Coregistration: Simultaneous Alignment and Modeling of Articulated 3D Shape

David A. Hirshberg, Matthew Loper, Eric Rachlin, Michael J. Black

Max Planck Institute for Intelligent Systems, Tübingen, Germany

Abstract. Three-dimensional (3D) shape models are powerful because they enable the inference of object shape from incomplete, noisy, or ambiguous 2D or 3D data. For example, realistic parameterized 3D human body models have been used to infer the shape and pose of people from images. To train such models, a corpus of 3D body scans is typically brought into registration by aligning a common 3D human-shaped template to each scan. This is an ill-posed problem that typically involves solving an optimization problem with regularization terms that penalize implausible deformations of the template. When aligning a corpus, however, we can do better than generic regularization. If we have a *model* of how the template can deform then alignments can be regularized by this model. Constructing a model of deformations, however, requires having a corpus that is already registered. We address this chicken-and-egg problem by approaching modeling and registration together. By minimizing a single objective function, we reliably obtain high quality registration of noisy, incomplete, laser scans, while simultaneously learning a highly realistic articulated body model. The model greatly improves robustness to noise and missing data. Since the model explains a corpus of body scans, it captures how body shape varies across people and poses.

1 Introduction

Strong 3D shape priors enable robust and accurate inference. Building strong shape priors from data, however, is difficult, even with dense accurate measurements from high-end 3D scanners. Methods for learning shape models depend on identifying corresponding points across many 3D scans; that is, registered data. Establishing such correspondences is particularly difficult for articulated shapes such as people with highly varied shapes and poses. Across such scans one can usually identify some key anatomical landmarks (*e.g.* the tip of the nose, the corners of the eyes) but there are large regions of the body where it is difficult to find useful correspondences, automatically or manually.

Registering a corpus of human body scans involves bringing them into alignment with a common human-shaped template. In this process, the template is typically aligned (i.e. deformed) to each scan in isolation (Fig. 1). The traditional data likelihoods used to guide template-to-scan alignment measure distance between the two surfaces and distance between corresponding localizable landmarks. This leaves the alignment severely under constrained away from landmarks; motion of non-landmarked template vertices along the scan surface is not
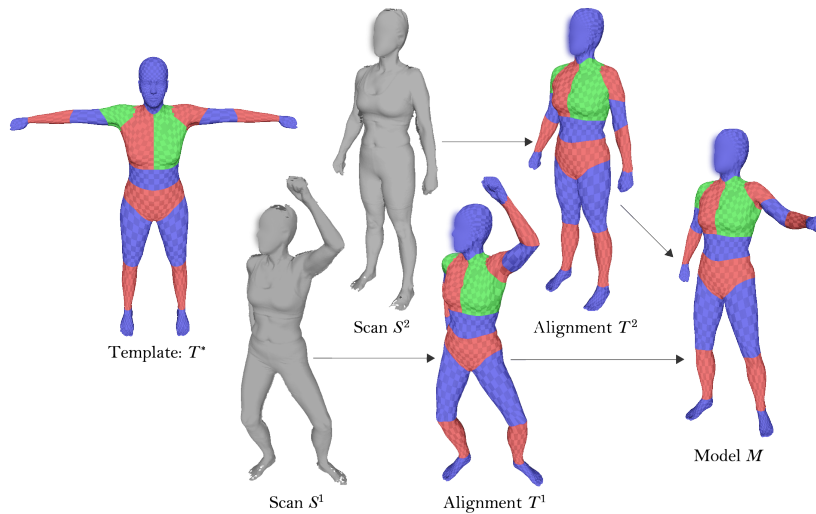
Fig. 1: **Corpus registration and learning.** A person-shaped template mesh T^* is aligned to several laser scans, S^1 and S^2, of a single person in multiple poses. The alignments T^1 and T^2 should accurately reflect the shape of each scan and be in good point-to-point correspondence. From the registered training data an articulated model, M, is learned that accurately expresses new poses.

measured, allowing these vertices to be positioned inconsistently (i.e. slide) across scans. If available, image texture can provide an additional cue, especially when motion between scans is small, but many scanners do not provide texture data, and it may not be useful when registering scans of different people.

To provide plausible alignments of ambiguous meshes, existing alignment algorithms tend to employ simple priors motivated by analogy to deformations of physical objects – the template should deform elastically like rubber, or smoothly like taffy. When registering scans with a common template mesh, such priors yield geometric regularization terms that prevent the template from undergoing wildly implausible deformations. Unfortunately, it is difficult to get adequate constraints from these priors while retaining the flexibility to fit a wide range of poses and shapes.

In this paper, we offer an alternative that regularizes alignments using an articulated 3D model of human shape and pose. Unlike previous work that uses a simple articulated model for initialization [1, 2], we use a richer, learned, model of body shape for accurate registration. Our key novelty is to *integrate the learning of this model with the registration process.*

This concurrent approach to scan alignment and model building, which we term *coregistration*, offers a significant advantage over the traditional two-stage approach in which one first aligns each training example with the template mesh, and then builds a shape model from the aligned data. It provides a well-defined

objective function for registering a corpus of scans. As explained in Sec. 4, a good set of alignments must tightly fit the surface of each scan, while simultaneously remaining consistent with an underlying articulated shape model. A surface fitting term ensures that each alignment accurately captures the shape of each scan, while a model term ensures that template registration remain consistent across scans. The value of simultaneously aligning a corpus comes from: 1) the ability to leverage good data to explain poor or missing data; 2) the consistency of each individual's shape across poses; 3) similarities in the deformations of different bodies as they change pose; 4) the presence of ambiguities in pairwise alignment that can be resolved by modeling a class of objects.

Consider an area like the armpit or crotch where data is often missing in laser scans. If the person is seen in multiple poses, some of these may have complete data in these regions. Coregistration allows the shape learned with good data to constrain the alignment, while ignoring missing data. This is in contrast to most traditional methods, which rely on generic regularization in each scan to deal with this missing data. The few methods which describe model-specific regularization do not propose or optimize an objective function which encompasses both the quality of the model and that of the alignments.

To demonstrate the effectiveness of coregistration we apply it to multiple collections of 3D body data and obtain accurate alignments and highly-realistic articulated models for all subjects.

2 Prior Work

There is a huge literature on general mesh registration methods which is too large to review here. Additionally many of these methods focus on registering (*i.e.* aligning) only two meshes. Here we focus on the human body and the simultaneous registration of many meshes with a common template; this *corpus registration* problem is quite different and admits our solution that integrates registration with model building. Furthermore, most methods for registering many human bodies focus on aligning a template to bodies of different shapes in a canonical pose [3–5]. Here we focus on registering many different bodies in different poses, which is a much harder problem and has received less attention [1, 2]. It is worth noting that the registration of human bodies remains a challenging task despite many published methods.

Building a model of the body typically has three stages: 1) initialization, 2) registration, and 3) model building. Initialization is often done using manually placed markers on the scans and the template mesh [1, 3] but there have been several automated methods proposed [2, 4]. While the problem of reliably obtaining rough initializations is important, we do not address it here. The registration stage deforms the template to closely match the scan and must deal with missing data and noise. Previous work relies on fairly generic spatial smoothness constraints. The third stage uses registered scans to learn a body model or perform statistical analysis of body shape. Here we combine these last two stages and show that this results in both better models and better alignments.

2.1   Registration

3D scans of articulated, highly non-planar objects like hands and bodies are typically aligned using nonrigid ICP [1, 3, 6] or variants using soft or sliding correspondences [7, 8]. Faces, more often than not, are aligned using nonrigid ICP as well [9, 10], but there is history of face registration in 2D parameterizations using optical flow [11]. When a template is deformed to fit scan data, good point-to-point registration is typically maintained using simple, physically-motivated regularization terms on the deformation of the template.

The common regularization terms used during mesh registration fall into two classes. Both act on the deformations of the template surface. These are 'smoothness' terms, which penalize deformations changing rapidly over the surface of the template, and 'as rigid as possible' terms, which penalize local estimates of the deformation as they deviate from rigidity. As rigid as possible regularization is used for near-isometric deformations, such as those that occur when aligning scans of the same person [6, 8], and for interactively manipulating a mesh [12, 13] while maintaining its perceived character. Maintaining rigidity is less desirable when aligning different body shapes, so corpus registrations [1, 3, 14] usually rely only on smoothness regularization. Smoothness provides only a second order constraint on mesh registration; adjacent triangles must deform similarly. Smooth template deformations can still result in inconsistent registration across scans. Neither 'smoothness' nor 'as rigid as possible' regularizations are as informative as having an accurate class-specific shape model.

Anguelov *et al.* [15] automatically detect initial correspondences between 70 scans of one person in different poses and then perform registration using standard regularization. They do not register multiple people in different poses but Wuhrer *et al.* [2] do. They use automatically-detected landmarks to estimate a rough body pose and then use this to deform a linear-blend-skinned template. This provides a reasonable starting point for traditional, regularized, pairwise mesh registration. They do not address model learning.

The above registration methods employ a static template for regularization. However, Amberg [9] incorporates a deformable model into his method for registering a head template to face scans. The model is coupled to the aligned template by a prior that measures smoothness of the deformation between the registration and an optimized fit of the deformed model.

2.2   Human Shape Models

Once a series of 3D scans have been registered to a common template, standard multivariate statistical methods can be used to model the distribution of shapes. In the case of faces [9–11] and bodies scanned in a single pose [3], low-dimensional models have been obtained by performing principal component analysis (PCA) on the aligned vertices. For scans of multiple poses, articulated body models have been proposed that represent both the shape and pose of each registered scan [1, 15, 16]. We employ the SCAPE model here [15], although our method readily generalizes to other models.
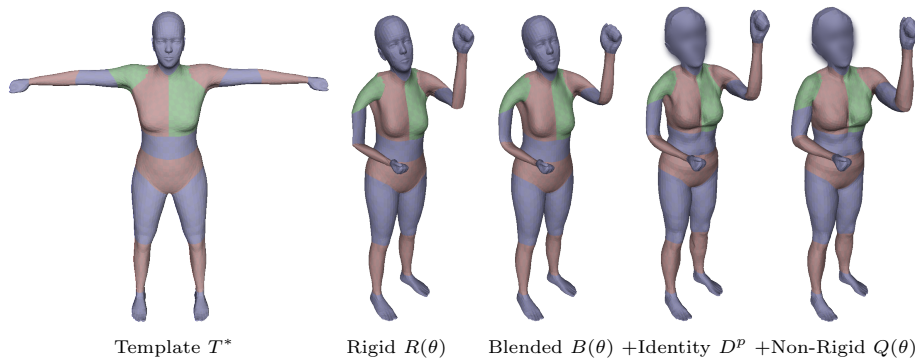
Template T^* Rigid $R(\theta)$ Blended $B(\theta)$ +Identity D^p +Non-Rigid $Q(\theta)$

Fig. 2: **SCAPE and BlendSCAPE.** Deforming a template with BlendSCAPE; see text.

When a model is learned from registered scan data, the quality of that model is heavily dependent on the quality of registration. In some cases, registrations have been improved by using them to learn a model, then using the model to initialize a second round of registration. Blanz and Vetter [11] use such iteration to improve correspondence in their optical flow-based registration method.

In contrast we define the first objective function that combines model building and registration. This approach brings a corpus of differently shaped bodies in varying pose into registration while learning a model of human pose-dependent shape variation. The result is quantitatively better registrations and a better articulated body shape model.

3 SCAPE and BlendSCAPE

SCAPE [15] is a model of human body shape learned from registered scans. It defines how to deform a human-shaped triangulated template mesh, T^*, to take on different poses and identities (body shapes). Let T^* be pre-segmented into parts (color coded in Fig. 2) connected in a kinematic tree structure. The relative rotations between neighboring parts in the tree are represented as Rodrigues vectors. Let θ be a vector collecting all the relative rotations and $R(\theta)$ represent the absolute rotations of the triangles in a part. In SCAPE, every triangle within a part has the same rotation. Let D represent the deformations that transform T^* into the shape of a specific person. Finally, let $Q(\theta)$ define non-rigid deformations of the triangles of T^* that capture shape change as a function of pose θ.

The template is deformed in three steps. First T^* is decomposed, or "unstitched," into disconnected triangles, T_f^*. Each unstitched triangle is represented by a pair of its edge vectors, "forgetting" its location but retaining its shape and orientation. Second, each unstitched triangle is individually deformed according to a sequence of pose- and shape-dependent 3×3 linear deformations. Each unstitched triangle T_f^* is "posed" by a rotation $R_f(\theta)$ and deformed to represent a

person's body shape using a 3×3 matrix D_f. It is also deformed by a 3×3 matrix $Q_f(\theta)$ that accounts for pose-dependent shape changes like muscle bulging and skin wrinkling and corrects for deviations between the rigidly posed model and the true shape. A transformed triangle is written

$$T_f = R_f(\theta)D_f Q_f(\theta)T_f^*. \tag{1}$$

These deformed triangles are recomposed, or "stitched," to define the vertices of a watertight mesh, $M(\theta, D, Q)$. Because triangles are transformed independently, and will disagree at shared edges, we solve for the final vertex locations of the mesh using least-squares (cf. [15, 17]).

SCAPE uses a partition of the template triangles into rigid parts to define its model for R. Since each part is independently rotated, the final stitched body surface can collapse, crease or fold near joints (see Fig. 2). Q can be trained to correct these artifacts given artifact-free alignments, but if these artifacts are sufficiently severe they can cause convergence problems during coregistration.

To address this we introduce a *BlendSCAPE* model, in which each triangle's rotation is a linear blend, $B_f(\theta) \stackrel{\text{def}}{=} \sum_i w_{fi} R^i$, of the rotations, R^i, of the parts, indexed by i, in the kinematic tree. These weights, w_{fi} can be estimated along with the other parameters of the model, but in this work we define them manually by smoothing our SCAPE segmentation across part boundaries. The template posed with BlendSCAPE is shown in Fig. 2. Clearly $B_f(\theta)$ itself does not provide a realistic model of body pose deformation, but rather reduces the work that Q must do to correct its errors. We have found that this improves model fitting to scans and, consequently, registration (Fig. 3)

We assume a corpus of body scans containing multiple people, each in multiple poses, and that we know which scans correspond to which people. After coregistration, each scan is modeled by a person-specific D^p that represents that individual's body shape, a scan specific pose, θ^s, and a pose-dependent $Q(\theta^s)$ for each scan in which the function Q is the same all across people. As in previous work [15], the deformation Q is a linear function of the Rodrigues vectors describing the relative orientations of adjacent parts: $Q(\theta) = Q^0 + \sum_c \theta_c Q^c$ where θ_c is the cth element of the pose vector θ, and Q^0, Q^c contain the linear coefficients and are learned from the corpus of registered bodies. This model is constrained so only the orientations of parts near a triangle contribute to its deformation (*i.e.* Q^c is kept sparse).

Previous SCAPE models [15, 18] have been built using two body scan corpora: one containing people of different shapes in roughly a fixed pose and the other containing one person in many poses. This is in contrast to Hasler *et al.* [1] who train a model with correlation between shape and pose using scans of several people in different poses. Here we learn the first SCAPE pose model, Q, trained from multiple people in multiple poses. This improves our ability to model the deformations of different people. In summary, we approximate a scan in the corpus with a model $M(\theta^s, D^p, Q)$ that poses the model using $B(\theta^s)$, deforms it to the identity of the person using D^p, and accounts for non-rigid shape changes using Q, which are a function of pose, θ^s.
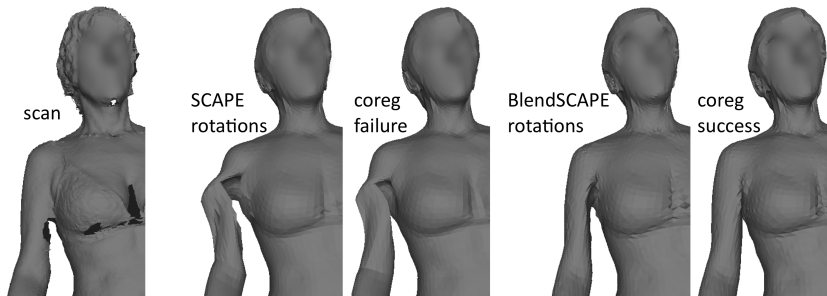
Fig. 3: **SCAPE vs BlendSCAPE.** Rigid part rotations, R, can introduce significant artifacts early in the fitting process, particularly when scan noise (*e.g.* holes) coincides with part boundaries of the model. Coregistration has difficultly eliminating these artifacts. In contrast, B from the BlendSCAPE model introduces smoother, less significant artifacts, which coregistration rapidly corrects

4   Coregistration

Coregistration aligns a triangulated template mesh to a corpus of 3D scans while simultaneously training a BlendSCAPE model. Below we define a data penalty term that seeks to deform our template to match a scan and a novel coupling term that constrains this deformation to be similar to a learned BlendSCAPE model. Optimization involves solving for both the alignment and the model parameters.

To train our model, we must estimate a pose θ^s for each scan in the corpus, a shape D^p for each person in our corpus, and a single linear pose-dependent deformation model $Q(\theta)$. Once coregistration is complete, each scan should be tightly fit by a deformed template mesh and should also closely match the corresponding BlendSCAPE body, $M(\theta^s, D^p, Q)$. Note that before training we have an "untrained" BlendSCAPE model in which D and Q are the identity. At the start of coregistration, we roughly align the template by posing and scaling the untrained BlendSCAPE model. For this step we use a set of landmarks associated with each scan (cf. [2]). Note, however, during coregistration the landmarks are discarded, in contrast to [3].

Given a scan S, we define the following data term, E_S, evaluating the fit of the deformed template T to the surface of the scan S

$$E_S(T; S) = \frac{1}{a_S} \int_{x_s \in S} \rho \left(\min_{x_t \in T} ||x_s - x_t|| \right) \tag{2}$$

where ρ is the Geman-McClure robust error function $\rho(x) = \frac{x^2}{\sigma^2 + x^2}$ [19], S is the scan surface, a_S is the scan's surface area, and T is the surface of the aligned template. We approximate the data error using a fixed set of locations x_s, uniformly sampled over the surface of the scan S. It is also possible to add a landmark term into E_S that would constrain known locations on the template to be close to measured locations on the scan.

To address the correspondence ambiguity inherent to E_S, we add a coupling term penalizing differences between the aligned template and the current model

$$E_C(T, \theta, D, Q) = \sum_f a_f \left\| \underbrace{T_f}_{\substack{\text{unstitched} \\ \text{alignment triangle}}} - \underbrace{B_f(\theta)D_f Q_f(\theta)T_f^*}_{\substack{\text{unstitched} \\ \text{model triangle}}} \right\|_F^2 \quad (3)$$

where T_f represents the pair of edge vectors of the unstitched triangle f of T, $B_f(\theta)D_f Q_f(\theta)T_f^*$ is the corresponding unstitched triangle of $M(\theta, D, Q)$, and a_f is the area of f on the template mesh, T^*. The squared Frobenius norm is used to measure the difference between corresponding unstitched triangles of T and $M(\theta, D, Q)$. This is simply the sum of squared distances between corresponding pairs of edge vectors.

Additionally, we use simple regularization terms to constrain the body shape deformations, D, and the pose-dependent deformation model, Q. The first term promotes spatial smoothness of the deformations, D, that map the template mesh to an observed person (cf. [9]). The second term penalizes the magnitude of the effect of the pose-dependent deformation model

$$E_D(D) = \sum_{\text{adjacent faces } i,j} a_{ij} \frac{||D_i - D_j||_F^2}{h_{ij}^2}$$

$$E_Q(Q) = \sum_{\text{faces } f} a_f \left(||Q_f^0 - I||_F^2 + \sum_c ||Q_f^c||_F^2 \right). \quad (4)$$

Here h_{ij} is the distance between the centroids of template triangles i and j, a_f is the area of triangle f, and $a_{ij} = \frac{a_i + a_j}{3}$ is the area of the diamond-shaped region defined by the centroids of triangles i and j and the endpoints of their shared edge.

We also use a weakly informative pose prior, E_θ, which penalizes deviation from the template pose. This regularizes the pose when the scan provides little useful information.

If D and the function Q were known, a single scan could be reliably aligned by optimizing $E_S(T; S) + E_C(T, \theta; D, Q)$. Since D and Q are not known, coregistration seeks to align all scans in parallel while simultaneously solving for D and Q across scans. Summing over all scans and adding our model regularization yields the following *coregistration optimization problem*

$$\min_{\{T^k\}, \{\theta^k\}, \{D^p\}, Q} \sum_{\text{scans } k} [E_S(T^k; S^k) + \lambda_C(E_C(T^k, \theta^k, D^{p_k}, Q) + \lambda_\theta E_\theta(\theta^k))]$$

$$+ \lambda_C[\lambda_D \sum_p E_D(D^p) + \lambda_Q E_Q(Q)]. \quad (5)$$

Here p indexes people, k indexes scans, and p_k identifies the person in each scan. The λ's control the relative influence of terms. λ_C is particularly important; it controls how much the alignments can deviate from the model.

5   Optimization

Our objective function is non-linear and the state space of solutions is very high-dimensional. Fortunately its structure admits a tractable alternating optimization scheme. Fixing the shapes D^p and the pose-dependent deformation model $Q(\cdot)$ decouples the scans; we minimize (5) by solving one non-linear problem of the form $\min_{T^k,\theta^k} E_S(T^k; S^k) + \lambda_C(E_C(T^k, \theta^k; D^{p_k}, Q) + \lambda_\theta E_\theta(\theta^k))$ for each scan. In essence, these subproblems are standard pairwise registration problems with a strong regularization toward the posable model (i.e. $\min_\theta E_C(\cdot, \theta; D^{p_k}, Q)$). We solve these subproblems using MATLAB's `lsqnonlin` (MathWorks, Natick MA). Solving 8 such problems takes 3 minutes on an 8-core Opteron processor.

With all T^k and $Q(\cdot)$ fixed, minimization with respect to each person's D^p is an independent linear least squares problem for each person p. Similarly, with all T^k and D^p fixed, minimization with respect to $Q_f(\cdot)$ is an independent linear least squares problem for each triangle f. These sparse least squares problems can be solved efficiently, thus our method's runtime largely depends on its rate of convergence and our ability to compute registration subproblems in parallel.

We initialize coregistration by fitting an untrained BlendSCAPE model to each scan using E_S and landmark correspondences. This simple model uses a trivial pose-dependent deformation model $Q_f(\theta) = I \quad \forall \theta$. Pose is allowed to vary freely, but shape varies only by isotropically scaling the template. The model fit to scan S^k initializes T^k and θ^k. Each person's shape D^p is initialized by averaging the scale of the fits for their scans. Q is initialized to the identity.

It is useful to perform the optimization in stages. We begin with a low coupling weight λ_C so that the crude initial model provides only a rough guide to the registration. We then increase λ_C from 0.25 to between 1 and 5 over several iterations, tightening the fit of the model to the scans. In each iteration, we minimize w.r.t. T^k and θ, then w.r.t. D and Q. As λ_C increases, the estimated model has more influence on the alignments, which enables information from good alignments to inform the registration of noisy scans. In addition, we gradually decrease the scale parameter σ of the robust error function in E_S, as is frequently done with non-convex error functions; σ, starts at 1 meter and decreases to $5cm$, $1cm$, and $5mm$. We observe that the result are not very sensitive to the precise sequence of values of these parameters, or to whether intermediate optimization steps are run to convergence.

6   Experiments

To demonstrate the accuracy and robustness of coregistration, we register several body scan corpora. Each corpus consists of multiple individuals in a wide range of poses. By visual appraisal, at least 96% of the scans in each corpus are registered well, and we obtain high quality models from both corpora. Images of the alignment to each scan, as well as an examination of our registration failures, are available in [20]. No scans were excluded due to registration failure.

Quantitative analysis. For quantitative evaluation we use a dataset of 124 scans of two females in a wide range of standing and sitting poses. One of the two women was scanned during two separate sessions two years apart with different clothing and different hair styles. For the purpose of coregistration, the corpus was treated as containing three individuals, each with distinct body shapes. This dataset has extremely challenging poses (more varied than [1]), scans with significant holes, and hand-placed landmarks that allow evaluation.

Initialization (see Sec. 5) used twelve hand-placed landmarks on each scan. Coregistration was then run as described in Sec. 4 without any landmarks. In eight iterations, good registrations were obtained for all but four scans. Hands were sometimes slightly misaligned, as hand scan data was often quite noisy. Figure 4 shows plots of two indicators of registration quality over the course of the coregistration process. The "Distance from scan" captures how faithful the alignment is to the scan surface. We use statistics of the distance between uniformly sampled points on the scan surface and their nearest points on the aligned template. Because an alignment can be close to a scan without having anatomically meaningful *correspondence*, we use "Landmark prediction error" to rate correspondence as well. Twenty-four hand-designated landmark vertices (which were not used in registration) are used to predict the location of anatomical landmarks on the scan. These predictions are compared to the locations of these landmarks marked on each scan by a human annotator.

In the first iteration, the alignment surface snaps to within about $1mm$ of the scan, but the alignment-scan gap widens afterward. The alignments are pulled toward shapes representable by the model as the alignment-model coupling constant λ_C increases between iterations 1 and 3. This results in alignments with better correspondence, as seen by the decrease in landmark prediction error and model to scan error. For evaluation, we withhold 30 scans of the same individuals. The model's ability to fit these held out scans improves with each iteration (see the green lines on Fig. 4).

Figure 5 shows some representative alignments and models. Note that many of the scans contain significant amounts of noise and missing data; e.g. the chest and back of the bent-over subject. Coregistration is able to use the data present in a scan and to propagate information learned from other scans. We consider the lower right example of a failure of our model. While the registration is good, the model does not fully capture the deformation of the shoulders.

In order to compare coregistration with existing corpus registration methods, we also registered our corpus of 124 scans using two algorithms representative of the methods discussed in Sec. 2.1. In `Algorithm I` each scan is registered independently using traditional "model-free" registration, and then all registrations are used to learn a model using the same optimization performed in the learning stage of coregistration. Model-free registration is performed using scan-to-mesh distance E_S, twelve landmark points, and a nonlinear smoothness regularization from [9]. In `Algorithm II`, `Algorithm I` is iterated as in [11]. After each iteration, the resulting model is fit to each scan and used to reinitialize a fresh run of `Algorithm I`.

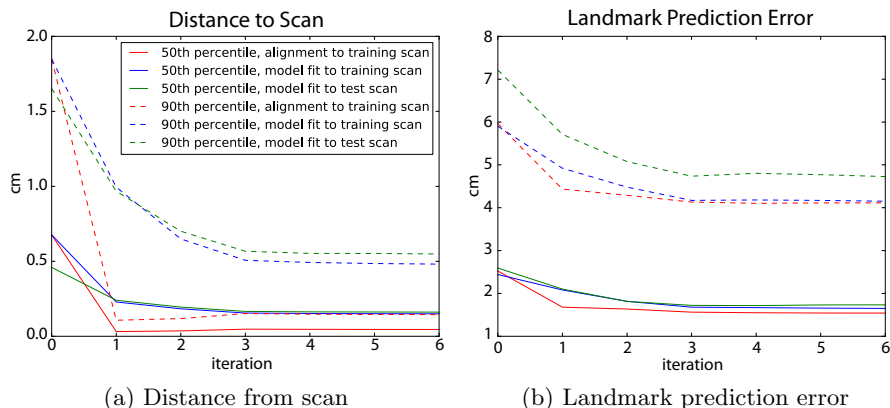(a) Distance from scan   (b) Landmark prediction error

Fig. 4: **Convergence of quality indicators.** Iteration 0 is initialization. Only six iterations are shown as there was no visible change afterward

All methods yield a registration of the model template and a model fit to each scan. Coregistration *alignments* give more accurate predictions of the 24 evaluation landmarks, with a mean landmark error of $2.0 \pm 2.1 cm$ versus 3.0 ± 2.8 for `Algorithm I` and 2.7 ± 2.7 for `Algorithm II`. Coregistration also yields better models. Models trained using coregistration are better able to fit scans, with a mean scan to model-fit distance of $0.25 \pm 0.32 cm$ on our 30 test scans. Algorithms `I` and `II` have distances of 0.34 ± 0.43 and 0.29 ± 0.31 respectively. Coregistration *models* give a mean landmark prediction error of $2.4 \pm 2.8 cm$ on the 30 test scans, whereas the models generated by `I` and `II` have errors of 3.4 ± 6.3 and 2.9 ± 2.3.

Large scale registration. To evaluate our method on a larger corpus with a wider range of body shapes, we register a publicly available set of scans provided by Hasler *et al.* [1]. The dataset contains 337 scans of 34 different women in 35 poses. Hasler *et al.* provide alignments as well, which we use to to obtain 36 rough landmark vertices on each scan for initialization. We observe only six bad registrations, each to a scan of a different woman. Five are in forward bend poses, in which large portions of the face and chest are missing from the scan. These failures do not appear to impact the model's ability to accurately capture the shapes, D, of the six women. Three successful examples are shown in Fig. 6. Note the detail and reduction in artifacts in the coregistered meshes.

Improving existing registrations. Because coregistration is able to integrate information from multiple scans of the same person and multiple people in different poses, it can be used to improve extant registered meshes without access to the original scans. We randomly selected 4 female subjects with 10 poses each from the Hasler *et al.* dataset. By fitting our model to a small number of these registrations, we estimate a correspondence between their template and ours. We use
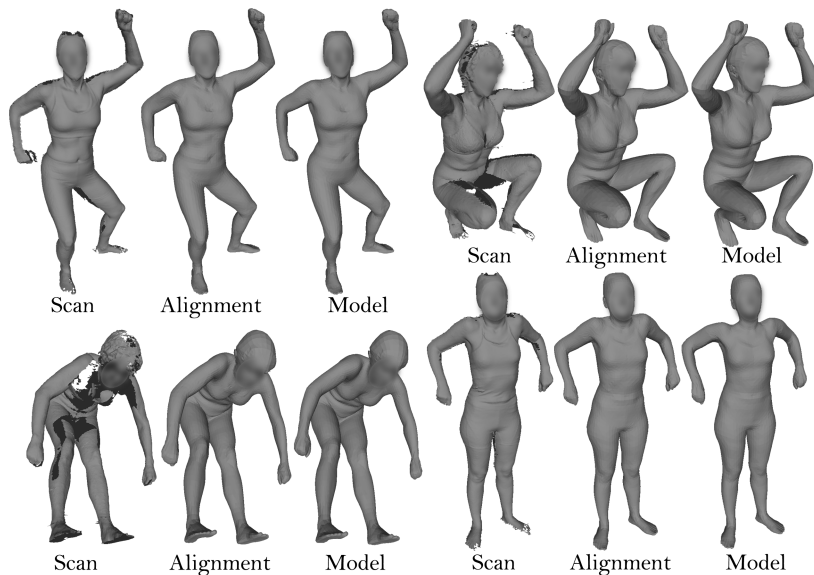
Fig. 5: **Examples from coregistration.** Sample scans, alignments, and model fits from coregistration of 124 scans of three subjects across a range of poses

this correspondence to initialize T^k for every body and then use coregistration to learn a model and registration to their *registered meshes*. Registering registered meshes may seem odd, but it has two effects: 1) it *denoises* the existing alignments and 2) it learns a model from them. Figure 7 shows examples of the original registrations and our refinement.

7   Conclusion

In this paper we address the corpus registration problem by approaching modeling and alignment simultaneously. Our algorithm for 'coregistration' incorporates a BlendSCAPE term into our registration objective function. This allows us to optimize over both aligned template meshes and over a shape model, offering a significant advantage over the traditional two-stage approach to model learning. By providing a well-defined, model-based objective function that a collection of registered meshes should minimize, coregistration allows shape information learned from good data to correct for missing data. To demonstrate the effectiveness of coregistration, we have registered several collections of 3D scans. Coregistration results in high quality alignments and a realistic BlendSCAPE model learned from multiple individuals.

Coregistration produces a high quality SCAPE model applicable only to the registered individuals. A natural extension of this work would be to learn not just the shape D of each individual, but also a low dimensional shape space capable
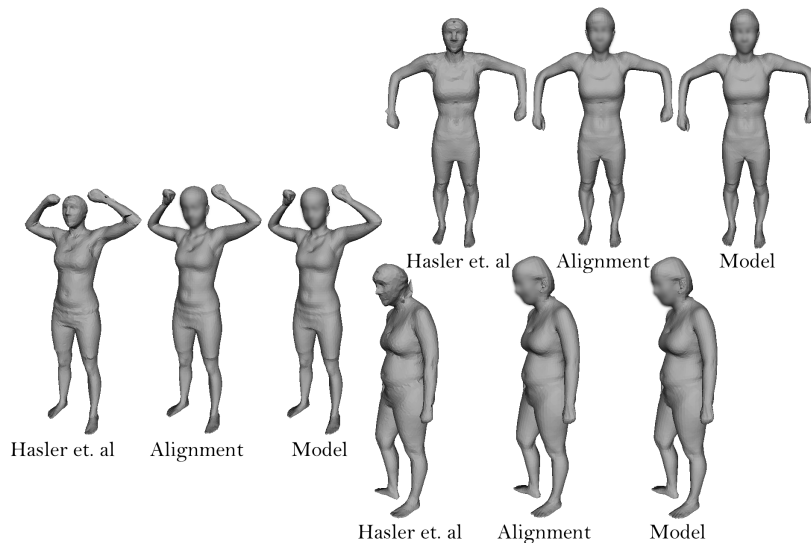
Fig. 6: **Coregistration of Hasler et al. data.** Each example shows, from left to right, one of the alignments from [1], our alignment to the same scan, and our model fit to that scan. See [20] for the full set of results. For privacy reasons, Hasler *et al.* replaced the heads of their alignments with generic heads and we blur over the faces we recover

of approximating all body shapes. This has been done previously with SCAPE [15, 18], but only using traditional registration techniques. Additionally, previous attempts to learn a shape space via PCA focus on single scans of individuals. Since our D estimates are learned across multiple scans of a person, they may be more reliable than those learned from a single scan. Future work will also address learning person-specific Q models.

References

1. Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., Seidel, H.-P.: A statistical model of human pose and body shape. Comp. Graphics Forum. **28**:2 (2009) 337–346
2. Wuhrer, S., Shu, C., Xi, P.: Landmark-free posture invariant human shape correspondence. The Visual Computer **27** (2011) 843–852
3. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: Reconstruction and parameterization from range scans. ACM ToG. (2003) 587–594
4. Hirshberg, D.A., Loper, M., Rachlin, E., Tsoli, A., Weiss, A., Corner, B., Black, M.J.: Evaluating the automated alignment of 3D human body scans. 2nd Int. Conf. 3D Body Scanning Technologies, (2011) 76–86

Fig. 7: **Registration denoising.** Each pair shows one pose for each of four women. The left mesh is the registration from [1]; these have noise and artifacts. The right mesh is our registration to the left registration using coregistration

5. Xi, P., Lee, W.S., Shu, C.: A data-driven approach to human-body cloning using a segmented body database. Proc. 15th Pacific Conf. Comp. Graph. and Apps. (2007) 139–147
6. Li, H., Adams, B., Guibas, L.J., Pauly, M.: Robust single-view geometry and motion reconstruction. ACM SIGGRAPH Asia. **28** (2009) 175:1–175:10
7. Granger, S., Pennec, X.: Multi-scale em-icp: A fast and robust approach for surface registration. ECCV, **LNCS 2353** Springer (2002) 418–432
8. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. Comp. Graphics Forum. **27**:5 (2008) 1421–1430
9. Amberg, B.: Editing Faces in Videos. PhD thesis, University of Basel (2011)
10. Breidt, M., Bülthoff, H.H., Curio, C.: Face models from noisy 3D cameras. ACM SIGGRAPH Asia Sketches. (2010) 12:1–12:2
11. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. ACM ToG SIGGRAPH (1999) 187–194
12. Sorkine, O., Alexa, M.: As-rigid-as-possible surface modeling. Proc. 5th Euro. Symp. Geom. Proc. **26** (2007) 109–116
13. Sumner, R.W., Schmid, J., Pauly, M.: Embedded deformation for shape manipulation. ACM ToG. **26**:3 (2007) 80:1–80:7
14. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS) (2009) 296–301
15. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: SCAPE: Shape completion and animation of people. ACM ToG. **24** (2005) 408–416
16. Allen, B., Curless, B., Popović, Z., Hertzmann, A.: Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. Proc. ACM SIGGRAPH/Eurographics Symp. on Comp. Anim. (2006) 147–156
17. Sumner, R.W., Popović, J.: Deformation transfer for triangle meshes. ACM ToG. **23** (2004) 399–405
18. Balan, A.: Detailed Human Shape and Pose from Images. PhD thesis, Brown University, Providence RI (2010)
19. Geman, S., McClure, D.: Statistical methods for tomographic image reconstruction. Bul. Int. Stat. Inst. **52** (1987) 5–21
20. Hirshberg, D.A., Loper, M., Rachlin, E., Black, M.J.: Coregistration: Supplemental material. MPI-IS-TR-004, Max Planck Institute for Intelligent Systems (2012)