SFLSH: Shape-Dependent Soft-Flesh Avatars

Pablo Ramón
Universidad Rey Juan Carlos
Madrid, Spain
pablo.ramon@urjc.es

Cristian Romero
Universidad Rey Juan Carlos
Madrid, Spain
crisrom002@gmail.com

Javier Tapia
Universidad Rey Juan Carlos
Madrid, Spain
fratapmer@gmail.com

Miguel A. Otaduy
Universidad Rey Juan Carlos
Madrid, Spain
miguel.otaduy@urjc.es

Figure 1: Simulated characters in a fitness class. The scene consists of 25 physics-based soft-body avatars, each one with different soft-flesh properties automatically estimated from their shape. The scene runs at 1.2 fps single-threaded.

ABSTRACT

We present a multi-person soft-tissue avatar model. This model maps a body shape descriptor to heterogeneous geometric and mechanical parameters of a soft-tissue model across the body, effectively producing a shape-dependent parametric soft avatar model. The design of the model overcomes two major challenges, the potential redundancy of geometric and mechanical parameters, and the complexity to obtain abundant subject data, which together induce major risk of overfitting the resulting model. To overcome these challenges, we introduce a local shape-dependent regularization of the model. We demonstrate accurate results, on par with independent per-subject estimation, accurate interpolation within the range of body shapes of the training subjects, and good generalization to unseen body shapes. As a result, we obtain a parametric soft-flesh avatar model easy to integrate in many existing applications.

CCS CONCEPTS

• **Computing methodologies → Physical simulation**.

KEYWORDS

Avatars, soft-tissue simulation

1 INTRODUCTION

Motion and deformation of the human body are of high interest in many computer-graphics applications, including capture and reconstruction [Bai et al. 2022; Işık et al. 2023], VFX [Smith et al. 2018], interactive avatar simulation [Komaritzan et al. 2021], or virtual try-on of clothing [Harrison et al. 2018]. A key element that has facilitated fast progress of research on avatar-related challenges is the availability of parametric body models [Anguelov et al. 2005; Loper et al. 2015], accounting for both pose and shape variation in a compact manner, as they simplify the application of research methods to very large populations of subjects. However, to date, parametric body models do not include a similarly compact parameterization of soft-tissue mechanics. As a result, parametric body models do not exhibit soft flesh deformation induced by body dynamics, and do not deform under contact.

In parallel, there have been efforts to characterize soft-flesh properties of humans and estimate soft-tissue simulation models [Kim et al. 2017; Pai et al. 2018; Romero et al. 2020]. Unfortunately, due to the lack of multi-person soft-flesh models, these efforts must be

reproduced for each new subject under study. Soft-tissue characterization and estimation is a complex and time-consuming process, not free of ethical challenges either.

In this paper, we present a multi-person soft-tissue avatar model, built on top of existing parametric body models, and trained from dynamics data of real subjects. Given a subject's shape descriptor, we produce a personalized physics-based model of soft flesh, covering both skin geometry and mechanical parameters. Notably, the model is heterogeneous across the body and the tissue parameters depend on local body shape, not just global shape. Thanks to this locality, the model can be trained with sparse data, i.e., measurements from just a few subjects.

In the design of the multi-person soft-flesh model we faced two important challenges. One is the high potential redundancy of geometric and mechanical parameters, e.g., the effective stiffness of a nonlinear tissue model can be modulated via both Young's modulus and tissue thickness. The other one is the scarcity of training data, due to the complexity to obtain personalized mechanical response data of the whole body. Together, these two challenges lead to strong risk of overfitting and the difficulty to design a parametric multi-person model.

The key component of our model is a mapping from a shape descriptor to heterogeneous soft-tissue parameters. In Section 4, we present a local shape-based regularization of the mapping, which avoids overfitting and provides robust generalization to unseen body shapes. Our soft-tissue model is built on top of the SMPL parametric body model [Loper et al. 2015], as described in Section 3, making it highly accessible for many applications. Moreover, we build our model on top of a reduced-order physics simulation approach [Tapia et al. 2021], to maximize performance. To estimate our soft-tissue model, we fit dynamic body motion of the DYNA dataset [Pons-Moll et al. 2015], as discussed in Section 5.

We demonstrate quantitatively and qualitatively the accuracy of the proposed model. We show that the results are as accurate as per-subject independent estimation of soft-flesh models. We also show that the model succeeds to accurately interpolate within the shape range of the training subjects, thanks to the robust interpretation of local shape. And finally, we show that the model generalizes well to shapes beyond the range of the training subjects, as in Figure 1. An open-source implementation of our soft-flesh multi-person avatar model can be found in https://gitlab.com/PabloRamonPrieto/flsh/-/tree/SIGAsia.

2 RELATED WORK

Our shape-dependent soft avatar model relies on an underlying parametric body model. SCAPE [Anguelov et al. 2005] set a breakthrough in body models, by finding statistical correlations in both shape and pose deformations for a large population of subjects. Later, SMPL [Loper et al. 2015] gained major adoption, due to its accuracy and amenity to graphics pipelines. Same as SCAPE, it parameterizes shape and pose separately. Shape is modeled as a linear deformation of a template mesh plus a correction to joint locations, and pose is modeled as an additional correction to the template mesh followed by linear blend skinning. In our work, we have used SMPL as the reference parametric body model. SMPL-X [Pavlakos et al. 2019] extends SMPL to include detailed face and

hand models. Deep learning methodologies have enabled a transition from explicit parametric models to neural models based on autoencoder architectures with a latent parameterization [Foti et al. 2022]. Alternatively, neural fields enable the animation of body dynamics without the need to fit a template to subject data [Bai et al. 2022]. The very recent work of Zheng et al. [2023] combines a template parametric model with an overlayed neural field for accurate representation. We build our soft avatar model on top of an explicit parametric body model, SMPL, but it would be possible to extend to neural latent models as long as the influence on local shape can be quantified.

In the simulation of body dynamics, the simplest models look only at the articulated or skeletal body structure. This is a common representation for character animation and control [Lee et al. 2019]. On top of the basic skeletal structure, multiple authors have worked on creating anatomically inspired models, including volumetric muscles [Fan et al. 2014], or recently even a full spine model with soft intervertebral disks [Lee et al. 2023]. As an intermediate approach, Komaritzan et al. [2021] introduced a layered model that can be personalized.

A different line of work has paid attention to the simulation of skin deformation and dynamics, by connecting a soft-tissue layer to skeletal motion [Liu et al. 2013; McAdams et al. 2011]. Unfortunately, adding detailed soft-flesh deformation increases the simulation cost by orders of magnitude in contrast to a pure skeletal simulation. As a middle-ground, to incorporate soft-flesh effects while retaining some of the speed of skeletal motion, the deformation of the soft flesh can be represented using reduced deformation models [Tapia et al. 2021; Xu and Barbič 2016]. In our work, we adopt this reduced simulation approach. A different approach for efficiency is the use of fast projective dynamics [Li et al. 2019].

Some works animate soft-tissue dynamics in a data-driven manner [Casas and Otaduy 2018; Pons-Moll et al. 2015; Santesteban et al. 2020]. The method of Zheng et al. [2021] learns how to generate secondary motion that is even transferable to other characters. These methods are very fast and accurate for skeletal dynamics, but they cannot support general deformation as produced by contact.

We opt for a physics-based simulation method that can generalize to arbitrary interactions. While data-driven methods estimate non-physical models, physics-based methods require the estimation of the parameters of a mechanical model.

For the estimation of soft-tissue mechanical parameters, we can distinguish different approaches. One approach uses dynamic motion data and maximizes the similarity of oscillatory dynamics [Kim et al. 2017; Romero et al. 2020]. Another approach uses skin compression profiles and maximizes the similarity of force-deformation profiles [Pai et al. 2018]. Outside computer graphics, the characterization of soft-tissue mechanics is of high interest for biomedical applications [Nava et al. 2008]. In our work, we estimate mechanical parameters from dynamic motion data, relying on the publically available DYNA data set [Pons-Moll et al. 2015].

3 BASELINE SOFT AVATAR MODEL

In this section, we describe the high-level design of a soft avatar model, including a discussion of design choices made in our implementation. These cover the underlying skeletal parametric avatar
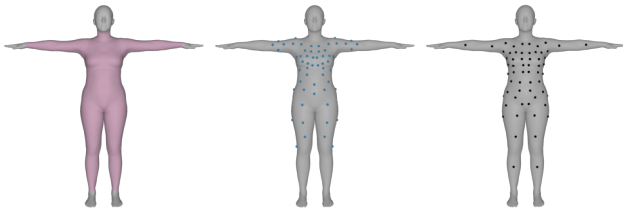
Figure 2: Some of the design settings of the soft avatar model. From left to right: (i) template body, showing the part modeled as soft flesh in pink, (ii) handles of the reduced model in blue, (iii) control points of the shape-based elasticity parameterization in black.

model, the formulation of continuum elasticity, its coupling with the skeletal model, the elasticity material model, spatial discretization, and reduced-order modeling for efficient simulation. Readers interested in our shape-dependent parameterization of elasticity may hop to Section 4. For a more detailed formulation of the simulation model, on the other hand, please see the supplementary document.

Let us consider a template body $\bar{\mathcal{X}}_0$ in rest or undeformed configuration, with surface $\bar{\mathcal{S}}_0 = \partial \bar{\mathcal{X}}_0$. The template body is transformed through a sequence of three transformations:

(1) Shape, which modifies the rest geometry in a personalized manner. Given a shape descriptor β, the rest geometry becomes $\bar{\mathcal{X}}(\bar{\mathcal{X}}_0, \beta)$.

(2) Elasticity, which applies a deformation due to mechanical effects such as contact or inertia. As a result of elastic deformation, a rest point $\bar{x} \in \bar{\mathcal{X}}$ transforms into a deformed point $x(\bar{x}) \in \mathcal{X}$, with $\mathcal{X}$ the deformed body.

(3) Pose, which accounts for the skeletal transformation, typically implemented through skinning. As several other works [Tapia et al. 2021; Xu and Barbič 2016], in our model we apply the elastic deformation first, and then we pose this deformation through skinning. Given a pose descriptor θ, the posed body becomes $\mathcal{X}_\theta(\mathcal{X}, \theta)$.

We use SMPL [Loper et al. 2015] as baseline parametric body model. In SMPL, the shape transformation is defined through PCA of a large set of body scans, and the shape descriptor $\beta \in \mathbb{R}^{10}$ corresponds to the weights of the principal components. The pose transformation is defined by joint angles and linear blend skinning, together with pose-based corrections added to the unposed geometry. Default SMPL is only defined on the body's surface, but it can be extended to the body's volume using a smooth correspondence mapping from the surface of the template avatar $\bar{\mathcal{S}}_0$ to the skeleton [Romero et al. 2020].

We model a soft-tissue layer on the torso, the upper arms and the thighs, as shown in Figure 2. We formalize the soft-tissue model using continuum mechanics and a general hyperelasticity formulation. Given a deformation gradient $F(\bar{x})$ and heterogeneous (i.e., spatially varying) mechanical parameters $\kappa(\bar{x})$, we can generally define a strain energy density function $\Psi(F(\bar{x}), \kappa(\bar{x}))$. This energy density is integrated on the whole body (using FEM on a tetrahedral mesh) to obtain the total elastic energy $E = \int_{\bar{\mathcal{X}}} \Psi(\bar{x}) \, d\bar{x}$. In our implementation, we compute the deformation gradient as $F = \frac{\partial x}{\partial \bar{x}}$, without
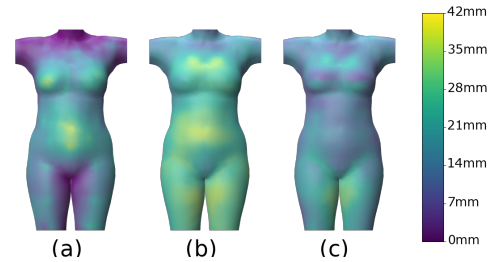

Figure 3: Dynamics of (a) the baseline soft avatar model vs. (b) the ground-truth DYNA data [Pons-Moll et al. 2015], and (c) the difference. The colormap shows the standard deviation of per-vertex amplitudes of deformations over all train subjects and train motions. The deformations are in the shape's neutral pose, before skinning, i.e., world-space deformations are pulled back with inverse skinning to the neutral pose.

the transformation introduced by pose. As shown by Romero et al. [2020], in this way the soft avatar model reproduces exactly the static poses of the SMPL data-driven model, yet elasticity captures deformations due to contact and dynamics. Moreover, we pick as constitutive model Ψ an orthotropic Saint Venant-Kirchhoff model with Fung-type saturation, which captures important anisotropy and nonlinearity aspects of skin [Romero et al. 2020]. We define three heterogeneous and personalized parameters in the model: skin thickness h (defined as a fraction of the maximum possible thickness between surface and skeleton), tangential Young modulus Y_t, and normal Young modulus Y_n. Conversely, we use uniform values for Poisson's ratio (= 0.2) and Fung saturation (= 4). It would be possible to use other material models and parameterizations, such as the one by Pai et al. [2018]. Our model also includes inertia, measured using velocities of the posed body $\mathcal{X}_\theta$. In our experiments, we execute dynamic simulations following an incremental-potential formulation of backward Euler integration [Kane et al. 2000].

In our implementation, we use a reduced-order approximation of full-space elasticity. In particular, we follow a handle-based model [Tapia et al. 2021; Wang et al. 2015], which uses as handles 24 skeletal bones together with 86 point handles distributed throughout the body, as shown in Figure 2.

This reduced model provides a seamless and efficient enhancement over the "rigid" SMPL model, yet at a fraction of the cost of a full-space simulation. In practice, the deformation x is encoded by a set of reduced degrees of freedom (DoFs) q, and elastic forces on these DoFs can be computed as $f = -\frac{\partial E}{\partial q} = -\int_{\bar{\mathcal{X}}} \frac{\partial \Psi}{\partial F} \frac{\partial F}{\partial q} \, d\bar{x}$. We approximate the integral with a data-oblivious cubature approach as proposed by Tapia et al. [2021].

We use a common discretization (of the volumetric tetrahedral mesh, the handles of the reduced model, and cubature points) for all bodies, computed on the undeformed template body $\bar{\mathcal{X}}_0$, at average skin thickness. To transform the discretization for an arbitrary skin thickness h and body shape β, we compute the displacements of the outer and inner surfaces, and apply Laplacian interpolation in between.

We also discretize the spatially-varying elasticity parameterization (skin thickness and Young moduli) on the surface of the

Table 1: Top row: The bodies of the DYNA dataset [Pons-Moll et al. 2015], used for training. Bottom row: New bodies used for testing. For each body, we show the 10 values of the β shape descriptor, ordered according to their importance in the SMPL model.

DYNA 1		DYNA 2		DYNA 3		DYNA 4		DYNA 5	
-0.38	-2.28	-1.66	0.38	-0.76	1.04	0.59	-0.73	0.17	0.34
-0.92	-0.05	0.12	-0.65	1.45	-1.42	-0.12	-0.85	-0.83	-0.65
2.74	-0.27	-0.27	-0.01	-0.26	1.73	-1.50	-0.05	-0.81	1.71
-1.12	2.28	0.51	0.02	1.21	1.06	-0.96	1.42	1.95	1.05
0.39	1.85	-1.62	-0.56	1.42	-1.79	2.07	0.75	0.86	-0.49

TEST 1		TEST 2		TEST 3		TEST 4		TEST 5	
-0.71	-0.31	-1.20	1.22	0.01	-2.32	0.91	2.23	0.58	-0.60
-2.11	-0.93	2.41	2.05	-0.91	-2.45	0.50	-0.92	-0.64	0.20
0.19	-0.32	-3.03	2.26	2.92	2.01	0.70	-0.64	0.10	-0.86
0.80	-0.90	-2.12	0.34	2.93	2.90	2.44	-1.11	0.23	0.63
-0.70	-0.39	-0.81	0.23	0.06	3.41	1.32	-0.45	-0.32	-0.13

template avatar, i.e., $\kappa(\bar{s}), \bar{s} \in \bar{S}_0$. In our final models, we use $N = 84$ control points $\{\bar{s}_i, \kappa_i\}$ (shown in Figure 2), but we start the optimization process with a single point (i.e., homogeneous parameters) and progressively introduce more points (Please see the details in Section 5). We design biharmonic basis functions $\{\phi_i(\bar{s})\}$ on $\bar{S}_0$ [Jacobson et al. 2011], and the elasticity parameter field becomes $\kappa(\bar{s}) = \sum_i \phi_i(\bar{s}) \kappa_i$. Skin thickness is naturally defined on the surface only, and tangential and normal Young moduli are transformed to the actual body shape using the Laplacian interpolation discussed earlier, yielding $\kappa(\bar{x})$.

As a baseline for our shape-dependent soft flesh model, we have optimized the model described so far, independently per subject, on a set of training subjects. Figure 3 depicts the dynamics recovered by the independently optimized soft avatar models, compared to the ground-truth data (and not captured by SMPL alone). In this figure and others in the paper, we measure spatially varying dynamics over the body as the standard deviation of per-vertex deformation amplitudes in the shape's neutral pose, i.e., before skinning.

The improvement of the optimized models over SMPL is large, but not perfect, as the limited expressiveness of the reduced model cannot represent all the dynamics in the ground-truth data. In particular, note how the reduced model reproduces dynamics more accurately on the torso (which has a higher density of handles) than on the thighs (which are sampled with fewer handles). We have confirmed that the major source of error between these baseline optimized models and the ground-truth dynamics is the use of a reduced model. We have optimized the elasticity parameters (skin thickness and stiffness) for one of the training subjects (DYNA4, shown in Table 1) using a full-space simulation model, and we have recovered 94% of the dynamics, while our reduced model recovers only 51%. Despite this difference, we opt for the reduced model because of its performance vs. accuracy trade-off.

The independently optimized results set optimal targets for our shape-dependent model, and we use them for comparisons throughout the rest of the paper. In Section 6 we provide more details about the ground-truth data and train and test subjects.

4 SHAPE-DEPENDENT ELASTICITY

Taking as starting point the soft avatar model described in the previous section, we extend it to allow a shape-dependent multi-person elasticity parameterization. In this section, we formulate the function of shape-dependent spatially-varying elasticity parameters $\kappa(\bar{s}, \beta)$, and we introduce a local shape-based regularization. For the formulation of the optimization, we consider the elasticity parameters κ (skin thickness and Young moduli) normalized by their maximum values.

To formulate the shape dependency in $\kappa(\bar{s}, \beta)$ we opt for the simplest choice, linear regression:

$$\kappa(\bar{s}, \beta) = A(\bar{s})^T \beta + b(\bar{s}). \tag{1}$$

The shape domain β is considerably high-dimensional, i.e., $\mathbb{R}^{10}$ in our case, and sparsely sampled with subject data. Then, correctly identifying possible nonlinearities appears hopeless, and higher-order regression is more prone to overfitting and oscillation. At first, we tried radial basis functions (RBFs), as a general scattered data interpolation method, but this did not work. We did not try neural networks, but one could expect the same overfitting issues as with RBFs. Our choice of linear regression is also supported by the results attained, as discussed in Section 6.

To further increase the robustness to overfitting, we regularize the regression coefficients $A(\bar{s})$. We do this by defining a Tikhonov regularization loss $\mathcal{L}_{reg}$ [Pighin and Lewis 2007], which is added in the optimization process detailed later in Section 5. However, simple

Table 2: Standard deviation (in mm.) of per-vertex dynamics for all train and test subjects, on train and test motions, for all the comparisons carried out in the paper. Ground-truth data DYNA data [Pons-Moll et al. 2015], independent optimizations (Section 3), leave-one-out cross-validations (Section 6.2), full optimizations (Section 6.3), and generalization to test subjects (Section 6.4).

	Train motions						Test motions					
	DYNA 1	DYNA 2	DYNA 3	DYNA 4	DYNA 5	Avg	DYNA 1	DYNA 2	DYNA 3	DYNA 4	DYNA 5	Avg
Ground truth	4.51	2.92	4.16	3.85	3.19	3.72	4.13	2.38	2.93	3.65	3.21	3.26
Independent	2.61	1.27	1.85	1.96	1.64	1.86	2.09	1.07	1.04	1.41	1.08	1.34
Cross-val. L^2	1.95	2.17	1.33	2.03	2.23	1.94	1.63	1.85	0.74	1.49	1.49	1.44
Cross-val. ours	1.85	1.48	1.40	2.13	2.05	1.78	1.57	1.30	0.77	1.52	1.38	1.30
Full opt. L^2	2.47	1.26	1.66	2.22	1.92	1.90	2.05	1.09	0.94	1.56	1.29	1.38
Full opt. ours	2.17	1.32	1.55	2.06	1.85	1.79	1.81	1.13	0.87	1.45	1.23	1.29
	TEST 1	TEST 2	TEST 3	TEST 4	TEST 5	Avg	TEST 1	TEST 2	TEST 3	TEST 4	TEST 5	Avg
Generaliz. L^2	1.53	1.37	1.88	1.17	1.61	1.51	1.24	1.01	1.56	0.86	1.26	1.18
Generaliz. ours	2.37	1.51	2.43	1.01	1.73	1.81	1.99	1.17	2.07	0.75	1.36	1.47

L^2 regularization $\mathcal{L}_{\text{reg}} = \int_{\bar{\mathcal{S}}_0} \text{tr}\left(A(\bar{s})^T A(\bar{s})\right) d\bar{s}$ is not satisfactory. It overdamps inter-subject variability, as we show in our results.

Instead, we propose a regularization approach that looks at the local importance of the shape coefficients. With our choice of SMPL parametric avatar model, the template body surface is personalized using shape blend-shapes as $\bar{s} + B(\bar{s})\beta$. We observe that shape coefficients in β are locally more important when their blend-shape coefficients $B(\bar{s})$ have larger local variance. Then, as we want to penalize shape coefficients with low importance, we can use as regularizer the inverse of their variance. Even better, we use the inverse of the covariance, which finds directions of high and low importance in parameter space that are not exactly aligned with the shape coefficients.

Formally, we implement this importance-based regularization using a generalized Tikhonov regularization of the form:

$$\mathcal{L}_{\text{reg}} = \int_{\bar{\mathcal{S}}_0} \text{tr}\left(A(\bar{s})^T Q(\bar{s}) A(\bar{s})\right) d\bar{s}. \qquad (2)$$

And based on the motivation above, we define the regularization metric $Q(\bar{s})$ as the inverse covariance of shape blend-shapes on a local patch:

$$Q(\bar{s}) = \left(\int_{\bar{\mathcal{S}}_0} w(\|\bar{t} - \bar{s}\|)\, B(\bar{t})^T B(\bar{t})\, d\bar{t}\right)^{-1}, \qquad (3)$$

with $w(\|\bar{t} - \bar{s}\|)$ a smoothly decaying weight function.

With our choice of regularization metric, directions of shape coefficients β with large local influence are weakly penalized in the regularization loss (2). Conversely, directions with little influence are strongly penalized, thus maximizing the robustness of linear regression. For comparison, L^2 regularization would amount to using an identity metric $Q(\bar{s}) = I$ instead of the inverse covariance.

We discretize the regularization loss (2) as:

$$\mathcal{L}_{\text{reg}} \approx \frac{1}{N} \sum_{i=1}^{N} \text{tr}\left(A_i^T Q_i A_i\right). \qquad (4)$$

We normalize this discrete approximation by the number of elasticity control points N, to ensure consistency of the loss as we progressively add control points in the optimization process.

Note that, for the discrete regularization (4), the metric is only computed at the control points. Therefore, we use as smoothly decaying weight function $w(\|\bar{t} - \bar{s}_i\|) = \phi_i(\bar{t})$, i.e., the biharmonic basis functions of the control points.

5 PARAMETER OPTIMIZATION

In this section, we describe the estimation of the shape-dependent elasticity model as a numerical optimization, and we discuss our solution strategy.

To estimate the shape-dependent elasticity parameters, we minimize the difference in estimated dynamic motion for a set of train subjects with respect to ground-truth motion sequences. We denote the aggregated fitting loss as $\mathcal{L}_{\text{fit}}$. Following previous works [Kim et al. 2017; Romero et al. 2020], we measure the difference in dynamic motion based on the amplitude of oscillations of surface points with respect to the SMPL baseline surface. In particular, we use separate normal and tangential motion variance as proposed by Romero et al. We discuss the specific motion sequences and train subjects in Section 6. An alternative to fitting the amplitude of dynamics is to minimize per-vertex error. However, in our experience this is such a hard problem that the optimization finds a compromise by making skin too stiff and largely clamping its dynamics.

We gather the parameters of the linear regression (1) for all control points, which we denote as $p = \{A_i, b_i\}$. With 10 shape coefficients, 3 elasticity parameters, and N control points, the number of parameters to optimize is $33 \times N$. Then, putting together the fitting loss and the regularization loss (4), the parameter estimation problem is formally defined as:

$$p = \arg\min \mathcal{L}, \quad \text{with } \mathcal{L} = \mathcal{L}_{\text{fit}} + \lambda\, \mathcal{L}_{\text{reg}}. \qquad (5)$$

The hyper-parameter λ balances fitting and regularization error. As discussed later in Section 6.2, we empirically optimize this hyper-parameter.
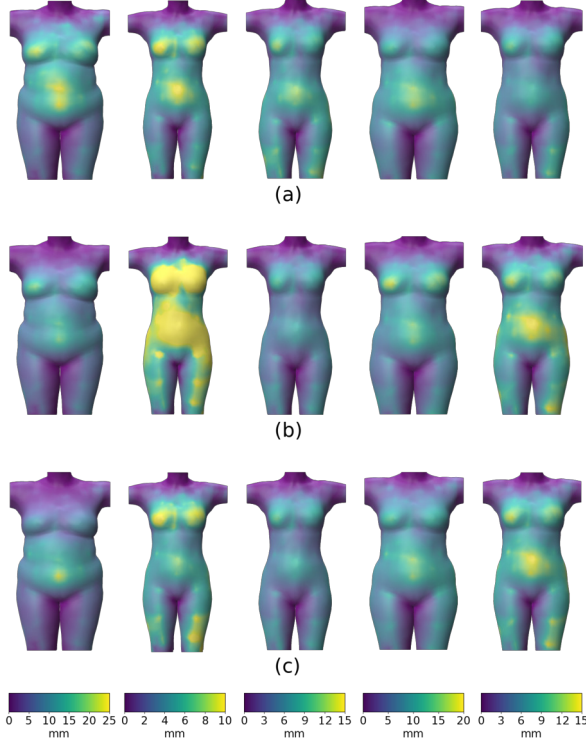
Figure 4: Body dynamics of the DYNA train subjects: (a) optimized independently, cross-validation (b) with L^2 regularization and (c) with our method.

We solve the optimization (5) using Scipy's trust-region solver for nonlinear least squares [Branch et al. 1999]. This requires expressing the loss terms as squared residuals:

$$\mathcal{L}_{\text{fit}} = r_{\text{fit}}^T r_{\text{fit}} \quad \text{with } r_{\text{fit}} = \sigma(p) - \sigma^*, \tag{6}$$

$$\lambda \mathcal{L}_{\text{reg}} = r_{\text{reg}}^T r_{\text{reg}} \quad \text{with } r_{\text{reg}} = \sqrt{\lambda} L p, \tag{7}$$

where $\sigma(p)$ and σ^* are, respectively, estimated and target motion variance vectors, and L is assembled from the Cholesky factorizations of the regularization metrics in (4) (scaled by $1/\sqrt{N}$ and with 0s for the $\{b_i\}$ terms in p).

For the efficient computation of residual gradients, it is also convenient to gather the elasticity parameters for all train subjects and their control points, which we denote as $q = \{\kappa_{i,j} = A_i^T \beta_j + b_i\}$, with $\{\beta_j\}$ the shape descriptors of the subjects. With M train subjects, the number of elasticity parameters for all subjects is $3 \times N \times M$. Then, the residual gradients can be expressed as:

$$\frac{dr_{\text{fit}}}{dp} = \frac{d\sigma}{dq} \frac{dq}{dp}, \qquad \frac{dr_{\text{reg}}}{dp} = \sqrt{\lambda} L. \tag{8}$$

dq/dp can be easily obtained analytically from (1). To compute $d\sigma/dq$, we have used a finite-difference approximation, which amounts to simulating the train motion sequences for $(1 + 3 \times N) \times M$ soft
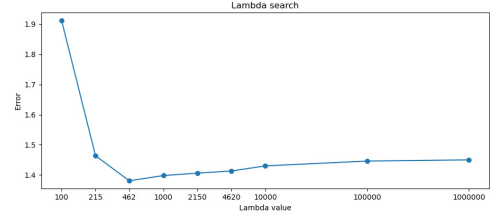

Figure 5: Cross-validation error as a function of the hyperparameter λ in (5).

avatars (with one-sided finite differences). This is one order of magnitude more efficient than estimating $d\sigma/dp$ through finite differences, which would require simulating the train motion sequences for $(1 + 33 \times N) \times M$ soft avatars. The computation of $d\sigma/dq$ could be accelerated with differentiable simulation, but in practice we have achieved acceptable performance by parallelizing the simulations for all soft avatars and motion sequences.

To further speed up the optimization, we have also increased the number of control points N progressively, with $N = \{1, 12, 24, 84\}$.

6 RESULTS

We show multiple examples where soft flesh is simulated on avatars of different body shapes. Figure 9 shows contact deformations produced by interaction with objects, Figure 10 shows skin dynamics produced by contact between avatars, and Figure 1 shows dynamics due to body motion. For this last example, we generated 25 random body shapes. In our implementation, the soft-flesh avatar simulation runs at 30 fps for one body, single-threaded, on a 3.4 GHz Intel Core i7-4770 CPU with 32GB of memory.

6.1 Subjects and Data

We have used as training data the bodies and animation captures of the DYNA data set [Pons-Moll et al. 2015]. These consist of 5 bodies, shown in the top row of Table 1. In the table, we also indicate the β shape coefficients of the SMPL parametric model fit to these bodies. We have used 3 motion sequences from the data set ("one leg jump", "running on spot", and "shake hips") as training data to optimize elasticity parameters, and 2 other motion sequence ("jiggle on toes" and "shake shoulders") as test data. For the optimization process, we have executed the simulations with a time step corresponding to 60 fps. We have validated that the dynamics obtained by the resulting model vary only by 1% with a time step of 120 fps, therefore the estimation is not significantly affected by numerical damping.

We have generated 5 test subjects (shown in the bottom row of Table 1) with specific body features to test the generalization capabilities of our method. TEST 1 has an upper body very similar to DYNA 1 despite the large difference in shape coefficients; TEST 2 has pear-shaped body with wide hips, which no train subject has; TEST 3 has a bigger chest than any train subject, TEST 4 is skinnier and taller than the train subjects, and TEST 5 has average shape.

Table 2 summarizes the quantitative analysis discussed in the following subsections.
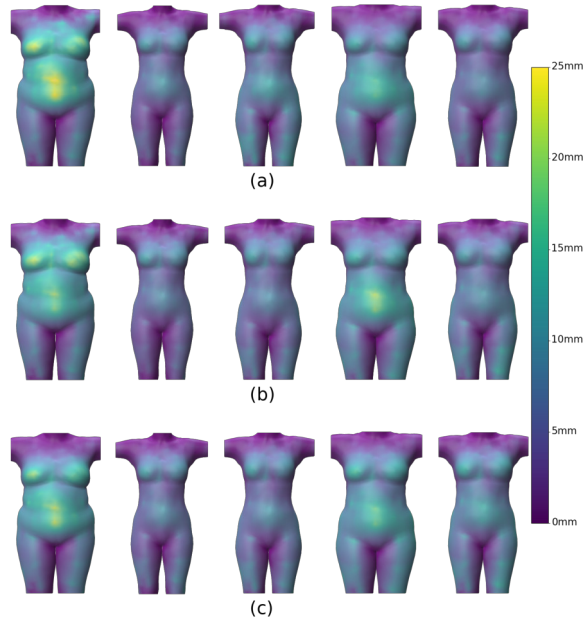
Figure 6: Body dynamics of the DYNA train subjects: (a) optimized independently, (b) with L^2 regularization and (c) with our method.

Figure 8: Body dynamics of (a) the DYNA train subjects, and the test subjects in Table 1 (b) with L^2 regularization and (c) with our method.

6.2 Leave-One-Out Cross-Validation

As a quantitative test for our model, we have performed a leave-one-out cross-validation on the train subjects. Note that this leaves only 4 subjects as train data each time, and the fifth subject is often far outside the range of train data. Figure 4 compares the standard deviation of per-vertex dynamics (a) of optimal targets fit independently, (b) of the fifth body modeled with L^2 regularization, and (c) of the fifth body modeled with our method. The colormaps clearly show the benefits of our model, as the results are very close to the optimal targets. With L^2 regularization, on the other hand, thinner subjects appear excessively soft.

We have also used the cross-validation to optimize the hyperparameter λ that balances fit vs. regularization in (5). Figure 5 shows the fit error aggregated on the left-out subjects, while sweeping λ.
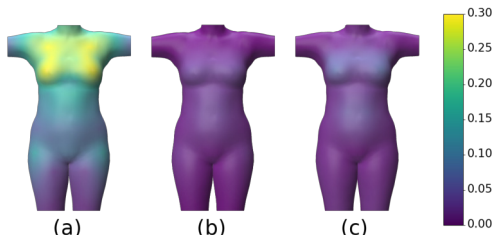

Figure 7: Norm of covariance of elasticity parameters on the train subjects: (a) optimized independently, (b) with our method, (c) with L^2 regularization.
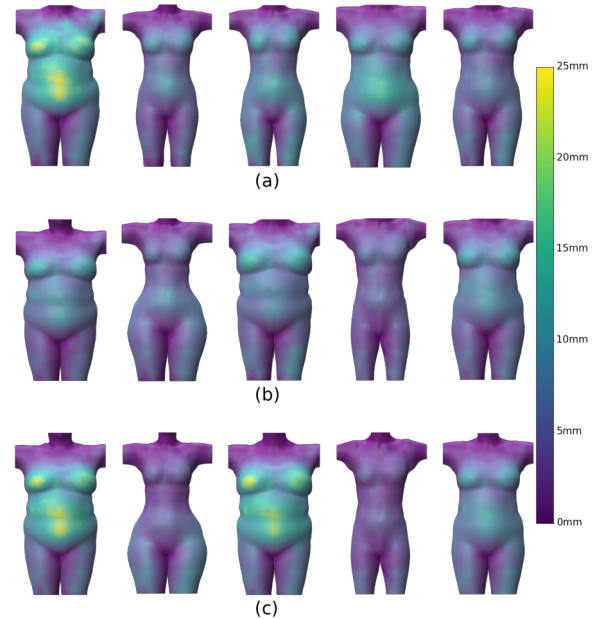
6.3 Full Optimization

With the model cross-validated and λ optimized, we have trained the final shape-dependent soft-flesh model on the 5 DYNA train subjects. The results are shown in Figure 6, compared against L^2 regularization and independently trained subjects. Figure 7, on the other hand, compares the covariance of the resulting elasticity parameters, again with our method, L^2 regularization, and independently trained subjects. To compute this covariance, we normalize each parameter within its allowed range in the optimization. The parameter set for each training subject at each control point represents a 3D point (skin thickness and two stiffness values). We compute the 3D covariance for the parameter sets of the 5 training subjects at each control point, and we display the Frobenius norm of this covariance.

The combined analysis of Figure 6 and Figure 7 denotes that our method achieves very similar dynamics to independently trained subjects, but with no overfitting, i.e., much lower parameter variance. The difference between our method and L^2 regularization is however small, but this is no surprise, as L^2 regularization works well on the training data. L^2 regularization does not suffer from overfitting either, but it fails to generalize as discussed in Section 6.2 above and in the next section.

6.4 Generalization

As a final qualitative test, we have applied our soft-flesh model to the test subjects in Table 1. Figure 8 compares the dynamics (i.e., per-vertex standard deviation of displacements) of our method vs. L^2 regularization. As a reference, we also show the dynamics of the

Figure 9: Example of object interaction with the body, producing contact deformations. Even though our soft-flesh model is trained from skeletal dynamics, it naturally supports contact deformations at runtime.

Figure 10: Example of dynamics produced by body-body contact. One avatar pushes the other one, and the contact forces produce both skeletal motion and soft-tissue dynamics (best seen in the video).

train subjects. Our method produces highly plausible results, even if the test bodies notably extend the range of shape coefficients. For instance, subject TEST 3 shows larger dynamics on the chest than any train subject, and TEST 4 shows less dynamics than any train subject. Despite their different shape coefficients, TEST 1 has an upper body very similar to DYNA 1, and our model succeeds at capturing local shape similarity and produces similar dynamics on both. The hips of TEST 2 are an extreme case of local shape not seen at training. In a nutshell, our method clearly outperforms L^2 regularization. With L^2 regularization, the dynamics of test subjects are strongly clamped (Figure 8-b); with our method they are much more expressive (Figure 8-c), comparable to those of the training subjects (Figure 8-a).

7 CONCLUSION

In this paper, we have presented the first soft-flesh avatar model that generalizes to bodies of different shape. We achieve this by fitting a shape-dependent model of elasticity parameters. As we show in the paper, it is key to design a shape-informed regularization, to avoid overfitting and to allow for rich generalization even under extremely sparse training data.

Our method suffers several limitations, and it also has room for interesting extensions. Due to the limited amount of data used in the optimization and tests, the generalization ability of the model requires further validation. The elasticity parameters are estimated from skeletal dynamics, and it is unclear if the resulting behavior is accurate for contact deformations. Ideally, a model should be estimated combining both sources of data.

As noted when comparing results to the ground-truth data, the accuracy of the model is largely limited by the expressive ability of the underlying reduced deformation model. This could be improved by building a learning-based model for dynamics, but it is important to retain the interaction capability of the physics-based model. Finally, given the visual quality achieved recently by NeRF models, it would be interesting to connect the deformation model to a

NeRF representation, which might require support for other shape parameterizations.

REFERENCES

Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. SCAPE: Shape Completion and Animation of People. *ACM Trans. Graph.* 24, 3 (July 2005), 408–416.

Ziqian Bai, Timur Bagautdinov, Javier Romero, Michael Zollhöfer, Ping Tan, and Shunsuke Saito. 2022. AutoAvatar: Autoregressive Neural Fields for Dynamic Avatar Modeling. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 222–239.

Mary Ann Branch, Thomas F. Coleman, and Yuying Li. 1999. A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM Journal on Scientific Computing* 21, 1 (1999), 1–23. https://doi.org/10.1137/S1064827595289108

Dan Casas and Miguel A Otaduy. 2018. Learning nonlinear soft-tissue dynamics for interactive avatars. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 1, 1 (2018), 10.

Ye Fan, Joshua Litven, and Dinesh K. Pai. 2014. Active Volumetric Musculoskeletal Systems. *ACM Trans. Graph.* 33, 4, Article 152 (2014), 9 pages.

S. Foti, B. Koo, D. Stoyanov, and M. J. Clarkson. 2022. 3D Shape Variational Autoencoder Latent Disentanglement via Mini-Batch Feature Swapping for Bodies and Faces. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 18709–18718.

Darcy Harrison, , Ye Fan, and Dinesh K Pai Egor Larionov. 2018. Fitting close-to-body garments with 3D soft body avatars. In *Proceedings of 3DBODY. TECH*. Lugano, Switzerland, 184–189.

Mustafa Işık, Martin Rünz, Markos Georgopoulos, Taras Khakhulin, Jonathan Starck, Lourdes Agapito, and Matthias Nießner. 2023. HumanRF: High-Fidelity Neural Radiance Fields for Humans in Motion. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–12. https://doi.org/10.1145/3592415

Alec Jacobson, Ilya Baran, Jovan Popović, and Olga Sorkine. 2011. Bounded Biharmonic Weights for Real-time Deformation. *ACM Trans. Graph.* 30, 4, Article 78 (July 2011), 8 pages.

C. Kane, J. E. Marsden, M. Ortiz, and M. West. 2000. Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *Internat.*

J. Numer. Methods Engrg. 49, 10 (2000), 1295–1325.

Meekyoung Kim, Gerard Pons-Moll, Sergi Pujades, Seungbae Bang, Jinwook Kim, Michael J. Black, and Sung-Hee Lee. 2017. Data-Driven Physics for Human Soft Tissue Animation. *ACM Trans. Graph.* 36, 4, Article 54 (2017), 12 pages. https://doi.org/10.1145/3072959.3073685

Martin Komaritzan, Stephan Wenninger, and Mario Botsch. 2021. Inside Humans: Creating a Simple Layered Anatomical Model from Human Surface Scans. *Frontiers in Virtual Reality* 2 (2021).

Seunghwan Lee, Yifeng Jiang, and C. Karen Liu. 2023. Anatomically Detailed Simulation of Human Torso. *ACM Trans. on Graphics* 42 (2023).

Seunghwan Lee, Moonseok Park, Kyoungmin Lee, and Jehee Lee. 2019. Scalable Muscle-Actuated Human Simulation and Control. *ACM Trans. Graph.* 38, 4, Article 73 (2019), 13 pages.

Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B. Tenenbaum, and Antonio Torralba. 2019. Learning Particle Dynamics for Manipulating Rigid Bodies, Deformable Objects, and Fluids. In *Proceedings of the International Conference on Learning Representations.*

Libin Liu, KangKang Yin, Bin Wang, and Baining Guo. 2013. Simulation and Control of Skeleton-Driven Soft Body Characters. *ACM Trans. Graph.* 32, 6, Article 215 (Nov. 2013), 8 pages.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-person Linear Model. *ACM Trans. Graph.* 34, 6, Article 248 (Oct. 2015), 16 pages.

Aleka McAdams, Yongning Zhu, Andrew Selle, Mark Empey, Rasmus Tamstorf, Joseph Teran, and Eftychios Sifakis. 2011. Efficient Elasticity for Character Skinning with Contact and Collisions. *ACM Trans. Graph.* 30, 4, Article 37 (July 2011), 12 pages.

A. Nava, E. Mazza, M. Furrer, P. Villiger, and W.H. Reinhart. 2008. In vivo mechanical characterization of human liver. *Medical Image Analysis* 12, 2 (2008), 203–216.

Dinesh K. Pai, Austin Rothwell, Pearson Wyder-Hodge, Alistair Wick, Ye Fan, Egor Larionov, Darcy Harrison, Debanga Raj Neog, and Cole Shing. 2018. The Human Touch: Measuring Contact with Real Human Soft Tissues. *ACM Trans. Graph.* 37, 4 (2018), 58:1–58:12.

Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR).* 10975–10985.

Fred Pighin and J. P. Lewis. 2007. Practical Least-Squares for Computer Graphics: Video Files Associated with This Course Are Available from the Citation Page. In *ACM SIGGRAPH 2007 Courses* (San Diego, California) *(SIGGRAPH '07).* Association for Computing Machinery, New York, NY, USA, 1–57.

Gerard Pons-Moll, Javier Romero, Naureen Mahmood, and Michael J. Black. 2015. Dyna: A Model of Dynamic Human Shape in Motion. *ACM Trans. Graph.* 34, 4, Article 120 (July 2015), 14 pages.

Cristian Romero, Miguel A. Otaduy, Dan Casas, and Jesus Perez. 2020. Modeling and Estimation of Nonlinear Skin Mechanics for Animated Avatars. *Computer Graphics Forum (Proc. Eurographics)* 39, 2 (2020).

Igor Santesteban, Elena Garces, Miguel A. Otaduy, and Dan Casas. 2020. SoftSMPL: Data-driven Modeling of Nonlinear Soft-tissue Dynamics for Parametric Humans. *Computer Graphics Forum* 39, 2 (2020), 65–75.

Breannan Smith, Fernando De Goes, and Theodore Kim. 2018. Stable Neo-Hookean Flesh Simulation. *ACM Trans. Graph.* 37, 2, Article 12 (March 2018), 15 pages.

Javier Tapia, Cristian Romero, Jesus Perez, and Miguel A. Otaduy. 2021. Parametric Skeletons with Reduced Soft-Tissue Deformations. *Computer Graphics Forum* (2021).

Yu Wang, Alec Jacobson, Jernej Barbič, and Ladislav Kavan. 2015. Linear Subspace Design for Real-Time Shape Deformation. *ACM Trans. Graph.* 34, 4, Article 57 (2015).

Hongyi Xu and Jernej Barbič. 2016. Pose-Space Subspace Dynamics. *ACM Trans. Graph.* 35, 4, Article 35 (July 2016), 14 pages.

Mianlun Zheng, Yi Zhou, Duygu Ceylan, and Jernej Barbic. 2021. A Deep Emulator for Secondary Motion of 3D Characters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 5932–5940.

Zerong Zheng, Xiaochen Zhao, Hongwen Zhang, Boning Liu, and Yebin Liu. 2023. AvatarReX: Real-Time Expressive Full-Body Avatars. *ACM Trans. Graph.* 42, 4, Article 158 (2023).