Vis Comput (2015) 31:19–33
DOI 10.1007/s00371-013-0901-6

ORIGINAL ARTICLE

Learning human shape model from multiple databases with correspondence considering kinematic consensus

Yao Yu · Yu Zhou · Sidan Du · Yuan Jie · Ziqiang Wang · Zhengyu Cai

Published online: 6 December 2013

Abstract In various applications of computer graphics and model-based computer vision, a human shape model cannot only model the kinematic properties of a subject to drive the mesh into various postures, but it can also be utilized to parameterize the shape variations across individuals. It is of great benefit to improve the diversity of the training databases by learning the model from multiple databases, once the correspondences among scans of these databases can be achieved. To accomplish this goal, we proposed a framework to match the scans from multiple databases, using the assistance of kinematic properties, to compute the correspondences. The resulting correspondence is accurate, robust, capable of handling scan incompletion, and is homogeneous across shapes and postures. In our approach, we start with evaluating how a correspondence, which is achieved via minimizing the deformation energy, agrees with the kinematic properties, and then, we jointly fit the source scans to the target scans to derive the correspondences between the databases. The extensive results show that our approach can generate a faithful correspondence even in extreme cases, without carefully selecting the deformation factors and markers. We also developed a method, with which a commendable and predictable result can be synthesized, to control the rendered shape in an intuitive way.

Keywords Deformations · Morphing · Non-rigid registration · Motion capture and synthesis · Correspondence

Y. Yu (✉) · Y. Zhou (✉) · S. Du · Y. Jie · Z. Wang · Z. Cai
School of Electronic Science and Engineering,
Nanjing University, Nanjing, Jiangsu 210046, China
e-mail: allanyu@nju.edu.cn

Y. Zhou
e-mail: nackzhou@nju.edu.cn

S. Du
e-mail: coff128@nju.edu.cn

1 Introduction

Human shape models are widely utilized in applications of motion capture [23,26], surface animation [29], shape completion [3], mesh deformation and synthesis [12,15]. They not only can model the kinematic properties of a subject as an articulated skeleton, which can be used to drive the mesh deforming into various postures, but are also capable of parameterizing the shape of different individuals. Therefore, an optimization-based motion capture system, e.g., [26], can utilize such models to track the motion of any subject regardless of how the shape changes. To accomplish this goal, approaches are proposed for learning human shape models from a series of scans for hundreds of individuals in various postures, such as EigenSkin [16], Linear Blending Skinning [12,17], SCAPE [3], and so forth.

In these human body models, the surface variations, which arise from varying postures and shapes, are learned from the training database in the form of meshes or point clouds. The performance of these learned models is thereby highly dependent on the diversity of postures and shapes in the input scans. Existing databases typically contain tens of scans, such as the partially opened SCAPE database [3] and the fully opened MPI database [9]. The opened subset of the SCAPE database provides scans for a single subject in different postures at a relatively high scan density, while the MPI database contains shapes for approximately 100 individuals at a relatively low scan resolution. Learning a model from either of these databases alone cannot achieve an attractive result in terms of generalizability. Nevertheless, the result can be improved via learning the model from scans that come from two or more databases, which is of great benefit in enhancing the diversity of the training scans.

Merging the databases to learn a human body model is not accomplished by simply placing all of the scans from

Fig. 1 We merge multiple databases via establishing correspondences with the assistance of the kinematic consensus to learn the human shape model, which can be used to generate surfaces across various shapes of humans in arbitrary postures

different databases together because all of the existing approaches for training human shape models are highly dependent on the correspondences between any pair of scans to measure and parameterize how the surface varies with the changes in postures and shapes. Specifically, the shape variations are investigated by calculating changes between a vertex/triangle and its corresponding vertex/triangle in scans for different postures or shapes. The correspondences are used here to indicate the mapping between the vertex and its counterpart. In the most common case, the correspondences in a database are usually derived with the assistance of markers, which are installed uniformly before the subjects are scanned. For lack of these markers, it imposes a significant difficulty in establishing the correspondences accurately between databases that are scanned separately compared with databases that are scanned uniformly; however, in this paper, we will show that this concern can be relaxed with the help of the kinematic properties.

On the other hand, the kinematic properties should be taken into account in achieving the correspondences across different postures and shapes, to allow the derived body model to describe the shape variations homogeneously. For example, to establish correspondences between a source database A and a target surface b via a correspondence algorithm, which is achieved by minimizing the deforming energy as [24], the correspondences should be different when diverse surfaces in A are chosen to be deformed to the target surface b. These differences will obviously affect the learned human body model; even in this case, the same database A plus a single surface b are used as input. Moreover, in some approaches, such as SCAPE, the shape variations are eventually parameterized via Principal Component Analysis (PCA) [13]. This heterogeneity, which results from scan-to-scan deformation without accounting for the kinematic properties, will greatly enlarge the PCA space to encode the variations and, thus, will correspondingly increase the parameter search space in applications of the model, such as the motion capture that jointly optimizes the shape and motion parameters. Last but not least, without considering the kinematic properties, the deforming energy, the deforming factors and the human-induced markers should be designed carefully. Otherwise, the synthesized shape can become drastic in some extreme cases.

Figure 1 illustrates the learned SCAPE model from a merged database. To this end, we proposed an approach in this paper to establish correspondences among multiple databases to enrich the diversity of the training datasets. Sections 3.1 and 3.2 show how the proposed approach incorporates the kinematic properties as the consensus in the postures, and a deformation process is performed to determine the correspondences between scans from different databases. Our method not only improves the accuracy of the correspondences but also narrows the parameter space that describe the shape variations. In Sect. 3.3, we demonstrate that incomplete scans can also be used to learn the model in the proposed approach; however, in contrast to traditional approaches that directly use the template model to fill the holes, we complete the missing parts via the learned experience on the shape variations. Then, the scans for all of the shapes and postures, from ballet dancers to sumo wrestlers, are used to learn the human shape model, which is further parameterized into the PCA space and translated into a direct way to control the generated shapes intuitively in Sect. 4, such as height, weight, arm length, and other measurements.

2 Related work

To learn a human shape model from multiple databases, we require the intrinsic correspondence information between the scans in these databases. The established correspondences will be used to measure the variation from a certain vertex/triangle to its corresponding part, which will be modeled and parameterized to generate the body shape deformation across different humans in different postures. The shape variations will be estimated per vertex/triangle; thus, we should obtain dense correspondences rather than compute sparse correspondences, such as [5].

Computing dense correspondences between two scans by feature matching and rigid transformation has been widely studied in [1,8,11] and [21], with an assumption that each scan can be perfectly aligned with a rigid transformation. These methods typically solve the correspondence problem based on verifying the rigid transformations that involve how the transformation matches the feature points [11] or closest points [21], and sometimes cooperate with RANdom SAmple Consensus (RANSAC) [7] for noise in the scans. Once the feature points or closest points are well aligned, the correspondences are constructed by applying the rigid transformation to the surface and assigning the correspondences of a vertex to be its closest counterpart. These approaches are not suitable for establishing the correspondences between two scans of human bodies, due to their non-rigid, deformed surface, and varying postures. Therefore, the correspondences between the scans for human bodies usually focus attention on only the non-rigid registration.

A common approach to model the non-rigid transformation that arises from human motion is usually to describe the motion of humans into a highly articulated 3D model. The parts of the model are tree-like linked, and the transformation of them is considered to be piecewise rigid [6,20,28]. Although such a simplified model works for scans of some artifact subjects, it can fail due to independent localized bending or stretching of the shapes in an elastic manner, especially at the joints between two rigid components. Moveover, to learn a human body model, these methods no longer work because of variations across different humans.

Recent methods [2,18] are proposed to address the elastic variations in shapes in applications in which the rigid and piecewise rigid assumption no longer makes sense. Allen et al. [2] employ an optimization-based approach to drive the source meshes to the target scan, while minimizing the deformation energy, and to iteratively find the closest counterpart of each vertex. These deformations can become stuck at a local minimum, once the pose and shape variation is large. Huang et al. [10] also optimize the deformation process to achieve a non-rigid registration and generate the correspondences using a pruning mechanism, which is similar to the approach proposed in [24]. However, a deformation-based approach without considering the kinematic properties can result in high levels of noise, which will in turn affect the learned model significantly. Zhang et al. [27] present an automatic feature-based correspondence algorithm to handle the non-rigid shape variations; nevertheless, similar to the method in [6], the results are not robust for noise and for incomplete scans of human bodies, and the features can vary greatly between bodies. Similar to [2], our approach also utilizes an optimization-based framework to achieve the correspondences. However, instead of performing pair-wise deformation, we achieve the correspondences by evaluating how such correspondences meet all of the pairs of meshes in the source and target databases.

More sophisticated approaches incorporate the geometry information and high-level shape semantics into the shape analysis pipeline, such as part styles [25], shape analogies [22], and prior knowledge [14], to establish the correspondences. Inspired by these approaches; however, we introduce more meaningful high-level information to compute the correspondences, which are accurate, robust, capable of handling scan incompletion, and homogeneous across shapes and postures. To meet these requirements, the proposed approach in this paper incorporates the kinematic properties, the consensus in postures, and the deformation energy to determine the correspondences. Even in some extreme cases, the derived correspondences are well established in human model learning. Moreover, it is homogeneous so that the learned model requires a relatively smaller PCA dimension to parameterize the shape variations.

3 Approach

As illustrated in Fig. 2, the purpose of our approach is to learn a human body model from multiple databases. Two databases are used in this paper: the SCAPE database [3] and the MPI database [9]. The SCAPE datasets are partially made available for research purposes, in which there are 71 scans of different postures for the same person. Each of the scans contains a set of original scanned meshes with 125K polygons and a set of hole-filled and simplified meshes with 25K polygons. The correspondences within the database have been established between each pair of scans using the correlated correspondence algorithm [4], and the indices of the vertices in each scan have been aligned in the order of the correspondences accordingly.

Because only a subset of scans in different postures for the same subject are available in SCAPE, to learn the human shape model, we merge the SCAPE database and the MPI database to enrich the diversity of the training datasets of scans across varying postures and individuals. The MPI database contains a set of relatively low-resolution scans of 111 individuals with 35 postures, and each scan is composed

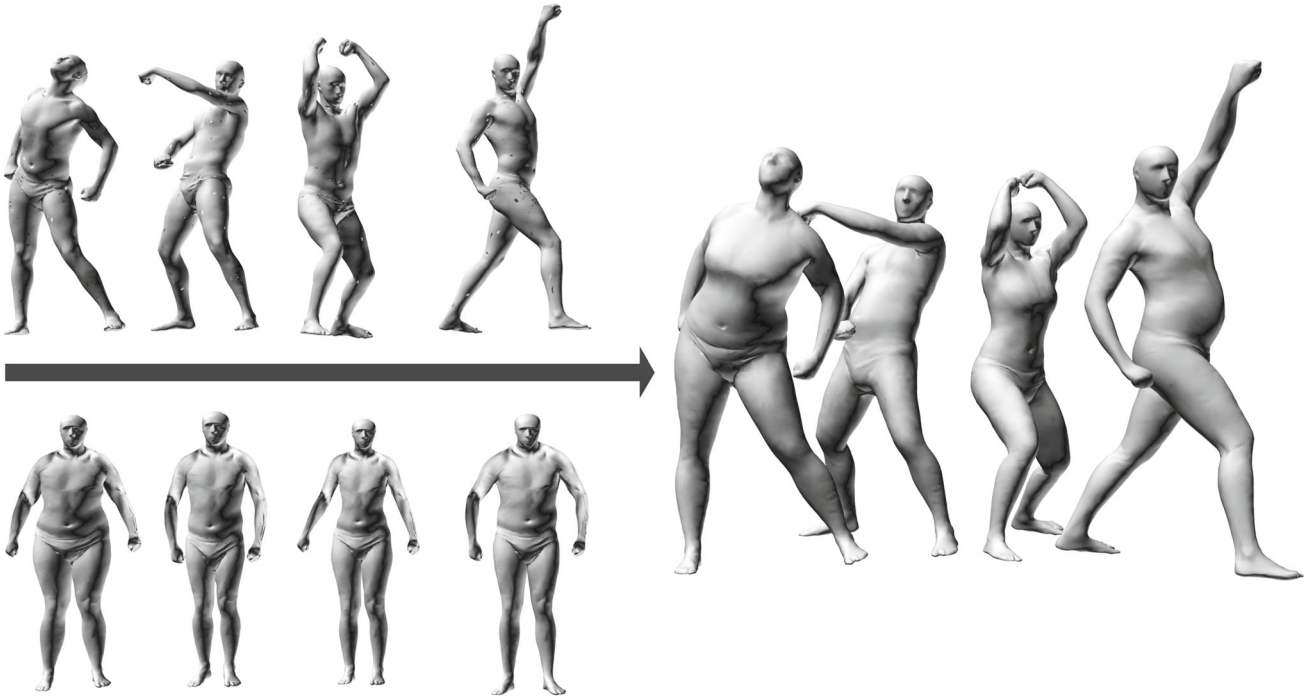


Fig. 2 We merge multiple databases, each of which contains tens of scans across different shapes and postures (*left*), via establishing the correspondences with the kinematic consensus to learn a human shape model, which can be used to generate various shapes of humans in arbitrary postures (*right*)

of 6, 449 vertices with 12, 894 polygons each. Similar to the SCAPE database, the vertices in the MPI database are also arranged in the order of the correspondences among the scans.

To learn the human body model from multiple databases, the correspondences between each pair of scans in different databases should be established as the first step. For the sake of this, we establish the correspondences with kinematic consensus in Sect. 3.2, which is different from the traditional correspondence approach. A further discussion of incomplete scans will be detailed in Sect. 3.3. To show the distinction between the correspondences within a database and the correspondences between databases, we call the former 'inner-correspondences' and the latter 'correspondences' for short. The inner-correspondences are assumed to be a part of the input data, and therefore, we focus on computing the correspondences between databases. Note that this assumption is not necessary to meet our approach; either our proposed approach or any other correspondence algorithms can achieve this goal, such as [4].

3.1 Optimization-based correspondences

Optimization-based approaches drive the source meshes to the target scan, while minimizing the deformation energy, to establish the correspondences between this pair of scans, such as the approaches illustrated in [2,18], that iteratively find the closest counterpart of each vertex, and shape feature-based approaches [10], which utilize the principal curvatures and geodesic distance to evaluate the correspondences in each iteration. Once the pose and shape variation is large between the source and target, it is easy for these deformations to become stuck at a local minimum. It is difficult to investigate the correctness of the derived correspondences, due to the local minimum of optimization. However, we found that the correctness could be judged among other pairs of scans in the source and target datasets because the incorrect correspondences usually fail to drive all of the other pairs of scans to be stuck at a local minimum. Inspired by this observation, we propose a pipeline to establish more robust correspondences with all of the scans in the source and target. In this section, we begin with the deformation-based approach, and an optimization framework to achieve the correspondences with kinematic consensus will be demonstrated in Sect. 3.2.

Because the datasets that are used for learning the human body model come from different sources, there is no uniformly pre-installed marker that is associated with the scanned data. Furthermore, the datasets are scanned across different individuals and postures, therefore, a set of markers, $\{z_1 \dots z_L\}$, are manually introduced into the input scans respectively. Because of the existence of the inner-correspondences, each marker can be introduced separately

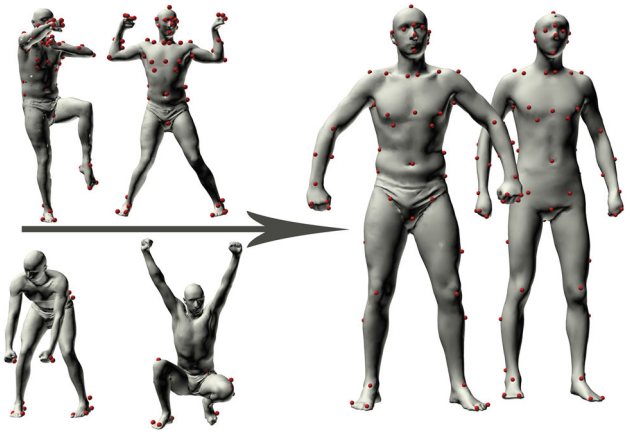


Fig. 3 The markers are introduced into the scans in different postures (*left*) and assembled via the method of voting (*right*)

into and located easily in the scans, where such a marker is easy to locate. Later, these markers are assembled via the method of voting to unify their associated vertices, as illustrated in Fig. 3.

To establish the correspondences between databases, we fit the source surface, $\mathbb{S}$, to the target surface, $\mathbb{T}$, in which $\mathbb{S}$ indicates the set of the scans, $\mathbb{S}_s, s \in [1 \ldots |\mathbb{S}|]$, of the source surface. Each of the scans $\mathbb{S}_s$ is composed of N vertices, which are denoted as $v_i^s, i \in [1 \ldots N]$. To accomplish the matching between $\mathbb{S}_s$ and $\mathbb{T}_t$, we employ an optimization framework to derive a set of affine transformations that minimize the deformation energy:

$$E_{s,t} = \alpha \mathrm{ED}_{s,t} + \beta \mathrm{ES}_{s,t} + \gamma \mathrm{EM}_{s,t} \tag{1}$$

To accomplish a good matching, corresponding meshes in $\mathbb{S}$ and $\mathbb{T}$ should be as close as possible. Hence, a distance error, $\mathrm{ED}_{s,t}$, is defined as the sum of squared distances between each vertex in $\mathbb{S}_s$ and its counterpart in $\mathbb{T}_t$:

$$\mathrm{ED}_{s,t} = \sum_{i=1}^{N} ||\hat{v}_i^s - c_i^s||^2 \tag{2}$$

where $\hat{v}_i^s$ is the deformed location for v_i^s, which is solved by minimizing the energy function, to fit the target meshes. Here, c_i^s is the closest vertex in the meshes $\mathbb{T}_t$ of the deformed vertex $\hat{v}_i^s$.

The purpose of the deformation process is to determine the closest counterpart for each vertex; however, using only the criterion $\mathrm{EM}_{s,t}$ could lead to an under-constrained objective function and an undesirable match that the vertices in $\mathbb{S}_s$ are mapped to disparate parts of $\mathbb{T}_t$, and vice-versa. To achieve a smooth matching, a smoothness error, $\mathrm{ES}_{s,t}$, is introduced to drive the meshes in $\mathbb{S}$ to the $\mathbb{T}$ smoothly along with the vertex in their neighborhood. Specifically, we constrain the Laplacian Coordinates [19] of the deformed vertex $\hat{v}_i^s$ to remain as similar as possible to v_i^s:

$$\mathrm{ES}_{s,t} = \sum_{i=1}^{N} ||T_i L(\hat{v}_i^s) - L(v_i^s)||^2 \tag{3}$$

in which $L(\cdot)$ denotes the Laplacian Coordinates, which can be derived by

$$L(v) = v - \frac{1}{\mathrm{Deg}(v)} \sum_{u \in Adj(v)} u \tag{4}$$

where $\mathrm{Adj}(v)$ denotes the set of vertices in the neighborhood of v, and $\mathrm{Deg}(v)$ represents the number of degrees of its neighborhood. Obviously, the Laplacian Coordinates are not rotation-invariant, therefore, the they should also change their orientations if the meshes rotate. T_i is the rotation matrix of vertex v_i^s.

The terms $\mathrm{ED}_{s,t}$ and $\mathrm{ES}_{s,t}$ drive the meshes in $\mathbb{S}_s$ moving toward the meshes in $\mathbb{T}_t$ as long as they are close enough to each other initially. In most common situations, however, the meshes in $\mathbb{S}_s$ have not been well aligned to the meshes in $\mathbb{T}_t$, which implies that the optimization will become stuck at a local minimum. To avoid local minima, we introduce a maker error $\mathrm{EM}_{s,t}$:

$$\mathrm{EM}_{s,t} = \sum_{i=1}^{L} ||\hat{v}_{Z_s(i)}^s - u_{Z_t(i)}^t||^2 \tag{5}$$

where the $Z_s(i)$ and $Z_t(i)$ are vertex indices of marker z_i in $\mathbb{S}$ and $\mathbb{T}$, respectively.

As illustrated in Fig. 4, the markers in $\mathbb{S}_s$ are transformed to the locations of their corresponding parts in $\mathbb{T}_t$ due to the

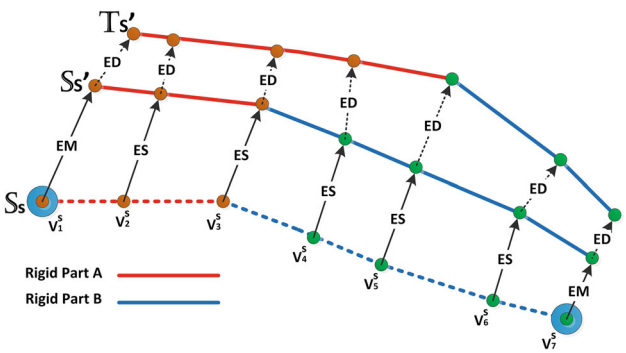


Fig. 4 Summary of establishing the correspondences based on the surface deformation. To establish the correspondences, the markers in the source $\mathbb{S}_s$ will be driven toward their counterparts in the target $\mathbb{T}_t$ by the EM energy term, while the other vertices will be carried along by the markers due to the ES term. As the deformation process iterates, the ED term in the energy function increasingly becomes dominant. Eventually, via minimizing the three energy terms, the source reaches the target and the correspondences are established after the deformation. It is also observed that, as a deformation energy-based correspondence algorithm, the resulting correspondence will minimize the deformation energy while ignoring the kinematics of human motion. For example, in this figure, the vertex v_4^s in the rigid bone B reaches its closest counterpart in terms of the minimized energy cost; however, the derived corresponding vertex belongs to the rigid part A

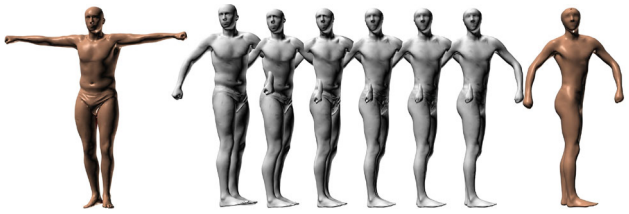


Fig. 5 Summary of the process of the deformation. The deformation process is performed iteratively. As the process is iterating, the source meshes increasingly deform toward the target

maker error term initially. As the markers are being transformed, the other vertices will be driven to move toward the meshes in $\mathbb{T}_t$ according to the smooth error $\mathrm{ES}_{s,t}$. In the next phases, the vertices in $\mathbb{S}_s$ will be further deformed to their closest vertex in $\mathbb{T}_t$, as the $\mathrm{ED}_{s,t}$ term becomes increasingly dominant. Figure 5 demonstrates an example of such a deformation process. We iteratively solve the optimization problem, and eventually the surfaces are matched and the correspondence is established. The details of solving the deformation energy function are given in Sect. 3.2.

3.2 Correspondence with a kinematic consensus

The process of deformation addressed in Sect. 3.1 is to drive the meshes in $\mathbb{S}_s$ to $\mathbb{T}_t$ with minimized deformation energy; however, it was found that the resulting match is heavily affected by the postures of the source and target scans. Specifically, the lowest deformation energy might not indicate a very attractive match. For example, in the example demonstrated in Fig. 4, the vertices $v_4^s \dots v_7^s$ are the components of the rigid part A, and other vertices are of the rigid part B. It is easy to verify that the derived match that minimizes the energy cost is the match that is depicted in Fig. 4, which means that the vertex v_4^s in the rigid part A near the joints is driven to a wrong location in part B. This inaccurate deformation will lead to a wrong correspondence that is used by learning the human body model. A better match is to align the components of rigid part A in $\mathbb{S}_s$ to their corresponding components of the rigid part A in $\mathbb{T}_t$; however, this goal cannot be achieved given such a pair of scans in Fig. 4. This finding implies that different postures will lead to different matches, and some postures could be suitable for some rigid parts of a shape to compute the correspondence, and others could be good for other parts.

Inspired by this observation, we improve the deformation process to incorporate a kinematic consensus into the deformation process across different postures and shapes to establish a more accurate correspondence as demonstrated in Fig. 6. Instead of transforming a single source in $\mathbb{S}$ to a single target in $\mathbb{T}$, we utilize the inner correspondence within $\mathbb{S}$ and $\mathbb{T}$ respectively, to jointly fit the scans in $\mathbb{S}$ to all of the scans in $\mathbb{T}$ to derive the correspondence. To accomplish this goal, we deform the scans from $\mathbb{S}$ to $\mathbb{T}$ to minimize the joint energy function:

$$
\begin{aligned}
&E = \alpha \sum_t \mathrm{ED}_{\kappa(t),t} + \beta \sum_t \mathrm{ES}_{\kappa(t),t} + \gamma \sum_t \mathrm{EM}_{\kappa(t),t} \\
&\text{s.t.} \\
&\mathrm{ED}_{s,t} = \sum_{i=1}^{N} \omega_i^t ||\hat{v}_i^s - c_i^s||^2 \\
&\mathrm{ES}_{s,t} = \sum_{i=1}^{N} \omega_i^t ||L(\hat{v}_i^s) - L(v_i^s)||^2 \\
&\mathrm{EM}_{s,t} = \sum_{i=1}^{L} \omega_i^t ||\hat{v}^s_{Z_s(i)} - u^t_{Z_t(i)}||^2
\end{aligned}
\tag{6}
$$

in which the $\kappa(\cdot)$ denotes the mapping from a scan in $\mathbb{T}$ to its best matched scan in $\mathbb{S}$. Thus, we compute the coordinates in the PCA space for any scan in $\mathbb{T}$ or $\mathbb{S}$. For each scan, the coordinates for those markers are filtered out and stacked into a vector according to their kinematic properties, such as the chain order in the skeleton, to describe the features of such a scan. The best matched scan for each of surfaces in $\mathbb{T}$ is then computed by searching the closest scan in $\mathbb{S}$ in terms of the distance between the feature vectors. The energy function in (6) then jointly considers the sum of the deformation energy to each scan $\mathbb{T}_t$ from its most similar scan $\mathbb{S}_{\kappa(t)}$. The resulting correspondences can be referred to as a weighted mean value that minimizes the deformation energy between all of the pairs of scans. That arrangement means that the derived correspondences will naturally follow

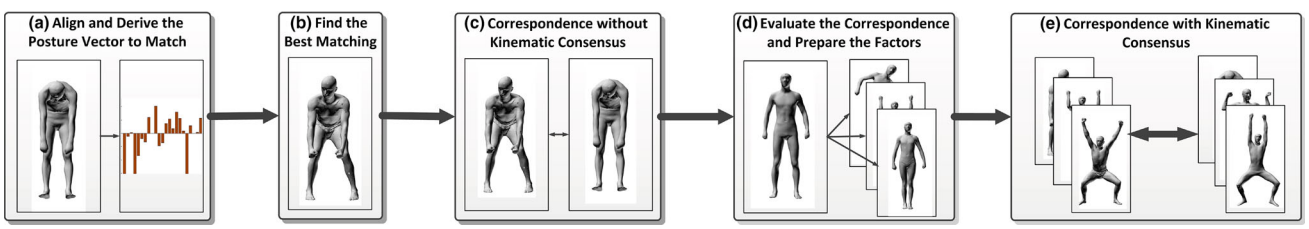


Fig. 6 The pipelines of computing correspondences with the kinematic consensus. **a** We extract and align the feature vectors for each scan in both the source and target databases via the PCA algorithm, and **b** the best matching scan for each target is determined for the next phase. **c** The correspondences without considering the kinematic properties is established for the matched scans to **d** prepare the impact factors ω_i^t. **e** Eventually, we jointly fit the source scans to the target scans to derive the correspondences between the two databases

the transitivity. For example, the correspondences between S_i and T_p and the correspondences between S_j and T_q are derived via the proposed method, in which S_i and S_j are from the same database, as well as T_p and T_q. One can also establish the same correspondences between S_i and T_p via $S_i \Leftrightarrow S_j \Leftrightarrow T_q \Leftrightarrow T_p$.

A factor ω_i^t is introduced to weight how the deformation from $\mathbb{S}_s$ to $\mathbb{T}_s$ impacts the vertex v_i^s. To derive these factors, we initially solve the optimization problem in (1) to match each pair of scans $\mathbb{S}_{\kappa(t)}$ and $\mathbb{T}_t$ to obtain a rough correspondence $\mathbb{C}_{\kappa(t),t}$ between them, $\forall t$. With each $\mathbb{C}_{\kappa(t),t}$, we follow the SCAPE model to learn the shape deformation model with the scan for the new shape. The learned SCAPE model is then used to generate an instance shape in each posture that is the same as $\mathbb{T}_u$, $\forall u$, and the error is derived between the generated instance and $\mathbb{T}_u$ at vertex i, denoted as $e_i^{t,u}$. The weight factor ω_i^t is obtained and latter normalized by

$$\omega_i^t = \phi\left(\frac{\sum_{u=1}^{|\mathbb{T}|} e_i^{t,u}}{|\mathbb{T}|}\right) \tag{7}$$

where $\phi(\varepsilon) = \exp(\frac{-\varepsilon^2}{2\sigma^2})$ and we adopt σ as the scan resolution of the database used in the experiments. The final correspondence is then achieved by applying $\kappa(\cdot)$ and ω_i^t to the energy function in (6).

We solve both the energy functions in (1) and (6) in an iterative way. In the first iteration, we ignore the distance error term using the weights $\alpha = 0, \beta = 1, \gamma = 10$; thus, the marker error term is the dominant constraint to drive the meshes in this phase. As the markers deform, other meshes will be carried along by the smoothness error term, and a set of valid closest points of each vertex can be found in the subsequent iteration. Then, in the second phase, the optimization problem is solved by increasing α from 0.5 to 100 in four steps and preserving $\beta = 1, \gamma = 1$. As α increases, the meshes in the source approximate the meshes in the target more and more closely after each iteration. In each iteration, we update the closest points for each vertex in $\mathbb{S}$ from the compatible vertices, whose normal vectors are no more than 90° apart from the normal of such a vertex in $\mathbb{S}$. It is noticeable that, in the process of deformation with a kinematic consensus, the closest vertex is computed among a set of super vertices. A super vertex here, more specifically, is generated by stacking all of the vertices $\hat{v}_i^{\kappa(t)}$, $\forall t$ into a column vector, whose dimension is $3 \times |\mathbb{T}|$, as well as the vertices u_i^t and $c_i^{\kappa(t)}$, $\forall t$.

The resulting correspondences in deformation-based methods can suffer from large shape and posture variations between the source and target scans. In our framework, however, we jointly compute the correspondences between all of the source and target scans; therefore, the algorithm can perform well as long as the distances are not too long. If this occurs, then the correspondences also can be derived among the other scans that can meet their close enough matching counterpart. Because the algorithm performs well when we consider all of the scans in the MPI and SCAPE databases, to investigate this situation, we artificially regard some scans as the scans who are too far away from their matching scans. To achieve this goal, we sort the pair-wise matching. Instead of using all of these pairs to establish the correspondences, we employ parts of them to solve the optimization problem in (6). Figure 7 illustrates the influence of using various portions of sorted pairs on the final matching results. Once the best parts of these pairs are chosen, the resulting correspondence appears to be acceptable even if only one-third of the pairs are enrolled because the derived correspondences can be achieved indirectly the hint of the chosen sorted matching. This result implies that in case some scans cannot match a close enough target scan, their correspondences can also be established from all of the other scans, while the process simply prevents them from participating in the optimization of the deformation energy. Those scans can be filtered out with a threshold when ω_i^t is computed in cases where the correspondences for some of the scans are not acceptable.

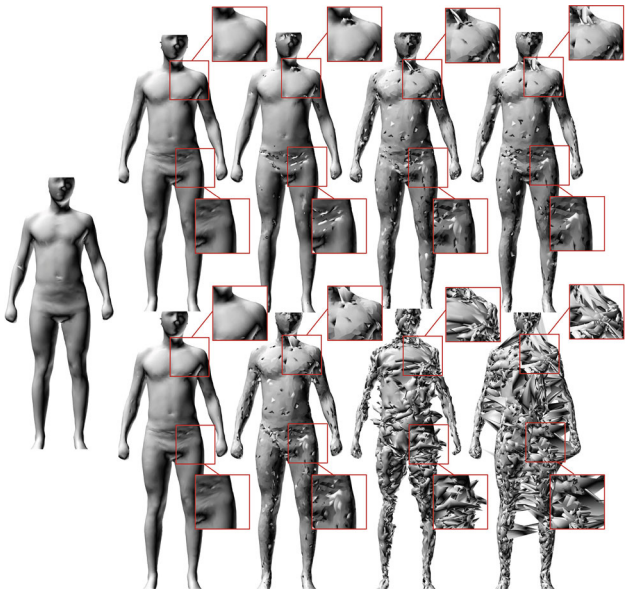


Fig. 7 The influence of using various portions of pairs to compute the correspondences. We examined the performance via two different strategies. One strategy is to choose the best portion (*top*), and the other strategy is to choose the worst portion. The first column is the target meshes. The results are derived by selecting various portions to determine the optimization in $[1\ \frac{1}{2}\ \frac{1}{3}\ \frac{1}{4}]$ (*from left to right*). The resulting correspondence appears to be acceptable even if only one-third of the best pairs are enrolled. In contrast, the performance degenerates quickly while choosing the worst matching pairs

3.3 Incomplete surface

The proposed approach, demonstrated in Sect. 3.2, can be utilized to merge multiple databases to learn the human body

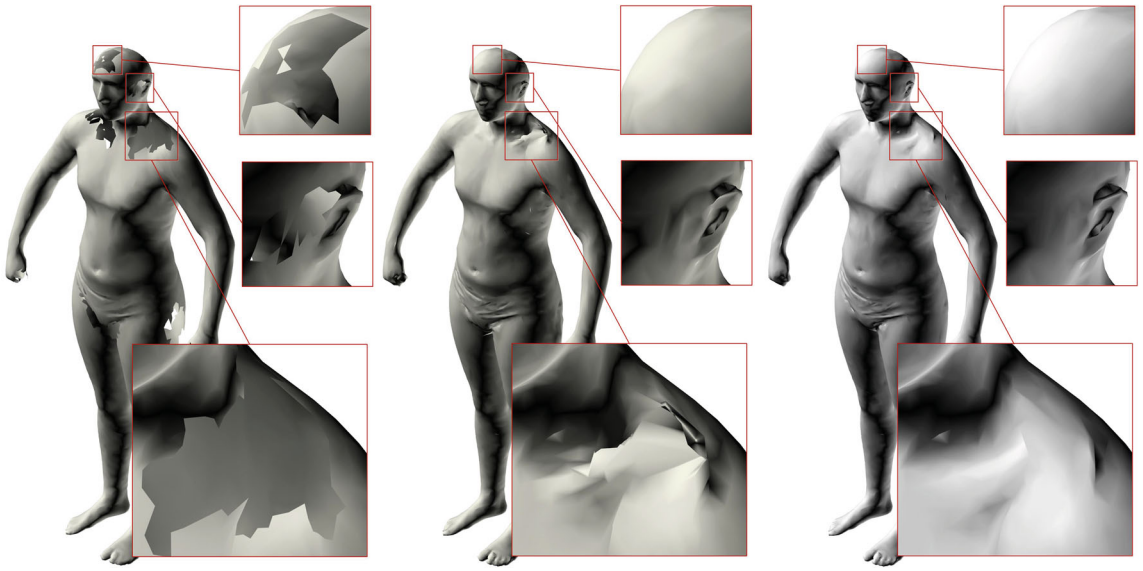


Fig. 8 Using the learned model to synthesize the missing parts. **a** The single input scan is not scanned well, and none of the missing parts can be extracted from other scans. Therefore, **b** one could fill the hole by deforming one of the scans in our database to fit the input surface by solving the deformation energy by ignoring the distance errors of the vertices near the holes. The deformation process not only achieves the correspondences but also fills the holes roughly with a slightly unattractive detail. **c** Once the correspondence is established, we use the learned model to synthesize polished and more attractive patches with derived approximate shape parameters for the input scan

model. However, some scans could be incomplete due to occlusions. In the most common cases, the scans are taken in different postures and, therefore, the missing part in some scans can be filled by making use of the part that exists in other scans via algorithms, such as the correlated correspondence algorithm [4]. In this case, the hole can be filled by the assumption that the missing part can be extracted from at least one of the scans in different postures. Once this assumption cannot be met, a hole-filling process is initiated. Instead of patching up the hole with the surface of the transformed template directly, we use the learned SCAPE model to generate the patches for the holes based on the partial information of the input scan. Figure 8 demonstrates the resulting meshes.

Recall that the SCAPE model parameterizes the variations in the body shapes, varying across different individuals and assuming that the body-shape variation is independent of the pose variation. The model introduces a linear transformation matrix S_k^i for each triangle k in instance i and learns the matrix by solving a least squares problem:

$$\begin{aligned}\underset{S^i}{\operatorname{argmin}} &\sum_k \sum_{j=2,3} \|R_k^i S_k^i Q_k^i \hat{v}_{k,j} - v_{k,j}^i\|^2 \\ &+ w_s \sum_{k_1,k_2 \text{ adj}} \|S_{k_1}^i - S_{k_2}^i\|^2. \end{aligned} \tag{8}$$

The details can be found in [23], and the derived S^i is assumed to be generated from a simple linear subspace, which is estimated using PCA:

$$S^i = \mathcal{S}_{U,\mu}(\beta^i) = \overline{U\beta^i + \mu} \tag{9}$$

which indicates that each shape of an individual is coded into a column vector β^i.

Inspired by the SCAPE model, to patch up the missing part, we solve the deformation energy in (1) to establish the correspondence between our model surface and the input incomplete scan in addition to ignoring the distance error of the vertices near the holes. After the correspondences are established, we derive the transformation matrices S^i by solving the optimization in (8), although we only can obtain the transformation matrices $\hat{S}^i$ for the triangles, which can be found in the input scan, due to the absent vertices of the missing part. An approximate PCA vector $\hat{\beta}^i$ is then obtained by solving:

$$\underset{\hat{\beta}^i}{\operatorname{argmin}} \, ||\overline{U\hat{\beta}^i + \mu} - \hat{S}^i||_F^2 \tag{10}$$

where U and μ are the learned parameters in the SCAPE model. Having $\hat{\beta}^i$, the hole-filled shape model can be generated via the SCAPE model with $S^i = \overline{U\hat{\beta}^i + \mu}$.

4 Results and discussion

With the derived correspondences, the shape variations can be learned into a human body model, which is utilized to realize the human shape animation across various postures and shapes. In the remainder of this section, we show the

correctness and efficiency of the proposed method in learning the shape model, synthesizing existent or nonexistent shapes into various poses and parameterizing shape variations. We also investigate how the advantages of learning model form a merged database and compare the proposed method with existing non-rigid registration methods.

4.1 Shape synthesis

To animate the human model, we embed an articulated skeleton into the model. This kinematic skeleton is composed of 16 rigid parts, and each part is associated with a subset of meshes in the model. The posture of the skeleton is synthesized via 51 parameters, which denote the global location, the global orientation and the rotation angles at each joint. Once we hope to drive the model to render a figure in a certain posture, we control the skeleton of the model by calculating the rigid transformation of each rigid part via updating the skeleton parameters. Sequentially, the associated meshes of each rigid part are transformed to form the surface into the desired shapes and postures. Figure 9 illustrates the rendered shapes of 4 individuals, which were learned from the merged database, into a posture that was given arbitrarily.

The learned model can also be used to synthesize a nonexistent shape from two or more subjects in the training datasets. Figure 10 demonstrates this application. To accomplish this goal, we generate the shapes with shape parameters of different individuals and linearly combine those shape

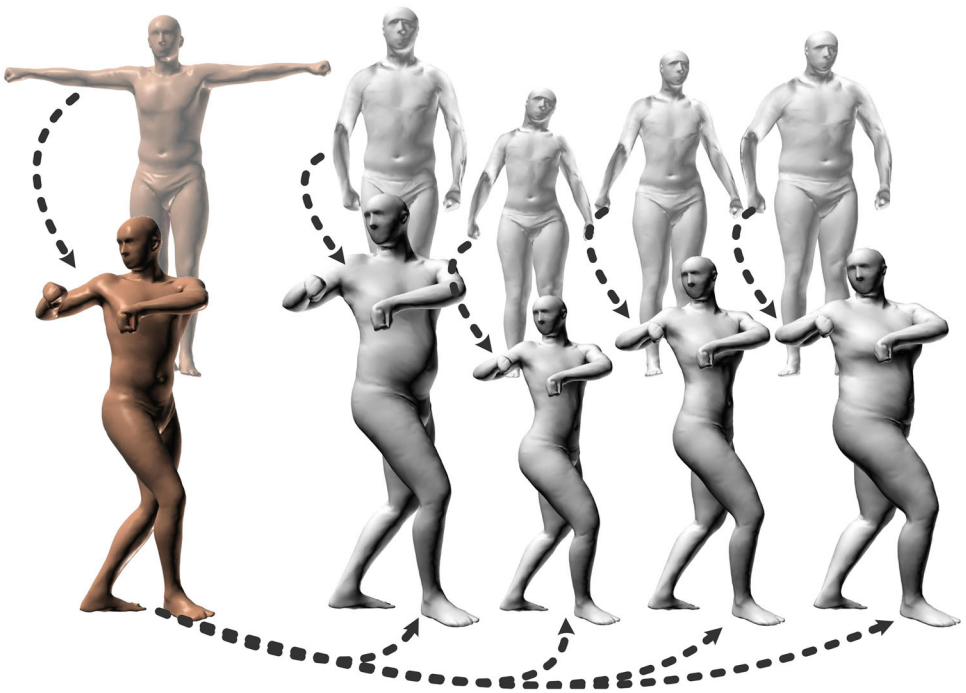


Fig. 9 Rendering a shape of arbitrary shape and posture. To render a shape for a certain individual in an arbitrary posture, we control the kinematic parameters of the skeleton to drive the template into the desired posture (*colored*). We integrate the rendered posture and the shape variations, which is learned from 4 individuals in this figure, to synthesize the final surface (*grey*) for those 4 individuals in the given posture

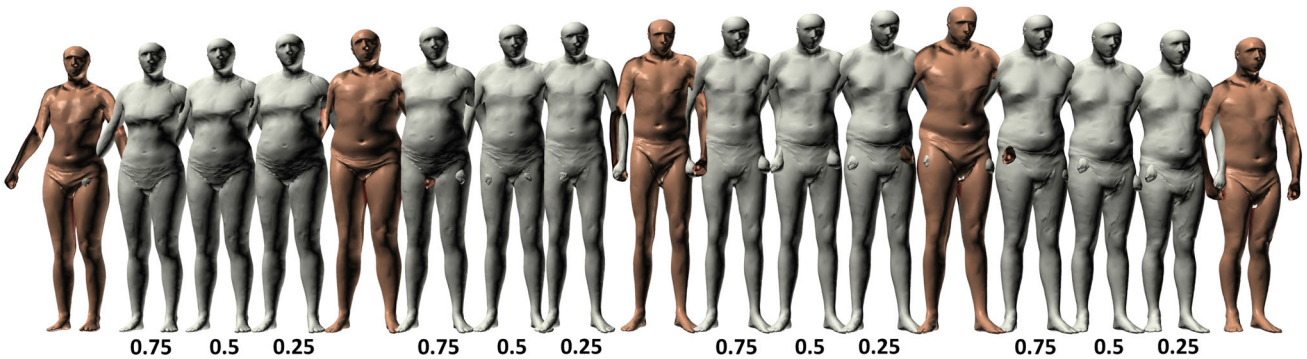


Fig. 10 Synthesizing shapes from existing scans. To demonstrate this application, we generate shapes (*colored*) of individuals in the training database. With these shapes, a series of synthetical shapes (*grey*) are rendered via linearly combining those shapes. Between each pair of colored shapes in the figure, the weight factor, used in the linear combination, increasingly varies from 0.75 to 0.25 (*from left to right*)

vertices. These results also imply that the correspondence is established correctly; thus, it can be used to align the scans from multiple databases, even for the meshes of the bottoms of the breasts and the waistline; otherwise, features will cross-fade instead of moving.

4.2 Comparison with the model learned from a single database

To show the advantages of learning a human shape model from a merged database, we also learned a model from a single database for comparison. Because the SCAPE datasets open only the scans across different postures, they cannot be applied to learn a complete SCAPE model. Nevertheless, we utilized the MPI database to train a human model. Figure 11 illustrates the resulting comparison. One of the drawbacks of the MPI database is its low scan quality. Thus, there is some loss of detail in the rendered shapes that were learned from MPI database. Because the number of samples used to train the model is a slightly small in the MPI database, the derived shapes could be drastic for some postures. In contrast, we merged the SCAPE and MPI databases via building the correspondences between these two datasets. Each scan in the merged database is composed of 12, 500 vertices and 25, 000 faces, which is the same as the scans in the SCAPE database. Therefore, we can model the pose and shape variations without loss of detail. Moreover, the pose variations can be modeled more accurately because it is of great benefit to improve the diversity of the training databases via learning the model from multiple databases.

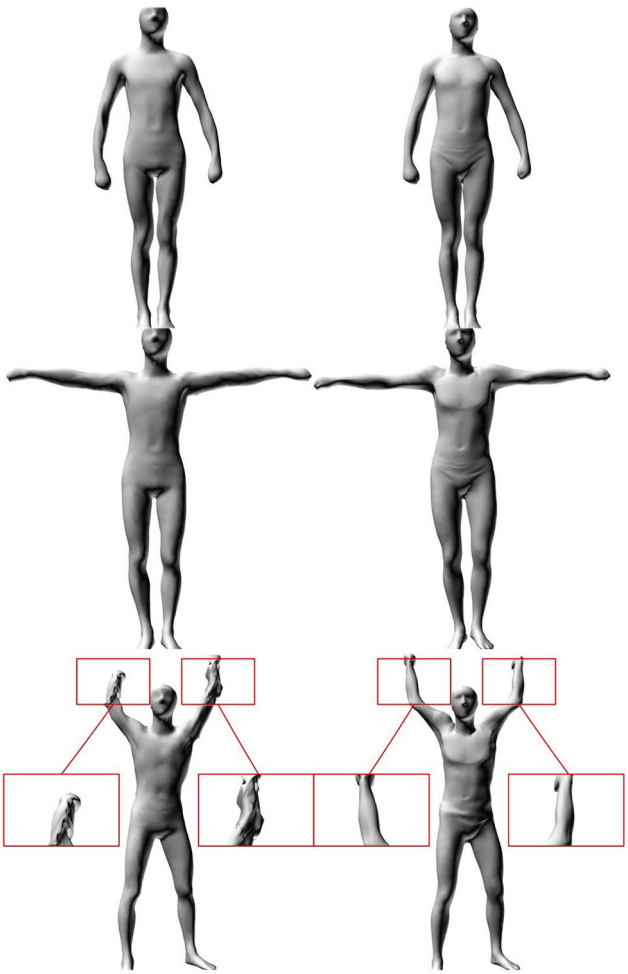


Fig. 11 Comparing the learned model using merged datasets (*right*) with the MPI database (*left*). The results show that the model using merged datasets provides rendered shapes with higher quality, and moreover, the variation due to pose changes can be modeled more accurately because it is of great benefit to improve the diversity of training databases via learning the model from multiple databases

4.3 Comparison with existing methods

Figure 12 demonstrates the performance comparison of our method and two non-rigid pair-wise shape matching approaches [10,18]. For each approach, we examine five test cases to demonstrate how effectively the approach can be performed on scans of different pose and shape variations. The Isometric Deformation approach [10] utilizes the principal curvatures to match the source and target vertices, which will be further pruned and propagated according to their pair-wise geodesic distances. This approach works well once the source mesh is very similar to the target mesh. In the most common cases, however, the source and target from different datasets are usually scanned from distinct individuals, which leads to unmatched principal curvatures and pair-wise geodesic distances between the source and target. These unmatched metrics drive the resulting meshes to form incorrect correspondences. Instead of using those shape-related descriptors such as the principal curvatures used in [10], Li's approach [18] utilizes the confidence weights to measure the reliability of each correspondence and identifies non-overlapping and overlapping areas. It performs robust even when there is a large difference between the source and target scans. However, it fail to establish the correspondences for those pairs of scans that have large pose variations. The optimization-based frameworks usually become stuck at a local minimum while the pose of the source scan is far away from the target scan because of closest points that are used in each iteration. In some extreme cases as illustrated in Fig. 13, such as when the deformation factors and markers are selected carelessly, these approaches could lead to drastic variation.

Our approach achieves similar results with the Isometric Deformation and Li's approach, when the source and target meshes are much similar in their postures. For those scans that have large pose and shape variations, moreover, our approach will consider the consensus among all of the scans in the datasets to achieve the correspondences, which

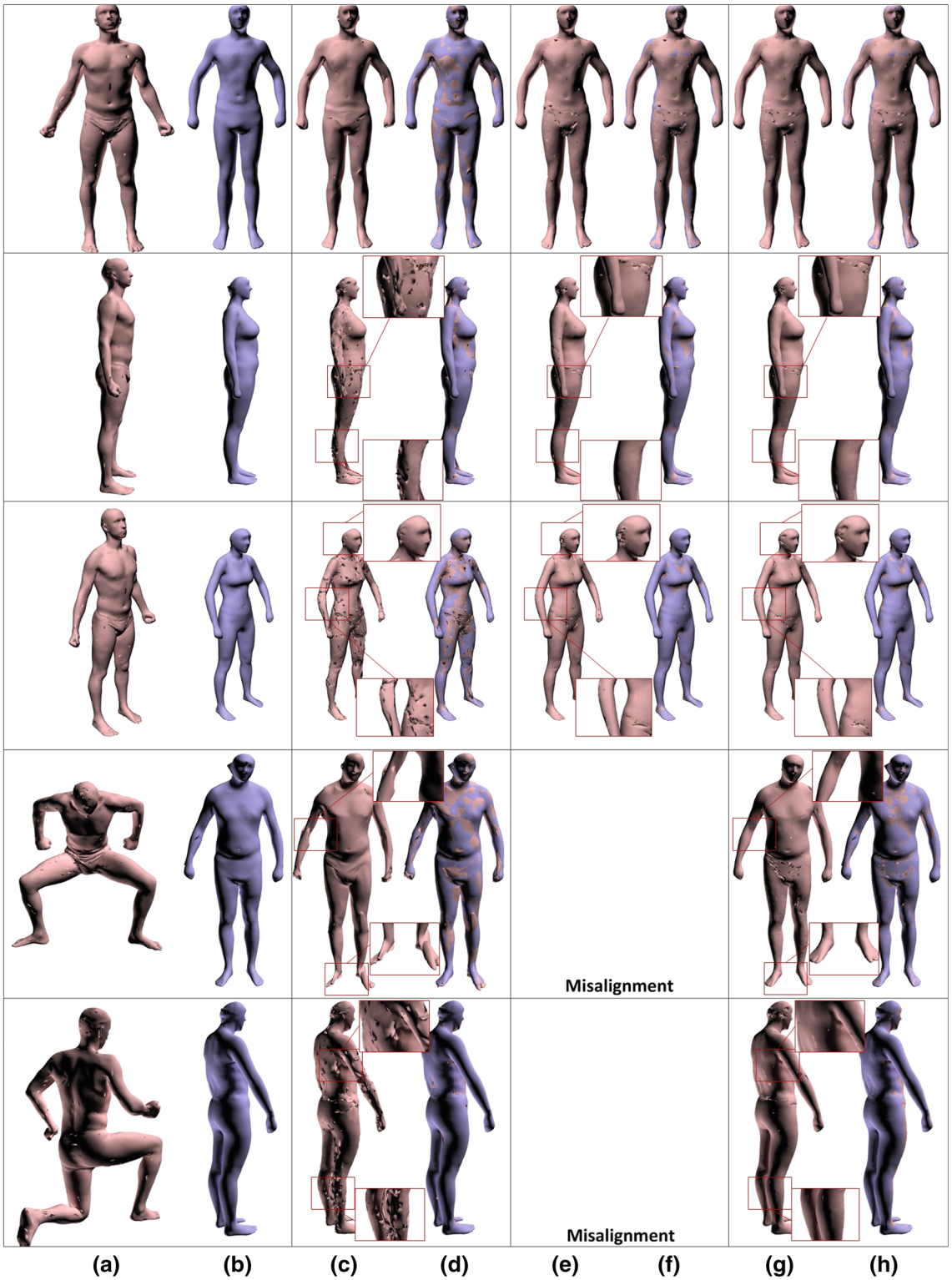


Fig. 12 We compare the performance of our method with two non-rigid methods. For each algorithm, we examine five test cases. *Columns a* and *b* demonstrate the source scans and target scans, respectively. *Columns c* and *d* show the resulting meshes and the difference between the resulting and target meshes, using Isometric Deformation [10]. *Columns e* and *f* are results from Li's approach [18], while columns *g* and *h* are given by our approach

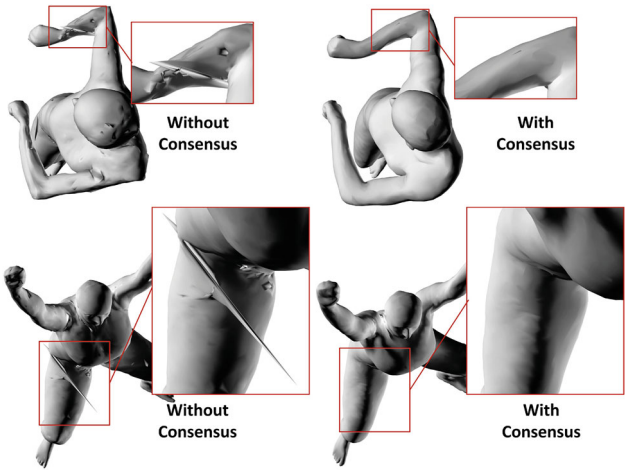


Fig. 13 In some extreme cases, such as when the deformation factors and markers are selected carelessly, the deformation-based correspondence algorithm could lead to drastic variation; however, our approach performs well in the same configuration

can fit the source scans to the target scans, while minimizing the energy of deformation with the kinematic consensus.

Another advantage of the proposed method is to achieve a set of homogeneous correspondences. For example, to parameterize the shape variations, SCAPE model utilizes PCA to extract the shape parameters from those shape variation matrices S. As an approximation of the shape variations, a higher PCA dimension leads to a much better approximation of the shape variations. Ideally, the transformation matrix should be affected only by the shape variations of each individual. However, because the matrices are derived by evaluating the triangle transformation based on the correspondences, the matrices are also affected by the established correspondences. To model two scans of different postures for a certain individual, for instance, the transformation matrices of such scans are denoted as S_1 and S_2 respectively. Ideally, S_1 should be equal to S_2, because the scans are from the same person. However, because of the pose variation, the correspondences between the SCAPE template and those scans could be different and the derived S_1 and S_2 are, thereby, not equal. Thus, to model the shape variation, a larger length is required for the PCA parameters to approach the shape variation. With the kinematic consensus, our approach can achieve a homogeneous correspondence for learning human model. As in the example shown in Figs. 14 and 15, our approach with the kinematic consensus can yield a much better resulting shape than the approach without the kinematic consensus, when the dimension of the PCA space is increasingly decreased. This finding implies that our approach is more suitable for model-based motion capture systems, which usually must optimize the motion and shape parameters jointly. Because this optimization is time-consuming, in these applications, a smaller shape parameters space can lead to a faster implementation in the phase for shape estimation. The results in Fig. 14 show

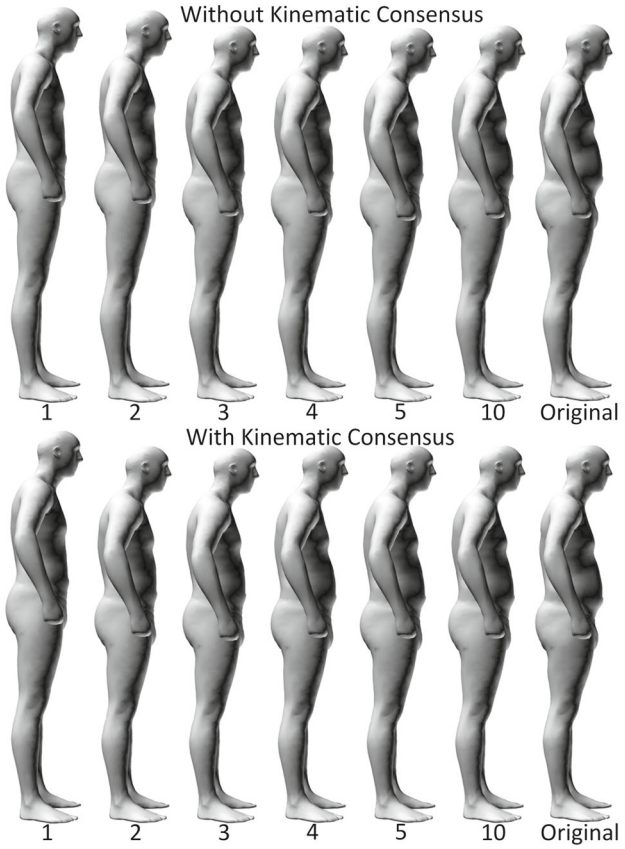


Fig. 14 The rendered shapes with different PCA dimensions. As the dimension increases, the model renders a shape that is more similar to the original one. With the kinematic consensus, a homogeneous correspondence is accomplished, and the PCA space can be narrowed substantially. The learned human model yields a much better resulting shape than that yielded by the approach without the kinematic consensus

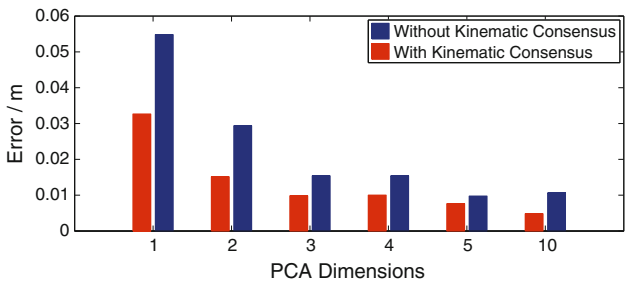


Fig. 15 The error between the rendered and ground-truth shapes. It is shown that our approach provides more faithful details and less error than the approach without considering the kinematic consensus

that our approach provides more faithful details, and less error than the approach without considering the kinematic consensus.

The physical meaning of those parameters in the PCA space is indistinct to understand. For example, it is difficult to configure one or more PCA parameters to control the intuitive properties of a shape, such as height, weight, leg length, etc. To achieve this goal, we developed a method to model

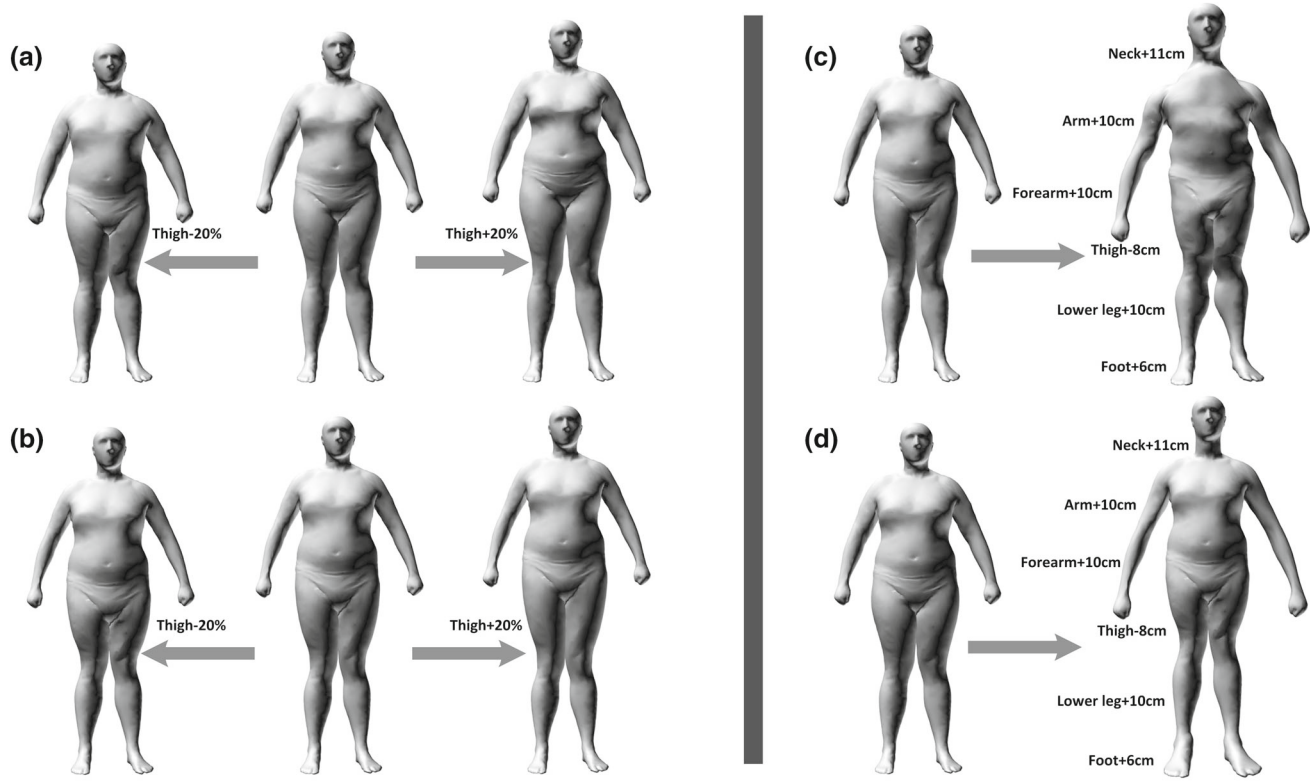


Fig. 16 Resulting shapes from configuring the shape parameters in an intuitive way. We changed the length of the thigh to synthesize two shapes. **a** Without limiting the impact of the shape parameters on the rigid parts of the human bodies, the changes in the thigh length could vary the meshes of the chest due to overfitting, and **c** with the intuitive parameters, an undesirable result could be yielded. However, in our method, **b** the overfitting can be avoided, and **d** the shape changes as the configurations of those intuitive parameters vary, which leads to a predictable result

the shape parameters in an intuitive way. Traditional methods to learn the shape parameters are to optimize the cost function:

$$\Psi = \underset{\Psi}{\operatorname{argmin}} \sum_i ||\beta^i - \Psi\theta^i||^2 \tag{11}$$

in which θ^i denotes the shape parameters for an individual i, and Ψ is the linear transformation matrix between the shape parameters and the PCA parameters. In our implementation, instead of learning the relationship between the intuitive parameters and the PCA parameters, we directly learn how those intuitive parameters impact on the elements of the transformation matrices S_k^i for shape variations in a linear manner. Moreover, to avoid the overfitting of the mapping, we constrain that a rigid body part is affected by a subset of those intuitive parameters:

$$\Psi^k = \underset{\Psi_k}{\operatorname{argmin}} \sum_i ||S_k^i - \overline{\Psi_k \Gamma_k(\theta^i)}||^2 \tag{12}$$

where $\Gamma_k(\cdot)$ is a mapping that indicates which shape parameters will affect the triangle k. For example, the leg length should not have an influence on the shape variation of the arm, and the waist width should not affect how to render the meshes for the head. Ψ^k is a matrix that shows how the mapped shape parameters form the shape transformation matrix S_k^i, which is learned for all of the triangles from the whole datasets. Figure 16 shows that, without the constraint being introduced, a small change in the length of the thigh can result in varying most parts of the body shapes. As illustrated in Fig. 16, given the intuitive parameters, an undesirable result could be yielded; while in our method, the shapes change according to how we configure those parameters and lead to a commendable and predictable result.

5 Conclusions

This paper presents a framework for learning a human body model from multiple databases. We proposed an approach to establish the correspondences while considering a kinematic consensus to learn the shape variations across individuals and postures. Our proposed approach incorporates the kinematic properties into the deformation process to compute the correspondence; therefore, the established correspondence is not only accurate, robust, and capable of handling scan incompletion but also homogeneous across shapes and postures. Those

properties make it possible to generate a faithful correspondence even in extreme cases, such as when the deformation factors and markers are selected carelessly. We also develop a method for controlling the rendered shapes in an intuitive way, with which a commendable and predictable result can be synthesized.

Acknowledgments This work was partially supported by Grant No. BE2011169, BK2011563 from the Natural Science Foundation of Jiangsu Province and Grant No. 61100111, 61300157, 61201425, 61271231 from the Natural Science Foundation of China.

References

1. Aiger, D., Mitra, N., Cohen-Or, D.: 4-points congruent sets for robust pairwise surface registration. In: ACM Transactions on Graphics (TOG), vol. 27, p. 85. ACM, New York (2008)
2. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. In: ACM Transactions on Graphics (TOG), vol. 22, pp. 587–594. ACM, New York (2003)
3. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. In: ACM Transactions on Graphics (TOG), vol. 24, pp. 408–416. ACM, New York (2005)
4. Anguelov, D., Srinivasan, P., Pang, H., Koller, D., Thrun, S., Davis, J.: The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces. Adv. Neural Inf. Process. Syst. **17**, 33–40 (2005)
5. Ben Azouz, Z., Shu, C., Mantel, A.: Automatic locating of anthropometric landmarks on 3D human models. (2006)
6. Chang, W., Zwicker, M.: Automatic registration for articulated shapes. In: Computer Graphics Forum, vol. 27, pp. 1459–1468. Wiley Online Library, New York (2008)
7. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
8. Gelfand, N., Mitra, N., Guibas, L., Pottmann, H.: Robust global registration. In: Proceedings of the Third Eurographics Symposium on Geometry Processing, pp. 197–206. (2005)
9. Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., Seidel, H.P.: A statistical model of human pose and body shape. In: Dutr'e, P., Stamminger, M. (eds.) Computer Graphics Forum (Proceedings of Eurographics 2008), vol. 2. Munich, Germany (2009)
10. Huang, Q., Adams, B., Wicke, M., Guibas, L.: Non-rigid registration under isometric deformations. In: Computer Graphics Forum, vol. 27, pp. 1449–1457. Wiley Online Library, New York (2008)
11. Irani, S., Raghavan, P.: Combinatorial and experimental results for randomized point matching algorithms. Comput. Geom. **12**(1–2), 17–31 (1999)
12. James, D., Twigg, C.: Skinning mesh animations. In: ACM Transactions on Graphics (TOG), vol. 24, pp. 399–407. ACM, New York (2005)
13. Jolliffe, I.: Principal Component Analysis. Wiley Online Library, New York (2005)
14. Kalogerakis, E., Hertzmann, A., Singh, K.: Learning 3D mesh segmentation and labeling. ACM Trans. Graph. **29**(4), 102 (2010)
15. Kraevoy, V., Sheffer, A.: Cross-parameterization and compatible remeshing of 3d models. In: ACM Transactions on Graphics (TOG), vol. 23, pp. 861–869. ACM, New York (2004)
16. Kry, P., James, D., Pai, D.: Eigenskin: real time large deformation character skinning in hardware. In: Symposium on Computer Animation: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, vol. 21, pp. 153–159 (2002)
17. Lewis, J., Cordner, M., Fong, N.: Pose space deformation: an unified approach to shape interpolation and skeleton-driven deformation. In: Proceedings of the 27th Annual Conference on Computer Graphics And Interactive Techniques, pp. 165–172. ACM Press/Addison-Wesley Publishing Co., New York (2000)
18. Li, H., Sumner, R.W., Pauly, M.: Global correspondence optimization for non-rigid registration of depth scans. In: Computer Graphics Forum, vol. 27, pp. 1421–1430. Wiley Online Library, New York (2008)
19. Lipman, Y., Sorkine, O., Cohen-Or, D., Levin, D., Rossi, C., Seidel, H.: Differential coordinates for interactive mesh editing. In: Shape Modeling Applications, 2004. Proceedings, pp. 181–190. IEEE, New York (2004)
20. Pekelny, Y., Gotsman, C.: Articulated object reconstruction and markerless motion capture from depth video. In: Computer Graphics Forum, vol. 27, pp. 399–408. Wiley Online Library, New York (2008)
21. Rusinkiewicz, S., Levoy, M.: Efficient variants of the icp algorithm. In: Third International Conference on 3-D Digital Imaging and Modeling, 2001. Proceedings, pp. 145–152. IEEE, New York (2001)
22. Shalom, S., Shapira, L., Shamir, A., Cohen-Or, D.: Part analogies in sets of objects. In: Eurographics Workshop on 3D Object Retrieval. (2008)
23. Straka, M., Hauswiesner, S., Ruther, M., Bischof, H.: Rapid skin: estimating the 3D human pose and shape in real-time. In: 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), pp. 41–48. IEEE, New York (2012)
24. Sumner, R., Popović, J.: Deformation transfer for triangle meshes. In: ACM Transactions on Graphics (TOG), vol. 23, pp. 399–405. ACM, New York (2004)
25. Xu, K., Li, H., Zhang, H., Cohen-Or, D., Xiong, Y., Cheng, Z.: Style-content separation by anisotropic part scales. In: ACM Transactions on Graphics (TOG), vol. 29, p. 184. ACM, New York (2010)
26. Ye, M., Wang, X., Yang, R., Ren, L., Pollefeys, M.: Accurate 3d pose estimation from a single depth image. In: IEEE International Conference on Computer Vision (ICCV), 2011, pp. 731–738. IEEE, New York (2011)
27. Zhang, H., Sheffer, A., Cohen-Or, D., Zhou, Q., Van Kaick, O., Tagliasacchi, A.: Deformation-driven shape correspondence. In: Computer Graphics Forum, vol. 27, pp. 1431–1439. Wiley Online Library, New York (2008)
28. Zheng, Q., Sharf, A., Tagliasacchi, A., Chen, B., Zhang, H., Sheffer, A., Cohen-Or, D.: Consensus skeleton for non-rigid space-time registration. In: Computer Graphics Forum, vol. 29, pp. 635–644. Wiley Online Library, New York (2010)
29. Zhou, S., Fu, H., Liu, L., Cohen-Or, D., Han, X.: Parametric reshaping of human bodies in images. ACM Trans. Graph. **29**(4), 126 (2010)

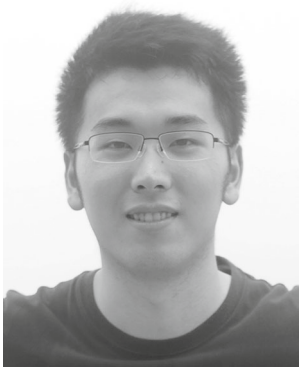
Yao Yu is currently a lecturer in School of Electronic Science and Engineering, Nanjing University, China. His research interests include computer graphics, 3D geometric modeling and computer vision. He received his B.E. and Ph.D. degree from Nanjing University, China in 2005 and 2010, respectively. In the past, he has worked as a Research Assistant at department of Electronic and Computer Engineering, Hong Kong University of Science & Technology, Hong Kong between 2007 and 2008. He is a member of the IEEE.

Yu Zhou received his M.E. degree in circuits and systems from Nanjing University, China in 2005, and the Ph.D. degree in signal and information processing from Nanjing University, China in 2008. He is an associate professor in School of Electronic Science and Engineering, Nanjing University, China. His research interests include computer vision and computer graphics. In 2012 and 2013, he worked as a visiting scholar in University of Kentucky, United States.

Sidan Du received the BS and MS degrees in electronic engineering from XiDian University, Xi'an, China in 1984 and 1987, respectively, and the Ph.D. degree in physics from Nanjing University, Nanjing, China, in 1997. She is currently a Professor in the School of Electronic Science and Engineering, Nanjing University. Her research interests include in digital imaging processing and computer vision.

Yuan Jie is an associate professor at Nanjing University in School of Electronic Science and Engineering. He got his bachelor in electronic science on 1997, master degree in signal process on 2000 and Ph.D. degree on 2003 at Nanjing University. His current research interest area includes Volumetric Display, Image and Video Process. In 2012 and 2013, he worked as a visiting scholar in Department of Radiology, University of Michigan, United States. He is a member of the IEEE.

Ziqiang Wang is an Associate Professor in School of Electronic Science and Engineering, Nanjing University, China. He earned his bachelor in electronic science from Nanjing University, China, in 1997. He is an experienced embedded system developer and familiar with parallel computing. His current research interest area includes parallel computing and optimization, signal processing.



Zhengyu Cai is currently a BS student in the School of Electronic Science and Engineering at Nanjing University, China. His research interests include digital image processing and human–computer interaction. He is now working as a research assistant in Vision-based Sensor and Graphics lab at Nanjing University. Besides his research, he has participated in the blue pathway internship of IBM and worked as a technical intern at IBM—Shenzhen.