Procrustean Normal Distribution for Non-Rigid Structure from Motion

Minsik Lee   Jungchan Cho   Chong-Ho Choi   Songhwai Oh

Department of EECS, ASRI, Seoul National University, 151-742, Seoul, Korea

mlee.paper@gmail.com   {cjc83, chchoi, songhwai}@snu.ac.kr

Abstract

Non-rigid structure from motion is a fundamental problem in computer vision, which is yet to be solved satisfactorily. The main difficulty of the problem lies in choosing the right constraints for the solution. In this paper, we propose new constraints that are more effective for non-rigid shape recovery. Unlike the other proposals which have mainly focused on restricting the deformation space using rank constraints, our proposal constrains the motion parameters so that the 3D shapes are most closely aligned to each other, which makes the rank constraints unnecessary. Based on these constraints, we define a new class of probability distribution called the Procrustean normal distribution and propose a new NRSfM algorithm, EM-PND. The experimental results show that the proposed method outperforms the existing methods, and it works well even if there is no temporal dependence between the observed samples.

1. Introduction

Finding 3D structure and motion from a set of 2D observations is one of the basic problems in computer vision. Many applications, such as stereo reconstruction [7], require this task. Structure from motion (SfM) [10], which estimates the 3D shape and pose of a rigid object from 2D-point tracks, is the most simple form of this problem. SfM has been well-studied in a past few decades and can be effectively solved by using the matrix factorization approach [7, 10]. However, its counterpart for deformable objects, non-rigid structure from motion (NRSfM) [13], is a difficult problem to solve. The difficulty is mainly due to the ambiguity of the solution [13], which is a natural consequence of the increased degree of freedom compared to SfM. Therefore, many researchers have tried to solve NRSfM by adding additional constraints [2, 4, 5, 8, 11, 13].

NRSfM is usually solved either by using the factorization algorithm or the EM algorithm that are combined with suitable constraints. Hence, how to choose the constraints is really important. Majority of the approaches [4, 8, 11, 13] fix the number of shape bases to restrict the 'degree' of de-

formation. Based on this rank constraint, Akhter *et al.* [1] showed that only an orthogonality constraint on the rotations is sufficient to find a unique solution. However, the choice on the number of shape bases greatly affects the reconstruction performance and it is difficult to know the right number. Meanwhile, dual approaches [2, 5] that constrain the number of trajectory bases, have been proposed recently. These approaches assume temporal dependence between frames, and incorporate the discrete cosine transform (DCT) bases in the model. However, the number of DCT bases should also be predetermined for these methods, which is another difficulty.

All of these difficulties are caused by constraining the number of bases. Most importantly, this number changes the solution of rotations, because the rotations are found based on the factorization results. As demonstrated in [1], the uniqueness of the solution of NRSfM is closely connected with the uniqueness of rotations. This means that finding the correct rotations are vital for NRSFM.

In this paper, we propose a novel algorithm that solves NRSfM without any rank constraint. Because finding the correct rotations is important, we consider NRSfM as an alignment problem and introduce an additional constraint to each rotation matrix. This constraint is derived from the generalized Procrustes analysis (GPA) [6, 14], which aligns a set of shapes most closely to each other. Owing to this new constraint, we do not need any rank constraint. We also modify the scale constraint in GPA to make the aligned shapes lie in a linear subspace, which makes the problem tractable. This subspace includes all possible deformation of shapes, and moreover, the null space of this subspace is 7-dimensional that is related to the variations due to rigid transformations. In other words, rigid and non-rigid shape variations are strictly separated under these constraints. This leads us to define a new class of probability distribution called the Procrustean normal distribution (PND), which is a special case of the normal distribution. The proposed algorithm, EM-PND, essentially fits the 2D-point tracks to a PND using the EM algorithm. Since there is no constraint on the deformation space, EM-PND loses less detail in the reconstructed shapes. Moreover, it does

not require any temporal dependence between the observed 2D tracks. EM-PND also works well when there are some missing data in the observation. Experimental results show that the proposed algorithm performs best among the existing methods in most of the cases.

The remainder of this paper is organized as follows: We explain the definitions and properties of PND in Section 2. Based on PND, the proposed algorithm is introduced in Section 3. The experiments are followed in Section 4, and finally, we conclude the paper in Section 5.

2. Procrustean Normal Distribution

Estimating rotations based on rank and orthonormality constraints as in other NRSfM algorithms can be troublesome, because of two reasons. First, knowing the correct rank is not easy and wrongly chosen rank can ruin the estimation. Second, how orthonormal the estimated rotations are is not directly connected to the accuracy of the rotations. Therefore, we propose another way to model the rotations by incorporating GPA. GPA finds the relative motions (including rotations) between similar shapes by aligning them as closely as possible. This principle determines rigid motions by minimizing non-rigid variations, which can improve the accuracy of the estimated rotations in NRSfM.

We show that the solution of GPA can be described by a set of simple conditions with respect to (w.r.t.) aligned shapes, and propose a new distribution based on the conditions. This distribution can be effectively used to separate rigid and non-rigid shape variations, and serves as a core component of the proposed algorithm.

2.1. GPA and its modification

GPA superimposes multiple landmark shapes to a common reference using rigid transformations. Let $\mathbf{X}_i \in \mathbb{R}^{n_d \times n_p}$, $s_i \in \mathbb{R}$, $\mathbf{R}_i \in \mathbb{R}^{n_d \times n_d}$, and $\mathbf{t}_i \in \mathbb{R}^{n_d}$ be the 3D shape, scale, rotation, and translation, respectively, for the ith sample, $1 \leq i \leq n_s$, where n_d, n_p, and n_s are the dimension of the coordinate system, the number of landmarks in a frame, and the number of frames, respectively. Then, the GPA problem is defined as

$$\min_{s_i, \mathbf{R}_i, \mathbf{t}_i, \overline{\mathbf{X}}} \quad \sum \left\| s_i \mathbf{R}_i \mathbf{X}_i + \mathbf{t}_i \mathbf{1}^T - \overline{\mathbf{X}} \right\|^2$$
$$\text{subject to} \quad \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \quad f(\mathbf{s}) = 1, \tag{1}$$

where $\mathbf{1}$ is a vector of ones, $\overline{\mathbf{X}}$ is the mean shape, and $\mathbf{s} = \begin{bmatrix} s_1 & \cdots & s_{n_s} \end{bmatrix}^T$. In this paper, $\|\cdot\|$ denotes the Frobenius norm for matrices, $i.e.$, $\|\mathbf{A}\|^2 = \mathrm{tr}(\mathbf{A}^T \mathbf{A}) = \|\mathrm{vec}(\mathbf{A})\|^2$, where $\mathrm{vec}(\cdot)$ is a vectorization operator. Here, $f(\mathbf{s}) = 1$ is a scale constraint and the two popular choices [6, 14] are (i) $\|s_i \mathbf{X}_i\| = 1$ and (ii) $\frac{1}{n_s} \sum \|s_i \mathbf{X}_i\|^2 = 1$. The aligned shapes will lie on a Procrustes shape space if the first constraint is used, and otherwise they will lie on a Kendall

shape space [14]. The procedure for solving a GPA problem is summarized as

1. Translate $\mathbf{X}_i$ as $\mathbf{X}_i \leftarrow \mathbf{X}_i - \frac{1}{n_p} \mathbf{X}_i \mathbf{1} \mathbf{1}^T$, and normalize it, $i.e.$, $\|\mathbf{X}_i\| = 1$. Set $s_i = 1$ and $\mathbf{R}_i = \mathbf{I}$.

2. Calculate the mean $\overline{\mathbf{X}} = \frac{1}{n_s} \sum s_i \mathbf{R}_i \mathbf{X}_i$.

3. Update the rotations for the new mean. Each new rotation is given as $\mathbf{R}_i = \mathbf{V}_i \mathbf{U}_i^T$ where $\mathbf{X}_i \overline{\mathbf{X}}^T = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^T$ is a singular value decomposition (SVD).

4. Update the scales so that the objective function is minimized and the scale constraint is satisfied.

5. Repeat Steps 2–4 until convergence.

Here, the translation component is removed from each shape in Step 1 and does not appear in the iterative procedure, because the optimal $\mathbf{t}_i$ is in fact given as $\mathbf{t}_i = -\frac{1}{n_p} s_i \mathbf{R}_i \mathbf{X}_i \mathbf{1}$ [6]. Accordingly, the optimal $\overline{\mathbf{X}}$ has a zero translation component because $\overline{\mathbf{X}} \mathbf{1} = \frac{1}{n_s} \sum s_i \mathbf{R}_i \mathbf{X}_i \mathbf{1} = \mathbf{0}$.

One may think of extending GPA to missing data cases as a way of solving the NRSfM problem. However, both the Procrustes and Kendall shape spaces are nonlinear manifolds, which make it hard to handle the distribution of the shape. We may drop the scale constraint to resolve this issue, but it is not a good idea because there can be scale changes due to camera motion that may significantly affect the reconstruction performance (as shown in the supplementary material). Therefore, we need to find a new scale constraint that makes the aligned shapes lie in a linear subspace. To do this, we propose another scale constraint so that each shape variation from the mean shape is orthogonal to the mean shape, $i.e.$,

$$\mathrm{vec}\left(s_i \mathbf{R}_i \mathbf{X}_i - \overline{\mathbf{X}}\right)^T \mathrm{vec}(\overline{\mathbf{X}}) = 0. \tag{2}$$

If we impose an additional constraint that the norm of the mean shape is one, $i.e.$, $\|\overline{\mathbf{X}}\| = 1$, then this constraint can be rewritten as

$$\mathrm{vec}(s_i \mathbf{R}_i \mathbf{X}_i)^T \mathrm{vec}(\overline{\mathbf{X}}) = s_i \mathrm{tr}\left(\mathbf{R}_i \mathbf{X}_i \overline{\mathbf{X}}^T\right) = 1. \tag{3}$$

Note that this is a linear constraint w.r.t. s_i, and it makes the aligned shapes lie in the stereographic projection [14], which is a mapping from an n-dimensional sphere to a hyperplane, of a Procrustes shape space. This let us describe the distribution of the aligned shapes more easily using typical probability distributions. Step 4 of the above algorithm can be restated incorporating these constraints as

4. Normalize the mean shape as $\overline{\mathbf{X}} \leftarrow \overline{\mathbf{X}} / \|\overline{\mathbf{X}}\|$ and update each scale as $s_i = 1 / \mathrm{tr}\left(\mathbf{R}_i \mathbf{X}_i^T \overline{\mathbf{X}}\right)$.

Then, the necessary condition for optimality of problem (1) is obtained as

$$\mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \qquad \|\overline{\mathbf{X}}\|^2 = 1,$$
$$s_i \operatorname{tr}\left(\mathbf{R}_i \mathbf{X}_i \overline{\mathbf{X}}^T\right) = 1, \quad \mathbf{R}_i \mathbf{X}_i \overline{\mathbf{X}}^T \in \mathbf{S}_+^{n_d}, \qquad (4)$$

where $\mathbf{S}_+^{n_d}$ is a set of n_d-dimensional positive semi-definite (PSD) matrices, which is convex. s_i, R_i, and $\overline{\mathbf{X}}$ must satisfy these conditions to be optimal. A more important insight can be drawn from these conditions: Note that the third condition is a linear equation w.r.t. an aligned shape $s_i \mathbf{R}_i \mathbf{X}_i$ and the last condition forces the set of the aligned shapes to be convex, which are relatively easy to handle. Hence, if we solve NRSfM based on a distribution constrained by these conditions, then we may enforce the reconstructed shapes to be a possible optimal solution for problem (1). This motivates us to define a new distribution. Because there are three more conditions in (4) besides the first equation, which is the orthonormality constraint in other NRSfM algorithms, the solution may be quite different from the other NRSfM algorithms. In fact, these conditions can concisely describe a set of non-rigid deformations, which will be explained in the next subsection.

2.2. Definition and Properties of PND

Let $\mathbf{Y}_i$ be the ith aligned shape expressed as $\mathbf{Y}_i = s_i \mathbf{R}_i \mathbf{X}_i$ and $\overline{\mathbf{Y}} = \overline{\mathbf{X}}$, then (4) can be written as

$$\|\overline{\mathbf{Y}}\|^2 = 1, \quad \operatorname{tr}\left(\mathbf{Y}_i \overline{\mathbf{Y}}^T\right) = 1, \quad \mathbf{Y}_i \overline{\mathbf{Y}}^T \in \mathbf{S}_+^{n_d}. \quad (5)$$

Here, the second and last conditions are linear equality and convex constraints w.r.t. $\mathbf{Y}_i$, respectively. However, if we relax the PSD constraint to a symmetric constraint, then the last one can be also expressed as a linear equality constraint w.r.t. $\mathbf{Y}_i$, i.e.,

$$\mathbf{Y}_i \overline{\mathbf{Y}}^T - \overline{\mathbf{Y}} \mathbf{Y}_i^T = \mathbf{0}. \qquad (6)$$

(In this paper, we use $\mathbf{0}$ to denote both matrices and vectors of zeros.) To set the centroid of an aligned shape at the origin, an additional constraint is introduced as

$$\mathbf{Y}_i \mathbf{1} = \mathbf{0}. \qquad (7)$$

These constraints can be simplified using the vectorization operator as

$$\left\|\operatorname{vec}(\overline{\mathbf{Y}})\right\|^2 = 1, \quad \mathbf{P}_N^T \operatorname{vec}(\mathbf{Y}_i - \overline{\mathbf{Y}}) = \mathbf{0}, \qquad (8)$$

based on an $n_Y \times n_N$-dimensional matrix $\mathbf{P}_N$, where $n_Y = n_d n_p$ and $n_N = n_d(n_d+1)/2+1$. $\mathbf{P}_N$ is a function of $\overline{\mathbf{Y}}$ and is defined as

$$\mathbf{P}_N = \begin{bmatrix} \operatorname{vec}(\overline{\mathbf{Y}}) & \mathbf{K}(\overline{\mathbf{Y}}) & (\mathbf{1} \otimes \mathbf{I}) \end{bmatrix},$$
$$\mathbf{K}(\overline{\mathbf{Y}}) = \begin{bmatrix} [\overline{\mathbf{y}}_1]_\times & \cdots & [\overline{\mathbf{y}}_{n_p}]_\times \end{bmatrix}^T, \qquad (9)$$

where $\overline{\mathbf{y}}_i$ is the ith column vector of $\overline{\mathbf{Y}}$ and $\otimes$ is the Kronecker product. $[\mathbf{y}]_\times \in \mathbb{R}^{\binom{n_d}{2} \times n_d}$ is a linear function of $\mathbf{y}$ and is a matrix equivalent of the exterior product [3] of $\mathbf{y}$ with another vector. For $n_d = 3$, this is related to the cross product and becomes

$$[\mathbf{y}]_\times = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix}. \qquad (10)$$

Note that n_N, the number of columns of $\mathbf{P}_N$, corresponds to the degree of freedom (DOF) of a rigid transform. For example, $n_N = 7$ for $n_d = 3$, which corresponds to the DOF of a rigid transform (1 for scale, 3 for rotation, and 3 for translation) in a 3-dimensional space. This follows from the fact that $\operatorname{vec}(\overline{\mathbf{Y}})$, $\mathbf{K}(\overline{\mathbf{Y}})$, and $(\mathbf{1} \otimes \mathbf{I})$ in (9) correspond to the scale, rotation, and translation constraints, respectively. Since the variation of $\mathbf{Y}_i$ is only allowed in the subspace perpendicular to $\mathbf{P}_N$, i.e., perpendicular to the subspace of rigid shape variations, the variations will only contain nonrigid shape variations. Furthermore, $\operatorname{vec}(\overline{\mathbf{Y}})$, $\mathbf{K}(\overline{\mathbf{Y}})$, and $(\mathbf{1} \otimes \mathbf{I})$ have the following interesting property.

Proposition 1. *The range spaces of* $\operatorname{vec}(\overline{\mathbf{Y}})$, $\mathbf{K}(\overline{\mathbf{Y}})$, *and* $(\mathbf{1} \otimes \mathbf{I})$ *are orthogonal to each other.*

Proof. First, we show that $(\mathbf{1} \otimes \mathbf{I})$ is orthogonal to the other terms. Note that

$$(\mathbf{1} \otimes \mathbf{I})^T \operatorname{vec}(\overline{\mathbf{Y}}) = \overline{\mathbf{Y}} \mathbf{1} = \mathbf{0}, \qquad (11)$$

because the centroid of $\overline{\mathbf{Y}}$ is at the origin. Similarly, because $[\mathbf{y}]_\times$ is a linear function of $\mathbf{y}$,

$$(\mathbf{1} \otimes \mathbf{I})^T \mathbf{K}(\overline{\mathbf{Y}}) = \sum [\overline{\mathbf{y}}_i]_\times^T = \left[\sum \overline{\mathbf{y}}_i\right]_\times^T = \mathbf{0}. \quad (12)$$

Finally, we show the relationship between $\operatorname{vec}(\overline{\mathbf{Y}})$ and $\mathbf{K}(\overline{\mathbf{Y}})$. Note that

$$\mathbf{K}(\overline{\mathbf{Y}})^T \operatorname{vec}(\overline{\mathbf{Y}}) = \sum [\overline{\mathbf{y}}_i]_\times \overline{\mathbf{y}}_i. \qquad (13)$$

Hence, this expression is equivalent to the sum of the exterior products of $\overline{\mathbf{y}}_i$ with themselves. Because the exterior product of a vector with itself is $\mathbf{0}$, i.e., $\mathbf{y} \wedge \mathbf{y} = \mathbf{0}$ and $\mathbf{y} \times \mathbf{y} = \mathbf{0}$ for $n_d = 3$, (13) is also $\mathbf{0}$. $\square$

This proposition states that each constraint specifies a different orthogonal subspace. In other words, variations due to scaling, rotation and translation are mutually orthogonal to each other under the constraints in (8).

Based on these constraints, we can define a new distribution that only includes non-rigid shape variations, eliminating rigid variations. We call this the Procrustean normal

distribution, and denote it as $\mathcal{N}_P(\overline{\mathbf{Y}}, \boldsymbol{\Sigma})$. The parameters of a PND satisfy the following conditions:

$$\left\|\overline{\mathbf{Y}}\right\|^2 = 1, \quad \overline{\mathbf{Y}}\mathbf{1} = \mathbf{0}, \quad \mathbf{P}_N^T \boldsymbol{\Sigma} = \mathbf{0}. \tag{14}$$

Hence, $\boldsymbol{\Sigma} \in \mathbb{R}^{n_Y \times n_Y}$ is singular. To handle the singularity, we denote $\mathbf{Q}_N$ as an orthogonalized version of $\mathbf{P}_N$. Based on Proposition 1, $\mathbf{Q}_N$ can be described as

$$\mathbf{Q}_N = \begin{bmatrix} \text{vec}(\overline{\mathbf{Y}}) & \mathbf{Q}_K & \left(\frac{1}{\sqrt{n_p}}\mathbf{1} \otimes \mathbf{I}\right) \end{bmatrix}, \tag{15}$$

where $\mathbf{Q}_K$ is an orthogonalized version of $\mathbf{K}(\overline{\mathbf{Y}})$, which can be calculated using the QR decomposition. We also define $\mathbf{Q}$ as an orthogonal matrix whose range space is the left null space of $\mathbf{Q}_N$. Then, the probability distribution function (pdf) of a PND random matrix $\mathbf{Y} \sim \mathcal{N}_P(\overline{\mathbf{Y}}, \boldsymbol{\Sigma})$ is

$$p(\mathbf{Y}) \propto \frac{1}{|\boldsymbol{\Sigma}_R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{v}^T \mathbf{Q}\boldsymbol{\Sigma}_R^{-1}\mathbf{Q}^T \mathbf{v}\right) \delta\left(\mathbf{Q}_N^T \mathbf{v}\right),$$
$$\boldsymbol{\Sigma}_R = \mathbf{Q}^T \boldsymbol{\Sigma} \mathbf{Q}, \qquad \mathbf{v} = \text{vec}(\mathbf{Y} - \overline{\mathbf{Y}}), \tag{16}$$

where $\boldsymbol{\Sigma}_R \in \mathbb{R}^{n_R \times n_R}$ is the covariance matrix of $\mathbf{Q}^T \mathbf{v}$, $\delta(\cdot)$ is the n_N-dimensional Dirac-delta function, and $n_R = n_Y - n_N$. The Dirac-delta function is introduced so that $p(\mathbf{Y})$ can be expressed with the degenerate $\boldsymbol{\Sigma}$. Note that $\mathbf{Q}\boldsymbol{\Sigma}_R^{-1}\mathbf{Q}^T = \boldsymbol{\Sigma}^+$, where $(\cdot)^+$ is the pseudo-inverse of $(\cdot)$.

Because PND does not include any rigid shape variations, it is possible to find relative motions between sample shapes by fitting them to a PND. This has the same effect as solving a GPA problem, and moreover, we can apply some statistical estimation techniques to PND. There are two important advantages of PND in solving NRSfM. First, because no low-rank constraint is needed for PND, there is no need to adjust the rank and less details will be lost in the fitting process. Second, PND strictly separates rigid and non-rigid variations in the fitting process, which will lead to more accurate motion parameters and reconstructed shapes. The idea of ruling out rigid variations is somewhat similar to that in [12], however, PND does not require accurate motion information in advance, unlike the approach. For PND, estimating the distribution of non-rigid variations determines the rigid motions as a by-product, because $\mathbf{Q}_N$ depends on the mean shape.

A rotation of a PND is also a PND as explained in the following proposition:

Proposition 2. *Let* $\mathbf{Y} \sim \mathcal{N}_P(\overline{\mathbf{Y}}, \boldsymbol{\Sigma})$ *be a PND random matrix, and* $\mathbf{Y}' = \mathbf{R}\mathbf{Y}$ *for an orthogonal matrix* $\mathbf{R}$. *Then,* $\mathbf{Y}'$ *is also a PND random matrix as* $\mathbf{Y}' \sim \mathcal{N}_P(\mathbf{R}\overline{\mathbf{Y}}, (\mathbf{I} \otimes \mathbf{R}) \boldsymbol{\Sigma} (\mathbf{I} \otimes \mathbf{R}^T))$.

Proof. First of all, a linear transform of a normal distribution is also Gaussian, hence $\mathbf{Y}'$ is Gaussian. Moreover, the

determinant $|\boldsymbol{\Sigma}_R|$ remains the same because the transform is orthogonal. By observing the squared error term in (16),

$$\text{vec}\left(\mathbf{R}^T\overline{\mathbf{Y}}' - \overline{\mathbf{Y}}\right)^T \boldsymbol{\Sigma}^+ \text{vec}\left(\mathbf{R}^T\overline{\mathbf{Y}}' - \overline{\mathbf{Y}}\right)$$
$$= \mathbf{v}'^T \left((\mathbf{I} \otimes \mathbf{R}) \boldsymbol{\Sigma} (\mathbf{I} \otimes \mathbf{R}^T)\right)^+ \mathbf{v}', \tag{17}$$
$$\mathbf{v}' = \text{vec}(\mathbf{Y}' - \mathbf{R}\overline{\mathbf{Y}}),$$

and this gives the mean and the covariance matrix stated in the proposition.

The only thing left to prove is whether the constraints on $\mathbf{Y}'$ hold after the transformation. To show this, we substitute $\mathbf{Y} = \mathbf{R}^T\mathbf{Y}'$ in (5), (6), and (7), *i.e.*,

$$\text{tr}\left((\mathbf{R}^T\mathbf{Y}')\overline{\mathbf{Y}}^T\right) = \text{tr}\left(\mathbf{Y}'(\mathbf{R}\overline{\mathbf{Y}})^T\right) = 1,$$
$$(\mathbf{R}^T\mathbf{Y}')\overline{\mathbf{Y}}^T - \overline{\mathbf{Y}}(\mathbf{R}^T\mathbf{Y}')^T = \mathbf{0}, \quad (\mathbf{R}^T\mathbf{Y}')\mathbf{1} = \mathbf{0}. \tag{18}$$

By multiplying $\mathbf{R}$ in the left and right sides of the second equation in (18), we obtain $\mathbf{Y}'(\mathbf{R}\overline{\mathbf{Y}})^T - (\mathbf{R}\overline{\mathbf{Y}})\mathbf{Y}'^T = \mathbf{R}\mathbf{0}\mathbf{R}^T = \mathbf{0}$, and similarly, $\mathbf{Y}'\mathbf{1} = \mathbf{R}\mathbf{0} = \mathbf{0}$ from the third equation. Thus, the constraints for $\mathbf{Y}'$ holds after the transform. $\square$

Note that (14) can be applied to distributions other than the normal distribution for the analysis of NRSfM, but we use the normal distribution for its simplicity. The DOF of PND is $n_Y - n_d - 1$ for $\overline{\mathbf{Y}}$ and $n_R(n_R + 1)/2$ for $\boldsymbol{\Sigma}$.[1] If a distribution satisfies the properties of PND except the first one of (14), we call it a scaled PND and denote it as $\mathcal{N}_P^s$.

3. The proposed algorithm: EM-PND

Let $\mathbf{D}_i \in \mathbb{R}^{n_d \times n_p}$ be the input landmark data, observed by an orthographic camera, of the ith sample, and $\mathbf{W}_i \in \mathbb{R}^{n_d \times n_p}$ be the weight matrix filled with ones and zeros that indicates whether the corresponding elements are observed or missing. In case of NRSfM, the last row of $\mathbf{W}_i$ should be filled with zeros because the z coordinates are unknown. We also assume that the translation component is initially removed from each $\mathbf{D}_i$ as

$$d_{ijk} \leftarrow \begin{cases} d_{ijk} - \frac{\sum_l w_{ijl}d_{ijl}}{\sum_l w_{ijl}} & \text{if } w_{ijk} = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{19}$$

where d_{ijk} and w_{ijk} are the (j, k)th elements of $\mathbf{D}_i$ and $\mathbf{W}_i$, respectively. This can be alternatively represented as

$$\mathbf{D}_i \leftarrow \mathbf{W}_i \odot \left(\mathbf{D}_i - (\mathbf{c}_i \odot ((\mathbf{W}_i \odot \mathbf{D}_i)\mathbf{1}))\mathbf{1}^T\right), \tag{20}$$

where $\odot$ is the Hadamard product and $\mathbf{c}_i$ is an n_d-dimensional vector whose jth element is $\frac{1}{\sum_l w_{ijl}}$ (0 if the jth row of $\mathbf{W}_i$ is filled with zeros). After this initialization, $\mathbf{D}_i$ satisfies $\mathbf{D}_i\mathbf{1} = \mathbf{0}$. EM-PND fits a PND to these input data using the EM algorithm. In the following subsections, we will explain the detailed procedure of EM-PND.

---

[1]The derivation is given in the supplementary material.

3.1. Objective function

The overall goal of EM-PND is to maximize the following log-likelihood function:

$$\sum \log\big(p(\mathbf{D}_i, \mathbf{X}_i | \sigma, s_i, \mathbf{R}_i, \overline{\mathbf{X}}, \mathbf{\Sigma})\big) \tag{21}$$
$$= \sum \log(p(\mathbf{D}_i | \mathbf{X}_i, \sigma)) + \log\big(p(\mathbf{X}_i | s_i, \mathbf{R}_i, \overline{\mathbf{X}}, \mathbf{\Sigma})\big),$$

where σ is the standard deviation of the Gaussian noise in $\mathbf{D}_i$. Here, $\mathbf{X}_i$ is a hidden variable representing the true 3D shape of the ith sample.

Let the likelihood of $\mathbf{D}_i$ be defined as

$$p(\mathbf{D}_i | \mathbf{X}_i, \sigma) = \frac{1}{(2\pi)^{n_i^W/2} \sigma^{n_i^W}} \exp\left(-\frac{1}{2\sigma^2} g_i^2\right),$$
$$g_i^2 = \min_{\mathbf{t}_i} \|\mathbf{D}_i - \mathbf{W}_i \odot (\mathbf{X}_i - \mathbf{t}_i)\|^2, \tag{22}$$
$$n_i^W = \sum_j \left(\sum_k w_{ijk}\right) - \text{sign}\left(\sum_k w_{ijk}\right),$$

where $\text{sign}(\cdot)$ denotes the signum function. g_i can be alternatively expressed as

$$g_i^2 = \|\text{vec}(\mathbf{D}_i) - \mathbf{F}_i \text{vec}(\mathbf{X}_i)\|^2,$$
$$\mathbf{F}_i = \widehat{\mathbf{W}}_i - \widehat{\mathbf{W}}_i (\mathbf{1} \otimes \mathbf{I}) \mathbf{c}_i (\mathbf{1}^T \otimes \mathbf{I}) \widehat{\mathbf{W}}_i, \tag{23}$$
$$\widehat{\mathbf{W}}_i = \text{diag}(\text{vec}(\mathbf{W}_i)).$$

Here, $\mathbf{F}_i$ is a projection matrix, *i.e.*, $\mathbf{F}_i^2 = \mathbf{F}_i$, and $\text{diag}(\mathbf{a})$ denotes a diagonal matrix with the elements of vector $\mathbf{a}$ on the main diagonal.

In order to represent the prior distribution of $\mathbf{X}_i$, we assume that the aligned shapes $\mathbf{Y}_i = s_i \mathbf{R}_i \mathbf{X}_i$ are independently and identically distributed PND random matrices, *i.e.*, $\mathbf{Y}_i \sim \mathcal{N}_P(\overline{\mathbf{X}}, \mathbf{\Sigma})$. According to Proposition 2, $\mathbf{X}_i$ follows a scaled PND, *i.e.*,

$$\mathbf{X}_i \sim \mathcal{N}_P^s\left(\frac{1}{s_i} \mathbf{R}_i^T \overline{\mathbf{X}}, \frac{1}{s_i^2}\left(\mathbf{I} \otimes \mathbf{R}_i^T\right) \mathbf{\Sigma}\left(\mathbf{I} \otimes \mathbf{R}_i\right)\right). \tag{24}$$

3.2. E-step

In E-step, we calculate the distribution of $\mathbf{X}_i$ given the current estimates of the parameters and $\mathbf{D}_i$. If we define the parameter set $\Phi = \{\sigma, s_i, \mathbf{R}_i, \overline{\mathbf{X}}, \mathbf{\Sigma}\}$,

$$p_E(\mathbf{X}_i) = p(\mathbf{X}_i | \mathbf{D}_i, \Phi) = \frac{p(\mathbf{X}_i, \mathbf{D}_i | \Phi)}{\int p(\mathbf{X}_i, \mathbf{D}_i | \sigma, \Phi) d\mathbf{X}_i}. \tag{25}$$

Since the distribution is Gaussian, we have to calculate the mean and covariance matrix. After some manipulation, we can find that

$$p_E(\mathbf{X}_i) \propto \exp\left(-\frac{1}{2} \text{vec}(\mathbf{X}_i)^T \mathbf{H}_i \text{vec}(\mathbf{X}_i) + \frac{1}{\sigma^2}\right.$$
$$\left. \text{vec}(\mathbf{D}_i)^T \text{vec}(\mathbf{X}_i)\right) \delta\left(\mathbf{Q}_N^T \text{vec}(\mathbf{R}_i \mathbf{X}_i) - \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}\right), \tag{26}$$

where $\mathbf{H}_i = s_i^2 \left(\mathbf{I} \otimes \mathbf{R}_i^T\right) \mathbf{\Sigma}^+ \left(\mathbf{I} \otimes \mathbf{R}_i\right) + \frac{1}{\sigma^2} \mathbf{F}_i$. However, in practice, the Dirac-delta term in this equation is too restrictive to allow a meaningful update in the EM procedure, *i.e.*, the parameters almost do not change in M-step. Hence, we ignore the Dirac-delta term, and finally, obtain the following distribution:

$$p_E(\mathbf{X}_i) = p_E(\text{vec}(\mathbf{X}_i)) \sim \mathcal{N}(\mathbf{m}_i, \mathbf{C}_i),$$
$$\mathbf{C}_i = \mathbf{H}_i^+, \quad \mathbf{m}_i = \frac{1}{\sigma^2} \mathbf{C}_i \text{vec}(\mathbf{D}_i). \tag{27}$$

Ignoring the Dirac-delta term makes the distribution $p_E(\mathbf{X}_i)$ closer to the observation $\mathbf{D}_i$. This has a positive effect to the convergence rate, but it also has a downside of underestimating σ in M-step.

3.3. M-step

In M-step, the parameter set Φ is updated so that it maximizes the expectation of the log-likelihood w.r.t. $p_E(\mathbf{X}_i)$, which is given as

$$J(\Phi) = E\left[\sum \log(p(\mathbf{D}_i, \mathbf{X}_i | \Phi)) \Big| p_E(\mathbf{X}_i)\right]$$
$$= \sum -n_i^W \log(\sigma) - \frac{1}{2\sigma^2} \|\text{vec}(\mathbf{D}_i) - \mathbf{F}_i \mathbf{m}_i\|^2$$
$$- \frac{1}{2\sigma^2} \text{tr}(\mathbf{F}_i \mathbf{C}_i) - \frac{1}{2} \log|\mathbf{\Sigma}_R| + n_R \log(s_i) - \frac{1}{2} \mathbf{h}_i^T \mathbf{\Sigma}_R^{-1} \mathbf{h}_i$$
$$- \frac{s_i^2}{2} \text{tr}\left(\left(\mathbf{I} \otimes \mathbf{R}_i^T\right) \mathbf{Q} \mathbf{\Sigma}_R^{-1} \mathbf{Q}^T \left(\mathbf{I} \otimes \mathbf{R}_i\right) \mathbf{C}_i\right) + c_0, \tag{28}$$

where $\mathbf{h}_i = \mathbf{Q}^T \left(s_i \left(\mathbf{I} \otimes \mathbf{R}_i\right) \mathbf{m}_i - \text{vec}(\overline{\mathbf{X}})\right)$ and c_0 is a constant. Then, the optimization problem for M-step is described as

$$\min_{\Phi} \quad J(\Phi)$$
$$\text{subject to} \quad \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \qquad \|\overline{\mathbf{X}}\|^2 = 1, \tag{29}$$
$$s_i \text{tr}\left(\mathbf{R}_i \mathbf{M}_i \overline{\mathbf{X}}^T\right) = 1, \quad \mathbf{R}_i \mathbf{M}_i \overline{\mathbf{X}}^T \in \mathbf{S}_+^{n_d},$$

where $\mathbf{M}_i$ is the expectation of $\mathbf{X}_i |_{p_E(\mathbf{X}_i)}$, *i.e.*, $\text{vec}(\mathbf{M}_i) = \mathbf{m}_i$. All the constraints in this problem are the same as the constraints in (4) except that $\mathbf{X}_i$ is replaced with its expectation $\mathbf{M}_i$. Here, the PSD condition has been reinstated in the last constraint because it makes the solution much easier to calculate.

Since this problem is very complicated, we solve it using the alternating update method, *i.e.*, we update each of the parameters assuming that the other parameters are fixed. In this framework, the most difficult parameter to update is $\overline{\mathbf{X}}$, on which $\mathbf{Q}$ as well as the last three constraints in (29) depend. To resolve this, we regard $\mathbf{Q}$ as an independent parameter and simply ignore the constraints in the update of $\overline{\mathbf{X}}$. Then, by differentiating the cost function w.r.t. $\overline{\mathbf{X}}$,

equating it to zero, and normalizing the solution, we obtain the following update equation:

$$\overline{\mathbf{X}} = \sum s_i \mathbf{R}_i \mathbf{M}_i / \left\| \sum s_i \mathbf{R}_i \mathbf{M}_i \right\|. \qquad (30)$$

The scales and rotations are relatively easy to update. According to the constraints in (29), the feasible s_i and $\mathbf{R}_i$ are unique if the other parameters are fixed and all the samples are non-degenerate. The corresponding update equations are

$$\mathbf{M}_i \overline{\mathbf{X}}^T = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^T, \quad \mathbf{R}_i = \mathbf{V}_i \mathbf{U}_i^T,$$
$$s_i = 1/\operatorname{tr}\left(\mathbf{R}_i \mathbf{M}_i \overline{\mathbf{X}}^T\right) = 1/\operatorname{tr}(\mathbf{\Lambda}_i), \qquad (31)$$

where the first equation is calculated using SVD.

To update $\mathbf{Q}_N$ and $\mathbf{Q}$ based on the new $\overline{\mathbf{X}}$, we calculate $\mathbf{P}_N$ using (9) and then compute $\mathbf{Q}_N$ and $\mathbf{Q}$ from it, as described in Section 2.2. $\mathbf{\Sigma}_R$ can be easily found by solving the first-order necessary condition of (29), *i.e.*,

$$\mathbf{\Sigma}_R = \frac{1}{n_s} \sum \mathbf{h}_i \mathbf{h}_i^T + s_i^2 \mathbf{Q}^T \left(\mathbf{I} \otimes \mathbf{R}_i\right) \mathbf{C}_i \left(\mathbf{I} \otimes \mathbf{R}_i^T\right) \mathbf{Q}. \qquad (32)$$

Accordingly, the covariance matrix is calculated as $\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Sigma}_R\mathbf{Q}^T$. σ can also be derived in a similar way. However, as mentioned in the previous section, σ tends to be underestimated due to the approximation of $p_E(\mathbf{X}_i)$, and we introduce a scaling constant $\alpha > 1$ to compensate this;

$$\sigma^2 = \frac{\alpha}{\sum n_i^W} \sum \|\operatorname{vec}(\mathbf{D}_i) - \mathbf{F}_i \mathbf{m}_i\|^2 + \operatorname{tr}(\mathbf{F}_i \mathbf{C}_i). \qquad (33)$$

For the experiments in Section 4, we used $\alpha = 2$.

Strictly speaking, this update procedure should be iterated several times in each M-step, but we have found in our experiments that a single iteration is sufficient.

3.4. Practical considerations

The problem we are dealing with is highly nonlinear, and providing good initial values for the parameters is important. The rotations are especially crucial. The proposed method works well for random initial rotations when the shape variations are moderate, but the performance may deteriorate for large variations. Hence, we adopt the initialization method used in [5] for the rotations, which calculates rotations based on the factorization results for all possible numbers of shape bases and then automatically chooses the most 'orthonormal' ones.[2] The scale parameters are initialized so that the norm of $\mathbf{D}_i$ becomes 1, *i.e.*, $s_i = \frac{1}{\|\mathbf{D}_i\|}$. Based on these initial scales and rotations, the initial mean shape is set to $\overline{\mathbf{X}} = \frac{1}{n_s} \sum s_i \mathbf{R}_i \mathbf{D}_i$, and then $\mathbf{P}_N$, $\mathbf{Q}_N$, and $\mathbf{Q}$ are calculated. The reduced covariance matrix $\mathbf{\Sigma}_R$ and

---

[2]The performance comparison with the random rotation initialization is shown in the supplementary material.

the noise standard deviation σ are initialized as $10^{-3}\mathbf{I}$ and 10^{-3}, respectively.

Although EM-PND alone is good enough for solving NRSfM problems, it is a bit slow because it has to handle numerous parameters in each EM iteration. Therefore, by assuming $\sigma = 0$ and $\mathbf{\Sigma}_R = \mathbf{I}$ in (21), we solve the following simplified problem before the main EM procedure to reduce the computation time.

$$\min_{s_i, \mathbf{R}_i, \mathbf{z}_i, \overline{\mathbf{X}}} \quad \sum \|s_i \mathbf{R}_i \left(\mathbf{D}_i + \mathbf{L}_i(\mathbf{z}_i)\right) \mathbf{B} - \overline{\mathbf{X}}\|^2$$
$$\text{subject to} \quad \mathbf{R}_i^T \mathbf{R}_i = \mathbf{I}, \qquad \|\overline{\mathbf{X}}\|^2 = 1, \qquad (34)$$
$$s_i \operatorname{tr}\left(\mathbf{R}_i \left(\mathbf{D}_i + \mathbf{L}_i(\mathbf{z}_i)\right) \mathbf{B} \overline{\mathbf{X}}^T\right) = 1,$$

where $\mathbf{z}_i$ is the vector of the missing elements for the ith sample and $\mathbf{B} = \mathbf{I} - \frac{1}{n_p}\mathbf{1}\mathbf{1}^T$. $\mathbf{L}_i(\mathbf{z}_i) \in \mathbb{R}^{n_d \times n_p}$ is a mapping that places each element of $\mathbf{z}_i$ to the corresponding location in a shape matrix. It is noted that $\operatorname{vec}(\mathbf{L}_i(\mathbf{z}_i))$ can be expressed as $\operatorname{vec}(\mathbf{L}_i(\mathbf{z}_i)) = \widetilde{\mathbf{W}}\mathbf{z}_i$ where $\widetilde{\mathbf{W}}$ is a truncated version of $\left(\mathbf{I} - \widehat{\mathbf{W}}\right)$ removing all-zero columns. Similar to the update equation in the previous section, the alternating update equations for this problem are

$$\mathbf{z}_i = \left(\widetilde{\mathbf{W}}^T \left(\mathbf{B} \otimes \mathbf{I}\right) \widetilde{\mathbf{W}}\right)^+ \widetilde{\mathbf{W}}^T \operatorname{vec}\left(\frac{1}{s_i}\mathbf{R}_i^T \overline{\mathbf{X}} - \mathbf{D}_i\right),$$
$$\overline{\mathbf{X}} = \frac{\sum s_i \mathbf{R}_i \left(\mathbf{D}_i + \mathbf{L}_i(\mathbf{z}_i)\right) \mathbf{B}}{\|\sum s_i \mathbf{R}_i \left(\mathbf{D}_i + \mathbf{L}_i(\mathbf{z}_i)\right) \mathbf{B}\|},$$
$$\mathbf{R}_i = \mathbf{V}_i \mathbf{U}_i^T, \qquad s_i = 1/\operatorname{tr}(\mathbf{\Lambda}_i), \qquad (35)$$

where $\left(\mathbf{D}_i + \mathbf{L}_i(\mathbf{z}_i)\right) \mathbf{B} \overline{\mathbf{X}}^T = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^T$. Since the cost function of (34) is quite different from the cost function of EM-PND, we have to be careful not to perform this simplified iteration too many times. Hence, we have to find a good stopping criterion that takes the cost function of EM-PND into consideration. Considering that the dominant term in the cost function of EM-PND is $\log|\mathbf{\Sigma}_R|$, we have found out empirically that the following measure is appropriate.

$$\mathbf{c}_{\text{simple}} = \sum \max(\log(\lambda_j/\epsilon), 0), \qquad (36)$$

where λ_j is the jth eigenvalue of the sample covariance matrix of $\operatorname{vec}(s_i \mathbf{R}_i \left(\mathbf{D}_i + \mathbf{L}_i(\mathbf{z}_i)\right))$ and $\epsilon (= 10^{-7}$ in the experiments of Section 4) is a small constant.

The overall procedure of EM-PND is summarized as follows:

1. Normalize the input data as (20) and initialize all the parameters as in the first paragraph of Section 3.4.

2. Repeat (35) until $\left(\mathbf{c}_{\text{simple}}^{k-1} - \mathbf{c}_{\text{simple}}^k\right)/\mathbf{c}_{\text{simple}}^k < 5 \times 10^{-4}$, where k is the iteration number.

Table 1. Average reconstruction errors without noise and missing data

data \ method	EM-PPCA	MP	CSF2	SPM	EM-PND
FRGC	0.1469	0.1395	0.1926	0.1094	**0.0727**
walking	0.1485	0.2699	0.0708	0.0861	**0.0465**
shark	0.0688	0.0874	0.0551	0.5475	**0.0135**
face	0.0208	0.0329	0.0209	0.0233	**0.0165**
yoga	0.61	0.632	0.0226	0.0224	**0.014**
stretch	0.5393	0.5792	0.0219	0.0288	**0.0156**
pickup	0.5149	0.3465	0.0607	**0.0356**	0.0372
drink	0.1292	0.2707	0.0123	0.0216	**0.0037**
dance	0.2325	0.4838	**0.1349**	0.1454	0.1834

Table 2. Average reconstruction errors with noise and without missing data

data \ method	EM-PPCA	MP	CSF2	SPM	EM-PND
FRGC	0.198	0.1408	0.2061	0.184	**0.0889**
walking	0.1364	0.321	0.0966	–	**0.077**
shark	**0.0486**	0.118	0.1043	0.1784	0.06
face	0.0464	0.0523	0.0543	0.1054	**0.0403**
yoga	0.5287	0.6318	0.0529	0.0822	**0.0409**
stretch	0.5479	0.5806	0.0543	0.0652	**0.0444**
pickup	0.5037	0.3695	0.0705	0.0581	**0.0409**
drink	0.1764	0.2719	0.0365	0.0407	**0.0339**
dance	0.2229	0.4354	0.1544	**0.151**	0.1806

Table 3. Average reconstruction errors with missing data and without noise

data \ method	EM-PPCA	MP	CSF2	EM-PND
FRGC	0.1764	0.1431	0.4505	**0.0805**
walking	0.1361	0.2819	0.1033	**0.0469**
shark	0.1374	0.1077	0.0653	**0.0166**
face	0.2978	0.0456	0.0412	**0.0177**
yoga	0.1463	0.5768	0.0854	**0.0181**
stretch	0.6735	0.6149	0.0597	**0.015**
pickup	0.4969	0.4053	0.0933	**0.0149**
drink	0.1692	0.2612	0.0357	**0.0055**
dance	0.2632	0.3951	**0.1415**	0.1766

Table 4. Average reconstruction errors with noise and missing data

data \ method	EM-PPCA	MP	CSF2	EM-PND
FRGC	0.2196	0.1469	0.4574	**0.0968**
walking	0.1541	0.2853	0.1095	**0.0842**
shark	0.1323	0.1101	0.0872	**0.0672**
face	0.2273	0.0666	0.0583	**0.0464**
yoga	0.1401	0.5765	0.1583	**0.0488**
stretch	0.6881	0.6211	0.0707	**0.0535**
pickup	0.4931	0.401	0.0948	**0.0486**
drink	0.1868	0.2644	0.0428	**0.0408**
dance	0.2732	0.3986	**0.1501**	0.1601

3. Calculate $\mathbf{m}_i$ and $\mathbf{C}_i$ as in (27).

4. Update Φ using (30), (31), (32), and (33).

5. Repeat Steps 3 and 4 until $\left\| \overline{\mathbf{X}}^k - \overline{\mathbf{X}}^{k-1} \right\|^2 < 10^{-10}$.

After finishing EM-PND, the final $\mathbf{M}_i$ can be used as a reconstructed 3D shape. If a shape model rather than a reconstructed shape is needed, then $\overline{\mathbf{X}}$ and $\boldsymbol{\Sigma}$ can be used instead. Similarly, s_i and $\mathbf{R}_i$ can be used to represent a rigid motion.

4. Experimental Results

We performed NRSfM experiments using EM-PND for various data sets. We used the motion capture data sets from [2] and [11]. To see the performance for the data with no temporal dependence, we also manually annotated 62 landmarks in each of 400 3D depth samples in the Face Recognition Grand Challenge (FRGC) 2.0 Database [9]. Random scaling and rotation were applied to each of these samples to form a new 3D facial-landmark data with no temporal dependence.

We conducted experiments with and without noise and/or missing data. The standard deviation of the Gaussian noise was set as $\sigma_{\text{noise}} = 0.02 \max_{i,j,k}\{|d_{ijk}|\}$. For the missing data, we randomly set 30 percents of the landmarks as missing. We compared the propose method with some other state-of-the-art schemes, which were EM-PPCA [11], MP [8], CSF2 [5], and SPM [4]. The parameters of these methods were set in accordance with their original paper, and we used the block matrix method for SPM, which is known to be the best algorithm according to [4]. The performance was evaluated in terms of normalized reconstruction error, i.e.,

$$e_i = \|\widehat{\mathbf{X}}_i - \mathbf{X}_i^*\|/\|\mathbf{X}_i^*\|, \qquad (37)$$

where $\mathbf{X}_i^*$ and $\widehat{\mathbf{X}}_i$ is the ith ground truth and reconstructed shapes, respectively. Since a reconstructed shape have reflection ambiguity, we also measured the error for the inverted shape and picked the smaller error. All the experiments were repeated ten times and the results were averaged.

Tables 1-4 show the reconstruction errors under various conditions. There are no experimental results for missing data using SPM because SPM can not handle the case of missing data. When an algorithm does not converge, its result is denoted as "–". Among the 32 cases, excluding four cases of the dance sequence, EM-PND gives the best performance except for two cases, and even for these two cases, EM-PND gives the second best performance. For the case in the pickup sequence without noise and missing data, the error difference between EM-PND and the best method is about five percent. For the shark sequence, EM-PPCA gives a smaller error than EM-PND for the case of no missing data but with noise. This seems to be attributed due to the nature of the shark sequence, which was artificially generated by superposing two basis shapes [11]. Because of this, EM-PPCA, which explicitly limits the number of shape bases in the reconstruction process, gives better performance for the shark data. However, this is hardly the case for the real-world applications. The dance sequence includes a large deformation, and CSF2, which enforces temporal dependence between frames, gives better results than the other schemes. We expect that the performance of EM-PND for the dance sequence can also be improved by enforcing temporal dependence. In the other cases, EM-PND reduces the errors
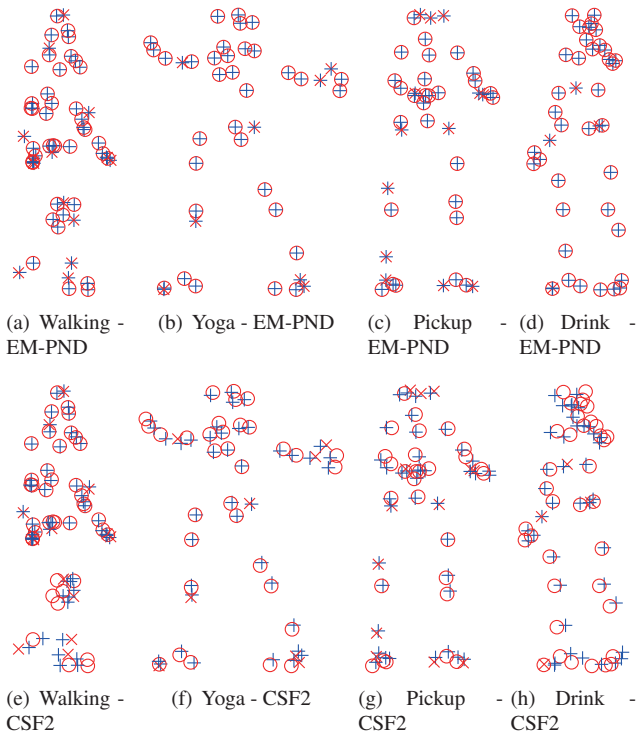
(a) Walking - EM-PND (b) Yoga - EM-PND (c) Pickup - EM-PND (d) Drink - EM-PND

(e) Walking - CSF2 (f) Yoga - CSF2 (g) Pickup - CSF2 (h) Drink - CSF2

Figure 1. Reconstructed results (top row: EM-PND, bottom row: the second best method, ○: observed ground truth, ×: missing ground truth, +: reconstructed points).

by up to 85 percents compared to the second best methods. Especially, EM-PND gives much better performance than CSF2 even though they use the same rotation initialization method. This indicates that the contribution of EM-PND to performance enhancement is large. This outstanding performance of EM-PND proves the effectiveness of using PND in NRSfM. Another advantage of EM-PND is that there is no parameter to be tuned, which will enable EM-PND to give superb performance in practical problems. Note that CSF2 gives relatively good performance, but not for the FRGC data because CSF2 assumes the existence of temporal dependence between frames. EM-PND shows a better fit between the reconstructed points and the corresponding ground truth than the second best method, as can be seen from the reconstruction results in Fig. 1. The videos of reconstructed shapes are also provided in the supplementary material to confirm the performance of EM-PND.

5. Conclusion

In this paper, we proposed EM-PND to solve NRSfM problems. Instead of rank constraints employed in the other methods, EM-PND imposes constraints on the motion parameters, following the practices in GPA, which makes the 3D shapes most closely aligned in a linear subspace. We defined PND based on this constraints, and used a slightly

modified version of the EM algorithm to fit input samples effectively to a PND. EM-PND gives state-of-the-art performance, as validated in the experimental results, by separating rigid and non-rigid variations and not using any rank constraint. Future work will consider the problems of adding temporal dependence into the model and designing a new factorization algorithm based on PND to reduce the computation time.

References

[1] I. Akhter, Y. Seikh, and S. Khan. In defense of orthonormality constraints for nonrigid structure from motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2009.

[2] I. Akhter, Y. Seikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(7):1442–1456, July 2011.

[3] R. L. Bishop and S. I. Goldberg. *Tensor Analysis on Manifolds*. Courier Dover Publications, 1980.

[4] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2012.

[5] P. F. U. Gotardo and A. M. Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2011.

[6] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, March 1975.

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 edition, 2004.

[8] M. Paladini, A. Del Bue, M. Stošić, M. Dodig, J. ao Xavier, and L. Agapito. Factorization for non-rigid and articulated structure using metric projections. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2009.

[9] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2005.

[10] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int'l J. Computer Vision*, 9(2):137–154, November 1992.

[11] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(5):878–892, May 2008.

[12] J. Valmadre and S. Lucey. General trajectory prior for nonrigid reconstruction. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2012.

[13] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. *Int'l J. Computer Vision*, 67(2):233–246, April 2006.

[14] M. L. Zelditch, D. L. Swiderski, H. D. Sheets, and W. L. Fink. *Geometric Morphometrics for Biologists: A Primer*. Elsevier/Academic Press, 2004.