Landmark-free posture invariant human shape correspondence

Wuhrer, Stefanie; Shu, Chang; Xi, Pengcheng

Publisher's version / Version de l'éditeur:
https://doi.org/10.1007/s00371-011-0557-z
The Visual Computer, 27, 9, pp. 843-852, 2011-09

Vis Comput (2011) 27:843–852
DOI 10.1007/s00371-011-0557-z

ORIGINAL ARTICLE

Landmark-free posture invariant human shape correspondence

Stefanie Wuhrer · Chang Shu · Pengcheng Xi

Published online: 17 March 2011

Abstract We consider the problem of computing accurate point-to-point correspondences among a set of human bodies in varying postures using a landmark-free approach. The approach learns the locations of the anthropometric landmarks present in a database of human models in strongly varying postures and uses this knowledge to automatically predict the locations of these anthropometric landmarks on a newly available scan. The predicted landmarks are then used to compute point-to-point correspondences between a rigged template model and the newly available scan.

Keywords Shape correspondence · Template fitting

1 Introduction

We aim to compute dense point-to-point correspondences for human shapes in varying postures. The human shapes are assumed to be represented by possibly incomplete triangular meshes, which can be acquired by 3D sensing devices such as laser or structured-light body scanners. This problem arises from building a statistical model that encodes posture and shape simultaneously using a database of human scans [17]. In order to build a statistical model of 3D shapes, the raw scans have to be parameterized in such a way that likewise anatomical parts correspond across the models [13]. Considering human posture when conducting shape analysis is important since the human body shape depends on the posture of the human due to local shape changes such as muscle bulging.

While many approaches have been proposed to compute point-to-point correspondences [25], only few of them have been applied to statistical model building and shape analysis. Hasler et al. [17] build a statistical model of human shape and posture variation. They obtain the correspondence results by using manually placed markers to guide the computation of the correspondences. Unfortunately, manually placing the markers is a tedious task, and it is impractical to use routinely in large surveys where several thousands of subjects are typically scanned.

The purpose of this paper is to provide a fully automatic solution to the problem of computing point-to-point correspondences among a set of human shapes in varying postures. These correspondences can then be used to conduct shape analysis while taking into account human posture. To the best of our knowledge, previous methods for analyzing the human body shape use known landmark positions when computing the correspondences as discussed in detail in Sect. 2. We integrate an extension of the landmark prediction method [26] and the template fitting method [17]. Our approach starts by automatically computing a set of landmark positions on a human body in arbitrary posture. To compute these landmarks, we first learn the characteristics and locations of the landmarks on a human model using a database of human models in different postures. This information is used to predict the landmark positions on a new human shape in arbitrary posture. Since we aim to compute the shape correspondence of human shapes, we assume the knowledge of a human template shape T represented by a triangular mesh. Furthermore, we assume the knowledge of the skeleton and rigging weights of T. Our approach fits the

S. Wuhrer (✉) · C. Shu · P. Xi
National Research Council of Canada, Ottawa, Canada
e-mail: stefanie.wuhrer@nrc-cnrc.gc.ca

C. Shu
e-mail: chang.shu@nrc-cnrc.gc.ca

P. Xi
e-mail: pengcheng.xi@nrc-cnrc.gc.ca

template T to the new human body shape in arbitrary posture as follows. First, the predicted landmarks are used to fit an initial skeleton to the new body shape. Second, the skeleton and rigging weights are used to adjust the posture of T to the posture of the new body shape. Third, the shape of T is changed to fit to the new body shape. A detailed overview of the approach is given in Sect. 3.

2 Related work

Correspondence of deformed shapes Computing dense point-to-point correspondences between two possibly deformed surfaces has received considerable attention in recent years [25]. Although many algorithms have been developed to solve the correspondence problem, only few of these algorithms are suitable for statistical model building and shape analysis.

Approaches that solve the correspondence problem by aligning two shapes using a transformation that is approximately rigid [1, 16, 24] are not suitable to align a set of human shapes in varying posture due to nonrigid shape and posture changes. Hence, we focus our attention on approaches that take nonrigid transformations into account.

Several authors suggested landmark-based methods to obtain the correspondence and applied them to shape analysis. Blanz and Vetter [9] use a set of landmarks to compute the correspondence between pairs of human faces. Allen et al. [2, 3] use a set of landmarks and a template model to deform a template model to human shapes in similar and varying postures. Anguelov et al. [4] present an approach that also works for varying postures. The approach computes the correspondences between a large database of humans and uses the result to build an animated surface model of a moving person. The database contains one subject in multiple postures and the remaining subjects in the standing posture of the CAESAR database. Hasler et al. [17] improve the approach by using fewer landmark positions and by using a database containing many subjects in multiple postures. Pauly et al. [22] compute the correspondences and the transformation between multiple views of a scan for the application of scan completion using a small set of landmark positions.

Recently, several landmark-free approaches have been proposed. Some of these methods align bending-invariant canonical forms directly to obtain dense point-to-point correspondences [10, 19]. These methods are currently not accurate enough to perform statistical analysis. When canonical forms are aligned directly, there is no guarantee that close-by points in one shape match close-by points in the other shape. Huang et al. [18] proceed by iteratively alternating between a correspondence optimization and a deformation optimization. The approach can be viewed as an extension of the Iterative Closest Point algorithm (ICP) [8] that is often used to solve the rigid correspondence problem. The method is shown to perform well if the two meshes are initially well aligned. If the initial alignment is poor, however, the method fails. Huang et al. show that the obtained correspondences yield visually pleasing shape interpolations. The main drawback of this method is that it relies heavily on nonintuitive user-defined parameters, which makes the method hard to use. Zhang et al. [27] propose a technique that solves the correspondence problem by finding a small set of features and by choosing the best feature correspondence as the one that minimizes a deformation energy. To improve the efficiency of the algorithm, the tree of all matching features is pruned if the features are too dissimilar. Nonetheless, the algorithm is not as efficient as the algorithm of Huang et al. [18]. Once the feature correspondences are computed, the full correspondence is found by deforming the full mesh based on the feature points. The main drawback of this method is its computational inefficiency. Results are only demonstrated for models with less than 4000 vertices. Furthermore, like the method of Huang et al., the tree pruning relies heavily on nonintuitive user-defined parameters. Chang and Zwicker [12] use a reduced deformable model to compute the correspondence and the transformation between two surfaces. Instead of operating on the surface directly, the approach needs to convert the surface into a voxel grid. This is computationally expensive. Furthermore, this step introduces the use of several input parameters. While all of these methods are landmark-free, they require a set of nonintuitive user-specified input parameters.

Methods that require neither landmark positions nor user-specified input parameters have been proposed for motion capture [11, 15]. The methods assume that the same shape was captured in several gradually changing poses and use this information to learn a deformation model. In our application, this type of input data is not available. Li et al. [20] propose an approach to register pairs of range images without using any landmark positions or input parameters. While the method is shown to perform well, the method makes use of the fact that each surface is a terrain and can be parameterized by projecting each point to a plane. Since our aim is to register the surface of full human bodies, this method cannot be applied.

To summarize, existing fully automatic methods are not accurate enough to produce results that can be used for shape analysis. Landmark-based methods yield accurate results but require manually placing the landmarks. In this paper, we aim to automatically predict the landmark positions and use them to find a correspondence of high quality.

Automatic prediction of landmark positions Ben Azouz et al. [6] propose to find reliable correspondences by automatically predicting marker positions and by using these marker positions to find correspondences. Their method is

based on statistical learning. This method works for models in similar postures but fails if the posture variation is large. Wuhrer et al. [26] extend this method to work for humans in varying postures by using statistical learning in a bending-invariant embedding space.

3 Overview

We aim to compute the shape correspondence of human models. Hence, we can use a template T of a human represented by a triangular mesh. We manually find the set of 14 landmarks shown in the left of Fig. 1 on T. Let $v_i^{(T)}, i = 0, \ldots, n$, denote the vertices of T and let $l_i^{(T)}, i = 0, \ldots, 13$, denote the landmarks of T. Denote the homogeneous coordinates of $v_i^{(T)}$ by $\tilde{v}_i^{(T)}$. Furthermore, we compute a skeleton $S^{(T)}$ consisting of 17 bones and skinning weights $W^{(T)}$ for T using the approach by Baran and Popovic [7]. This allows us to deform the template into an arbitrary posture by deforming each vertex as $v_i^{*(T)} = \sum_{j=0}^{16} W_{i,j}^{(T)} \mathbf{T}_j \tilde{v}_i^{(T)}$, where $W_{i,j}^{(T)}$ is the weight for the jth bone and the ith vertex of T, and where $\mathbf{T}_j$ is the 3×4 transformation matrix applied to the jth bone. Figure 1 shows the template model. The left of the figure shows the template with landmarks, and the middle of the figure shows the template and the fitted skeleton. The right of the figure shows the rigging weights by assigning a color to every bone of the skeleton and by coloring each vertex of the template with the color of the bone that has the largest influence on this vertex.

Given a scan P of a human in arbitrary posture, we aim to compute the shape correspondence of this scan and the template T. To achieve this goal, we first predict the locations of the 14 landmarks on P using probabilistic inference. We use a database of humans in varying postures to train a Markov network and use this network to predict the landmark locations. This is based on Wuhrer et al.'s approach [26] and is explained in Sect. 4. The approach proceeds by embedding the intrinsic geometry of the human body shape in $\mathbb{R}^3$ and by predicting the landmarks in this embedding space. The embedding is invariant with respect to rotations, translations, and reflections. After centering the embedding and after aligning it along its principal axes, the embedding is invariant with respect to flipping the axes. Since there are three possible axes to be flipped, this results in eight possible alignments. Hence, the approach by Wuhrer et al. computes eight possible solutions. We denote the landmark positions by $l_i^{o\ (P)}$ in the following, where $i = 0, \ldots, 13$ is the index of the landmark, and $o = 0, \ldots, 8$ is the index of the option.

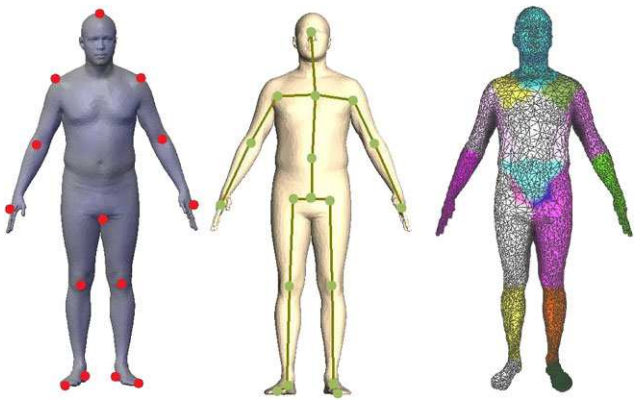


Fig. 1 Template model T with landmarks $l_i^{(T)}$, skeleton $S^{(T)}$, and rigging weights $W_{i,j}$

For each of the eight possible options, we solve a shape deformation problem as follows. We use the locations of the predicted landmarks $l_i^{o\ (P)}$ to find an initial skeleton $S^{(P)}$ with the same structure as $S^{(T)}$. We use each learned landmark as a vertex of the skeleton. We find the remaining vertices of the skeleton using linear combinations of the learned landmark positions. This skeleton is then refined to fit the posture of P using an optimization method similar to the one used by Hasler et al. [17]. Section 5 outlines this step. This step aligns the posture of T to the posture of P. Afterwards, we refine the shape of T to match the shape of P using an optimization method similar to the one used by Allen et al. [2]. Section 6 outlines this step. Since we perform these steps for each of the eight possible options, we obtain eight possible solutions. Finally, we find the correct result as the one of the eight that minimizes the symmetric shape distance to P. We compute the *shape distance* between two models P_0 and P_1 as

$$d(P_0, P_1) = \frac{1}{2|P_0|} \sum_{v_i^{(P_0)} \in P_0} \left\| v_i^{(P_0)} - NN^{P_1}\left(v_i^{(P_0)}\right) \right\| + \frac{1}{2|P_1|} \sum_{v_i^{(P_1)} \in P_1} \left\| v_i^{(P_1)} - NN^{P_0}\left(v_i^{(P_1)}\right) \right\|,$$

where $|P_i|$ is the number of vertices in model P_i, $NN^{P_j}(v_i^{(P_k)})$ is the nearest neighbor of $v_i^{(P_k)}$ in P_j, and $\|\cdot\|$ denotes the Euclidean length.

Figure 2 gives an overview of the approach.

4 Landmark prediction

We assume the knowledge of a database of scans of humans in varying postures represented by triangular meshes. Furthermore, for each scan, we assume the knowledge of the location of a set of landmarks $l_0, \ldots, l_{13}$. The landmarks are shown as red points in Fig. 3. We use this knowledge to learn relative locations and local surface properties of the landmark points using the approach by Wuhrer et al. [26].

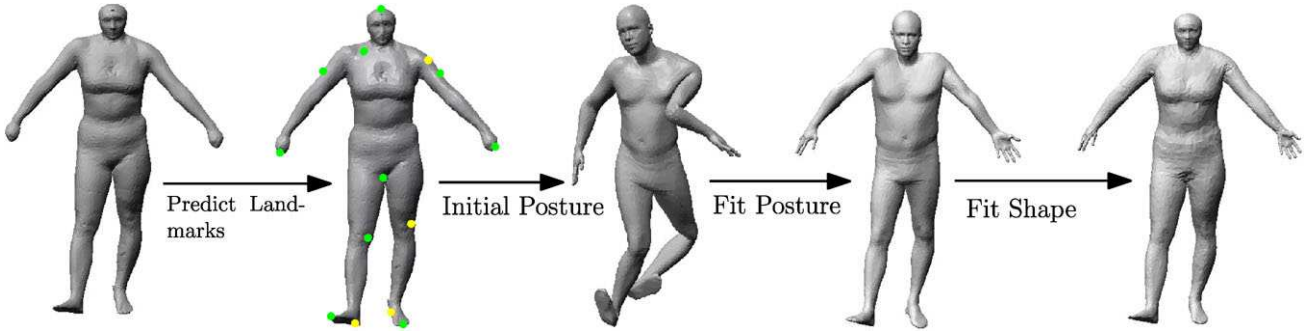


Fig. 2 Overview of the approach. For each of the eight possible landmark predictions, the approach fits the posture and the shape. Finally, the approach selects the best option based on the shape distance. Predicted landmarks on the front are shown in *green*, and predicted landmarks on the back are shown in *yellow*

Fig. 3 Location of the landmarks and structure of landmark graph

The approach is based on statistical learning and models the structure of the landmarks as a Markov network. The network structure we use is shown in Fig. 3. Each red landmark point represents a node of the Markov network. Each black edge connecting two landmark points represents an edge of the Markov network.

It is difficult to spatially align models of human subjects in different postures due to the large posture variation. Hence, the approach computes the canonical form [14] of each of the models in the database. The canonical forms of all the models have a similar posture and can be spatially aligned using the known landmark positions. This allows one to learn the locations and relative positions of the landmarks in the space of canonical forms. The approach uses this information to restrict the search space of the method and to compute the edge potentials of the Markov network.

Furthermore, the approach learns a surface property for each landmark based on the area of a geodesic neighborhood of the landmark. This information is learned on the original surface and not in canonical form space. Note that the area of a geodesic neighborhood of a landmark is isometry-invariant. The learned information is used as node potential in the Markov network. Since all of the information contributing to the Markov network is isometry-invariant, this approach allows the prediction of landmarks in arbitrary postures.

When a new scan P becomes available, the approach predicts the 14 landmark positions by performing probabilistic inference on the learned Markov network. The search space of the method is restricted using the learned average locations of the landmarks in canonical form space as follows. The canonical form of P is computed and spatially aligned with the training data. Note that since the canonical form is invariant with respect to flipping, there are eight possible alignments. For each possible alignment, only vertices in the neighborhood of the learned average location of a landmark are considered as candidates for this landmark.

Since the canonical form of a shape is invariant with respect to flipping, this approach produces eight results $l_i^{o\ (P)}$. In the original approach, it is up to the user to pick the index o that yields the best result. In this work, we use each of the eight results to compute a shape deformation of the template, and we report the result that minimizes the shape distance between the deformed template and the scan.

We execute the steps outlined in the following sections for each option o. As we consider o to be fixed during these steps, we denote the landmarks by $l_i^{(P)}$ in the following.

5 Posture fitting

This section describes how to change the posture of the template model to fit the posture of the scan P. Posture fitting starts with the initial skeleton $S^{(P)}$ of P computed based on the predicted landmarks and aims to optimize the location of $S^{(P)}$ to optimally fit the posture of the model in scan P.

The skeleton $S^{(T)}$ has a tree structure. Hence, by picking one arbitrary but fixed bone as the root, we can order the bones using a depth first order. We model the deformation of the skeleton $S^{(T)}$ as follows. We express the transformation of the root using a rigid transformation consisting of a quaternion rotation, a scale factor, and a translation vector.

The relative transformation of every other bone with respect to its parent is expressed as a quaternion rotation. Hence, the deformation is defined using $8 + 4 \cdot 16 = 72$ parameters b_i. This deformation restricts the deformation of the skeleton to deform using a single uniform scale factor and a single translation vector. Furthermore, each bone can only rotate with respect to its parent. If we know the parameters b_i, it is straightforward to compute the global transformations $\mathbf{T}_i$ of each bone of $S^{(T)}$ using composite transformations.

Given the initial skeleton $S^{(P)}$ and $S^{(T)}$, we compute the parameters b_i that deform $S^{(T)}$ close to $S^{(P)}$ as follows. The global scale factor is computed as the average scale factor between $S^{(T)}$ and $S^{(P)}$ of the bones of the torso. The global translation and rotation are computed to align the bone of the torso. Every other rotation is computed based on the relative positions of adjacent bones.

In a first step, we optimize the location of $S^{(P)}$ using the predicted landmark positions by minimizing the energy

$$E_{\text{lnd}} = \sum_{i=0}^{13} \left(\left(\sum_{j=0}^{16} W_{i,j}^{(T)} \mathbf{T}_j \tilde{l}_i^{(T)} \right) - l_i^{(P)} \right)^2$$

with respect to the parameters b_i, where $W_{i,j}$ is the weight for the jth bone and the ith landmark of T, and $\tilde{l}_i^{(T)}$ contains the homogeneous coordinates of $l_i^{(T)}$. During this optimization we restrict the scaling so that the height of the person is between 1.40 m and 2.10 m. Furthermore, we restrict the angle of the rotation of the head so that the head cannot face backwards. Note that the transformations $\mathbf{T}_i$ depend on the parameters b_i.

In a second step, we optimize the location of $S^{(P)}$ using all vertex positions by minimizing the energy

$$E_{\text{nn}} = \sum_{i=0}^{n} \left(\left(\sum_{j=0}^{16} W_{i,j}^{(T)} \mathbf{T}_j \tilde{v}_i^{(T)} \right) - NN^{(P)} \left(\sum_{j=0}^{16} W_{i,j}^{(T)} \mathbf{T}_j \tilde{v}_i^{(T)} \right) \right)^2$$

with respect to the parameters b_i, where $W_{i,j}(T)$ is the weight for the jth bone and the ith vertex of T, and where $NN^{(P)}(\sum_{j=0}^{16} W_{i,j}^{(T)} \mathbf{T}_j \tilde{v}_i^{(T)})$ is the nearest neighbor of the transformed vertex $\sum_{j=0}^{16} W_{i,j}^{(T)} \mathbf{T}_j \tilde{v}_i^{(T)}$ in P. Note that we only consider the term corresponding to $v_i^{(T)}$ if the angle between the outer normal vectors of the transformed vertex on the template and its nearest neighbor in the scan is at most 110 degrees. We use k-d trees [5] to speed up the nearest neighbor search and minimize E_{lnd} and E_{nn} using a quasi-Newton approach [21].

6 Shape fitting

This section describes how to change the body shape of the template model to fit the shape of the scan P. We first deform the template T to match the posture $S^{(P)}$ computed in the previous section. Denote this deformed skeleton model by T^*.

The problem that remains to be solved is to fit a template model T^* to a scan P, where T^* and P are in approximately the same posture. We solve this problem using an optimization method similar to the one by Allen et al. [2]. That is, we allow each vertex $v_i^{(T^*)}$ of T^* to deform using a 3×4 transformation matrix $\mathbf{A}_i$. The goal is to fit T^* to the scan P while preserving the overall shape of the surface. This is achieved by minimizing the energy

$$E_{\text{shape}} = \alpha \sum_{i=0}^{n} \left(\mathbf{A}_i \tilde{v}_i^{(T^*)} - NN^{(P)} \left(\mathbf{A}_i \tilde{v}_i^{(T^*)} \right) \right)^2 + \beta \sum_{(i,j) \in E^{(T^*)}} (\mathbf{A}_i - \mathbf{A}_j)^2$$

with respect to the transformations $\mathbf{A}_i$, where $NN^{(P)}(\mathbf{A}_i \tilde{v}_i^{(T^*)})$ is the nearest neighbor of the transformed vertex $\mathbf{A}_i \tilde{v}_i^{(T^*)}$ in P, $E^{(T^*)}$ is the set of edges of T^*, and α and β are weights. As before, we only consider the nearest neighbor term if the angle between the outer normal vectors of the transformed vertex and its nearest neighbor is at most 110 degrees. We minimize E_{shape} using a quasi-Newton approach [21]. We initially set $\alpha^0 = 1$ and $\beta^0 = 10$, and we relax β^t as $\beta^t = 0.5\beta^{t-1}$ whenever the energy does not change much. This relaxation scheme ensures that the details of the target mesh are fitted.

7 Results

This section evaluates the proposed method using the MPI database [17]. This database contains the surface scans of different subjects in up to 35 different postures. We manually placed the fourteen landmarks on 300 of the models. First, we evaluate the impact of the training data set on the correspondence performance. Second, we conduct an evaluation of the correspondence performance.

7.1 Influence of training data

We use a subset of the 300 models with manually placed landmarks to learn the properties and relative locations of the landmarks. We then compute the correspondences for a different subset of the 300 models with manually placed landmarks. This experiment evaluates the influence of the data set picked for training on the performance of the correspondence computation. We choose three training sets; one

that contains primarily shape variation, one that contains primarily posture variation, and one that contains both shape and posture variation. This allows us to evaluate the relative importance of the presence of shape and posture variations in the training set.

The data set used to compute the correspondences consists of 50 models of 7 subjects and 28 different postures. The first training set T_s consists of 50 models of 43 subjects in similar postures. Hence, T_s covers the shape variability well, while the posture variability is not well represented. The second training set T_p consists of 50 models of seven subjects in mostly different postures. The data set contains at least one model in all of the 35 postures present in the MPI database. Hence, T_p covers the posture variability well, while the shape variability is not well represented. The third training set T_{p+s} consists of 200 models of 43 different subjects and 35 different postures. Hence, T_{p+s} covers both the posture variability and the shape variability well.

When using T_s for training, the algorithm produces visually pleasing results for 72% of the models. When using T_p for training, the algorithm produces visually pleasing results for 74% of the models. When using T_{p+s} for training, the algorithm produces visually pleasing results for 78% of the models. This shows the importance of both pose and shape variations in the training set. Note that pose variation appears to be slightly more important than shape variation since the local surface area close a landmark varies more for subjects in different postures than for different subjects in the same posture.

7.2 Evaluation

In the following, we use the aforementioned set of 200 models to learn the properties and relative locations of the landmarks. We compute the correspondences for the remaining 100 models of the MPI database to evaluate the accuracy of the predicted landmarks with respect to the manually picked landmarks. We first evaluate the accuracy of the predicted landmarks. In the best case, our automatic algorithm always picks the best of the eight available options. To demonstrate the accuracy in this case, we manually pick the best of the eight results for landmark prediction. The accuracy obtained in this case is summarized in Table 1. We measure the error as Euclidean distance between the predicted landmarks and the ground truth landmarks. Note that most of the average errors are under 10 cm. The predicted landmarks at the shoulders, the elbows, and the knees have the largest average errors. This is a result of the nonisometric deformations of the human body in these areas.

Note that for the remaining experiments, the landmarks are picked automatically. Hence, we can expect the predicted landmarks to be at most as accurate as the ones in Table 1.

Table 1 Error of landmark prediction computed over 100 test human scans

Landmark	Average (mm)	Standard deviation (mm)	Maximum (mm)
1 Head	60.18	6.018	140.8
2 Crotch	75.77	4.908	205.6
3 Right shoulder	122.3	5.652	286.9
4 Right elbow	76.78	5.6	147.2
5 Right hand	12.89	1.289	66.7
6 Left shoulder	127.6	2.678	273.9
7 Left elbow	79.43	3.94	155.1
8 Left hand	10.24	1.024	57.06
9 Right knee	97	3.04	179.8
10 Right heel	20.85	0.8441	102.1
11 Right toe	0.4487	0.04487	39.46
12 Left knee	98	0.4846	185.8
13 Left heel	24.24	2.424	103.8
14 Left toe	0.8894	0.08894	39.81

The algorithm produces visually pleasing results for 76 of the models. For these 76 models, we compute the shape distance between the deformed template mesh and the original model. The mean of the shape distance over all 76 models is 5.94 mm, and its standard deviation is 0.20. We consider an average error of under 6 mm as sufficient since slight movements of the person during the acquisition of the scan caused by breathing or slight posture changes can lead to an error of the same magnitude in the acquired data.

Figure 4 shows some of the results. The first column shows the model with both the predicted landmarks and the ground truth landmarks. Ground truth landmarks are shown in red, and predicted landmarks are shown in green. When the prediction and the ground truth are identical, only the green point is shown. The second column shows the result after posture fitting, and the third column shows the result after shape fitting. The algorithm finds visually pleasing results in spite of noise in the original models (see neck of second row) and inaccuracies of the predicted landmarks (see shoulder of fourth row). The difference in shape of the hands (fists versus extended hands) comes from the shape of the hands in the template model.

The quality of the correspondences is visualized in Fig. 5. We manually applied a texture to the template model (left of Fig. 5) and transferred the texture to the bodies shown in Fig. 4 using the correspondences obtained with our algorithm. We can see that the texture map is preserved for models with different body postures and body shapes. Furthermore, we manually selected a set of feature points on the template model and assigned a unique color to each feature point. The features were then transferred to the bodies shown in Fig. 4 using the correspondences obtained with our

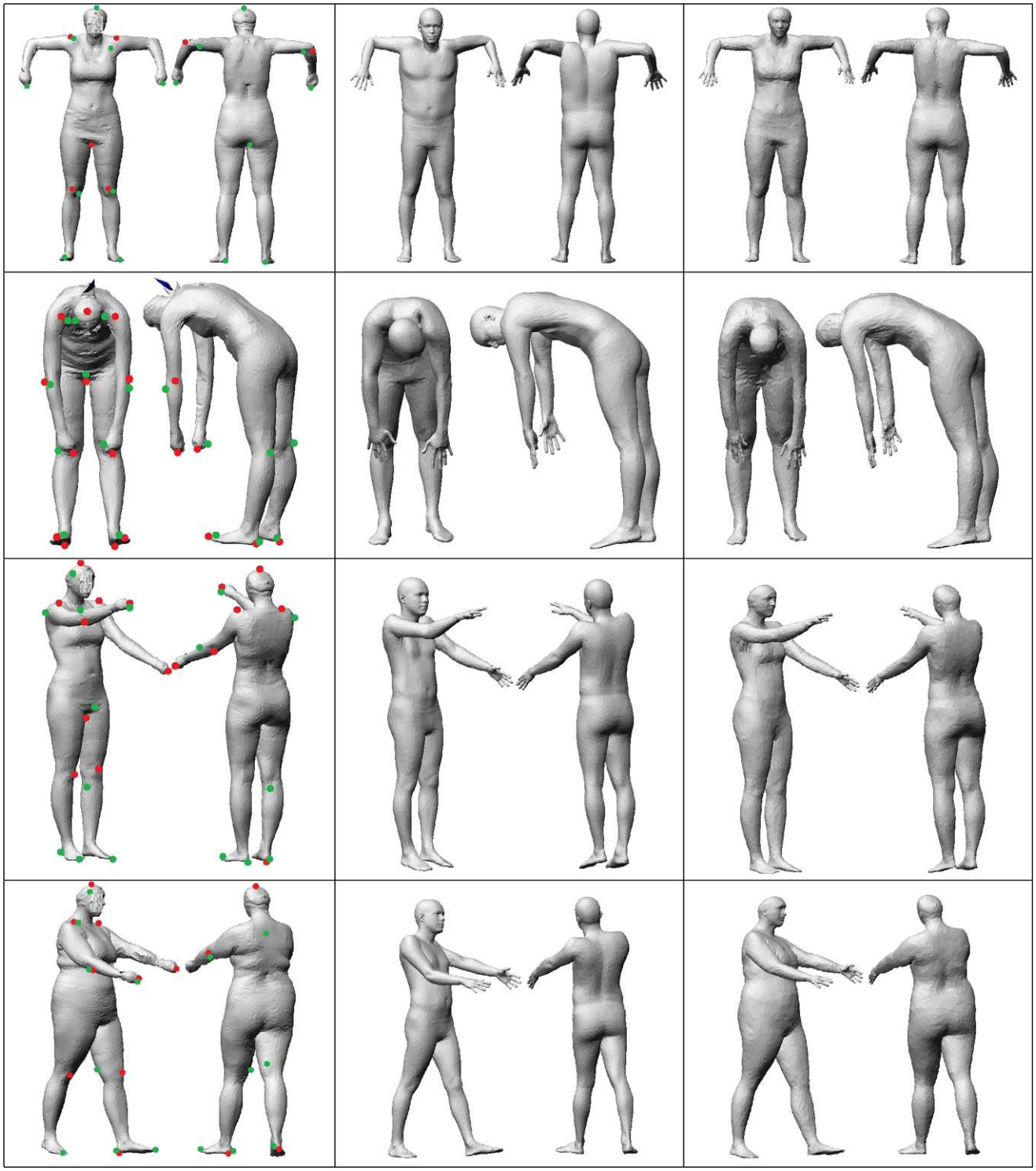


Fig. 4 The *first column* shows the model with both the predicted landmarks (in *green*) and the ground truth landmarks (in *red*). The *second column* shows the result after posture fitting, and the *third column* shows the result after shape fitting

algorithm. Note that the locations of the features are on corresponding anatomical parts of the bodies.

Next, we analyze the cases for which our algorithm fails to find visually pleasing correspondences. In 17 of the cases, the algorithm picked the wrong landmark option. In this case, body parts are erroneously corresponded to symmetric body parts, which results in a globally erroneous registration. An example of this case is shown in the first row of

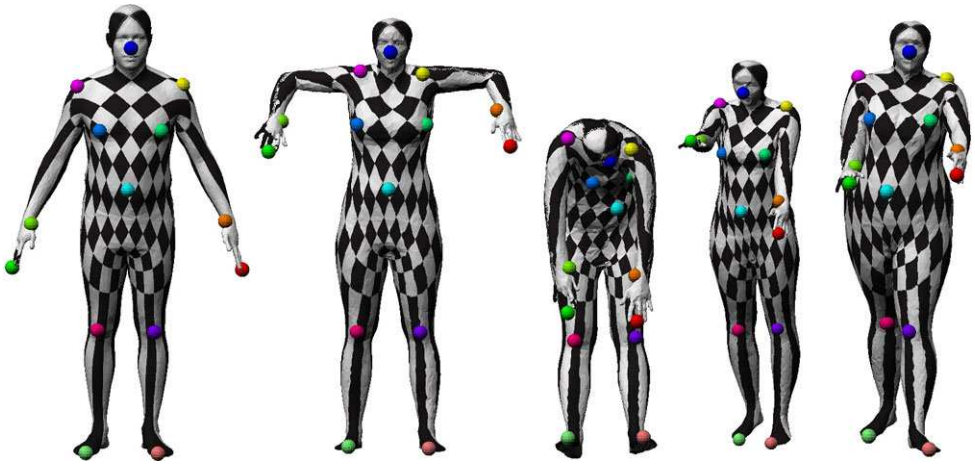


Fig. 5 Texture mapping of the corresponded models. The template mesh with feature points and a texture map is shown on the *left*. The feature points and the texture map are transferred to the four models shown on the *right* using the computed correspondences

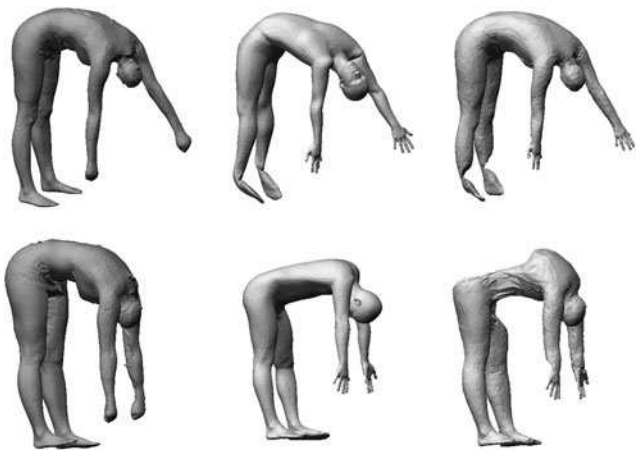


Fig. 6 Two cases where the proposed algorithm fails to find a visually pleasing correspondence. From *left* to *right*: original model, result after posture fitting, and result after shape fitting

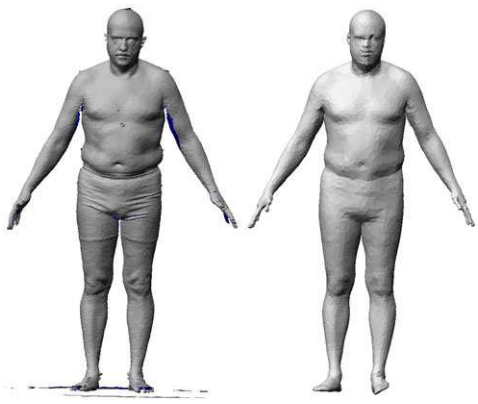


Fig. 7 *Left*: incomplete model of the CAESAR database with holes shown in *blue*. *Right*: result of our algorithm

Fig. 6. Here, the front of the body is matched to the back of the template mesh. In the remaining seven cases, the algorithm picked the correct landmark option, but the landmarks are inaccurate. In this case, the result after fitting the posture using linear blending is too far from the true surface for the nearest neighbor energy to work. An example of this case is shown in the second row of Fig. 6.

Finally, we demonstrate the performance of the algorithm when the input model is incomplete. Note that since our approach is based on canonical forms, it is not suitable for shapes with missing body parts or large holes. However, the approach is robust with respect to relatively small holes that do not alter the global shape of the canonical form. We use a model of the Civilian American and European Anthropometric Resource (CAESAR) database [23] that was acquired using a laser-range scanner and that contains a number of small holes. Figure 7 shows that a globally satisfactory result is obtained for this incomplete model. The local artifact on the left foot of the model is due to missing data at the back of the foot, which results in an erroneously estimated location of the landmark point at the heel.

8 Conclusions

We proposed an automatic method to compute accurate point-to-point correspondences between a set of human models in varying postures. We showed that in most cases, accurate correspondences were found. This method eliminates the tedious task of manually placing markers on the models to guide the correspondence computation.

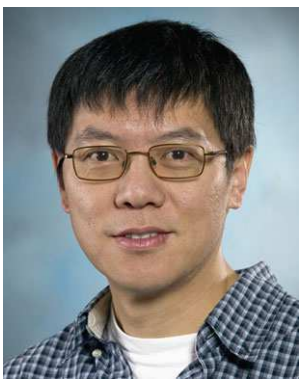
Since the presented approach is based on numerical solutions of optimization problems, there is no guarantee that a satisfactory result is obtained. In our experiments, we obtained satisfactory results for 76% of all cases. We leave it for future work to find a formulation of solving the posture

invariant correspondence problem for human shapes that is guaranteed to converge.

References

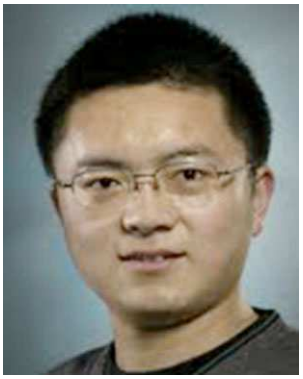
1. Aiger, D., Mitra, N., Cohen-Or, D.: 4-points congruent sets for robust surface registration. ACM Trans. Graph. **27**(3), 1–10 (2008). Proceedings of SIGGRAPH
2. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. ACM Trans. Graph. **22**(3), 587–594 (2003). Proceedings of SIGGRAPH
3. 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: Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 147–156 (2006)
4. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. ACM Trans. Graph. **24**(3), 408–416 (2005). Proceedings of SIGGRAPH
5. Arya, S., Mount, D.M.: Approximate nearest neighbor queries in fixed dimensions. In: Symposium on Discrete Algorithms, pp. 271–280 (1993)
6. Azouz, Z.B., Shu, C., Mantel, A.: Automatic locating of anthropometric landmarks on 3D human models. In: 3D Data Processing, Visualization and Transmission (2006)
7. Baran, I., Popović, J.: Automatic rigging and animation of 3D characters. ACM Trans. Graph. **26**(3) (2007). Proceedings of SIGGRAPH
8. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
9. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of SIGGRAPH, pp. 187–194 (1999)
10. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. PNAS **103**(5), 1168–1172 (2006)
11. Chang, W., Zwicker, M.: Automatic registration for articulated shapes. Comput. Graph. Forum (Special Issue of SGP 2008) **27**(5), 1459–1468 (2008)
12. Chang, W., Zwicker, M.: Range scan registration using reduced deformable models. Comput. Graph. Forum (Special Issue of Eurographics 2009) **28**(2) (2009)
13. Dryden, I., Mardia, K.: Statistical Shape Analysis. Wiley, New York (2002)
14. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10), 1285–1295 (2003)
15. Gall, J., Stoll, C., de Aguiar, E., Theobalt, C., Rosenhahn, B., Seidel, H.-P.: Motion capture using joint skeleton tracking and surface estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
16. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: Symposium on Geometry Processing, p. 197 (2005)
17. Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., Seidel, H.-P.: A statistical model of human pose and body shape. Comput. Graph. Forum (Special Issue of Eurographics 2008) **2**(28) (2009)
18. Huang, Q., Adams, B., Wicke, M., Guibas, L.J.: Non-rigid registration under isometric deformations. Comput. Graph. Forum (Special Issue of SGP 2008) **27**(5) (2008)
19. Jain, V., Zhang, H., van Kaick, O.: Non-rigid spectral correspondence of triangle meshes. IJSM **13**(1), 101–124 (2007)
20. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. Comput. Graph. Forum **27**(5) (2008)
21. Liu, D.C., Nocedal, J.: On the limited memory method for large scale optimization. Math. Program. B **45**, 503–528 (1989)
22. Pauly, M., Mitra, N., Giesen, J., Gross, M., Guibas, L.: Example-based 3d scan completion. In: Symposium on Geometry Processing, p. 23 (2005)
23. Robinette, K., Daanen, H., Paquet, E.: The CAESAR project: A 3-D surface anthropometry survey. In: 3-D Digital Imaging and Modeling, pp. 180–186 (1999)
24. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Conference on 3D Digital Imaging and Modeling, June 2001
25. van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D.: A survey on shape correspondence. In: Eurographics State-of-the-Art Report (2010)
26. Wuhrer, S., Azouz, Z.B., Shu, C.: Semi-automatic prediction of landmarks on human models in varying poses. In: Canadian Conference on Computer and Robot Vision (2010)
27. Zhang, H., Sheffer, A., Cohen-Or, D., Zhou, Q., van Kaick, O., Tagliasacchi, A.: Deformation-driven shape correspondence. Comput. Graph. Forum (Special Issue of SGP 2008) **27**(5) (2008)

Stefanie Wuhrer received a Ph.D. in Computer Science in 2009 and an M.Sc. in 2006 from Carleton University, Canada. Currently, she is a Research Associate at the National Research Council of Canada. Her research interests include geometry processing, shape analysis, and computational geometry.



Chang Shu received the Ph.D. degree in computer science from Queen Mary College, University of London, UK, in 1992, and the B.Sc. from Harbin Institute of Technology, China, in 1985. He is currently a senior research scientist at the Institute for Information Technology, National Research Council of Canada (NRC). He is also an adjunct research professor at the School of Computer Science, Carleton University, Ottawa, Canada. From 1992 to 1996, he was a research associate in the Department of Mechanical and Aerospace Engineering at the Carleton University. At NRC, he leads the Digital Human Modeling project which focuses on developing geometric and statistical methods for understanding shape variations in human and other biological forms. The results of this research have been applied to ergonomic product design, medical and biological research, gaming and animation, and security applications. In 2009, he received the Federal Partners in Technology

Transfer Award (FPTT) for his work in geometry processing. He was a program co-chair of the IEEE International Workshop on 3D Digital Imaging and Modeling (3DIM) held in Kyoto, Japan, in 2009. He is a founding member of the World Engineering Anthropometry Resource Association (WEAR). He is a member of the IEEE.



Pengcheng Xi received the Master's Degree in Computer Science from the University of Ottawa, Canada, in 2007. After that he started working at the Institute for Information Technology, National Research Council of Canada (NRC). He conducts research and development in digital human modeling, including consistent parameterization, statistical shape analysis, and knowledge-based applications.