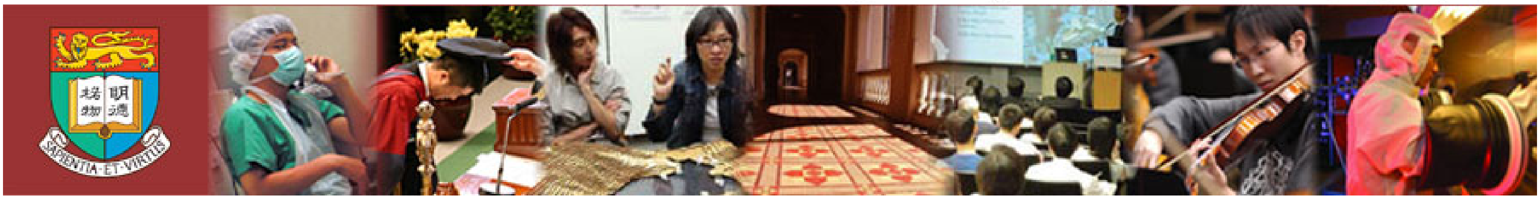

# A decomposition method for non-rigid structure from motion with orthographic cameras

**X. B. Zhang** **A. W. K. Tang** **Y. S. Hung**
Department of Electrical and Electronic Engineering, The University of Hong Kong
Postal address: Room 501, CYC Building, The University of Hong Kong, Hong Kong

## Abstract

*In this paper, we propose a new approach to non-rigid structure from motion based on the trajectory basis method by decomposing the problem into two sub-problems. The existing trajectory basis method requires the number of trajectory basis vectors to be specified beforehand, and then camera motion and the non-rigid structure are recovered simultaneously. However, we observe that the camera motion can be derived from a mean shape without recovering the non-rigid structure. Hence, the camera motion can be recovered as a sub-problem to optimize an error indicator without a full recovery of the non-rigid structure or the need to pre-define the number of basis required for describing the non-rigid structure. With the camera motion recovered, the non-rigid structure can then be solved in a second sub-problem together with the determination of the basis number by minimizing another error indicator. The solutions to these two sub-problems can be combined to solve the non-rigid structure from motion problem in an automatic manner, without any need to pre-define the number of basis vectors. Experiments show that the proposed method improves the reconstruction quality of both the non-rigid structure and camera motion.*

**Keywords**: non-rigid structure, orthographic camera, structure from motion, automatic recovery

## 1. Introduction

Structure from motion is one of the most important problems in computer vision. For a 3D structure projected to a set of cameras, the structure from motion problem is to recover the structure in 3D from the 2D image projections [1]. Traditionally, 3D structures are assumed to be rigid and stationary. Such an assumption incurs a rank-3 (rank-4 in perspective camera case) condition on image measurements. With such a condition, various methods have been proposed [2, 3], most of which are based on rank constrained factorization [4, 5].

In recent years, more and more attention is paid to the non-rigid structure from motion problem [6-8], where the 3D structure is allowed to move and deform. Based on an assumption that the deformation of a non-rigid structure can be modeled by a linear combination of a set of rigid shapes, the traditional factorization approach for rigid structure reconstruction has been extended to handle non-rigid structure recovery [9, 10]. As the non-rigid structure is represented as a linear combination of a shape basis, the method is referred to as the shape basis method in literature.

Then, Akhter et al. developed a trajectory basis method for non-rigid structure representation [11], which, as shown by the authors, is dual to the shape basis method. By tracking trajectories of corresponding points of a non-rigid structure and modeling them using a DCT (Discrete Cosine Transform) basis, the trajectory basis method recovers the structure in 3D space using not only the rank constraint, but also an implicit "smooth deforming trajectory" constraint. The introduction and enforcement of the "smooth deforming trajectory" constraint effectively prevents meaningless solutions and significantly reduce the gap between recovered structure and original structure. Despite so, the trajectory basis method has two persisting problems inherited from shape basis method:

a. the number of basis for non-rigid structure representation, which is normally unknown in advance, has to be pre-defined
b. there is no criteria for quality evaluation of the recovered structure and camera motion

In this paper, we proposed a new method based on trajectory basis representation that solves both of these two problems. By disassociating camera motion recovery with structure recovery and proposing a criterion for quality evaluation, we are able to obtain better solutions for camera matrices. At the same time, a criterion reflecting the error of fitting a non-rigid structure using trajectory basis representation with different number of basis is proposed, leading to a method for automatic determination of the best basis number for non-rigid structure representation.

The rest of the paper is organized as follows: In Section 2, some preliminaries about trajectory basis method are introduced, together with notations used in this paper. In Section 3, the non-rigid structure from motion problem with orthographic cameras is reformulated and a new method is proposed. In Section 4, experimental evaluations are presented, including comparisons with existing algorithms. Some concluding remarks are given in Section 5.

# 2. Trajectory basis for non-rigid structure from motion

## 2.1. Non-rigid structure from motion

Let $\{\boldsymbol{x}_{ij} \in R^{2\times 1} \mid i = 1,2,\cdots,F, j = 1,2,\cdots,N\}$ be the orthographic projections of $N$ 3D points $\{\boldsymbol{X}_{ij} \in R^{3\times 1}\}$ projected to $F$ frames of a moving camera. Then

$$\boldsymbol{x}_{ij} = \mathrm{R}_i \boldsymbol{X}_{ij} \tag{1}$$

where $\mathrm{R_i} \in \mathrm{R}^{2\times 3}$ is the camera matrix associated with the $i$th frame. It follows that

$$\mathrm{W} = \begin{bmatrix} \boldsymbol{x}_{11} & \cdots & \boldsymbol{x}_{1N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}_{F1} & \cdots & \boldsymbol{x}_{FN} \end{bmatrix} = \begin{bmatrix} \mathrm{R}_1\mathrm{X}_1 \\ \vdots \\ \mathrm{R}_F\mathrm{X}_F \end{bmatrix} \tag{2}$$

where $\mathrm{X}_i = [\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \cdots, \boldsymbol{X}_{iN}] \in R^{3\times N}$ is called structure matrix and W is called measurement matrix.

The structure from motion problem refers to the problem of recovering camera matrix $\mathrm{R}_i$ and structure matrices $\mathrm{X}_i$ given the measurement matrix W.

For a rigid stationary object, the structure matrices $\mathrm{X}_i$ are equal across all frames. Thus a rank-3 constraint can be enforced in the measurement matrix, which is the basis for all factorization based algorithms for solving the rigid structure from motion problem.

For a non-rigid object, the rank-3 constraint does not hold anymore. Thus other constraints are necessary in order to make the problem solvable. A common assumption made is that the structure deformation can be modeled by a linear combination of a fixed set of $K$ shape basis $\mathrm{B}_k \in R^{3\times N}$ $(k = 1,2,\cdots,K)$. Formally, the assumption can be written as

$$\boldsymbol{X}_{ij} = \sum_{k=1}^{K} c_{ik}\boldsymbol{B}_{kj}, \qquad i = 1,2,\cdots,F, j = 1,2,\cdots,N \tag{3}$$

where $\mathrm{B}_k = [\boldsymbol{B}_{k1}, \boldsymbol{B}_{k2}, \cdots, \boldsymbol{B}_{kN}]$ and $c_{ik}$ is the coefficient of the $i$th frame at the $k$th shape basis.

Such an assumption imposes a rank-$3K$ constraint on the measurement matrix and thus makes it possible to extend the factorization approach from rigid structure to non-rigid structure recovery.

## 2.2. Trajectory basis for non-rigid structure representation

Trajectory basis representation can be regarded as the dual of the shape basis representation by taking $c_{ik}$ in equation (3) as the $k$th trajectory basis entry at the $i$th frame and $\boldsymbol{B}_{kj}$ as a corresponding vector of weighting coefficients. The trajectory basis method has been demonstrated to be more stable in non-rigid structure recovery by restricting the recovered non-rigid structure to be smoothly deforming in consecutive frames. Moreover, the trajectory basis is pre-defined and is independent of the non-rigid structure dataset, thus reducing the parameters to be solved in the optimization problem.

With trajectory basis method, the measurement matrix in equation (2) can be rewritten as

$$\mathrm{W} = \begin{bmatrix} \mathrm{R}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{R}_F \end{bmatrix} \begin{bmatrix} c_{11}\mathrm{I}_3 & \cdots & c_{1K}\mathrm{I}_3 \\ \vdots & \ddots & \vdots \\ c_{F1}\mathrm{I}_3 & \cdots & c_{FK}\mathrm{I}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{B}_{K1} & \cdots & \boldsymbol{B}_{KN} \end{bmatrix} \tag{4}$$

$$= \begin{bmatrix} \boldsymbol{c}_1 \otimes \mathrm{R}_1 \\ \vdots \\ \boldsymbol{c}_F \otimes \mathrm{R}_F \end{bmatrix} \begin{bmatrix} \mathrm{B}_1 \\ \vdots \\ \mathrm{B}_K \end{bmatrix} \tag{5}$$

$$= \mathrm{MB} \tag{6}$$

where $\boldsymbol{c}_i = [c_{i1}, c_{i2}, \cdots, c_{iK}] \in R^{1\times K}$ is a row vector consisted of trajectory basis entries of the $i$th frame, $\mathrm{M} \in R^{2F\times 3K}$ represents a scaled motion matrix and $\mathrm{B} \in R^{3K\times N}$ is a coefficient matrix.

# 3. Non-rigid structure and motion recovery

## 3.1. Problem decomposition

Equation (6) suggests that the measurement matrix of a non-rigid structure with orthographic cameras is subject to a rank-$3K$ constraint, and thus both structure and camera matrices can be obtained using a factorization method similar to that for rigid structure reconstruction. The idea can be summarized with the following equations.

$$\mathrm{W} = \widehat{\mathrm{M}}\,\widehat{\mathrm{B}} \tag{7}$$

$$= \widehat{\mathrm{M}}\,\mathrm{GG}^{-1}\,\widehat{\mathrm{B}} \tag{8}$$

where W is first factorized into rank 3K matrices $\widehat{\mathrm{M}}$ and $\widehat{\mathrm{B}}$ in a projective space, and then a metric upgrade is performed by finding $\mathrm{G} \in R^{3K\times 3K}$ that satisfies

$$\widehat{\mathrm{M}}_i\mathrm{G} = \boldsymbol{c}_i \otimes \widehat{\mathrm{R}}_i, i = 1,2,\cdots,F \tag{9}$$

where $\widehat{\mathrm{R}}_i \in R^{2\times 3}$ represents recovered camera matrix of the $i$th frame with row-orthonormal properties.

An inherent problem with the above method is that the number of basis required for non-rigid structure description needs to be specified beforehand. Notice however that the camera matrices in (4) are independent of the structure model in 3D space. It may be possible to recover camera matrices without the recovery of non-rigid structure. Also, the recovery of 3D structure is independent of the method used for camera motion recovery. As long as camera matrices are given, structure in 3D space can be recovered optimally from 2D measurements under the trajectory basis model.

Hence, we propose to decompose the non-rigid structure from motion problem into two sub-problems. The first one is to recover camera motion, and the second problem is to recover the structure in 3D space when camera motion is already known.

The benefit of the problem decomposition is that optimal solutions can be obtained in both of these two sub-problems. Furthermore, both sub-problems can be solved in an automatic manner with the help of error indicators that may not be available when solving the two sub-problems as a whole.

## 3.2. Recovery of camera projection matrices

If the trajectory basis in (4) is generated using DCT, the first basis vector is given by $[c_{11}, c_{21}, \cdots, c_{F1}]^T = \frac{1}{\sqrt{F}}[1,1,\cdots,1]^T$. Hence we may rewrite (5) as:

$$\mathrm{W} = \begin{bmatrix} \mathrm{R}_1 \\ \vdots \\ \mathrm{R}_F \end{bmatrix} \frac{\mathrm{B}_1}{\sqrt{F}} + \begin{bmatrix} \boldsymbol{c}_{1,2\cdots K} \otimes \mathrm{R}_1 \\ \vdots \\ \boldsymbol{c}_{F,2\cdots K} \otimes \mathrm{R}_F \end{bmatrix} \begin{bmatrix} \mathrm{B}_2 \\ \vdots \\ \mathrm{B}_K \end{bmatrix} \quad (10)$$

$$= \mathrm{RS} + \mathrm{D} \quad (11)$$

where $\boldsymbol{c}_{i,2\cdots K} = [c_{i2}, c_{i3}, \cdots, c_{iK}]$, R consists of the stacked rotation matrices, and S and D are defined in an obvious manner.

We may interpret (11) as decomposing W into two components: (i) the projection of a mean shape S with orthographic camera matrices in R, and (ii) the projection of deformations D of the non-rigid structure. Furthermore, we note that the camera matrices can be recovered from the first component (i) alone if W is decomposed as in (11). In view of (11), we propose to recover the camera matrices by a partial upgrade of only the mean shape component of the non-rigid structure, by writing (7) – (9) as:

$$\mathrm{W} = \widehat{\mathrm{M}}\,\widehat{\mathrm{B}} \quad (12)$$

$$= \widehat{\mathrm{M}}[\mathrm{G}_{13} \quad \mathrm{G}_{4K}][\mathrm{G}_{13} \quad \mathrm{G}_{4K}]^{-1}\,\widehat{\mathrm{B}} \quad (13)$$

$$= [\widehat{\mathrm{M}}\mathrm{G}_{13} \quad \widehat{\mathrm{M}}\mathrm{G}_{4K}] \begin{bmatrix} \widehat{\mathrm{S}} \\ \widehat{\mathrm{V}} \end{bmatrix} \quad (14)$$

$$= [\widehat{\mathrm{R}} \quad \widehat{\mathrm{M}}\mathrm{G}_{4K}] \begin{bmatrix} \widehat{\mathrm{S}} \\ \widehat{\mathrm{V}} \end{bmatrix} \quad (15)$$

$$= \widehat{\mathrm{R}}\widehat{\mathrm{S}} + \widehat{\mathrm{M}}\mathrm{G}_{4K}\widehat{\mathrm{V}} \quad (16)$$

subject to

$$\widehat{\mathrm{M}}_i \mathrm{G}_{13}\mathrm{G}_{13}^T\widehat{\mathrm{M}}_i^T = \mathrm{I}_2, \qquad i = 1,2,\cdots,F \quad (17)$$

where W is factorized into two rank-K matrices $\widehat{\mathrm{M}} \in R^{2F\times K}$ and $\widehat{\mathrm{B}} \in R^{K\times N}$, $\mathrm{G}_{13} \in R^{K\times 3}$ and $\mathrm{G}_{4K} \in R^{K\times(K-3)}$ are the 1st-3rd columns and 4th-$K$th columns of an upgrade matrix, respectively, $\widehat{\mathrm{R}} \in R^{2F\times 3}$ is the recovered stacked camera matrix, $\widehat{\mathrm{S}} \in R^{3\times N}$ is mean shape of the 3D structure, and $\widehat{\mathrm{V}} \in R^{(K-3)\times N}$ is a matrix spanning the deformation space of the non-rigid structure.

Equation (16) suggests that camera matrices with non-rigid structure projections can be obtained by a rank-K factorization followed by applying a metric upgrading matrix $\mathrm{G}_{13} \in \mathrm{R}^{\mathrm{K}\times 3}$ which is subject to a non-linear constraint stated in equation (17). An algorithm to solve for camera matrices is described in Table 1.

Table 1: Algorithm for camera motion recovery

<u>Objective</u>

Given a set of image measurements of a non-rigid structure projected by orthographic cameras and a factorization rank $K$, compute a set of rotation matrices that enforce condition (17).

<u>Algorithm 1</u>

**Camera motion recovery**

1. Factorize measurement matrix with equation (12).
2. Find a triple column metric upgrading matrix $\mathrm{G}_{13}$ that satisfying equation (17).
3. Output stacked rotation matrices $\widehat{\mathrm{R}} = \widehat{\mathrm{M}}\mathrm{G}_{13}$

At a given factorization rank, Algorithm 1 looks for a set of rotation matrices satisfying the orthonormality constraint stated in (17) which is shown to be sufficient to recover camera projection matrices [12]. As factorization rank in (12) is unknown, it is necessary to identify the best factorization rank for each dataset. To achieve such a goal, we propose to employ 2D reprojection error as an error indicator at different factorization rank, as described in the next subsection.

## 3.3. Error indicator using 2D reprojection error

Using different factorization rank in (12) results in different level of approximation to the non-rigid structure. The approximation of non-rigid structure leads to errors in the mean shape, and thus affects the error of recovered camera matrices. Hence it is necessary to have an indicator reflecting the quality of recovered camera matrices at different factorization rank. Noting that the error of recovered camera matrices based on (16) is correlated to the reprojection error evaluated with a full projection model shown in (5), we propose to take the difference norm of equation (5) as an error indicator for recovered projection matrices. The key problem of evaluating the difference norm of equation (5) is that the coefficient matrix $\widehat{\mathrm{B}}$ is unknown. Here we use a coarse to fine approach for obtaining the coefficient matrix. Let $\boldsymbol{c}^k \in R^{F\times 1}$ be the $k$th trajectory basis vector, and

$$K_m = \lfloor \mathrm{rank}(\mathrm{W})/3 \rfloor \quad (18)$$

be the maximum number of basis that can be chosen for current non-rigid structure description where $\lfloor a \rfloor$ means the maximum integer $\leq a$. The coefficient matrix $\widehat{\mathrm{B}}$ can be obtained by iteratively solving the following problem:

$$\widehat{\mathrm{B}}_k = \arg\min_{\widehat{\mathrm{B}}_k} \left\| \mathrm{W}_r^k - [diag(\boldsymbol{c}^k) \otimes \mathrm{I}_2]\widehat{\mathrm{R}}\widehat{\mathrm{B}}_k \right\|^2 \quad k = 1,2,\cdots,K_m \quad (19)$$

where

$$\mathrm{W}_r^1 = \mathrm{W} \quad (20)$$

$$\mathrm{W}_r^{k+1} = \mathrm{W}_r^k - [diag(\boldsymbol{c}^k) \otimes \mathrm{I}_2]\widehat{\mathrm{R}}\widehat{\mathrm{B}}_k \quad k = 1,2,\cdots,K_m - 1 \quad (21)$$

Thus error indicator for recovered camera matrices using 2D reprojection error can be defined as

$$e = \sum_{i=1}^{F} \left\| \mathrm{W}_i - [\boldsymbol{c}_i \otimes \widehat{\mathrm{R}}_i] \widehat{\mathrm{B}} \right\|^2 \qquad (22)$$

Although the correct number of basis remains unknown, we add basis vectors one by one and solve for optimal coefficients for each newly added basis vector until the maximum number of basis allowed is reached. This procedure has the effect of avoiding over-fitting caused by unnecessary basis vectors because the optimization of equation (19) would not disturb the trajectory coefficients that have already been recovered using a smaller number of basis vectors. Hence, $e$ is defined without the need to specify the number of basis, and is an indicator of the quality of the camera matrices alone.

## 3.4. Cross validation for automatic basis number decision

Given rotation matrices, the problem of non-rigid structure recovery using a trajectory basis model is to find suitable coefficient matrix $\widehat{\mathrm{B}}$ that solves the following problem:

$$\min_{\widehat{\mathrm{B}}} \sum_{i=1}^{F} \left\| \mathrm{W}_i - [\boldsymbol{c}_i \otimes \widehat{\mathrm{R}}_i] \widehat{\mathrm{B}} \right\|^2 \qquad (23)$$

The key issue in solving the above problem is that the number of basis $K_b$ for the non-rigid structure description is unknown. As long as $K_b$ is defined, the above problem can be solved with standard least square techniques. In order to decide the best number of basis for the non-rigid structure representation, it is necessary to have an error indicator signifying the quality of recovered non-rigid structure at different number of basis.

Similar to [13], we use cross validation score as an error indicator. The idea is to partition the image measurements of each frame into training data and testing data. While the collection of training data is used to recover the non-rigid structure; that of testing data is used to quantify how well the recovered non-rigid structure is by evaluating the distance between testing data measurements and testing data reprojections. The cross validation score is taken to be the average distance of several such partitions. An algorithm for automatic basis number decision is given in Table 2.

Table 2: Algorithm for automatic basis number decision

Objective

Given a set of image measurements of a non-rigid structure and a set of camera matrices, compute a suitable number of trajectory basis such that its cross validation score is the smallest.

Algorithm 2

**Automatic bases number decision**

1. Randomly partition measurements of each frame into training data and testing data for $K_p$ times.
2. For $k = 2 : K_m$
   a) Define trajectory basis matrix c with $k$ basis vectors.
   b) For each partitioned dataset, solving for coefficient matrix $\widehat{\mathrm{B}}$ in (23) with the collection of training data.
   c) Obtain non-rigid structure using recovered coefficients and defined trajectory basis.
   d) Reproject the recovered non-rigid structure onto images; evaluate the distance between measurements of testing data and reprojections of testing data.
   e) Average the distance over $K_p$ trials and record it as cross validation score $s$.
3. Find the smallest cross validation score $s$ and output its corresponding basis number.

Table 3: Algorithm for non-rigid structure recovery

Objective

Given a set of image measurements of a non-rigid structure projected by orthographic cameras, reconstruct the non-rigid structure in 3D space.

Algorithm 3

**Non-rigid structure recovery**

1. For $\mathrm{k} = 3 : \mathrm{rank(W)}$
   a) Seek for rotation matrices $\widehat{\mathrm{R}}_k$ using Algorithm 1 with rank-$k$ factorization.
   b) Evaluate the 2D reprojection error $e$ using equation (22) with rotation matrices $\widehat{\mathrm{R}}_k$.
2. Select rotation matrices $\widehat{\mathrm{R}}$ with smallest 2D reprojection error $e$.
3. Find the optimal basis number $K_b$ using Algorithm 2 with recovered rotation matrices $\widehat{\mathrm{R}}$.
4. Define trajectory basis matrix $\mathrm{c}_b$ with $K_b$ basis vectors for non-rigid structure description.
5. Solve for optimal coefficients $\widehat{\mathrm{B}}$ by solving (23) with camera matrices $\widehat{\mathrm{R}}$ and trajectory basis matrix $\mathrm{c}_b$.
6. Evaluate equation (3) with $\mathrm{c}_b$ and $\widehat{\mathrm{B}}$, and obtain the non-rigid structure in 3D space.

## 3.5. Algorithm for non-rigid structure recovery

The solutions to the two sub-problems, namely Algorithms 1 and 2, can now be combined to recover the non-rigid structure using trajectory basis representation in an automatic manner. An overview of the algorithm for non-rigid structure recovery is given in Table 3.

# 4. Experimental results

The proposed method is evaluated with both synthetic images of non-rigid structures and images of deforming structures in the real world. Comparisons with existing trajectory basis method are also made in this section.

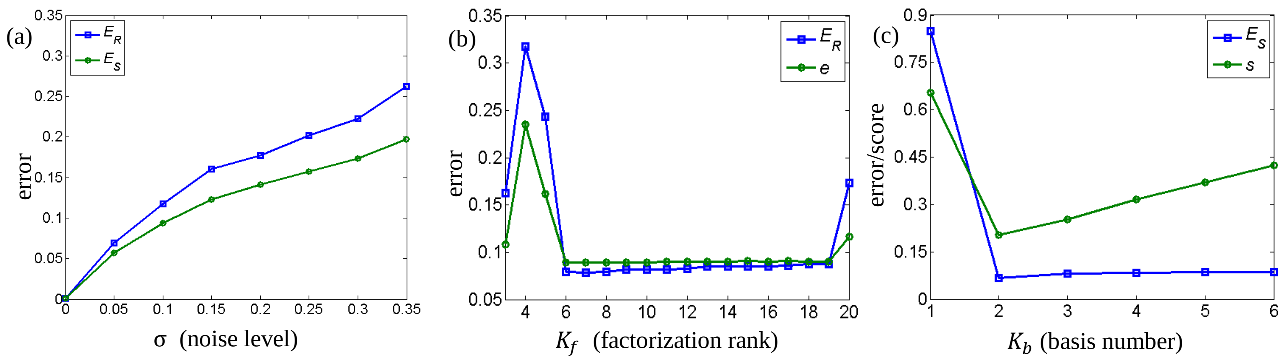

Figure 1: Algorithm performance evaluation with synthetic data. (a) Algorithm performance with regard to noise contamination of measurement matrix. (b) Correlation between the error of rotation matrices and the 2D reprojection error evaluated using equation (22). (c) Correlation between structure error and cross validation score.

### 4.1. Performance on synthetic non-rigid structure

A total of 20 points are randomly generated in a $1.0 \times 1.0 \times 1.0$ cube in 3D space. The points are allowed to move in such a way that their locations are defined by equation (3) using trajectory basis model with $K = 2$. Around those points, 36 randomly positioned orthographic cameras are pointing at them, producing 36 images of size $1.0 \times 1.0$. Each image contains a set of measurements of the non-rigid structure. Those measurements are stacked together forming a noise-free measurement matrix. Different levels of Gaussian noise (standard deviation σ ranging from 0 to 0.35) are added to the measurement matrix, which is then used for motion and structure recovery.

Let $\mathrm{R}_i$ and $\widehat{\mathrm{R}}_i$ be the ground truth and recovered camera matrix of the $i$th image frame, respectively, and $\mathrm{S}_i$ and $\hat{\mathrm{S}}_i$ be ground truth structure and recovered structure at the $i$th frame, respectively. We define the rotation matrix error $E_{\mathrm{R}}$ and structure error $E_{\mathrm{S}}$ as

$$E_{\mathrm{R}} = \min_{R} \sqrt{\frac{1}{F}\sum_{i=1}^{F}\left\|\mathrm{R}_i - \widehat{\mathrm{R}}_i R\right\|^2} \tag{24}$$

$$E_{\mathrm{S}} = \min_{R} \sqrt{\frac{1}{F \cdot N}\sum_{i=1}^{F}\left\|\mathrm{S}_i - R\hat{\mathrm{S}}_i\right\|^2} \tag{25}$$

where $R$ is a 3×3 matrix aligning recovered camera (or structure) matrices with ground truth.

Figure 1(a) shows the proposed algorithm's performance with regard to noise. At each noise level, a total of 30 trials are performed, and their average errors are plotted in Figure 1(a). It shows that both camera matrix error and structure error increase almost linearly with regard to noise contamination in the measurement matrix, indicating the structure and motion can be recovered robustly.

The camera matrix error recovered using Gaussian noise (σ=0.05) contaminated measurement matrix at different factorization rank is shown in Figure 1(b) in blue squares, together with 2D reprojection error shown in green circles. The correlation between 2D reprojection error and camera matrix error is evident, and thus the best factorization rank can be identified using 2D reprojection error, in case of real images where the camera error is unknown. In this example, as the measurements are generated exactly using equation (5), it is expected that the best factorization rank is $3K$.

For a given set of camera matrices (recovered from $\sigma = 0.05$ Gaussian noise contaminated measurements with factorization rank $K_f = 6$), the structure error for different number of trajectory basis vectors is shown in Figure 1(c). Also shown in the figure is the cross validation score which indicates the quality of recovered non-rigid structures. It can be seen that cross validation score correlates with structure error well and thus is a good indicator of the quality of recovered non-rigid structure.

### 4.2. Real non-rigid structure

The proposed algorithm is also quantitatively evaluated with images of deforming object in the real world. Datasets containing real world object deformations are obtained from the project website of Akhter et al. [14]. In each dataset, a sequence of synthetic orthographic cameras are rotating 5 degrees per frame around the z-axis, pointing to the object and generating image measurements. In our experiments, noises are added in such a way that the standard deviation of Gaussian noise is 5% of the standard deviation of measurement matrix. And, in order to make error comparison more meaningful, non-rigid structure is centroid removed and normalized with standard deviation being equal to 1 before evaluating structure error using equation (25).

Table 4 shows a quantitative comparison with existing trajectory basis method [6] whose code is provided by the authors at their project website [14]. Both recovered

Table 4: Quantitative evaluation of the proposed method with datasets containing real world deformations

| Dataset | noise free | | | | | | | 5% Gaussian noise | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Proposed method | | | | Method of [6] | | | Proposed method | | | | Method of [6] | | |
| | $E_R$ | $E_S$ | $K_f$ | $K_b$ | $E_R$ | $E_S$ | $K$ | $E_R$ | $E_S$ | $K_f$ | $K_b$ | $E_R$ | $E_S$ | $K$ |
| Drink | 0.0052 | 0.0247 | 33 | 13 | 0.0058 | 0.0250 | 13 | 0.0292 | 0.0544 | 9 | 13 | 0.0335 | 0.053 | 13 |
| PickUp | 0.1363 | 0.2129 | 13 | 10 | 0.1549 | 0.2369 | 12 | 0.1465 | 0.2331 | 35 | 12 | 0.1477 | 0.2315 | 12 |
| Yoga | 0.0792 | 0.1125 | 5 | 10 | 0.1059 | 0.1622 | 11 | 0.0796 | 0.1238 | 5 | 8 | 0.1263 | 0.1801 | 11 |
| Stretch | 0.0487 | 0.0702 | 37 | 11 | 0.0549 | 0.1088 | 12 | 0.0785 | 0.1317 | 39 | 11 | 0.0861 | 0.1516 | 12 |

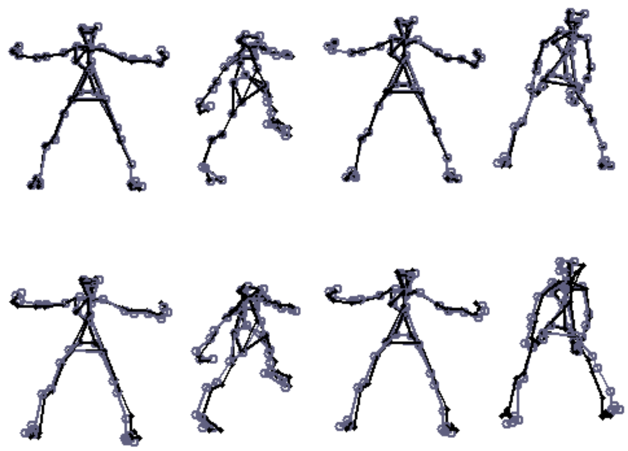

Figure 2: Reconstructed Yoga structures using 5% Gaussian noise contaminated image measurements. The upper row is the structure obtained by the proposed decomposition method, and the lower row is structure obtained by method of [6].

rotation matrices and recovered structure are compared with ground truth, and the errors are evaluated using equation (24) and (25). Also, in Table 4 we show the factorization rank $K_f$ and basis number $K_b$ that are generated by the proposed method; whereas the basis number $K$ for the method of [6] is suggested by the authors (with the same basis number chosen in noisy case).

An example of recovered non-rigid structure (shown in grey circles) using noise contaminated Yoga dataset is shown in Figure 2, with ground truth structure (shown in black dots) superimposed. The upper row of Figure 2 is the structure recovered using the proposed decomposition method, and the lower row is the structure recovered using the method of [6].

# 5. Conclusions

In this paper, we proposed a new method to recover camera motion and non-rigid structure with a trajectory basis representation for the non-rigid structure. By decomposing the problem into two sub-problems and solving for optimal solution to each sub-problem, the method first recovers camera motion without the need to pre-define the basis number. Then, with recovered camera motion, the method finds the best number of basis that should be used for non-rigid structure representation. Hence, the proposed method leads to a completely automatic algorithm for non-rigid structure reconstruction. Experiments demonstrate that the method improves the reconstruction quality of both the non-rigid structure and the camera motion.

# Acknowledgements

The work described in this paper was supported by a grant from the Research Grants Council (project no. HKU 712911E) of the Hong Kong Special Administrative Region, China and CRCG of the University of Hong Kong.

# References

[1] R. I. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2nd ed.: Cambridge University Press, 2004.

[2] Y. S. Hung and W. K. Tang, "Projective reconstruction from multiple views with minimization of 2D reprojection error," *International Journal of Computer Vision,* vol. 66, pp. 305-317, 2006.

[3] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," in *European Conference on Computer Vision*, 1996, pp. 709-720.

[4] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision,* vol. 9, pp. 137-154, 1992.

[5] B. Triggs, "Factorization methods for projective structure and motion," in *Computer Vision and Pattern Recognition*, 1996, pp. 845-851.

[6] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Trajectory space: a dual representation for nonrigid structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, pp. 1442-1456, 2011.

[7] M. Brand and R. Bhotika, "Flexible flow for 3D nonrigid tracking and shape recovery," in *Computer Vision and Pattern Recognition*, 2001, pp. I-315-I-322 vol.1.

[8] J. Xiao, J. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," *International Journal of Computer Vision,* vol. 67, pp. 233-246, 2006.

[9] M. Brand, "A direct method for 3D factorization of nonrigid motion observed in 2D," in *Computer Vision and Pattern Recognition*, 2005, pp. 122-128 vol. 2.

[10] R. Hartley and R. Vidal, "Perspective nonrigid shape and motion recovery," in *European Conference on Computer Vision*, Marseille, France, 2008, pp. 276-289.

[11] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Nonrigid structure from motion in trajectory space," in *Neural Information Processing Systems*, 2008.

[12] I. Akhter, Y. Sheikh, and S. Khan, "In defense of orthonormality constraints for nonrigid structure from motion," in *Computer Vision and Pattern Recognition*, 2009, pp. 1534-1541.

[13] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd, "Coarse to fine low rank structure from motion," in *Computer Vision and Pattern Recognition*, 2008, pp. 1-8.

[14] http://cvlab.lums.edu.pk/nrsfm/