Non-Rigid Shape and Motion Recovery: Degenerate Deformations

Jing Xiao **Takeo Kanade**

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{jxiao, tk}@cs.cmu.edu

Abstract

This paper studies the problem of 3D non-rigid shape and motion recovery from a monocular video sequence, under the degenerate deformations. The shape of a deformable object is regarded as a linear combination of certain shape bases. When the bases are non-degenerate, i.e. of full rank 3, a closed-form solution exists by enforcing linear constraints on both the camera rotation and the shape bases [18]. In practice, degenerate deformations occur often, i.e. some bases are of rank 1 or 2. For example, cars moving or pedestrians walking independently on a straight road refer to rank-1 deformations of the scene. This paper quantitatively shows that, when the shape is composed of only rank-3 and rank-1 bases, i.e. the 3D points either are static or independently move along straight lines, the linear rotation and basis constraints are sufficient to achieve a unique solution. When the shape bases contain rank-2 ones, imposing only the linear constraints results in an ambiguous solution space. In such cases, we propose an alternating linear approach that imposes the positive semi-definite constraint to determine the desired solution in the solution space. The performance of the approach is evaluated quantitatively on synthetic data and qualitatively on real videos.

1. Introduction

Recovery of 3D shape and motion from a monocular video sequence is an important task for applications like human computer interaction and robot navigation. The decades of work has led to significant successes on this problem. When the scene is static, reliable systems exist for 3D reconstruction of the scene structure. In reality, many scenes are dynamic and non-rigid: expressive faces, cars moving beside buildings, etc. Such scenes often deform with a class of basis structures. For example, the shape of a face can be regarded as a weighted sum of some shape bases, which correspond to various facial expressions [3].

Bregler and his colleagues [5] first introduced the basis representation to the problem of non-rigid structure from motion. Using this representation, in [18], we presented two sets of linear metric constraints, orthonormality constraints on camera rotations (*rotation constraints*) and uniqueness constraints on shape bases (*basis constraints*). We proved that, when the shape deformation is non-degenerate, *i.e.* all bases are of full rank 3, enforcing the linear constraints leads to a closed-form solution [18]. In practice, many scenes deform with degenerate bases of rank 1 or 2. Such bases limit the shape to deform only in a 2D plane. For instance, if a scene contains pedestrians walking independently along straight lines, the bases referring to those rank-1 translations are degenerate. A simple illustration of rank-3, 2, and 1 bases is shown in Figure 1. Under degenerate deformations, enforcing the linear metric constraints is not necessarily sufficient to determine a unique solution.

This paper demonstrates that, when the shape involves rank-2 bases, the linear constraints leads to an ambiguous solution space that contains invalid solutions. The degree of freedom of the space is determined by the number of the rank-2 bases. Under such situations, we show that a valid solution in the space must be positive semi-definite. We then present an alternating linear optimization approach that combines the linear metric constraints and the positive semi-definite constraint to determine the desired solution. When the shape bases are of either rank 3 or rank 1, *i.e.* all the 3D points in the scene either are static or independently move along straight lines, the linear metric constraints provide a unique solution to reconstructing the dynamic scene structure and camera motion. Note that such special degenerate deformations often occur in real applications. For example, when several people walk independently along different directions, each of the independent motions refers to a shape basis and all of them are of rank 1. Most of previous approaches [1, 7, 17] on degenerate deformations were proposed for this special case. However they require either the moving velocities are constant [7, 17] or the camera projection matrices are given [1].

2. Previous Work

The problem of 3D shape and motion recovery from 2D image sequences has attracted a lot of attention. Various

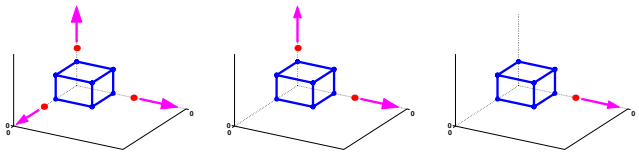

Figure 1: (Left): Three points (red) simultaneously move along fixed directions in the 3D space. Their trajectories form a deformation basis of rank 3. (Middle): Two points move along fixed directions within a 2D plane. Their trajectories form a rank-2 shape basis. (Right): One point move along a fixed direction. Its trajectory forms a rank-1 basis.

approaches have been proposed for different applications [12, 10, 18]. Our discussion will focus on the factorization methods that are closely related to our work.

The factorization method was first proposed by Tomasi and Kanade [12]. First it applies the rank constraint to factorize a set of feature locations tracked across the entire sequence. Then it uses the orthonormality constraints on the camera rotations to reconstruct the shape and motion in one step. This approach and its extensions to various camera projection models [9, 14] work for static scenes.

Costeira and Kanade [6] proposed a method that factorizes the image measurement to segment multiple independently moving objects and individually recover their shapes. Wolf and Shashua [16] derived a geometrical constraint, called the segmentation matrix, to reconstruct a scene containing two independently moving objects from two perspective views. Vidal and his colleagues [15] generalized this approach to the case of multiple independently moving objects. For reconstruction of scenes consisting of both static objects and objects moving along fixed directions, Han and Kanade [7] proposed a factorization method that achieves a unique solution assuming constant velocities. A more generalized solution to reconstructing the shapes that deform at constant velocity is presented in [17].

Bregler and his colleagues [5] first introduced the basis representation of non-rigid shapes to embed the deformation constraints into the scene structure. By analyzing the low rank of the image measurements, they enforce the orthonormality constraints on camera rotations to factorize the non-rigid shape and motion. This method was extended to the nonlinear optimization approaches in [13, 4]. These three methods impose only the constraints on rotations. In [18], we proved that enforcing only the rotation constraints leads to ambiguous and invalid solutions. We then introduced the uniqueness constraints on the shape bases and proved that imposing both the basis and the rotation constraints results in a linear closed-form solution, assuming the shape deformations are non-degenerate [18]. To reconstruct the degenerate deformations, most of previous approaches [1, 7, 17] assume strong prior knowledge on either shape or motion. The methods in [7, 17] require that the deformation velocity is constant. The method in [1] assumes that the trajectory of each 3D point is either a straight line or a conic and the camera projection matrices are all given.

3. Problem Statement

Given 2D locations of P feature points across F frames, $\{(u,v)_{fp}^T | f = 1,...,F, p = 1,...,P\}$, our goal is to recover the motion of the non-rigid object relative to the camera, including rotations $\{R_f | f = 1,...,F\}$ and translations $\{\mathbf{t}_f | f = 1,...,F\}$, and its 3D deforming shapes $\{(x,y,z)_{fp}^T | f = 1,...,F, p = 1,...,P\}$, under the assumption of weak-perspective projection model.

We follow the representation of [3, 5]. The non-rigid shape is represented as linear combination of K shape bases $\{B_i, i = 1,...,K\}$. The bases are $3\times P$ matrices controlling the deformation of P points. Then the 3D coordinate of the point p at the frame f is,

$$\mathbf{X}_{fp} = (x,y,z)_{fp}^T = \Sigma_{i=1}^K c_{fi}\mathbf{b}_{ip} \tag{1}$$

where $\mathbf{b}_{ip}$ is the p_{th} column of B_i and c_{if} is its combination coefficient at the frame f. The image coordinate of $\mathbf{X_{fp}}$ under the weak perspective projection model is,

$$\mathbf{x}_{fp} = (u,v)_{fp}^T = s_f(R_f \cdot \mathbf{X}_{fp} + \mathbf{t}_f) \tag{2}$$

where R_f stands for the first two rows of the f_{th} camera rotation and $\mathbf{t}_f = (t_{fx}t_{fy})^T$ is its translation relative to the world origin. s_f is the nonzero scalar of the weak perspective projection.

Replacing $\mathbf{X}_{fp}$ using Eq. (1) and absorbing s_f into c_{fi} and $\mathbf{t}_f$, we have

$$\mathbf{x}_{fp} = \begin{pmatrix} c_{f1}R_f & \dots & c_{fK}R_f \end{pmatrix} \cdot \begin{pmatrix} \mathbf{b}_{1p} \\ \dots \\ \mathbf{b}_{Kp} \end{pmatrix} + \mathbf{t}_f \tag{3}$$

Suppose the image coordinates of all P feature points across F frames are obtained. We form a $2F \times P$ *measurement matrix* W by stacking all image coordinates. Then $W = MB + T(11...1)$. where M is a $2F \times 3K$ scaled rotation matrix, B is a $3K \times P$ bases matrix, and T is a $2F \times 1$ translation vector,

$$M = \begin{pmatrix} c_{11}R_1 & \dots & c_{1K}R_1 \\ \vdots & \vdots & \vdots \\ c_{F1}R_F & \dots & c_{FK}R_F \end{pmatrix}$$
$$B = \begin{pmatrix} \mathbf{b}_{11} & \dots & \mathbf{b}_{1P} \\ \vdots & \vdots & \vdots \\ \mathbf{b}_{K1} & \dots & \mathbf{b}_{KP} \end{pmatrix}, \quad T = \begin{pmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_F \end{pmatrix} \tag{4}$$

As in [7, 5], we position the world origin at the scene center and compute the translation vector by averaging the image projections of all points. We then subtract it from W

and obtain the *registered* measurement matrix $\tilde{W} = MB$. Since $\tilde{W}$ is the product of the $2F \times 3K$ matrix M and the $3K \times P$ matrix B, its rank is at most $min\{3K, 2F, P\}$. In practice, the frame number F and point number P are usually much larger than the basis number K. Thus under non-degenerate deformations, *i.e.* each of the K bases is of full rank 3, the rank of $\tilde{W}$ is $3K$.

Under degenerate deformations, suppose of K shape bases K_1 bases are of rank 1, K_2 are of rank 2, and K_3 are of rank 3, the rank of B is $K_d = K_1 + 2K_2 + 3K_3$ and thus $\tilde{W}$ is of rank K_d. We then perform SVD on $\tilde{W}$ to get its best possible rank K_d approximation, $\tilde{M}\tilde{B}$, where $\tilde{M}$ is a $2F \times K_d$ matrix and $\tilde{B}$ is a $K_d \times P$ matrix. This decomposition is only determined up to a non-singular $K_d \times K_d$ linear transformation. In such cases, the true scaled rotation matrix $\hat{M}$ and bases matrix $\hat{B}$ are of the form,

$$\hat{M} = \tilde{M}G, \qquad \hat{B} = G^{-1}\tilde{B} \tag{5}$$

where G is called the *corrective transformation* matrix. K_3 triple-columns of G refer to the non-degenerate shape bases. We assume $K_3 > 0$, *i.e.* there is at least one non-degenerate basis, because we are not studying degenerate shapes, *e.g.* planar objects, but degenerate shape deformations. Without the loss of generality, we denote them as the first K_3 triple-columns of G, $\tilde{g}_1, ..., \tilde{g}_{K_3}$. The other columns, $g_{3K_3+1}, ..., g_{K_d}$ correspond to the degenerate bases. Let us denote them in such a way that the former columns refer to the K_2 rank-2 bases and the latter columns correspond to the K_1 rank-1 bases. Each of $\tilde{g}_i$, $i = 1, \ldots, K_3$, consists of three columns and each of $g_{j,\, j=3K_3+1,\ldots,K_d}$ contains only one column. We have,

$$\tilde{M}\tilde{g}_i = \begin{pmatrix} c_{1i}R_1 \\ \vdots \\ c_{Fi}R_F \end{pmatrix}, \quad \tilde{M}g_j = \begin{pmatrix} c_{1j}R_1 \\ \vdots \\ c_{Fj}R_F \end{pmatrix} r_j \tag{6}$$

where r_j is a unitary 3×1 vector. According to Eq. (6), the first K_3 triple-columns of $\hat{M}$ correspond to the non-degenerate bases and are the same as those of M in Eq. (4). The other columns of $\hat{M}$ refer to the degenerate bases and they are unitary projections of the corresponding triple-columns of M. The unitary vector r_j is the eigenvector of the corresponding degenerate shape basis. For example, a degenerate basis B_j of rank 1 can be factorized as $r_j\hat{B}_j$, where r_j is the unitary 3×1 eigenvector and $\hat{B}_j$ is the $1\times P$ projection vector. Then r_j is absorbed into $\hat{M}$ as in Eq. (6) and $\hat{B}_j$ becomes one row of $\hat{B}$.

According to Eq. (5, 6), once G is recovered, the rotations, shape bases, and combination coefficients are all determined. Therefore the problem is reduced to: given the measurement matrix W, how can we determine the *corrective transformation* matrix G?

4. Constraints

In [18], we presented that two types of constraints should be imposed to compute G: orthonormality constraints on camera rotations (*rotation constraints*) and uniqueness constraints on shape bases (*basis constraints*).

4.1. Rotation Constraints

The orthonormality constraints on the rotation matrices are one of the most powerful metric constraints and they have been used in reconstructing the shape and motion for static objects [12], multiple moving objects [7], and deformable objects [5, 18].

Denote $\tilde{g}_i\tilde{g}_i^T$ by Q_i. According to Eq. (6), We have,

$$\tilde{M}_{2m-1:2m}Q_i\tilde{M}_{2n-1:2n}^T = \Sigma_{k=1}^K c_{mk}c_{nk}R_mR_n^T \tag{7}$$

where $\tilde{M}_{2m-1:2m}$ represents the m_{th} two-row of $\tilde{M}$. Due to the orthonormality of the rotation matrices,

$$\tilde{M}_{2m-1:2m}Q_i\tilde{M}_{2m-1:2m}^T = \Sigma_{k=1}^K c_{mk}^2\mathbf{I}_{2\times2} \tag{8}$$

where $\mathbf{I}_{2\times2}$ is a 2×2 identity matrix. Since Q_i is symmetric, the number of unknowns in Q_i is $(K_d^2+K_d)/2$. For a frame m, Eq. (8) yields two linear constraints on Q_i,

$$\tilde{M}_{2m-1}Q_i\tilde{M}_{2m-1}^T = \tilde{M}_{2m}Q_i\tilde{M}_{2m}^T \tag{9}$$

$$\tilde{M}_{2m-1}Q_i\tilde{M}_{2m}^T = 0 \tag{10}$$

For F frames, we have $2F$ linear constraints on $(K_d^2 + K_d)/2$ unknowns. It appears that, given enough images, *i.e.* $F \geq (K_d^2 + K_d)/2$, the rotation constraints in Eq. (9,10) would be enough to determine Q_i via the least-square method. However, this is not true in general. Many of these constraints are redundant. No matter how many frames are given, the solution obtained using only these constraints is inherently ambiguous.

Since the corrective transformation G is a non-singular matrix, we denote Q_i as GHG^T. H is a $K_d \times K_d$ symmetric matrix, of which we only need to determine the upper triangle matrix. As shown in Figure 2, the upper triangle of H consists of 6 partitions, which correspond to the 6 types of compositions between the columns of G when constructing Q_i. For example, a 3×3 block H_{mn} in Ω_1 refers to the composition of $\tilde{g}_m H_{mn}\tilde{g}_n^T$; A 3×1 block H_{ml} in Ω_3 refers to the composition of $\tilde{g}_m H_{ml}g_l^T$, where $\tilde{g}_m$ and $\tilde{g}_n$ correspond to non-degenerate rank-3 bases and g_l refers to a rank-1 basis. Accordingly, we represent Ω_1 as 3×3 blocks, Ω_2 as 3×2 blocks, Ω_3 as 3×1 blocks, Ω_4 as 2×2 blocks, Ω_5 as 2×1 blocks, and Ω_6 as individual elements.

Theorem 1 *The general solution of the rotation constraints in Eq. (9,10) can be expressed as GHG^T, where G is the*

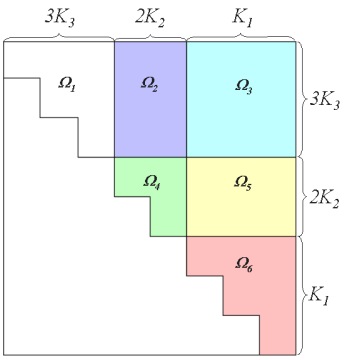

Figure 2: The upper triangle of H consists of 6 partitions, which refer to 6 types of compositions between the columns of G.

desired corrective transformation matrix and H satisfies,

$$H_{mn} = \begin{cases} \lambda_{mn}\mathbf{I}_{3\times 3}, & (m,n) \in \Omega_1, m = n \\ Y_{mn(3\times 3)} + \lambda_{mn}\mathbf{I}_{3\times 3}, & (m,n) \in \Omega_1, m < n \\ \begin{pmatrix} Y_{mn(2\times 2)} & \\ 0 \quad 0 & \end{pmatrix}, & (m,n) \in \Omega_2 \\ 0, & (m,n) \in \Omega_4, m = n \\ Y_{mn(2\times 2)}, & (m,n) \in \Omega_4, m < n \\ 0, & (m,n) \in \Omega_3 \bigcup \Omega_5 \bigcup \Omega_6 \end{cases} \tag{11}$$

where λ_{mn} is an arbitrary scalar. Y_{mn} is an arbitrary skew-symmetric matrix, *i.e.* $Y_{mn} = -Y_{mn}^T$. Note that the size of H_{mn} varies in different partitions. The proving procedure is similar with that of Theorem 1 in [18]. For details, refer to [18]. The diagonal elements of a skew-symmetric matrix are all zeros. Thus a 2×2 skew-symmetric matrix includes only 1 free element and a 3×3 skew-symmetric matrix includes 3 free elements. According to Eq. (11), it is easy to show that the degree of freedom of the solution by enforcing only the rotation constraints is $2K_3^2 - K_3 + K_3K_2 + K_2^2$.

4.2. Basis Constraints

In [18] we demonstrate that the ambiguity of rotation constraints arises from the non-uniqueness of shape bases because any non-singular linear transformation on the bases yields a new set of eligible bases. To eliminate the ambiguity, we need to determine a unique set of bases. In non-degenerate cases, this is done by selecting K frames including independent shapes and treating those shapes as a set of bases [18]. Specifically, for any group of K frames, we compute the condition number of the corresponding image measurements (sub-matrix of $\tilde{W}$). The group with the smallest condition number is chosen to determine the bases because smaller condition numbers refer to more independent shapes. However, this process can only determine the non-degenerate bases.

In degenerate cases, even if some bases are degenerate, their linear combination is in general non-degenerate. Thus the shape in any frame is non-degenerate as a combination of all bases and we cannot determine the degenerate bases as we did for the non-degenerate ones. We can only choose the group of K_3 frames of which the image measurements have the smallest condition number and treat the associated shapes as the K_3 non-degenerate bases. Note that although the non-degenerate bases are implicitly determined by the above procedure, their explicit values are unknown at this point.

Without the loss of generality, we denote the chosen frames as the first K_3 images in the sequence and the corresponding coefficients are

$$\begin{array}{l} c_{mm} = 1, \ m \leq K_3 \\ c_{mn} = 0, \ m \neq n, \ m \leq K_3, \ n \leq K \end{array} \tag{12}$$

According to Eq. (7), we obtain $4F(K_3 - 1)$ linear constraints on Q_i. They are called the *basis constraints*:

$$\tilde{M}_{2m-1}Q_i\tilde{M}_{2n-1}^T = \begin{cases} 1, & (m,n) \in \omega_1 \\ 0, & (m,n) \in \omega_2 \end{cases} \tag{13}$$

$$\tilde{M}_{2m}Q_i\tilde{M}_{2n}^T = \begin{cases} 1, & (m,n) \in \omega_1 \\ 0, & (m,n) \in \omega_2 \end{cases} \tag{14}$$

$$\tilde{M}_{2m-1}Q_i\tilde{M}_{2n}^T = 0, \ (m,n) \in \omega_1 \bigcup \omega_2 \tag{15}$$

$$\tilde{M}_{2m}Q_i\tilde{M}_{2n-1}^T = 0, \ (m,n) \in \omega_1 \bigcup \omega_2 \tag{16}$$

where $\omega_1 = \{(m,n) | m = n = i\}$ and $\omega_2 = \{(m,n) | m \leq K_3, \ n \leq F, \ m \neq i\}$.

Theorem 2 *Enforcing both basis and rotation constraints, i.e. Eq. (9,10,13∼16), the general solution of $Q_i = GHG^T$ satisfies,*

$$H_{mn} = \begin{cases} \mathbf{I}_{3\times 3}, & m = n = i \\ \begin{pmatrix} Y_{mn(2\times 2)} & 0 \\ 0 \quad 0 & 0 \end{pmatrix}, & m = i \neq n, \ (m,n) \in \Omega_1 \\ 0, & (m,n) \neq i, \ (m,n) \in \Omega_1 \\ \begin{pmatrix} Y_{mn(2\times 2)} & \\ 0 \quad 0 & \end{pmatrix}, & (m,n) \in \Omega_2 \\ 0, & (m,n) \in \Omega_4, \ m = n \\ Y_{mn(2\times 2)}, & (m,n) \in \Omega_4, \ m < n \\ 0, & (m,n) \in \Omega_3 \bigcup \Omega_5 \bigcup \Omega_6 \end{cases} \tag{17}$$

where Y_{mn}'s are skew-symmetric matrices that satisfy,

$$\begin{array}{l} Y_{im} + \Sigma_{l=K_3+1}^{K_3+K_2} c_{ml} Y_{il} = \ 0, \quad m \leq K_3, \ m \neq i \\ Y_{mn} + \Sigma_{l=K_3+1}^{K_3+K_2} c_{ml} Y_{ln} = \ 0, \quad (m,n) \in \Omega_2, \ m \neq i \end{array} \tag{18}$$

The proving procedure is similar with that of Theorem 2 in [18]. Since a 2×2 skew-symmetric matrix consists of one free element, H consists of $K_3 - 1 + K_3K_2 + (K_2^2 - K_2)/2$ free elements. Eq. (18) provides $K_3 - 1 + K_3K_2 - K_2$ independent linear constraints. Therefore enforcing both the metric constraints leads to a solution space of which the degree of freedom is $N_D = (K_2^2 + K_2)/2$. When the shape

bases are either rank-3 or rank-1 ($K_2 = 0$), the metric constraints generate a unique solution ($N_D = 0$). Otherwise when there exist rank-2 bases ($K_2 > 0$), the solution is ambiguous ($N_D > 0$).

5. Solutions

5.1. Determine the Number of the Bases

To utilize the rotation and basis constraints, we need to know the number of rank-3, 2, and 1 bases. First let us determine K_d, the rank of $\tilde{W}$. We perform SVD on $\tilde{W}$ and obtain the singular values. In noiseless settings, K_d equals the number of the non-zero singular values. When noise exists, K_d is estimated as the smallest number of the singular values whose sum is larger than some percentage (99% in our experiments) of the sum of all the singular values.

We then decide K_3, the number of non-degenerate bases. Because these bases are of rank 3, $1 \leq K_3 \leq K_d/3$. In previous section, we show that the basis constraints only determine the rank-3 bases, *i.e.* only rank-3 bases satisfy the basis constraints. Thus we choose K_3 as the largest number from 1 to $K_d/3$ for which the linear constraints (Eq. (9,10,13$\sim$16)) are satisfied.

We now determine K_2, the number of rank-2 bases. According to Theorem 2, the rank of the linear constraints is a quadratic function of K_2. Because K_3 is known, we can compute the rank of Eq. (9,10,13$\sim$16) and calculate K_2 as a root of the function. Finally K_1, the number of rank-1 bases, is $K_d - 2K_2 - 3K_3$.

5.2. An Alternating Linear Solution under the Existence of Rank-2 Shape Bases

Due to Theorem 2, when rank-2 shape bases exist ($K_2 > 0$), imposing the metric constraints (Eq. (9,10,13$\sim$16)) leads to an ambiguous solution space. By definition $Q_i = \tilde{g}_i \tilde{g}_i^T$ is positive semi-definite. According to Eq. (17), if any of the skew-symmetric matrices (Y_{mn}) in H is not zero, H is not positive semi-definite and nor is Q_i that equals GHG^T. Thus the solution space contains invalid solutions. Y_{mn}'s have to be zeros so that Q_i is a valid solution. We thus develop an alternating linear method that enforces this constraint to uniquely determine a valid solution in the space.

Because the linear solution space has the degree of freedom of N_D, we represent Q_i as a weighted sum of a particular solution and N_D homogeneous solutions,

$$Q_i = \Lambda_0 + \Sigma_{m=1}^{N_D} \lambda_m \Lambda_m \tag{19}$$

where Λ_0 is the particular solution and Λ_1,...,Λ_{N_D} are the homogeneous solutions. The scalars λ_m are the only unknowns to solve for. Our algorithm consists of three steps:

1. Use the particular solution Λ_0 as the initial estimate of Q_i.
2. Apply SVD on Q_i to compute its best possible rank 3 approximation $\tilde{g}_i \tilde{g}_i^T$.
3. Given $\tilde{g}_i$, calculate the coefficients λ_m in Eq. (19) by the linear least-square method. Then update Q_i via Eq. (19).

The last two linear processes are repeated alternatively till they converge. Note that the positive semi-definite constraint $Q_i = \tilde{g}_i \tilde{g}_i^T$ is explicitly enforced. Once $\tilde{g}_i$, $i = 1, ..., K_3$ are determined, according to Eq. (6), we reconstruct the rotations and the associated coefficients.

So far we have recovered the columns of G that refer to the non-degenerate bases and the camera rotations. We now recover the other columns, $g_{3K_3+1}, ..., g_{K_d}$, which correspond to the degenerate bases. From the second equation in Eq. (6), we cancel the unknown coefficients and achieve F constraints on g_j and r_j,

$$(\tilde{M}_{2m-1} g_j R_{m,2} - \tilde{M}_{2m} g_j R_{m,1}) r_j = 0,\ m = 1, ..., F \tag{20}$$

where $R_{m,1}$ means the first row of the rotation matrix R_m. Due to Eq. (12), we obtain another $2K_3$ constraints on g_j,

$$\tilde{M}_m g_j = 0, \quad m = 1, ..., 2K_3 \tag{21}$$

We then apply the following alternating linear approach to determine g_j and r_j,

1. Calculate a particular solution of Eq. (21) as the initial estimate of g_j.
2. Given g_j, calculate the rank-1 null space of Eq. (20) as the solution of r_j.
3. Given r_j, solve Eq. (20) and (21) to update g_j.

The last two linear processes are repeated alternatively till they converge. In these processes, we constrain G to be non-singular by forcing its columns independent on each other. This way prevents the algorithm from converging to some trivial solutions, *e.g.* g_j and r_j are both zeros. Now we have completely recovered the corrective transformation G. The associated coefficients and the shape bases are computed using Eq. (6) and (5) respectively. Their composition then reconstructs the non-rigid shapes as in Eq. (1).

5.3. A Unique Solution when Rank-2 Shape Bases do not Exist

A special case of degenerate deformations, *i.e.* all the points on the non-rigid shape either are static or independently move along straight lines, often occurs in practice. For example, cars drive or pedestrians walk independently along straight lines and beside a house. Several approaches [1, 7, 17] have been developed specifically for such degenerate deformations. However they require strong prior knowledge on either shape or motion. For example, assuming the camera projection matrices and the feature correspondence are given across five or more views, [1] presents the trajectory triangulation technique that uniquely reconstructs the 3D shape and motion trajectories.

In such cases, the shape bases are either rank-3 or rank-1. In the above example, the bases referring to the independent motions of cars or pedestrians are of rank 1 and the basis corresponding to the static house is of rank 3. Because $K_2 = 0$, according to Theorem 2, enforcing the linear metric constraints (Eq. (9,10,13$\sim$16)) leads to a unique solution of Q_i. Using SVD, we can factorize Q_i to compute $\tilde{g}_i$. Then the camera rotations can be recovered using Eq. (6).

Under the weak-perspective projection model, given the recovered rotations, we can construct the projection matrix up to a scalar,

$$\Omega_i = \begin{pmatrix} & R_i & & 0 \\ & & & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{22}$$

where the translation has been eliminated by moving the origin to the center of all points. We then apply the trajectory triangulation technique [1, 11] to uniquely reconstruct the 3D shapes and motion trajectories. For details of the trajectory triangulation technique, refer to [1, 11]. Note that we do not require the assumptions as previous approaches did.

6. Performance Evaluation

The performance of our approach is evaluated in a number of experiments. First, we evaluate its robustness and accuracy quantitatively on synthetic data. Second, we apply it on real image sequences to examine it qualitatively.

6.1. Quantitative Evaluation on Synthetic Data

Our approach is first quantitatively evaluated on the synthetic data. We test its accuracy and robustness on two factors: number of degenerate bases and strength of noise. Since the number of the unknowns involved in the alternating linear algorithm only depends on the number of the rank-2 bases, we choose all the degenerate bases to be of rank 2 in the experiments. Thus more degenerate bases result in a more complex optimization process. Assuming a Gaussian white noise, we represent the noise strength level by the ratio between the Frobenius norm of the noise and the measurement, *i.e.* $\frac{\|noise\|}{\|\tilde{W}\|}$. In general, when noise exists, the larger the number of degenerate bases is, the more complicated the optimization process is and thus the worse its performance is.

Figure 3 shows the evaluation on a 10 bases setting. The number of degenerate bases is respectively 1,..., or 9, shown as the horizontal axes. Four levels of Gaussian white noise are imposed. Their strength levels are 0%, 5%, 10%, and 20% respectively. We test a number of trials on each setting. The average reconstruction errors on the rotations and 3D shapes relative to the ground truth are shown in Figure 3. In the experiments when the noise level is 0%, regardless of how many bases are degenerate, our method converges to the exact rotations and shapes with zero error. When there is noise, it achieves reasonable accuracy, *e.g.* the maximum reconstruction error is less than 20% when the noise level is 20% and 9 out of 10 bases are degenerate. As we expected, under the same noise level, the performance is better when more bases are non-degenerate.

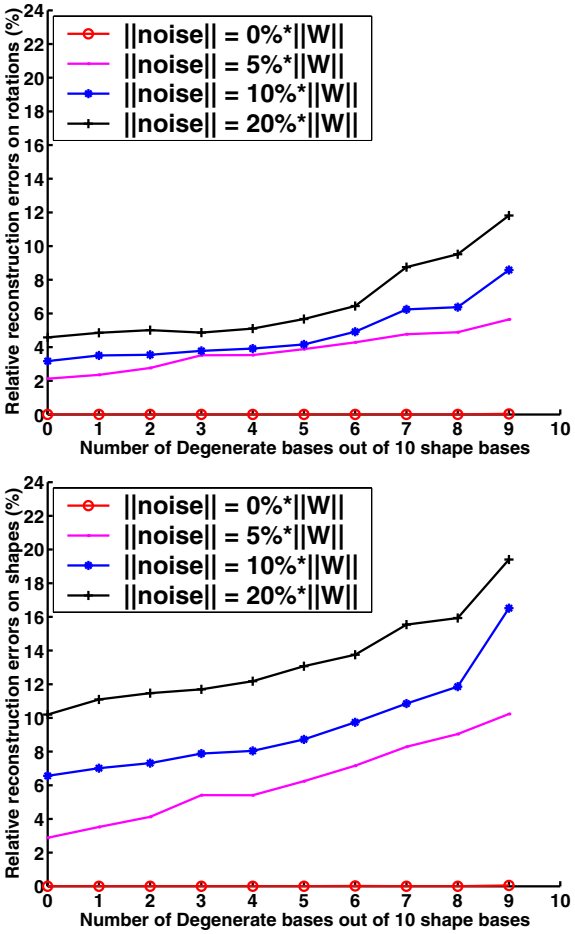

Figure 3: The relative reconstruction errors under different levels of noise and various number of degenerate bases. Each curve refers to a respective noise level.

6.2. Qualitative Evaluation on Real Video Sequences

We then examine our approach qualitatively on a number of real video sequences. One example is shown in Figure 4. The sequence was taken of an indoor scene by a handhold camera. The dynamic scene consisted of a static table and two boxes moving on top of the table. The boxes moved independently along the straight borders on the table top and at varying velocities. The scene structure is thus composed of three shape bases, one representing the static table and the initial locations of the two boxes and the other two rep-

resenting the two linear motion vectors respectively. Since the boxes vertices and the table corners are not located in the same plane, the first shape basis is of rank 3. The other two bases are both of rank 1. Thus the rank of the image measurement $\tilde{W}$ is 5.

18 feature points, consisting of the table corners and visible vertices of the boxes, across 30 images are given for reconstruction. Two of them are shown in Figure 4.(a,b). The numbers of the three types of bases are determined as described in Section 5.1. The camera rotations and dynamic scene structure are then reconstructed by the alternating linear algorithm. To evaluate the reconstruction, we synthesize the scene appearance viewed from one side, as shown in Figure 4.(c,d). The wireframes show the structure and the yellow lines show the trajectories of the moving boxes from the beginning of the sequence until the present frames. The recovered structure is consistent with our observation, *e.g.* the boxes approximately move along the table top borders. Figure 4.(e,f) show the reconstructed scene viewed from the top. Because the scene structure is composed of rank-1 and rank-3 bases, we also tested the unique solution described in Section 5.2 on this setting and achieved the similar results. Occlusion was not taken into account when rendering these images. So in the regions that should be occluded, *e.g.* the areas behind the boxes, the stretched texture of the occluding objects appears. Our approach assumes the weak-perspective projection model that requires the scene to be far from the camera. However in this experiment, the images were not taken from a long distance. Due to the perspective effect, the recovered object shapes are somewhat distorted, *e.g.* the shapes of the boxes are not precisely cuboid.

Human faces are highly non-rigid objects and 3D face shapes can be represented as linear combinations of certain shape bases that refer to various facial expressions. Under some facial motions, *e.g.* eye opening, the deformations along horizontal and vertical directions are dominant and those along depth direction are relatively subtle. Under the expressions where these degenerate motions are accompanied with other facial deformations such that the corresponding bases for the entire face shape are non-degenerate, the non-rigid shapes can be recovered using the method in [18]. Under some expressions, *e.g.* yawning and blinking, the facial deformations are mainly composed of these degenerate motions and thus the corresponding bases are close to degenerate. In such cases, we have to utilize the alternating linear method. One example is shown in Figure 5. The sequence consists of 180 face images that contain expressions like blinking and smiling. 68 feature points were tracked using an efficient Active Appearance Model (AAM) method [2]. Figure 5.(a,b) display two input images with marked features. Their corresponding shapes are reconstructed and shown from a novel view in Figure 5.(c,d). The overlapped wireframes demonstrate the recovered facial deformations such as mouth widening when smiling and eye closure when blinking.

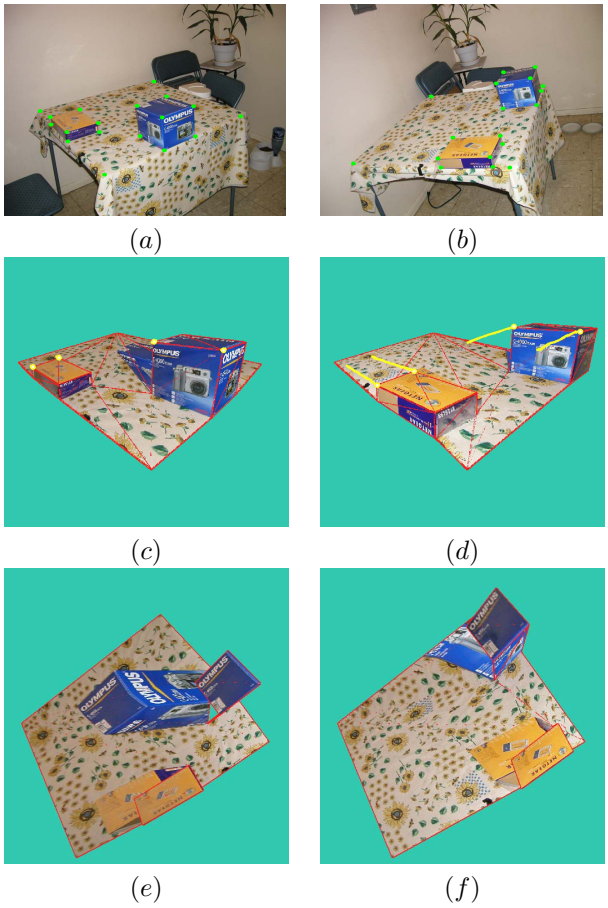

Figure 4: Reconstruction of two boxes independently moving along the borders of a static table top. $(a)\&(b)$: Two input images with marked features. $(c)\&(d)$: Reconstructed scene appearance viewed from one side. The wireframes show the structure and the yellow lines show the trajectories of the boxes from the beginning of the sequence until the present frames. $(e)\&(f)$: Reconstructed scene appearance viewed from the top.

7. Conclusion and Discussion

This paper studies the problem of non-rigid structure from motion under degenerate deformations. We quantitatively demonstrate that when rank-2 bases exist, imposing only the linear metric constraints (Eq. (9,10,13$\sim$16)) results in an ambiguous solution space. To eliminate the ambiguity, we develop an alternating linear approach that combines the metric constraints with the positive semi-definite constraint. When the points on the shape either are static or independently move along straight lines, we present a unique solution to reconstructing the 3D shape and motion trajectories.

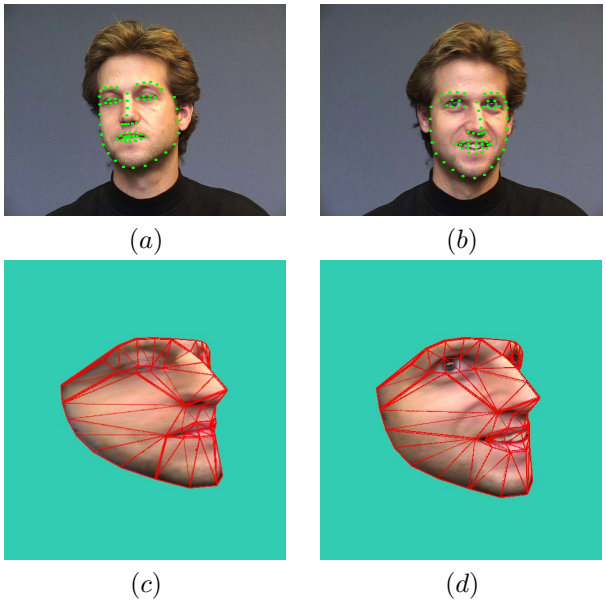

Figure 5: Reconstruction of face shapes with expressions of blinking and smiling. $(a)\&(b)$: Input images with marked features. $(c)\&(d)$: Reconstructed face structures seen from a novel view. The overlapped wireframes demonstrate the recovered facial deformations such as mouth widening and eye closure.

A limitation of our approach is the assumption of the weak-perspective projection model. When the depth variance of the scene points is not small relative to the distance from the scene to the camera, the reconstructed shape will be distorted due to the perspective effect. An example is shown in Figure 4. To compensate for this effect, we are working on the extension of our method to the full perspective projection model. Another problem is how to deal with outliers such as occlusion and missing data that often occur in practice. One idea is to combine the factorization process with a robust estimator that detects and handles the outliers [8]. An alternative way is to incorporate a prior knowledge on the specific object structure, *e.g.* symmetry, into the factorization process.

Acknowledgments

We would like to thank Simon Baker and Iain Matthews for providing the AAM tracking software, and thank Jinxiang Chai and Qifa Ke for fruitful discussions. The authors were partly supported by grant R01 MH51435 from the National Institute of Mental Health.

References

[1] S. Avidan, A. Shashua, " Trajectory Triangulation: 3D Reconstruction of Moving Points from a Monocular Image Sequence ," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 4, 2000.

[2] S. Baker, I. Matthews, " Equivalence and Efficiency of Image Alignment Algorithms," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2001.

[3] V. Blanz, T. Vetter, " A morphable model for the synthesis of 3D faces," *SIGGRAPH'99*, pp. 187-194, 1999.

[4] M. Brand, " Morphable 3D Models from Video," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2001.

[5] C. Bregler, A. Hertzmann, H. Biermann, " Recovering Non-Rigid 3D Shape from Image Streams," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2000.

[6] J. Costeira, T. Kanade, " A multibody factorization method for independently moving objects," *Int. Journal of Computer Vision*, 29(3):159-179, 1998.

[7] M. Han, T. Kanade, " Reconstruction of a Scene with Multiple Linearly Moving Objects," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2000.

[8] D. Q. Huynh, R.Hartley, A. Heyden," Outlier Correction in Image Sequences for the Affine Camera," *Proc. Int. Conf. Computer Vision*, pp. 585-590, 2003.

[9] C. Poelman, T. Kanade, " A paraperspective factorization method for shape and motion recovery," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(3):206-218, 1997.

[10] M. Pollefeys, R. Koch, L. Van Gool, " Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters," *Int. Journal of Computer Vision*, 32(1), pp. 7-25, 1999.

[11] S. Teller, " Computing the Antipenumbra Cast by an Area Light Source," *SIGGRAPH'92*, Vol. 26, No. 2, pp. 139-148, 1992.

[12] C. Tomasi, T. Kanade, " Shape and motion from image streams under orthography: A factorization method," *Int. Journal of Computer Vision*, 9(2):137-154, 1992.

[13] L. Torresani, D. Yang, G. Alexander, C. Bregler, " Tracking and Modeling Non-Rigid Objects with Rank Constraints," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2001.

[14] B.Triggs, " Factorization Methods for Projective Structure and Motion," *Proc. Int. Conf. Computer Vision and Pattern Recognition*,1996.

[15] R. Vidal, S. Soatto, Y. Ma, S. Sastry, " Segmentation of Dynamic Scenes from the Multibody Fundamental Matrix," *ECCV Workshop on Vision and Modeling of Dynamic Scenes*, 2002.

[16] L. Wolf, A. Shashua, " Two-body Segmentation from Two Perspective Views," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, 2001.

[17] L. Wolf, A. Shashua, " On Projection Matrices $P^k \rightarrow P^2, k = 3, \ldots, 6$, and their Applications in Computer Vision," *Int. Journal of Computer Vision*, 48(1):53-67, 2002.

[18] J. Xiao, J. Chai, T. Kanade, " A Closed-Form Solution to Non-Rigid Shape and Motion Recovery," *Proc. Euro. Conf. Computer Vision*, 2004.