

Simultaneous Registration and Modeling of Deformable Shapes

Jing Xiao*
Epson Palo Alto Lab, CA

Bogdan Georgescu Xiang Zhou Dorin Comaniciu
Siemens Corporate Research, NJ

Takeo Kanade
Robotics Institute, CMU, PA

Abstract

Many natural objects vary the shapes as linear combinations of certain bases. The measurement of such deformable shapes is coupling of rigid similarity transformations between the objects and the measuring systems and non-rigid deformations controlled by the linear bases. Thus registration and modeling of deformable shapes are coupled problems, where registration is to compute the rigid transformations and modeling is to construct the linear bases. The previous methods [3, 2] separate the solution into two steps. The first step registers the measurements regarding the shapes as rigid and the deformations as random noise. The second step constructs the linear model using the registered shapes. Since the deformable shapes do not vary randomly but are constrained by the underlying model, such separate steps result in registration biased by non-rigid deformations and shape models involving improper rigid transformations. We for the first time present this bias problem and formulate that, the coupled registration and modeling problems are essentially a single factorization problem and thus require a simultaneous solution. We then propose the Direct Factorization method that extends a structure from motion method [16]. It yields a linear closed-form solution that simultaneously registers the deformable shapes at arbitrary dimensions ($2D \rightarrow 2D$, $3D \rightarrow 3D$, \dots) and constructs the linear bases. The accuracy and robustness of the proposed approach are demonstrated quantitatively on synthetic data and qualitatively on real shapes.

1 Introduction

The inter- and intra-variability are often inherent in biological or biomedical object shapes or dynamic scene structures. For example, the shapes of human faces involve both inter-variability across individual persons of different ages, genders, and races, and intra-variability across different expressions. Similar variabilities also occur in objects such as beating hearts, growing tumors, and dynamic scenes includ-

*This work was conducted while the author was at Siemens Corporate Research. Please contact with xiaoj@erd.epson.com for further discussion.

ing moving cars or pedestrians. With the expeditious development of computer and sensor technologies, statistical modeling of such deformable shapes has shown enormous importance for many tasks in computer vision and medical image interpretation, e.g. segmenting the 2D MRI brain images [14], tracking the myocardial wall motion from the 3D ultrasound images [2, 9], and tracking and recognizing human faces from 2D videos [3, 4, 5].

Each of the statistical models is constituted by a number of bases, of which the linear combinations comprise the deformable shapes. Constructing such a model usually requires a collection of pose-free shapes that involve purely the non-rigid deformations. Since the sensors measure the shapes with respect to respective local coordinate systems, e.g. from different view directions, the measured shapes are composed of both the non-rigid deformations controlled by the linear bases and the rigid similarity transformations relative to the local coordinate systems. Constructing the model from the measured shapes requires that the measurements are registered into a common coordinate system such that the rigid pose variations are eliminated. On the other hand, the statistical parameterization of the shape models provides global constraints that allow the shapes to vary only in a limited number of styles. Fitting the models to the measurements has achieved robust performance in registering deformable shapes [3, 2]. Thus registration and modeling of deformable shapes are coupled as chicken-egg problems.

The previous methods [3, 14, 9, 2] proposed to solve the coupled problems separately in two consecutive steps. The first step registers the measured shapes, where Generalized Procrustes Analysis [7, 8, 10, 11, 15] has been widely used. This technique treats the deformable shapes as a rigid shape contaminated by Gaussian noise, and iteratively registers the individual shapes with a common reference shape that is the mean of the already-registered shapes. The iteration stops when a stable mean shape is achieved. The second step then extracts the linear deformable model from the registered shapes using Principal Component Analysis.

Generalized Procrustes Analysis is a perfect fit for registering rigid shapes, of which the variations come from either the similarity transformations or the noise. However it results in registration of deformable shapes biased by the

non-rigid deformations, because the deformable shapes do not vary randomly but the deformations are constrained by the underlying linear model. Specifically, when the deformations are insignificant, *i.e.* the mean shape is dominant compared with the deformations, or when they are symmetric with respect to the mean shape, *i.e.* the positive and negative offsets cancel each other, they can be regarded as Gaussian noise and Generalized Procrustes Analysis works successfully on registering the deformable shapes. In practice the shape deformations are often significant and asymmetric, *e.g.* a dynamic scene consisting of static buildings and moving cars, or two eyes blinking asynchronously. Under such situations, the deformations cannot be regarded as Gaussian noise and the linear deformable model has to be taken into account. By attenuating the deformations as denoising, Generalized Procrustes Analysis leads to the registered shapes still containing rigid similarity transformations. Thus the registration is incomplete and consequently the linear model constructed from the registered shapes does not describe purely the shape deformations.

This paper for the first time points out this bias problem and formulates that, registration and modeling of deformable shapes are essentially a single factorization problem and thus have to be solved simultaneously. The similar factorization problem has attracted many interests recently for solving the problem of 3D reconstruction of non-rigid structures from 2D images [1, 16]. This paper extends the method in [16] and develops the Direct Factorization method that simultaneously registers the measured shapes and extracts the underlying model. Following the idea in [16], this method enforces linear constraints on orthonormality of the rigid rotations and on uniqueness of the linear bases and yields a linear closed-form solution. However our method is different from the method in [16]. The presented method aims at registering deformable shapes generally at arbitrary dimensions ($2D \rightarrow 2D$, $3D \rightarrow 3D$, \dots), while the method in [16] is for structure from motion ($2D \rightarrow 3D$ only). Note that besides presenting the Direct Factorization method, the more important contributions of this paper are to point out the bias problem of the separate two-step solution and to propose the simultaneous factorization formulation for the coupled registration and modeling problems.

2 The Coupled Registration and Modeling Problems

This paper regards the shape of a deformable object as a linear combination of K shape bases $\{B_k, k = 1, \dots, K\}$. Each basis is a $D \times P$ matrix, with the same dimension as the shape. They direct the P points on the shape to alter the locations in the D -dimensional object shape space. Accordingly the coordinate of the shape at time i is,

$$S_i = \sum_{k=1}^K l_{ik} B_k \quad (1)$$

where l_{ik} is the combination weight of the basis B_k . The deformable shape is often measured from different views, in different distances, and at different scales, *i.e.* with respect to different coordinate systems. Thus the measured shape is a similarity transformation of the shape S_i ,

$$W_i = c_i R_i S_i + T_i \cdot \mathbf{1} \quad (2)$$

where c_i is a nonzero scalar, R_i is a $D \times D$ orthonormal matrix, and T_i is a $D \times 1$ vector. They respectively stand for the scaling, rotation, and translation transformations, and together form the similarity transformation. $\mathbf{1}$ is a $1 \times P$ vector, of which all elements are 1s.

Given a collection of N observations of such deformable shapes, we would like to register them to recover the similarity transformations (c_i , R_i , and T_i) with respect to the constraint in Eq. (1), and extract from them the underlying bases B_k and coefficients l_{ik} with respect to the constraint in Eq. (2). Therefore registration and modeling of deformable shapes are coupled chicken-egg problems.

3 The Bias Problem of the Separate Two-Step Method

The previous solutions to the coupled registration and modeling problems generally consist of two separate steps [3, 14, 9, 2]. The first step registers the deformable shape measurements using Generalized Procrustes Analysis (GPA) [7, 15] and the second step computes the linear bases from the registered shapes using Principal Component Analysis (PCA).

GPA was originally developed for registering rigid shapes contaminated with Gaussian noise. When the shapes are deformable, GPA regards the non-rigid deformations as part of the noise and estimates the similarity transformations by minimizing the following objective function,

$$(\tilde{c}, \tilde{R}, \tilde{T}) = \underset{c, R, T}{\operatorname{argmin}} \sum_{i=1}^N \|c_i R_i S_i + T_i - S_C\|^2 \quad (3)$$

where S_C stands for the centroid shape, *i.e.* average of the aligned shapes in the previous iteration. The optimization is achieved using an iterative algorithm as follows,

1. Specify one of the shapes as the initial centroid shape S_C .
2. Estimate the similarity transformation between each shape and S_C using Orthogonal Procrustes Analysis (OPA) [13].
3. Align the shapes using the estimated similarity transformations and update S_C as the average of the registered shapes.

The last two steps are repeated until a least square solution is converged. The least square solution will be biased from the true answer when the noise is not symmetric, *i.e.* the positive and negative variances between the noisy data and the ground truth do not cancel each other. Since the optimization process does not take into account the deformable model, *i.e.* Eq. (3) is not minimized with respect to the constraint in Eq. (1), the measured shapes are registered with

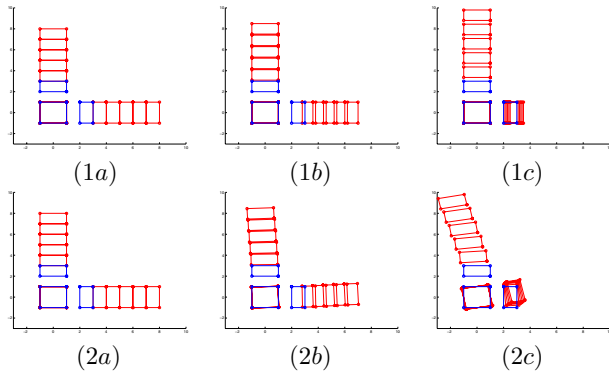


Figure 1. Three settings of one rectangle staying static and two rectangles moving along straight lines. (1a~1c) Ground truth shapes. (2a~2c) Registered shapes by Generalized Procrustes Analysis.

the rigid centroid shape and the deformations are regarded as noise. Thus when the deformations are asymmetric with respect to the centroid shape, the accuracy of the recovered similarity transformations will be “sacrificed” to minimize the registration error according to Eq. (3). In another word, even though GPA minimizes the registration errors, the recovered rigid similarity transformations are biased from the true answer such that the registered shapes are more rigid, *i.e.* contain fewer deformations, than the true deformable shapes. Consequentially the deformable model constructed using such registered shapes in the second step does not represent purely the desired deformable shapes.

We demonstrate the bias problem using simple 2D examples in three noiseless settings. In these examples, the 2D shape consisted of three rectangles, represented by the 12 corners. The middle rectangle was static. The top one moved up and the right one moved right simultaneously along straight lines. Fig.1.(1a) shows the first setting, where the two rectangles moved at the same speed, *i.e.* the shapes deformed symmetrically. The blue lines refer to the initial shape and the red ones stand for the shape deformations. 6 shapes were observed from different views. We applied GPA to estimate the transformations between the measured shapes. As shown in Fig.1.(2a), the registered shapes were exactly the same as the ground truth shapes with zero errors.

In the second example, the speed of moving up was slightly different from that of moving right, *i.e.* the deformations were slightly asymmetric, as shown in Fig.1.(1b). 6 measured shapes from different views were aligned by GPA. As shown in Fig.1.(2b), the registered shapes were slightly different from the ground truth. The average registration error on the shapes was 2.63%. The third setting contained strongly asymmetric shape deformations, *i.e.* the two speeds were very different, as shown in Fig.1.(1c). These shapes were observed as is, without any

similarity transformation. Fig.1.(2c) demonstrates that GPA yielded notable registration errors, of which the average was 12.18%. The aligned shapes were rotated and they did not move straightly, while the true shapes did not rotate and the true trajectories were straight lines. These experiments confirm that, GPA achieves biased registration of the non-rigid shapes that contain asymmetric deformations. The stronger the asymmetry, the greater the registration error.

4 Simultaneous Registration and Modeling

4.1 The Factorization Formulation

To address the bias problem, in this section we reformulate the coupled registration and modeling problems. We first rewrite Eq. (2) by replacing S_i using Eq. (1) and absorbing c_i into l_{ik} ,

$$W_i = (l_{i1}R_i \quad \dots \quad l_{iK}R_i \quad T_i) \cdot \begin{pmatrix} B_1 \\ \vdots \\ B_K \\ \mathbf{1} \end{pmatrix} \quad (4)$$

We stack all the N shape measurements into a $DN \times P$ matrix W . Each of the N D -rows in W contains a measured shape respectively. Due to Eq. (4), since all the measured shapes refer to the same set of shape bases, we have,

$$W = M \cdot B + T \cdot \mathbf{1} \quad (5)$$

where M is a $DN \times DK$ scaled rotation matrix, B is a $DK \times P$ basis matrix, and T is a $DN \times 1$ translation vector,

$$M = \begin{pmatrix} l_{11}R_1 & \dots & l_{1K}R_1 \\ \vdots & \vdots & \vdots \\ l_{N1}R_N & \dots & l_{NK}R_N \end{pmatrix} \quad (6)$$

$$B = \begin{pmatrix} B_1 \\ \vdots \\ B_K \end{pmatrix}, \quad T = \begin{pmatrix} T_1 \\ \vdots \\ T_N \end{pmatrix}$$

The coupled problems of deformable shape registration and modeling are now reformulated into a single factorization problem, *i.e.* how to factorize the measurement matrix W to simultaneously recover the rigid similarity transformations in M and T and reconstruct the linear bases in B .

4.2 The Direct Factorization Method

The similar factorization problem has been studied extensively for 3D reconstruction of non-rigid structures from 2D images [1, 16]. In this section we extend the method in [16] and develop a Direct Factorization method that simultaneously solves the registration and modeling problems.

Following the idea in [16], we position the coordinate origin at the center of the measured shape. Then the translations in T equal the average coordinates over the shape.

Subtracting them from the measurements, we have that $\tilde{W} = W - T \cdot \mathbf{1} = M \cdot B$. According to Eq. (6), the rank of \tilde{W} is $\min\{DK, DN, P\}$. In general the shape number N and point number P are much larger than the basis number K such that $DN > DK$ and $P > DK$. The rank of \tilde{W} is thus DK and K is determined by $K = \frac{\text{rank}(\tilde{W})}{D}$.

Using Singular Value Decomposition (SVD), we decompose \tilde{W} into the product of a $DN \times DK$ matrix \tilde{M} and a $DK \times P$ matrix \tilde{B} . This decomposition is only determined up to a non-singular $DK \times DK$ linear transformation G . The true scaled rotations M and bases B are of the forms,

$$M = \tilde{M} \cdot G, \quad B = G^{-1} \cdot \tilde{B} \quad (7)$$

G consists of K D -columns, denoted as $g_k, k = 1, \dots, K$. They are $DK \times D$ matrices and independent on each other since G is non-singular. According to Eq. (6,7), g_k satisfies,

$$\tilde{M}g_k = \begin{pmatrix} l_{1k}R_{11} \\ \vdots \\ l_{Nk}R_{N1} \end{pmatrix} \quad (8)$$

\tilde{M} consists of N D -rows, denoted as $\tilde{M}_i, i = 1, \dots, N$. Each of them is a $D \times DK$ matrix. Let Q_k denote the $DK \times DK$ symmetric matrix $g_k g_k^T$. Then,

$$\tilde{M}_i Q_k \tilde{M}_j^T = l_{ik} l_{jk} R_i R_j^T, \quad i, j = 1, \dots, N \quad (9)$$

As in [16], computing Q_k requires imposing two types of constraints. First, due to orthonormality of the rotations,

$$\tilde{M}_i Q_k \tilde{M}_i^T = l_{ik}^2 \mathbf{I}_{D \times D}, \quad i = 1, \dots, N \quad (10)$$

where $\mathbf{I}_{D \times D}$ is a $D \times D$ identity matrix. The D diagonal elements of Eq. (10) are equivalent and yield $D - 1$ linear constraints on Q_k . The off-diagonal elements are all zeros. Since Q_k is symmetric, Eq. (10) provides $\frac{D^2 - D}{2}$ linear constraints. From N measured shapes we obtain $\frac{N}{2}(D^2 + D - 2)$ linear constraints.

As shown in [16], since any non-singular linear transformation applied on the deformable bases leads to another set of eligible bases, enforcing the orthonormality constraints alone is inherently insufficient and results in ambiguous and invalid solutions. Thus it is necessary to enforce the second constraints that ensure the uniqueness of the bases. Similar to [16], we need to determine K measured shapes that contain independent deformable shapes, *i.e.* after registration the K shapes without rigid transformations are independent on each other. Such independence is measured by the condition number of the $DK \times P$ sub-matrix of \tilde{W} corresponding to the K shape measurements. A smaller condition number refers to stronger independence.

We compute the condition number of each possible set of K measured shapes and select the set with the smallest condition number, which contains the most independent K

deformable shapes. Since any K independent shapes in the linear shape space form a set of eligible bases, we specify the deformable shapes contained in the selected K measurements as the unique bases. In another word, the selected K measurements are the respective scaled and rotated shape bases. Since scaling does not influence the independence of the shapes, we absorb the scalars into the bases and then the chosen K basis measurements are simply the rotated bases. Note that so far we have not recovered the bases, but only determined their measurements after unknown rotations.

We denote the selected K measurements as the first K measurements in the data set, *i.e.* $\tilde{W}_i = R_i B_i, i = 1, \dots, K$. The corresponding coefficients are thus,

$$\begin{aligned} l_{ii} &= 1, \quad i = 1, \dots, K \\ l_{ij} &= 0, \quad i, j = 1, \dots, K, \quad i \neq j \end{aligned} \quad (11)$$

According to Eq. (9,11), we have,

$$\tilde{M}_i Q_k \tilde{M}_j^T = \begin{cases} \mathbf{I}_{D \times D}, & i = j = k \\ \mathbf{0}_{D \times D}, & (i, j) \in \Phi \end{cases} \quad (12)$$

where Φ means $\{(i, j) | i = 1, \dots, K; j = 1, \dots, N; i \neq k\}$.

Similar to the proof in [16], we prove that solving both the linear equations in Eq. (10) and (12) leads to a closed-form solution of $Q_k = g_k g_k^T, k = 1, \dots, K$. We then recover g_k by decomposing Q_k via SVD. The decomposition of Q_k is up to an arbitrary $D \times D$ orthonormal transformation Ψ , since $(g_k \Psi)(g_k \Psi)^T$ also equals Q_k . This ambiguity arises from the fact that g_1, \dots, g_K are estimated independently under different coordinate systems. To resolve the ambiguity, we need to transform g_1, \dots, g_K to be under a single reference coordinate system.

Due to Eq. (8), $\tilde{M}_i g_k = l_{ik} R_i, i = 1, \dots, N$. Because the rotation matrix R_i is orthonormal, *i.e.* $\|R_i\| = 1$, we have $R_i = \pm \frac{\tilde{M}_i g_k}{\|\tilde{M}_i g_k\|}$. The sign is determined by the orientation of the reference coordinate system. We compute K sets of rotations using g_1, \dots, g_K respectively. Because of the decomposition ambiguity, there is a $D \times D$ orthonormal transformation between each two sets. We specify one of the rotation sets as the reference. The signs of the other rotations are determined such that they are consistent with the corresponding reference rotations. The orthonormal transformations are computed by OPA [12] to transform the signed rotation sets to the reference set. They also transform g_1, \dots, g_K to be under a common coordinate system, *i.e.* the desired transformation G is achieved. The coefficients are then computed by Eq. (8), and the shape bases are recovered by Eq. (7). Their combinations reconstruct the true deformable shapes without rigid transformations.

Our algorithm is summarized as follows,

1. Compute the translation T as the average of the measurements W and then eliminate it by $\tilde{W} = W - T \cdot \mathbf{1}$.

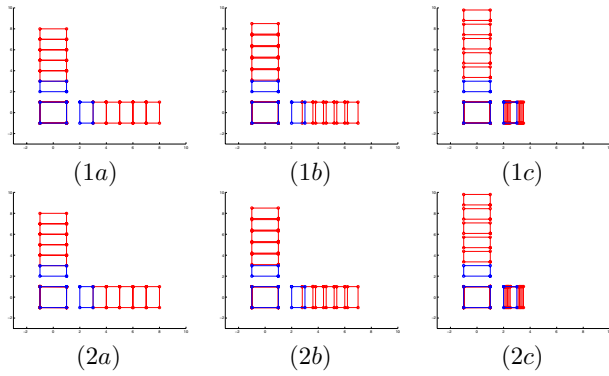


Figure 2. Three settings of one rectangle staying static and two rectangles moving along straight lines. (1a~1c) Ground truth shapes. (2a~2c) Registered shapes by the Direct Factorization method.

2. Perform SVD to determine the rank of \tilde{W} that dominates the energy. The basis number K is computed by $\text{rank}(\tilde{W})/D$.
3. Determine the K basis measurements, reorder \tilde{W} , and compute rank- DK approximation of $\tilde{W} = \tilde{M}\tilde{B}$ by SVD;
4. Compute $Q_k, k = 1, \dots, K$ by solving the linear equations in Eq. (10) and (12) via the least squares method.
5. Calculate $g_k, k = 1, \dots, K$ via SVD, then recover the rotations by Eq. (8) and transform them to a common coordinate system using OPA.
6. Recover the linear bases by Eq. (7) and the coefficients by Eq. (8). Their combinations reconstruct the registered deformable shapes.

It is not necessary to check all possible K shape measurements and specify the most independent deformable shapes as the bases. When the number of measurements is large, we can save the computational cost by only searching for K measurements with a condition number below certain small threshold. Note that when the basis number K is set as 1, the Direct Factorization method also regards the deformable shapes as a rigid shape contaminated with random noise and achieves the least square registration similar as GPA in the two-step method does. The only difference between their solutions is that GPA uses the mean shape and the Direct Factorization method computes the most dominant shape mode as the reference for registration. Thus the two-step method is a special case of the Direct Factorization method where the shapes are assumed rigid.

5 Performance Evaluation

The performance of the proposed method was evaluated in a number of experiments.

5.1 Accuracy on Noiseless Examples

We first tested the proposed method on the simple noiseless examples described in Section 3. These settings involved 2 bases, since the two rectangles simultaneously

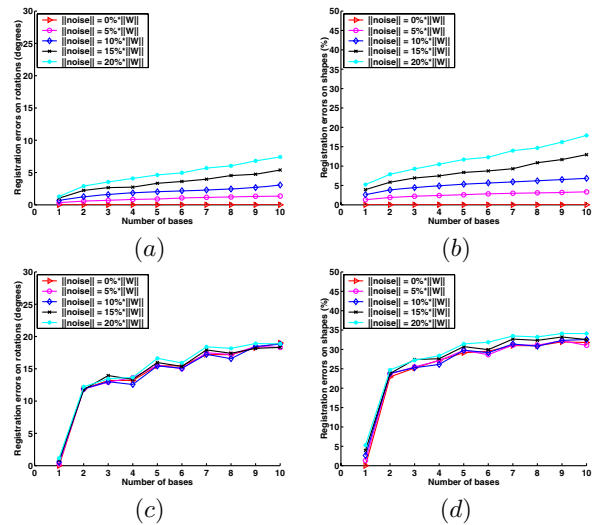


Figure 3. Registration and reconstruction errors on rotations and shapes under different noise levels and basis numbers. (a)&(b) By Direct Factorization. (c)&(d) By separate two-step method. Each curve refers to a noise level. The scalings of the error axis are $[0^\circ, 30^\circ]$ for rotations and $[0\%, 50\%]$ for shapes.

moved along straight lines. One basis referred to the initial shape and another stood for the linear motion vector. The Direct Factorization method automatically determined $K = 2$. It yielded the registered shapes identical to the ground truth with zero errors, no matter whether the shape deformations are symmetric or asymmetric, as shown in Fig.2. When the shape deformed asymmetrically, in contrast to the crooked point trajectories of the registration by GPA, our method recovered the 2-basis deformable model that reconstructed the accurate linear trajectories.

5.2 Quantitative Evaluation

We then quantitatively evaluated the accuracy and robustness of the Direct Factorization method with respect to different measurement noise levels and basis numbers. Assuming a Gaussian white noise, we represent the noise strength by the ratio between the Frobenius norms of the noise and the measurements, $\frac{\|noise\|}{\|\tilde{W}\|}$. The algorithm was tested under 5 different noise levels, 0%, 5%, 10%, 15%, and 20% respectively. On each level, we examined the rigid settings involving 1 basis and the non-rigid settings involving 2, ..., and 10 bases respectively in the 2D space.

For each of the settings, we tested 100 trials. At each trial, the bases were randomly generated and normalized ($\|B_i\| = 1$). Thus they were generally asymmetric. 66 shapes were constructed as linear combinations of the bases. The combination weights were randomly generated in such

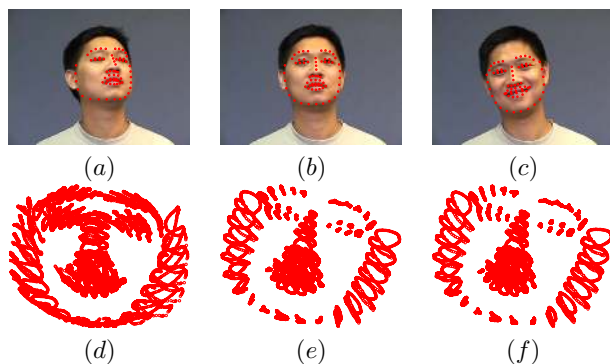


Figure 4. Registration and modeling of 2D face shapes carrying various expressions. (a~c): Three face images and shapes. (d): shape measurements after subtracting the translations. (e,f): Reconstructed 3-basis deformable shapes by Direct Factorization and the two-step method respectively. Their results were comparable because the deformations were similar to Gaussian noise.

a way that the weights for different bases had the same order of magnitude, *i.e.* none of the bases was dominant. These shapes were then randomly scaled, rotated, translated, and contaminated with noise to generate the measurements. Since the bases equally contributed to the shape composition, the bigger the basis number, the more “non-rigid” the shapes and the stronger the noise relative to each individual basis. Thus, in general, the registration is more sensitive to the noise when more bases were involved.

The shape measurements were registered by the proposed method and GPA respectively. Their average registration errors on the rigid rotations are shown in Fig.3.(a, c). Because the space of rotations is a manifold, the errors were measured as the Riemannian distance, $d(R_{est}, R_{truth}) = \arccos(\frac{\text{trace}(R_{est}R_{truth}^T)}{2})$, in degrees. To compare the modeling performance with the separate two-step method, the registered shapes by GPA were used to reconstruct the shape model and the true deformable shapes using PCA. The relative reconstruction errors on shapes were computed as, $\frac{\|Reconstruction-Truth\|}{\|Truth\|}$, as shown in Fig.3.(b, d).

When the noise level was 0%, the Direct Factorization method always recovered the exact rigid rotations and the true deformable shapes with zero errors. When noise was imposed, as expected, the registration was more sensitive to the noise and the errors were greater when the basis number was bigger. Yet our method achieved reasonable accuracy, *e.g.* in the worst case of 10 bases and 20% of the noise level, the average error on shapes was less than 18% and that on rotations was less than 7.5 degrees. In the rigid settings, both methods treated the shapes as rigid and the deformations as Gaussian noise. Their performances were

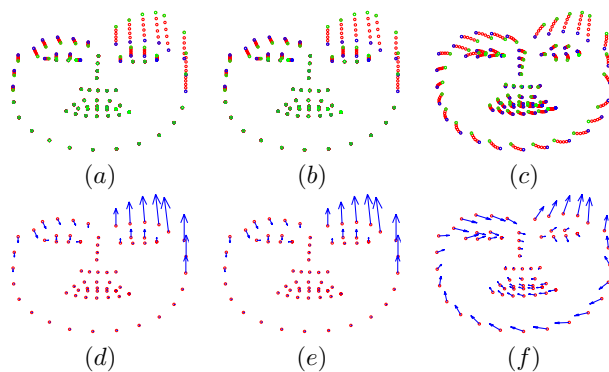


Figure 5. Registration and modeling of 2D face shapes, where the right eye lid and brow opened up, the left ones closed down, and the rest were mostly static. (a): Measured shapes. (d): Underlying bases: the initial shape and the linear motions. (b,e): Correct registration of the shapes and reconstruction of the 2-basis model by Direct Factorization. (c,f): Distorted registration and modeling by the two-step method.

almost identical. In the non-rigid trials the two-step method was always outperformed. Even in the noiseless cases, it yielded notable errors, 22% on shapes and 11 degrees on rotations. The errors by the two-step method were also bigger when the basis number was bigger. But its performance was less sensitive to the noise, because GPA computes the least squares solution that is insensitive to the Gaussian noise.

5.3 Qualitative Evaluation

Finally we examined our approach qualitatively on registering and modeling the shapes of three real deformable objects: human faces, myocardium, and dynamic scenes.

Human faces are highly non-rigid objects. It has been shown in [3] that the 2D images of the face shapes are linear combinations of certain basis shapes. This linear statistical model is important for tasks like face tracking and recognition. Our approach is capable of registering the deformable face shapes and reconstructing the linear shape model. One example is shown in Fig.4. The sequence consisted of 180 images that contained face rotations and facial expressions like blinking and smiling. The 2D face shapes were represented by 68 feature points. Fig.4.(a~c) display three of the images. The shape points are marked with red circles.

Subtracting from the shape measurements the translations, *i.e.* the average coordinates, the measurements were then composed of the deformations and the rotations. The feature points of the translated measurements are shown together in Fig.4.(d). The number of bases was estimated as 3 so that 99% of the energy could remain after the decomposition according to the rank constraint. Then the

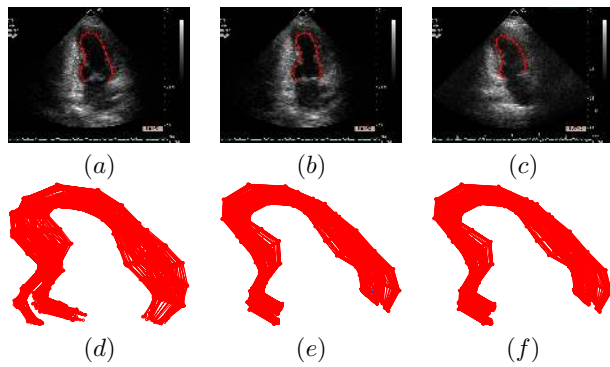


Figure 6. Registration and modeling of 2D myocardium shapes. (a~c): Three myocardium images and shapes. (d): Measured myocardium shapes. (e, f): Reconstructed 3-basis deformable myocardium shapes by Direct Factorization and the two-step method. Their results were comparable because the deformations were nearly symmetric.

rigid rotations and the 3-basis linear deformable model were computed using our method. The reconstructed deformable shapes in the 3-basis shape space are shown in Fig.4.(e). According to the shape point distributions in Fig.4.(d), the variations of the point coordinates were insignificant and their distributions were similar to Gaussian noise. Therefore the two-step method was able to achieve comparable registration and modeling, as shown in Fig.4.(f). The reconstructed 3-basis deformable shapes preserved very little rigid transformations. That is why the separate two-step method worked successfully in [3, 4, 5].

Fig.5.(a) shows a stack of deformable face shapes, where the deformations were asymmetric and not random: the right eye lid and brow opened up and the left ones closed down simultaneously. These shapes were observed from a fixed view point, without any rigid transformations. The measured shapes were thus composed of two linear bases alone, as shown in Fig.5.(d). One is the initial shape (red circles) and the other is the linear motions (blue arrows). The registered shapes by our method and GPA are demonstrated in Fig.5.(b) and (c) respectively, where the blue circles represent the initial shape, the green ones refer to the last shape, and the red ones mean the shapes in between. The 2-basis model reconstructed by our method and the two-step method are displayed in Fig.5.(e) and (f). The registration and reconstruction by the two-step method apparently involved rotations that did not occur in reality, while the Direct Factorization method correctly registered the shapes and reconstructed the deformable model.

Human organs such as heart and brain usually vary their shapes dynamically or across persons. The statistical shape models have been utilized for interpreting the images of

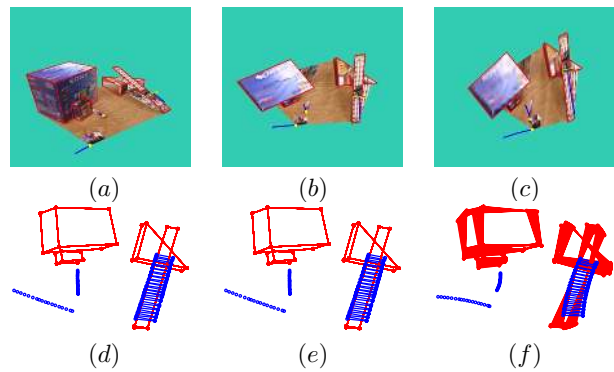


Figure 7. Registration and modeling of 3D dynamic scene shapes. (a~c): Three textured scene shape measurements, red lines for static shapes and blue for moving trajectories. (d): Ground truth shapes of the dynamic scene. (e): Exact reconstruction by Direct Factorization. (f): Distorted reconstruction by the two-step method.

these objects in many applications. We tested the proposed method on registering the shapes of myocardium in 2D echocardiography images and extracting the linear shape model. 120 myocardium images were captured. Three of them are shown in Fig.6.(a~c). Fig.6.(d) displays the 120 measured myocardium shapes after subtracting the translations. Our method determined the basis number as 3 and recovered the rotations and the 3-basis deformable shapes, as shown in Fig.6.(e). The most dominant basis occupied about 90% of the total energy, *i.e.* the deformations were insignificant like random noise. In addition, the shapes deformed somewhat symmetrically. Thus the two-step method was able to achieve the comparable registration and modeling performance, as shown in Fig.6.(f). This explains why the two-step method worked well in [2, 9].

3D modeling of dynamic scenes is important for tasks like robot navigation and visual surveillance. When a 3D scene consists of static buildings and vehicles or pedestrians moving straight, its shape is a linear combination of deformable bases. One example is shown in Fig.7. The sequence contained three toys moving along respective directions simultaneously, two on top of the table and one along the slope. The rest of the scene were static. The scene shape was composed of two bases, one for the initial shape and another representing the linear motions.

Eighteen shapes were measured from different view points, three of which with mapped texture are shown in Fig.7.(a~c). Red wireframes represent the static objects, yellow ones refer to the moving objects, and blue lines stand for the moving trajectories. Fig.7.(d) shows the ground truth 3D deformable shapes. Fig.7.(e) demonstrates that the 2-basis dynamic scene shapes were correctly reconstructed by

the Direct Factorization method. Since the linear 3D motions were significant and asymmetric, the corresponding shape deformations should not be treated as Gaussian noise. Fig.7.(f) illustrates that the reconstruction by the two-step method was apparently distorted with improper rotations, e.g. the motion trajectories were “curved”, even for this simple setting where only 6 out of 32 shape points moved.

In these experiments when the basis number was forced to be 1, i.e. regarding the shape as rigid and the deformations as noise, the rigid transformations and the registered shapes computed by our method were almost the same as those by the two-step method. This supports our analysis that the two-step method is a special case of the Direct Factorization method. Please refer to the supplementary videos for more details about the experiments.

6 Conclusion and Discussion

This paper has studied the problems of registration and modeling of deformable shapes. The previous methods separate the solutions into two steps and regard the deformations as Gaussian noise in the registration step. When the deformations do not act like Gaussian noise but as linear combinations of certain bases, the two-step methods result in registration biased by non-rigid deformations and deformable models distorted by rigid transformations. We for the first time point out this bias problem and propose the formulation that the registration and modeling problems are essentially coupled as a single factorization problem and require a simultaneous solution. We extend the method for non-rigid structure from motion [16] and develop a Direct Factorization method that simultaneously registers the deformable shapes and constructs the linear shape bases.

Currently the linear constraints on orthonormality of rotation transformations and uniqueness of deformation bases are enforced equivalently. However, the former constraints come from all the measured shapes and the latter ones arise from only the first K shapes. When noise exists, they might have different stabilities and thus different importance to the solution. Within either group of the constraints, the importance of individual constraints also varies because of noise and outliers, such as missing data, associated with the measurement. Currently we are exploring how to weight the constraints properly to improve the accuracy and robustness of the Direct Factorization method. Another benefit of having different weights for the constraints is that we can sample them to improve the efficiency of the algorithm.

Acknowledgments

This work was partly supported by Siemens Corporate Research. Jing Xiao and Takeo Kanade were partly supported by grant R01MH51435 from the National Institute of Mental Health. Jing Xiao was also supported partly by Epson R&D, Inc.

References

- [1] C. Bregler, A. Hertzmann, H. Biermann, “Recovering Non-Rigid 3D Shape from Image Streams,” *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp. 690-696, 2000.
- [2] D. Comaniciu, X.S. Zhou, S. Krishnan, “Robust Real-Time Tracking of Myocardial Border: An Information Fusion Approach,” *IEEE Trans. Medical Imaging*, Vol. 23, No. 7, pp 849-860, 2004.
- [3] T. Cootes, C. Taylor, “Active Shape Models- ‘Smart Snakes’,” *Proc. British Machine Vision Conference*, pp 266-275, 1992.
- [4] T. Cootes, C. Taylor, “Statistical Models of Appearance for Medical Image Analysis and Computer Vision,” *Proc. SPIE Medical Imaging*, pp 236-248, 2001.
- [5] T. Cootes, C. Taylor, D. Cooper, J. Graham, “Active Shape Models-Their Training and Application,” *Computer Vision and Image Understanding*, Vol. 61, No. 1, pp 38-59, 1995.
- [6] I. L. Dryden, K. V. Mardia, “Statistical Shape Analysis,” *J. Wiley, Chichester*, 1998.
- [7] C. Goodall, “Procrustes Methods in the Statistical Analysis of Shape,” *Journal Royal Statist. Soc.*, B. 53, No. 2, pp 285-339, 1991.
- [8] J. C. Gower, “Generalized Procrustes Analysis,” *Psychometrika*, Vol. 40, No. 1, pp 33-51, 1975.
- [9] G. Jacob, J.A. Noble, C. Behrenbruch, A.D. Kelion, A.P. Banning, “A Shape-Space-Based Approach to Tracking Myocardial Borders and Quantifying Regional Left-Ventricular Function Applied in Echocardiography,” *IEEE Trans. Medical Imaging*, Vol. 21, No. 3, pp 226-238, 2002.
- [10] M. A. Koschat, D. F. Swayne, “A Weighted Procrustes Criterion,” *Psychometrika*, Vol. 56, No. 2, pp 229-239, 1991.
- [11] R. W. Lissitz, P. H. Schönemann, J. C. Lingoes, “A Solution to the Weighted Procrustes Problem in Which the Transformation is in Agreement with the Loss Function,” *Psychometrika*, Vol. 41, No. 4, pp 547-550, 1976.
- [12] P. H. Schönemann, “A Generalized Solution of the Orthogonal Procrustes Problem,” *Psychometrika*, Vol. 31, No. 1, pp 1-10, 1966.
- [13] P. H. Schönemann, R. M. Carroll, “Fitting One Matrix to Another under Choice of A Central Dilation and A Rigid Motion,” *Psychometrika*, Vol. 35, No. 2, pp 245-255, 1970.
- [14] D. Shen, C. Davatzikos, “An Adaptive-Focus Deformable Model Using Statistical and Geometric Information,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp 906-913, 2000.
- [15] J. M. F. Ten Berge, “Orthogonal Procrustes Rotation for Two or More Matrices,” *Psychometrika*, Vol. 42, No. 2, pp 267-276, 1977.
- [16] J. Xiao, J. Chai, T. Kanade, “A Closed-Form Solution to Non-Rigid Shape and Motion Recovery,” *The 8th European Conference on Computer Vision*, 2004.