of the cloth motion in Section 3.4, it will be necessary to understand in more detail how the cloth simulator works. We give these details here.

The dynamics of the cloth motion is described using the Newton's second law in generalized coordinates:

$$\mathbf{r}_t(\mathcal{M}_t) + \mathbf{e}_t(\mathbf{x}_t) + \mathbf{g} = \mathbf{M}_t \frac{\Delta \mathbf{v}_t}{\Delta t} \tag{2}$$

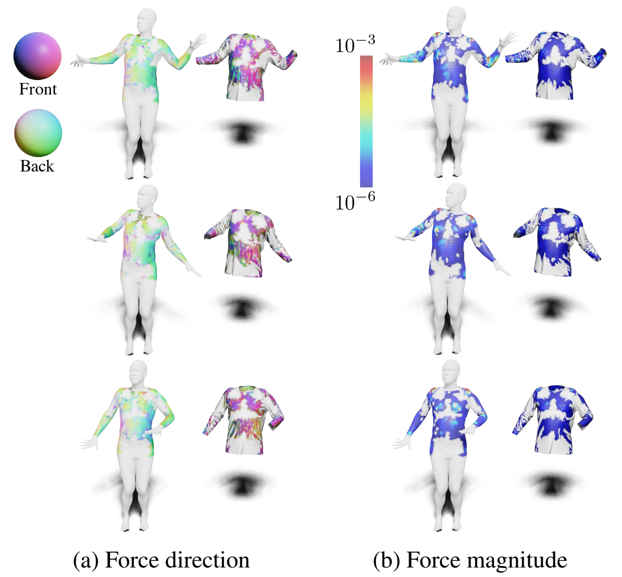
where $\mathbf{r}_t \in \mathbb{R}^{3N_c}$, $\mathbf{e}_t \in \mathbb{R}^{3N_c}$ and $\mathbf{g} \in \mathbb{R}^{3N_c}$ are the external force, internal force induced by elasticity [41], and gravity exerted at the cloth points, respectively. Further, $\mathbf{M}_t = \mathrm{blkdiag}(\{m_t^i \mathbf{I}_3\}_i^{N_c}) \in \mathbb{R}^{3N_c \times 3N_c}$ is the block diagonal mass matrix made of the mass of each point m, with $\mathbf{I}_3$ being the 3×3 identity matrix[1]. Finally, $\Delta \mathbf{v}_t / \Delta t = (\mathbf{v}_t - \mathbf{v}_{t-1})/\Delta t$ is the acceleration of the cloth points over the time step.

The external force $\mathbf{r}_t(\mathcal{M}_t) = \begin{bmatrix} \mathbf{r}_t^{1\mathsf{T}} & \cdots & \mathbf{r}_t^{N_c\mathsf{T}} \end{bmatrix}^\mathsf{T}$ is the contact force exerted by the body $\mathcal{M}$, i.e., $\mathbf{r}_t^i \in \mathbb{R}^3$ is the contact force applied at the i^{th} point at the t^{th} time instant. We employ Signorini–Coulomb law [12] to model the friction force with a non-penetration constraint, i.e., the friction force is equal or less than the normal force with the friction coefficient along the opposite direction of the surface velocity. This nonlinear condition can be expressed as an implicit constraint [38]:

$$\Phi(\mathbf{r}_t^i, \widehat{\mathbf{v}}_t^i) = \mathbf{0}, \tag{3}$$

where $\widehat{\mathbf{v}}_t^i$ is the velocity of the body point $\widehat{\mathbf{x}}_t^i$ that is in contact with the i^{th} cloth point, i.e., $\widehat{\mathbf{v}}_t^i = \mathbf{v}_t^i + (\mathbf{x}_t^i - \widehat{\mathbf{x}}_t^i)/\Delta t$. The full expression and derivation of the implicit constraint can be found in the Supplementary Material.

[1] We assume each vertex of the cloth as a point mass.



(a) Force direction (b) Force magnitude

Figure 4: The movement of the cloth is controlled by the body motion that exerts a force in the form of contacts. Extruded regions of body (e.g., shoulder, belly, and lower arm) are in contact with the cloth. The force for each vertex is measure in Newton (N).

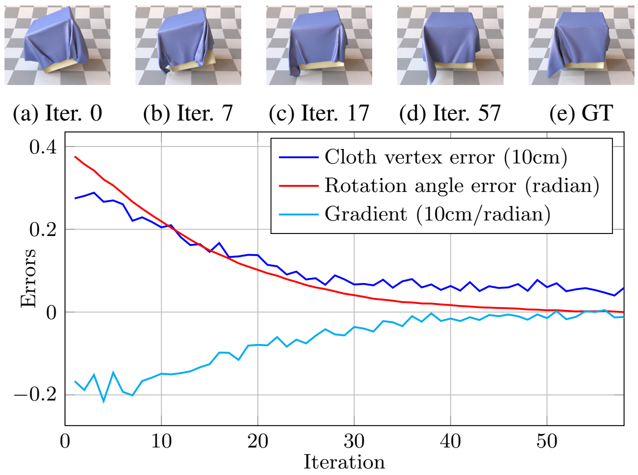
Figure 3 shows the geometry of the cloth and body annotated with the physical properties of the i^{th} cloth point at the t^{th} time instant. The velocity of the cloth mesh vertex $\mathbf{v}_t^i$ is influenced by contact force $\mathbf{r}_t^i$, elasticity force $\mathbf{e}_t^i$ and gravity $\mathbf{g}$. The cloth vertex is in contact with the body mesh face. Note that the contact point on the body is not a body mesh vertex in general. The velocity of the body contact point $\widehat{\mathbf{v}}_t^i$ is computed by the weighted average of the velocities of the body vertices on the contact face based on barycentric weights.

The new cloth state $(\mathbf{x}_t, \mathbf{v}_t)$ is determined by Equations (2) and (3) along with the kinematic relation between position and velocity:

$$\begin{cases} \mathbf{M}_t \left(\frac{\mathbf{v}_t - \mathbf{v}_{t-1}}{\Delta t} \right) = \mathbf{r}_t(\mathcal{M}_t) + \mathbf{e}_t(\mathbf{x}_t) + \mathbf{g}, \\ \Phi(\mathbf{r}_t^i, \widehat{\mathbf{v}}_t^i) = \mathbf{0}, \\ \mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t \Delta t \end{cases} . \tag{4}$$

This is a system of nonlinear equations in three unknowns $\mathbf{x}_t, \mathbf{v}_t, \mathbf{r}_t$. The cloth simulator solves this system using an iterative method; details can be found in [20].

Figure 4 illustrates the contact force on both body and cloth vertices during the cloth simulation, where the force direction and magnitude are visualized. The extruded regions such as shoulder and belly are in contact with the clothes to apply the external forces.



(a) Iter. 0 (b) Iter. 7 (c) Iter. 17 (d) Iter. 57 (e) GT

(f) Error over optimization

Figure 5: We test the feasibility of our method using a toy example of cloth motion on a cube. The control parameter is the rotation angle (1D) of the cube from the horizontal plane. A simulation with 0 rotation angle is used as the ground truth. Our gradient method minimizes the cloth error. (a)-(d) Last frame over iterations. (e) The ground truth. (f) The gradient of the error reduces the cloth vertex error that leads to the rotation error reduction.

3.4. Fitting Cloth Physics via Optimal Control

We formulate the 4D reconstruction of dynamic cloth and body using an optimal control scheme, i.e., the body states are optimized such that the simulated cloth motion should be matched to the point measurements. Given the point measurements $\{\mathcal{X}_t\}$, we minimize the following objective function over the body states $(\boldsymbol{\beta}, \{\boldsymbol{\theta}_t, \mathbf{t}_t\}_{t=0}^{T})$ where T is the total number of time instances:

$$\underset{\boldsymbol{\beta}, \{\boldsymbol{\theta}_t, \mathbf{t}_t\}_{t=0}^{T}}{\text{minimize}} \sum_{t}^{T} L(\mathbf{x}_t, \mathcal{X}_t), \tag{5}$$

where $L = L_c(\mathbf{x}_t, \mathcal{X}_t) + \lambda_s L_b(\mathcal{M}_t, \mathcal{X}_t) + \lambda_t L_t(\boldsymbol{\theta}_t, \mathbf{t}_t)$. The objective function is composed of three terms. (i) $L_c(\mathbf{x}_t, \mathcal{X}_t) = \sum_i^{N_c} D_{\mathcal{X}_t}(\mathbf{x}_t^i)$ measures the cloth reconstruction error where $D_{\mathcal{X}}(\mathbf{x}^i)$ is the Chamfer distance of the i^{th} point to the measurement point set $\mathcal{X}$. This accounts for the reconstruction error that enforces the reconstructed cloth geometry to be matched to the the measurement point set. (ii) $L_b(\mathcal{M}_t, \mathcal{X}_t) = \sum_{\mathbf{y}_t \in \widetilde{\mathcal{M}_t}} D_{\mathcal{X}_t}(\mathbf{y}_t)$ measures the body reconstruction error for the exposed body (e.g., the lower arm surface of whom wearing a short sleeve T-shirt) where $\widetilde{\mathcal{M}}_t$ is the set of points on the exposed body surface. This objective prevents from overfitting of the cloth geometry by ignoring the other body geometry. (iii) $L_t(\boldsymbol{\theta}_t, \mathbf{t}_t) = \|\nabla^2 \boldsymbol{\theta}_t\|^2 + \|\nabla^2 \mathbf{t}_t\|^2$ where ∇^2 is a discrete Laplacian operator. It enforces minimizing time difference of body states, which produces temporally smooth body movement.

We use the gradient decent method to minimize the objective function where the gradient of body states can be obtained by applying the chain rule:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \sum_{t}^{T} \frac{\partial L}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \boldsymbol{\beta}}$$
$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}_t} = \frac{\partial L}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \boldsymbol{\theta}_t}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{t}_t} = \frac{\partial L}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathcal{M}_t} \frac{\partial \mathcal{M}_t}{\partial \boldsymbol{t}_t},$$

where $\mathcal{L} = \sum_{t=0}^{T} L(\mathbf{x}_t, \mathcal{X}_t)$.

Here we have made one notable simplification: we assume that variations in the time-varying body states $(\boldsymbol{\theta}_t, \mathbf{t}_t)$ only affect the current state of the cloth geometry $\mathbf{x}_t$. That is to say, although future cloth states $\mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \ldots$ also depend on the current body state via the cloth dynamics, we ignore this dependence in our optimization since accounting for it would require propagating gradients across the entire duration of the simulation, which is intractable. We justify this approximation by noting that the geometry of a garment is determined to a great extent by the shape and pose of the wearer, especially if the garment is not very loose-fitting and the motion is not highly energetic.

The gradients of the loss terms L_c, L_b, L_t can be trivially derived explicitly. We use the automatic differentiation to compute the body gradient with respect to the shape and pose parameters ($\partial\mathcal{M}_t/\partial\boldsymbol{\beta}$, $\partial\mathcal{M}_t/\partial\boldsymbol{\theta}_t$, and $\partial\mathcal{M}_t/\partial\boldsymbol{t}_t$). The remaining term is the cloth gradient with respect to the body, $\partial\mathbf{x}/\partial\mathcal{M}$. We derive the cloth gradient by differentiating the implicit constraint in Equation (3):

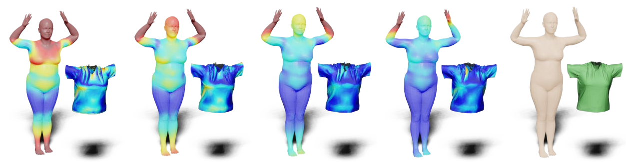
$$\left(\frac{\partial \Phi}{\partial \mathbf{r}_t} \frac{\partial \mathbf{r}_t}{\partial \mathbf{v}_t} + \frac{\partial \Phi}{\partial \widehat{\mathbf{v}}_t} \frac{\partial \widehat{\mathbf{v}}_t}{\partial \mathbf{v}_t}\right) \frac{\partial \mathbf{v}_t}{\partial \mathcal{M}_t} = -\frac{\partial \Phi}{\partial \widehat{\mathbf{v}}_t} \frac{\partial \widehat{\mathbf{v}}_t}{\partial \widehat{\mathbf{x}}_t} \frac{\partial \widehat{\mathbf{x}}_t}{\partial \mathbf{x}_t}. \tag{6}$$

We solve this linear system for $\partial\mathbf{v}_t/\partial\mathcal{M}_t$ and derive the cloth gradient $\partial\mathbf{x}_t/\partial\mathcal{M}_t = \Delta t \partial\mathbf{v}_t/\partial\mathcal{M}_t$.

Figure 5 illustrates the feasibility of our method using a toy example of cloth motion on a cube. The pose of the cube is controlled by one dimensional rotation angle with respect to the horizontal plane. 0 rotation angle is the ground truth. Over the optimization, we minimize the cloth error controlled by the cube rotation, resulting in convergence to the ground truth. The gradient of the rotation angle is negative, which indicates the update direction. Figure 6 is another example illustrating how the body and cloth error changes over optimization

3.5. Body and Cloth Initialization

The objective function in Equation (5) is highly nonlinear, which requires a proper initialization. For the body, we initialize the shape and pose parameters for each frame independently. Given the measured point set $\mathcal{X}$ at each time instant, we find the shape and pose that best describe it by



(a) Iter. 0 (b) Iter. 5 (c) Iter. 10 (d) Iter. 15 (e) GT

Figure 6: The error of body and cloth is decreased over optimization. This example demonstrates the error map of body and clothes over optimization.

minimizing the following objective:

$$\underset{\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{t}}{\text{minimize}} \sum_{i=1}^{N_b} D_{\mathcal{X}}(\mathcal{M}^i), \tag{7}$$

where $\mathcal{M}^i \in \mathbb{R}^3$ is the i^{th} point on the body surface $\mathcal{M}$. Given the time series of the body parameters, we apply a temporal Gaussian filter to produce a smooth parameter trajectory.

For cloth initialization, we ensure the physical validity, i.e., the cloth and body must be interpenetration free. We propose a new cloth relaxation scheme to initialize the cloth by transforming from a canonical state to the state of the first time instant. Consider the body model of canonical pose (e.g., T-pose) $\mathcal{M}_C$. We linearly interpolate between $\mathcal{M}_C$ and $\mathcal{M}_0$ where $\mathcal{M}_0$ is the body pose at the first time instant. In the meantime, we align the position and orientation of a loose cloth with that of the body at T-pose $\mathcal{M}_C$ while ensuring no interpenetration between the body and cloth. With the alignment, we simulate the cloth motion according to the interpolated sequence of the body. We extend the simulation with constant pose $\mathcal{M}_0$ to obtain the steady state cloth geometry at the first time instant (relaxation).

4. Experiments

4.1. Implementation Details

We extend the frictional contact cloth simulator [20] by implementing analytic computation of cloth gradient. The body gradient with respect to shape and pose parameters are computed by automatic differentiation of PyTorch [29]. The parameters for cloth simulation are empirically set based on [20]. We run our experiments on 16-core Intel Haswell E5-2680v3 nodes equipped with 6GB memory in a distributed cluster. The slow computation is a limitation of the proposed method, where the forward simulation constitutes dominant computation (see the Supplementary Material). Potentially, we can significantly reduce the running time by changing the simulator to faster ones [19] that utilize GPU.

4.2. Evaluation Data

Synthetic Data We generate a sequence of SMPL body models from CMU motion capture database. We use the following motion sequences: subject 49 trial 22, subject 55 trial 1, subject 103 trial 2, and subject 131 trial 1. Given the body model (shape and pose), we simulate the cloth motion of T-shirt, shirt, dress, tank and pants from Berkeley Garment Library [27, 41]. The generated body and cloth motions are used as the ground truth, and the sequence of randomly sampled 3D point sets on the clothed human with a Gaussian noise is used for the measurements. To sample the 3D point sets, we first remove the occluded parts of the body which are covered by the cloth. The occluded body vertices are identified by ray casting from the body vertices along their normal direction, where we perform a proximity search for their intersection with the cloth mesh. Once the occluded body parts are removed, we increase the density of the body and cloth vertices by subdivision. Then we add Gaussian noise to the mesh vertices to obtain 3D point sets.

Real Data We show the effectiveness of the proposed method qualitatively on example sequences from BUFF (Bodies Under Flowing Fashion) [46] and HUMBI [45] datasets. BUFF contains high-quality 3D body scan sequences, where the human body and the clothes are represented as a single mesh for each frame. HUMBI consists of multi-view images of human body sequences, from which we reconstruct 3D point cloud of the human body using multi-view stereo (MVS) [35, 36].

4.3. Comparison with Existing Methods

We compare the proposed method to the state-of-the-art human body reconstruction methods, including Tex2Shape [3], PIFuHD [34], IF-Net [11], IP-Net [7] and TailorNet [30]. Note that these methods do not share the same format of input and output. Tex2Shape and PIFuHD take single-view image as input. We provide an input image to these methods by rendering the synthetic human body from the frontal view. TailorNet requires the SMPL shape and pose parameters as input, where we directly provide the ground truth shape and pose parameters. IP-Net produces four different outputs, including a inner surface reconstruction for the undressed body, an SMPL model fitted to the inner surface, an outer surface reconstruction for the clothed body, and an SMPL+D model registered to the outer surface. We denote these models as IP-Net-Inner, IP-Net-SMPL, IP-Net-Outer, and IP-Net-SMPL+D, respectively.

Tex2Shape does not directly produce the reconstruction surface. Instead, it predicts a displacement map for the SMPL body in the UV texture space, which can be applied to an arbitrary SMPL body. We provide the ground truth shape and pose parameters to Tex2Shape to make it comparable. PIFuHD reconstruct the 3D clothed human body in a canonical coordinate space instead of the world coordinate space. We apply ICP (Iterative Closest Point) [6] to

	T-shirt	Shirt	Dress	Tank	Pants
IP-Net [7]-Inner	5.149 ±0.146	4.014 ±0.559	4.164 ±0.550	**2.816 ±0.176**	4.755 ±0.141
IP-Net [7]-SMPL	8.933 ±0.498	3.821 ±1.225	9.930 ±4.083	6.151 ±0.763	11.600 ±0.610
Ours-body	**3.634 ±0.232**	**3.691 ±0.761**	**3.303 ±0.234**	3.583 ±0.322	**3.442 ±0.106**
TailorNet [30]	2.759 ±0.069	**1.974 ±0.173**	N/A	N/A	3.303 ±0.305
Ours-cloth	**2.380 ±0.072**	2.359 ±0.514	2.752 ±0.313	2.135 ±0.320	**2.728 ±0.143**
Tex2Shape [3]	**2.450 ±0.046**	5.282 ±0.408	2.741 ±0.197	5.895 ±0.156	3.164 ±0.236
PIFuHD [34]	5.982 ±0.334	5.019 ±0.669	6.316 ±1.131	6.026 ±1.163	7.771 ±0.903
IF-Net [11]	3.749 ±0.089	2.899 ±0.105	3.048 ±0.150	2.605 ±0.173	3.280 ±0.696
IP-Net [7]-Outer	3.147 ±0.080	2.140 ±0.193	**2.289 ±0.163**	**1.690 ±0.045**	2.492 ±0.162
IP-Net [7]-SMPL+D	3.797 ±0.209	**2.060 ±0.150**	2.783 ±0.410	2.075 ±0.106	3.114 ±0.289
Ours-body/cloth	2.980 ±0.130	2.933 ±0.599	3.014 ±0.190	3.144 ±0.281	**3.086 ±0.091**

Table 1: (White rows) Comparison on body reconstruction (mean±std cm). (Light gray rows) Comparison on cloth reconstruction. Tank and Dress cloth models are not available for TailorNet. (Dark gray rows) Comparison on combined reconstruction of body and cloth. These baseline methods are designed to fit to the data without understanding of physics.

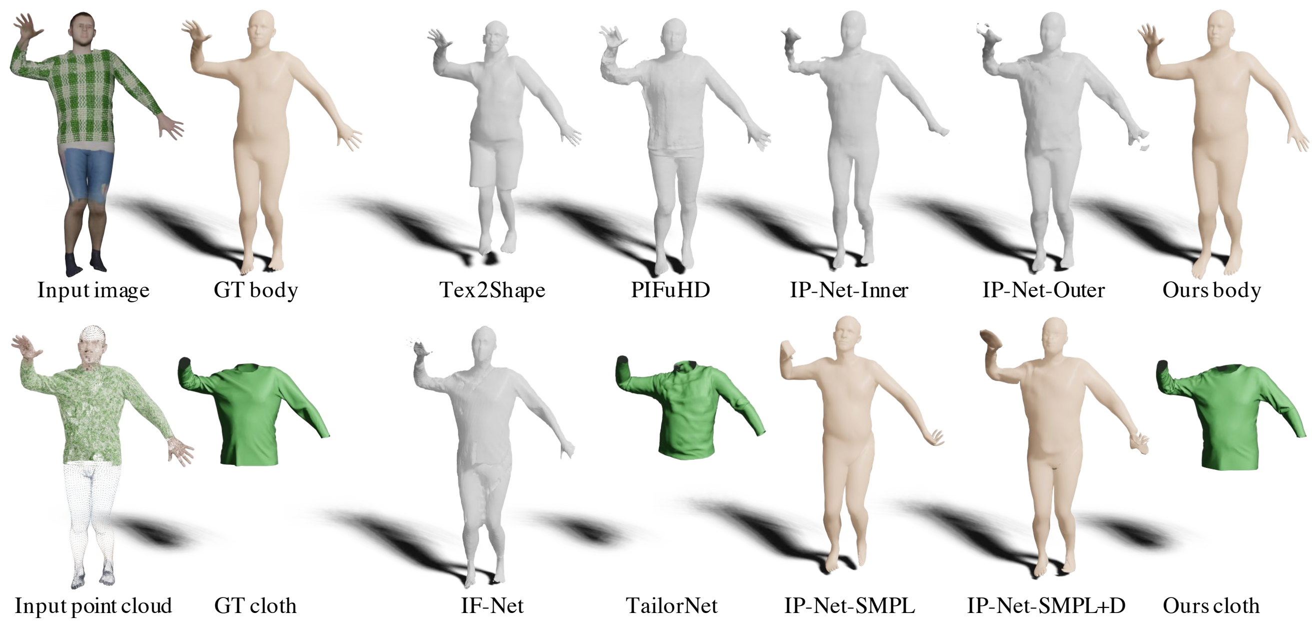
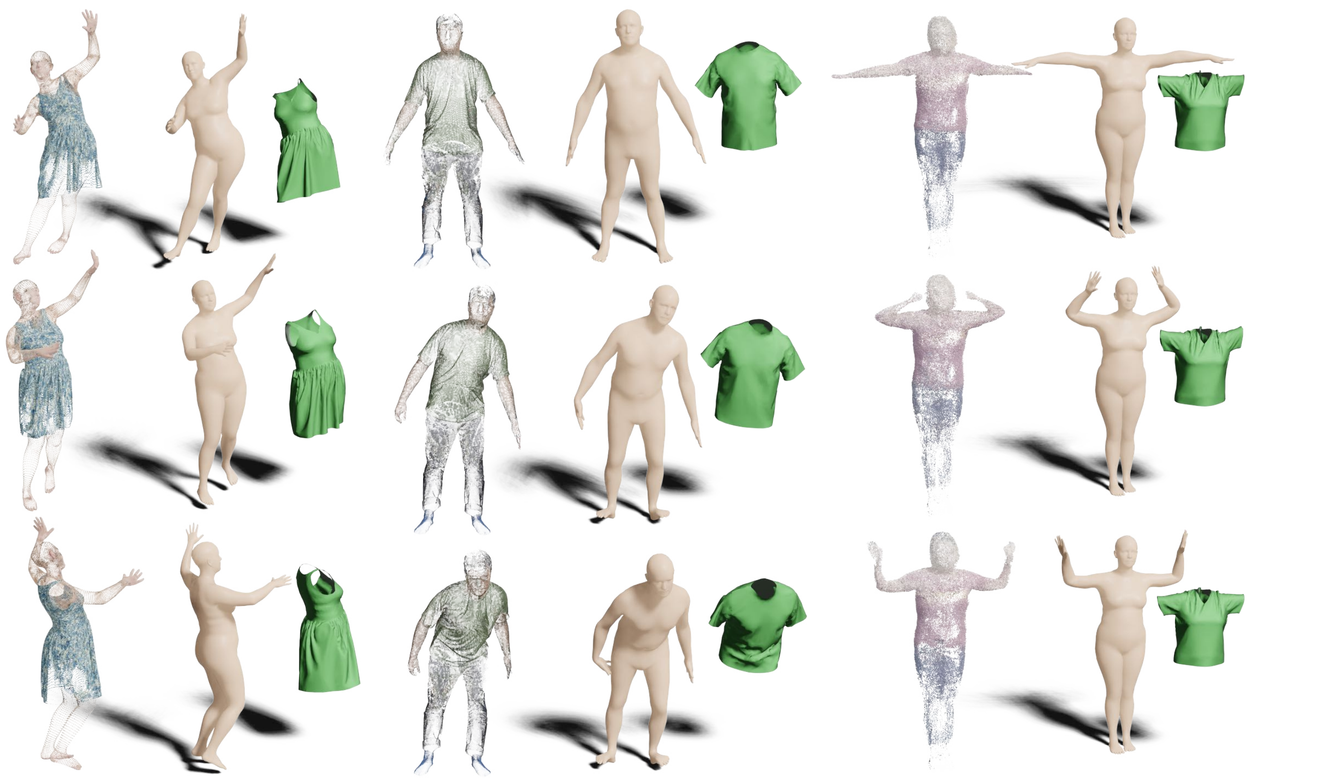


Figure 7: Comparison with the state-of-the-art human body reconstruction methods. The input image is used by Tex2Shape and PIFuHD. Tex2Shape and TailorNet have access to the ground truth body parameters. Other methods only use the input point cloud.

align PIFuHD's reconstruction with the ground truth body for evaluation purpose.

We evaluate the performance of each method using the Chamfer distance between the reconstruction and the ground truth. The error is measured in centimeters. We report the mean and standard deviation of the error over frames. Specifically, we compare the undressed body estimation of our method with IP-Net-Inner, IP-Net-SMPL for the body errors, and report the results in Table 1 (white rows). We compare the cloth reconstruction of our method with TailorNet for the cloth error, and report the results in Table 1 (light gray rows). Our method show strong performance on predicting the underlying body shape. Note that our method takes 3D point sets as input without knowing the ground truth body. However, TailorNet takes ground truth SMPL body parameters as input. Although it does not use any information from the 3D point sets, it has the advantage of knowing part of the ground truth.

For the other methods, we evaluate on the point cloud that represents both cloth and body as summarized in Table 1 (dark gray rows). To make a fair comparison in this setting, we combine our body and cloth results into a single surface by removing the occluded body parts, which are detected by ray casting from body to cloth along vertex nor-



(a) Synthetic sequence (b) BUFF [46] sequence (c) HUMBI [45] sequence

Figure 8: Qualitative results for synthetic data and real-world data.

mal. Note that our method is not always the best, i.e., it is highly competitive but the baseline methods show equal or stronger performance. This is because these baseline methods are designed to fit to the data without any understanding of physics. However, because of static prediction and lack of physics, their resulting motions are unrealistic, temporally jittery, and not physically interpretable (see Figure 7 and Supplementary Materials).

4.4. Qualitative Results

We show more qualitative results of the proposed method on both the synthetic data and real data in Figure 8. These results demonstrate that the proposed method can handle diverse types of clothes on different shapes of bodies. The physically plausible time-varying cloth geometry is faithfully matched to the point sets. BUFF dataset provides high-resolution 4D scan of the clothed human body, where the geometric details like wrinkles and folds are captured. In this case, our method is able to reconstruct the details to some extent, even if the rest shape of the cloth template may not match the actual cloth in the scan. HUMBI dataset provides multi-view capture of a human body without depth information. The point sets in Figure c is reconstructed by MVS from the multi-view images, whose quality is much lower than the BUFF 4D scan. Despite the lack of visible details, our method can generate physically plausible results that are matched to the input point sets.

5. Conclusion

We have presented an inverse simulation method to reconstruct time-varying geometry of the clothes and the underlying body given a time series of 3D point clouds. We model the dynamics of the cloth geometry using a dynamical system that is controlled by the movement history of the body states (shape and pose). Based on the idea of optimal control, we optimize the body state such that the resulting time-varying geometry of the clothes are matched to the 3D point clouds. The analytic gradient of the cloth simulator is derived, which allows us to back-propagate the errors to update the body states. We also proposed a robust cloth initialization scheme to ensure the physical validity of the cloth. Experimental results show that the proposed method can produce physically plausible reconstruction of diverse types of clothes faithful to the real-world measurements.

Acknowledgements This work is supported by NSF CAREER IIS-1846031 and NSF CNS-1919965.

References

[1] T. Alldieck, M. Magnor, B. L. Bhatnagar, C. Theobalt, and G. Pons-Moll. Learning to reconstruct people in clothing from a single rgb camera. In *CVPR*, 2019. 1, 2

[2] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll. Detailed human avatars from monocular video. In *3DV*, 2018. 1, 2

[3] T. Alldieck, G. Pons-Moll, C. Theobalt, and M. Magnor. Tex2shape: Detailed full human body geometry from a single image. In *ICCV*, 2019. 6, 7

[4] M. Athans and P. L. Falb. *Optimal control: an introduction to the theory and its applications*. 2013. 2, 3

[5] D. Baraff and A. Witkin. Large steps in cloth simulation. In *SIGGRAPH*, 1998. 1, 2

[6] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, 1992. 6

[7] B. L. Bhatnagar, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. Combining implicit function learning and parametric models for 3d human reconstruction. In *ECCV*, 2020. 2, 6, 7

[8] B. L. Bhatnagar, G. Tiwari, C. Theobalt, and G. Pons-Moll. Multi-garment net: Learning to dress 3d people from images. In *ICCV*, 2019. 2

[9] S. Bouaziz, S. Martin, T. Liu, L. Kavan, and M. Pauly. Projective dynamics: Fusing constraint projections for fast simulation. *TOG*, 2014. 2

[10] R. Bridson, R. Fedkiw, and J. Anderson. Robust treatment of collisions, contact and friction for cloth animation. 2002. 1, 2

[11] J. Chibane, T. Alldieck, and G. Pons-Moll. Implicit functions in feature space for 3d shape reconstruction and completion. In *CVPR*, 2020. 1, 2, 6, 7

[12] G. Daviet, F. Bertails-Descoubes, and L. Boissieux. A hybrid iterative solver for robustly capturing coulomb friction in hair dynamics. In *SIGGRAPH Asia*, 2011. 4

[13] Q. Guo, X. Han, C. Fu, T. Gast, R. Tamstorf, and J. Teran. A material point method for thin shells with frictional contact. *TOG*, 2018. 3

[14] Y. Hu, J. Liu, A. Spielberg, J. B. Tenenbaum, W. T. Freeman, J. Wu, D. Rus, and W. Matusik. Chainqueen: A real-time differentiable physical simulator for soft robotics. In *ICRA*, 2019. 3

[15] C. Jiang, T. Gast, and J. Teran. Anisotropic elastoplasticity for cloth, knit and hair frictional contact. *TOG*, 2017. 3

[16] H. Joo, T. Simon, and Y. Sheikh. Total capture: A 3d deformation model for tracking faces, hands, and bodies. In *CVPR*, 2018. 1, 2

[17] J. M. Kaldor, D. L. James, and S. Marschner. Simulating knitted cloth at the yarn level. In *SIGGRAPH*, 2008. 3

[18] J. M. Kaldor, D. L. James, and S. Marschner. Efficient yarn-based cloth with adaptive contact linearization. *TOG*, 2010. 3

[19] C. Li, M. Tang, R. Tong, M. Cai, J. Zhao, and D. Manocha. P-cloth: interactive complex cloth simulation on multi-gpu systems using dynamic matrix assembly and pipelined implicit integrators. *TOG*, 2020. 6

[20] J. Li, G. Daviet, R. Narain, F. Bertails-Descoubes, M. Overby, G. E. Brown, and L. Boissieux. An implicit frictional contact solver for adaptive cloth simulation. *TOG*, 2018. 3, 4, 6

[21] J. Liang, M. Lin, and V. Koltun. Differentiable cloth simulation for inverse problems. In *NeurIPS*, 2019. 3

[22] T. Liu, A. W. Bargteil, J. F. O'Brien, and L. Kavan. Fast simulation of mass-spring systems. *TOG*, 2013. 2

[23] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. *TOG*, 2015. 2, 3

[24] M. Ly, J. Jouve, L. Boissieux, and F. Bertails-Descoubes. Projective dynamics with dry frictional contact. *TOG*, 2020. 2

[25] M. Müller, B. Heidelberger, M. Hennix, and J. Ratcliff. Position based dynamics. *Journal of Visual Communication and Image Representation*, 2007. 2

[26] R. Narain, M. Overby, and G. E. Brown. Admm projective dynamics: Fast simulation of general constitutive models. In *SCA*, 2016. 2

[27] R. Narain, A. Samii, and J. F. O'brien. Adaptive anisotropic remeshing for cloth simulation. *TOG*, 2012. 2, 6

[28] M. Overby, G. E. Brown, J. Li, and R. Narain. Admm projective dynamics: Fast simulation of hyperelastic models with dynamic constraints. *TVCG*, 2017. 2

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*. 2019. 6

[30] C. Patel, Z. Liao, and G. Pons-Moll. Tailornet: Predicting clothing in 3d as a function of human pose, shape and garment style. In *CVPR*, 2020. 6, 7

[31] G. Pons-Moll, S. Pujades, S. Hu, and M. J. Black. Clothcap: Seamless 4d clothing capture and retargeting. *TOG*, 2017. 2

[32] X. Provot. Deformation constraints in a mass-spring model to describe rigid cloth behaviour. In *Graphics interface*, 1995. 1, 2

[33] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, and H. Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *ICCV*, 2019. 1, 2

[34] S. Saito, T. Simon, J. Saragih, and H. Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *CVPR*, 2020. 1, 2, 6, 7

[35] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 6

[36] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm. Pixelwise view selection for unstructured multi-view stereo. In *ECCV*, 2016. 6

[37] A. Stomakhin, C. Schroeder, L. Chai, J. Teran, and A. Selle. A material point method for snow simulation. *TOG*, 2013. 3

[38] J. Sun and L. Zhang. A globally convergent method based on fischer–burmeister operators for solving second-order cone constrained variational inequality problems. *Computers & Mathematics with Applications*, 2009. 4

[39] A. P. Tampubolon, T. Gast, G. Klár, C. Fu, J. Teran, C. Jiang, and K. Museth. Multi-species simulation of porous sand and water mixtures. *TOG*, 2017. 3

[40] B. Thomaszewski, M. Wacker, W. Straßer, E. Lyard, C. Luible, P. Volino, M. Kasap, V. Muggéo, and N. Magnenat-Thalmann. Advanced topics in virtual garment simulation. 2007. 2

[41] H. Wang, J. F. O'Brien, and R. Ramamoorthi. Data-driven elastic models for cloth: modeling and measurement. *TOG*, 2011. 4, 6

[42] S. Weiss, R. Maier, D. Cremers, R. Westermann, and N. Thuerey. Correspondence-free material reconstruction using sparse surface constraints. In *CVPR*, 2020. 3

[43] T. Yu, Z. Zheng, K. Guo, J. Zhao, Q. Dai, H. Li, G. Pons-Moll, and Y. Liu. Doublefusion: Real-time capture of human performances with inner body shapes from a single depth sensor. In *CVPR*, 2018. 2

[44] T. Yu, Z. Zheng, Y. Zhong, J. Zhao, Q. Dai, G. Pons-Moll, and Y. Liu. Simulcap: Single-view human performance capture with cloth simulation. In *CVPR*, 2019. 2

[45] Z. Yu, J. S. Yoon, I. K. Lee, P. Venkatesh, J. Park, J. Yu, and H. S. Park. Humbi: A large multiview dataset of human body expressions. In *CVPR*, 2020. 6, 8

[46] C. Zhang, S. Pujades, M. J. Black, and G. Pons-Moll. Detailed, accurate, human shape estimation from clothed 3d scan sequences. In *CVPR*, 2017. 6, 8