EUROGRAPHICS 2013 / I. Navazo, P. Poulin
(Guest Editors)

Volume 32 (2013), Number 2

Capture and Statistical Modeling of Arm-Muscle Deformations

T. Neumann[1], K. Varanasi[3], N. Hasler[3], M. Wacker[1], M. Magnor[2] and C. Theobalt[3]

[1]HTW Dresden, Germany
[2]Computer Graphics Lab, TU Braunschweig, Germany
[3]Max-Planck-Institut Informatik, Saarbrücken, Germany

Abstract

We present a comprehensive data-driven statistical model for skin and muscle deformation of the human shoulder-arm complex. Skin deformations arise from complex bio-physical effects such as non-linear elasticity of muscles, fat, and connective tissue; and vary with physiological constitution of the subjects and external forces applied during motion. Thus, they are hard to model by direct physical simulation. Our alternative approach is based on learning deformations from multiple subjects performing different exercises under varying external forces. We capture the training data through a novel multi-camera approach that is able to reconstruct fine-scale muscle detail in motion. The resulting reconstructions from several people are aligned into one common shape parametrization, and learned using a semi-parametric non-linear method. Our learned data-driven model is fast, compact and controllable with a small set of intuitive parameters - pose, body shape and external forces, through which a novice artist can interactively produce complex muscle deformations. Our method is able to capture and synthesize fine-scale muscle bulge effects to a greater level of realism than achieved previously. We provide quantitative and qualitative validation of our method.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—I.4.8 [Image Processing]: Digitization and Image Capture—Scanning

1. Introduction

Realistic virtual humans are becoming commonplace in movies and interactive CG applications. Achieving realistic appearance depends not only on accurate rendering and motion, but also on accurate modeling of subtle pose-dependent deformations which are due to muscles and other soft tissue. Currently, creating such subtle effects is a challenging process - usually requiring tedious manual work by experienced animators. State-of-the-art techniques typically rely on a complex multi-layered animation model - comprising an animation rig, attached muscles and a physics based simulation of deformation effects [LGK*12]. Despite producing believable results, these approaches have several drawbacks. Customization of muscle models to specific virtual subjects is not easy. Control and simulation of these models may be computationally expensive and require tuning of many parameters. Further, since the underlying muscle model is still an approximation of the anatomical reality, truly fine-scale muscle deformation may not be reproducible.

In this paper, we present an alternative data-driven approach for muscle deformation which addresses some of these limitations. Rather than designing a model *a priori*, we learn the complex inter-dependencies between muscle activity of real subjects and their physical constitution, motion, and external forces. From a database of multi-view video observations on male human subjects performing a range of arm motions, we build a data-driven muscle simulation model for the shoulder-arm complex. Our model is efficient and easy-to-use, controllable with only a small number of intuitive parameters. It is capable of reproducing a wide range of arm movements and muscle bulges in various poses as well as under the action of various external forces. We note that similar concepts as presented in this work can be applied to other body parts, or across wider populations.

Our model takes as input standard skeletal motion parameters to specify the motion to be executed. In addition, the parameters that control the appearance of the subject in terms of physique and training level - the subject's body

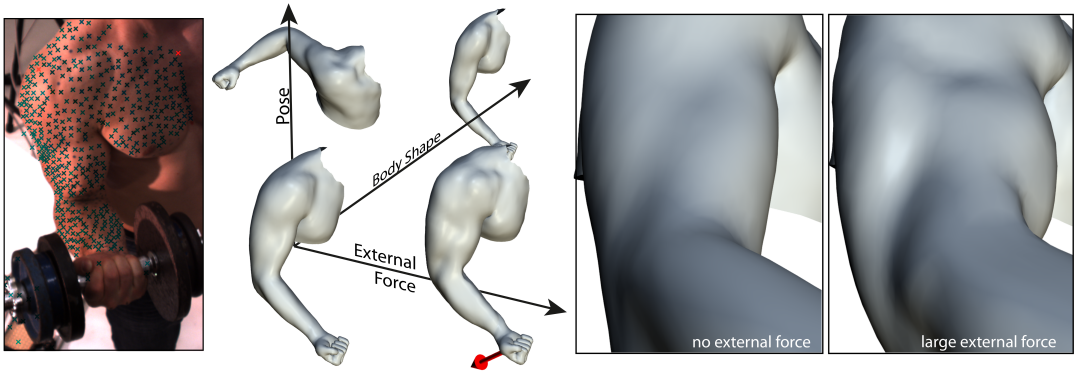


Figure 1: *Multi-view recordings of human subjects performing various muscle exercises (left) are used to build a statistical model for synthesizing muscle deformations (middle) that can be altered according to pose, body shape and external forces. Our model can simulate fine-scale bulges and dimples on the skin which are caused by various muscle strands (right).*

mass index (BMI), muscularity, height are considered. We are also able to smoothly vary the muscle deformation between subjects of very different physique, training level and muscularity. Finally, the model takes as input an external force vector on the hand. Ultimately, our model reproduces realistic muscle shape as an interplay of all the above factors, and generalizes well to other motions, people and mechanical scenarios beyond the training dataset.

Building such a data-driven model poses new challenges in acquisition, model learning, and parametrization. In this paper, we make these principal contributions

1. A 3D acquisition method for highly detailed and spatio-temporally coherent geometry of deforming skin - which captures individual muscle strands, fat tissue, as well as tangential stretching and shifting of skin. The method is robust and scalable, e.g. for registering thousands of scans of a person it requires hand-labeling landmarks in just a single scan.
2. A semi-parametric learning approach to build a data-driven model from the above input data. The model needs little memory, can handle huge training datasets, and can be evaluated in real-time. At the same time, it can produce non-linear muscle deformation effects that cannot be achieved by previous data driven methods. Our prototype application rapidly synthesizes new animations of the shoulder-arm complex by controlling a small set of intuitive parameters.

To our knowledge, our method is the first data-driven approach for reproducing detailed pose-dependent skin and muscle deformation that spans the space of motion, body shape and physique, as well as external forces. In the following, we review related work (Sec. 2) and formally state the problem (Sec. 3). In Sec. 4, we describe our acquisition method. We describe our shape parameterization and learning method in Sec. 5 and Sec. 6. We show quantitative evaluation with groundtruth and a variety of deformation results atop new motions in Sec. 7 and in the supplementary video.

2. Related Work

Data-driven Muscle Modeling was studied by Allen et al. [ACP02] by acquiring static range scans of the torso and the arm of a single person in different poses. Performing nearest-neighbor interpolation allows them to generate deformations at new poses. Park and Hodgins [PH06] used a VICON mocap system and a set of 350 reflective markers to capture skin deformation. In [PH08], they describe a non-linear kernel regression method for obtaining a surface model at this resolution (350 markers) from a smaller set of $40-50$ skeletal markers. They later describe a method to generate a better skeletal model for the shoulder with additional virtual joints, using a capture set-up of 200 markers [HPH10]. Drawing inspiration from these works, we build a model with a wider scope and higher detail. Specifically, unlike these methods, we model the statistical variations of muscle deformations across multiple people. We also incorporate external forces which enables modeling isometric muscle contractions in static poses, an effect largely overlooked previously. Additionally, our acquisition strategy is less intrusive and yields high resolution of capture.

Data-driven Face Capture and Modeling Vlasic et al. [VBPP05] present a multi-linear face model that models personal identity, facial expression and visemes as discrete variables. Attempts have also been made to capture the elastic properties of facial muscles: Beeler et al. [BBO*09] propose a marker-based framework for capturing muscle strain in response to stress induced by a force probe. In our current work, we synthesize such force-dependent deformations not by parameters of elasticity, but directly through a data-driven model.

Statistical Body Modeling Allen et al. [ACP03] analyze shape variations of 3D scans of people standing in the same pose. The SCAPE body model [ASK*05] generalizes this work by modeling surface deformation as a function of both body shape and pose, by assuming that both effects

are independent (using 70 scans from one subject in multiple poses and 37 scans of multiple subjects in one pose). Later, it was shown by Allen et al. [ACPH06] and Hasler et al. [HSS*09] that body pose and shape can also be modeled simultaneously, which allows scanning several subjects in several poses to analyze a broader range of variations. Statistical body models like these have applications in image and video understanding [HAR*10], animation and shape completion [ASK*05]. Unlike these methods which are based on laser scans, we capture and model minute muscle deformation and tangential skin-shifting on multiple subjects. We use a large training database of more than 32,000 meshes and contribute to this area by modeling the effect of external forces on muscle deformation and stretching of skin.

Anatomical and Physics-based Modeling Anatomically accurate modeling of human muscle tissue and bone deformation has applications in sport medicine and kinesiology, alongside computer graphics. In the recent past, Teran et al. [TSB*05] show impressive results by simultaneously modeling muscles, bones and tendon properties using the Visible Human dataset, and simulating them using a Finite Element method. A similar approach is also used for building an anatomically based face muscle model [SNF05] and for realistic hand animation [SKP08]. Lee et al. [LST09] describe a comprehensive system for the biomechanical modeling of the upper human body. For a more comprehensive review of physiologically based modeling, we refer to the survey of [LGK*12]. Such physiologically based simulation methods are complementary to our approach. Although they are more accurate and yield elaborate shape and elastic deformation priors, tuning these parameters to different subjects is not straightforward. Simulating these complex mechanical systems is also computationally demanding. In this paper, we describe a data-driven approach that is easy-to-use, computationally efficient and which adapts quickly to new people.

Example-based Deformation Lewis et al. [LCF00] presented one of the first approaches to smoothly blend deformations defined at specific training poses. [WPP07] present a fast example-based skinning method based on predicting deformation gradients [SP04] from example meshes. Context-aware skeletal shape deformation [WSLG07] takes a similar path by learning polar-decomposed deformation gradients as residuals on top of a skeleton that is rigged by an artist. These methods provide deformation control using limited number of training poses modeled by the artist, while our approach provides deformation control by intuitive parameters and learns from a large dataset of 3D reconstructions of real people.

Shape Parametrization It is interesting to note how all these related approaches parametrize 3D shape. [PH08] model surface vertex displacements as offsets from the bone skeleton. [ASK*05] model surface triangle deformations using deformation gradients [SP04]. [HSS*09] use the relative rotations between triangles as a rotation-invariant encoding, which are, however, hard to use alongside positional constraints on vertices. Our deformation gradient encoding is related to this prior work, and is combined with a skinning prior as in [WSLG07] to make it usable alongside an artist-controllable skeleton. We also compress the encoding further, which reduces the burden on the learning method, especially for large training datasets like ours.

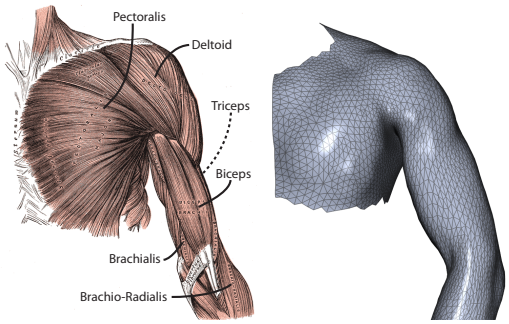


Figure 2: *Left: Anatomy of the shoulder-arm complex (*Gray's Anatomy - *Henry Gray,* 1858*) , depicting a few participating muscles. Right: Surface mesh of the same area as obtained by our method.*

3. Problem Statement

When a person moves, a complex biological control system of muscles, bones and soft tissue is activated. In the shoulder-arm complex alone, more than 20 muscles contribute to the motion - not taking into account multiple branches of pennation for muscles such as the *deltoid*, cf. Fig. 2. The largest arm muscles are the *biceps brachii* and the *triceps brachii* in the upper arm, which show large pose and force-dependent deformations. On the one hand, muscles are intentionally or sub-consciously activated through neural signalling. On the other hand, muscles react to external influences, i.e., forces or torques acting on joints. As a result, the body moves, the muscles bulge, and the soft tissue and skin deform, a process whose biological mechanisms are well-studied in anatomy and kinesiology [Eno08]. However, muscle shape and deformation also depend on a person's physical constitution, training level, tissue and fibre composition, etc. It is no wonder that the design of biologically accurate physics-based forward simulation models that account for all these parameters influencing the muscle and skin deformation is challenging, if not impossible, even today.

It is a reasonable abstraction of the true biomechanical process to assume that the main external influence on muscle bulges are forces acting on the rigid skeletal parts that the muscles are connected to. Such forces also create secondary

effects such as the in-taking of air into the lungs, the deformation of the spine to stabilize the body, and slight translations of bones. In this paper, we concentrate on building a model of the shoulder-arm-complex, as by its motion range and anatomical complexity it can be considered a down-scaled version of the full body [ACP02]. We consider slow to medium-fast motions which can be interpreted as *quasi-static* according to biomechanics literature [Eno08, *"When is a movement fast ?"*]. We acquire a training set of 3D skin surface measurements of several subjects doing a variety of muscle exercises slowly and naturally, where the quasi-static assumption is justified. We propose a data-driven approach to learn a mapping Ψ between a few intuitive yet biologically motivated input parameters (body pose θ, body shape β, external forces γ) and a body surface $\mathbb{M}$ that exhibits plausible muscle bulges and skin deformation, as follows:

$$\mathbb{M} = \Psi(\theta, \beta, \gamma) \tag{1}$$

Body Pose: The base of our model is a simplified bone skeleton with a shoulder and an elbow joint. Each joint is parametrized as a 3-degree of freedom ball- and socket joint, i.e., in total six parameters θ determine the skeleton pose.

Body Shape: We consider the body mass index (BMI), muscle proportion (as a percentage of body mass), and the height of the person as important physiological parameters β that influence muscle bulges.

External Forces: We add the force magnitude and direction vector rotated into the coordinate frame of the lower and upper arm (in all, $2 \cdot 3 + 1$ scalars) as model parameters. Please note that, although gravity acts always downwards, we capture deformation effects due to forces at different directions because of the relative rotations of the torso, upper and lower arm (supplementary document shows the validation).

The kind of deformations that are induced by external force are visually analyzed in Fig. 4, which shows two reconstructions acquired using our method.

4. Acquisition

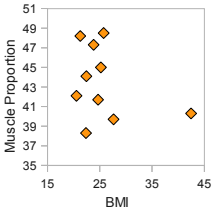
To capture a training set for learning muscle deformations as specified in our problem statement, we developed a new vision-based reconstruction approach that measures dense dynamic 3D arm geometry with a multi-camera system.

4.1. Capturing Setup

We use a multi-camera acquisition setup consisting of 16 synchronized and calibrated FireWire video cameras, each recording at a resolution of 1600×1200 pixels and a frame rate of 30 Hz, Fig 3(a). Cameras are arranged in a convergent setup around the shoulder-arm region. To facilitate space-time reconstruction, we apply a pattern of dense colored dots to the arm, shoulder and torso (random peppering of black, dark green and blue dots) using film make-up, Fig. 3(b). Depending on the body shape of the subject, around 800 to 1200 markers are applied with a maximum inter-marker distance of about 5*mm*. While most markers are randomly placed on the subject, we place 11 markers at anatomically equivalent locations with a red pen. These anatomical markers will be used in a template initialization step, and enable template matching across captured subjects. Application of make-up takes around $30-45$ minutes per person.

4.2. Training Database

We recruited 10 male subjects (between 20 to 34 years of age, of different body shapes) to be captured. Five of the subjects practice body building regularly, the remaining subjects exercise less and cover a wide variety of body shapes, from corpulent to thin. For each subject, we record some physiological parameters, such as weight and height. With a body fat scale we also get estimates of bone mass, body fat ratio, and muscle proportion. Based on this data, body shape can be roughly characterized with a scalar body mass index (BMI), height, and the muscle proportion. On the left, we show the measured BMI and muscularity of our 10 subjects.

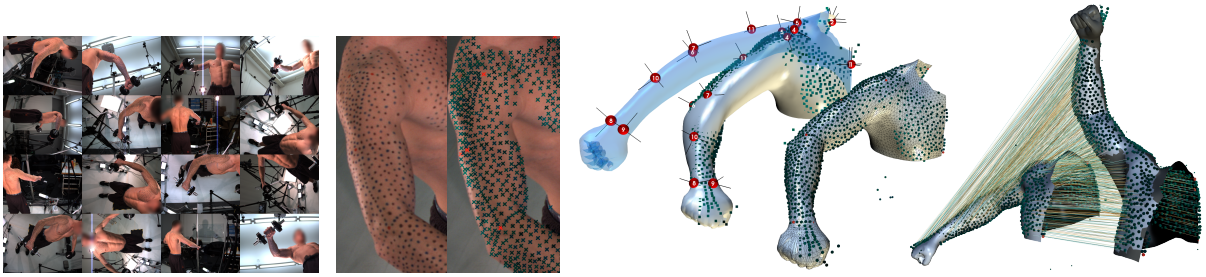


Each subject performs the same 11 predefined arm motions. An exact description of the motions is provided as supplementary document. In some motions, the shoulder-arm joints are articulated separately. In others, articulations are combined. Five of these motions are repeated with a barbell in the hand, to simulate external forces. For each subject, we capture the motion with 8 barbells between 0.5 kg and 17.5 kg, or up to the maximum he could lift. Additionally, some free motions like boxing, dancing, or flexing are performed, which we use for cross validation. Per subject, around $30-40$ motions, each consisting of approximately 100 frames, were captured. Our dataset is designed to sufficiently sample the pose, body shape and force parameter dimensions serving as input to our model.

4.3. Multi-view Reconstruction

The reconstruction pipeline consists of four stages: 3D marker reconstruction, template initialization, template registration across motions, and articulated motion estimation.

3D Marker Reconstruction In each frame of video, markers are detected with sub-pixel precision as Gaussian blobs. To match markers across views, we exploit epipolar constraints, and additionally impose soft constraints to enforce consistently matched marker neighborhoods. To this end, we extended the approximate Graduated Assignment algorithm [GR96], a graph-matching algorithm, from its original two



(a) (b) (c) (d)

Figure 3: *Capturing the skin deformation dataset: a) Subjects are recorded by 16 synchronized video cameras. b) Dot pattern on skin used for reconstruction of point clouds. c) Template (light blue) registered to a single initialization point cloud per subject, first using user-selected anatomic landmarks, followed by automatic alignment to the remaining* $\approx 1,000$ *points. d) Dense and robust correspondences (lines) between marker point clouds automatically align the initialization template to all other* $\approx 3,000$ *poses of the subject, occluded areas (dark grey) are interpolated.*

camera setting to the multi-camera case. The result is a dense cloud of 3D markers at each time step. These point clouds are not yet in temporal correspondence over frames. Henceforth, we use the terms points and markers interchangeably.

Template Initialization To bring all point clouds of the same subject, i.e., data from all motions, into correspondence, we register them to a common template mesh $\mathbb{M}$ (see Sect. 3) of the arm with $n = 5152$ vertices. We use the arm section of the average mesh of human body scans, kindly provided to us by [HSS*09]. From all the recorded frames per subject, we select a single *initialization time step* having the least occlusions.

To fit the template to the initialization point cloud, correspondences between the 11 anatomical markers in the point cloud and the respective locations on the template are manually marked, Fig. 3(c), using a simple user interface. The rest of the fitting procedure is fully-automatic and consists of a rough pose alignment followed by a non-rigid fine registration. For rough alignment, we deform the template into the point cloud by means of Laplacian deformation with rotation correction to compensate for the rotation-variance of differential coordinates [SKR*06].

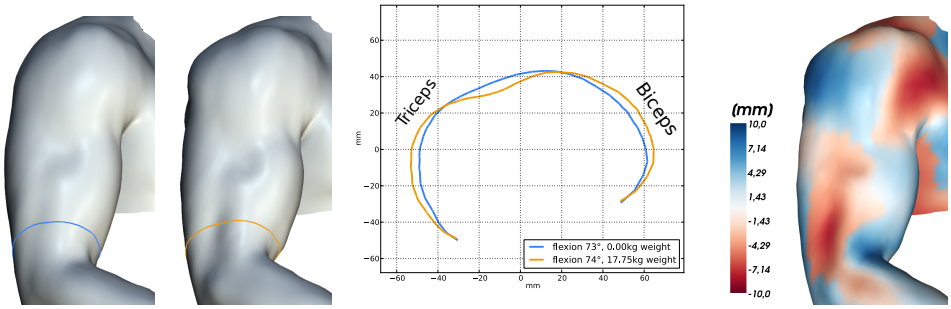
To reproduce the fine-scale skin deformations in the data, we subsequently perform a non-rigid registration. We consider point matches to barycentric coordinates relative to enclosing triangles on the template mesh. The mesh is then displaced towards these constraints with As-rigid-as-possible surface deformation [SA07] as regularization, Fig. 3(c).

Marker Matching and Template Registration Across Motions Once the template is initialized, it has to be aligned to all captured frames of a subject. Instead of relying on template tracking and thereby suffering from drift, we propose a novel strategy to robustly match 3D markers across any two poses and use these matches to deform the template into each target pose independently - thereby creating a fast and parallelized pipeline for template registration across all poses.

For each marker, we construct a series of Shape Context descriptors, each having a distinct orientation of the histogram axes. The coordinate frame of each descriptor is defined by a triplet consisting of the central and two neighboring points. The matching distance between two points is then the minimum distance of all descriptors attached to those points. An important benefit of this approach is that it implicitly takes care of skin deformations by deforming the histogram axes accordingly. By filtering out erroneous correspondences using non-maximal suppression, and by checking geometric consistency in the local spatial-neighborhoods of matches, this strategy yields a robust set of matches across even large pose differences, Fig 3(d).

We use these to deform the template mesh to fit to any pose in the captured dataset. Similar to the template initialization step, we achieve this by Laplacian deformation with rotation correction for the differential coordinates [SKR*06]. As shown in Fig. 4 and in the accompanying video, this yields highly accurate 3D reconstructions of the arm in movement.

In fact, the template mesh gives a strong prior for extending marker correspondences. After the template is deformed using the initial set of marker matches, new correspondences are found by searching within the local neighborhood of the deformed template and matching closest predicted marker locations for each unmatched marker. This process of marker matching, template deformation and correspondence update is iterated until no new correspondences can be added, usually around $5-10$ iterations. Please note that since this template alignment procedure is applied to the data of each individual, alignment of the entire database comes for free. User intervention is needed only for the template initialization step, once for each subject in the database.



(a) (b) (c) (d)

Figure 4: *Muscle deformations due to external load in elbow flexing motion, as captured on one subject in our dataset. (a) Flexing with no external load (b) Flexing with* 17.75 kg *of barbell weight in one hand (c) Cross-sectional cut along the horizontal plane at the bottom of the upper arm. Subtle bulges can be seen on triceps and biceps muscles, which work as an agonist-antagonist pair to keep the arm in equilibrium against the external load. (d) Vertex displacements from 'a' to 'b' visualized on mesh 'b'. Bulges are shown in blue, and dimples are shown in red (see also video for an animation).*

Articulated Motion Estimation All data are now registered based on a surface model, but we still need to infer joint angles for each pose to train our model. We consider a user-supplied segmentation of the shoulder-arm complex into 4 parts: Torso, upper arm, lower arm, and hand, as shown on the right. A rigid body rotation $\mathbf{R}$ and translation $\mathbf{t}$ is now found that best maps the rest-pose vertices $\mathbf{v}_i$ of each body part to the vertices $\mathbf{v'}_i$ in the observed pose, by minimizing the distance $\sum \|\mathbf{R}\mathbf{v}_i + \mathbf{t} - \mathbf{v}'_i\|^2$ for each body part separately. An initial rotation is found by solving the orthogonal procrustes problem using SVD. From this starting point, a more accurate rigid transformation is estimated by iterative minimization using a Newton-Raphson method and the twist representation of the rotation, similar to [BM98] but without the need for a kinematic chain. These segment transformations are aligned between different subjects by expressing them relative to a skeleton rigged inside the template mesh. Similar to conventional motion-capture, rotation angles are used to represent the body part rotations.

5. Shape Parametrization

As described in our problem statement, we would like to represent the surface deformation as a function of body pose (θ), shape (β) and external forces representing triggers for muscle activation (γ). Of these, parametrizing the surface deformation through skinning from joint angles θ is widely used practice in animation. We adopt a two layered representation for surface deformation that enables us to accommodate the additional input dimensions: The first layer is provided by a skeleton and triangle-based quaternion blend skinning [WSLG07]. The second layer of residual transformations that connects the skinned mesh to the actual detailed skin and muscle deformation is to be learnt by our mathematical model Ψ.

To this end, we need to represent a shape (that is, the vertex coordinates $\mathbf{v}_i$ of our template mesh $\mathbb{M}^t$ in a pose t) in a suitable encoding. Using 3D Euclidean positions directly cannot model rotations of the arm properly, and even when modeled in pose space [LCF00, SRC01] causes problems due to varying arm lengths and shapes between subjects. We therefore build upon and extend ideas on deformation gradient encoding [SP04], to learn a rotation and stretch transformation per template triangle from the training data.

Deformation Gradient Encoding Let us consider a triangle in the current pose $\mathbb{M}^t$ with vertices $\mathbf{v}^t_1, \mathbf{v}^t_2, \mathbf{v}^t_3$. Following [SP04], we represent its deformation gradient from a rest pose s as

$$\mathbf{D}^\mathbf{t} = \mathbf{T}^\mathbf{t} \cdot (\mathbf{T}^\mathbf{s})^{-1} \tag{2}$$

where

$$\mathbf{T}^\mathbf{t} = \begin{bmatrix} \mathbf{v}^\mathbf{t}_2 - \mathbf{v}^\mathbf{t}_1 & \mathbf{v}^\mathbf{t}_3 - \mathbf{v}^\mathbf{t}_1 & \mathbf{n}^\mathbf{t} \end{bmatrix} \tag{3}$$

and correspondingly for $\mathbf{T}^\mathbf{s}$. Here, $\mathbf{n}^\mathbf{t}$ is the triangle face normal. We simplify the formulation by setting $(\mathbf{T}^\mathbf{s})^{-1} = \mathbf{T}^\mathbf{s} = \mathbf{I}$, i.e. by constraining the rest-pose triangle to lie in the xy-plane with edge lengths 1. In the following discussion, we remove the pose-specific superscripts t, s.

We can separate the rotation and stretch components of the deformation gradient $\mathbf{D} = \mathbf{RS}$ by polar decomposition. The rotation $\mathbf{R}_j$ can be represented in axis-angle representation $\mathbf{r}_j$ with 3 components (using the log map to convert to, and the exponential map to convert from rotation matrices). In the particular case of $\mathbf{T}^\mathbf{s} = \mathbf{I}$, the stretch matrix $\mathbf{S}_j$ of a given triangle j has only 3 relevant components for representing xy-plane scaling and shear. These can be assembled

into a vector $\mathbf{s}_j = \left[S_{j,1,1}\, S_{j,2,2}\, S_{j,1,2}\right]^T$. Thus, the deformation gradient encoding maps from $\mathbb{R}^{3n}$ to $\mathbb{R}^{6m}$ where m is the number of triangles.

The deformation gradient encoding is translation invariant. To decode the deformation gradients and obtain the mesh, a reconstruction process stitches together the unconnected edge vectors found in the deformation gradients by providing positional information p_a for a set A of (at least) one anchor point. By expressing Eq. (2) as linear operator $\mathbf{G}$, the stitching process can be shown to be equivalent to Poisson reconstruction with boundary constraints [SP04],

$$\arg\min_{\mathbf{v}} \quad \|\mathbf{G}\mathbf{v} - \mathbf{D}\|_2 + \lambda \sum_{a \in A} (\mathbf{p}_a - \mathbf{v}_a)^2 . \tag{4}$$

For our special case of $\mathbf{T}^s = \mathbf{I}$, $\mathbf{G}$ has a particularly simple form with a row with 2 nonzero entries $(-1, 1)$ per edge in the mesh. As shown by Botsch et al. [BSPG06], the triangle normal in the third column of $\mathbf{D}$ is unnecessary for the solution of Eq. (4).

Decomposition into residual transformations The deformation gradient (2) of a triangle in each pose is decomposed further to yield a shape encoding better suited for learning. Let $\mathbf{R}_a$ be the pose-dependent rotation for the triangle given by blending articulated segment rotations using the blending weights and estimated body part rotations as discussed previously. The deformation of the triangle with respect to this simple pose-dependent transformation is given as

$$\mathbf{D}_a = \mathbf{R}_a \mathbf{D}_o = \mathbf{R}_a (\mathbf{R}_o \mathbf{S}_o) \tag{5}$$

where $\mathbf{D}_o = \mathbf{R}_o \mathbf{S}_o$ is the deformation gradient from the unit triangle in the xy-plane to the triangle in the template pose.

However, our reconstructed shape shows residual deformations that appear as both the rotation ($\mathbf{R}^*$) and stretch ($\mathbf{S}^*$) components

$$\mathbf{D} = \mathbf{R}_a (\mathbf{R}^* \mathbf{R}_o)(\mathbf{S}^* \mathbf{S}_o) \tag{6}$$

As discussed before, these residual deformations can be represented compactly using only 6 scalars per triangle: 3 for representing the rotation in axis-angle representation, and 3 for the stretch matrix. Together, these parameters in effect compose a residual deformation gradient $\mathbf{D}^* = \mathbf{R}^* (\mathbf{R}_o \mathbf{S}^* \mathbf{R}_o^T)$ that maps the skinned mesh to the observed mesh, such that $\mathbf{D} = \mathbf{R}_a \mathbf{D}^* \mathbf{D}_0$.

The final shape representation vector $\mathbf{y}$ that represents the residual transformations $(\mathbf{R}^*, \mathbf{S}^*)$ for all the triangles is given by

$$\mathbf{y} = \mathrm{vec}([\mathbf{s}_0 \cdots \mathbf{s}_m \; \mathbf{r}_0 \cdots \mathbf{r}_m]) . \tag{7}$$

6. Deformation Learning

The aim is now to learn a model that, given input parameters such as pose, body shape, and external load, can generate the shape that realistically shows the deformations as seen in the recorded data. For this, we assemble the dataset of f example meshes, encoded as vectors $\mathbf{y}_i$ with Eq. (7), into a training matrix $\mathbf{Y} \in \mathbb{R}^{f \times 6m}$; and equivalently with the corresponding l-dimensional input parameters $\mathbf{x}_i = [\theta_i, \beta_i, \gamma_i]$ to form $\mathbf{X} \in \mathbb{R}^{f \times l}$. The learning task is to estimate a function $\Psi(\mathbf{x})$ that is able to generate a shape vector from a novel, previously unseen $\mathbf{x}_a$. The function Ψ is learnt separately for each output parameter, i.e, the deformation mode of each triangle in the mesh.

An obvious choice for function Ψ is a linear regression model which can be trained by minimizing the square loss to the training examples and therefore results in the following linear least squares problem

$$\arg\min_{\mathbf{W} \in \mathbb{R}^{l \times 6m}} \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2 \tag{8}$$

The evaluation phase is then a simple matrix multiplication, $\Psi(\mathbf{x}_a) = \mathbf{W}^\top \mathbf{x}_a$. This linear model avoids overfitting and generalizes well along dimensions where limited training data is available, e.g. the body shape of the person. However, for the subset $l_\theta < l$ of pose-specific parameters (the joint angles θ), this fails to produce details such as fine concavities at certain elbow poses.

This necessitates learning additional non-linearities for a more accurate non-linear model. Since our captured dataset covers the space of l_θ pose parameters well, we can build such a model without overfitting. Non-parametric methods such as kernel ridge regression are well suited in this situation. The main idea is to replace the l_θ-dimensional input parameters by a higher dimensional feature vector $\theta_i \to \phi(\theta_i)$. The well-known kernel trick is applied and the dual problem is solved. To avoid overfitting with a huge number of features, regularization is needed,

$$\arg\min_{\mathbf{C} \in \mathbb{R}^{f \times 6m}} \|\mathbf{K}\mathbf{C} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{C}^T \mathbf{K} \mathbf{C}\|_F^2 , \tag{9}$$

where $\mathbf{K}(i,j) = \kappa(\theta_i, \theta_j)$ is the precomputed Gaussian kernel matrix, that is related to the non-linear map ϕ as follows

$$\phi(\theta_i)^T \phi(\theta_j) = \kappa(\theta_i, \theta_j) = \exp(-\|\theta_i - \theta_j\| / \sigma) . \tag{10}$$

During evaluation, the kernel ridge regression synthesizes pose-dependent shapes by computing $\Psi(\theta_a) = \mathbf{C}^\top \kappa(\theta, \mathbf{x}_a)$.

Training the kernel ridge regression, Eq. (9), requires solving a linear system of size $f \times f$ (for $6m$ right hand sides), which in our case of $f \approx 32{,}000$ is prohibitively expensive. Inspired by the work of [KK10], who face a similar problem in the scope of image super-resolution, we find a sparse approximation of the kernel ridge regression. The approach is based on finding a set of $b \ll f$ basis points $\mathbf{B} = [\mathbf{b}_1 \; \ldots \; \mathbf{b}_b]^\top, \mathbf{B} \in \mathbb{R}^{b \times l_\theta}$ from the pose input parameters. Then, the solution is estimated only with respect to those few basis points,

$$\arg\min_{\mathbf{C}_b \in \mathbb{R}^{b \times 6m}} \|\mathbf{K}_b \mathbf{C}_b - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{C}_b^T \mathbf{K}_0 \mathbf{C}_b\|_F^2 . \tag{11}$$

where $\mathbf{K}_0 = \kappa(\mathbf{B}, \mathbf{B})$ is the $b \times b$ pairwise kernel matrix in

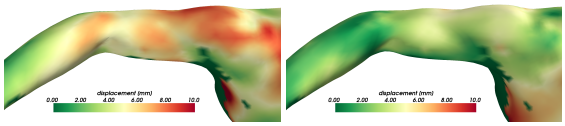


Figure 5: *Results on* shoulder lift *motion, colored by vertex displacement error from ground-truth. Left: Linear model. Right: our model.*

	ours	linear	[HSS*09]
curl low weight	5.57	9.29	24.23
curl high weight	5.94	8.86	24.87
shoulder spin	14.66	14.19	28.04
arm rotate	9.33	11.30	29.72
boxing	14.28	15.58	49.52
freestyle	9.00	10.20	30.80

Table 1: *Evaluation on motions not used for training. Average errors in vertex displacements (in* mm*) for our method, a linear model with our proposed shape encoding, and a linear regression with the encoding from [HSS*09].*

the basis set, and $\mathbf{K}_b(i,j) = \kappa(\mathbf{b}_i, \theta_j)$. The algorithm is also faster during synthesis, since the kernel needs to be evaluated only with the few basis vectors, $\Psi(\theta_a) = \mathbf{C}_b^\top \kappa(\mathbf{B}, \theta_a)$.

To obtain $\mathbf{B}$, we run K-Means clustering on the pose parameters of the training set and use the cluster center locations as the basis points. In the final model, we combine linear regression with the sparse kernel regression, and optimize

$$\underset{\substack{\mathbf{C}_b \in \mathbb{R}^{b\times 6m} \\ \mathbf{W}\in\mathbb{R}^{l\times 6m}}}{\arg\min} \left\| \begin{bmatrix} \mathbf{K_b} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{C_b} \\ \mathbf{W} \end{bmatrix} - \mathbf{Y} \right\|_F^2 + \lambda \|\mathbf{C}_b^T \mathbf{K}_0 \mathbf{C}_b\|_F^2 , \tag{12}$$

where $\mathbf{K_b}$ and $\mathbf{K}_0$ are formed by evaluating the pairwise kernel function κ only for the pose parameters θ_i of the training set; and only kernelized pose parameters are regularized. Given a new set of parameters $\mathbf{x}_a$ (with the subset θ_a denoting the pose parameters), the corresponding deformation can be synthesized as follows.

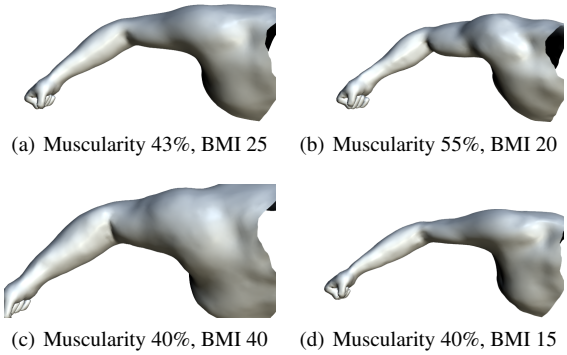
$$\Psi(\mathbf{x}_a) = \begin{bmatrix} \mathbf{C_b} \\ \mathbf{W} \end{bmatrix}^\top \begin{bmatrix} \kappa(\mathbf{B}, \theta_a) \\ \mathbf{x}_a \end{bmatrix} . \tag{13}$$

7. Results

Quantitative Evaluation As described in Sect. 4, we captured 10 male subjects in $30-40$ motions each. Each motion is roughly $80-100$ frames long, and our dataset consists of ca. $32,000$ meshes. Our model is trained on a set of 16 input parameters - 3 angles of rotation each at the shoulder and elbow joints, 3 body shape parameters given by height, BMI and muscularity, and 7 force parameters given by the $3-$dimensional force vector of the barbell weight in the local coordinate frames of the lower and upper arm and their magnitude. We quantitatively evaluate the accuracy of our model by cross-validating our result with the ground-truth on 6 reconstructed motions, each of a different subject, by leaving that motion out of the training set. We report the accuracy in the $\mathbb{R}^3$ Euclidean space of vertex coordinates. In order to recover 3D Euclidean vertex coordinates from our shape encoding, we need to re-align the generated mesh with ground truth rigidly before comparison, as our shape encoding is translation-invariant. Quantitative evaluation with respect to both these error metrics is reported in Tab. 1. For comparison purposes, we also train a purely linear model (given by Eq. (8)) and a purely linear model using a different shape encoding [HSS*09] on the same dataset. The shape encoding of [HSS*09] fails to provide comparative quantitative results since their encoding does not contain the base layer of a skinned skeleton, which results in misaligned body parts. Our method is also robust against flickering in some reconstructions, as visible in the supplementary video.

Our sparse kernel-learning method has two hyperparameters - the kernel size σ and the regularization strength λ. We observed that the learning method is robust to different hyperparameter settings, but the best results are obtained at $\sigma = 0.1$ and $\lambda = 10.0$. We tested different cluster numbers k in K-Means for finding the pose space basis points (64, $256, 512$ and 1024). More clusters yield slightly better accuracy, but take longer to evaluate. We found $k = 512$ as a good compromise. In Fig. 5, we visualize the error of our result against the linear model, both compared against the ground-truth 3D reconstruction on the *shoulder-lift* motion.

Visual Results As shown in the accompanying video, our model can be interactively used by an artist to synthesize complex muscle effects by changing pose, shape or external forces, and smoothly vary between the parameters. In Fig. 6, we show the results expressing plausible variation in arm shape with respect to BMI and muscularity parameters. In Fig. 7, we show the major strength of our model i.e, to simultaneously model external forces along with the other parameters. Our model generalizes well - the pose and external force parameters given to produce these effects are beyond the capture range in the training dataset. One can see physiological effects such as the sharp pronouncement of muscle strands in reaction to external forces (the *deltoid* and *biceps* muscles in Fig. 7(b), the *biceps* and *triceps* muscles in Fig. 7(c)), complex skin deformation around joint capsules (*scapula joint* at the back of the shoulder in Fig. 7(a)), and the marked difference in skin deformation with respect to the direction of the force vector (the loose skin around the elbow joint in Fig. 7(d)). These effects are achieved rapidly by the artist using our prototype tool through editing the skeletal pose, body shape and the force vector.



(a) Muscularity 43%, BMI 25 (b) Muscularity 55%, BMI 20

(c) Muscularity 40%, BMI 40 (d) Muscularity 40%, BMI 15

Figure 6: *Example body shape variations as produced by our statistical model, by controling BMI and muscularity.*

Performance At run time, new poses with our trained model can be produced extremely fast, requiring only simple matrix operations. Even our unoptimized Python implementation on the CPU provides realtime feedback. Owing to the strategy of sparse-kernel learning and the need to keep only a few basis vectors, it also has a low memory footprint (100 MB for 512 basis points). This makes it suitable for deployment in resource constrained applications, such as character animation in game consoles. The learning phase of our unoptimized code took 25 minutes over the dataset of ca. 32,000 meshes.

Discussion Our model generalizes well beyond the range of inputs captured in the training set, but starts to fail when these parameters are far off (e.g, external force that is twice the maximum captured barbell weight fails in many poses). Synthesized deformation is not perfect for poses not well-captured in the dataset due to occlusions, such as with the arm vertically upward. Capturing more people and training on a larger dataset would further improve the results. Capturing the effects of dynamics, such as the wobbling of body fat and soft tissue, is another important future work which can benefit from our contributions on 3D data acquisition. We used a simple skeleton structure, akin to those in animations for computer games, but achieved very natural and detailed animation results. Certain motions like shrugging of the shoulders cannot be modeled by this skeleton. Physiological skeletons and physics-based muscle models will further improve the accuracy of our method - e.g, subtle motions of clavicle and scapula that currently cannot be described by our control skeleton and only moderately visible when changing arm pose and external force, can be better modeled. Similarly, data-driven models like ours help to better initialize physiological muscle models to novel subjects. This potential for cross-fertilization needs to be explored in future work. Capturing with higher resolution cameras will help model even finer deformations (e.g, vertebral notches, clavicle bone). The acquisition and modeling paradigms of this paper can also be applied for capturing the whole human body, such that force-dependent muscle deformations are simulated by a data-driven model that is easy to control.

8. Conclusion

In this paper, we present the first comprehensive data-driven model for muscle deformation of the shoulder-arm region, spanning the parameter spaces of body pose, body shape and external forces. We propose an acquisition method and setup to capture a large training database of detailed muscle and skin deformations from real subjects. We further contribute by an effective semi-parametric non-linear approach to learn an expressive data-driven model, which can be evaluated in real-time and has a low memory footprint. The trained model can be used in a fast and intuitive-to-use prototype application to reproduce fine-scale, anatomically realistic muscle deformations at a high quality not shown by previous data-driven methods. Our approach is built on shape representations that easily fit into the work-flow of animators.

References

[ACP02] ALLEN B., CURLESS B., POPOVIĆ Z.: Articulated body deformation from range scan data. *ACM TOG 21*, 3 (2002), 612–619. 2, 4

[ACP03] ALLEN B., CURLESS B., POPOVIĆ Z.: The space of human body shapes: reconstruction and parameterization from range scans. *ACM TOG 22*, 3 (2003), 587–594. 2

[ACPH06] ALLEN B., CURLESS B., POPOVIĆ Z., HERTZMANN A.: Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In *Proc. SCA* (2006), pp. 147–156. 3

[ASK*05] ANGUELOV D., SRINIVASAN P., KOLLER D., THRUN S., RODGERS J., DAVIS J.: Scape: shape completion and animation of people. *ACM TOG 24*, 3 (2005), 408–416. 2, 3

[BBO*09] BICKEL B., BÄCHER M., OTADUY M. A., MATUSIK W., PFISTER H., GROSS M.: Capture and modeling of non-linear heterogeneous soft tissue. *ACM TOG 28*, 3 (2009). 2

[BM98] BREGLER C., MALIK J.: Tracking people with twists and exponential maps. In *Proc. CVPR* (1998), pp. 8–15. 6

[BSPG06] BOTSCH M., SUMNER R., PAULY M., GROSS M.: Deformation transfer for Detail-Preserving surface editing. In *Proc. VMV* (2006), pp. 357–364. 7

[Eno08] ENOKA R. M.: *Neuromechanics of Human Movement*, 4 ed. 2008, ch. 3, p. 122. 3, 4

[GR96] GOLD S., RANGARAJAN A.: A graduated assignment algorithm for graph matching. *IEEE Trans. PAMI 18* (1996), 377–388. 4

[HAR*10] HASLER N., ACKERMANN H., ROSENHAHN B., THORMAHLEN T., SEIDEL H.-P.: Multilinear pose and body shape estimation of dressed subjects from image sets. In *Proc. CVPR* (2010), pp. 1823–1830. 3

[HPH10] HONG Q. Y., PARK S. I., HODGINS J. K.: A data-driven segmentation for the shoulder complex. In *CGF (eurographics)* (2010). 2

[HSS*09] HASLER N., STOLL C., SUNKEL M., ROSENHAHN B., SEIDEL H.-P.: A statistical model of human pose and body shape. In *CGF (eurographics)* (2009), vol. 2. 3, 5, 8

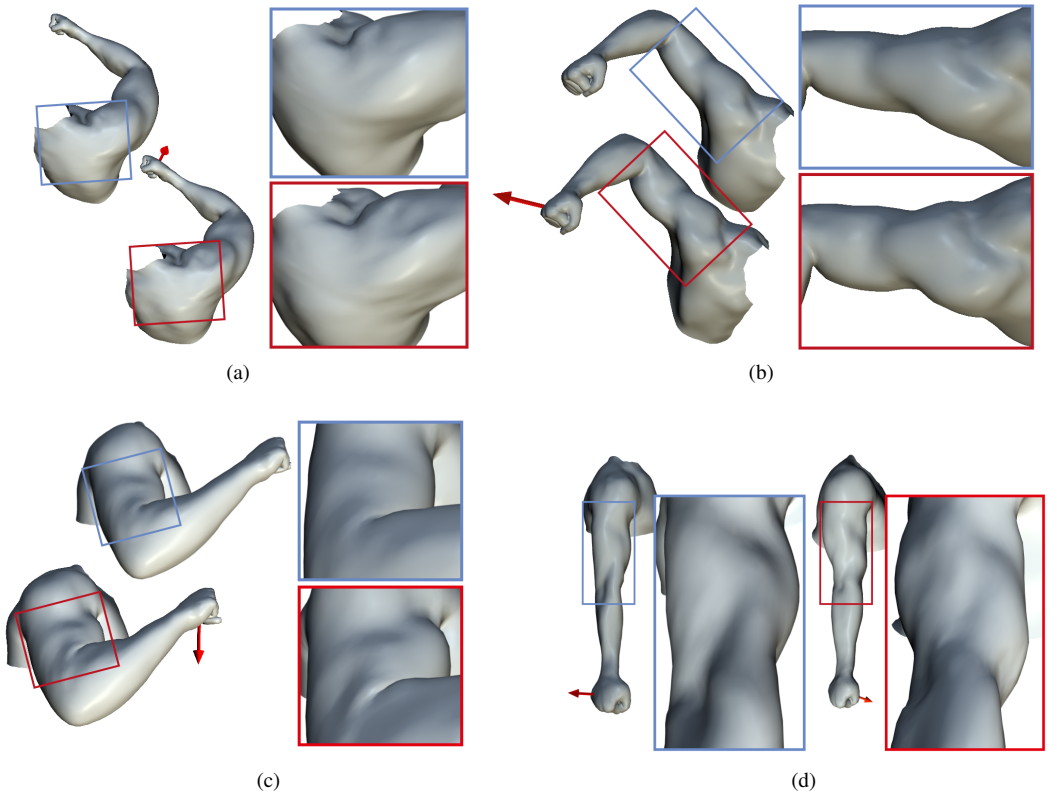


Figure 7: *Artist-driven muscle deformations to poses and large external forces not captured in our dataset: a) Deformation around the* scapula *bone that counteracts some of the force; b) the* deltoid *and* biceps *muscles get sharply defined by force; c)* triceps lateral head *and* biceps long head *get pronounced with force; d) opposing activations of* biceps *and* triceps *as the force vector reverses direction.*

[KK10] KIM K. I., KWON Y.: Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans. PAMI 32*, 6 (2010), 1127–1133. 7

[LCF00] LEWIS J. P., CORDNER M., FONG N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *ACM TOG* (2000), pp. 165–172. 3, 6

[LGK*12] LEE D., GLUECK M., KHAN A., FIUME E., JACKSON K.: Modeling and simulation of skeletal muscle for computer graphics: A survey. *Foundations and Trends in Computer Graphics and Vision 7*, 4 (2012), 229–276. 1, 3

[LST09] LEE S.-H., SIFAKIS E., TERZOPOULOS D.: Comprehensive biomechanical modeling and simulation of the upper body. *ACM TOG 28*, 4 (2009), 99:1–99:17. 3

[PH06] PARK S. I., HODGINS J. K.: Capturing and animating skin deformation in human motion. *ACM TOG 25*, 3 (2006). 2

[PH08] PARK S. I., HODGINS J. K.: Data-driven modeling of skin and muscle deformation. *ACM TOG* (2008). 2, 3

[SA07] SORKINE O., ALEXA M.: As-rigid-as-possible surface modeling. In *Proc. SGP* (2007), pp. 109–116. 5

[SKP08] SUEDA S., KAUFMAN A., PAI D. K.: Musculotendon simulation for hand animation. *ACM TOG 27*, 3 (2008). 3

[SKR*06] STOLL C., KARNI Z., RÃŰSSL C., YAMAUCHI H., SEIDEL H.-P.: Template deformation for point cloud fitting. In *Proc. SPBG* (2006), pp. 27–35. 5

[SNF05] SIFAKIS E., NEVEROV I., FEDKIW R.: Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM TOG* (2005), 417–425. 3

[SP04] SUMNER R. W., POPOVIĆ J.: Deformation transfer for triangle meshes. *ACM TOG 23*, 3 (2004), 399–405. 3, 6, 7

[SRC01] SLOAN P.-P. J., ROSE III C. F., COHEN M. F.: Shape by example. In *Proc. I3D* (2001), pp. 135–143. 6

[TSB*05] TERAN J., SIFAKIS E., BLEMKER S. S., NG-THOW-HING V., LAU C., FEDKIW R.: Creating and simulating skeletal muscle from the visible human data set. *TVCG 11*, 3 (2005), 317–328. 3

[VBPP05] VLASIC D., BRAND M., PFISTER H., POPOVIĆ J.: Face transfer with multilinear models. *ACM TOG* (2005), 426–433. 2

[WPP07] WANG R. Y., PULLI K., POPOVIĆ J.: Real-time enveloping with rotational regression. *ACM TOG 26*, 3 (July 2007). 3

[WSLG07] WEBER O., SORKINE O., LIPMAN Y., GOTSMAN C.: Context-aware skeletal shape deformation. *CGF (eurographics) 26*, 3 (2007), 265–273. 3, 6