Volumetric Template Fitting for Human Body Reconstruction from Incomplete Data

Tsz-Ho Kwok, Kwok-Yun Yeung, and Charlie C.L. Wang

***Abstract*—In this paper, we present a method for reconstructing 3D human body from incomplete data, which are point clouds captured by inexpensive RGB-D cameras. Making use of the volumetric mesh in a template, the fitting process is robust. This method produces high quality fitting results on incomplete data, which are hard to be offered by the surface fitting based methods. The method is formulated as an optimization procedure, so that the results of volumetric fitting rely on the quality of initial shape (i.e., the shape of template). In order to find a good initial shape, we develop a template selection algorithm to choose a template in an iterative manner by using the statistical models of human bodies. Experimental results show that our method can successfully reconstruct human body with good quality to be used in design and manufacturing applications.**

***Index Terms*—Template fitting, volumetric mesh, incomplete data, human body reconstruction, RGB-D camera**

I. Introduction

THREE-dimensional models of human body are widely used in many applications of robotics, virtual reality, design and manufacturing. At present, the most standard and reliable method for reconstructing 3D human bodies is to use structured-light or laser-based scanners, which can produce accurate results of 3D scanning. However, these devices usually occupy large space and are very expensive (e.g., the Cyberware system [1] costs more than $240,000). Moreover, the scanning procedure takes a long time (e.g., >10 seconds). Another common approach for human body modeling is the image-based algorithms (e.g., [2], [3]); however, as lack of depth information, they always produce results with low accuracy.

Recently, RGB-D cameras (e.g., Microsoft Kinect at a price of ∼$100), have drawn much attention in the community of computer graphics, design and manufacturing. The cameras can provide both depth (D) and color (RGB) information. This kind of small and inexpensive device allows consumers being able to afford it out-of-laboratory. Therefore, many researchers (e.g., [4], [5], [6]) have started to use Kinect in their 3D human modeling applications. Unfortunately, these RGB-D cameras provide noisy information in low resolution and the accuracy of depth values drops tremendously when the distance between camera and subject is large (see the detail analysis given by Khoshelham [7]). In the recent work of Tong et al. [5], the cameras are placed very close to the subjects. Three Kinect sensors and one turntable platform are utilized to compensate for the narrow visible region. However, the major disadvantage of their method is that the subject has to stand statically on the platform about half a minute during the data collection.

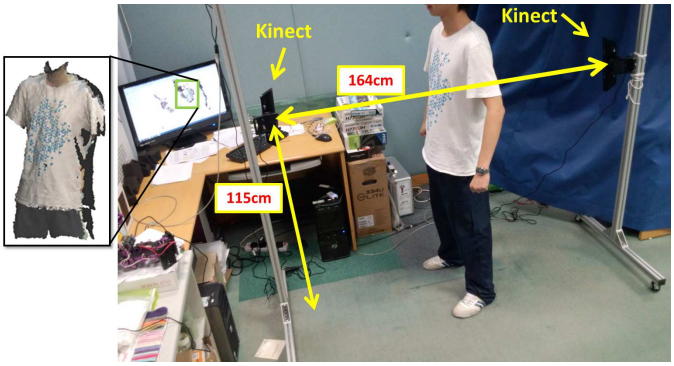


Fig. 1. Hardware setup of our scanning system: two Kinect sensors are placed to capture front and back of the major region (between neck and thigh) of a human body.

Our work is motivated by building up a human body scanner with inexpensive RGB-D cameras. This scanner is "instant" because the step of data collection in our platform is just taking a camera shot. Instant data-collection is a very important feature for scanning impatient subjects (e.g., children). When Kinect sensors are used to capture a full human body, they must be placed at least three meters away from the subject. However, on the other aspect, we need to place the cameras as close as possible to obtain depth information in high accuracy. We also try to reduce the number of sensors in the system. This is because too many sensors will increase the complexity of hardware installation (e.g., calibration) and meanwhile lead to interference as mentioned by Maimone and Fuchs [8]. In our system, two Kinect sensors are installed to obtain the human shape. The relative position and orientation of the sensors would affect the result of reconstruction. In order to capture the major information of the subject's shape, the sensors are installed in the front and at the back of the subject to harvest 3D information of the major body (i.e., from neck to thigh) - see Fig.1 for an illustration. The two sensors are calibrated by using a rectangular box which has some color marker on its planar faces (see Fig.2). Each sensor should be able to view at least 4 markers. Note that the dimension of the box and the coordinates of each marker are predefined and also known for the calibration algorithm. Based on the correspondences between marks extracted from the color images, the rigid transformation matrix of each sensor can be obtained. With the help of these transformation matrices,

This work was supported by the Hong Kong RGC/GRF Grant (CUHK/417508 and CUHK/417109) and the Direct Research Grant (CUHK/2050518).

The authors are with Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong. Corresponding Author: Charlie C. L. Wang (E-mail: cwang@mae.cuhk.edu.hk)

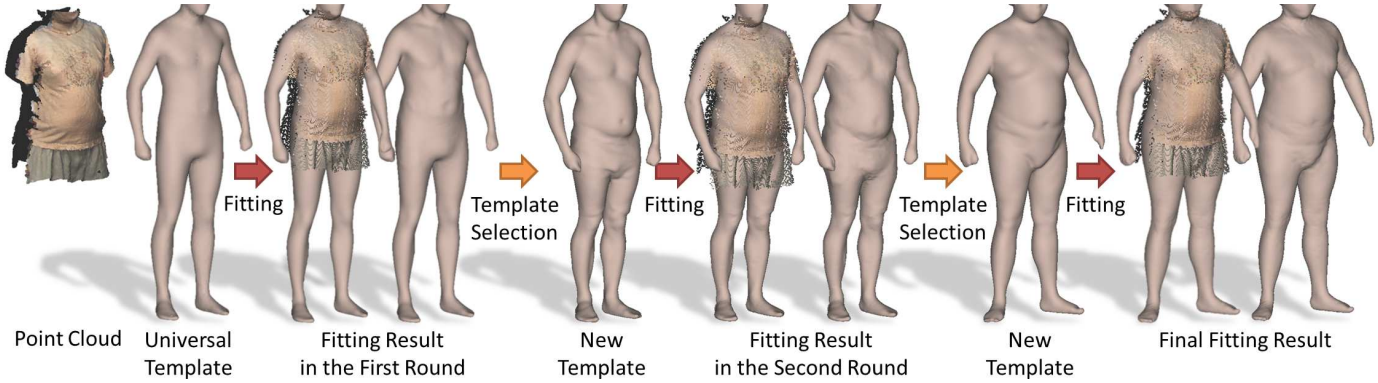


Fig. 3. Overview of our framework for human body reconstruction: By iteratively applying the interlaced selection and fitting algorithms developed in this paper, the 3D shape of a full human body can be reconstructed from the incomplete data captured by inexpensive RGB-D cameras.

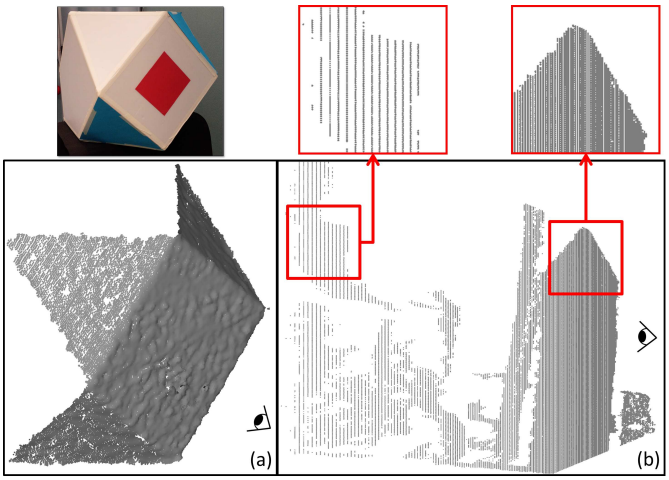


Fig. 2. The figures show the noisiness and accuracy of the depth image captured by RGB-D camera. The icon of eye shows the viewing direction of camera. The surface of the point cloud in (a) captured by the camera is very harsh, but the subject showed in photograph (upper) actually have a smooth surface. It is showed the data from RGB-D sensor is noisy. The figure (b) showing the RGB-D camera capturing an indoor envirnment. Two zoom views show the objects placed at around 120cm (upper, right) and 260cm (upper, left) away from the camera. It is clearly shown that the gap between each pixels increasing with respect to the depth value.

the RGB-D images captured by front and back sensors can be integrated together to represent the 3D point cloud of a subject. The 3D model of a subject will be reconstructed from this point cloud. Specifically, a volumetric template fitting based algorithm is developed for reconstructing human models from incomplete data. The template models are represented as volumetric meshes, which provide volumetric information to preserve the shape of human body robustly during the fitting procedure. Features of human body (e.g., the feature curves used in clothing industry) can be automatically extracted for the subject when the template fitting is completed.

For template fitting algorithms, a good template (as an initial guess) can always help to improve the quality of fitting. Taking advantage of statistical analysis in shape space, we propose an iterative selection algorithm in this paper to find a good template. A “good” template is defined as the one that are highly similar to the target shape. Overview of the proposed framework is shown in Fig.3.

A. Related work

Literature review is taken in two categories including human body reconstruction and statistical models.

Human Body Reconstruction

There are many approaches in literature that focus on the human body reconstruction from point clouds generated by structured-light-based or laser-based scanners (ref. [1], [9]). However, such scanners are too expensive to be used out-of-laboratory. Therefore, we mainly focus on the camera-based approaches below.

Tong et al. [5] used three Kinect sensors to collect the 3D data from different parts of human body, and perform pair-wise non-rigid registration and global registration iteratively to combine the data sets collected in different time instances. A full human body can be reconstructed after the registrations. Weiss et al. [4] estimated body shape by fitting image silhouettes and depth data to SCAPE models [10]. The resultant model is a best match among all candidates instead of being generated from the data itself. The optimization step in their framework takes more than one hour. Wang et al. [6] used a single fixed 3D camera to scan a full body. They propose a part-based cylindrical representation for the human model, and estimate 3D shape of a human body from four key views extracted from a depth video sequence (captured by RGB-D camera). However, the data collection time for all above methods is too long for practical usage.

When the collected data is incomplete or with clothes, we need to estimate the underlying body shape. Hasler et al. [9] employed ICP method and proposed a human body estimation approach for dressed subjects by inputting the point clouds obtained from 3D laser scanners. Guan et al. [11] estimated pose and shape of human models from single image by giving a number of manually specified correspondences between the image and the subject to be reconstructed. Hasler et al. [12] proposed a multilinear model of human pose and body shape estimation from a set of images. Bălan and Black [13] introduced a solution to reconstruct loosely dressed subjects. Each subject is photographed several times with a setup

of multiple cameras, and the subject is wearing different clothes and in different poses at each time. By combining the gathered constraints, their method is able to generate a 3D shape of the underlying subject. The optimization is improved by performing skin color detection in the images. A major limitation is that multiple (more than two) calibrated cameras are required.

Statistical Model

Statistical models have been used in prior researches (e.g., [14], [15]) to reconstruct or estimate human body models. A large database of 3D human bodies makes it possible to build a statistical model that guides the shape deformation of human bodies based on silhouettes, 3D point clouds, or depth images. Given a database of scanned 3D human bodies, the major issue for formulating a statistical model is to estimate the parameters (normally in a lower dimension) of a model that explains the variation among all scans. As a famous prior work, SCAPE [10] encoded shape variations by applying the *Principal Component Analysis* (PCA) on the 3D positions of vertices of human models in the training set. Hasler et al. [15] expressed pose and shape variations by a different encoding method that is invariant to rotation and translation. Once the statistical model is well-defined, new 3D shapes that are not in the database can be created by substituting new parameters. This is known as model synthesis. However, the main problem in model synthesis is that we can hardly determine the range of parameters so that the newly generated model is in a reasonable shape (i.e., not being a "monster"). Another common application using statistical model is individual recognition. An input data can be projected to the space spanned by the statistical model, and a set of parameters is obtained. The parameters are used to determine a match in the database by the ranged search. Related techniques can be found in face recognition [16] and image processing [17]. In this paper, we also employ the ranged search in the reduced dimensions of statistical models for template selection. The aforementioned "monster" problems in model synthesis can be avoided and a "good" template can be chosen for the final fitting.

B. Main Results

In this paper, we develop algorithms for reconstructing a full human body from incomplete data captured by two RGB-D sensors. As a result, an instant scanning system for human bodies is developed. Our technical contributions are as follows.

- A template fitting method using volumetric meshes is proposed for reconstructing 3D human body. It is found that employing volumetric mesh in template fitting is more robust than using surface mesh, especially when the input data are incomplete. A new algorithm is investigated for fitting volumetric mesh onto a sparse point cloud with the presence of noises and missing regions.
- We study the statistical model of shape-space by using *Principal Component Analysis* (PCA), and propose some practical ways to improve the quality of body shape evaluation.
- An iterative method is investigated to select a "good" template to reduce deflects in template fitting.

The paper is organized as follows. In Section II, we will first introduce the template fitting methods based on surface meshes and volumetric meshes, and then compare the performance of these techniques. After that, our volumetric template fitting algorithm is described in Section III. In Section IV, we study the statistical analysis of human body's shape-space and propose an iterative selection algorithm to find a good template from the database of human models. Experimental tests and examples obtained by this framework are presented in Section V. Finally, our paper ends with the conclusion section.

II. Surface v.s. Volumetric Fitting

This section studies the techniques of surface and volumetric fitting, and evaluates their performance on the incomplete point sets.

A. Fitting by SDM

Squared Distance Minimization (SDM) was originally proposed for curve and surface fitting (ref. [18], [19]). When fitting a surface S to a target shape Γ represented by a polygonal mesh or a point cloud, the problem is defined as minimizing the following functional.

$$E(\mathbf{X}) = F_{S\mapsto\Gamma}(\mathbf{X}) + \lambda R(\mathbf{X}) \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_i]_{i=1}^n$ are the variables to control the shape of S. In this formulation, $F_{S\mapsto\Gamma}(\mathbf{X})$ is a fitting error term as

$$F_{S\mapsto\Gamma}(\mathbf{X}) = \sum_i \|\mathbf{x}_i - \Pi(\mathbf{x}_i)\|^2 \tag{2}$$

with $\Pi(\mathbf{x}_i)$ denoting the projection of $\mathbf{x}_i$ on Γ. $R(\mathbf{X})$ is a regularization term with the regularization factor λ controlling the trade-off between smoothness and fitting criteria. One of the most commonly used form of the regularization term is Laplacian [20].

B. Regularization by Surface ARAP energy

Using a Laplacian-based regularization in SDM can produce smooth and regular result. However, it cannot preserve spatial relationship between vertices on the input model (see Fig.4 for an example). In order to maintain the model shape, the *As-Rigid-As-Possible* (ARAP) energy of the boundary surface (akin to [21], [22]) can be employed for regularization. Specifically, for each triangle $t \in S$ with three points $(\mathbf{x}_1^t, \mathbf{x}_2^t, \mathbf{x}_3^t)$ and their original positions on the template $(\hat{\mathbf{x}}_1^t, \hat{\mathbf{x}}_2^t, \hat{\mathbf{x}}_3^t)$, an accessory point is placed along the unit normals of triangles before and after deformation as $\hat{\mathbf{x}}_4^t$ and $\mathbf{x}_4^t$ respectively. The transformation matrix $\mathbf{T}_t$ for the deformation is $\mathbf{T}_t = \mathbf{P}\hat{\mathbf{P}}^{-1}$ with $\mathbf{P} = [\mathbf{x}_1^t - \mathbf{x}_4^t \;\; \mathbf{x}_2^t - \mathbf{x}_4^t \;\; \mathbf{x}_3^t - \mathbf{x}_4^t]$ and $\hat{\mathbf{P}} = [\hat{\mathbf{x}}_1^t - \hat{\mathbf{x}}_4^t \;\; \hat{\mathbf{x}}_2^t - \hat{\mathbf{x}}_4^t \;\; \hat{\mathbf{x}}_3^t - \hat{\mathbf{x}}_4^t]$. As mentioned by Liu et al. [23], $\mathbf{T}_t$ can be decomposed by the "signed version" of *singular value decomposition* (SVD) into $\mathbf{T}_t = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, and the rigid transformation can be enforced by letting $\mathbf{T}_t \mapsto \mathbf{L}_t$ with

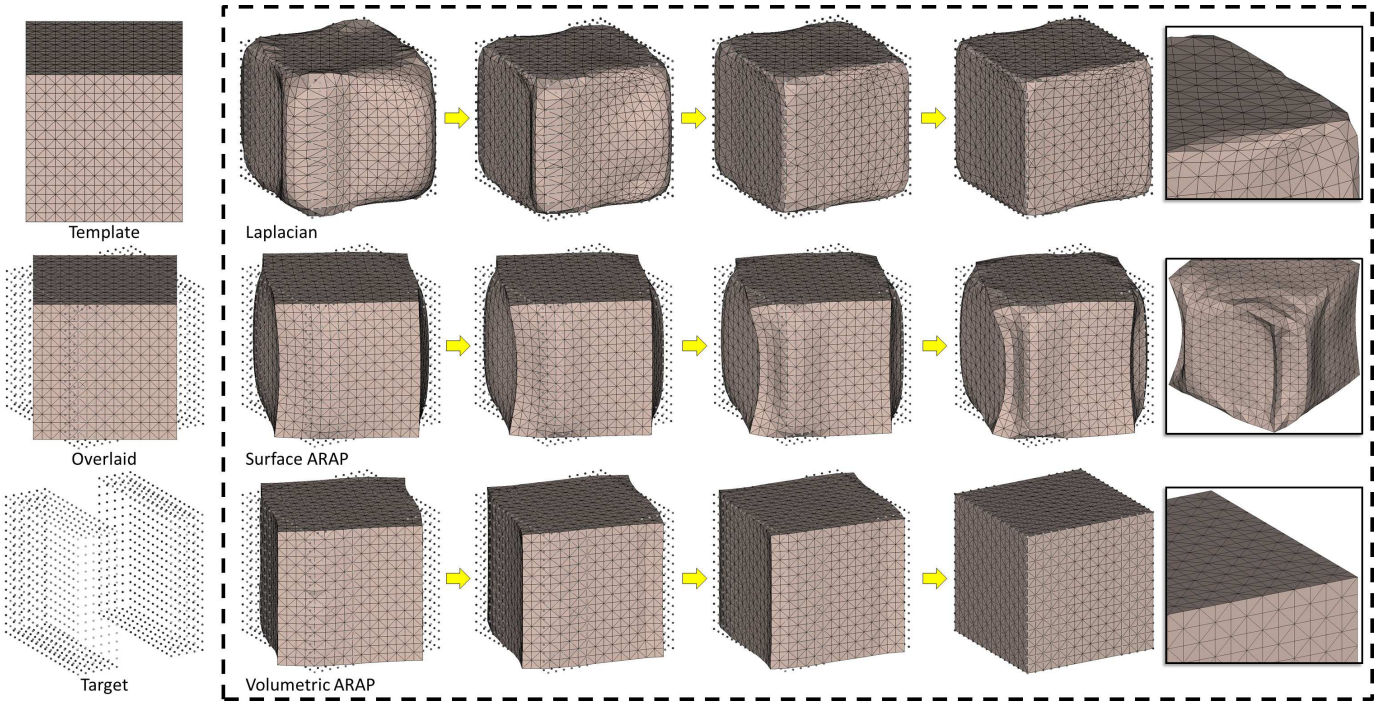


Fig. 4. Fitting tests on a template with cubic shape. Target is a point cloud sampled from a part of the cube (with middle part missing) and rotated 30 degrees around y-axis - therefore, it is mis-aligned with the template. Progressive results of fitting using different regularization terms are shown. The one using volume ARAP energy is the most robust.

$\mathbf{L}_t = \mathbf{U}\mathbf{V}^T$. By this formulation, the regularization term using surface ARAP energy is expressed as follows.

$$R_S(\mathbf{X}) = \sum_t A_t \|\mathbf{T}_t - \mathbf{L}_t\|_F^2, \tag{3}$$

where $\|\cdot\|_F$ is the Frobenius norm, and A_t is the area of triangle. Minimizing $R_S(\mathbf{X})$ is to release triangles from shearing and scaling (i.e., only rotation and translation are allowed). The energy is defined for all triangles on the surface S, so it is called surface ARAP.

C. Regularization by volumetric ARAP energy

Surface ARAP can help in preserving the rigidity of the triangles in S; however, the solution of this minimization is not unique. For example, if a piece of paper is pushed from two ends, the paper can be bended to form either a convex or a concave shape even the rigidity of each local small region (i.e., the triangles when the paper is represented by a triangular mesh) is maintained. The lack of uniqueness can lead to an abnormal fitting result. For the example shown in Fig.5, even after using the surface ARAP energy as regularization term, the fitting error is still high. Allen et al. [14] introduce a smoothness term by minimizing the differences of transformation matrices between triangles. However, the smoothness error term cannot solve the aforementioned problem of uniqueness. Here, we solve the problem by introducing volumetric information into the regularization term.

Specifically, surface ARAP energy is extended to a volumetric ARAP energy defined on a tetrahedral mesh, M, that is compatible to S on its boundary (i.e., $\partial M = S$). The volumetric ARAP energy is then defined as

$$R_V(\mathbf{X}) = \sum_{tet} \Delta_{tet} \|\mathbf{T}_{tet} - \mathbf{L}_{tet}\|_F^2. \tag{4}$$

It is applied to every tetrahedron tet (having four vertices) instead of surface triangles, and Δ_{tet} is the volume of the tetrahedron tet.

D. Comparisons

Template fitting using 1) Laplacian, 2) surface ARAP energy, and 3) volume ARAP energy as the regularization terms are evaluated in this section. As shown in Fig.4, the first test tries to fit the mesh surface of a cube (as the template) onto a point cloud sampled from the cube. To evaluate the robustness of fitting, the target point cloud is rotated 30 degrees around y-axis and removed the middle region. The test is to simulate the input of our scanner that the subject could be mis-aligned with the template model. For surface fitting with Laplacian as the regularization term, a result with low fitting error can be obtained. However, the fitting result does not respect the spatial relationship of vertices in the original cube, and the shape of template has been significantly distorted. For a fitting using surface ARAP energy, although it can preserve the rigidity of triangles, surface of the template cube is highly distorted. These fitting results have large distortion, this is because the regularizations are not robust enough to drive the template to rotate during the iterative fitting procedure. For a fitting using volumetric ARAP energy, the stiffness of interior tetrahedra strongly contributes to the regularization term. Therefore, the distortions on interior tetrahedra must be absorbed by rotating the template model, so that the fitting procedure can achieve

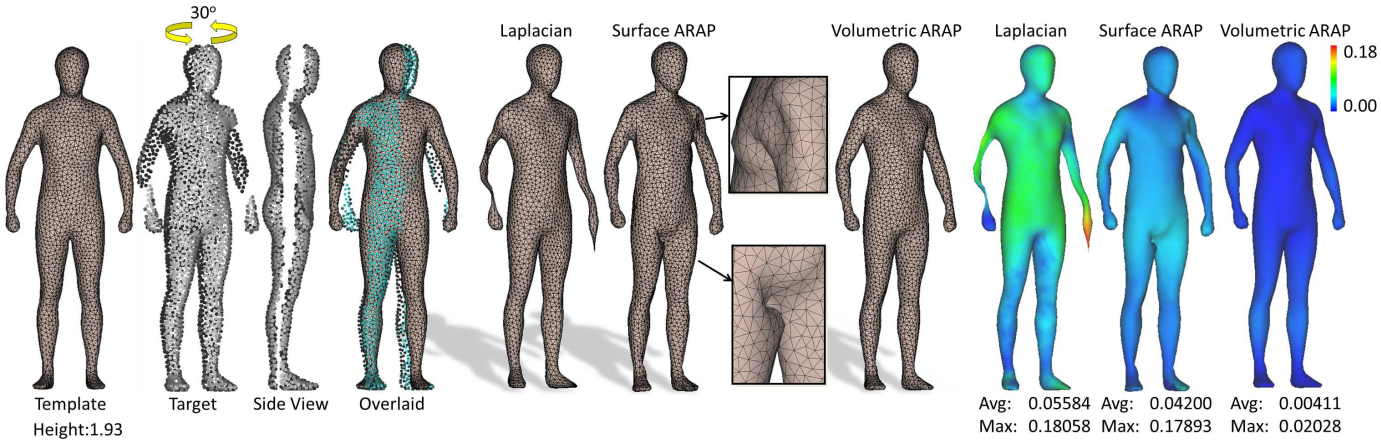


Fig. 5. Fitting tests on a template with freeform shape - human body. Target is a point cloud sampled from the mesh model by simulating the incomplete data to be obtained by our RGB-D camera based scanner. Again, the point cloud is rotated 30 degrees around y-axis to mis-align with the template. The color maps show the shape-approximation-errors on the fitting results using different regularization terms. The 'Avg' and 'Max' show the average and the maximal point-to-point distances. Fitting with volume ARAP energy as regularization term presents the best result.

a better optimum. Meanwhile, cubic shape of the template is fully preserved on the fitting result.

To evaluate the performance on freeform models, another fitting experiment is conducted on a human model (see Fig.5). Similarly, a point cloud is sampled from the model with artificially removed regions to simulate incomplete data, and the point cloud is rotated 30 degrees along y-axis. As the ground truth is the shape of the template model itself, we are able to measure the distance errors at every vertex by comparing to those in the ground truth. This error is named as *shape-approximation-error* in this paper. The shape-approximation-error is computed on the fitting results using different regularization terms. As shown in Fig.5, average errors on the fitting result with volumetric ARAP energy are less than $1/10$ of the errors on the other two methods.

III. Template Fitting

Although our template-fitting framework can be applied to general inputs with 3D points, we mainly focus on the inputs from RGB-D sensors. To build an instant body scanner by using two RGB-D sensors, each sensor will only capture a part of the body like the point cloud shown in the zoom-view of Fig.1. We assume that the input of our template-fitting framework only covers the main body of a subject. The alignment and scaling of template models can be automatically conducted by using this prior information. All points are assigned with an initial orientation that is pointing to its corresponding camera. The orientation-aware PCA is then applied to estimate accurate normal vectors, which are important to improve reliability of the fitting process.

A. Algorithm of fitting

The 3D model of a human body can be reconstructed from an incomplete point cloud by fitting a template model. This is in fact an optimization procedure to minimize the energy function defined in Eq.(1). Without loss of generality, Eq.(1) can be rewritten into the form of

$$E(\mathbf{X}) = \|\mathbf{A}_{fit}\mathbf{X} - \mathbf{b}_{fit}\|^2 + \lambda\|\mathbf{A}_{reg}\mathbf{X} - \mathbf{b}_{reg}\|^2, \qquad (5)$$

where $\mathbf{A}_{fit}$ and $\mathbf{b}_{fit}$ are derived from the fitting energy $F_{S\mapsto\Gamma}(\mathbf{X})$, and $\mathbf{A}_{reg}$ and $\mathbf{b}_{reg}$ are derived from the regularization term $R_V(\mathbf{X})$. In other words, the function is in a least-square form. As a result, the positions of vertices, $\mathbf{X}$, on a template M are updated by solving the following over-determined linear system

$$\begin{bmatrix} \mathbf{A}_{fit} \\ \sqrt{\lambda}\mathbf{A}_{reg} \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{b}_{fit} \\ \mathbf{b}_{reg} \end{bmatrix} \qquad (6)$$

during the optimization. After each update of $\mathbf{X}$, $\Pi(\mathbf{x}_i)$ in Eq.(2) and $\mathbf{L}_v$ in Eq.(4) must also be updated.

Target Position

Generally, for a vertex $\mathbf{x}_i \in \Re^3$ on the template S, the target positions $\Pi(\mathbf{x}_i)$ in Eq.(2) can be its *nearest neighbor* (NN) in the point cloud Γ. If Γ is a complete data, some approaches (e.g., [13]) have proposed to use bidirectional search to determine $\Pi(\mathbf{x}_i)$ in the fitting process. For Γ with missing data, Hasler et al. [9] proposed to use point distance and normal difference between $\mathbf{x}_i$ and its NN to filter out mismatched pairs. However, this filtering cannot always avoid matching several vertices in S onto the same target in Γ that may produce unwanted stretches. This scenario happens when some regions of the point cloud, Γ, are sparser than the distribution of vertices on the template.

As illustrated in the top row of Fig.6, when Γ is dense and complete, all vertices (i.e., $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_6$) of the template model can search their corresponding NN around them on Γ. The fitting result is satisfactory. However, when Γ is incomplete (middle row), e.g., the region around $\mathbf{x}_3$ and $\mathbf{x}_4$ is missing, the NNs found for $\mathbf{x}_3$ and $\mathbf{x}_4$ may lead to stretched result even when the *Tangent Distance* (TD) [19] is used. Note that, point distance and normal difference filters cannot eliminate the mis-matched pairs of $\mathbf{x}_3$ (or $\mathbf{x}_4$).

An inverse NN search strategy is developed to solve the target-tracking problem on an incomplete data set. Instead of searching NNs in Γ for $\mathbf{x}_i \in S$, we search NNs on S for all the points in Γ. In this case, $\mathbf{x}_3$ and $\mathbf{x}_4$ will not be associated to any points by mistake (see the bottom row of Fig.6).

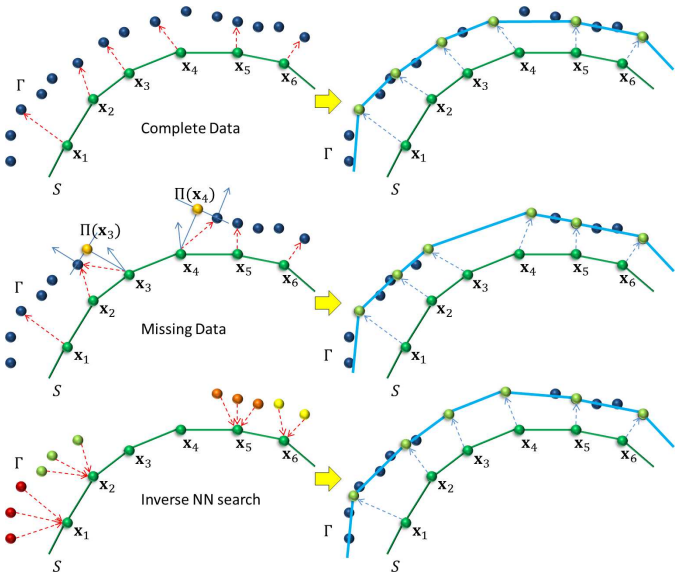


Fig. 6. Illustration of comparison between the NN search, $S \mapsto \Gamma$, and the inverse NN search, $\Gamma \mapsto S$, in incomplete data.

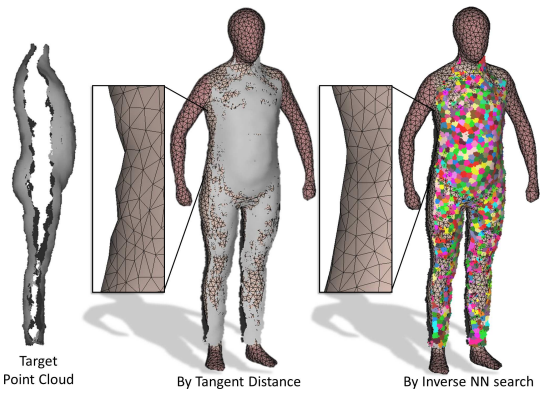


Fig. 7. When the scanned point cloud is incomplete (see left), the NN search $S \mapsto \Gamma$ gives poor fitting result (see middle) near the boundary of missing region. The inverse NN search $\Gamma \mapsto S$ results in a human body reconstruction with better surface quality.

When the point cloud is incomplete, this inverse NN search is more reasonable since the template is always complete. Moreover, point distance and normal difference can filter out outliers in the point cloud when conducting this inverse NN search. Those vertices in Γ sharing the same NN are grouped together and form a cluster. If a cluster (corresponding to $\mathbf{x}_i$) on Γ consists of more than one point, we fit a *Moving Least Square* (MLS) surface and $\Pi(\mathbf{x}_i)$ is the projection of $\mathbf{x}_i$ on the MLS surface; otherwise, TD of $\mathbf{x}_i$ is used for $\Pi(\mathbf{x}_i)$. Figure 7 shows an example to demonstrate the functionality of the inverse NN searching.

Factor of Regularization

The only parameter in this framework is the factor of regularization, λ, in Eq.(1). We allow this factor to change in the whole process. Generally, it is known that if a larger value is chosen for λ, a smoother result is obtained. On the other aspect, using a smaller λ will obtain a fitting result with lower shape-approximation-error. We adjust the value of λ in different phases of the fitting process as follows:

- **In Alignment Phase**: a large factor ($\sim$1/200) is used for the orientation and posture alignment.
- **In Fitting Phase**: a smaller factor ($\sim$1/500) is applied to generate a smooth result in the final fitting.

Moreover, a different regularization factor ($\sim$1/5000) is adopted for the template selection (details can be found in Section IV-A).

Dense Mesh for Visualization

For the sake of efficiency, a template mesh cannot be too dense. Therefore, we use volumetric meshes that have only around 3.4k vertices on the boundary surface in the template fitting. In order to give a better illustration for the details of human bodies, each coarse mesh in the database is paired up with a dense surface mesh, which has around 33k vertices. After fitting the coarse mesh to a target shape, the dense surface mesh associated with the template model is deformed onto the fitting result by using t-FFD [24].

B. Dressed subjects

Our template fitting framework can also reconstruct human models for dressed subjects by taking minor modifications. To achieve this goal, the basic change of above fitting algorithm is that the fitting result should entirely lie in the region bounded by the scanned point cloud - called *feasible region*. Meanwhile, the reconstruction must be in the shape of a human body (i.e., does not have the "monster" problem).

To prevent the fitting result running out of the feasible region, we penalize the matching pairs in $F_{S\mapsto\Gamma}(\mathbf{X})$ with a higher weighting ($\times$10) when the vertices of S go outside. Whether a vertex $\mathbf{x}_i \in S$ is outside the feasible region can be detected by computing the dot product: $((\Pi(\mathbf{x}_i) - \mathbf{x}_i) \cdot \mathbf{n}(\Pi(\mathbf{x}_i)))$; here, $\Pi(\mathbf{x}_i)$ is the projection of $\mathbf{x}_i$ on the input point cloud Γ and $\mathbf{n}(\cdots)$ gives the oriented normal vector. If a negative value is returned by this check, $\mathbf{x}_i$ is considered as being outside of the feasible region.

Another problem to be solved is how to ensure the fitting result being a shape of human body. For a similar purpose, Hasler et al. [9] applied a so-called *humanization* step by projecting the result onto the solution manifold (represented by a statistical model). Bălan and Black [13] controlled the shape coefficients within three standard deviations from the mean. Chu et al. [25] constructed a convex hull for the human body database and projected the infeasible parameters onto the surface of convex hull. However, the projection based methods have several drawbacks.

- It is computational expensive for constructing convex hull in high dimensional space.
- It is over-constrained to restrict the solution being within the convex hull or falling in three standard deviations. Counterexamples can be easily found.
- Such projection can induce large fitting error.

Our volumetric ARAP fitting framework shows a good property that can well preserve the shape of models. Therefore, we

do not need to project the fitting result by a post-processing step. However, on the down side, the volumetric ARAP energy based regularization term can be too strong that the fitting result may be quite different from the shape of target point cloud. To overcome this difficulty, we develop a template selection method in next section to automate the procedure of selecting a similar template.

IV. Template Selection

In this section, we first brief how to build a statistical model by *Principle Component Analysis* (PCA). After that, we will discuss the limitations of PCA-based selection and investigate methods for improvement. Finally, a template selection algorithm will be presented.

A. Template searching

Principle Component Analysis (PCA) has been used to establish statistical models for analyzing high-dimensional datasets including 3D human models (e.g., [14], [10], [9], [26]). The major advantage of PCA is that the relationship between exemplars with low variance can be discarded after analysis. The full dataset can be greatly reduced to represent the original exemplars in an approximation form. As a result, both the computational time and the size of storage can be reduced.

Given a point cloud Γ as the target shape to be reconstructed, we are going to search for a model that is the most similar to the target. This model will be used as the template for fitting. Suppose that the surface representation, S_Γ, of Γ has been reconstructed, and then the corresponding coefficients of this model in terms of the *principle components* (PCs) can be obtained by the bijective mapping between S_Γ and all exemplars in the database. In other words, the representation of S_Γ (the query model) by the statistical model is obtained. An efficient method is needed to take search by the PCs of a query model. Different distance metrics could be used for the search (e.g., L^1-norm or Mahalanobis distance metrics were studied by Draper et al. [27] to provide the best result in face recognition). For an efficient implementation, we employ kD-tree in the *Approximate Nearest Neighbor* (ANN) library [28] for searching, which is based on L^2-norm.

An example is shown in Fig.8. Without loss of generality, the statistical model of all human bodies in a database is represented by three PCs. Only four examples (Models A, B, C and D) are stored in the database here and their corresponding coefficients of the first three PCs are shown in Fig.8. For a query model (e.g., E, F, G or H), once its triangular mesh, which is compatible to the models in the database, is obtained (e.g., by fitting), its projection onto the PCs can be obtained. As a result, the most similar model in the database can be found by the L^2-norm based search using the coefficients of PCs (see the illustration shown at the bottom of Fig.8). By a well-defined statistical model, shape of the searching results is quite similar to the query model.

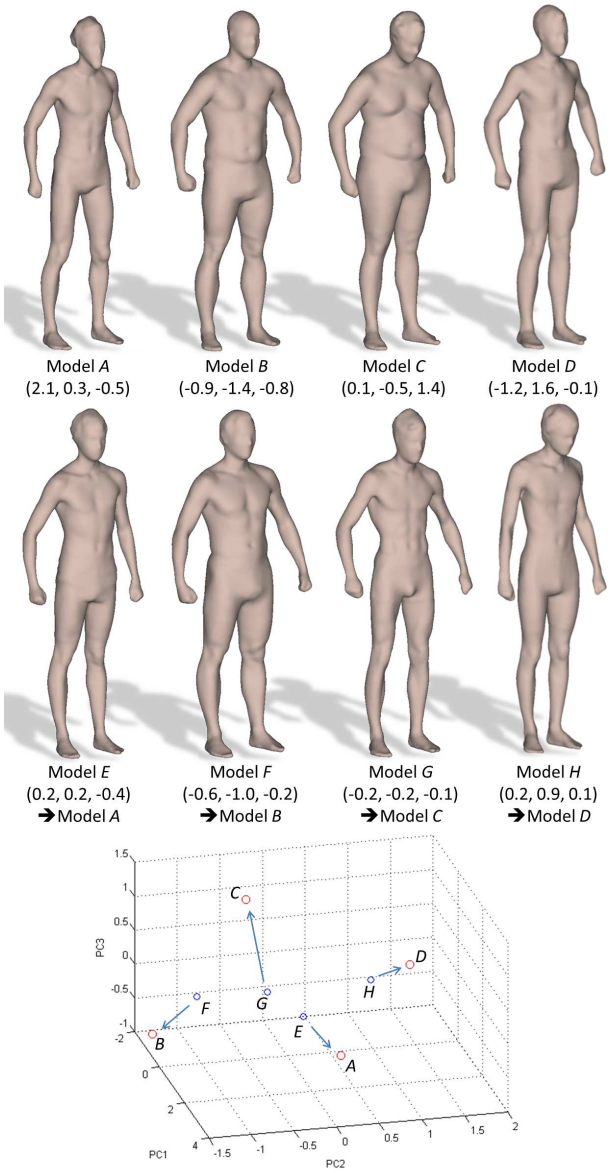


Fig. 8. Four example models (A, B, C and D) are used to form a database; their corresponding coefficients for the first three PCs are shown in the top row. Four query models (E, F, G and H) and their corresponding coefficients for the first three PCs are shown in the middle row. These coefficients are used to search the most similar models from the database by L^2-norm - see the illustration shown in the bottom row where the three PCs are used as three axes in $\Re^3$.

B. Limitations and adjustments

PCA is a powerful tool that can identify the most meaningful bases to approximate a large dataset with fewer coefficients. Approximating the given dataset by the new PCs (in lower dimensions) can reveal the hidden structure and filter out the noises. However, when positions of vertices on 3D human bodies are used for PCA, the analysis is highly sensitive to the distribution of vertices. Specifically, after scaling, translating, rotating a human body, or changing its posture, the analysis could result in very different coefficients for PCs. As demonstrated in Fig.9, when scaling model A smaller ($\times 0.8$) to get A', translating model B along x-axis to a new position B', rotating model C around y-axis 90 degrees to C', and

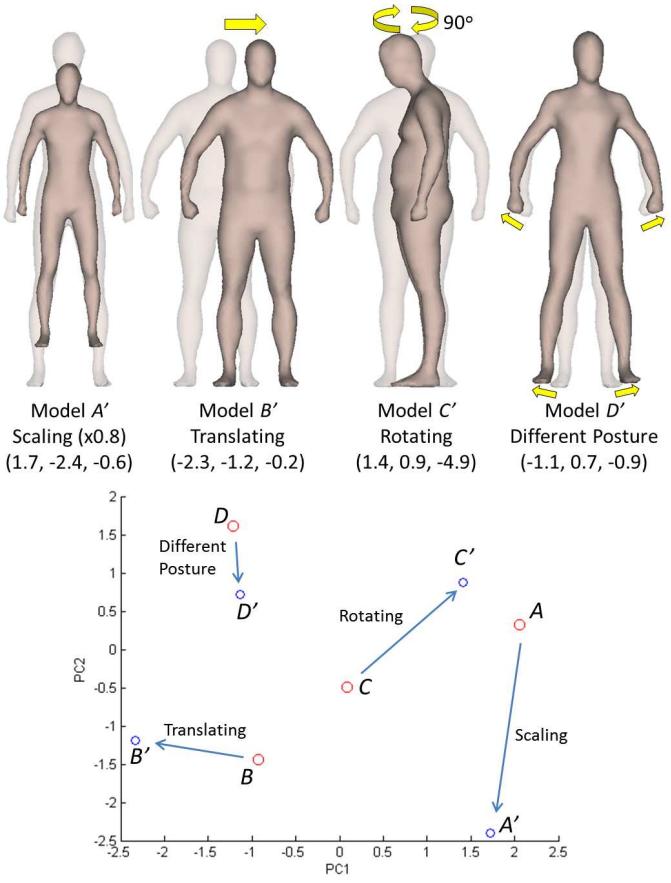


Fig. 9. Principal component analysis of human bodies using vertices is very sensitive to the change of vertex-positions. (Top) From left to right, scaling, translating, rotating and changing poses will make models (A', B', C' and D') to have very different coefficients of PCs comparing to their original shape (A, B, C and D). (Bottom) Changes according to the most significant two PCs are plotted in a 2D chart for better understanding.

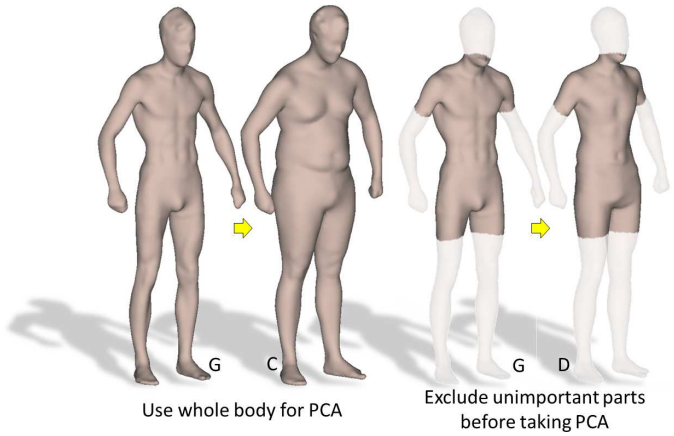


Fig. 10. When using all vertices of human models in PCA, the NN search for the query input G will return the model C as the 'most similar' model by mistake. However, the shapes of G and C are quite different from each other. After excluding unimportant parts as suggested in this paper, the NN search for G will return the model D. The shape of model D is more similar to G. The reason why G is not close to D before applying the normalization is that the poses of their lower bodies are different from each other.

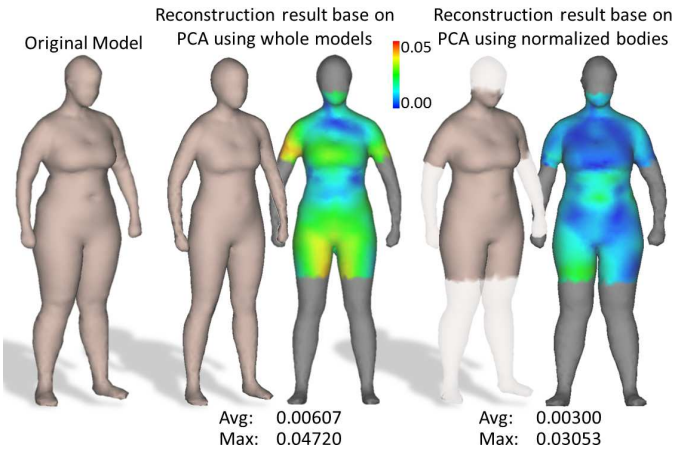


Fig. 11. The shape-approximation-error (visualized as the color map) can be reduced when the reconstruction is conducted on the PCs obtained by the normalized human bodies. Five PCs are used in the reconstruction.

deforming model D to a different posture D', their coefficients are changed significantly.

To solve these problems in PCA-based statistical models, human models stored in a database are recommended to take the following changes:

1) Be scaled to the same height (e.g., the average height);
2) Be adjusted to have the same orientation (e.g., facing front);
3) Be shifted to the same center (e.g., the middle point between the belly-button and its corresponding point at the back);
4) Be posed in a similar posture.

This is called *normalization*. The processes 1-3 can be automatically conducted. However, it is tedious to adjust the posture of every human model in the database. Instead of adjusting the human models' posture one by one, Hasler et al. [9] encode the triangles' relative rotation matrices for doing PCA, but their method creates a very high dimensional space that is expensive in both storage and processing time. Here, a simple and practical solution is used to solve this problem. Our study finds that the shapes of some unimportant parts (e.g., forearms, legs, and head) have significant effects on PCA. To avoid biasing the analysis, unimportant parts are excluded from the PCA and in the template search. This modification can be easily implemented because that the bijective mapping has been defined among all models in database. Figure 10 shows an example to demonstrate how these steps of normalization can improve the quality of PC-based template search. Again, the search is based on L^2-norm in the space of PCs.

To further verify the functionality of the normalization, we compare the results of PCA-based synthesis before and after applying the normalization in Fig.11. In the synthesis of human models, five most significant PCs are kept from 30 PCs. PCA without taking the normalization is biased by the poses of exemplar. As a result, the human model synthesized by the set of parameters that is same as the original model presents large shape-approximate-error. After taking the normalization, the remained PCs mainly focus on the variation of body shapes; therefore, the body shape with small shape-approximation-error can be reconstructed even when five PCs are used only.

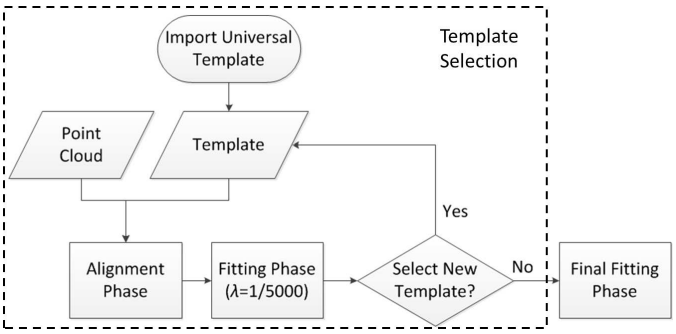


Fig. 12. Flow chart of template selection.

C. Iteration procedure

To search the best template in database by a query model (as explained in Section IV-A), one needs to find the surface bijective mapping between the query model and human models stored in the database. However, for a query input obtained by RGB-D camera, this mapping is not available. We propose an algorithm to solve this problem by running the fitting step and the template selection step iteratively. Flow chat of the algorithm is shown in Fig.12. We first use a universal template, which is the average model of all models in the database, to fit the target shape Γ (by a small regularization factor $\lambda = 1/5000$). Then, the fitting result is used to select a new template, and the new template is applied to fit the target shape again. Here, the fitting result owns the bijective mapping to models in the database so that it can be used to search a new template by the method presented in Section IV-A. During these iterations, templates that are more and more similar to the input Γ will be selected and the fitting results will be progressively improved. Illustration of the whole process has been shown in Fig.3.

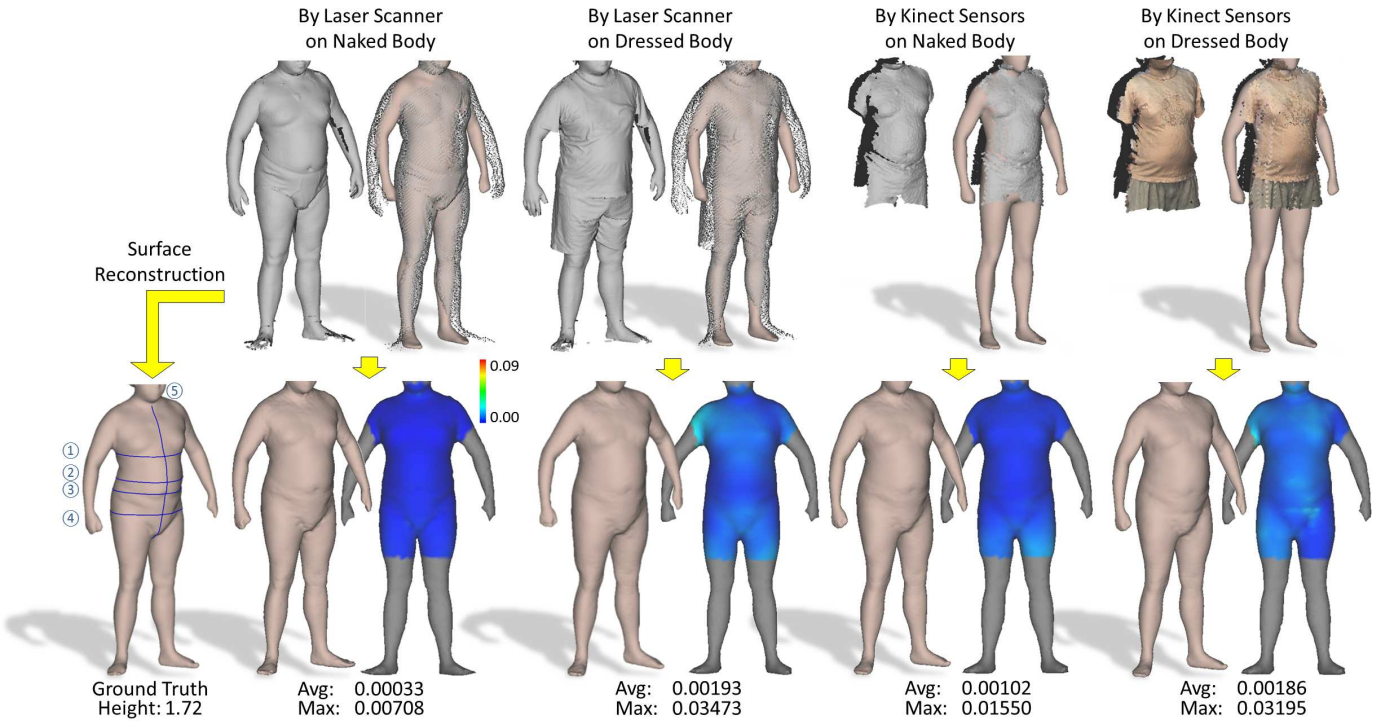
V. Results

We have implemented the proposed framework by Visual C++ and run the experimental tests on a PC with Intel Core(TM) i7-3770 CPU @ 3.40GHz, 8GB RAM and Win7 64bit OS. Our database of human models has 43 male subjects and 36 female subjects, which come from a European database [15]. The point clouds of all subjects are converted into a mesh surface by the Poisson reconstruction [29]. The surfaces are remeshed to have the same connectivity by cross-parameterization[30], [31]. each model is represented as a tetrahedral mesh with 7k vertices and 31k tetrahedra. Among these 7k vertices, 3.4k are on the boundary surface and about half of them are used in the fitting term (as the input is a set of incomplete data). Several experimental tests have been taken on this platform. With the point clouds (with about 400k points) generated by Kinect sensors, the whole procedure of human reconstruction can be completed in around 30 seconds. A video to illustrate and demonstrate our approach can be accessed at: http://youtu.be/ZFeunA_B41I.

The first example is shown in Fig.3. An incomplete data, Γ, is captured by our setup of body scanner. The reconstruction starts from the pose alignment (fitting with large regularization factor $\lambda = 1/200$) by using the universal template. When the displacements of all vertices are less than 10^{-3}, the template selection starts and uses a very small regularization factor $\lambda = 1/5000$ to fit the volumetric mesh onto Γ in around 10 fitting steps. After that, a new template is selected from the database by the most significant PCs. In our tests, 35 PCs are used for male subjects while 30 PCs are used for female. Then, this new template is fit onto Γ again with $\lambda = 1/5000$. The interlaced template selection and fitting steps are repeated until the newly selected template is same as the previous one. Usually, the template selection converges in 2-3 iterations. The last phase of our framework is the final fitting procedure with the 'best' template and a regularization factor $\lambda = 1/500$.

In order to verify the quality of reconstruction, we test our approach in four different scenarios: with point clouds captured by laser scanner on 1) naked and 2) dressed bodies, with point clouds obtained from Kinect RGB-D cameras on 3) naked and 4) dressed bodies. The laser scanner used in this experiment is Vitus Smart XXL manufactured by Human Solution [32]. Although we do not carry simultaneous capture for both laser and Kinect scans, they are taken one after one in a very short time, so the shape difference between scans can be neglected. In order to verify the quality of reconstruction, we compare the reconstructed result with the ground truth. To build a ground truth for the experimental tests, the point cloud captured by laser scanner is reconstructed into a surface and remeshed in the same way as that in preparing the database. First of all, starting from the universal template, we test if our template selection algorithm can pick the ground truth out from the dataset as the final optimal template. We add the model of ground truth into the database in the example of Fig.13, and the result is encouraging, where the ground truth is successfully selected for all testing scenarios (see Fig.3 for the selection process). Note that, we does not have the ground truth in general cases, but this test verifies the success of our iterative template selection method. Another example using the database without ground truth is shown in Fig.14. In this example, even European database is used for our Asian input, we can still get reconstruction results in good quality. Second, we check the shape-approximation-error of the reconstructed human body w.r.t the ground truth. The color map of shape-approximation-error has been shown in Fig.13. Finally, we also conduct anthropometry measurements on five feature curves to verify the results. The measurements are shown in the table of Figs.13 and 14. From these tests, we can find that:

- Our iterative template selection routine can well-capture the target body shape, and the best template is selected.
- Our framework can accurately reconstruct the body shape from the point cloud generated on naked body by laser scanner (i.e., the shape-approximation-errors and anthropometry measurement errors are very small).
- The result obtained from the naked body captured by RGB-D cameras is not as accurate as the result from laser scanner, but it is better than the result obtained by dressed body scanned by laser scanner.
- The results obtained from dressed subjects are relatively poor. The region with the highest error is around the hip. It is because that the shirt and pants worn by subjects



	Ground Truth	Laser on Naked		Laser on Dressed		Kinect on Naked		Kinect on Dressed	
	Length	Length	Error	Length	Error	Length	Error	Length	Error
① **under-bust girth**	1.081	1.078	0.28%	1.088	0.65%	1.082	0.09%	1.087	0.56%
② **waist girth**	1.080	1.076	0.37%	1.084	0.37%	1.085	0.46%	1.103	2.13%
③ **lower-waist girth**	1.126	1.123	0.27%	1.127	0.09%	1.129	0.27%	1.143	1.51%
④ **hip girth**	1.138	1.138	0.00%	1.152	1.23%	1.142	0.35%	1.150	1.05%
⑤ **vertical trunk**	1.685	1.680	0.30%	1.675	0.59%	1.693	0.47%	1.698	0.77%

Fig. 13. To evaluate the error of our volumetric fitting based human reconstruction, we first construct the ground truth of a subject by scanning its human body using a laser scanner. The surface mesh of this human body is reconstructed by the Poisson's reconstruction [29] from the point cloud. Our algorithm is applied to the point clouds obtained from four different setups: by laser scanner on 1) naked and 2) dressed bodies, by Kinect RGB-D cameras on 3) naked and 4) dressed bodies. Shape-approximation-errors (w.r.t. the ground truth) at the main body are displayed as the color maps, and anthropometry measurements are taken at five feature curves to further verify the quality of our reconstruction.

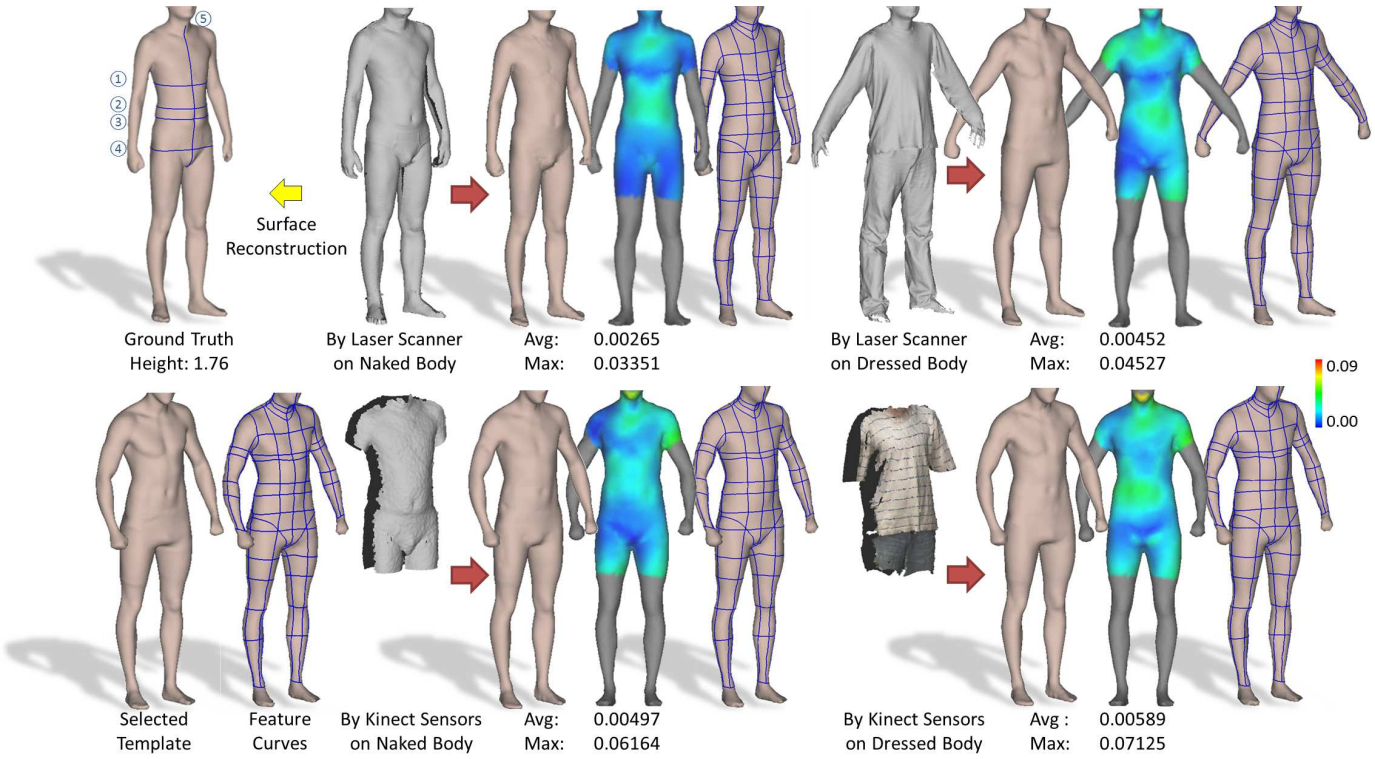
loosely cover the region of hip.

Improvements could be made if the database is enhanced and includes more variations of body shapes. Currently, the European database [15] is used in the tests, but the query inputs for template selection are Asian. More reconstruction results from dressed subjects scanned by RGB-D cameras are shown in Fig.15.

The reconstructed human models own the same connectivity with models in the database. Therefore, they can be used in a variety of applications.

Application I: *Extraction of Semantic Features*
The problem to extract semantic feature, especially those cannot be obtained by geometric extremities, is illy posed. Our work provides a robust tool to extract semantic features on 3D human models. Since our fitting framework can generate results having the same connectivity as the template, the feature curves defined on the template (see bottom-left in Fig.14) can be easily transferred to the reconstructed results (see the right-most model for all the four cases in Fig.14). Based on the transferred feature curves, we can automatically take anthropometry measurements on the reconstructed models. Examples are shown in Fig.13 and 14. Specifically, we measure (1) the under-bust girth, (2) the waist girth, (3) the lower-waist girth at the belly-button, (4) the hip girth, and (5) the vertical trunk, which are the most important features for body shape in clothing industry.

Application II: *Human-centric Design Automation*
After reconstructing a human model, the correspondences between triangles on the template model and the reconstructed model have been established. By this information, we can use deformation techniques (e.g. [22], [33], [34]) to warp an apparel product that is designed for the template model to a new shape that fit for the new model. Examples are shown in Fig.16. A suit and a wet-suit are designed for the male and female template respectively, and they can be automatically warped to fit the reconstructed human models. The techniques developed in the paper enable a lot of downstream applications like virtual try-on [35], real-time clothing synthesis [36], garment resizing [37] and product customization [38]. More discussions about this application can be found in Wang's book [26].



	Ground Truth	Laser on Naked		Laser on Dressed		Kinect on Naked		Kinect on Dressed	
	Length	Length	Error	Length	Error	Length	Error	Length	Error
① **under-bust girth**	0.837	0.835	0.24%	0.856	2.27%	0.847	1.19%	0.849	1.43%
② **waist girth**	0.792	0.787	0.63%	0.782	1.26%	0.791	0.13%	0.783	1.14%
③ **lower-waist girth**	0.827	0.826	0.12%	0.825	0.24%	0.826	0.12%	0.830	0.36%
④ **hip girth**	0.945	0.947	0.21%	0.996	5.40%	0.963	1.90%	1.003	6.14%
⑤ **vertical trunk**	1.683	1.682	0.06%	1.681	0.12%	1.664	1.13%	1.669	0.83%

Fig. 14. Another example is shown for verifying the quality results generated by our algorithm. Similar to Fig.13, the tests are conducted on four setups: by laser scanner on naked and dressed bodies, by Kinect RGB-D camera on naked and dressed bodies. In this example, the ground truth is not inserted into the database. Instead, an existing model in the database is selected as the template. By our algorithm, the same template is automatically selected in all the four scenarios. Again, shape-approximation-errors and anthropometry measurements are shown on the results obtained in these four scenarios. For the template model with pre-defined feature curves, these features can be automatically reconstructed on the fitting results.

Application III: *Interior Structures Transferring*
One interesting property of our volumetric fitting framework is that the correspondences are not only defined on boundary surface, but also defined in the whole volume. An example of transferring the interior structures (e.g., muscles) to the reconstructed models is shown in Fig.17. This can be used to create human models used in biomedical analysis and simulations. A very good feature of such a kind of mapping is that, as long as we can guarantee there is no degenerated (i.e., zero or negative volume) tetrahedron produced in the fitting, the mapping is bijective. This is different from the deformation techniques used above; the wrapping function does not guarantee an intersection-free transferring.

VI. Conclusion and Discussion

In this paper, we present a volumetric template fitting based method for reconstructing 3D human models from incomplete data, which are point clouds captured by inexpensive RGB-D cameras. With the help of this new approach, we develop a 3D body scanner by using two Kinect sensors. The data acquisition can be completed by taking a camera shot, so that our scanner is much faster than the conventional laser scanners for human bodies (taking more than 10 seconds for a scan). Although the data captured by this scanner are incomplete and have relatively low resolution (e.g., 640×480 RGB-D images are captured by Kinect sensors), the human bodies reconstructed by our algorithm have high quality and the fitting procedure is reliable. To further improve the result of template fitting, statistical model is employed in this paper. It is used to develop an iterative template selection routine for getting a best template from the database to generate a fitting result with smaller shape-approximation-errors. The whole procedure of human body reconstruction can be finished in around 30 seconds on a consumer-level PC. Results have been tested and demonstrated on the RGB-D images captured from both naked and dressed subjects. It shows that this method can generate 3D human models with high quality, and it can benefit a variety of applications.

In our current implementation, the selected template model is automatically scaled according to the remained points on the main body of the subjects. If the input point cloud is not cropped as what is expected, the scaling could be inaccurate

and lead to a poor fitting result. Fortunately, if our human body reconstruction framework is used for a well calibrated setup (e.g., our instant scanner based on RGB-D cameras), the scaling problem can be solved by specifying a feature point (e.g., the neck point) on the point cloud (ref. [39]). Moreover, our algorithm relies pretty much on the exemplars stored in the database of human models. If the shape of the subject cannot be spun by our human model database, the fitting result will not be satisfactory. In order to solve this limitation, we will add more exemplars with a variety of shapes into the database in our future work. For example, a database with Asian subjects can be built.

Another limitation of the current implementation is that the regularization factor, λ, is selected in a heuristic manner. Although the suggestion about how to choose the values in different phases of the algorithm (i.e., pose alignment, template selection, and final fitting) has been given, it could be different when the numbers of vertices and tetrahedra of the template are changed. A better method to select an appropriate regularization factor will be considered in our near future work.

ACKNOWLEDGMENT

The authors would like to thank TPC (HK) Limited providing the clothing designs shown in Fig.16.

REFERENCES

[1] Cyberware, "Whole body color 3D scanner," 2007, http://cyberware.com/.

[2] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun, "Performance capture from sparse multi-view video," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 98:1–98:10, Aug. 2008.

[3] K. Kolev, M. Klodt, T. Brox, and D. Cremers, "Continuous global optimization in multiview 3D reconstruction," *International Journal of Computer Vision*, vol. 84, pp. 80–96, 2009.

[4] A. Weiss, D. Hirshberg, and M. Black, "Home 3D body scans from noisy image and range data," in *2011 IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1951–1958.

[5] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using kinects," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 643–650, 2012.

[6] R. Wang, J. Choi, and G. Medioni, "Accurate full body scanning from a single fixed 3D camera," *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, vol. 0, pp. 432–439, 2012.

[7] K. Khoshelham, "Accuracy analysis of kinect depth data," *International Society for Photogrammetry and Remote Sensing (ISPRS) Workshop Laser Scanning*, vol. 38, no. 5/W12, p. 1, 2010.

[8] A. Maimone and H. Fuchs, "Reducing interference between multiple structured light depth sensors using motion," in *IEEE Virtual Reality*, 2012, pp. 51–54.

[9] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, and H.-P. Seidel, "Estimating body shape of dressed humans," *Computers & Graphics*, vol. 33, no. 3, pp. 211–216, 2009.

[10] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: shape completion and animation of people," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 408–416, Jul. 2005.

[11] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," in *IEEE International Conference on Computer Vision*, oct 2009, pp. 1381–1388.

[12] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormahlen, and H. Seidel, "Multilinear pose and body shape estimation of dressed subjects from image sets," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, june 2010, pp. 1823–1830.

[13] A. O. Bălan and M. J. Black, "The naked truth: Estimating body shape under clothing," in *Proceedings of the 10th European Conference on Computer Vision: Part II*, ser. ECCV '08, 2008, pp. 15–29.

[14] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: reconstruction and parameterization from range scans," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 587–594, Jul. 2003.

[15] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, "A statistical model of human pose and body shape," *Computer Graphics Forum*, vol. 28, no. 2, pp. 337–346, 2009.

[16] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ser. CVPR '91, 1991, pp. 586–591.

[17] F. Shih, *Image Processing and Pattern Recognition: Fundamentals and Techniques*. Wiley, 2010.

[18] H. Pottmann and S. Leopoldseder, "A concept for parametric surface fitting which avoids the parametrization problem," *Computer Aided Geometric Design*, vol. 20, no. 6, pp. 343–362, sep 2003.

[19] W. Wang, H. Pottmann, and Y. Liu, "Fitting b-spline curves to point clouds by curvature-based squared distance minimization," *ACM Transactions on Graphics*, vol. 25, no. 2, pp. 214–238, April 2006.

[20] V. Nivoliers, D.-M. Yan, and B. Levy, "Fitting polynomial surfaces to triangular meshes with voronoi squared distance minimization," *Engineering with Computers*, pp. 1–12, 2012.

[21] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Proceedings of the fifth Eurographics symposium on Geometry processing*, ser. SGP '07, 2007, pp. 109–116.

[22] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 399–405, Aug. 2004.

[23] L. Liu, L. Zhang, Y. Xu, C. Gotsman, and S. J. Gortler, "A local/global approach to mesh parameterization," in *Proceedings of the Symposium on Geometry Processing*, ser. SGP '08, 2008, pp. 1495–1504.

[24] K. G. Kobayashi and K. Ootsubo, "t-FFD: free-form deformation by using triangular mesh," in *Proceedings of the eighth ACM Symposium on Solid Modeling and Applications*, ser. SM '03, 2003, pp. 226–234.

[25] C.-H. Chu, Y.-T. Tsai, C. C. L. Wang, and T.-H. Kwok, "Exemplar-based statistical model for semantic parametric design of human body," *Computers in Industry*, vol. 61, no. 6, pp. 541–549, 2010.

[26] C. C. L. Wang, *Geometric Modeling and Reasoning of Human-Centered Freeform Products*. Springer London, 2013.

[27] B. A. Draper, K. Baek, M. S. Bartlett, and J. Beveridge, "Recognizing faces with PCA and ICA," *Computer Vision and Image Understanding*, vol. 91, no. 1–2, pp. 115–137, 2003, special Issue on Face Recognition.

[28] D. M. Mount and S. Arya, "ANN: A Library for Approximate Nearest Neighbor Searching," 1997.

[29] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the fourth Eurographics symposium on Geometry processing*, ser. SGP '06, 2006, pp. 61–70.

[30] T.-H. Kwok, Y. Zhang, and C. C. L. Wang, "Efficient optimization of common base domains for cross parameterization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 10, pp. 1678–1692, oct. 2012.

[31] ——, "Constructing common base domain by cues from voronoi diagram," *Graph. Models*, vol. 74, no. 4, pp. 152–163, 2012.

[32] Human Solution, "Vitus Smart XXL," 2010, www.human-solutions.com/fashion/front_content.php?idcat=139&lang=7.

[33] C. C. L. Wang, Y. Wang, and M. M. F. Yuen, "Design automation for customized apparel products," *Computer-Aided Design*, vol. 37, no. 7, pp. 675–691, Jun. 2005.

[34] C. C. L. Wang, K.-C. Hui, and K.-M. Tong, "Volume parameterization for design automation of customized free-form products," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 1, pp. 11–21, 2007.

[35] Y. Meng, P. Y. Mok, and X. Jin, "Interactive virtual try-on clothing design systems," *Computer-Aided Design*, vol. 42, no. 4, pp. 310–321, 2010.

[36] W. Xu, N. Umentani, Q. Chao, J. Mao, X. Jin, and X. Tong, "Sensitivity-optimized rigging for example-based real-time clothing synthesis," *ACM Transactions on Graphics (Proc. Siggraph'2014)*, vol. 33, no. 4, 2014.

[37] Y. Meng, C. C. L. Wang, and X. Jin, "Flexible shape control for automatic resizing of apparel products," *Computer-Aided Design*, vol. 44, no. 1, pp. 68–76, 2012.

[38] T. H. Kwok and C. C. Wang, "Shape optimization for human-centric product with standardized components," *Computer-Aided Design*, vol. 53, pp. 51–63, 2014.

[39] C. C. L. Wang, Y. Wang, T. K. Chang, and M. M. Yuen, "Virtual human modeling from photographs for garment industry," *Computer-Aided Design*, vol. 35, no. 6, pp. 577–589, 2003.

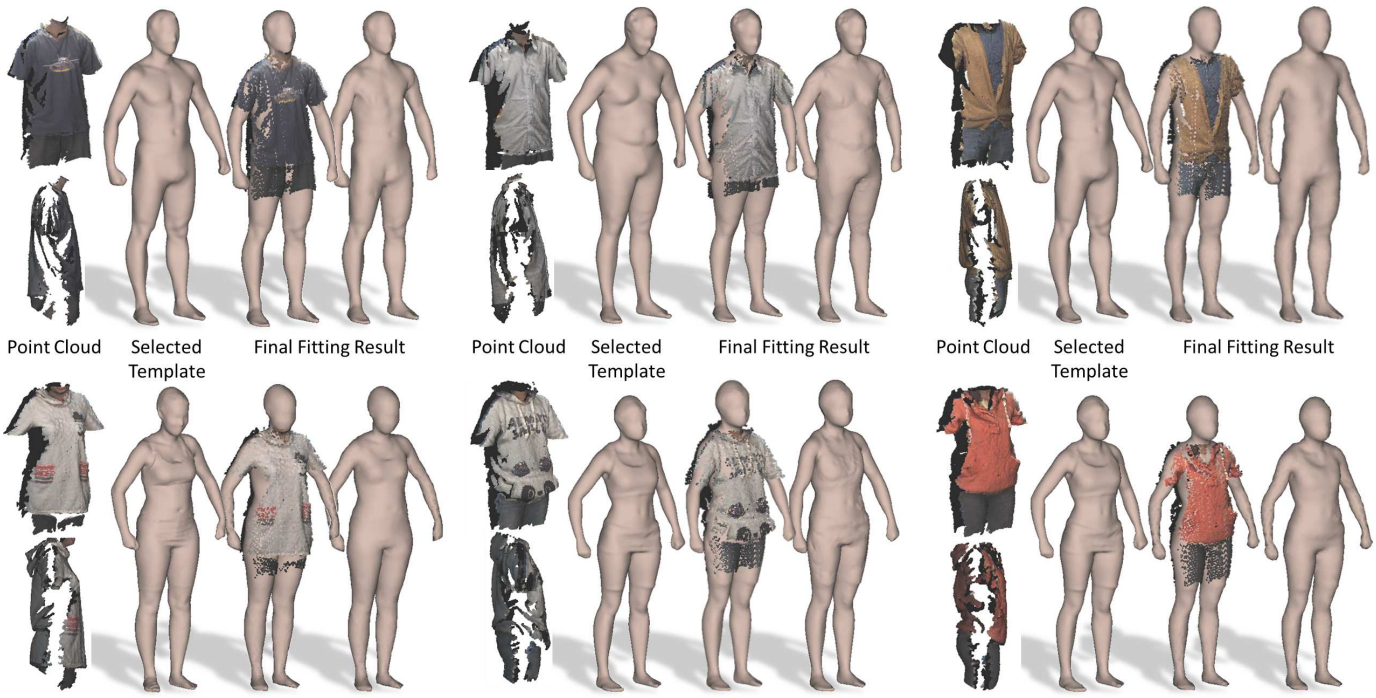


Fig. 15. The estimation of 3D human bodies from different dressed subjects (captured by RGB-D cameras) obtained by our volumetric template fitting based framework.

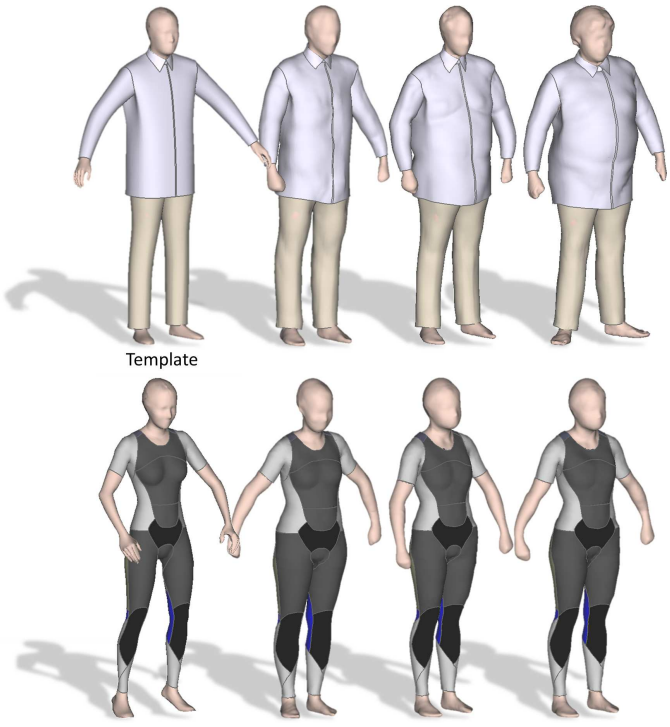


Fig. 16. A suit is designed for the male template, and a wet-suit is designed for the female template (left). The designs can be transferred to fit the reconstructed human models automatically.

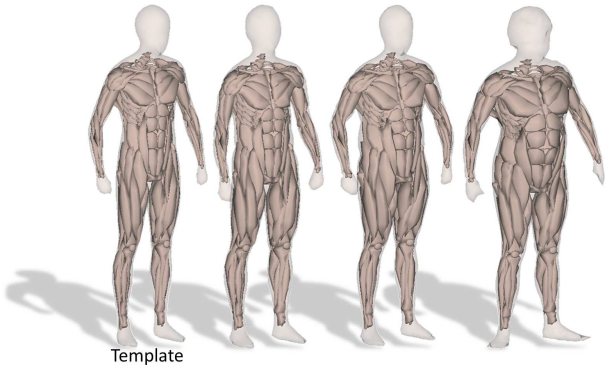


Fig. 17. Volumetric template fitting can support the function of transferring interior structures (e.g., muscles) onto the reconstructed human models.