

# Nonrigid Shape and Motion from Multiple Perspective Views

René Vidal<sup>1,2</sup> and Daniel Abretske<sup>2</sup>

<sup>1</sup> Center for Imaging Science, Department of BME, Johns Hopkins University

<sup>2</sup> Department of Computer Science, Johns Hopkins University

**Abstract.** We consider the problem of nonrigid shape and motion recovery from point correspondences in multiple perspective views. It is well known that the constraints among multiple views of a rigid shape are multilinear on the image points and can be reduced to bilinear (epipolar) and trilinear constraints among two and three views, respectively. In this paper, we generalize this classic result by showing that the constraints among multiple views of a nonrigid shape consisting of  $K$  shape bases can be reduced to multilinear constraints among  $K + \lceil (K + 1)/2 \rceil, \dots, 2K + 1$  views. We then present a closed form solution to the reconstruction of a nonrigid shape consisting of two shape bases. We show that point correspondences in five views are related by a nonrigid quintifocal tensor, from which one can linearly compute nonrigid shape and motion. We also demonstrate the existence of intrinsic ambiguities in the reconstruction of camera translation, shape coefficients and shape bases. Examples show the effectiveness of our method on nonrigid scenes with significant perspective effects.

## 1 Introduction

The past few decades have witnessed significant advances on the reconstruction of *static scenes* observed by a moving camera under the assumption that the scene is *Lambertian*, *rigid* and *static*. The Lambertian assumption is crucial to the problems of tracking, optical flow and correspondences, because the intensity of a point is independent of the view point. Given optical flow or point correspondences, the assumption of a rigidly moving camera observing a static world enables us to both recover the camera motion as well as reconstruct the rigid shape of the scene.

Recently, there have been attempts to relax each one of these assumptions. For example, the generalized constant brightness constraint allows one to compute optical flow for non Lambertian scenes. Likewise, the multibody fundamental matrix [11] allows one to reconstruct *dynamic scenes* consisting of multiple rigid motions. As for the third assumption, there have been two main approaches to dealing with *nonrigid scenes*. In direct approaches [5, 12], a static camera observes a nonrigid scene whose temporal evolution exhibits certain stationarity, e.g., water, foliage, steam, etc. These scenes are called *dynamic textures*, and have been successfully modeled as the output of a time invariant linear dynamical system. In feature-based methods [2, 3, 4, 9, 10, 13], a rigidly moving *affine* camera observes a nonrigid shape that deforms as a linear combination of rigid shapes with time varying coefficients. This assumption allows one to recover non-rigid shape and motion using extensions of the classical rigid factorization algorithm

[8]. For instance, [4] uses multiple matrix factorizations to enforce orthonormality constraints on camera rotations. [2] uses a non-linear optimization method called flexible factorization. [10] uses a trilinear optimization algorithm that alternates the computation of shape bases, shape coefficients, and camera rotations. Unfortunately, all these methods fail to reconstruct the correct shape and motion, because rotation constraints are not sufficient to guarantee a unique solution. [13] not only provides a complete characterization of the space of ambiguous solutions, but also proposes a closed form solution by enforcing additional *shape constraints* on the shape bases.

A key assumption of these approaches is that the projection model is *affine*. Although one can use nonlinear optimization to extend affine methods to the perspective case, e.g., [1], it is well known that iterative schemes applied to multilinear problems are very sensitive to initialization. The objective of this paper is to understand the algebraic constraints among multiple views of a nonrigid shape and to develop algebraic methods for nonrigid shape reconstruction that can be used for initializing optimization-based schemes. To the best of our knowledge, there is no prior work addressing these issues.

In this paper, we look at the problem of nonrigid shape and motion recovery from multiple *perspective views*. We first study the geometry of the problem, particularly the nature of the constraints among shape, motion and point correspondences. We show that the constraints among multiple views of a nonrigid scene can be derived from a rank constraint on the so-called *nonrigid multiple view matrix*. In the case of  $K$  shape bases, we prove that these algebraic constraints can be reduced to multilinear constraints among  $K + \lceil (K + 1)/2 \rceil, \dots, 2K + 1$  views of the image points,<sup>1</sup> thus ruling out the existence of epipolar or trilinear geometry for nonrigid scenes. We then show how to exploit these multilinear constraints for reconstructing a nonrigid shape consisting of  $K = 2$  rigid shapes. We demonstrate the existence of a *nonrigid quintifocal tensor*, which can be linearly estimated from the given point correspondences. We exploit algebraic properties of this tensor to compute nonrigid fundamental matrices among pairs of views. This leads to a linear algorithm for computing camera rotation and point depths. We also discuss the existence of intrinsic ambiguities in the reconstruction of camera translations, shape bases and shape coefficients. We then present examples showing the effectiveness of our method on nonrigid scenes with significant perspective effects.

## 2 Nonrigid Multiple View Geometry

Consider a nonrigid shape consisting of  $K$  shape bases, i.e. each 3-D point  $\mathbf{X}_f$  at frame  $f$  is a linear combination of  $K$  rigid shapes  $\{B_k \in \mathbb{R}^3\}_{k=1}^K$

$$\mathbf{X}_f = \sum_{k=1}^K (c_{fk} B_k), \quad (1)$$

where  $\{c_{fk}\}$  are the *shape coefficients*. Assume now that this nonrigid shape is observed by a moving perspective camera whose pose in the  $f$ th frame is given by  $(R_f, T_f)$

<sup>1</sup> Classical multilinear constraints in structure from motion show up as the special case  $K = 1$ .

$\in SE(3)$ . Therefore, the projection  $\mathbf{x}_f \in \mathbb{P}^2$  of  $\mathbf{X}_f$  is related to its depth  $\lambda_f$ , the camera motion  $(R_f, T_f)$ , the shape bases  $\{B_k\}$  and the shape coefficients  $\{c_{fk}\}$  by the equation

$$\lambda_f \mathbf{x}_f = R_f \sum_{k=1}^K (c_{fk} B_k) + T_f. \quad (2)$$

In this section, we show that one can algebraically eliminate depth and shape bases from the above equations, and derive algebraic constraints relating image points  $\{\mathbf{x}_f\}$ , camera motion  $\{R_f, T_f\}$  and shape coefficients  $\{c_{fk}\}$  only. Furthermore, we show that the constraints among multiple views of a nonrigid shape can be reduced to multilinear constraints among  $K + \lceil (K + 1)/2 \rceil, \dots, 2K + 1$  views.

## 2.1 One Shape Basis

For the sake of simplicity, we first review the well-known results in the case of one rigid shape. We refer the reader to [7] for further details. Note that if  $K = 1$  we do not need to consider shape coefficients, hence we can assume without loss of generality that  $c_{f1} = 1$  and  $\mathbf{X} = B_1$ . Also, we assume without loss of generality that  $(R_1, T_1) = (I, 0)$ . Combining these observations with (2) we obtain  $\lambda_1 \mathbf{x}_1 = B_1$  and  $\lambda_f \mathbf{x}_f = \lambda_1 R_f \mathbf{x}_1 + T_f$  for  $f = 2, 3, \dots, F$ . We can eliminate  $\lambda_f$  from this equation by multiplying by  $\widehat{\mathbf{x}}_f$  on both sides, where  $\widehat{\mathbf{x}} \in so(3)$  is the skew-symmetric matrix generating the cross product by  $\mathbf{x}$ . This multiplication yields  $\lambda_1 \widehat{\mathbf{x}}_f R_f \mathbf{x}_1 + \widehat{\mathbf{x}}_f T_f = 0$ . Since this equation holds for all  $f = 2, 3, \dots, F$ , we can write the motion equations for all frames in terms of a single linear equation

$$M_1 \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{x}}_2 R_2 \mathbf{x}_1 & \widehat{\mathbf{x}}_2 T_2 \\ \vdots & \vdots \\ \widehat{\mathbf{x}}_F R_F \mathbf{x}_1 & \widehat{\mathbf{x}}_F T_F \end{bmatrix} \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} = 0. \quad (3)$$

The matrix  $M_1 \in \mathbb{R}^{3(F-1) \times 2}$  is called the *multiple view matrix* [7].

From (3), note that the vector  $\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$  lives in the right null space of  $M_1$ , hence

$$\text{rank}(M_1) \leq 1. \quad (4)$$

The implication of this result is that the determinant of any  $2 \times 2$  sub-matrix of  $M_1$  is equal to zero. By simply counting the ways we can choose two rows from  $M_1$ , it becomes immediately obvious that we can only select rows in such a way as to depend on either two or three views: Two view constraints are obtained by considering two rows from the same block of three frames, e.g., rows 1 and 2, while three view constraints are obtained by considering two rows from two different blocks of three frames, e.g., rows 1 and 4. Two view constraints can be reduced to the well-known epipolar constraint as shown in [7], while three view constraints are the well-known trilinear constraints [6]. This shows that the constraints among multiple views of a rigid scene are multilinear and algebraically dependent on the constraints among two and three views.

## 2.2 Multiple Shape Bases

Consider now a nonrigid shape consisting of  $K$  shape bases. Note that there is an ambiguity in the definition of the shape bases and coefficients, because for any invertible  $L \in \mathbb{R}^{K \times K}$  one can choose a new set of shape bases  $[B_1 \ B_2 \ \cdots \ B_K] \rightarrow [B_1 \ B_2 \ \cdots \ B_K]L$

and coefficients  $\begin{bmatrix} c_{f1} \\ \vdots \\ c_{fK} \end{bmatrix} \rightarrow L^{-1} \begin{bmatrix} c_{f1} \\ \vdots \\ c_{fK} \end{bmatrix}$  that yield the same point in 3D space  $\mathbf{X}_f$ .

[13] proposes to resolve this ambiguity by enforcing the following *basis constraints*

$$c_{kk} = 1, k = 1, \dots, K \quad c_{jk} = 0, j \neq k = 1, \dots, K. \quad (5)$$

Combining the basis constraints with the motion equations in (2) for  $f = 1 \dots K$  leads to  $\lambda_k \mathbf{x}_k = R_k B_k + T_k$  for  $k = 1, \dots, K$ , hence we can solve for the shape bases as  $B_k = R_k^\top (\lambda_k \mathbf{x}_k - T_k)$ . After choosing the reference frame so that  $(R_1, T_1) = (I, 0)$ , we can express the motion equations for frames  $f = K + 1, \dots, F$  as

$$\lambda_f \mathbf{x}_f = \lambda_1 c_{f1} R_f \mathbf{x}_1 + R_f \sum_{k=2}^K c_{fk} R_k^\top (\lambda_k \mathbf{x}_k - T_k) + T_f. \quad (6)$$

We now proceed as before using the cross product with  $\mathbf{x}_f$  to eliminate the depths  $\{\lambda_f\}$  for  $f = K + 1, \dots, F$ . The final result is a matrix equation of the following form

$$M_K \boldsymbol{\lambda}_K = \begin{bmatrix} \widehat{\mathbf{x}_{K+1}} Q_{K+1} \mathbf{x}_1 & \widehat{\mathbf{x}_{K+1}} S_{K+1}^2 \mathbf{x}_2 & \cdots & \widehat{\mathbf{x}_{K+1}} S_{K+1}^K \mathbf{x}_K & \widehat{\mathbf{x}_{K+1}} V_{K+1} \\ \vdots & \vdots & & \vdots & \vdots \\ \widehat{\mathbf{x}_F} Q_F \mathbf{x}_1 & \widehat{\mathbf{x}_F} S_F^2 \mathbf{x}_2 & \cdots & \widehat{\mathbf{x}_F} S_F^K \mathbf{x}_K & \widehat{\mathbf{x}_F} V_F \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_K \\ 1 \end{bmatrix} = 0, \quad (7)$$

where

$$Q_f = c_{f1} R_f, \quad S_f^k = c_{fk} R_f R_k^\top, \quad V_f = T_f - \sum_{k=2}^K c_{fk} R_f R_k^\top T_k.$$

The *nonrigid multiple view matrix*  $M_K \in \mathbb{R}^{3(F-K) \times (K+1)}$  has the vector of depths  $\boldsymbol{\lambda}_K$  in the first  $K$  frames in its right null space, hence it satisfies the rank constraint

$$\text{rank}(M_K) \leq K. \quad (8)$$

Therefore, we can eliminate the vector of depths in the first  $K$  frames,  $\boldsymbol{\lambda}_K$ , by enforcing that the determinant of each  $(K + 1) \times (K + 1)$  sub-matrix of  $M_K$  be zero. Since each block of three rows of  $M_K$  provides only two linearly independent equations, in choosing  $K + 1$  rows we need at least  $\lceil \frac{K+1}{2} \rceil$  blocks. Therefore, the determinant involving the minimum number of views contains  $K + \lceil \frac{K+1}{2} \rceil$  views. Note that this is much smaller than the minimum number of affine views, which is  $(3K^2 + 3K)/2$  [13]. On the other hand, if we choose one row per block, then the determinants involve  $K + (K + 1) = 2K + 1$  views. We have shown the following.

**Theorem 1 (Algebraic dependency of multiple view constraints for  $K$  shape bases).** *Consider a moving camera observing a nonrigid shape consisting of  $K$  shape bases. The equations relating camera motion, shape bases, shape coefficients and image points can be reduced to a set of algebraic constraints that do not depend on the shape bases and involve only  $K + \lceil \frac{K+1}{2} \rceil, \dots, 2K + 1$  views at a time.*

**Corollary 1.** *The constraints among multiple views of a rigid shape ( $K = 1$ ) can be reduced to constraints among two and three views.*

The next step is to understand whether the multiple view constraints are multilinear on the image points, as in the rigid case. To this end, note first that image points in one of the first  $K$  frames appear in only one column of  $M_K$  at a time, hence multiple view constraints are necessarily *linear* in each one of the first  $K$  views. However, the constraints on a point in the remaining  $F - K$  frames can be either linear or quadratic, depending on whether we choose one or two rows per block, respectively. This can be seen by considering how one might choose rows from  $M_K$  when forming the determinant of a  $(K + 1) \times (K + 1)$  submatrix. One can choose either one or two rows corresponding to each frame. If a single row is chosen from a frame then the constraints must be linear in points from that frame since that point only appears in a single row of the submatrix. However, when two rows are chosen from a single frame, it may still be possible that the resulting constraint remains linear and does not become quadratic on the point from that frame. The following theorem shows that this is indeed the case.

**Theorem 2 (Multilinear constraints for  $K$  shape bases).** *The algebraic constraints among multiple views of a nonrigid shape consisting of  $K$  shape bases can be reduced to a set of multilinear constraints on  $K + \lceil \frac{K+1}{2} \rceil, \dots, 2K + 1$  views of the image points. The coefficients of these multilinear constraints depend on the camera motion and the shape coefficients, but not on the shape bases.*

In what follows, we prove the theorem in the particular cases  $K = 2$  and  $K = 3$ , to then extend the proof to arbitrary  $K$ .

**Multilinear constraints for two shape bases.** We already know that in this case the algebraic constraints among multiple views can be reduced to those among four and five views. Moreover, we have already shown that the constraints among five views are multilinear in the point correspondences, because all minors of  $M_2$  involve one row per view. We are left with proving that the constraints among four views are also multilinear.

Without loss of generality, consider views 1 through 4, and choose two rows of  $M_2$  from the 3rd view and one from the 4th view. As choosing these three rows is equivalent to choosing three lines  $\ell_{31}, \ell_{32}$  and  $\ell_4$  such that  $\ell_{31}^\top \mathbf{x}_3 = \ell_{32}^\top \mathbf{x}_3 = \ell_4^\top \mathbf{x}_4 = 0$ , the algebraic constraint among these four views can be written as

$$\Delta_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = \det \begin{pmatrix} \ell_{31}^\top Q_3 \mathbf{x}_1 & \ell_{31}^\top S_3 \mathbf{x}_2 & \ell_{31}^\top V_3 \\ \ell_{32}^\top Q_3 \mathbf{x}_1 & \ell_{32}^\top S_3 \mathbf{x}_2 & \ell_{32}^\top V_3 \\ \ell_4^\top Q_4 \mathbf{x}_1 & \ell_4^\top S_4 \mathbf{x}_2 & \ell_4^\top V_4 \end{pmatrix}. \quad (9)$$

Before proceeding further, we need the following technical lemma, whose proof follows by direct calculation.

**Lemma 1.** *Let  $\mathbf{x} = (x_1, x_2, 1)^\top$ ,  $\ell_1 = (1, 0, -x_1)^\top$  and  $\ell_2 = (0, 1, -x_2)^\top$ . Then for all  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$  we have that*

$$\det \left( \begin{bmatrix} \ell_1^\top \mathbf{a} & \ell_1^\top \mathbf{b} \\ \ell_2^\top \mathbf{a} & \ell_2^\top \mathbf{b} \end{bmatrix} \right) = \mathbf{b}^\top \widehat{\mathbf{x}} \mathbf{a}. \quad (10)$$

After expanding  $\Delta_2$  along the bottom row of the matrix in (9) and applying Lemma 1 three times, we see that  $\Delta_2$  is in fact equal to

$$\Delta_2 = \ell_4^\top Q_4 \mathbf{x}_1 (V_3^\top \widehat{\mathbf{x}}_3 S_3 \mathbf{x}_2) - \ell_4^\top S_4 \mathbf{x}_2 (V_3^\top \widehat{\mathbf{x}}_3 Q_3 \mathbf{x}_1) + \ell_4^\top V_4 ((S_3 \mathbf{x}_2)^\top \widehat{\mathbf{x}}_3 Q_3 \mathbf{x}_1), \quad (11)$$

which is multilinear in  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ell_4)$ , hence in  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ , as claimed.

**Multilinear constraints for three shape bases.** In this case, the multiple view matrix  $M_3$  has four columns, hence one can form constraints on 5, 6 or 7 views. The case of 7 views is obviously multilinear, as one chooses a single row from each frame (4,5,6,7).

In the case of 6 views one must choose two rows from one frame and two more rows from two other frames. Without loss of generality consider choosing two rows from the 4th frame, a row from the 5th frame and a row from the 6th frame. This is equivalent to choosing lines  $\ell_{41}, \ell_{42}, \ell_5$  and  $\ell_6$  such that  $\ell_{41}^\top \mathbf{x}_4 = \ell_{42}^\top \mathbf{x}_4 = \ell_5^\top \mathbf{x}_5 = \ell_6^\top \mathbf{x}_6 = 0$ . Such a choice leads to the following determinant

$$\Delta_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6) = \det \left( \begin{bmatrix} \ell_{41}^\top Q_4 \mathbf{x}_1 & \ell_{41}^\top S_4^2 \mathbf{x}_2 & \ell_{41}^\top S_4^3 \mathbf{x}_3 & \ell_{41}^\top V_4 \\ \ell_{42}^\top Q_4 \mathbf{x}_1 & \ell_{42}^\top S_4^2 \mathbf{x}_2 & \ell_{42}^\top S_4^3 \mathbf{x}_3 & \ell_{42}^\top V_4 \\ \ell_5^\top Q_5 \mathbf{x}_1 & \ell_5^\top S_5^2 \mathbf{x}_2 & \ell_5^\top S_5^3 \mathbf{x}_3 & \ell_5^\top V_5 \\ \ell_6^\top Q_6 \mathbf{x}_1 & \ell_6^\top S_6^2 \mathbf{x}_2 & \ell_6^\top S_6^3 \mathbf{x}_3 & \ell_6^\top V_6 \end{bmatrix} \right). \quad (12)$$

We know that  $\Delta_3$  is linear in each of  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_5$  and  $\mathbf{x}_6$ . The question is whether  $\Delta_3$  is also linear in  $\mathbf{x}_4$ . Let  $\mathbf{x}_4 = (x_4, y_4, 1)^\top$ ,  $\ell_{41} = (1, 0, -x_4)^\top$  and  $\ell_{42} = (0, 1, -y_4)^\top$ . If we expand  $\Delta_3$  along the last row of the matrix in (12), we obtain

$$\begin{aligned} \Delta_3 &= (\ell_6^\top Q_6 \mathbf{x}_1) \Delta_{21}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5) - (\ell_6^\top S_6^2 \mathbf{x}_2) \Delta_{22}(\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5) + \\ &\quad (\ell_6^\top S_6^3 \mathbf{x}_3) \Delta_{23}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4, \mathbf{x}_5) - (\ell_6^\top V_6) \Delta_{24}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5), \end{aligned} \quad (13)$$

where each  $\Delta_{2i}$  is of the same form as the determinant seen in equation (9), thus multilinear in its entries. Therefore,  $\Delta_3$  is also multilinear in  $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6)$ .

In the case of 5 views, without loss of generality choose two rows from the 4th frame and two rows from the 5th frame, and let  $\mathbf{x}_i = (x_i, y_i, 1)^\top$ ,  $\ell_{i1} = (1, 0, -x_i)^\top$  and  $\ell_{i2} = (0, 1, -y_i)^\top$ . Such a choice gives the following determinant

$$\tilde{\Delta}_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5) = \det \left( \begin{bmatrix} \ell_{41}^\top Q_4 \mathbf{x}_1 & \ell_{41}^\top S_4^2 \mathbf{x}_2 & \ell_{41}^\top S_4^3 \mathbf{x}_3 & \ell_{41}^\top V_4 \\ \ell_{42}^\top Q_4 \mathbf{x}_1 & \ell_{42}^\top S_4^2 \mathbf{x}_2 & \ell_{42}^\top S_4^3 \mathbf{x}_3 & \ell_{42}^\top V_4 \\ \ell_{51}^\top Q_5 \mathbf{x}_1 & \ell_{51}^\top S_5^2 \mathbf{x}_2 & \ell_{51}^\top S_5^3 \mathbf{x}_3 & \ell_{51}^\top V_5 \\ \ell_{52}^\top Q_5 \mathbf{x}_1 & \ell_{52}^\top S_5^2 \mathbf{x}_2 & \ell_{52}^\top S_5^3 \mathbf{x}_3 & \ell_{52}^\top V_5 \end{bmatrix} \right). \quad (14)$$

After expanding  $\tilde{\Delta}_3$  along the first column of the matrix, we obtain

$$\begin{aligned} \tilde{\Delta}_3 &= \ell_{41}^\top c \Delta_{25}(\mathbf{x}_2, \mathbf{x}_3, \ell_{42}, \mathbf{x}_5) - \ell_{42}^\top c \Delta_{25}(\mathbf{x}_2, \mathbf{x}_3, \ell_{41}, \mathbf{x}_5) \\ &\quad + \ell_{51}^\top d \Delta_{26}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ell_{52}) - \ell_{52}^\top d \Delta_{26}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ell_{51}) \end{aligned} \quad (15)$$

where  $\mathbf{c} = Q_4 \mathbf{x}_1$ ,  $\mathbf{d} = Q_5 \mathbf{x}_1$ , and  $\Delta_{25}(\cdot, \cdot, \cdot)$  and  $\Delta_{26}(\cdot, \cdot, \cdot)$  are determinants of 3 by 3 matrices that depend linearly on each of the quantities in the parentheses by direct application of Lemma 1. Since Lemma 1 also implies that  $(\ell_{f1}^\top \mathbf{y})(\ell_{f2}) - (\ell_{f2} \mathbf{y})(\ell_{f1}) = \widehat{\mathbf{y}} \mathbf{x}_f$ , the expression for  $\tilde{\Delta}_3$  reduces to

$$\begin{aligned} \tilde{\Delta}_3 &= \Delta_{25}(\mathbf{x}_2, \mathbf{x}_3, \ell_{41}^\top \mathbf{c} \ell_{42} - \ell_{42}^\top \mathbf{c} \ell_{41}, \mathbf{x}_5) + \Delta_{26}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ell_{51}^\top \mathbf{d} \ell_{52} - \ell_{52}^\top \mathbf{d} \ell_{51}) \\ &= \Delta_3(\mathbf{x}_2, \mathbf{x}_3, \widehat{\mathbf{c}} \mathbf{x}_4, \mathbf{x}_5) + \tilde{\Delta}_3(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \widehat{\mathbf{d}} \mathbf{x}_5), \end{aligned}$$

which is in fact linear in  $\mathbf{x}_4$  and  $\mathbf{x}_5$  as claimed.

**Multilinear constraints for multiple shape bases.** In the case of  $K$  shape bases, the constraints among multiple views are simply minors of the multiple view matrix  $M_K$ . Each minor is formed by choosing  $K + 1$  rows from  $M_K$ . Without loss of generality, assume we choose two rows from each one of the first  $m$  blocks, and one row from each one of the next  $K + 1 - 2m$  blocks. We obtain the following determinant

$$\Delta_K = \det \left( \begin{array}{ccc|ccc} \ell_{K+1,1}^\top Q_{K+1} \mathbf{x}_1 & \cdots & \ell_{K+1,1}^\top S_{K+1}^K \mathbf{x}_K & \ell_{K+1,1}^\top V_{K+1} & & \\ \ell_{K+1,2}^\top Q_{K+1} \mathbf{x}_1 & \cdots & \ell_{K+1,2}^\top S_{K+1}^K \mathbf{x}_K & \ell_{K+1,2}^\top V_{K+1} & & \\ \vdots & \cdots & \vdots & & & \\ \ell_{K+m,1}^\top Q_{K+m} \mathbf{x}_1 & \cdots & \ell_{K+m,1}^\top S_{K+m}^K \mathbf{x}_K & \ell_{K+m,1}^\top V_{K+m} & & \\ \ell_{K+m,2}^\top Q_{K+m} \mathbf{x}_1 & \cdots & \ell_{K+m,2}^\top S_{K+m}^K \mathbf{x}_K & \ell_{K+m,2}^\top V_{K+m} & & \\ \hline \ell_{K+m+1}^\top Q_{K+m+1} \mathbf{x}_1 & \cdots & \ell_{K+m+1}^\top S_{K+m+1}^K \mathbf{x}_K & \ell_{K+m+1}^\top V_{K+m+1} & & \\ \vdots & \cdots & \vdots & & & \\ \ell_{2K-m+1}^\top Q_{2K-m+1} \mathbf{x}_1 & \cdots & \ell_{2K-m+1}^\top S_{2K-m+1}^K \mathbf{x}_K & \ell_{2K-m+1}^\top V_{2K-m+1} & & \end{array} \right) \quad (16)$$

It is clear that  $\Delta_K$  is linear in each  $\mathbf{x}_i$  with  $i \leq K$ , because each image point appears only in one column of the matrix in (16). Similarly, it is clear that  $\Delta_K$  is linear in each  $\mathbf{x}_i$  with  $i > K + m$ , because each image point appears in only one row of the matrix in (16). The fact that  $\Delta_K$  is also linear in each  $\mathbf{x}_i$  with  $K + 1 \leq i \leq K + m$  follows by repeated application of the following lemma since the upper portion of the matrix in (16) is of the exact form called for by the lemma.

**Lemma 2.** Let  $\mathbf{a}_{ij} \in \mathbb{R}^3$ ,  $\mathbf{x}_i = (x_i, y_i, 1)^\top$ ,  $\ell_{i1} = (1, 0, -x_i)^\top$  and  $\ell_{i2} = (0, 1, -y_i)^\top$ . Then, for  $k$  even,  $\Delta_k = \det(\mathcal{M}_k)$  is linear in each  $\mathbf{x}_i$ , where

$$\mathcal{M}_k = \begin{bmatrix} \ell_{11}^\top \mathbf{a}_{11} & \cdots & \ell_{11}^\top \mathbf{a}_{1k} \\ \ell_{12}^\top \mathbf{a}_{11} & \cdots & \ell_{12}^\top \mathbf{a}_{1k} \\ \vdots & \cdots & \vdots \\ \ell_{\frac{k}{2}1}^\top \mathbf{a}_{\frac{k}{2}1} & \cdots & \ell_{\frac{k}{2}k}^\top \mathbf{a}_{\frac{k}{2}k} \\ \ell_{\frac{k}{2}2}^\top \mathbf{a}_{\frac{k}{2}1} & \cdots & \ell_{\frac{k}{2}2}^\top \mathbf{a}_{\frac{k}{2}k} \end{bmatrix} \in \mathbb{R}^{k \times k}. \quad (17)$$

*Proof.* The proof proceeds by strong induction. The example of three shape bases and five views proves the case of  $k = 2$ . Now assume that this holds for up to  $k = n - 2$ .

We will proceed to show that it holds for  $k = n$ . Consider  $\mathcal{M}_n$  and consider the result of expanding  $\det(\mathcal{M}_n)$  along the first column of  $\mathcal{M}_n$ . Without loss of generality we may consider just minors associated with the last two entries of this column. If the result holds for these minors then it must hold for all other minors since  $\mathcal{M}_n$  is composed of similar 2 by  $k$  blocks. These minors have the form

$$\begin{aligned} \ell_{n2}^\top \mathbf{a}_{n1} \Delta_{n-1}(\ell_{n1}, \ell_{(n-1)1}, \ell_{(n-1)2}, \dots, \ell_{11}, \ell_{12}) - \ell_{n1}^\top \mathbf{a}_{n1} \Delta_{n-1}(\ell_{n2}, \ell_{(n-1)1}, \ell_{(n-1)2}, \dots, \ell_{11}, \ell_{12}) \\ = \Delta_{n-1}(\mathbf{x}_n \widehat{\mathbf{a}}_{n1}, \ell_{(n-1)1}, \ell_{(n-1)2}, \dots, \ell_{11}, \ell_{12}), \end{aligned}$$

where the last equality follows from Lemma 1. By direct calculation it can be seen that

$$\Delta_{n-1}(\mathbf{x}_n \widehat{\mathbf{a}}_{n1}, \ell_{(n-1)1}, \ell_{(n-1)2}, \dots, \ell_{11}, \ell_{12}) = \sum_{j=2}^n \mathbf{a}_{n1}^\top \widehat{\mathbf{x}}_n \mathbf{a}_{nj} \Delta_{n-2}(\ell_{(n-1)1}, \ell_{(n-1)2}, \dots, \ell_{11}, \ell_{12}).$$

By the induction hypothesis  $\Delta_{n-2}$  is multilinear, hence  $\det(\mathcal{M}_n)$  is multilinear as claimed.  $\blacksquare$

### 3 Reconstruction of Two Shape Bases

Given that the constraints among multiple views of a nonrigid shape are multilinear, the next question is how to exploit such constraints in order to recover camera motion and nonrigid shape. In this section, we show how to do so in the case of a nonrigid shape consisting of two shape bases seen in five views. First, we demonstrate that the quintilinear constraints can be expressed in terms of a single tensor, which can be linearly estimated from the given point correspondences. Next, we study properties of this tensor that, surprisingly, demonstrate the existence of geometric entities analogous to epipolar lines and fundamental matrices for nonrigid motions. We exploit such properties in order to linearly solve for camera motion. Finally, we demonstrate the existence of ambiguities in the reconstruction of camera translation, shape coefficients and shape bases. These ambiguities are intrinsic to the nonrigid shape and motion problem, in the sense that they show up in the case of affine cameras as well. Surprisingly, they have not received wide attention in the literature, being only briefly discussed in [1].

#### 3.1 The Nonrigid Quintifocal Tensor

From the previous section, we know that in the case of 5 views, the multilinear constraints are determinants of  $3 \times 3$  sub-matrices of the multiple view matrix  $M_2 \in \mathbb{R}^{9 \times 3}$ . Furthermore, in the case of quintilinear constraints, each sub-matrix is formed by choosing three rows from each one of the three blocks of  $M_2$ . Therefore, we can write a single quintilinear constraint for a point-point-line-line-line correspondence as

$$\mathcal{T}(\mathbf{x}_1, \mathbf{x}_2, \ell_3, \ell_4, \ell_5) = \det \left( \begin{bmatrix} \ell_3^\top Q_3 \mathbf{x}_1 & \ell_3^\top S_3 \mathbf{x}_2 & \ell_3^\top V_3 \\ \ell_4^\top Q_4 \mathbf{x}_1 & \ell_4^\top S_4 \mathbf{x}_2 & \ell_4^\top V_4 \\ \ell_5^\top Q_5 \mathbf{x}_1 & \ell_5^\top S_5 \mathbf{x}_2 & \ell_5^\top V_5 \end{bmatrix} \right) = 0. \quad (18)$$

By expanding this determinant as a polynomial in  $\mathbf{x}_1, \mathbf{x}_2, \ell_3, \ell_4, \ell_5$ , we may write the quintilinear constraint as

$$\sum_{ijkmn=1}^3 \mathcal{T}_{ijkmn} x_{1i} x_{2j} \ell_{3k} \ell_{4m} \ell_{5n} = 0, \quad (19)$$

where  $\mathcal{T}_{ijkmn}$  are the coefficients of the polynomial. We define the *nonrigid quintifocal tensor*  $\mathcal{T} \in \mathbb{R}^{3 \times 3 \times 3 \times 3 \times 3}$  as the collection of all these coefficients.

Note that each point correspondence provides  $2^3 = 8$  linear equations in the 242 unknowns in  $\mathcal{T}$ , because we can choose two lines  $\ell_f$  for each  $\mathbf{x}_f$  for  $f = 3, 4, 5$ . Therefore, in order to determine  $\mathcal{T}$  linearly, we need at least 31 point correspondences.

**Notation.** For ease of notation, we will drop the summation and subscripts in multilinear expressions such as  $\sum \mathcal{T}_{ijkmn} x_{1i} x_{2j} \ell_{3k} \ell_{4m} \ell_{5n}$  and write them as  $\mathbf{x}_1 \mathbf{x}_2 \ell_3 \ell_4 \ell_5 \mathcal{T}$ . We will also write the matrix whose  $(ij)$ th entry is  $\sum_{kmn} \mathcal{T}_{ijkmn} \ell_{3k} \ell_{4m} \ell_{5n}$  as  $\ell_3 \ell_4 \ell_5 \mathcal{T}$  (or whichever indices are being considered), and the vector whose  $i$ th entry is given by  $\sum_{jkmn} \mathcal{T}_{ijkmn} x_{2j} \ell_{3k} \ell_{4m} \ell_{5n}$  as  $\mathbf{x}_2 \ell_3 \ell_4 \ell_5 \mathcal{T}$  (similarly for other indices).

### 3.2 Recovering Camera Motion Via Nonrigid Epipolar Geometry

In this subsection, we present some algebraic and geometric properties of the quintifocal tensor  $\mathcal{T}$ . We show that even though epipolar geometry is not defined for nonrigid shapes, there still exist algebraic entities that play the analogous role of geometric entities such as epipolar lines and essential matrices, which are only defined for a single rigid shape. These properties lead to a linear algorithm for recovering nonrigid epipolar lines, nonrigid essential matrices, and camera rotations from the quintifocal tensor.

At the core of the proposed method, we find a set of rank constraints on slices of  $\mathcal{T}$ , as stated in the following lemma.

**Lemma 3 (Rank constraints on slices of the quintifocal tensor).** *Let  $\mathcal{T}$  be a nonrigid quintifocal tensor. Then,  $\text{rank}(\mathbf{x}_1 \mathbf{x}_2 \ell_i \mathcal{T}) \leq 2$  for  $i = 3, 4, 5$ , and  $\text{rank}(\ell_3 \ell_4 \ell_5 \mathcal{T}) \leq 2$ .*

*Proof.* It follows by direct calculation that  $\mathbf{x}_1 \mathbf{x}_2 \ell_3 \ell_4 \ell_5 \mathcal{T} = \ell_4^\top \mathcal{M} \ell_5$ , where

$$\begin{aligned} \mathcal{M} = & Q_4 \mathbf{x}_1 (V_5^\top (\ell_3^\top S_3 \mathbf{x}_2) - \mathbf{x}_2^\top S_5^\top (\ell_3^\top V_3)) - S_4 \mathbf{x}_2 (V_5^\top (\ell_3^\top Q_3 \mathbf{x}_1) - \mathbf{x}_1^\top Q_5^\top (\ell_3^\top V_3)) \\ & + V_4 (\mathbf{x}_2^\top S_5^\top (\ell_3^\top Q_3 \mathbf{x}_1) - \mathbf{x}_1^\top Q_5^\top (\ell_3^\top S_3 \mathbf{x}_2)). \end{aligned}$$

Taking  $\ell_5$  as  $(V_5^\top (\ell_3^\top S_3 \mathbf{x}_2) - \mathbf{x}_2^\top S_5^\top (\ell_3^\top V_3)) \times (V_5^\top (\ell_3^\top Q_3 \mathbf{x}_1) - \mathbf{x}_1^\top Q_5^\top (\ell_3^\top V_3))$  gives the right null space of  $\mathcal{M} = \mathbf{x}_1 \mathbf{x}_2 \ell_3 \mathcal{T}$ . One may compute the left and right null spaces of  $\mathbf{x}_1 \mathbf{x}_2 \ell_4 \mathcal{T}$ ,  $\mathbf{x}_1 \mathbf{x}_2 \ell_5 \mathcal{T}$  and  $\ell_3 \ell_4 \ell_5 \mathcal{T}$  in an analogous fashion. ■

It follows from Lemma 3 that  $\mathbf{x}_1 \mathbf{x}_2 \ell_3 \mathcal{T}$  is of the form

$$\mathbf{x}_1 \mathbf{x}_2 \ell_3 \mathcal{T} = \mathbf{a}(\mathbf{x}_1) \mathbf{b}(\mathbf{x}_2)^\top + \mathbf{c}(\mathbf{x}_2) \mathbf{d}(\mathbf{x}_1)^\top + \mathbf{e} \mathbf{f}(\mathbf{x}_1, \mathbf{x}_2)^\top, \quad (20)$$

where  $\mathbf{f}(\mathbf{x}_1, \mathbf{x}_2)$  must be a linear combination of  $\mathbf{b}(\mathbf{x}_2)$  and  $\mathbf{d}(\mathbf{x}_1)$  so that  $\mathbf{x}_1 \mathbf{x}_2 \ell_3 \mathcal{T}$  be rank 2. This implies that the null space of  $\mathbf{x}_1 \mathbf{x}_2 \ell_3 \mathcal{T}$  is of the form  $\mathbf{b}(\mathbf{x}_2) \times \mathbf{d}(\mathbf{x}_1)$ .

As the choice of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  is arbitrary, we may fix  $\mathbf{x}_2$  and compute the null space of  $\mathbf{x}_1\mathbf{x}_2\ell_3\mathcal{T}$  for two different values of  $\mathbf{x}_1$ , say  $\mathbf{x}_{11}$  and  $\mathbf{x}_{12}$ . Taking the cross product of the two null space vectors gives the following linear combinations of  $V_5$  and  $S_5\mathbf{x}_2$ :

$$(\mathbf{b}(\mathbf{x}_2) \times \mathbf{d}(\mathbf{x}_{11})) \times (\mathbf{b}(\mathbf{x}_2) \times \mathbf{d}(\mathbf{x}_{12})) \sim \mathbf{b}(\mathbf{x}_2) \sim (\ell_3^\top S_3 \mathbf{x}_2) V_5 - (\ell_3^\top V_3) S_5 \mathbf{x}_2. \quad (21)$$

By repeating the above procedure for another choice of  $\ell_3$ , we obtain a second linear combination of  $V_5$  and  $S_5\mathbf{x}_2$ . The cross product of these two linear combinations is  $\widehat{V}_5 S_5 \mathbf{x}_2$ , which is the *nonrigid epipolar line* of  $\mathbf{x}_2$  in the 5th view according to the *nonrigid fundamental matrix*  $\widehat{V}_5 S_5$  relating the 2nd and 5th views. This leads to the following algorithm for recovering the camera rotations and the  $V$  vectors:

1. Choose  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  and  $\ell_3$  and compute the right null space  $\mathbf{a}_1$  of  $\mathbf{x}_1\mathbf{x}_2\ell_3\mathcal{T}$ . Repeat for another choice of  $\mathbf{x}_1$  to obtain  $\mathbf{a}_2$ . Set  $\mathbf{b}_1 = \mathbf{a}_1 \times \mathbf{a}_2$ . The vector  $\mathbf{b}_1$  is now proportional to  $V_5^\top (\ell_3^\top S_3 \mathbf{x}_2) - \mathbf{x}_2^\top S_5^\top (\ell_3^\top V_3)$ .
2. Repeat step 1 for a new choice of  $\ell_3$  to get  $\mathbf{b}_2$ . Set  $\mathbf{c} = \mathbf{b}_1 \times \mathbf{b}_2 \sim \widehat{V}_5 S_5 \mathbf{x}_2$ .
3. Repeat steps 1 and 2 for multiple choices of  $\mathbf{x}_2$  and linearly solve for the fundamental matrix  $\widehat{V}_5 S_5$  from  $\mathbf{c} \times \widehat{V}_5 S_5 \mathbf{x}_2 = 0$ . Subsequently solve for  $V_5$  and  $S_5$  using a modified version of the 8-point algorithm that enforces  $\lambda_f > 0$  for  $f = 1, 2, \dots, 5$ .<sup>2</sup>
4. Recover  $Q_5$  using steps 1-3, but allowing  $\mathbf{x}_2$  to vary instead of  $\mathbf{x}_1$ .
5. Recover  $S_4$ ,  $Q_4$  and  $V_4$  from the left null space of  $\mathbf{x}_1\mathbf{x}_2\ell_3\mathcal{T}$  in an analogous way.
6. Recover  $Q_3$ ,  $S_3$  and  $V_3$  from the left null space of  $\mathbf{x}_1\mathbf{x}_2\ell_5\mathcal{T}$  in an analogous way.

### 3.3 Recovering Shape Coefficients and Depths Via Factorization

Once we have the rotations and the  $V$  vectors we can return to our original system of equations (6) and solve for the shape coefficients and depths. To this end, let  $\mathbf{x}_{fp}$  be the image of point  $p = 1, \dots, P$  in frame  $f = 1, \dots, 5$  and let  $\lambda_{fp}$  be its depth. Also, let  $\gamma_f$  be the unknown scale up to which  $V_f$  is recovered. From (7) we have that

$$\begin{bmatrix} \widehat{\mathbf{x}}_{fp} R_f \mathbf{x}_{1p} & \widehat{\mathbf{x}}_{fp} R_f R_2^\top \mathbf{x}_{2p} & \widehat{\mathbf{x}}_{fp} V_f \end{bmatrix} \begin{bmatrix} c_{f1} \lambda_{1p} \\ c_{f2} \lambda_{2p} \\ \gamma_f \end{bmatrix} = 0, \quad f = 3, 4, 5. \quad (22)$$

We can solve these linear system for all  $f = 3, 4, 5$  and  $p = 1, \dots, P$  and build  $3 \times P$  matrices  $W_1$  and  $W_2$  whose  $(f, p)$  entries are given by

$$W_1(f, p) = \frac{c_{f1} \lambda_{1p}}{\gamma_f} \quad \text{and} \quad W_2(f, p) = \frac{c_{f2} \lambda_{2p}}{\gamma_f}, \quad (23)$$

<sup>2</sup> Given  $\widehat{V}_5 S_5$ , the 8-point algorithm for rigid scenes gives 4 solutions for  $(S_5, V_5)$ . The correct solution must satisfy  $\lambda_2, \lambda_5 > 0$ , which can be easily checked, because there are closed form formulae for  $\lambda_2$  and  $\lambda_5$  given  $(S_5, V_5)$  and  $(\mathbf{x}_2, \mathbf{x}_5)$ . In the nonrigid case, however,  $(\mathbf{x}_2, \mathbf{x}_5)$  are *not* images of the same point in 3D space, hence one cannot obtain closed form formulae for the depths. In fact, from the equation  $\lambda_f \mathbf{x}_f = \lambda_1 c_{1f} R_f \mathbf{x}_1 + \lambda_2 c_{2f} R_f R_2^\top \mathbf{x}_2 + V_f$ , we see that one can only solve for  $c_{1f} \lambda_1$ ,  $c_{2f} \lambda_2$ ,  $\lambda_3$ ,  $\lambda_4$  and  $\lambda_5$ . In order to check if  $\lambda_1, \lambda_2 > 0$ , we make the additional assumption that  $c_{if} > 0$ , and look for the pair  $(S_5, V_5)$  that results in the maximum number of positive depths. The assumption that  $c_{if} > 0$  corresponds to rearranging the frames so that the 1st and 2nd frame form a convex basis for the 3rd, 4th and 5th frames.

respectively. Notice that both  $W_1$  and  $W_2$  are rank-1 matrices, hence we can obtain the depths  $\{\lambda_{1p}\}$  and  $\{\lambda_{2p}\}$ , each one up to a different scale factor, from the SVD of  $W_1$  and  $W_2$ , respectively. Similarly, we can obtain the vector  $(c_{f1}, c_{f2})^\top$  up to a scale factor  $\gamma_f$ , also from the SVD of  $W_1$  and  $W_2$ .

### 3.4 Refining Shape Coefficients and Depths Via an Iterative Approach

The previous section detailed an algorithm for recovering the depths and shape coefficients using only factorization. However, factorization algorithms can and often do perform poorly in the presence of noise. To that end, we suggest the following iterative method (initialized by the previous factorization approach). First notice that if one knew the  $\lambda_{fp}$ 's or the  $c_{fi}$ 's in (22), then it should be possible to recover the others. In fact, by knowing the depths we can build a matrix  $G \in \mathbb{R}^{3P \times 3}$  whose rows have the form

$$G_p = [\lambda_{1p} \widehat{\mathbf{x}}_{fp} R_f \mathbf{x}_{1p} \quad \lambda_{2p} \widehat{\mathbf{x}}_{fp} R_f R_2^\top \mathbf{x}_{2p} \quad \widehat{\mathbf{x}}_{fp} V_f] \in \mathbb{R}^{3 \times 3}.$$

Similarly if we knew the coefficients, we could build a matrix  $H \in \mathbb{R}^{3F \times 3}$  with rows

$$H_f = [c_{f1} \widehat{\mathbf{x}}_{fp} R_f \mathbf{x}_{1p} \quad c_{f2} \widehat{\mathbf{x}}_{fp} R_f R_2^\top \mathbf{x}_{2p} \quad \widehat{\mathbf{x}}_{fp} V_f] \in \mathbb{R}^{3 \times 3}.$$

The null space of  $G$  or  $H$  gives the shape coefficients or the depths, respectively, in the first two frames. Thus one can iterate between these two steps until convergence. While this iterative method will give a correct estimate of the depths, we would like to point out that it does not give accurate coefficients due to the existence of intrinsic ambiguities which we discuss in the following section.

### 3.5 Ambiguities in Nonrigid Reconstruction

In this section, we discuss various ambiguities in nonrigid motion and shape recovery from multiple perspective views and relate them to those discussed in previous works such as [1]. It is important to understand that these ambiguities are not specific to our algorithm, but rather intrinsic to the problem of nonrigid shape and motion recovery under the assumption that a nonrigid shape is a linear combination of shape basis.

**Scale ambiguity.** It was very briefly discussed in [1] that there exists a scale ambiguity between the bases. However, the implications of this scale ambiguity or the primary cause of its existence were not discussed in any detail. We now refer the reader back to (23) and point out that when recovering the  $\lambda_1$ 's and  $\lambda_2$ 's, each quantity is being estimated independently from the other. Therefore, we can only recover the coefficients and the depths up to an unknown scale factor for each of the five frames. In the case of  $K$  shape bases, one may eliminate  $K + 1$  of these scales (assuming a maximal number of frames are used) by imposing the constraint that the coefficients sum to one. To a degree this is a physically meaningful constraint which simply enforces that the shapes in the scene be barycentric combinations of the shape bases. As a simple example one can consider the case of rigid motion. Rigid motion can be thought of as a scene with a single shape bases and in this case the shape coefficients must be one. Unfortunately, we can only eliminate the scales of the frames after the  $K$ th frame in this manner so

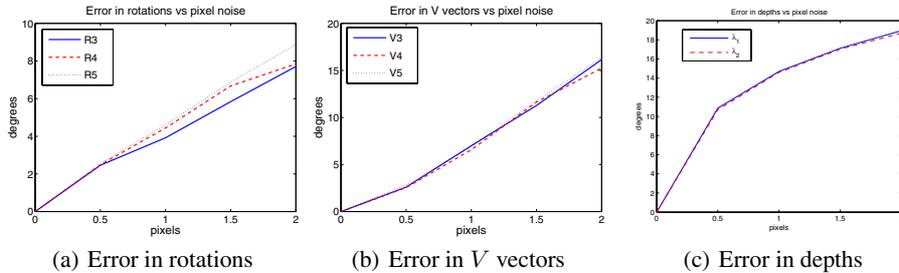
we are left with  $K$  extra scale factors. The problem is that it is impossible to determine the relative scale of one basis to another. Due to this inherent ambiguity, it is actually impossible to recover the correct shape coefficients and thus the shape bases. Therefore, our experimental results have focused upon the recovery of rotations,  $V$  vectors and the depths of the first two frames. For ground truth comparison, however, the true  $V$  vectors cannot be determined without knowing both the coefficients and the rotation and translation of the second frame. Therefore, in our real world examples, we focus solely on camera rotations.

**Translational ambiguity.** Note also that there is an ambiguity in the simultaneous reconstruction of the translations  $T_f$  and bases  $B_k$ , because  $B'_{kp} = B_{kp} + B$ , for  $p = 1, \dots, P$ , and  $T'_f = T_f - R_f B \sum_{k=1}^K c_{k,f}$  are also valid solutions for all  $B \in \mathbb{R}^3$ .

## 4 Experimental Results

To evaluate our algorithm effectiveness, we tested it on random synthetic experiments, a structured synthetic experiment similar to the two bases case in [13], and on a real world video sequence. As per the preceding analysis of ambiguities, we focus on the recovered rotations for the structured synthetic data and the real world experiments.

**Random synthetic data.** We randomly generated bases, coefficients, rotations and translations for an image of size 1000 by 1000. We run the iterative algorithm with 135 iterations. The error in rotation is calculated as  $\cos^{-1}((\text{trace}(R\hat{R}^\top) - 1)/2)$ , and the error in depth as the angle between true and estimated vector of all depths. Fig. 1 shows the mean errors averaged over 400 trials, except for a small percentage of outliers.<sup>3</sup>



**Fig. 1.** Mean error in depths, rotations, and  $V$  vectors after outlier removal

There are two primary sources of error: the noisy estimation of the tensor and the scale check of the eight point algorithm. If one uses the correct tensor, even with noisy data, the recovered estimates of rotation and the  $V$  vectors are virtually error free. The

<sup>3</sup> Due to the ambiguities in reconstruction, the algorithm gives large errors in a small percentage of trials. A trial was considered to be an outlier when the depth or rotation error was greater than 20 degrees or when the error in the  $V$  vectors was greater than 25 degrees.

**Table 1.** Rotation errors for a structured synthetic scene

	$R_3$ perspective error	$R_4$ perspective error	$R_5$ perspective error
Xiao’s approach	6.41°	4.43°	4.62°
Our approach	0.00°	0.00°	0.00°

primary cause of outliers would seem to be the scale check of the eight point algorithm. This scale check can in some cases fail to choose the correct solution. When this happens we generally see errors becoming quite close to 90 degrees.

**Structured synthetic data.** In this experiment we positioned 31 points in the following manner: 7 points were positioned on the corners of a unit cube at the origin, the remaining 24 points were divided into three groups of 8 points and then placed evenly along each of the coordinate axes. The 7 points were held fixed while the points along the axes translated in a positive direction along the axes and were perturbed by a small random amount in the respective off-axes directions. The camera was initially positioned to be at (20,20,20) facing the origin. The camera was then allowed to pan around the origin, and translate towards and away from the origin as the structure of the scene deformed. The data was projected in a perspective manner. We compared our algorithm to the one proposed in [13]. The mean errors over 400 trials are displayed in Table 1.

**Real world experiments.** The sequence shown in Fig. 2 was used to test our algorithm. 32 points were chosen by hand to generate the tensor estimates and another 8 static scene points were chosen in each frame to generate an 8-point algorithm estimate of the rotations, which was then used as our ground truth. The 1st and 5th images were taken as the reference, rather than the 1st and 2nd, hence rotations errors were measured for the 2nd, 3rd and 4th frame, relative to the 1st. The respective errors were 0.16°, 5.94° and 2.55°, which is expected, as frames 3 and 4 were the noisiest frames in the sequence.

**Fig. 2.** Real world test sequence

## 5 Summary and Conclusions

We have presented a geometric approach to nonrigid shape and motion recovery from multiple *perspective views*. We demonstrated that the constraints among multiple views of a nonrigid scene are multilinear, and proposed an algorithm for the reconstruction of two shape bases in five perspective views. We also examined the existence of intrinsic ambiguities in the reconstruction of nonrigid scenes.

**Acknowledgements.** This work has been supported by Hopkins WSE startup funds, and by grants NSF CAREER IIS-0447739 and ONR N000140510836.

## References

1. H. Aanaes and F. Kahl. Estimation of deformable structure and motion. In *ECCV Workshop on Vision and Modelling of Dynamic Scenes*, 2002.
2. M. Brand. Morphable 3D models from video. In *IEEE Conf. on Computer Vision & Pattern Recognition*, pages 456–463, 2001.
3. Matthew Brand and Rahul Bhotika. Flexible flow for 3D nonrigid tracking and shape recovery. In *IEEE Conf. on Computer Vision & Pattern Recognition*, pages 315–322, 2001.
4. C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *IEEE Conf. on Computer Vision & Pattern Recognition*, pages 2690–2696, 2000.
5. G. Doretto, A. Chiuso, Y. Wu, and S. Soatto. Dynamic textures. *International Journal of Computer Vision*, 51(2):91–109, 2003.
6. R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2000.
7. Y. Ma, Kun Huang, R. Vidal, J. Košecká, and S. Sastry. Rank conditions on the multiple view matrix. *International Journal of Computer Vision*, 59(2):115–137, 2004.
8. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography. *International Journal of Computer Vision*, 9(2):137–154, 1992.
9. L. Torresani and C. Bregler. Space-time tracking. In *European Conference on Computer Vision*, pages 801 – 812, 2002.
10. L. Torresani, D. Yang, E. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 493–500, 2001.
11. R. Vidal, Y. Ma, S. Soatto, and S. Sastry. Two-view multibody structure from motion. *International Journal of Computer Vision*, 2006.
12. R. Vidal and A. Ravichandran. Optical flow estimation and segmentation of multiple moving dynamic textures. In *IEEE Conf. on Computer Vision & Pattern Recognition*, pages 516–521, 2005.
13. Jing Xiao, Jin-Xiang Chai, and Takeo Kanade. A closed-form solution to non-rigid shape and motion recovery. In *European Conference on Computer Vision*, pages 573–587, 2004.