

# Non-Rigid Self-Calibration Of A Projective Camera

Hanno Ackermann and Bodo Rosenhahn

Leibniz University Hannover

**Abstract.** Rigid structure-from-motion (SfM) usually consists of two steps: First, a projective reconstruction is computed which is then upgraded to Euclidean structure and motion in a subsequent step. Reliable algorithms exist for both problems. In the case of non-rigid SfM, on the other hand, especially the Euclidean upgrading has turned out to be difficult. A few algorithms have been proposed for upgrading an *affine* reconstruction, and are able to obtain successful 3D-reconstructions. For upgrading a non-rigid *projective* reconstruction, however, either simple sequences are used, or no 3D-reconstructions are shown at all.

In this article, an algorithm is proposed for estimating the self-calibration of a projectively reconstructed non-rigid scene. In contrast to other algorithms, neither prior knowledge of the non-rigid deformations is required, nor a subsequent step to align different motion bases. An evaluation with synthetic data reveals that the proposed algorithm is robust to noise and it is able to accurately estimate the 3D-reconstructions and the intrinsic calibration. Finally, reconstructions of a challenging real image with strong non-rigid deformation are presented.

## 1 Introduction

Approaches for *rigid* structure-from-motion (SfM) usually consist of two steps. Given 2D-feature correspondences between several images, a projective reconstruction is estimated which is identical to the true solution up to a projective transformation. In a second step, usually referred to as *self-calibration* or *auto-calibration*, this projective distortion is removed by imposing a certain structure on the motion matrices [1]. Assuming the basis model introduced by Bregler *et al.* in [2], we consider the problem of computing the self-calibration of a *projective* camera which observes a *non-rigidly* deforming body or scene. We assume that this camera has an unknown focal length which may vary or be constant, zero skew and principal point at the origin. Furthermore, the proposed algorithm is more general than other works as particular non-rigid deformations need not be known.

Self-calibrating a projective camera can be considered a mature field if the observed body is *rigid* [3–6].

In the case of a *non-rigid* body observed by an *affine* camera, Xiao *et al.* [7] proposed a linear solution. Brand [8] suggested an algorithm in which the

motion constraints are first imposed for a particular, arbitrarily chosen deformation mode, and all other deformation modes are corrected with respect to the initially chosen one, an approach which is non-optimal as the error is concentrated in all deformation modes but the reference one. Olsen and Bartoli [9] used a smoothness prior on the camera motion to determine the self-calibration. Torresani *et al.* [10] imposed the prior knowledge that the coefficients of non-rigid deformation satisfy a Gaussian distribution. In a seminal work, Paladini *et al.* [11] introduced an iterative projection algorithm which alternates unconstrained optimization with projection of the motion matrices to the required structure.

To this day, only two algorithms consider the problem of self-calibrating a *projective* camera observing a body deforming non-rigidly. Xiao and Kanade [12] extended their work from [7] to a projective camera with constant focal length. Hartley and Vidal [13] proposed a method which requires that the intrinsic camera parameters are fixed and known. Similar to [8] they first correct a particular, arbitrarily chosen deformation mode. Remaining modes are subsequently estimated with respect to the previously corrected ones. While being an elegant, non-iterative solution, no 3D-reconstructions are shown in this article.

In this article, an algorithm is presented for self-calibration of a projective camera observing a non-rigidly deforming object. It is assumed that the skew is zero, the focal length unknown while varying or being constant throughout the sequence, and the principal point is at the origin. Though seemingly similar to the requirements in [12], the current work does not demand particular non-rigid deformation coefficients to be known. Furthermore, the proposed algorithm does not require a second step (*Orthogonal Procrustes Analysis*) to enforce identical rotations. The advantage is that the error should be more fairly distributed between the bases. To align the bases, additional constraints are necessary. We therefore generalize the equations introduced by Brand [14] to the projective camera model. It is proven that the solution is unique up to a global rotation and reflection of the world coordinate system and individual scalings of each basis. The accuracy of the proposed algorithm is evaluated with experiments on synthetic data. Furthermore, 3D-reconstructions are presented for a challenging real-image sequence showing a body with strong local and global non-rigid deformation.

This work is structured as follows: In Section 2, the problem of self-calibrating a projective camera observing a non-rigidly deforming body or scene is defined. Constraints by which the problem can be determined are derived in Section 3. It will be proven that these constraints are necessary and sufficient to obtain the required structure of the motion matrices. Synthetic and real image experiments are presented in Section 4 before a summary and conclusions in Section 5.

Capital letters denote matrices, bold capital letter scalar constants and bold lower-case letters vectors. Normal lower-case letters denote scalar variables or counters.

## 2 Problem Definition

Let there be  $\mathbf{K}$   $4 \times n$  *basis shape matrices*  $X_k$ ,  $k = 1, \dots, \mathbf{K}$ , consisting of  $n$  homogeneous 3D-points  $X^j$ ,  $j = 1, \dots, \mathbf{N}$ , each,  $\mathbf{M}$  images with the  $3 \times 4$  projection matrices  $P^i$ ,  $i = 1, \dots, \mathbf{M}$  and mixing coefficients  $\alpha_k^i$  blending the  $\mathbf{K}$  basis shapes

$$\lambda_{ij} \mathbf{x}_{ij} = P^i \left( \sum_{k=1}^{\mathbf{K}} \alpha_k^i X_k \right). \quad (1)$$

The linear mixing model was introduced by Bregler *et al.* [2] for an affine camera model. Here, the scalars  $\lambda_{ij}$  are the *projective depths* necessary for Eq. (1) to hold true under perspective projection. The projection matrices  $P^i$  consist of the orientations  $R^i$ , positions  $\mathbf{t}^i$  and calibrations  $K^i$  of the cameras<sup>1</sup>

$$P^i = K^i [R^i | \mathbf{t}^i], \quad K^i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

with  $f_i$  being the unknown focal length of the  $i$ th camera.

It can be seen that the measurement matrix  $W$  consisting of all 2D-features  $\mathbf{x}_{ij}$  rescaled with the correct projective depths  $\lambda_{ij}$  has rank  $3\mathbf{K} + 1$  if the two matrices  $P$  and  $X$  each have rank  $3\mathbf{K} + 1$

$$W = \begin{bmatrix} \lambda_{11} \mathbf{x}_{11} & \cdots & \lambda_{1n} \mathbf{x}_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{\mathbf{M}1} \mathbf{x}_{\mathbf{M}1} & \cdots & \lambda_{\mathbf{M}n} \mathbf{x}_{\mathbf{M}n} \end{bmatrix} = \underbrace{\begin{bmatrix} \alpha_1^1 K^1 R^1 & \cdots & \alpha_{\mathbf{K}}^1 K^1 R^1 & K^1 \mathbf{t}^1 \\ \vdots & & \vdots & \vdots \\ \alpha_1^{\mathbf{M}} K^{\mathbf{M}} R^{\mathbf{M}} & \cdots & \alpha_{\mathbf{K}}^{\mathbf{M}} K^{\mathbf{M}} R^{\mathbf{M}} & K^{\mathbf{M}} \mathbf{t}^{\mathbf{M}} \end{bmatrix}}_P \cdot \underbrace{\begin{bmatrix} X_1 \\ \vdots \\ X_{\mathbf{K}} \\ \mathbf{1} \end{bmatrix}}_X \quad (3)$$

Given all projective depths  $\lambda_{ij}$ , for instance by the algorithms proposed in [15, 16], the matrix  $W$  can be factorized by singular value decomposition by Eq. (1)

$$W = U \Sigma V^T, \quad (4)$$

where  $U \in \mathbb{R}^{3m \times (3\mathbf{K}+1)}$ ,  $\Sigma \in \mathbb{R}^{(3\mathbf{K}+1) \times (3\mathbf{K}+1)}$ , and  $V \in \mathbb{R}^{(3\mathbf{K}+1) \times n}$ . We may consider  $U$  as projectively distorted camera matrices  $P$ , and  $\Sigma V$  as structure matrix  $X$  perturbed by the inverse distortion.

The problem of non-rigid projective self-calibrating is to determine a  $(3\mathbf{K} + 1) \times (3\mathbf{K})$  matrix  $A$  which transforms  $U$  such that  $UA$  satisfies the required structure of the first  $3\mathbf{K}$  columns of  $P$ , *i.e.* each row triple of  $UA$  must consist of scaled instances of a rotation  $R^i$  distorted by some  $K^i$ .

<sup>1</sup> With some risk of confusion, we use the symbol  $K^i$  for the intrinsic camera calibration in the  $i$ th image whereas the bold letter  $\mathbf{K}$  denotes the number of basis shapes.

### 3 Deriving Constraints on Non-Rigid Self-Calibration

Let  $U^i$  denote the  $i$ th row triple of  $U$ . Straightforwardly applying the derivation of the dual absolute quadric of rigid scenes to the non-rigid case, we arrive at

$$\omega_i = \mathbf{K} \begin{bmatrix} f_i^2 & 0 & 0 \\ 0 & f_i^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{\gamma_i^2 \beta_i} U^i A A^\top U^{i\top} \quad (5)$$

where  $\omega_i$  denotes the dual image of the absolute conic  $\omega_i = K_i K_i^\top$  at image  $i$ ,  $\beta_i = ((\alpha_1^i)^2 + \dots + (\alpha_K^i)^2)$ , and the scalars  $\gamma_i$  account for the perspective projection in image  $i$ . The positive-semidefinite  $(3\mathbf{K}+1) \times (3\mathbf{K}+1)$  matrix  $\Omega_\infty = A A^\top$  of rank  $3\mathbf{K}$  is the extension of the dual absolute quadric to the non-rigid case.

It is obvious that Eq. (5) is ambiguous: any change in  $\gamma_i$ , for instance can be compensated by a scaling of  $\beta_i$ . Similarly, and scaling of all  $\alpha_k^i$ ,  $i = 1, \dots, \mathbf{M}$  requires an inverse scaling on the  $k$ th structure basis  $X_k$ .

Given  $\omega$  as defined in Eq. (5), we can obtain four equations per image for determining  $\Omega_\infty = A A^\top$

$$\mathbf{u}_a^{i\top} A A^\top \mathbf{u}_b^i = 0, \quad (6a)$$

$$\mathbf{u}_a^{i\top} A A^\top \mathbf{u}_a^i - \mathbf{u}_b^{i\top} A A^\top \mathbf{u}_b^i = 0 \quad (6b)$$

where  $\mathbf{u}_{\{a,b\}}^{i\top}$ ,  $a \neq b$ , denotes the first, second, or third row of  $U^i$ . Equations (6) are the so-called *orthogonality constraints* derived by Xiao *et al.* for the problem of self-calibrating an affine [7] or projective camera [12].

While it seems straightforward to determine  $\Omega_\infty$  by solving Eq. (6), it was shown that even the affine problem is indeterminate [7, 17]. With a slight risk of confusion, denote by  $P^i$  the row triple corresponding to image  $i$  of matrix  $P$  in Eq. (3). In the case of a projective camera, we obtain for the ambiguity:

**Lemma 1.** *Let there be a  $3\mathbf{K} \times 3\mathbf{K}$  matrix  $D$ ,*

$$D = \begin{bmatrix} d_{11}O_1 & d_{12}O_2 & d_{13}O_3 & & \\ d_{21}O_1 & d_{22}O_2 & d_{23}O_3 & \cdots & \\ & & \vdots & & \end{bmatrix}, \quad (7)$$

where  $d_{ab}$  are scalar factors and the  $3 \times 3$  matrices  $O_c$ ,  $c = 1, \dots, \mathbf{K}$ , are arbitrary elements of the orthogonal group, i.e.  $O_c O_c^\top = I$ .

Then, Eqs. (6) are always satisfied for  $\Omega_\infty = D D^\top$ , yet  $P^i$  and  $P^i D$  are not invariant up to a similarity transformation.

*Proof.* Assume a general deformation matrix

$$D = \begin{bmatrix} d_{11}D_{11} & \cdots & d_{1K}D_{1K} \\ \vdots & \ddots & \vdots \\ d_{K1}D_{K1} & \cdots & d_{KK}D_{KK} \end{bmatrix}, \quad (8)$$

where the  $3 \times 3$  matrices  $D_{ab}$ , and the scalars  $d_{ab}$  are arbitrary. Letting  $S = DD^\top$ ,  $S_{ab}$   $3 \times 3$  blocks of  $S$ ,  $l_{ab}$  be sums of the  $d_{ab}$ , and

$$S' = [(l_1^i l_1^i S_{11} + \dots + l_K^i l_1^i S_{K1}) + \dots + (l_1^i l_K^i S_{1K} + \dots + l_K^i l_K^i S_{KK})], \quad (9)$$

we obtain the three equations

$$\gamma_i^2 \beta_i f_i^2 = \mathbf{r}_1^{i\top} S' \mathbf{r}_1^i \quad (10a)$$

$$\gamma_i^2 \beta_i f_i^2 = \mathbf{r}_2^{i\top} S' \mathbf{r}_2^i \quad (10b)$$

$$0 = \mathbf{r}_a^{i\top} S' \mathbf{r}_b^i, \quad a \neq b, \quad (10c)$$

where  $\mathbf{r}_{\{1,2,3\}}^i$  denotes the first, second, or third row vector of  $R^i$ .

If we take the  $3 \times 3$  matrices  $D_{ak} = O_k$ ,  $a = 1, \dots, \mathbf{K}$ , all the matrices  $S_{ab}$  are scaled identity matrices,  $S_{ab} = s_{ab}I$ , for arbitrary scalars  $s_{ab}$ , hence Equations (10) are always satisfied.  $\square$

Please notice that the third rows of the rotation matrices are only constrained by the orthogonality constraint (10c). Since  $\mathbf{r}_3^{i\top} S' \mathbf{r}_3^i = \gamma_i^2 \beta_i$ , the lengths of the third rows are arbitrary. As the equations including the focal length depend on depend on the third row (by  $\gamma_i^2$  and  $\beta_i$ ), the focal lengths are also arbitrary, therefore<sup>2</sup>.

Furthermore, the Equations (10) do not define constraints *between*  $A_{k_1}$  and  $A_{k_2}$ ,  $k_1 \neq k_2$ ,  $A = [A_1 \dots A_{\mathbf{K}}]$ . Brand gave such constraints in [14] for an affine camera. Due to the affine model, they only define constraints on the first two rows, hence the ambiguity between focal lengths and projective depths as well as non-rigid mixing coefficients remains.

The problem is thus to define constraints between the different  $A_{k_1}$  and  $A_{k_2}$ , and on the third rows  $\mathbf{u}_3^{i\top} A_k$ . We now arrive at the central contribution of this article, namely additional constraints for constraining the self-calibration matrix  $A$  of a projective camera.

**Theorem 1.** *Given projectively distorted  $3 \times (3\mathbf{K} + 1)$  matrices  $U^i$ , a matrix  $A = [A_1 \dots A_{\mathbf{K}}]$  satisfying Eqs. (6) and*

$$\left( \mathbf{u}_a^{i\top} A_{k_1} A_{k_2}^\top \mathbf{u}_a^i \right)^2 - \left( \mathbf{u}_a^{i\top} A_{k_1} A_{k_1}^\top \mathbf{u}_a^i \right) \cdot \left( \mathbf{u}_a^{i\top} A_{k_2} A_{k_2}^\top \mathbf{u}_a^i \right) = 0 \quad (11a)$$

$$\begin{aligned} & \left( \mathbf{u}_1^{i\top} A_{k_1} A_{k_1}^\top \mathbf{u}_1^i \right) \cdot \left( \mathbf{u}_3^{i\top} A_{k_2} A_{k_2}^\top \mathbf{u}_3^i \right) - \\ & \left( \mathbf{u}_1^{i\top} A_{k_2} A_{k_2}^\top \mathbf{u}_1^i \right) \cdot \left( \mathbf{u}_3^{i\top} A_{k_1} A_{k_1}^\top \mathbf{u}_3^i \right) = 0 \end{aligned} \quad (11b)$$

for  $a = \{1, 2, 3\}$  and  $k_1 \neq k_2$  in the unknown column triples  $A_k$  of  $A$  transforms a projectively distorted  $U$  to the structure required by Eq. (3). Equations (11) are necessary and sufficient to transform matrices  $U^i A$  such that the column

<sup>2</sup> Such an indeterminacy could be attractive to fit a non-rigid model if some or all focal lengths are *a-priorily* known.

triples  $U^i A_k$  constitute aligning orthogonal systems, and the lengths of the first two vectors  $\mathbf{u}_{\{1,2\}A_{k_1}}^{i\top}$  of any basis  $k_1$  and the lengths of the first two vectors of any other basis  $k_2$  are related by multiplication with  $(\alpha_{k_2}^i)^2$  and  $(\alpha_{k_1}^i)^2$ .

*Proof. Necessity:* By Eqs. (6), the six vectors  $\mathbf{u}_{\{1,2,3\}A_{k_1}}^{i\top}$  and  $\mathbf{u}_{\{1,2,3\}A_{k_2}}^{i\top}$  form two systems of orthogonal vectors. Provided sufficiently many images,

$$\frac{\mathbf{u}_a^{i\top} A_{k_1} A_{k_2}^\top \mathbf{u}_a^i}{\|\mathbf{u}_a^{i\top} A_{k_1}\| \cdot \|\mathbf{u}_a^{i\top} A_{k_2}\|} = 1 \quad (12)$$

holds true if and only if each pair of vectors  $\mathbf{u}_a^{i\top} A_{k_1}$  and  $\mathbf{u}_a^{i\top} A_{k_2}$  points into the same direction, thus Equation (11a) imposes that the two systems of orthogonal vectors align for  $a = \{1, 2, 3\}$ . Equations (3) and (5) further require that

$$\frac{\mathbf{u}_1^{i\top} A_{k_1} A_{k_1}^\top \mathbf{u}_1^i}{\mathbf{u}_3^{i\top} A_{k_1} A_{k_1}^\top \mathbf{u}_3^i} = \frac{\mathbf{u}_1^{i\top} A_{k_2} A_{k_2}^\top \mathbf{u}_1^i}{\mathbf{u}_3^{i\top} A_{k_2} A_{k_2}^\top \mathbf{u}_3^i} = (\phi^i)^2 \quad (13)$$

for some scalar variables  $\phi^i$  from which we obtain Eq. (11b).

*Sufficiency:* If  $A$  satisfies the Eqs. (11), the matrix  $U^i A$  has the following structure

$$U^i A = \begin{bmatrix} \phi^i & 0 & 0 \\ 0 & \phi^i & 0 \\ 0 & 0 & 1 \end{bmatrix} [\sigma_1^i R^i \cdots \sigma_K^i R^i] \quad (14)$$

for some scalars  $\sigma^i$ . □

Please notice that Eq. (11a) has to be imposed for all three vectors  $\mathbf{u}_a^i$ ,  $a = \{1, 2, 3\}$  in order to define a constraint on  $\gamma_i^2 \beta_i$ .

If the focal length is known to be constant yet unknown, we can impose that constraint by requiring that  $\sigma^1 \phi^1 = \cdots = \sigma^M \phi^M$ . In the following, denote by  $i_1$  and  $i_2$  two different image numbers.

**Corollary 1.** *The equation*

$$\left( \mathbf{u}_{1A_k}^{i_1\top} A_k A_k^\top \mathbf{u}_{1A_k}^{i_1} \right) \cdot \left( \mathbf{u}_{3A_k}^{i_2\top} A_k A_k^\top \mathbf{u}_{3A_k}^{i_2} \right) - \left( \mathbf{u}_{1A_k}^{i_2\top} A_k A_k^\top \mathbf{u}_{1A_k}^{i_2} \right) \cdot \left( \mathbf{u}_{3A_k}^{i_1\top} A_k A_k^\top \mathbf{u}_{3A_k}^{i_1} \right) = 0 \quad (15)$$

for  $i_1 \neq i_2$  imposes constant focal length throughout the images.

*Proof.* We must require that any  $\phi^{i_1}$  equals any other  $\phi^{i_2}$  for  $i_1 \neq i_2$ , hence we obtain from Eq. (11)

$$\frac{\mathbf{u}_{1A_k}^{i_1\top} A_k A_k^\top \mathbf{u}_{1A_k}^{i_1}}{\mathbf{u}_{3A_k}^{i_1\top} A_k A_k^\top \mathbf{u}_{3A_k}^{i_1}} = \frac{\mathbf{u}_{1A_k}^{i_2\top} A_k A_k^\top \mathbf{u}_{1A_k}^{i_2}}{\mathbf{u}_{3A_k}^{i_2\top} A_k A_k^\top \mathbf{u}_{3A_k}^{i_2}} \quad (16)$$

from which Eq. (15) follows directly. □

The set of Eqs. (6) and (11) impose the required structure on the matrices  $U^i$ . The question is the remaining ambiguity.

**Lemma 2.** *Given a transformation  $A$  satisfying the sets of Eqs. (10) and (11) which brings each  $U^i$  to the required structure, scalars  $d_k$ ,  $k = 1, \dots, K$ , and an arbitrary  $3 \times 3$  matrix  $O_g$  which is an element of the orthogonal group, i.e.  $O_g O_g^\top = I$ , then  $A$  is ambiguous up to multiplication with a matrix  $D$*

$$D = \begin{bmatrix} d_{11}O_g & d_{12}O_g & & \\ d_{21}O_g & d_{22}O_g & \cdots & \\ & & \vdots & \end{bmatrix}. \quad (17)$$

*Proof.* To satisfies Eqs. (10), we may assume that  $D$  has the structure as defined lemma (1). Let  $D_k$  denote the  $k$ th column triple of  $D$ , and let

$$S_{kk} = D_k D_k^\top = \begin{bmatrix} d_{11}^2 I & d_{1k} d_{2k} I & \cdots & d_{1k} d_{Kk} I \\ & \vdots & & \\ d_{Kk} d_{1k} I & d_{Kk} d_{2k} I & \cdots & d_{Kk}^2 I \end{bmatrix} \quad \text{and} \quad (18a)$$

$$S_{k_1 k_2} = D_{k_1} D_{k_2}^\top = \begin{bmatrix} d_{1k_1} d_{1k_2} O_{k_1} O_{k_2}^\top & \cdots & d_{1k_1} d_{Kk_2} O_{k_1} O_{k_2}^\top \\ & \vdots & \\ d_{Kk_1} d_{1k_2} O_{k_1} O_{k_2}^\top & \cdots & d_{Kk_1} d_{Kk_2} O_{k_1} O_{k_2}^\top \end{bmatrix} \quad (18b)$$

where  $I$  denotes the  $3 \times 3$  identity matrix, and  $O_{k_1}$  and  $O_{k_2}$ ,  $k_1 \neq k_2$ , are  $3 \times 3$  matrices of the orthogonal group.

Let  $P^i$  denote the row triple of  $P$  corresponding to the  $i$ th image. Then, we have

$$P^i S_{k_1 k_2} P^{i\top} = ((\alpha_1^i)^2 d_{1k_1} d_{1k_2} + \dots + (\alpha_K^i)^2 d_{Kk_1} d_{Kk_2}) K^i R^i O_{k_1} O_{k_2}^\top R^{i\top} K^{i\top} \quad \text{and} \quad (19a)$$

$$P^i S_{kk} P^{i\top} = ((\alpha_1^i)^2 d_{1k}^2 + \dots + (\alpha_K^i)^2 d_{Kk}^2) K^i K^{i\top} \quad (19b)$$

From Eq. (19a) and Eq. (11a), we can see that  $O_{k_1}$  and  $O_{k_2}$  must be identical if there are sufficiently many images. Equation (11b) imposes no further constraints on the structure of  $D$ .  $\square$

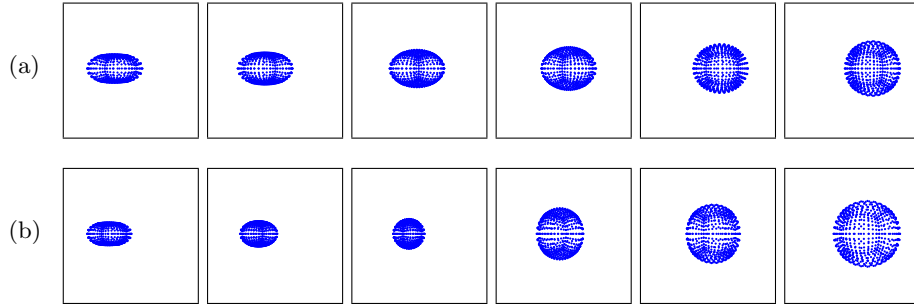
Lemma 2 implies that any matrix  $A$  satisfying Eqs. (6) and (11) is unique up to a global rotation and reflection of the world coordinate system. Furthermore, the bases are unique up an individual scaling of each basis.

Minimizing Eqs. (10) and (11) amounts to minimizing the Frobenius-norm

$$\left\| U^i A A^\top U^{i\top} - K^i K^{i\top} \right\|_F. \quad (20)$$

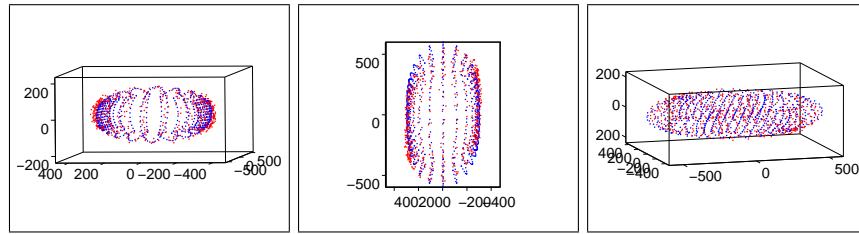
Since minimizing the Frobenius-norm of  $A$  is equivalent to minimizing its singular values<sup>3</sup>, it is necessary to prevent a rank-degeneracy of  $A$ . We therefore impose

<sup>3</sup> Since  $\|A\|_F = \sqrt{\sum_i \sigma(A)_i^2}$  where  $\sigma(A)_i$  is the  $i$ th singular value of  $A$ .



**Fig. 1.** (a) Six images of a sequence of 25 images showing an ellipsoid morphing into a sphere. At each image the 3D-shape rotates by  $7.2^\circ$  around the  $y$ -axis (upwards) and translates in direction of the  $x$ -axis. The focal length is constant throughout the sequence. (b) Same structure and motion while the focal length changes between images 1-12 and 13-25.

the constraint that the smallest singular value of  $A$  is larger than 0.1. This constraint also prevents the trivial solution due to the scalar factors  $\gamma_i$  and  $\beta_i$  in Eq. (5).



**Fig. 2.** Example of a 3D-reconstruction if the data is contaminated with normally distributed noise (blue: reconstructed shape; red: ground truth shape). The standard deviation was set to 1% of the maximum variation in  $x$ ,  $y$ , and  $z$ -direction.

## 4 Experiments

### 4.1 Synthetic Image Experiments

For synthetic evaluation we created a 25-image sequence consisting of 726 3D-points of an ellipsoid morphing into a sphere. Six images of this sequence are shown in Fig. 1(a). At each image the 3D-shape rotates by  $7.2^\circ$  around the  $y$ -axis while translating in direction of the  $x$ -axis.



To measure the influence of noise, we added normally distributed noise with standard deviation set to 0% to 3.0% in steps of 0.5% of the maximum variation in  $x$ ,  $y$  and  $z$ -direction. For each noise level, we created 10 contaminated data sets to compute average errors. As error measure, we took the average of the Euclidean distance between the 3D-points of the ground truth shape and the reconstruction after translating it so that the centroids of both point clouds coincide since the dual absolute quadric constraint ignores the  $(3\mathbf{K}+1)$ st column of matrix  $P$  in Eq. (3). We normalized this number by the Frobenius norm of the ground truth shape

$$\epsilon = \frac{1}{n} \frac{\|X^{gt} - X^{est}\|_F}{\|X^{gt}\|_F}. \quad (21)$$

Here,  $X^{gt}$  denotes the matrix consisting of the ground truth 3D-points (for simplicity we omitted an index denoting the image number), and  $X^{est}$  the matrix consisting of the the estimated 3D-points. The symbol  $\|\cdot\|_F$  denotes the Frobenius norm.

For optimization, we use *semi quadratic programming*. Since the algorithm is susceptible to local minima, we randomly initialize it 40 times and take the best result.

We reconstructed 3D-shapes using two basis shapes ( $K = 2$ ). Figure 3, left plot, shows a the average error as the noise increases. As can be seen, the proposed method is quite robust with respect to noise. In the right plot of Fig. 3, we show average errors *per image* for noise levels 0%, 1% and 2%. The error is not evenly distributed yet there are no exceptional spikes.

For a second experiment, we used the same structure and motion shown in Fig. 1(a) yet changed the focal length between images 1-12 and 13-24. This sequence is shown in Fig. 1(b). The left plot of Fig. 4 shows the reconstruction errors. The right plot of this figure shows the reconstruction errors per image.

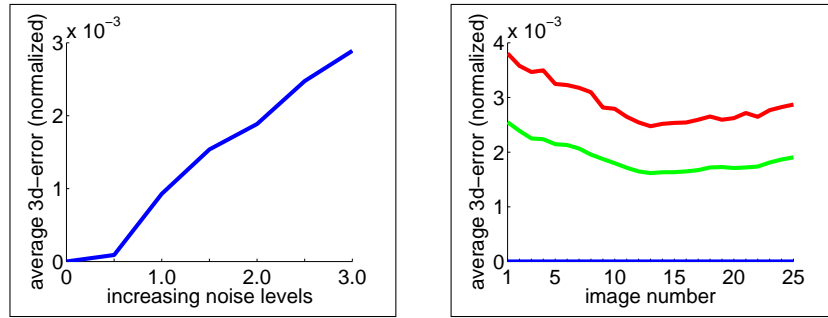
To evaluate the estimated calibration matrices we computed the following error metric

$$\epsilon_i = \frac{1}{9} \left\| \frac{1}{\gamma_i^2 \beta_i} U^i A A^\top U^{i\top} - K^i K^{i\top} \right\|_F. \quad (22)$$

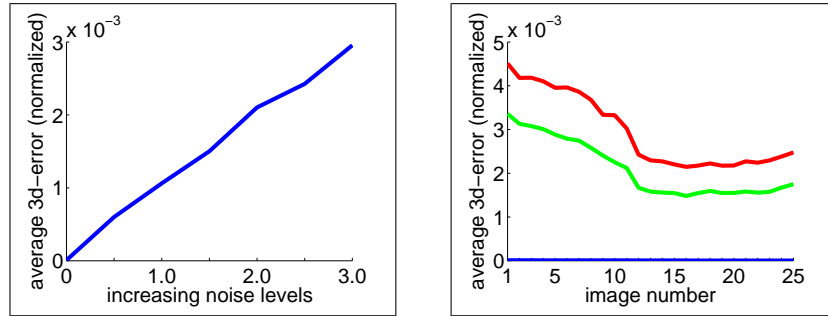
The left plot in Fig. 5 shows the calibration errors for constant focal length (corresponding to the sequence shown in Fig. 1(a)), the right plot for varying  $f$  (Fig.1(b)). Apparently, the proposed algorithm can handle constant and changing focal lengths well.

The average estimated focal lengths per image are shown in Fig. 6. The left plot shows the estimations for constant  $f = 5$  whereas the right plot shows them for  $f = 4$  in images 1 until 12 and  $f = 6$  in images 13 until 25. It can be seen that under noise, the algorithm deviates more from the true values as each image induces its own estimate of the focal length.

Figure 2 shows an example of the reconstructed 3D-shape in the first if the data is perturbed with noise of standard deviation 1%. Blue points denote estimated 3D-points, red points the ground truth. Apparently, the estimated points and the ground truth points almost coincide.



**Fig. 3.** Left: Average 3D-errors for increasing levels of noise with constant yet unknown focal length (corresponding to the sequence shown in Fig. 1(a)). Right: Average 3D-error per image for noise levels of 0% (solid blue line), 1% (dash-dotted green line) and 2% (dashed red line).



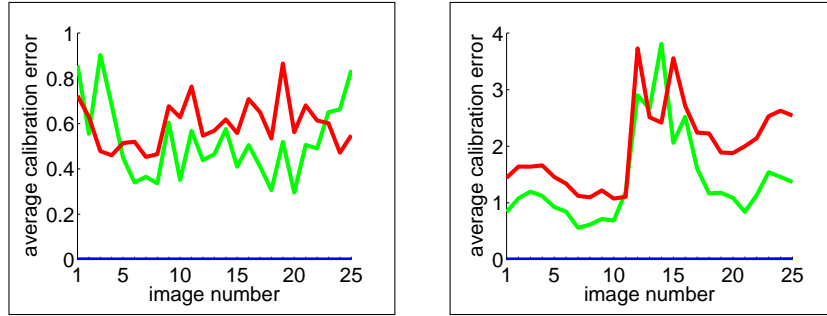
**Fig. 4.** Average 3D-errors for a changing focal length and the sequence shown in Fig. 1(b). Left: Average 3D-errors for increasing levels of noise. Right: Average 3D-error per image for noise levels of 0% (solid blue line), 1% (dash-dotted green line) and 2% (dashed red line).

## 4.2 Real Image Experiments

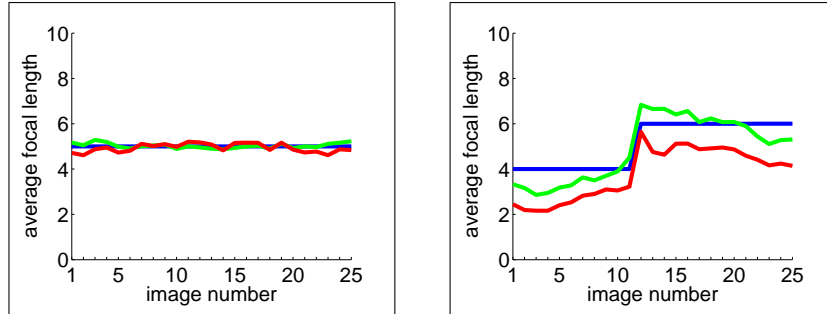
Figure 7 shows six images of a 25-image sequence. It shows a box whose sides and top paper deform non-rigidly. Please notice that the top paper exhibits strong deformations which cannot be explained by a multi-body or articulated chain model. A total of 375 points were tracked throughout the sequence.

For projective 3D-reconstruction we used the algorithms proposed in [15, 16] which amounts to camera resectioning and intersectioning. We assumed two rigid basis shapes ( $K = 2$ ) and thus optimized for a rank of 7 of the observation matrix.

3D-reconstructions of the shapes observed in every fifth image are shown in Fig. 8. From left to right are shown the image number, the 3D-reconstruction



**Fig. 5.** Left: Average calibration errors for different levels of noise (blue: no noise, green:  $\sigma = 1.0$ , red:  $\sigma = 2.0$ ) per image. Left: constant focal length; right: focal length varies between images 12 and 13.



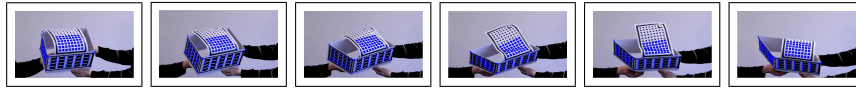
**Fig. 6.** Left: Average focal lengths for different levels of noise (blue: no noise, green:  $\sigma = 1.0$ , red:  $\sigma = 2.0$ ) per image. Left: constant focal length  $f = 5$ ; right: focal length varies:  $f = 4$  in images 1 until 12 and  $f = 6$  in images 13 until 25.

corresponds to, the image, a top view of the estimated shape, a side view (from left), a frontal view, and another side view from the right.

The planar sides of the box show a strong perspective distortion. This is due to the estimated projective depths. The configuration of the frontal and the left plane to each other closely reflect the shape of the box in the images. The non-rigid bending of the 3D-points on the top structure also closely resembles the shape of the top paper in the images. Overall, the reconstruction looks reasonable.

## 5 Summary and Conclusions

The contributions made in this article can be summarized as follows: Considering a pinhole camera with unknown focal length which may be varying or constant, the problem considered in this work was to determine the Euclidean upgrading



**Fig. 7.** Six images of a 25-image sequence with 375 trajectories showing a box deforming non-rigidly. The top paper deforms non-rigidly, so a multi-body model would not be satisfied.

if this camera observes a non-rigidly deforming object or scene. To align all motion bases simultaneously during optimization, *i.e.* enforce identical rotations, constraints were derived which allow joint estimation of all motion bases. In terms of error distribution such a joint estimation should be more fair with respect to the different bases.

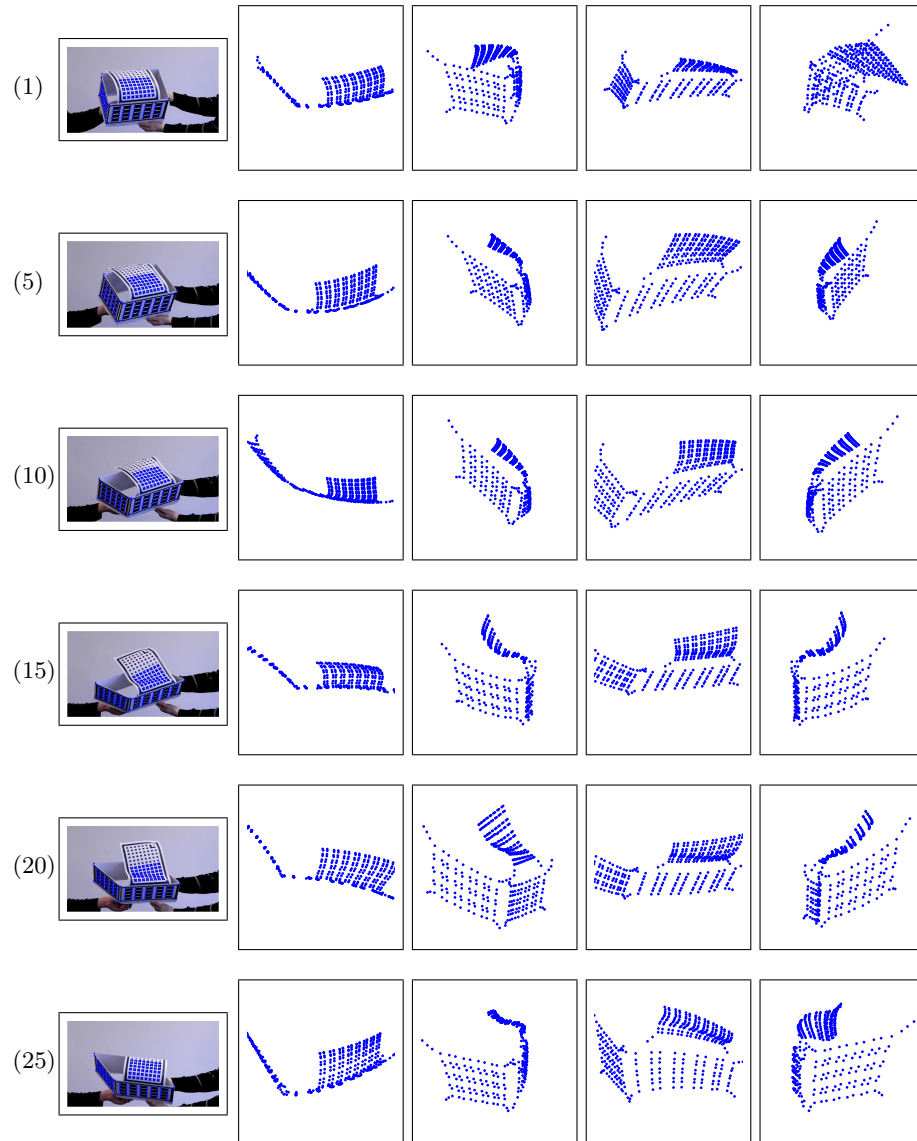
It was proven that the upgrading transformation is unique up to rotation and reflection of the world coordinate system and individual scalings of each basis. By evaluation of synthetic data as well as a 3D-reconstruction of a difficult real image sequence in which the object exhibits highly non-rigid distortion, it was shown that the proposed algorithm is indeed quite robust to increasing noise and able to reconstruct accurate 3D-shapes.

In future works we will focus on generalizing the camera model to a fully projective model whose intrinsic parameters are all varying and unknown. Furthermore, means of global optimization will be investigated.

## References

1. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)
2. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: IEEE Computer Vision and Pattern Recognition (CVPR), Hilton Head, SC, USA (2000) 690–696
3. Triggs, B.: Autocalibration and the absolute quadric. In: Conf. Comp. Vis. and Pat. Recog. (CVPR). (1997)
4. Pollefeys, M., Koch, R., van Gool, L.: Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. Int. J. Comp. Vis. (IJCV) **32** (1999) 7–25
5. Seo, Y., Heyden, A.: Auto-calibration by linear iteration using the DAC equation. Img. Vis. Comp. **22** (2004) 919–926
6. Chandraker, M., Agarwal, S., Kahl, F., Nistér, D., Kriegman, D.: Autocalibration via rank-constrained estimation of the absolute quadric. In: Conf. Comp. Vis. and Pat. Recog. (CVPR). (2007)
7. Xiao, J., Chai, J., Kanade, T.: A closed-form solution to non-rigid shape and motion recovery. International Journal of Computer Vision **67** (2006) 233–246
8. Brand, M.: A direct method for 3D factorization of nonrigid motion observed in 2d. In: IEEE Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA (2005) 122–128
9. Olsen, S., Bartoli, A.: Implicit non-rigid structure-from-motion with priors. Journal of Mathematical Imaging and Vision **31** (2008) 233–244

10. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Pattern Analysis and Machine Intelligence (PAMI)* **30** (2008) 878–892
11. Paladini, M., Del Bue, A., Stosic, M., Dodig, M., Xavier, J., Agapito, L.: Factorization for non-rigid and articulated structure using metric projections. In: *IEEE Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA (2009) 2898–2905
12. Xiao, J., Kanade, T.: Uncalibrated perspective reconstruction of deformable structures. In: *Proceedings of the 10th International Conference on Computer Vision (ICCV)*. Volume 2. (2005) 1075–1082
13. Hartley, R., Vidal, R.: Perspective nonrigid shape and motion recovery. In: *Proceedings of the 10th European Conference on Computer Vision (ECCV)*. (2008) 276–289
14. Brand, M.: Morphable 3D models from video. In: *IEEE Computer Vision and Pattern Recognition (CVPR)*. (2001) 456–463
15. Heyden, A., Berthilsson, R., Sparr, G.: An iterative factorization method for projective structure and motion from image sequences. *International Journal on Computer Vision* **17** (1999) 981–991
16. Mahamud, S., Hebert, M.: Iterative projective reconstruction from multiple views. In: *The Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Volume 2. (2000) 430 – 437
17. Akhter, I., Sheikh, Y., Khan, S.: In defense of orthonormality constraints for non-rigid structure from motion. In: *IEEE Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA (2009) 1534–1541



**Fig. 8.** Six images of a 25-image sequence with 375 trajectories showing a box deforming non-rigidly. The top paper deforms non-rigidly, so a multi-body model would not be satisfied. Shown from left to right are image number, image, top view, left side view, frontal view and right side view of the reconstructed 3D-shape corresponding to each image.