SoftSMPL: Data-driven Modeling of Nonlinear Soft-tissue Dynamics for Parametric Humans

Igor Santesteban   Elena Garces   Miguel A. Otaduy   Dan Casas

Universidad Rey Juan Carlos, Madrid, Spain

Figure 1: *Our method regresses soft-tissue dynamics for parametric avatars. Here we see five different body shapes performing a running motion, each of them enriched with soft-tissue dynamics. We depict the magnitude of the regressed displacements using colormaps (right). See supplementary video for full animation.*

Abstract

We present SoftSMPL, a learning-based method to model realistic soft-tissue dynamics as a function of body shape and motion. Datasets to learn such task are scarce and expensive to generate, which makes training models prone to overfitting. At the core of our method there are three key contributions that enable us to model highly realistic dynamics and better generalization capabilities than state-of-the-art methods, while training on the same data. First, a novel motion descriptor that disentangles the standard pose representation by removing subject-specific features; second, a neural-network-based recurrent regressor that generalizes to unseen shapes and motions; and third, a highly efficient nonlinear deformation subspace capable of representing soft-tissue deformations of arbitrary shapes. We demonstrate qualitative and quantitative improvements over existing methods and, additionally, we show the robustness of our method on a variety of motion capture databases.

CCS Concepts

• *Computing methodologies → Animation;*

1. Introduction

Soft-tissue dynamics are fundamental to produce compelling human animations. Most of existing methods capable of generating highly dynamic soft-tissue deformations are based on physics-based approaches. However, these methods are challenging to implement due to the inner complexity of the human body, and the expensive simulation process needed to animate the model. Alter-

natively, data-driven models can potentially learn human soft-tissue deformations as a function of body pose directly from real-world data (*e.g.*, 3D reconstructed sequences). However, in practice, this is a very challenging task due to the highly nonlinear nature of the dynamic deformations, and the scarce of datasets with sufficient reconstruction fidelity.

In this work we propose a novel learning-based method to ani-

mate parametric human models with highly expressive soft-tissue dynamics. SoftSMPL takes as input the shape descriptor of a body and a motion descriptor, and produces dynamic soft-tissue deformations that generalize to unseen shapes and motions. Key to our method is to realize that humans move in a highly personalized manner, *i.e.*, motions are shape and subject dependent, and these subject-dependant features are usually entangled in the pose representation.

Previous methods fail to disentangle body pose from shape and subject features; therefore, they overfit the relationship between tissue deformation and pose, and generalize poorly to unseen shapes and motions. Our method overcomes this limitation by proposing a new representation to disentangle the traditional pose space in two steps. First, we propose a solution to encode a compact and *deshaped* representation of body pose which eliminates the correlation between individual *static* poses and subject. Second, we propose a motion transfer approach, which uses person-specific models to synthesize animations for pose (and style) sequences of other persons. As a result, our model is trained with data where pose and subject-specific *dynamic* features are no longer entangled. We complement this contribution with a highly efficient nonlinear subspace to encode tissue deformations of arbitrary bodies, and a neural-network-based recurrent regressor as our learning-based animation model. We demonstrate qualitative and quantitative improvements over previous methods, as well as robust performance on a variety of motion capture databases.

2. Related work

The 3D modeling of human bodies has been investigated following two main trends: data-driven models, which learn deformations directly from data; and physically-based models, which compute body deformations by solving a simulation problem, usually consisting of a kinematic model coupled with a deformable layer. In this section we discuss both trends, with special emphasis on the former, to which our method belongs.

Data-driven models. Pioneering data-driven models interpolate manually sculpted static 3D meshes to generate new samples [SRC01]. With the appearance of laser scanning technologies, capable of reconstructing 3D static bodies with great level of detail, the data-driven field became popular. Hilton *et al.* [HSC02] automatically fit an skeleton to a static scan to generate animated characters. Allen *et al.* proposed one of the first methods to model upper body [ACP02] and full body [ACCP03] deformations using a shape space learned from static scans and an articulated template. Anguelov *et al.* [ASK*05] went one step further and modeled both shape *and* pose dependent deformations directly from data. Many follow-up data-driven methods have appeared [HSS*09, JTST10, HLRB12, CLZ13, YYZ*14, FCS15, ZB15, LMR*15, PWH*17], but all of these are limited to modeling *static* deformations.

Data-driven models have also been explored to model soft-tissue deformations, which is our main goal too. Initial works used *sparse* marker-based systems to acquire the data. The pioneering work of Park and Hodgins [PH06] reconstructs soft-tissue motion of an actor by fitting a 3D mesh to 350 tracked points. In subsequent work [PH08], they proposed a second-order dynamics model to

synthesize skin deformation as a function of body motion. Similar to us, they represented both body pose and dynamic displacements in a low-dimensional space. However, their method does not generalize to different body shapes. Neumann *et al.* [NVH*13] also used sparse markers to capture shoulder and arm deformations of multiple subjects in a multi-camera studio. They were able to model muscle deformations as a function of shape, pose, and external forces, but their method is limited to the shoulder-arm area, and cannot learn temporal dynamics. Similarly, Loper *et al.* [LMB14] did not learn dynamics either, but they were able to estimate full body pose *and* shape from a small set of motion capture markers. Remarkably, despite their lack of explicit dynamics, their model can reproduce soft-tissue motions by allowing body shape parameters to change over time.

More recently, 3D/4D scanning technologies and mesh registration methods [BPS*08, CBI10, DTF*15, BRPMB17, RCDAT17, PMPHB17] allow to reconstruct high-quality dynamic sequences of human performances. These techniques have paved the way to data-driven methods that leverage *dense* 3D data, usually in the form of temporally coherent 3D mesh sequences, to extract deformation models of 3D humans. Neumann *et al.* [NVW*13] used 3D mesh sequences to learn sparse localized deformation modes, but did not model temporal dynamics. Tsoli *et al.* reconstructed 3D meshes of people breathing [TMB14] in different modes, and built a statistical model of body surface deformations as a function of lung volume. In contrast, we are able to model far more complex deformations, with higher frequency dynamics, as a function of body shape and pose. Casas and Otaduy [CO18] modeled full-body soft-tissue deformations as a function of body motion using a neural-network-based nonlinear regressor. Their model computes per-vertex 3D offsets encoded in an efficient subspace, however it is subject-specific and does not generalize to different body shapes. Closest to our work is Dyna [PMRMB15], a state-of-the-art method that relates soft-tissue deformations to motion and body shape from 4D scans. Dyna uses a second-order auto-regressive model to output mesh deformations encoded in a subspace. Despite its success in modeling surface dynamics, we found that its generalization capabilities to unseen shapes and poses are limited due to the inability to effectively disentangle pose from shape and subject-style . Furthermore, Dyna relies on a linear PCA subspace to represent soft-tissue deformations, which struggles to reproduce highly nonlinear deformations. DMPL [LMR*15] proposes a soft-tissue deformation model heavily-inspired in Dyna, with the main difference that it uses a vertex-based representation instead of triangle-based. However, DMPL suffers from the same limitations of Dyna mentioned above. We also propose a vertex-based representation, which eases the implementation in standard character rigging pipelines, while achieving superior generalization capabilities and more realistic dynamics.

Garment and clothing animation have also been addressed with data-driven models that learn surface deformations as a function of human body parameters [dASTH10, WSFM19]. Guan *et al.* [GRH*12], use linear models to regress wrinkles from pose and shape parameters, but each source of deformation is modeled independently. Similarly, other data-driven methods are limited to represent garment shape variations as linear scaling factors [PM-PHB17, YFHWW18, LCT18] and therefore do not reproduce re-

alistic deformations. Closer to ours is the work of Santesteban *et al.* [SOC19] which effectively disentangles garment deformations due to shape and pose, allowing to train a nonlinear deformation regressor that generalizes to new subjects and motions. Alternatively, Gundogdu *et al.* [GCS*19], extracts geometric features of human bodies to use them to parameterize garment deformations.

Physically-based models. The inherent limitation of data-driven models is their struggle to generate deformations far from the training examples. Physically-based models overcome this limitation by formulating the deformation process within a simulation framework. However, these approaches are not free of difficulties: defining an accurate and efficient mechanical model to represent human motions, and solving the associated simulations is hard.

Initial works used layered representations consisting of a deformable volume for the tissue layer, rigidly attached to a kinematic skeleton [CGC*02, LCA05]. Liu *et al.* [LYWG13] coupled rigid skeletons for motion control with a pose-based plasticity model to enable two-way interaction between skeleton, skin, and environment. McAdams *et al.* [MZS*11] showed skin deformations with a discretization of corotational elasticity on a hexahedral lattice around the surface mesh, but did not run at real-time rates. Xu and Barbič [XB16] used secondary Finite Element Method (FEM) dynamics and model reduction techniques to efficiently enrich the deformation of a rigged character. To speed up simulations, Position-Based Dynamics (PBD) [BMM17] solvers have been widely used for different physics systems, also for human soft tissue [DB13, KB18] and muscle deformation [RMSQ19]. Projective Dynamics, another common approach to accelerate simulations, has also been used for simulating deformable characters [LLK19]. Meanwhile, Pai *et al.* [PRWH*18] presented a novel hand-held device to estimate the mechanical properties of real human soft-tissue.

Subspaces for Simulation. Subspace simulation methods attempt to find a low-dimensional representation of an initial dense set of equations in order to facilitate the computations. PCA has been widely used to this end [KLM01, BJ05, TLP06], while an alternative set of works aims to find new and more efficient bases [TOK14, TMDK15]. For clothing, Hahn *et al.* [HTC*14] built a linear subspace using temporal bases distributed across pose space. Holden *et al.* [HDDN19] also built a linear subspace using PCA and machine learning to model external forces and collisions at interactive rates. Finally, Fulton *et al.* [FMD*19] built a non-linear subspace using an auto-encoder on top of an initial PCA subspace to accelerate the solver.

3. Overview

Our animation model for soft-tissue dynamics takes as input descriptors of body shape and motion, and outputs surface deformations. These deformations are represented as per-vertex 3D displacements of a human body model, described in Section 4.1, and encoded in an efficient nonlinear subspace, described in Section 4.2. At runtime, given body and motion descriptors, we predict the soft-tissue deformations using a novel recurrent regressor proposed in Section 4.3. Figure 2 depicts the architecture of our run-

time pipeline, including the motion descriptor, the regressor, and a soft-tissue decoder to generate the predicted deformations.

In addition to our novel subspace and regressor, our key observation to achieve highly expressive dynamics with unprecedented generalization capabilities is an effective disentanglement of the pose space. In Section 5, we argue and demonstrate that the standard pose space (*i.e.*, vector of joint angles θ) used in previous methods is entangled with subject-specific features. This causes learning-based methods to overfit the relationship between tissue deformation and pose. In Section 5.1 we identify *static* features, mostly due to the particular anatomy of each person, that are entangled in the pose space, and propose a *deshaped* representation to effectively disentangle them. Furthermore, in Section 5.2 we identify *dynamic* features that manifest across a sequence of poses (also known as *style*), and propose a strategy to eliminate them.

4. SoftSMPL

4.1. Human Model

We build our soft-tissue model on top of standard human body models (*e.g.*, [FCS15, LMR*15]) controlled by shape parameters $\beta \in \mathbb{R}^{|\beta|}$ (*e.g.*, principal components of a collection of body scans in rest pose) and pose parameters $\theta \in \mathbb{R}^{|\theta|}$ (*e.g.*, joint angles). These works assume that a deformed body mesh $M(\beta, \theta) \in \mathbb{R}^{3 \times V}$, where V is the number of vertices, is obtained by

$$M(\beta, \theta) = W(T(\beta, \theta), \beta, \theta, \mathcal{W}) \tag{1}$$

where $W(\cdot)$ is a skinning function (*e.g.*, linear blend skinning, dual quaternion, etc.) with skinning weights $\mathcal{W}$ that deforms an unposed body mesh $T(\beta, \theta) \in \mathbb{R}^{3 \times V}$.

Inspired by Loper *et al.* [LMR*15], who obtain the unposed mesh $T(\beta, \theta)$ by deforming a body mesh template $\mathbf{T} \in \mathbb{R}^{3 \times V}$ to incorporate changes in shape $B_s(\beta)$ and pose corrective displacements $B_p(\theta)$, we propose to further deform the body mesh template to incorporate soft-tissue dynamics. More specifically, we define our unposed body mesh as

$$T(\beta, \theta, \gamma) = \mathbf{T} + B_s(\beta) + B_p(\theta) + B_d(\gamma, \beta), \tag{2}$$

where $B_d(\gamma, \beta) = \Delta \in \mathbb{R}^{3 \times V}$ is a soft-tissue regressor that outputs per-vertex displacements required to reproduce skin dynamics given a shape parameter β and a motion descriptor γ. Notice that, in contrast to previous model-based works that also predict soft-tissue displacements [PMRMB15, LMR*15, CO18], our key observation is that such regressing task cannot be formulated directly as function of pose θ (and shape β), because subject-specific information is entangled in that pose space. See Section 5 for a detailed description of our motion descriptor γ and full details on our novel pose disentanglement method.

4.2. Soft-Tissue Representation and Deformation Subspace

We represent soft-tissue deformations Δ as per-vertex 3D offsets of a body mesh $\mathbf{T}$ in an unposed state. This representation allows to isolate the soft-tissue deformation component from other deformations, such as pose or shape.
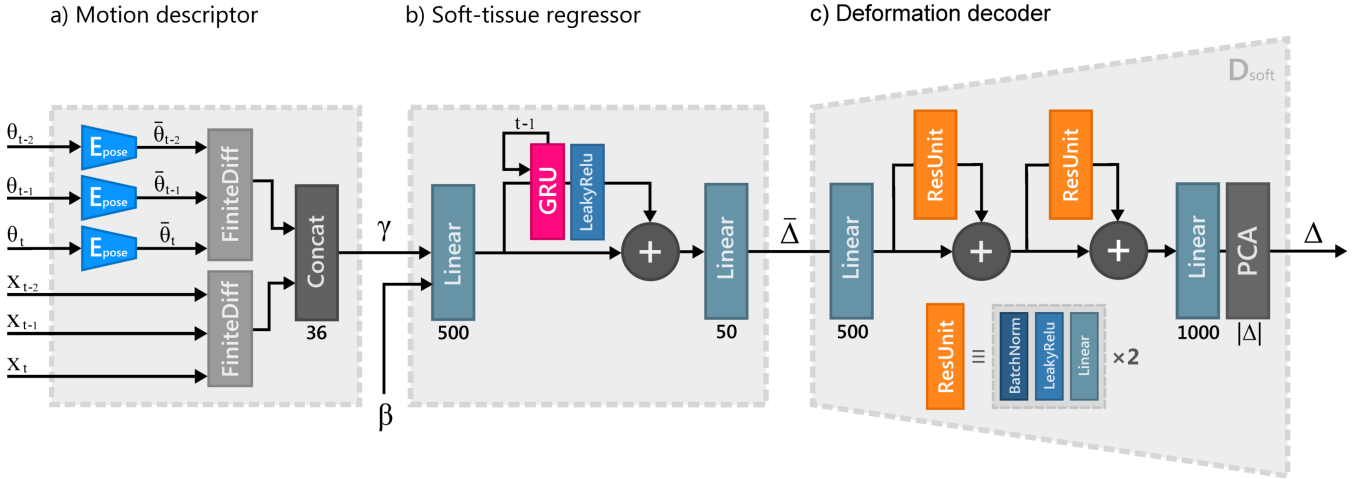
a) Motion descriptor   b) Soft-tissue regressor   c) Deformation decoder

Figure 2: *Runtime pipeline of our approach. First, the temporal motion data is encoded in our novel disentangled pose descriptor. Then, the resulting low dimensional vector is concatenated with the skeleton root offsets to form the motion descriptor. This descriptor along with the desired shape parameters are passed through the soft-tissue regressor, which predicts the nonlinear dynamic behaviour of the soft-tissue deformation in a latent space. Finally, the deformation decoder recovers the original full space of deformation offsets for each vertex of the mesh.*
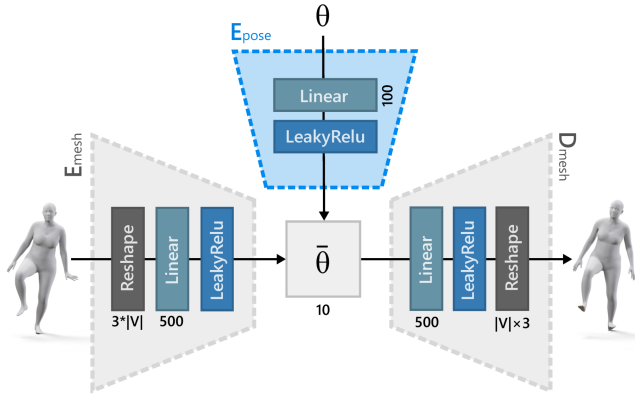

Figure 3: *Architecture of the multi-modal pose autoencoder.*

Given the data-driven nature of our approach, in order to train our model it is crucial that we define a strategy to extract ground truth deformations $\Delta^{GT} \in \mathbb{R}^{3 \times V}$ from real world data. To this end, in a similar spirit to [PMRMB15, LMR*15, PMPHB17], given a dataset $\mathcal{S} = \{\mathbf{S}_t\}_{t=0}^{T-1}$ of 4D scans with temporally consistent topology, we extract the soft-tissue component of each mesh $\mathbf{S} \in \mathbb{R}^{3 \times V}$ as

$$\Delta^{GT} = W^{-1}(\mathbf{S}, \theta, \mathcal{W}) - \mathbf{T} - B_P(\theta) - B_S(\beta), \quad (3)$$

where $W^{-1}(\cdot)$ is the inverse of the skinning function, $B_P(\theta)$ a corrective pose blendshape, and $B_S(\beta)$ a shape deformation blendshape (see [LMR*15] for details on how the latter two are computed). Solving Equation 3 requires estimating the pose θ and shape β parameters for each mesh $\mathbf{S}$, which is a priori unknown (*i.e.*, the dataset $\mathcal{S}$ contains only 3D meshes, no shape or pose parameters).

Similar to [PMRMB15], we solve the optimization problem:

$$\underset{\theta, \beta}{\operatorname{argmin}} ||\mathbf{S} - M(\theta, \beta)||_2 \quad (4)$$

to estimate the shape β and pose θ parameters of each scan $\mathbf{S}$ in the dataset $\mathcal{S}$.

Despite the highly-convenient representation of encoding soft-tissue deformations as per-vertex 3D offsets $\Delta \in \mathbb{R}^{3 \times V}$, this results in a too high-dimensional space for an efficient learning-based framework. Previous works [LMR*15, PMRMB15] use linear dimensionality reduction techniques (*e.g.*, Principal Component Analysis) to find a subspace capable of reproducing the deformations without significant loss of detail. However, soft-tissue deformations are highly nonlinear, hindering the reconstructing capabilities of linear methods. We mitigate this by proposing a novel autoencoder to find an efficient nonlinear subspace to encode soft-tissue deformations of parametric humans.

Following the standard autoencoder pipeline, we define the reconstructed (*i.e.*, encoded-decoded) soft-tissue deformation as

$$\Delta_{rec} = D_{soft}(E_{soft}(\Delta)), \quad (5)$$

where $\bar{\Delta} = E_{soft}(\Delta)$ and $D_{soft}(\bar{\Delta})$ are encoder and decoder networks, respectively, and $\bar{\Delta} \in \mathbb{R}^{|\bar{\Delta}|}$ soft-tissue displacements projected into the latent space. We train our deformation autoencoder by using a loss function $\mathcal{L}_{rec}$ that minimizes both surface and normal errors between input and output displacements as follows

$$\mathcal{L}_{surf} = ||\Delta - \Delta_{rec}||_2 \quad (6)$$

$$\mathcal{L}_{norm} = \frac{1}{F} \sum_{f=1}^{F} ||1 - N_f(\Delta) \cdot N_f(\Delta_{rec})||_1 \quad (7)$$

$$\mathcal{L}_{rec} = \mathcal{L}_{surf} + \lambda_{norm} \mathcal{L}_{norm} \quad (8)$$

where F is the number of faces of the mesh template, $N_f(\Delta)$ the

normal of the fth face, and λ_{norm} is set to 1000. Notice that, during training, we use ground truth displacements Δ^{GT} from a variety of characters which enables us to find a subspace that generalizes well to encode soft-tissue displacements of *any* human shape. This is in contrast to previous works [CO18] that need to train shape-specific autoencoders.

We implement the encoder E_{soft} and decoder D_{soft} using a fully-connected neural network architecture composed of several residual units [HZRS16b] . Inspired by the work of Fulton *et al.* [FMD*19], we initialize the first and last layers of the autoencoder with weights computed using PCA, which eases the training of the network. In Figure 2 (right) we depict the decoder D_{soft}. The encoder E_{soft} uses an analogous architecture.

4.3. Soft-Tissue Recurrent Regressor

In this section we describe the main component of our runtime pipeline: the soft-tissue regressor R, illustrated in Figure 2 (center). Assuming a motion descriptor γ (which we discuss in detail in Section 5.1) and a shape descriptor β, our regressor outputs the predicted soft tissue displacements $\bar{\Delta}$. These encoded displacements are subsequently fed into the decoder D_{soft} to generate the final per-vertex 3D displacements

$$\Delta = D_{\text{soft}}(R(\gamma, \beta)). \quad (9)$$

To learn the naturally nonlinear dynamic behavior of soft-tissue deformations, we implement the regressor R using a recurrent architecture GRU [CvMG*14]. Recurrent architectures learn which information of previous frames is relevant and which not, resulting in a good approximation of the temporal dynamics. This is in contrast to modeling temporal dependencies by explicitly adding the output of one step as the input of the next step, which is prone to instabilities specially in nonlinear models. Furthermore, our regressor also uses a residual shortcut connection to skip the GRU layer altogether, which improves the flow of information [HZRS16a]. We initialize the state of the GRU to zero at the beginning of each sequence.

We train the regressor R by minimizing a loss $\mathcal{L}_{\text{reg}}$, which enforces predicted vertex positions, velocities, and accelerations to match the latent space deformations $\bar{\Delta}$,

$$\mathcal{L}_{\text{reg}} = \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{vel}} + \mathcal{L}_{\text{acc}} \quad (10)$$

5. Disentangled Motion Descriptor

To efficiently train the soft-tissue regressor $R(\gamma, \beta)$, described earlier in Section 4.3, we require a pose-disentangled and discriminative motion descriptor γ. To this end, in this section we propose a novel motion descriptor. It encompasses the velocity and acceleration of the body root in world space X, a novel pose descriptor $\bar{\theta}$, and the velocity and acceleration of this novel pose descriptor, as follows:

$$\gamma = \{\bar{\theta}, \frac{d\bar{\theta}}{dt}, \frac{d^2\bar{\theta}}{dt^2}, \frac{dX}{dt}, \frac{d^2X}{dt^2}\}. \quad (11)$$

In the rest of this section we discuss the limitation of the pose descriptors used in state-of-the-art human models, and introduce

a new disentangled space $\bar{\theta}$ to remove *static* subject-specific features (Section 5.1). Moreover, we also propose a strategy to remove *dynamic* subject-specific features (Section 5.2) from sequences of poses.

5.1. Static Pose Space Disentanglement

The regressor R proposed in Section 4.3 relates body motion and body shape to soft-tissue deformations. To represent body motion, a standard parameterization used across many human models [FCS15, ASK*05, LMB14, LMR*15] is the joint angles of the kinematic skeleton, θ. However, our key observation is that this pose representation is entangled with shape- and subject-specific information that hinders the learning of a pose-dependent regressor. Additionally, Hahn *et al.* [HTC*14] also found that using joint angles to represent pose leads to a high-dimensional space with redundancies, which makes the learning task harder and prone to overfitting. We hypothesize that existing data-driven parametric human models are less sensitive to this entanglement and overparameterization because they learn simpler deformations with much more data. In contrast, we model soft-tissue with a limited dataset of 4D scans, which requires a well disentangled and discriminative space to avoid overfitting tissue deformation and pose. Importantly, notice that removing these features manually is not feasible, not only because of the required time, but also because these features are not always apparent to a human observer.

We therefore propose a novel and effective approach to *deshape* the pose coefficients, *i.e.*, to disentangle subject-specific anatomical features into a normalized and low-dimensional pose space $\bar{\theta}$:

$$\bar{\theta} = E_{\text{pose}}(\theta). \quad (12)$$

We find $E_{\text{pose}}(\theta) \in \mathbb{R}^{|\bar{\theta}|}$ by training a multi-modal encoder-decoder architecture, shown in Figure 3. In particular, having a mesh scan $\mathbf{S}$ and its corresponding pose θ and shape β parameters (found by solving Equation 4), we simultaneously train two encoders and one decoder minimizing the loss

$$\mathcal{L} = \|M(\theta, \mathbf{0}) - D_{\text{mesh}}(E_{\text{mesh}}(M(\theta, \mathbf{0}))\|_2 \\ + \|M(\theta, \mathbf{0}) - D_{\text{mesh}}(E_{\text{pose}}(\theta)\|_2, \quad (13)$$

where $M(\theta, \mathbf{0})$ are the surface vertices of a skinned mesh in pose θ and mean shape (*i.e.*, vector of shape coefficients is zero). The intuition behind this multi-modal autoencoder is the following: the encoder E_{mesh} takes as input *skinned vertices* to enforce the similarity of large deformations (*e.g.*, lifting arms, where many vertices move) in the autoencoder loss. By using a significantly small latent space, we are able to simultaneously train it with the encoder E_{pose} such that the latter learns to remove undesired local pose articulations (and keep global deformations) directly in the pose vector θ. In contrast, notice that without the loss term that uses E_{mesh} we would not be able to distinguish between large and small deformations, because in the pose parameterization space of θ all parameters (*i.e.*, degrees of freedom) contribute equally.

The effect of the encoder E_{pose} is depicted in Figure 4, where subject- and shape-specific features are effectively removed, producing a *normalized* pose. In other words, we are disentangling
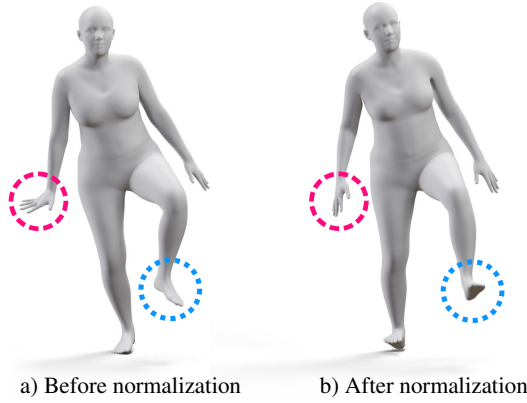
a) Before normalization   b) After normalization

Figure 4: *Result after static pose disentanglement. Our approach effectively removes subject- and shape-dependent features, while retaining the main characteristics of the input pose. See supplementary material for a visualisation of the pose disentanglement across a sequence.*

features originally present in the pose descriptor θ (*e.g.*, wrist articulation) that are related to that particular subject or shape, but we are keeping the overall pose (*e.g.*, raising left leg).

We found 10 to be an appropriate size of the latent space for a trade-off between capturing subtle motions and removing subject-specific features.

5.2. Avoiding Dynamic Pose Entanglement

The novel pose representation $\bar{\theta}$ introduced earlier effectively disentangles *static* subject-specific features from the naive pose representation θ, however, our motion descriptor γ also takes temporal information (velocities and accelerations) into account. We observe that such temporal information can encode *dynamic* shape- and subject-specific features, causing an entanglement potentially making our regressor prone to overfitting soft-tissue deformations to subject-specific pose dynamics.

We address this by extending our 4D dataset by transferring sequences (encoded using our motion descriptor) across the different subjects. In particular, given two sequences of two different subjects

$$\mathcal{S}_A^i = \{\mathbf{S}_{A,t}^i(\theta_t^i)\}_{t=0}^{N_A^i} \quad (14)$$

$$\mathcal{S}_B^j = \{\mathbf{S}_{B,t}^j(\theta_t^j)\}_{t=0}^{N_B^j} \quad (15)$$

where $\mathbf{S}_{A,t}^i(\theta_t^i)$ is the mesh of the subject A performing the sequence identity i at time t, we transfer the sequence of poses θ_t^i to a subject B by training a subject-specific regressor R_B. This process generates a new sequence

$$\mathcal{S}_B^i = R_B(\gamma_A^i) = \{\mathbf{S}_{B,t}^i(\theta_t^i)\}_{t=0}^{N_A^i} \quad (16)$$

with the shape identity of the subject B performing the motion θ_t^i (notice, a motion originally performed by subject A). By transferring all motions across all characters, we are enriching our dataset

in a way that effectively avoids overfitting soft-tissue deformations to subject and shape-specific dynamics (*i.e.*, style).

In Section 7 we detail the number of sequences and frames that we transfer, and evaluate the impact of this strategy. Specifically, Figure 8 shows an ablation study on how the generalization capabilities of our method improve when applying the pose disentangling methods introduced in this section.

6. Datasets, Networks and Training

In this section we provide details about the datasets, network architectures, and parameters to train our models.

6.1. Soft-tissue Autoencoder and Regressor

Data. Our soft-tissue autoencoder and soft-tissue regressor (Section 4.3) are trained using the 4D sequences provided in the Dyna dataset [PMRMB15]. This dataset contains highly detailed deformations of registered meshes of 5 female subjects performing a total of 52 dynamic sequences captured at 60fps (42 used for training, 6 for testing). Notice that we do not use the Dyna provided meshes directly, but preprocess them to *unpose* the meshes. To this end, we solve Equation 4 for each mesh, and subsequently apply Equation 3 to find the ground truth displacements for all Dyna meshes.

Moreover, in addition to the motion transfer technique described in Section 5.2, we further synthetically augment the dataset by mirroring all the sequences.

Setup. We implement all networks in TensorFlow, including the encoder-decoder architecture of E_{soft} and D_{soft}, and the R regressor. We also leverage TensorFlow and its automatic differentiation capabilities to solve Equation 4. In particular, we optimize β using the first frame of a sequence and then optimize θ while leaving β constant. We use Adam optimizer with a learning rate of 1e-4 for the autoencoder and 1e-3 for the regressor. The autoencoder is trained during 1000 epochs (around 3 hours) with a batch size of 256, and a dropout rate of 0.1. The regressor is trained during 100 epochs (around 25 minutes) with batch size of 10, and no dropout. The details of the architecture are shown in Figure 2.

6.2. Pose Autoencoder

Data. To train our pose autoencoder presented in Section 5.1 we are not restricted to the data of 4D scans because we do not need dynamics. We therefore leverage the SURREAL dataset [VRM*17], which contains a vast amount of Motion Capture (MoCap) sequences, from different actors, parameterized by pose representation θ. Our training data consists of 76094 poses from a total of 298 sequences and 56 different subjects, including the 5 subjects of the soft-tissue dataset (excluding the sequences used for testing the soft-tissue networks).

Setup. We use Adam optimizer with a learning rate of 1e-3, and a batch size of 256, during 20 epochs (20 min). The details of the architecture are shown in Figure 3.
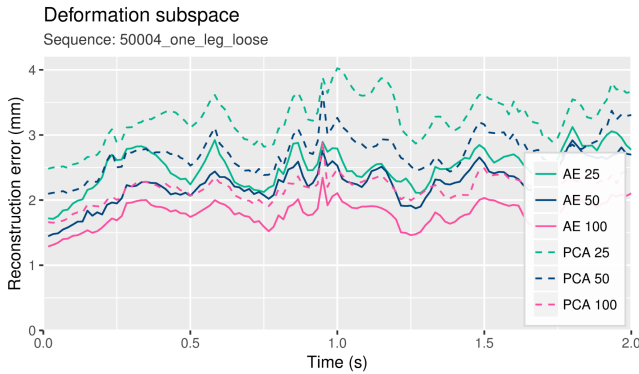
Figure 5: *Soft-tissue autoencoder quantitative evaluation*

Figure 6: *Reconstruction errors of our soft-tissue autoencoder and PCA, for two different body shapes. Notice that our subspace efficiently encodes soft-tissue displacements for parametric shapes, in contrast to previous works [CO18] that required an autoencoder per subject.*

7. Evaluation

In this section we provide qualitative and quantitative evaluation of both the reconstruction accuracy of our soft-tissue deformation subspace, described in Section 4.2, and the regressor proposed in Section 4.3.

7.1. Soft-tissue Autoencoder Evaluation

Quantitative Evaluation. Figure 5 shows a quantitative evaluation of the reconstruction accuracy of the proposed nonlinear autoencoder (AE) for soft-tissue deformation, for a variety of subspace sizes. We compare it with linear approaches based on PCA used in previous works [LMR*15, PMRMB15], in a test sequence (*i.e.*, not used for training). Furthermore, Table 1 shows the reconstruction error in the full test dataset. These results demonstrate that our autoencoder consistently outperforms the reconstruction accuracy of the subspaces used in previous methods.

	25D	50D	100D
PCA	3.82mm	3.17mm	2.38mm
AE	**3.02mm**	**2.58mm**	**2.09mm**

Table 1: *Reconstruction error of our soft-tissue autoencoder and PCA evaluated in the full test dataset. The autoencoder (AE) performs better than the linear approach (PCA) in all tested subspace sizes.*

Qualitative Evaluation. Figure 6 depicts a qualitative evaluation of the soft-tissue deformation autoencoder for a variety of subspace dimensions. Importantly, we also show that the reconstruction accuracy is attained across different shapes. The realism of the autoencoder is better appreciated in the supplementary video, which includes highly dynamic sequences reconstructed with our approach.

7.2. Soft-tissue Regressor Evaluation

We follow a similar evaluation protocol as in Dyna [PMRMB15], and evaluate the following scenarios to exhaustively test our
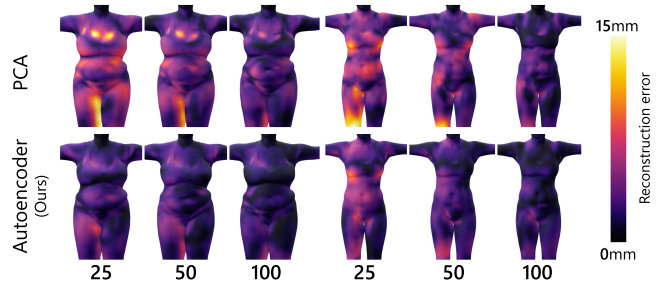
method. Additionally, we provide novel quantitative insights that demonstrate significantly better generalization capabilities of our regression approach with respect to existing methods.

Generalization to New Motions. In Figure 7 and in the supplementary video we demonstrate the generalization capabilities of our method to unseen motions. In particular, at train time, we left out the sequence `one_leg_jump` of the Dyna dataset, and then use our regressor to predict soft-tissue displacements for this sequence, for the shape identity of the subject 50004. Leaving ground truth data out at train time allows us to quantitatively evaluate this scenario. To this end, we also show a visualization of the magnitude of soft-tissue displacement for both ground truth Δ^{GT} and regressed Δ displacements, and conclude that the regressed values closely match the ground truth.

Additionally, in the supplementary video we show more test sequences of different subjects from the Dyna dataset animated with MoCap sequences from the CMU dataset [VRM*17]. Notice that for these sequences there is no ground truth soft-tissue available (*i.e.*, actors were captured in a MoCap studio, only recording joint positions). Our animations show realistic and highly expressive soft-tissue dynamics that match the expected deformations for different body shapes.

Generalization to New Subjects. We quantitatively evaluate the generalization capabilities to new subjects by looking at the magnitude of the predicted soft-tissue displacements for different body shapes. Intuitively, subjects with larger body mass (*i.e.*, more fat), which map to the smaller $\beta[1]$ parameters, should exhibit larger soft-tissue velocities. In contrast, thin subjects, which maps to mostly positive values in $\beta[1]$, should exhibit much lower soft-tissue velocities due to the high rigidity of their body surface. We exhaustively evaluate this metric in Figure 8, where we show an ablation study comparing our full method, our method trained with each of the contributions alone, and Dyna. Although Dyna [PMRMB15] produces different deformation modes for different subjects, the resulting motion is significantly attenuated. In contrast, our full model (in pink) regresses a higher dynamic range of deformations, outputting larger deformations for small values of $\beta[1]$ (*i.e.*, fat subjects), and small surface velocities for larger values of
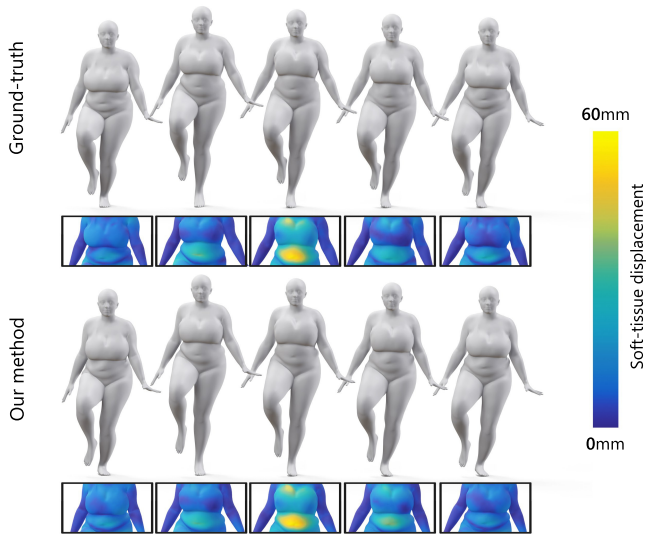
Figure 7: *Evaluation of generalization to new motions. The sequence* `one_leg_jump` *was left out at train time, and used only for testing, for subject 50004. We show ground truth meshes and vertex displacements* Δ^{GT} *(top), and the regressed deformations* Δ *(bottom). Notice how the magnitude of the regressed displacement closely matches the ground truth.*

$\beta[1]$ (*i.e.*, thin subjects). Importantly, we show that each contribution of our model (the static and dynamic pose disentangling methods introduced in Section 5) contributes to our final results, and that all together produce the highest range of deformations.

In the supplementary video we further demonstrate our generalization capabilities. We also show an interactive demo where the user can change the shape parameters of an avatar in real-time, and our method produces the corresponding and very compelling soft tissue deformation.

Generalization to New Motion and New Subject. We finally demonstrate the capabilities of our model to regress soft-tissue deformations for new body shapes and motions. To this end, we use MoCap data from SURREAL and AMASS datasets [VRM*17, MGT*19] and arbitrary body shape parameters. Figure 9 shows sample frames of sequences `01_01` and `09_10` for two different shapes. Colormaps on 3D meshes depict per-vertex magnitude regressed offsets to reproduce soft-tissue dynamics. As expected, frames with more dynamics exhibit larger deformations. Please see the supplementary video for more details.

7.3. Runtime performance

We have implemented our method on a regular desktop PC equipped with an AMD Ryzen 7 2700 CPU, a Nvidia GTX 1080 GPU, and 32GB of RAM. After training the model, we use TensorRT [NVI18] to optimize the neural networks for faster inference at runtime. On average, a forward pass of the optimized model takes 4.8ms. This cost is distributed across the components of the
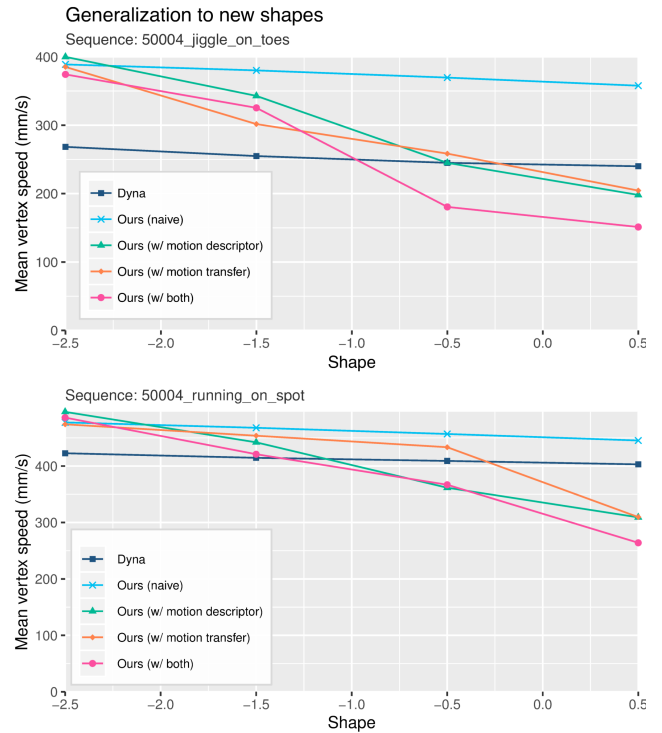

Figure 8: *We quantitatively evaluate the generalization to new shapes of our regressor by looking at the mean vertex speed of the predicted soft-tissue offsets in unposed state in two test sequences. Our model (pink) produces a higher range of dynamics, with large velocities for fat subjects (shape parameter -2.5) and small velocities for thin subjects (shape parameter 0.5). In contrast, previous works (Dyna, in dark blue) produce a much smaller range, resulting in limited generalization capabilities to new subjects. Furthermore, here we also demonstrate that all components of our method contribute to getting the best generalization capabilities.*

model as follows: 0.6ms the pose encoder, 1.9ms the soft-tissue regressor and 2.3ms the soft-tissue decoder.

8. Conclusions

We have presented SoftSMPL, a data-driven method to model soft-tissue deformations of human bodies. Our method combines a novel motion descriptor and a recurrent regressor to generate per-vertex 3D displacements that reproduce highly expressive soft-tissue deformations. We have demonstrated that the generalization capabilities of our regressor to new shapes and motions significantly outperform existing methods. Key to our approach is to realize that traditional body pose representations rely on an entangled space that contains static and dynamic subject-specific features. By proposing a new disentangled motion descriptor, and a novel subspace and regressor, we are able to model soft-tissue deformations as a function of body shape and pose with unprecedented detail.

Despite the significant step forward towards modeling soft-tissue dynamics from data, our method suffers for the following limita-
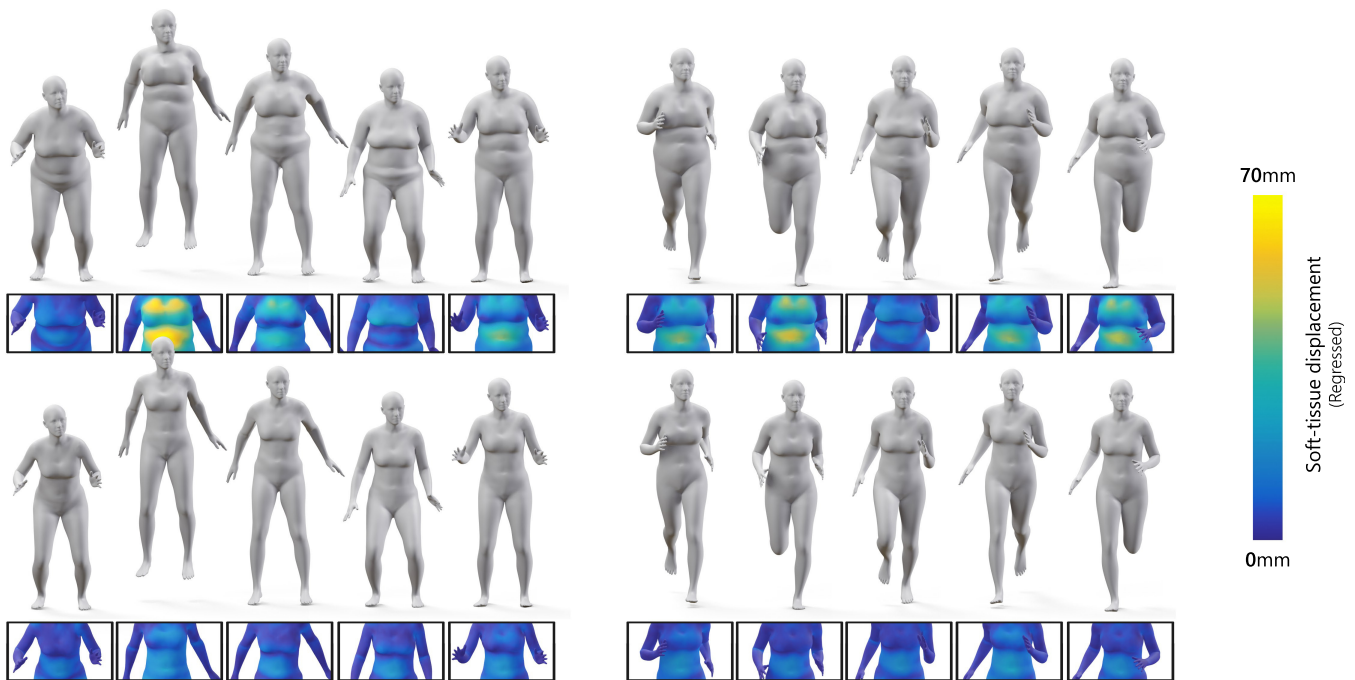
Figure 9: *Sample frames of soft-tissue regression on two test sequences and two test subjects. Colormap depicts the magnitude of the regressed deformation. Notice how our method successfully regresses larger deformations on highly dynamic poses such as in the middle of a jump or when a foot steps on the ground. See supplementary video for full animation and more examples.*

tions. With the current 4D datasets available, which contain very few subjects and motions, it is not feasible to learn a model for a high-dimensional shape space. Furthermore, subtle motions that introduce large deformations are also very difficult to reproduce. Finally, as in most data-driven methods, our model cannot interact with external objects and does not support different topologies. Physics-based models can handle arbitrary meshes and react to external forces [KPMP*17, KIL*16, KB18], but they come at the expense of significantly higher computational cost.

Our approach to static pose disentanglement depends on compression, which is not always reliable and requires choosing an appropriate size for the pose space. Since the dataset contains several subjects performing similar motions, future works could make use of this information to find more robust ways to disentangle pose from static subject features.

References

[ACCP03] ALLEN B., CURLESS B., CURLESS B., POPOVIĆ Z.: The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics (Proc. SIGGRAPH) 22*, 3 (July 2003), 587–594. doi:10.1145/882262.882311. 2

[ACP02] ALLEN B., CURLESS B., POPOVIĆ Z.: Articulated body deformation from range scan data. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3 (2002), 612–619. doi:10.1145/566654.566626. 2

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: SCAPE: Shape completion and animation of people. *ACM Transactions on Graphics (TOG) 24*, 3 (2005), 408–416. doi:10.1145/1073204.1073207. 2, 5

[BJ05] BARBIČ J., JAMES D. L.: Real-time subspace integration for st. venant-kirchhoff deformable models. *ACM Transactions on Graphics (TOG) 24*, 3 (2005), 982–990. doi:10.1145/1073204.1073300. 3

[BMM17] BENDER J., MÜLLER M., MACKLIN M.: A survey on position based dynamics. In *Eurographics Tutorials* (2017). doi:10.2312/egt.20171034. 3

[BPS*08] BRADLEY D., POPA T., SHEFFER A., HEIDRICH W., BOUBEKEUR T.: Markerless garment capture. *ACM Transactions on Graphics (Proc. SIGGRAPH) 27*, 3 (2008), 99. doi:10.1145/1360612.1360698. 2

[BRPMB17] BOGO F., ROMERO J., PONS-MOLL G., BLACK M. J.: Dynamic FAUST: Registering human bodies in motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2017). doi:10.1109/CVPR.2017.591. 2

[CBI10] CAGNIART C., BOYER E., ILIC S.: Probabilistic deformable surface tracking from multiple videos. In *European con-*

ference on computer vision (2010), pp. 326–339. doi:10.1007/978-3-642-15561-1_24. 2

[CGC*02] CAPELL S., GREEN S., CURLESS B., DUCHAMP T., POPOVIĆ Z.: Interactive skeleton-driven dynamic deformations. *ACM Transactions on Graphics (Proc. SIGGRAPH) 21*, 3 (2002), 586–593. doi:10.1145/566654.566622. 3

[CLZ13] CHEN Y., LIU Z., ZHANG Z.: Tensor-based human body modeling. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2013), pp. 105–112. doi:10.1109/CVPR.2013.21. 2

[CO18] CASAS D., OTADUY M. A.: Learning nonlinear soft-tissue dynamics for interactive avatars. *Proc. of the ACM on Computer Graphics and Interactive Techniques 1*, 1 (2018), 10. doi:10.1145/3203187. 2, 3, 5, 7

[CvMG*14] CHO K., VAN MERRIËNBOER B., GULCEHRE C., BAHDANAU D., BOUGARES F., SCHWENK H., BENGIO Y.: Learning phrase representations using RNN encoder–decoder for statistical machine translation. doi:10.3115/v1/D14-1179. 5

[dASTH10] DE AGUIAR E., SIGAL L., TREUILLE A., HODGINS J. K.: Stable spaces for real-time clothing. *ACM Transactions on Graphics (TOG) 29*, 4 (July 2010), 106:1–106:9. doi:10.1145/1778765.1778843. 2

[DB13] DEUL C., BENDER J.: Physically-based character skinning. In *Workshop on Virtual Reality Interaction and Physical Simulation* (2013), Bender J., Dequidt J., Duriez C., Zachmann G., (Eds.), The Eurographics Association. doi:10.2312/PE.vriphys.vriphys13.025-034. 3

[DTF*15] DOU M., TAYLOR J., FUCHS H., FITZGIBBON A., IZADI S.: 3D scanning deformable objects with a single RGBD sensor. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 493–501. doi:10.1109/CVPR.2015.7298647. 2

[FCS15] FENG A., CASAS D., SHAPIRO A.: Avatar reshaping and automatic rigging using a deformable model. In *Proc. of the 8th ACM SIGGRAPH Conference on Motion in Games* (2015), pp. 57–64. doi:10.1145/2822013.2822017. 2, 3, 5

[FMD*19] FULTON L., MODI V., DUVENAUD D., LEVIN D. I. W., JACOBSON A.: Latent-space dynamics for reduced deformable simulation. *Computer Graphics Forum (Proc. Eurographics)* (2019). doi:10.1111/cgf.13645. 3, 5

[GCS*19] GUNDOGDU E., CONSTANTIN V., SEIFODDINI A., DANG M., SALZMANN M., FUA P.: GarNet: A two-stream network for fast and accurate 3D cloth draping. In *Proc. International Conference on Computer Vision (ICCV)* (2019). URL: http://arxiv.org/abs/1811.10983. 3

[GRH*12] GUAN P., REISS L., HIRSHBERG D. A., WEISS A., BLACK M. J.: DRAPE: Dressing any person. *ACM Transactions on Graphics (TOG) 31*, 4 (2012), 35:1–35:10. doi:10.1145/2185520.2185531. 2

[HDDN19] HOLDEN D., DUONG B. C., DATTA S., NOWROUZEZAHRAI D.: Subspace neural physics: Fast data-driven interactive simulation. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)* (2019), pp. 6:1–6:12. doi:10.1145/3309486.3340245. 3

[HLRB12] HIRSHBERG D. A., LOPER M., RACHLIN E., BLACK M. J.: Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In *Proc. of European Conference on Computer Vision* (2012), pp. 242–255. doi:10.1007/978-3-642-33783-3_18. 2

[HSC02] HILTON A., STARCK J., COLLINS G.: From 3D shape capture to animated models. In *IEEE Conference on 3D Data Processing, Visualisation and Transmission* (2002), pp. 246–255. doi:10.1109/TDPVT.2002.1024069. 2

[HSS*09] HASLER N., STOLL C., SUNKEL M., ROSENHAHN B., SEIDEL H.-P.: A statistical model of human pose and body shape. *Com-*

puter Graphics Forum 28, 2 (2009), 337–346. doi:10.1111/j.1467-8659.2009.01373.x. 2

[HTC*14] HAHN F., THOMASZEWSKI B., COROS S., SUMNER R. W., COLE F., MEYER M., DEROSE T., GROSS M.: Subspace clothing simulation using adaptive bases. *ACM Transactions on Graphics (Proc. SIGGRAPH) 33*, 4 (2014), 105:1–105:9. doi:10.1145/2601097.2601160. 3, 5

[HZRS16a] HE K., ZHANG X., REN S., SUN J.: Deep residual learning for image recognition. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778. doi:10.1109/CVPR.2016.90. 5

[HZRS16b] HE K., ZHANG X., REN S., SUN J.: Identity mappings in deep residual networks. In *European Conference on Computer Vision (ECCV)* (2016), pp. 630–645. doi:10.1007/978-3-319-46493-0_38. 5

[JTST10] JAIN A., THORMÄHLEN T., SEIDEL H.-P., THEOBALT C.: Moviereshape: Tracking and reshaping of humans in videos. *ACM Trans. Graph. 29*, 6 (2010), 148:1–148:10. doi:10.1145/1882261.1866174. 2

[KB18] KOMARITZAN M., BOTSCH M.: Projective skinning. *Proceedings of the ACM on Computer Graphics and Interactive Techniques 1*, 1 (2018), 12:1–12:19. doi:10.1145/3203203. 3, 9

[KIL*16] KADLEČEK P., ICHIM A.-E., LIU T., KŘIVÁNEK J., KAVAN L.: Reconstructing personalized anatomical models for physics-based body animation. *ACM Trans. Graph. 35*, 6 (2016), 213:1–213:13. doi:10.1145/2980179.2982438. 9

[KLM01] KRYSL P., LALL S., MARSDEN J. E.: Dimensional model reduction in non-linear finite element dynamics of solids and structures. *International Journal for numerical methods in engineering 51*, 4 (2001), 479–504. doi:10.1002/nme.167. 3

[KPMP*17] KIM M., PONS-MOLL G., PUJADES S., BANG S., KIM J., BLACK M. J., LEE S.-H.: Data-driven physics for human soft tissue animation. *ACM Transactions on Graphics 36*, 4 (2017), 54:1–54:12. doi:10.1145/3072959.3073685. 9

[LCA05] LARBOULETTE C., CANI M.-P., ARNALDI B.: Dynamic skinning: adding real-time dynamic effects to an existing character animation. In *Proc. of Spring Conference on Computer graphics* (2005), ACM, pp. 87–93. doi:10.1145/1090122.1090138. 3

[LCT18] LÄHNER Z., CREMERS D., TUNG T.: DeepWrinkles: Accurate and realistic clothing modeling. In *European Conference on Computer Vision (ECCV)* (2018). doi:10.1007/978-3-030-01225-0_41. 2

[LLK19] LI J., LIU T., KAVAN L.: Fast simulation of deformable characters with articulated skeletons in projective dynamics. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)* (2019), pp. 1:1–1:10. doi:10.1145/3309486.3340249. 3

[LMB14] LOPER M., MAHMOOD N., BLACK M. J.: MoSh: Motion and shape capture from sparse markers. *ACM Trans. Graph. 33*, 6 (2014), 220:1–220:13. doi:10.1145/2661229.2661273. 2, 5

[LMR*15] LOPER M., MAHMOOD N., ROMERO J., PONS-MOLL G., BLACK M. J.: SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 34*, 6 (2015), 248:1–248:16. doi:10.1145/2816795.2818013. 2, 3, 4, 5, 7

[LYWG13] LIU L., YIN K., WANG B., GUO B.: Simulation and control of skeleton-driven soft body characters. *ACM Transactions on Graphics 32*, 6 (2013), 215:1–215:8. doi:10.1145/2508363.2508427. 3

[MGT*19] MAHMOOD N., GHORBANI N., TROJE N. F., PONS-MOLL G., BLACK M. J.: AMASS: Archive of motion capture as surface shapes. *CoRR abs/1904.03278* (2019). URL: http://arxiv.org/abs/1904.03278, arXiv:1904.03278. 8

[MZS*11] MCADAMS A., ZHU Y., SELLE A., EMPEY M., TAMSTORF R., TERAN J., SIFAKIS E.: Efficient elasticity for character skinning with contact and collisions. *ACM Transactions on Graphics (Proc. SIGGRAPH) 30*, 4 (July 2011), 37:1–37:12. doi:10.1145/2010324.1964932. 3

[NVH*13] NEUMANN T., VARANASI K., HASLER N., WACKER M., MAGNOR M., THEOBALT C.: Capture and statistical modeling of arm-muscle deformations. *Computer Graphics Forum 32*, 2 (2013), 285–294. doi:10.1111/cgf.12048. 2

[NVI18] NVIDIA CORPORATION: TensorRT: Programmable inference accelerator, 2018. URL: https://developer.nvidia.com/tensorrt. 8

[NVW*13] NEUMANN T., VARANASI K., WENGER S., WACKER M., MAGNOR M., THEOBALT C.: Sparse localized deformation components. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia) 32*, 6 (Nov. 2013). doi:10.1145/2508363.2508417. 2

[PH06] PARK S. I., HODGINS J. K.: Capturing and animating skin deformation in human motion. *ACM Transactions on Graphics (TOG) 25*, 3 (2006), 881–889. doi:10.1145/1141911.1141970. 2

[PH08] PARK S. I., HODGINS J. K.: Data-driven modeling of skin and muscle deformation. *ACM Transactions on Graphics (TOG) 27*, 3 (2008), 96:1–96:6. doi:10.1145/1360612.1360695. 2

[PMPHB17] PONS-MOLL G., PUJADES S., HU S., BLACK M. J.: ClothCap: Seamless 4D clothing capture and retargeting. *ACM Transactions on Graphics (Proc. of SIGGRAPH) 36*, 4 (2017), 73:1–73:15. doi:10.1145/3072959.3073711. 2, 4

[PMRMB15] PONS-MOLL G., ROMERO J., MAHMOOD N., BLACK M. J.: Dyna: A model of mic human shape in motion. *ACM Trans. Graph. 34*, 4 (2015), 120:1–120:14. doi:10.1145/2766993. 2, 3, 4, 6, 7

[PRWH*18] PAI D. K., ROTHWELL A., WYDER-HODGE P., WICK A., FAN Y., LARIONOV E., HARRISON D., NEOG D. R., SHING C.: The human touch: Measuring contact with real human soft tissues. *ACM Transactions on Graphics (Proc. SIGGRAPH) 37*, 4 (2018), 58:1–58:12. doi:10.1145/3197517.3201296. 3

[PWH*17] PISCHULIN L., WUHRER S., HELTEN T., THEOBALT C., SCHIELE B.: Building statistical shape spaces for 3D human modeling. *Pattern Recognition 67* (2017), 276–286. doi:10.1016/j.patcog.2017.02.018. 2

[RCDAT17] ROBERTINI N., CASAS D., DE AGUIAR E., THEOBALT C.: Multi-view performance capture of surface details. *International Journal of Computer Vision (IJCV) 124*, 1 (Aug 2017), 96–113. doi:10.1007/s11263-016-0979-1. 2

[RMSQ19] ROMEO M., MONTEAGUDO C., SÁNCHEZ-QUIRÓS D.: Muscle and fascia simulation with extended position based dynamics. *Computer Graphics Forum* (2019). doi:10.1111/cgf.13734. 3

[SOC19] SANTESTEBAN I., OTADUY M. A., CASAS D.: Learning-based animation of clothing for virtual try-on. *Computer Graphics Forum (Proc. Eurographics) 38*, 2 (2019), 355–366. doi:10.1111/cgf.13643. 3

[SRC01] SLOAN P.-P. J., ROSE C. F., COHEN M. F.: Shape by example. In *Proc. of Symposium on Interactive 3D Graphics (I3D)* (2001), pp. 135–143. doi:10.1145/364338.364382. 2

[TLP06] TREUILLE A., LEWIS A., POPOVIĆ Z.: Model reduction for real-time fluids. In *ACM Transactions on Graphics (TOG)* (2006), vol. 25, ACM, pp. 826–834. doi:10.1145/1141911.1141962. 3

[TMB14] TSOLI A., MAHMOOD N., BLACK M. J.: Breathing life into shape: Capturing, modeling and animating 3D human breathing. *ACM Transactions on Graphics (Proc. of SIGGRAPH) 33*, 4 (2014), 52:1–52:11. doi:10.1145/2601097.2601225. 2

[TMDK15] TENG Y., MEYER M., DEROSE T., KIM T.: Subspace condensation: full space adaptivity for subspace deformations. *ACM Transactions on Graphics (TOG) 34*, 4 (2015), 76. doi:10.1145/2766904. 3

[TOK14] TENG Y., OTADUY M. A., KIM T.: Simulating articulated subspace self-contact. *ACM Transactions on Graphics (TOG) 33*, 4 (2014), 106. doi:10.1145/2601097.2601181. 3

[VRM*17] VAROL G., ROMERO J., MARTIN X., MAHMOOD N., BLACK M. J., LAPTEV I., SCHMID C.: Learning from synthetic humans. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 109–117. doi:10.1109/CVPR.2017.492. 6, 7, 8

[WSFM19] WANG T. Y., SHAO T., FU K., MITRA N. J.: Learning an intrinsic garment space for interactive authoring of garment animation. *ACM Transactions on Graphics (TOG) 38*, 6 (Nov. 2019). doi:10.1145/3355089.3356512. 2

[XB16] XU H., BARBIČ J.: Pose-space subspace dynamics. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 35:1–35:14. doi:10.1145/2897824.2925916. 3

[YFHWW18] YANG J., FRANCO J.-S., HÉTROY-WHEELER F., WUHRER S.: Analyzing clothing layer deformation statistics of 3D human motions. In *Proc. of the European Conference on Computer Vision (ECCV)* (2018), pp. 237–253. doi:10.1007/978-3-030-01234-2_15. 2

[YYZ*14] YANG Y., YU Y., ZHOU Y., DU S., DAVIS J., YANG R.: Semantic parametric reshaping of human body models. In *International Conference on 3D Vision (3DV)* (2014), vol. 2, IEEE, pp. 41–48. doi:10.1109/3DV.2014.47. 2

[ZB15] ZUFFI S., BLACK M. J.: The stitched puppet: A graphical model of 3D human shape and pose. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 3537–3546. doi:10.1109/CVPR.2015.7298976. 2