

# A Unifying Model for Camera Calibration

Srikumar Ramalingam, Peter Sturm

► **To cite this version:**

Srikumar Ramalingam, Peter Sturm. A Unifying Model for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, Institute of Electrical and Electronics Engineers, 2017, 39 (7), pp.1309-1319. <10.1109/TPAMI.2016.2592904>. <hal-01396809>

**HAL Id: hal-01396809**

**<https://hal.inria.fr/hal-01396809>**

Submitted on 15 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Unifying Model for Camera Calibration

Srikumar Ramalingam and Peter Sturm, *Members, IEEE*

**Abstract**—This paper proposes a unified theory for calibrating a wide variety of camera models such as pinhole, fisheye, cata-dioptic, and multi-camera networks. We model any camera as a set of image pixels and their associated camera rays in space. Every pixel measures the light traveling along a (half-) ray in 3-space, associated with that pixel. By this definition, calibration simply refers to the computation of the mapping between pixels and the associated 3D rays. Such a mapping can be computed using images of calibration grids, which are objects with known 3D geometry, taken from unknown positions. This general camera model allows to represent non-central cameras; we also consider two special subclasses, namely central and axial cameras. In a central camera, all rays intersect in a single point, whereas the rays are completely arbitrary in a non-central one. Axial cameras are an intermediate case: the camera rays intersect a single line. In this work, we show the theory for calibrating central, axial and non-central models using calibration grids, which can be either three-dimensional or planar.

**Index Terms**—camera calibration, generic imaging model, non-central, cata-dioptic, omni-directional

## 1 INTRODUCTION

Several applications in medicine, surveillance, virtual reality, autonomous navigation and robotics involve cameras which go beyond the pinhole model: stereo, fisheye cameras, cata-dioptic systems (combinations of lenses and mirrors), multi-camera setups and other non-central cameras [1], [2]. We propose and develop a calibration approach that would work for any type of camera model, and especially for cameras without a single viewpoint. To do so, we first renounce on parametric models, and adopt the following general model: a camera acquires images consisting of pixels; each pixel captures light that travels along a ray in 3D. The camera is fully described by:

- the coordinates of these rays (given in some local coordinate system).
- the mapping between pixels and rays; this is basically a simple indexing.

The non-parametric nature of this model adds one difficulty: how to compute 3D rays for an image point with non-integer image coordinates? To do so, the only possibility is to add continuity assumptions, e.g. that neighboring pixels have neighboring 3D rays. Under this or more restrictive assumptions, 3D rays for arbitrary image points can be computed by interpolation. Similarly the projection of 3D points onto images, is not straightforward, but can for example be solved analogously, by interpolation. The general camera model allows to describe any camera that captures light rays traveling along straight lines. A few examples are:

- a camera with any type of geometrical distortion, such as radial or tangential.
- a camera looking at a reflective surface, e.g. as often used in surveillance, a camera looking at a spherical

or otherwise curved mirror [3]. Such systems, as opposed to central catadioptric systems [4] composed of cameras and parabolic mirrors, do not in general have a single viewpoint.

- multi-camera stereo systems: put together the pixels of all image planes – they “catch” light rays that definitely do not travel along lines that all pass through a single point. Nevertheless, in the above general camera model, a stereo system (with rigidly linked cameras) may be considered as a **single** camera.
- other acquisition systems, see e.g. [5], [6], [7], or eyes of some insects.

### 1.1 Relation to previous work

The generic imaging model was used in more or less explicit form, in various works [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], and is best described in [9], [19]. There are conceptual links to other works: acquiring an image with a camera of our general model may be seen as sampling the plenoptic function [20], and a light field [21] or lumigraph [22] may be interpreted as a single image, acquired by a camera of an appropriate design.

In the last decade, there has been a lot of progress in using non-parametric camera models for calibration. It has been shown that generic imaging models may give better calibration results than parametric approaches when the distortion is very high [23], [24]. Miraldo and Araujo showed interesting approaches for generic cameras using radial basis functions [25], [26]. Absolute pose and relative motion algorithms have been developed for general cameras [12], [27], [28], [29], [30], [31], [32], [33]. A few approaches have been proposed for the self-calibration of general camera models [34], [35], [36], [37], [38], which do not use known calibration grids. Several robust calibration toolboxes exist for omni-directional cameras based on lenses and mirrors [39], [40]. An elaborate survey of the camera models used in the geometrical vision community can be found in the review by Sturm et al. [41].

In this article, we focus on the generalized theory for calibrating cameras using calibration grids. In particular,

---

• S. Ramalingam is with Mitsubishi Electric Research Labs (MERL), USA.  
E-mail: srikumar.ramalingam@gmail.com

• P. Sturm is with INRIA Grenoble Rhône-Alpes, Univ. Grenoble Alpes, LJK, France.  
E-mail: peter.sturm@inria.fr

we provide the theory for computing the camera rays for a generalized family of cameras based on 2D to 3D point correspondences obtained using calibration grids.

## 1.2 Overview

In prior work we showed the theory for one or two cases to illustrate the main idea [19], [42], [43]. This work is a more complete version where we provide a complete overview and the theory for all the important cases in the generic imaging model. For the purpose of illustration we first introduce the underlying idea for 2D cameras in section 2.1 and 2.2. Then we focus on the more practical scenario involving 3D cameras in section 2.3. By 3D cameras we do not mean cameras providing three-dimensional images, but cameras evolving in 3D space and acquiring 2D images. The calibration algorithms are developed for two kinds of calibration grids: 3D and planar.

We use a simple idea to calibrate our generic imaging model. Three points that correspond to a single pixel in the image, are sampled on different calibration grids whose relative position is not assumed to be known. Since every pixel samples light ray passing through a single straight line in space, the grid points have to be collinear. We formulate a collinearity constraint on the grid points to first compute the relative position of the calibration grids. Once this is done, we complete the calibration by computing the camera rays passing through the sampled grid points. This idea is used to calibrate different camera models. Although the basic principle is the same for different cameras, the specific equation systems and solution methods it gives rise to, depends on the geometry of the camera rays. We consider three classes of general cameras (depending on the geometry of the camera rays): fully non-central (the general case), central, and axial. In a central model, all camera rays pass through a single point, the optical center. Examples are pinhole cameras with and without radial distortion, fisheye lenses and central catadioptric systems using parabolic, hyperbolic and elliptical mirrors. A non-central camera may have completely arbitrary camera rays. The so-called *axial cameras* [43] are non-central but their rays are constrained by the existence of a line that cuts all of them. First we provide the theory for various cases of the general, non-central model, followed by the central case and finally the axial one.

## 2 GENERIC CALIBRATION OF NON-CENTRAL CAMERAS

### 2.1 Calibration of 2D cameras

We consider here a camera and scene living in a 2D plane, i.e. camera rays are lines in that plane. Input are two images of an object undergoing some motion. Consider here a single pixel and its camera ray, as illustrated in figure 1(a). Figures 1 (b) and (c) show the two points on the object that are seen by that pixel in the two images. We suppose to be able to determine the coordinates of these two points, in some local coordinate frame attached to the object ("matching").

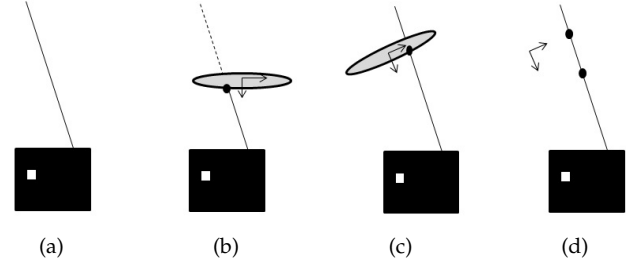


Fig. 1. (a) The camera as black box, with one pixel and its camera ray. (b) The pixel sees a point on a calibration object, whose coordinates are identified in a frame associated with the object. (c) Same as (b), for another position of the calibration object. (d) If the object's motion is known, the two points on the calibration object can be placed in the same coordinate frame (here the same one as in (c)). The camera ray is then determined by joining them.

#### 2.1.1 The case of known motion

If the object's motion between the image acquisitions is known, then the two object points can be mapped to a single coordinate frame, e.g. the object's coordinate frame for the second image, as shown in figure 1 (d). Computing our pixel's camera ray is then simply done by joining the two points. This summarizes the calibration approach proposed by Grossberg and Nayar [9], applied here for the 2D case. Camera rays are thus initially expressed in a coordinate frame attached to one of the calibration object's positions. This does not matter (all that counts are the relative positions of the camera rays), but for convenience, one would typically try to choose a better frame. For a central camera for example, one would choose the optical center as origin or for a non-central camera, the point that minimizes the sum of distances to the set of camera rays (if it exists).

#### 2.1.2 The case of unknown motion

The above approach is no longer applicable, and we need to estimate, implicitly or explicitly, the unknown motion. We now show how to do this, given three images instead of only two as above. Let us note the three points on the calibration objects, that are seen in the same pixel, by  $\mathbf{Q}$ ,  $\mathbf{Q}'$  and  $\mathbf{Q}''$ . These are 3-vectors of homogeneous coordinates, expressed in the respective local coordinate frame. Without loss of generality, we choose the coordinate frame associated with the object's first position, as common frame. The unknown (relative) motions, that allow to map the second and third frames onto the first one, are given by  $2 \times 2$  rotation matrices  $\mathbf{R}'$  and  $\mathbf{R}''$  and translation vectors  $\mathbf{t}'$  and  $\mathbf{t}''$ . Note that for the rotation matrices we have  $R'_{11} = R'_{22}$  and  $R'_{12} = -R'_{21}$  (and similarly for  $\mathbf{R}''$ ). The calibration points, after mapping them to the common frame, are given as:

$$\mathbf{Q} \quad \begin{pmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{Q}' \quad \begin{pmatrix} \mathbf{R}'' & \mathbf{t}'' \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{Q}''$$

They must all lie on the pixel's camera ray, hence they must be collinear. Algebraically, this collinearity constraint is expressed by the fact that the determinant of the  $3 \times 3$  matrix composed of the above three coordinate vectors,

i	$C_i$	$V_i$
1	$Q_1 Q'_1 Q''_3 + Q_2 Q'_2 Q''_3$	$R'_{21}$
2	$Q_1 Q'_2 Q''_3 - Q_2 Q'_1 Q''_3$	$R'_{11}$
3	$Q_1 Q'_3 Q''_1 + Q_2 Q'_3 Q''_2$	$-R'_{21}$
4	$Q_1 Q'_3 Q''_2 - Q_2 Q'_3 Q''_1$	$-R'_{11}$
5	$Q_3 Q'_1 Q''_1 + Q_3 Q'_2 Q''_2$	$R'_{11} R'_{21} - R'_{21} R'_{11}$
6	$Q_3 Q'_1 Q''_2 - Q_3 Q'_2 Q''_1$	$R'_{11} R'_{11} + R'_{21} R'_{21}$
7	$Q_1 Q'_2 Q''_3$	$t'_2 - t'_1$
8	$Q_2 Q'_1 Q''_3$	$-t'_1 + t'_1$
9	$Q_3 Q'_1 Q''_3$	$R'_{11} t'_{21} - R'_{21} t'_{11}$
10	$Q_3 Q'_2 Q''_3$	$-R'_{21} t'_{21} - R'_{11} t'_{11}$
11	$Q_3 Q'_3 Q''_1$	$R'_{21} t'_{11} - R'_{11} t'_{21}$
12	$Q_3 Q'_3 Q''_2$	$R'_{21} t'_{11} + R'_{11} t'_{21}$
13	$Q_3 Q'_3 Q''_3$	$t'_{11} t'_{21} - t'_{11} t'_{21}$

TABLE 1

Coupled variables in the trifocal calibration tensor for the general 2D camera.

vanishes:

$$\begin{vmatrix} Q_1 & R'_{11} Q'_1 - R'_{21} Q'_2 + t'_{11} Q'_3 & R'_{11} Q''_1 - R'_{21} Q''_2 + t'_{11} Q''_3 \\ Q_2 & R'_{21} Q'_1 + R'_{11} Q'_2 + t'_{21} Q'_3 & R'_{21} Q''_1 + R'_{11} Q''_2 + t'_{21} Q''_3 \\ Q_3 & Q'_3 & Q''_3 \end{vmatrix} = 0 \quad (1)$$

This equation is of the form  $\sum_{i=1}^3 \sum_{i'=1}^3 \sum_{i''=1}^3 V_{i,i',i''} Q_i Q'_{i'} Q''_{i''} = 0$ , i.e. it is trilinear in the calibration point coordinates. The equation's coefficients  $V$  may be interpreted as elements of a trilinear "calibration tensor" (somewhat related to the *homography tensor* derived in [44]); they depend on the unknown motions' coefficients. Among the  $3 \cdot 3 \cdot 3 = 27$  coefficients of the calibration tensor, 8 are always zero and among the remaining 19 ones, there are 6 pairs of identical ones. There are thus in general 13 different coefficients to consider; these are shown in table 1. The table and the entries  $C_i$  and  $V_i$  it contains, are interpreted as follows:

$$\sum_{i=1}^{13} C_i V_i = 0 \quad (2)$$

is merely a reformulation of the collinearity constraint (1). It is linear in the coefficients  $V_i$ . Given sufficiently many triplets of points  $\mathbf{Q}$ ,  $\mathbf{Q}'$  and  $\mathbf{Q}''$ , we may expect to compute the trilinear tensor up to an unknown scale  $\lambda$  using linear least squares. Note that we have verified using simulated data, that in general we indeed can obtain a unique solution (up to scale) for the tensor. The minimum number of triplets is 12 (estimation of 13 coefficients, up to scale).

The main question now is, if the motion coefficients  $R'$ ,  $R''$ ,  $t'$  and  $t''$  can be extracted from the  $V_i$ . This is indeed possible, as shown below. Once the motions are determined, the above approach for known motions, can be readily applied to compute camera rays.

We now describe an algorithm for extracting the motion parameters. Let the estimated tensor coefficients be  $V'_i$ ; they are equal to the coefficients of table 1 up to an unknown scale:  $V'_i = \lambda V_i$ ,  $i = 1 \dots 13$ . The proposed algorithm works as follows:

- Estimate  $\lambda$  :  $\lambda = \pm \sqrt{(V'_1)^2 + (V'_2)^2}$  (exploiting orthonormality of  $R'$ ).  $\lambda$  is defined up to sign.
- Compute  $V_i = \frac{V'_i}{\lambda}$ ,  $i = 1 \dots 13$ .
- Compute  $R'$ :  $R'_{11} = R'_{22} = V_2$  and  $R'_{21} = -R'_{12} = V_1$

i	$C_i$	$V_i$
1	$Q_1 Q'_1 Q''_3$	$R'_{21}$
2	$Q_1 Q'_2 Q''_3$	$-R'_{21}$
3	$Q_1 Q'_3 Q''_3$	$t'_2 - t'_1$
4	$Q_3 Q'_1 Q''_1$	$R'_{11} R'_{21} - R'_{21} R'_{11}$
5	$Q_3 Q'_1 Q''_2$	$R'_{11} t'_{21} - R'_{21} t'_{11}$
6	$Q_3 Q'_2 Q''_1$	$R'_{21} t'_{11} - R'_{11} t'_{21}$
7	$Q_3 Q'_3 Q''_3$	$t'_{11} t'_{21} - t'_{11} t'_{21}$

TABLE 2

Coefficients of the trifocal calibration tensor for the general 2D camera and a linear calibration object.

- Compute  $R''$ :  $R''_{11} = R''_{22} = -V_4$  and  $R''_{21} = -R''_{12} = -V_3$
- Compute  $t$  and  $t'$ : From table 1 we have:

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & -R'_{21} & R'_{11} \\ 0 & 0 & -R'_{11} & -R'_{21} \\ -R'_{21} & -R'_{11} & 0 & 0 \\ R'_{11} & R'_{21} & 0 & 0 \end{pmatrix}}_{\mathbf{M}} \begin{pmatrix} t'_{11} \\ t'_{21} \\ t'_{11} \\ t'_{21} \end{pmatrix} = \begin{pmatrix} V_7 \\ V_8 \\ V_9 \\ V_{10} \\ V_{11} \\ V_{12} \end{pmatrix}$$

The least-squares solution for  $t$  and  $t'$  is always well-defined:  $\mathbf{M}^T \mathbf{M}$  is always a non-singular matrix due to the orthonormality of  $R'$  and  $R''$ . Finally, we resolve the sign ambiguity in  $\lambda$  by recomputing all tensor coefficients from the motion parameters and verifying if they are equal to all computed tensor coefficients.

## 2.2 Calibration of 2D cameras using linear calibration objects

It is equally worthwhile to specialize our concept to the case of a linear calibration object. We now consider again the general, non-central camera model. Without loss of generality, we suppose that the calibration points lie on a line expressed as  $Y = 0$  in the local coordinate frame of the object, i.e.  $Q_2 = Q'_2 = Q''_2 = 0$ . The collinearity equation (1) gives then rise to a  $2 \times 2 \times 2$  trifocal calibration tensor. Among its 8 coefficients, only 1 is always zero, and among the others, none are identical to one another, cf. table 2.

We observe that the rotation coefficients  $R'_{11}$  and  $R''_{11}$  do not appear individually, contrary to the tensor for the general case (cf. table 1). Hence, the scale factor  $\lambda$  can no longer be determined as easily as in the above algorithm. Due to lack of space, we omit the details of this algorithm.

We have proven theoretically (not shown due to lack of space, but the proof can be sent by the authors upon request) that there is in general a unique solution for the tensor (in the absence of noise) unless two or all three of the calibration object's poses correspond to collinear lines.

## 2.3 3D cameras

We now extend the calibration concept from 2D to 3D cameras, i.e. cameras evolving in 3D space. This is first done for the case of non-central cameras undergoing **unknown motion** while capturing images of a 3D calibration object. Input are now, for each pixel, three 3D points  $\mathbf{Q}$ ,  $\mathbf{Q}'$  and  $\mathbf{Q}''$ ,

given by 4-vectors of homogeneous coordinates, relative to the calibration object's local coordinate system. Again, we adopt the coordinate system associated with the first image as global coordinate frame. The motion for the other two images is represented by  $3 \times 3$  rotation matrices  $R'$  and  $R''$  and translation vectors  $t'$  and  $t''$ , all unknown. We stack the 3D points, after transforming them to the global coordinate system, in the following  $4 \times 3$  matrix:

$$\begin{pmatrix} Q_1 & \sum_{i=1}^3 R'_{1i} Q'_i + t'_1 Q'_4 & \sum_{i=1}^3 R''_{1i} Q''_i + t''_1 Q''_4 \\ Q_2 & \sum_{i=1}^3 R'_{2i} Q'_i + t'_2 Q'_4 & \sum_{i=1}^3 R''_{2i} Q''_i + t''_2 Q''_4 \\ Q_3 & \sum_{i=1}^3 R'_{3i} Q'_i + t'_3 Q'_4 & \sum_{i=1}^3 R''_{3i} Q''_i + t''_3 Q''_4 \\ Q_4 & Q'_4 & Q''_4 \end{pmatrix} \quad (3)$$

The collinearity constraint means that this matrix must be of rank less than 3, which implies that all sub-determinants of size  $3 \times 3$  vanish. There are 4 of them, obtained by leaving out one row at a time from the original matrix. Each of these corresponds to a trilinear equation in point coordinates and thus to a trifocal calibration tensor whose coefficients depend on the motion parameters.

$i$	$C_i$	$V_i$	$W_i$
1	$Q_1 Q'_1 Q''_4$	0	$R'_{31}$
2	$Q_1 Q'_2 Q''_4$	0	$R'_{32}$
3	$Q_1 Q'_3 Q''_4$	0	$R'_{33}$
4	$Q_1 Q'_4 Q''_1$	0	$-R''_{31}$
5	$Q_1 Q'_4 Q''_2$	0	$-R''_{32}$
6	$Q_1 Q'_4 Q''_3$	0	$-R''_{33}$
7	$Q_1 Q'_4 Q''_4$	0	$t'_3 - t''_3$
8	$Q_2 Q'_1 Q''_4$	$R'_{31}$	0
9	$Q_2 Q'_2 Q''_4$	$R'_{32}$	0
10	$Q_2 Q'_3 Q''_4$	$R'_{33}$	0
11	$Q_2 Q'_4 Q''_1$	$-R''_{31}$	0
12	$Q_2 Q'_4 Q''_2$	$-R''_{32}$	0
13	$Q_2 Q'_4 Q''_3$	$-R''_{33}$	0
14	$Q_2 Q'_4 Q''_4$	$t'_3 - t''_3$	0
15	$Q_3 Q'_1 Q''_4$	$-R'_{21}$	$-R''_{11}$
16	$Q_3 Q'_2 Q''_4$	$-R'_{22}$	$-R''_{12}$
17	$Q_3 Q'_3 Q''_4$	$-R'_{23}$	$-R''_{13}$
18	$Q_3 Q'_4 Q''_1$	$R'_{21}$	$R''_{11}$
19	$Q_3 Q'_4 Q''_2$	$R'_{22}$	$R''_{12}$
20	$Q_3 Q'_4 Q''_3$	$R'_{23}$	$R''_{13}$
21	$Q_3 Q'_4 Q''_4$	$t''_2 - t'_2$	$t''_1 - t'_1$
22	$Q_4 Q'_1 Q''_1$	$R'_{21} R''_{31} - R'_{31} R''_{21}$	$R'_{11} R''_{31} - R'_{31} R''_{11}$
23	$Q_4 Q'_1 Q''_2$	$R'_{21} R''_{32} - R'_{32} R''_{21}$	$R'_{11} R''_{32} - R'_{32} R''_{11}$
24	$Q_4 Q'_1 Q''_3$	$R'_{21} R''_{33} - R'_{33} R''_{21}$	$R'_{11} R''_{33} - R'_{33} R''_{11}$
25	$Q_4 Q'_1 Q''_4$	$R'_{21} t''_3 - R'_{31} t'_2$	$R'_{11} t''_3 - R'_{31} t'_1$
26	$Q_4 Q'_2 Q''_1$	$R'_{22} R''_{31} - R'_{31} R''_{22}$	$R'_{12} R''_{31} - R'_{31} R''_{12}$
27	$Q_4 Q'_2 Q''_2$	$R'_{22} R''_{32} - R'_{32} R''_{22}$	$R'_{12} R''_{32} - R'_{32} R''_{12}$
28	$Q_4 Q'_2 Q''_3$	$R'_{22} R''_{33} - R'_{33} R''_{22}$	$R'_{12} R''_{33} - R'_{33} R''_{12}$
29	$Q_4 Q'_2 Q''_4$	$R'_{22} t''_3 - R'_{32} t'_2$	$R'_{12} t''_3 - R'_{32} t'_1$
30	$Q_4 Q'_3 Q''_1$	$R'_{23} R''_{31} - R'_{31} R''_{23}$	$R'_{13} R''_{31} - R'_{31} R''_{13}$
31	$Q_4 Q'_3 Q''_2$	$R'_{23} R''_{32} - R'_{32} R''_{23}$	$R'_{13} R''_{32} - R'_{32} R''_{13}$
32	$Q_4 Q'_3 Q''_3$	$R'_{23} R''_{33} - R'_{33} R''_{23}$	$R'_{13} R''_{33} - R'_{33} R''_{13}$
33	$Q_4 Q'_3 Q''_4$	$R'_{23} t''_3 - R'_{33} t'_2$	$R'_{13} t''_3 - R'_{33} t'_1$
34	$Q_4 Q'_4 Q''_1$	$R''_{31} t'_2 - R'_{31} t'_1$	$R''_{31} t'_1 - R'_{31} t'_1$
35	$Q_4 Q'_4 Q''_2$	$R''_{32} t'_2 - R'_{32} t'_1$	$R''_{32} t'_1 - R'_{32} t'_1$
36	$Q_4 Q'_4 Q''_3$	$R''_{33} t'_2 - R'_{33} t'_1$	$R''_{33} t'_1 - R'_{33} t'_1$
37	$Q_4 Q'_4 Q''_4$	$t'_2 t''_3 - t'_3 t''_2$	$t'_1 t''_3 - t'_3 t''_1$

TABLE 3

Coupled variables in the trifocal calibration tensors for a general 3D camera. Coefficients not shown here are always zero.

Table 3 shows the coefficients of the first two tensors. The associated equations are  $\sum_i V_i C_i = 0$  and  $\sum_i W_i C_i = 0$ . In both, 34 of the 64 coefficients are always zero. One

may observe that the tensors share several coefficients, e.g.  $V_8 = W_1 = R'_{31}$ . The situation is similar for the other two tensors, which are not shown here since the first two are sufficient to compute the motion parameters and thus to perform calibration.

The tensors can be estimated by solving linear equation systems, and we verified using simulated random experiments that in general unique solutions (up to scale) are obtained, if 3D points for sufficiently many pixels (29 at least) are available. In the following, we give an algorithm for computing the motion parameters. Let  $V'_i = \lambda V_i$  and  $W'_i = \mu W_i, i = 1 \dots 37$  be the estimated (up to scale) tensors. The algorithm proceeds as follows.

- 1) Estimate scale factors based on the orthonormality of  $R'$ :  $\lambda = \sqrt{(V'_8)^2 + (V'_9)^2 + (V'_{10})^2}$ ,  $\mu = \sqrt{(W'_1)^2 + (W'_2)^2 + (W'_3)^2}$ . This defines  $\lambda$  and  $\mu$  each up to sign; this is solved in the following two steps.
- 2) Compute  $V_i = \frac{V'_i}{\lambda}$  and  $W_i = \frac{W'_i}{\mu}, i = 1 \dots 37$ . If  $(V_8 W_1 + V_9 W_2 + V_{10} W_3) < 0$  then set  $\mu = -\mu$  and multiply  $W$  by  $-1$ .
- 3) Compute  $R'$  and  $R''$ :

$$R' = \begin{pmatrix} -W_{15} & -W_{16} & -W_{17} \\ -V_{15} & -V_{16} & -V_{17} \\ V_8 & V_9 & V_{10} \end{pmatrix},$$

$$R'' = \begin{pmatrix} W_{18} & W_{19} & W_{20} \\ V_{18} & V_{19} & V_{20} \\ -V_{11} & -V_{12} & -V_{13} \end{pmatrix}$$

If  $\det(R') < 0$  then scale  $V$  and  $W$  by  $-1$  and update  $R'$  and  $R''$  accordingly. In the presence of noise these matrices will in general not be orthonormal. We "correct" this by computing the orthonormal matrices that are closest to the original matrices (in the sense of the Frobenius norm). To do so, let  $U\Sigma V^T$  be the SVD of one of the original matrices. The required orthonormal matrix is then given by  $UV^T$ .

- 4) Compute  $t'$  and  $t''$  by solving the following linear system using least squares:

$$A \begin{pmatrix} t'_1 \\ t'_2 \\ t'_3 \\ t''_1 \\ t''_2 \\ t''_3 \end{pmatrix} = \underbrace{\begin{pmatrix} V_{14} \\ V_{21} \\ W_{21} \\ V_{25} \\ W_{25} \\ V_{29} \\ W_{29} \\ V_{33} \\ W_{33} \\ V_{34} \\ W_{34} \\ V_{35} \\ W_{35} \\ V_{36} \\ W_{36} \end{pmatrix}}_b$$

where  $A$  is given by:

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -R'_{31} & R'_{21} \\ 0 & 0 & 0 & -R'_{31} & 0 & R'_{11} \\ 0 & 0 & 0 & 0 & -R'_{32} & R'_{22} \\ 0 & 0 & 0 & -R'_{32} & 0 & R'_{12} \\ 0 & 0 & 0 & 0 & -R'_{33} & R'_{23} \\ 0 & 0 & 0 & -R'_{33} & 0 & R'_{13} \\ 0 & R'_{31} & -R'_{21} & 0 & 0 & 0 \\ R'_{31} & 0 & -R'_{11} & 0 & 0 & 0 \\ 0 & R'_{32} & -R'_{22} & 0 & 0 & 0 \\ R'_{32} & 0 & -R'_{12} & 0 & 0 & 0 \\ 0 & R'_{33} & -R'_{23} & 0 & 0 & 0 \\ R'_{33} & 0 & -R'_{13} & 0 & 0 & 0 \end{pmatrix}$$

The least squares solution  $(A^T A)^{-1} A^T \mathbf{b}$  is well defined since it can be shown that, due to the orthonormality of the rotation matrices  $R'$  and  $R''$ , the product  $A^T A$  is always an invertible matrix. This implies that a unique solution for motion can always be estimated and that it is correct if the calibration tensors are correct.

## 2.4 Calibration of 3D cameras with a planar calibration object

If we feed the general calibration algorithm of the previous section with data stemming from a planar calibration object, then the linear systems for estimating the tensors are more rank-deficient than they ought to be, i.e. the tensors cannot be estimated uniquely. In order to remove this extra rank deficiency, we have to make the following change to incorporate the planarity constraint:  $Q_3 = Q'_3 = Q''_3 = 0$ . The  $4 \times 3$  matrix containing the calibration point triplet, cf. equation (3), then becomes:

$$\begin{pmatrix} Q_1 & \sum_{i=1}^2 R'_{1i} Q'_i + t'_1 Q'_4 & \sum_{i=1}^2 R''_{1i} Q''_i + t''_1 Q''_4 \\ Q_2 & \sum_{i=1}^2 R'_{2i} Q'_i + t'_2 Q'_4 & \sum_{i=1}^2 R''_{2i} Q''_i + t''_2 Q''_4 \\ 0 & \sum_{i=1}^2 R'_{3i} Q'_i + t'_3 Q'_4 & \sum_{i=1}^2 R''_{3i} Q''_i + t''_3 Q''_4 \\ Q_3 & Q'_4 & Q''_4 \end{pmatrix}$$

We obtain 4 matching constraints by setting the minors of degree 3 of this matrix equal to zero. Each such equation gives rise to a trifocal tensor, analogously to the previous section. Using simulations, we found that the minor corresponding to the removal of the third row, does not provide a unique solution for the associated tensor. The reason might be that, while removing the third row, we do not actually utilize the planarity constraint. In the following we use only the first two tensors for computing the motion parameters. Their defining equations are  $\sum_i V_i C_i = 0$  and  $\sum_i W_i C_i = 0$ . The coefficients  $C_i$  are the same as the ones for the non-planar case, see Table 3, however all entries of the table corresponding to  $C_i$ 's that contain one or more of the coefficients  $Q_3$ ,  $Q'_3$  or  $Q''_3$ , disappear in the case of planar calibration objects. Only 19 coefficients  $C_i$  remain.

The extraction of individual motion parameters is slightly more complicated compared to the earlier cases. First of all we do not have any direct constraint to compute the scale parameters. However we may compute  $\mathbf{V}$  and  $\mathbf{W}$

up to a common scale  $\lambda$ , since they share some common variables. Let  $\mathbf{V}' = \lambda \mathbf{V}$  and  $\mathbf{W}' = \lambda \mathbf{W}$  be the estimated coefficients. In the following, we explain how to extract the scale  $\lambda$  and the rotation parameters from  $\mathbf{V}'$  and  $\mathbf{W}'$ , then the translation parameters.

We have the following relations, see table 3:

$$R'_{3,1} = \frac{1}{\lambda} V'_8, R'_{3,2} = \frac{1}{\lambda} V'_9, R''_{3,1} = -\frac{1}{\lambda} V'_{11}, R''_{3,2} = -\frac{1}{\lambda} V'_{12} \quad (4)$$

$$M \begin{pmatrix} R'_{2,1} & R'_{1,1} \\ R'_{2,2} & R'_{1,2} \\ R''_{2,1} & R''_{1,1} \\ R''_{2,2} & R''_{1,2} \end{pmatrix} = \begin{pmatrix} V'_{22} & W'_{22} \\ V'_{23} & W'_{23} \\ V'_{26} & W'_{26} \\ V'_{27} & W'_{27} \end{pmatrix}$$

where  $M$  is given by:

$$M = \begin{pmatrix} -V'_{11} & 0 & -V'_8 & 0 \\ -V'_{12} & 0 & 0 & -V'_8 \\ 0 & -V'_{11} & -V'_9 & 0 \\ 0 & -V'_{12} & 0 & -V'_9 \end{pmatrix}$$

The rank of  $M$  is 3 and hence we obtain the solution for the rotation coefficients in the above equation in the following subspaces:

$$\begin{pmatrix} R'_{2,1} \\ R'_{2,2} \\ R''_{2,1} \\ R''_{2,2} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} + l_1 \begin{pmatrix} V'_8 \\ V'_9 \\ -V'_{11} \\ -V'_{12} \end{pmatrix} \quad (5)$$

$$\begin{pmatrix} R'_{1,1} \\ R'_{1,2} \\ R''_{1,1} \\ R''_{1,2} \end{pmatrix} = \begin{pmatrix} a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix} + l_2 \begin{pmatrix} V'_8 \\ V'_9 \\ -V'_{11} \\ -V'_{12} \end{pmatrix} \quad (6)$$

where the  $a_i$  depend on the  $V'$  and  $W'$  (closed-form expressions exist but are not given due to lack of space).

We estimate the values of  $l_1$  and  $l_2$  using orthonormality constraints on the rotation matrices  $R'$  and  $R''$ , in particular:  $R'_{1,1} R'_{1,2} + R'_{2,1} R'_{2,2} + R'_{3,1} R'_{3,2} = 0$  and  $R'^2_{1,1} + R'^2_{2,1} + R'^2_{3,1} = R'^2_{1,2} + R'^2_{2,2} + R'^2_{3,2} = 1$  and analogously for  $R''$ . When substituting the rotation variables in these constraints and using (4), and after simplifying the expressions, we get the following system:

$$V \begin{pmatrix} l_1 \\ l_2 \\ l_1^2 + l_2^2 + \frac{1}{\lambda^2} \end{pmatrix} = \begin{pmatrix} -a_5 a_6 - a_1 a_2 \\ -a_7 a_8 - a_3 a_4 \\ 1 - a_1^2 - a_2^2 \\ 1 - a_2^2 - a_6^2 \\ 1 - a_3^2 - a_7^2 \\ 1 - a_4^2 - a_8^2 \end{pmatrix}$$

where  $V$  is given by:

$$V = \begin{pmatrix} a_1 V'_9 + V'_8 a_2 & a_5 V'_9 + V'_8 a_6 & V'_8 V'_9 \\ -a_3 V'_{12} - V'_{11} a_4 & -a_7 V'_{12} - V'_{11} a_8 & V'_{11} V'_{12} \\ 2a_1 V'_8 & 2a_5 V'_8 & V'^2_8 \\ 2a_2 V'_9 & 2a_6 V'_9 & V'^2_9 \\ -2a_3 V'_{11} & -2a_7 V'_{11} & V'^2_{11} \\ -2a_4 V'_{12} & -2a_8 V'_{12} & V'^2_{12} \end{pmatrix}$$

By solving this system to least squares, we get the solutions for  $l_1, l_2$  and, up to sign, for  $\lambda$ . The sign of  $\lambda$  cannot be recovered; this gives rise to two possible solutions, as described below. Equations (4), (5) and (6), allow to compute

the first two columns of  $R'$  and  $R''$ . The third column of each rotation matrix is simply computed as the cross-product of the first two. Now, the translation parameters can be computed by linear least squares, based on equations associated with coefficients  $V_{21}, V_{25}, V_{29}, V_{34}, V_{35}, W_{21}, W_{25}, W_{29}, W_{34}$  and  $W_{35}$ , cf. table 3.

Overall, we obtain two solutions for the motion parameters (corresponding to positive and negative  $\lambda$ ) and ultimately, of the camera rays. These two solutions are symmetric to one another: roughly speaking, one is obtained by mirroring the poses of the second and third calibration objects, in the plane of the first one. They thus correspond to the same set of camera rays, expressed in two different coordinate systems. Note that this is analogous to what happens for the pose estimation of a perspective camera from an image of a planar reference object, which also gives rise to two mirror solutions.

### 3 GENERIC CALIBRATION OF A CENTRAL CAMERA

#### 3.1 2D central cameras

In this section we study *central cameras*, i.e. all camera rays pass through a single point in space called the *optical center*  $\mathbf{O}$  ( $O_1, O_2$ ). Since the point  $\mathbf{O}$  lies on all the rays we only need two calibration points to formulate collinearity constraints. In other words two calibration objects are sufficient to calibrate a central 2D camera. We use the constraint that with the correct estimates of the object's poses and the optical center, the latter is collinear with the two calibration points associated to any pixel:

$$\begin{vmatrix} O_1 & Q_1 & R'_{11}Q'_1 - R'_{21}Q'_2 + t'_1Q'_3 \\ O_2 & Q_2 & R'_{21}Q'_1 + R'_{11}Q'_2 + t'_2Q'_3 \\ 1 & Q_3 & Q'_3 \end{vmatrix} = 0$$

i	$C_i$	$V_i$
1	$Q_1Q'_1 + Q_2Q'_2$	$R'_{21}$
2	$Q_1Q'_2 - Q_2Q'_1$	$R'_{11}$
3	$Q_1Q'_3$	$t'_2 - O_2$
4	$Q_2Q'_3$	$O_1 - t'_1$
5	$Q_3Q'_1$	$-O_1R'_{21} + O_2R'_{11}$
6	$Q_3Q'_2$	$-O_1R'_{11} - O_2R'_{21}$
7	$Q_3Q'_3$	$-O_1t'_2 + O_2t'_1$

TABLE 4

Coupled variables in the bifocal matching tensor for a central 2D camera.

This can be written as  $\sum_{i=1} C_i V_i = 0$ , where  $C_i$  and  $V_i$  are given in Table 4. There are only 7 coupled variables, which we may compute up to scale from the extracted calibration points using linear least squares. Let  $\mathbf{V}' = \lambda \mathbf{V}$  be the estimated variables. The motion parameters can be extracted as follows:

- 1)  $\lambda = \pm \sqrt{(V'_1)^2 + (V'_2)^2}$ ;  $V_i = \frac{V'_i}{\lambda}$ . Do steps 2 and 3 for both signs of  $\lambda$ .
- 2) From table 4 we get:

$$R' = \begin{pmatrix} V_2 & -V_1 \\ V_1 & V_2 \end{pmatrix}$$

- 3) From table 4 we get:

$$\begin{pmatrix} 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & -V_1 & V_2 \\ 0 & 0 & -V_2 & -V_1 \end{pmatrix} \begin{pmatrix} t'_1 \\ t'_2 \\ O_1 \\ O_2 \end{pmatrix} = \begin{pmatrix} V_3 \\ V_4 \\ V_5 \\ V_6 \end{pmatrix}$$

We obtain the following closed-form solution, due to the relation between  $V_1, V_2$  and  $R'$  as per step 2.

$$\begin{pmatrix} t'_1 \\ t'_2 \\ O_1 \\ O_2 \end{pmatrix} = \begin{pmatrix} 0 & -1 & -V_1 & -V_2 \\ 1 & 0 & V_2 & -V_1 \\ 0 & 0 & -V_1 & -V_2 \\ 0 & 0 & V_2 & -V_1 \end{pmatrix} \begin{pmatrix} V_3 \\ V_4 \\ V_5 \\ V_6 \end{pmatrix}$$

This implies that the solution is always well-defined.

- 4) Select the sign of  $\lambda$  for which  $V_7 = -O_1t'_2 + O_2t'_1$ .

The projection rays are then computed for all pixels. In fact only their directions are computed (they are constrained to pass through the computed optical center).

#### 3.2 3D central cameras

As in the 2D case, we have a 3D point  $\mathbf{O}$  ( $O_1, O_2, O_3$ ), which lies on all rays. We construct a matrix with  $\mathbf{O}$ ,  $\mathbf{Q}$  and  $\mathbf{Q}'$  in the same coordinate frame as shown below.

$$\begin{pmatrix} O_1 & Q_1 & \sum_{i=1}^3 R'_{1i}Q'_i + t'_1Q'_4 \\ O_2 & Q_2 & \sum_{i=1}^3 R'_{2i}Q'_i + t'_2Q'_4 \\ O_3 & Q_3 & \sum_{i=1}^3 R'_{3i}Q'_i + t'_3Q'_4 \\ 1 & Q_4 & Q'_4 \end{pmatrix}$$

The collinearity of these three points gives four constraints by removing one of the 4 rows at a time from the above matrix and equating the determinant of the remaining submatrix to zero. Two of these constraints are shown in Table 5. Using simulations, we observed that the third constraint failed to give a unique solution for the coupled variables. To extract the motion variables we used only the first and the second constraints.

i	$C_i$	$V_i$	$W_i$
1	$Q_1Q'_1$	0	$R'_{31}$
2	$Q_1Q'_2$	0	$R'_{32}$
3	$Q_1Q'_3$	0	$R'_{33}$
4	$Q_1Q'_4$	0	$-O_3 + t'_3$
5	$Q_2Q'_1$	$R'_{31}$	0
6	$Q_2Q'_2$	$R'_{32}$	0
7	$Q_2Q'_3$	$R'_{33}$	0
8	$Q_2Q'_4$	$-O_3 + t'_3$	0
9	$Q_3Q'_1$	$-R'_{21}$	$-R'_{11}$
10	$Q_3Q'_2$	$-R'_{22}$	$-R'_{12}$
11	$Q_3Q'_3$	$-R'_{23}$	$-R'_{13}$
12	$Q_3Q'_4$	$O_2 - t'_2$	$O_1 - t'_1$
13	$Q_4Q'_1$	$-O_2R'_{31} + O_3R'_{21}$	$-O_1R'_{31} + O_3R'_{11}$
14	$Q_4Q'_2$	$-O_2R'_{32} + O_3R'_{22}$	$-O_1R'_{32} + O_3R'_{12}$
15	$Q_4Q'_3$	$-O_2R'_{33} + O_3R'_{23}$	$-O_1R'_{33} + O_3R'_{13}$
16	$Q_4Q'_4$	$-O_2t'_3 + O_3t'_2$	$-O_1t'_3 + O_3t'_1$

TABLE 5

Coupled variables in the bifocal matching tensors for a 3D single center camera.

Now we explain how to extract the individual motion parameters from the estimated variables  $\mathbf{V}'$  and  $\mathbf{W}'$ . First, we estimate and undo the scale ambiguity in their

estimation. We compute the scale for  $\mathbf{V}'$  up to sign as  $\lambda = \sqrt{(V'_5)^2 + (V'_6)^2 + (V'_7)^2}$  and set  $V_i = \frac{V'_i}{\lambda}$ . Next we compute the scale  $\mu$  such that  $\mu W'_1, \mu W'_2$  respectively  $\mu W'_3$  are equal to  $V_5, V_6$  respectively  $V_7$ , and compute  $\mathbf{W} = \mu \mathbf{W}'$ . Then, the rotation matrix is estimated:

$$\mathbf{R}' = \begin{pmatrix} -W_9 & -W_{10} & -W_{11} \\ -V_9 & -V_{10} & -V_{11} \\ V_5 & V_6 & V_7 \end{pmatrix},$$

We remove the ambiguity on the sign of the scale parameter  $\lambda$  by checking if  $\det(\mathbf{R}') = 1$  or  $\det(\mathbf{R}') = -1$ . In the latter case, we change the sign of both  $\mathbf{V}$  and  $\mathbf{W}$  and recompute  $\mathbf{R}'$ . Finally, the optical center and the translation of the second calibration object, are estimated:

$$\mathbf{O} = \begin{pmatrix} \frac{V_{13}R'_{32} - V_{14}R'_{31}}{R'_{21}R'_{32} - R'_{22}R'_{31}} \\ \frac{O_3R'_{21} - V_{13}}{R'_{31}} \\ \frac{O_3R'_{11} - W_{13}}{R'_{31}} \end{pmatrix}, \quad \mathbf{t}' = \begin{pmatrix} O_1 - W_{12} \\ O_2 - V_{12} \\ W_4 + O_3 \end{pmatrix}$$

### 3.3 Calibration of Central 3D cameras using planar calibration patterns

It is possible to obtain constraints with just two views, like in the 2D case (section 3.1). However the constraints are insufficient to estimate the motion variables. This is to be expected since even for the pinhole model, full calibration using a planar calibration grid requires three views at least [45], [46]. We thus have to consider three views at least. We build a  $4 \times 4$  matrix consisting of  $\mathbf{O}, \mathbf{Q}, \mathbf{Q}'$  and  $\mathbf{Q}''$ :

$$\begin{pmatrix} Q_1 & \sum_{i=1}^2 R'_{1i}Q'_i + t'_1Q'_4 & \sum_{i=1}^2 R''_{1i}Q''_i + t''_1Q''_4 \\ Q_2 & \sum_{i=1}^2 R'_{2i}Q'_i + t'_2Q'_4 & \sum_{i=1}^2 R''_{2i}Q''_i + t''_2Q''_4 \\ Q_3 & \sum_{i=1}^2 R'_{3i}Q'_i + t'_3Q'_4 & \sum_{i=1}^2 R''_{3i}Q''_i + t''_3Q''_4 \\ Q_4 & Q'_4 & Q''_4 \end{pmatrix} \quad (7)$$

The collinearity of these four points implies that the  $4 \times 4$  matrix has to be of rank less than three. As a result all its 16 minors of degree 3 vanish. Please refer to Section 5.3 of [47] for the details on computing the motion parameters.

## 4 AXIAL CAMERAS

The axial model is a rather useful one. Many misaligned cata-dioptric configurations fall under this model (cf. Table 6.4 in [47]). Such configurations, which are slightly non-central, are usually classified as a non-central camera and calibrated using an iterative nonlinear algorithm [1], [2], [42]. For example, whenever the mirror is a surface of revolution and the central camera looking at the mirror lies anywhere on the revolution axis, the system is of axial type. Furthermore, two-camera stereo systems or systems consisting of three or more aligned cameras, are axial, as well as typical underwater cameras [48], [49]. Pushbroom cameras [50] are another example, although they are of a more restricted class (there exist two camera axes [51]).

We will call **camera axis** the line cutting all rays of an axial camera. It will be represented by a 6-vector  $\mathbf{L}$  and the associated  $4 \times 4$  skew-symmetric Plücker matrix  $[\mathbf{L}]_\times$ . The product  $[\mathbf{L}]_\times \mathbf{Q}$  gives the plane spanned by the line  $\mathbf{L}$  and the point  $\mathbf{Q}$ . Consider further the two 3-vectors  $\mathbf{A} = (L_5 \ L_6 \ L_4)^T$  and  $\mathbf{B} = (L_2 \ L_3 \ L_1)^T$ . The

Plücker constraint holds:  $\mathbf{B}^T \mathbf{A} = 0$ .  $\mathbf{A}$  represents the point at infinity of the line. The Plücker matrix can be written as:

$$[\mathbf{L}]_\times = \begin{pmatrix} 0 & -L_4 & L_6 & -L_2 \\ L_4 & 0 & -L_5 & -L_3 \\ -L_6 & L_5 & 0 & -L_1 \\ L_2 & L_3 & L_1 & 0 \end{pmatrix} = \begin{pmatrix} [\mathbf{A}]_\times & -\mathbf{B} \\ \mathbf{B}^T & 0 \end{pmatrix}$$

Camera rays will be constrained to cut the camera axis. Calibration thus consists in computing the position of the camera axis and of the camera rays. The proposed approach proceeds by first estimating the camera axis and the pose of all calibration objects.

## 5 CALIBRATION OF AXIAL CAMERAS WITH A 3D OBJECT

### 5.1 What can be done with only two views of 3D calibration objects?

Consider some pixel and let  $\mathbf{Q}$  and  $\mathbf{Q}'$  be the corresponding points on the two calibration objects, given as 3D points in the objects' local coordinate systems.

We have the following constraint on the pose of the second object ( $\mathbf{R}', \mathbf{t}'$ ) as well as the unknown camera axis  $\mathbf{L}$ : the line spanned by  $\mathbf{Q}$  and  $\mathbf{Q}'$  cuts  $\mathbf{L}$ , hence is coplanar with it. Hence, for the correct pose and camera axis, we must have:

$$\mathbf{Q}^T [\mathbf{L}]_\times \begin{pmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{Q}' = 0$$

Hence:

$$\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{pmatrix}^T \begin{pmatrix} 0 & -L_4 & L_6 & -L_2 \\ L_4 & 0 & -L_5 & -L_3 \\ -L_6 & L_5 & 0 & -L_1 \\ L_2 & L_3 & L_1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} Q'_1 \\ Q'_2 \\ Q'_3 \\ Q'_4 \end{pmatrix} = 0$$

We thus have the following  $4 \times 4$  binocular matching tensor that can be estimated linearly from calibration point correspondences:

$$\mathbf{F} \sim \begin{pmatrix} 0 & -L_4 & L_6 & -L_2 \\ L_4 & 0 & -L_5 & -L_3 \\ -L_6 & L_5 & 0 & -L_1 \\ L_2 & L_3 & L_1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R}' & \mathbf{t}' \\ \mathbf{0}^T & 1 \end{pmatrix} \quad (8)$$

One can extract, from  $\mathbf{F}$ , the complete camera axis  $\mathbf{L}$ , in the coordinate frame of each of the two calibration objects. For simplicity we apply, to each object, a rotation and translation such that the camera axis becomes the  $Z$ -axis, i.e.,

$$[\mathbf{L}]_\times = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

These rotations and translations are not unique – the pose of the second object is determined up to a rotation about the camera axis (the  $Z$ -axis) and translation along it:

$$\mathbf{R}' = \begin{pmatrix} c & -s & 0 \\ s & c & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{t}' = \begin{pmatrix} 0 \\ 0 \\ Z \end{pmatrix}$$

with  $c = \cos(\theta)$  and  $s = \sin(\theta)$  for some angle  $\theta$ .



F is now of the form

$$F = [L]_X \begin{pmatrix} R' & t' \\ 0^T & 1 \end{pmatrix} = \begin{pmatrix} -s & -c & 0 & 0 \\ c & -s & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We may compute the angle  $\theta$  by  $\theta = \arctan \frac{F_{11}}{F_{12}}$ ; this gives two solutions, separated by a  $180^\circ$  rotation about the camera axis. However, as for the translation along the camera axis, there is no possibility to compute it.

Hence, a complete calibration is not possible from only two views, even with three-dimensional calibration objects. Only one degree of freedom remains to be estimated (translation  $t'_3$  of the second object along the camera axis) and there remains a two-fold ambiguity on the rotational pose.

## 5.2 Full calibration using three views of 3D calibration objects

We use the technique described in the previous section on the image pairs (1, 2) and (1, 3). Our goal is to compute the remaining unknowns  $t'_3$  and  $t''_3$ , and fix the 2-fold ambiguities for  $R'$  and  $R''$ . Since  $Q$ ,  $Q'$  and  $Q''$  are collinear we form the following matrix and exploit the fact that its minors of degree 3 must vanish:

$$M = \begin{pmatrix} Q_1 & \pm Q'_1 & \pm Q''_1 \\ Q_2 & \pm Q'_2 & \pm Q''_2 \\ Q_3 & \pm Q'_3 + t'_3 Q'_3 & \pm Q''_3 + t''_3 Q''_3 \\ Q_4 & \pm Q'_4 & \pm Q''_4 \end{pmatrix}$$

- Concerning the 2-fold sign ambiguities: We consider the following minor, after removing the third row of  $M$ .

$$\det \begin{pmatrix} Q_1 & \pm Q'_1 & \pm Q''_1 \\ Q_2 & \pm Q'_2 & \pm Q''_2 \\ Q_4 & \pm Q'_4 & \pm Q''_4 \end{pmatrix} = 0$$

There are four variables ( $Q'_1, Q'_2, Q''_1, Q''_2$ ) with sign ambiguities. We take several point triplets ( $Q_i, Q'_i, Q''_i$ ) and compute the value of the above determinant. We try all 16 combinations (two possibilities for every variable) and only one combination will give zero determinant values for all the general triplets.

- Computing  $t'_3$  and  $t''_3$ . By removing one of the rows 1, 2 or 4 from matrix  $M$  and equating the determinant of the remaining matrix to zero, we get three constraints of the following form.

$$\det \begin{pmatrix} Q_i & Q'_i & Q''_i \\ Q_j & Q'_j & Q''_j \\ Q_3 & Q'_3 + t'_3 Q'_3 & Q''_3 + t''_3 Q''_3 \end{pmatrix} = 0$$

They can be rewritten as:

$$\begin{pmatrix} Q'_4(Q_j Q''_i - Q_i Q''_j) \\ Q''_4(Q_i Q'_j - Q_j Q'_i) \\ Q_i Q'_j Q''_3 + Q_3 Q'_i Q''_j + Q_j Q'_3 Q''_i \end{pmatrix}^T \begin{pmatrix} t'_3 \\ t''_3 \\ 1 \end{pmatrix} = 0$$

We stack all such equations for all calibration point triplets, and solve for the two translation coefficients using linear least squares.

## 6 CALIBRATION OF AXIAL CAMERAS WITH PLANAR OBJECTS

### 6.1 What can be done with two views of planar calibration objects?

Consider some pixel and let  $Q$  and  $Q'$  be the corresponding points on the two calibration grids, given as 3D points in the objects' local coordinate systems. Since we consider planar objects, we impose  $Q_3 = Q'_3 = 0$ .

We have the following constraint on the pose of the second object ( $R', t'$ ) as well as the unknown camera axis  $L$ : the line spanned by  $Q$  and  $Q'$  cuts  $L$ , hence is coplanar with it. Hence, for the correct pose and camera axis, we must have:

$$Q^T [L]_X \begin{pmatrix} R' & t' \\ 0^T & 1 \end{pmatrix} Q' = 0$$

Hence:

$$\begin{pmatrix} Q_1 \\ Q_2 \\ Q_4 \end{pmatrix}^T \begin{pmatrix} 0 & -L_4 & L_6 & -L_2 \\ L_4 & 0 & -L_5 & -L_3 \\ L_2 & L_3 & L_1 & 0 \end{pmatrix} \begin{pmatrix} \bar{R}' & t' \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} Q'_1 \\ Q'_2 \\ Q'_4 \end{pmatrix} = 0$$

where  $\bar{R}'$  refers to the  $3 \times 2$  sub-matrix of  $R'$  containing only the first and the second columns. We thus have the following  $3 \times 3$  tensor that can be estimated linearly from point correspondences:

$$F \sim \begin{pmatrix} 0 & -L_4 & L_6 & -L_2 \\ L_4 & 0 & -L_5 & -L_3 \\ L_2 & L_3 & L_1 & 0 \end{pmatrix} \begin{pmatrix} \bar{R}' & t' \\ 0^T & 1 \end{pmatrix} \quad (9)$$

$F$  is in general of rank 2. One could give the following interpretation of it:  $F$  can be interpreted as the fundamental matrix of a pair of perspective images, as follows. Imagine the two planar calibration objects as image planes of two perspective cameras, and the camera axis of the axial camera as the baseline containing the optical centers of these two perspective cameras. The epipolar geometry of this stereo system is given by  $F$ . It has only 7 degrees of freedom (9 - 1 for scale, -1 for rank-deficiency) so the 10 unknowns (4 for the camera axis, 3 for  $R'$  and 3 for  $t'$ ) cannot be recovered from it.

We now look at what can actually be recovered from  $F$ . Let us first notice that its left null-vector is  $(L_3, -L_2, L_4)^T$  (it truly is the null-vector, as can be easily verified when taking into account the Plücker constraint given above). We thus can recover 2 of the 4 parameters of the camera axis. That null-vector contains actually the coordinates of the camera axis' intersection with the first planar calibration object (in plane coordinates). Its 3D coordinates are given by  $(L_3, -L_2, 0, L_4)^T$ . Similarly, the right null-vector of  $F$  gives the plane coordinates of the axis' intersection with the second object. Besides this  $F$  also gives constraints on  $R'$  and  $t'$ . For example  $R'$  can be extracted up to 2 to 4 solutions, see [47].

To summarize, two views of planar calibration objects are not sufficient to uniquely calibrated an axial camera. Again, this is not surprising since even in the more constrained case of central cameras, two planar objects are not sufficient for calibration. The details of using three calibration objects for a full calibration, are given in Section 6.3.2 of [47].

## 7 EXPERIMENTS

**Central model:** We devised a method for taking into account multiple calibration grids at the same time, to obtain a good initial calibration for a sub-region of the image. During this phase, we obtain the poses, that is the orientation and position of these grids w.r.t. a common coordinate system. In Figure 2 we show multiple calibration grids used to compute the calibration for the omni-directional image. Details on how to extract the individual motion parameters from collinearity constraints on more than 3 images can be found in [47] (cf. Chapter 7). For example, in Figure 2(middle) we show the partial calibration using 5 planar grids. We can compute the position and orientation of a new grid by using this partial calibration and treating this problem as a generic pose estimation problem [27], [43]. We incrementally add new images of grids and increase the region of calibration as shown in Figure 2(right), which shows an almost completely calibrated image using 23 images. In Figure 3 we show an accurate distortion correction result for a fisheye image, obtained as a straightforward result of our calibration approach.

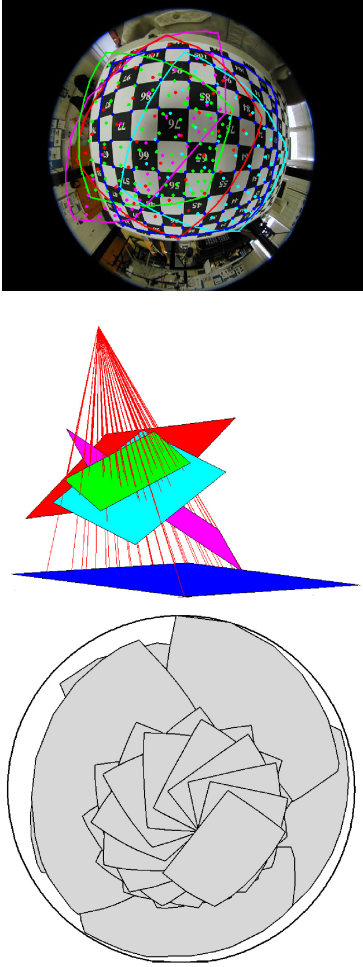


Fig. 2. Left: An omnidirectional image taken with a fisheye and the region of calibration grids occupied in 4 other images (shown using convex hulls of grid points). Middle: We show the 5 calibrated grid positions, which are used to compute the camera rays. Right: Example of a complete calibration for a fisheye camera from 23 overlapping grids.

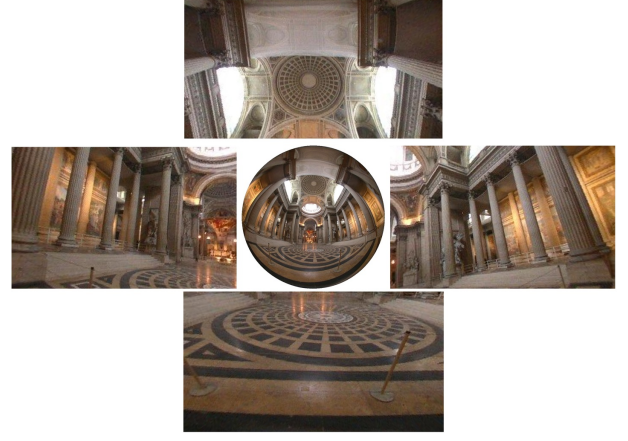


Fig. 3. Complete distortion correction of a fisheye image shown in the middle. Note that heavily distorted lines are corrected in the undistorted images. Only very close to the image border, residual distortions remain, which seems acceptable for such a generic calibration approach.

**Stereo camera:** We tested our axial calibration algorithm on a stereo system and 3 pairs of images of a calibration grid. Once we compute the poses of the grids we can compute rays corresponding to individual cameras in the stereo system. The reconstructed rays and axis are shown in Figure 4(a). For optimization we use a ray-point bundle adjustment minimizing the 3D distance between 3D points and camera rays [43]. The final RMS error for the stereo system is 0.04 with respect to the overall size of the scene. After clustering the two centers of the stereo camera, the RMS error is 0.07.

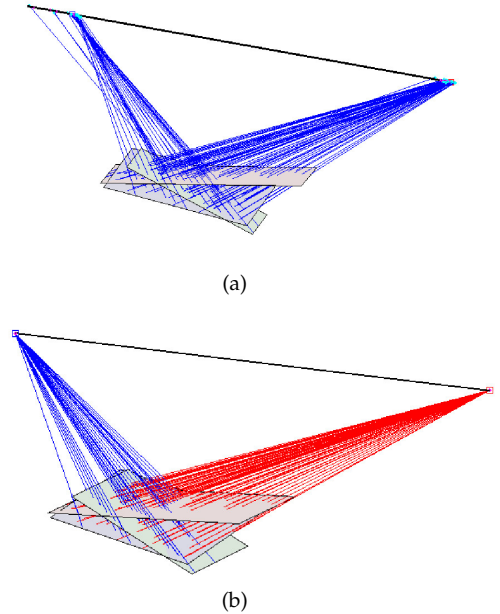


Fig. 4. Calibration of a stereo system using the axial algorithm with 3 images and a total of 481 rays. (a) Computed camera rays and axis. (b) Camera rays of the stereo system after clustering them and enforcing rays in each cluster to cut a single 3D point.

## 8 CONCLUSION

In this paper, we explain the theory and basic algorithms for calibrating a highly generic camera model consisting of just a collection of pixels and associated camera rays, from images of calibration objects with unknown poses. We have developed the theory and algorithms for three subclasses: fully non-central, axial and central cameras. Algorithms for usage of three-dimensional as well as planar calibration objects, are provided. Table 6 summarizes the nature of the results when applying the various calibration algorithms for different camera models. Additional experiments for non-central cameras are provided in [42], [47].

Algorithm	Actual type of camera and calibration object				
	C-2D	C-3D	A-2D	NC-2D	NC-3D
C-2D	Unique	NS	NS	NS	NS
C-3D	7/12	Unique	NS	6/12	NS
A-2D	2/10	4/10	Unique	NS	NS
NC-2D	5/14	8/14	2/14	Unique	NS
NC-3D	21/30	19/30	18/30	17/30	Unique

TABLE 6

Nature of solutions on applying the calibration algorithms on cameras of specific types. The labels 'C', 'NC', and 'A' denote central, non-central and axial respectively and the keywords '3D' and '2D' refer to the type of calibration object. 'NS' means that no solution exists: for instance, when applying the 'C-2D' algorithm (designed for a central camera and images of a planar object) to images taken actually with an axial or non-central camera, there is no solution that is consistent with the images. In the inverse case, applying an algorithm for a camera model that is more general than the actual camera, the solution is underconstrained: 'r/n' designates a rank deficiency of  $r$  when the linear system underlying the calibration has  $n$  rows. The rank deficiency can also be useful to detect whether a given camera is central, non-central or axial.

## ACKNOWLEDGMENTS

We thank Tomaš Pajdla, Branislav Mičušík, Bertrand Holveck, Suresh Lodha and Thomas Bonfort for useful discussions.

## REFERENCES

- [1] D. Aliaga, "Accurate catadioptric calibration for real-time pose estimation in room-size environments," in *ICCV*, 2001, pp. 127–134.
- [2] B. Mičušík and T. Pajdla, "Autocalibration and 3D reconstruction with non-central catadioptric cameras," in *CVPR*, 2004, pp. 748–753.
- [3] R. Hicks and R. Bajcsy, "Catadioptric sensors that approximate wide-angle perspective projections," in *CVPR*, 2000, pp. 545–551.
- [4] S. Baker and S. Nayar, "A theory of catadioptric image formation," in *ICCV*, 1998.
- [5] H. Bakstein and T. Pajdla, "An overview of non-central cameras," in *CVWW*, 2001.
- [6] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic stereo imaging," *PAMI*, pp. 279–290, 2001.
- [7] H. Shum, A. Kalai, and S. Seitz, "Omnivergent stereo," in *ICCV*, 1999, pp. 22–29.
- [8] C. Geyer and K. Daniilidis, "A unifying theory of central panoramic systems and practical implications," in *ECCV*, 2000.
- [9] M. Grossberg and S. Nayar, "A general imaging model and a method for finding its parameters," in *ICCV*, vol. 2, 2001, pp. 108–115.
- [10] J. Neumann, C. Fermüller, and Y. Aloimonos, "Polydioptric camera design and 3d motion estimation," in *CVPR*, vol. 2, 2003, pp. 294–301.
- [11] T. Pajdla, "Stereo with oblique cameras," in *IJCV*, vol. 47(1), 2002.
- [12] R. Pless, "Using many cameras as one," in *CVPR*, 2003, pp. 587–594.
- [13] S. Seitz, "The space of all stereo images," in *ICCV*, vol. 1, 2001, pp. 26–33.
- [14] R. Swaminathan, M. Grossberg, and S. Nayar, "A perspective on distortions," in *CVPR*, 2003.
- [15] Y. Wexler, A. Fitzgibbon, and A. Zisserman, "Learning epipolar geometry from image sequences," in *CVPR*, vol. 2, 2003, pp. 209–216.
- [16] D. Wood, A. Finkelstein, J. Hughes, C. Thayer, and D. Salesin, "Multiperspective panoramas for cell animation," in *SIGGRAPH*, 1997, pp. 243–250.
- [17] J. Ponce, "What is a camera?" in *CVPR*, 2009.
- [18] J. Yu and L. McMillan, "General linear cameras," in *ECCV*, 2004.
- [19] P. Sturm and S. Ramalingam, "A generic concept for camera calibration," in *ECCV*, vol. 2, 2004, pp. 1–13.
- [20] E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," *Computational Models of Visual Processing*, 1991.
- [21] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH*, 1996, pp. 31–42.
- [22] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The Lumigraph," in *SIGGRAPH*, 1996.
- [23] A. Dunne, J. Mallon, and P. Whelan, "A comparison of new generic camera calibration with the standard parametric approach," *MVA*, 2007.
- [24] —, "Efficient generic calibration method for general cameras with single centre of projection," in *ICCV*, 2007.
- [25] P. Miraldo, H. Araujo, and J. Queiro, "Point-based calibration using a parametric representation of the general imaging model," in *ICCV*, 2011.
- [26] P. Miraldo and H. Araujo, "Calibration of smooth camera models," *PAMI*, 2013.
- [27] D. Nistér, "A minimal solution to the generalized 3-point pose problem," in *CVPR*, 2004.
- [28] S. Ramalingam, S. Lodha, and P. Sturm, "A generic cross-camera structure-from-motion analysis," in *OMNIVIS*, 2004.
- [29] P. Sturm, "Multi-view geometry for general camera models," in *CVPR*, 2005.
- [30] H. Li, R. Hartley, and J. Kim, "A linear approach to motion estimation using generalized camera models," in *CVPR*, 2008.
- [31] G. Lee, F. Fraundorfer, and M. Pollefeys, "Motion estimation for self-driving cars with a generalized camera," in *CVPR*, 2013.
- [32] G. Lee, B. Li, M. Pollefeys, and F. Fraundorfer, "Minimal solutions for pose estimation of a multi-camera system," in *ICRA*, 2013.
- [33] P. Miraldo and H. Araujo, "A simple and robust solution to the minimal general pose estimation," in *ICRA*, 2014.
- [34] S. Thirithala and M. Pollefeys, "1D radial cameras and its application to omnidirectional geometry," in *ICCV*, 2005.
- [35] D. Nistér, H. Stewenius, and E. Grossmann, "Non-parametric self-calibration," in *ICCV*, 2005.
- [36] J. Tardif, P. Sturm, M. Trudeau, and S. Roy, "Calibration of cameras with radially symmetric distortion," *PAMI*, vol. 31, no. 9, 2008.
- [37] E. Grossmann, J. Gaspar, and F. Orabona, "Calibration from statistical properties of the visual world," in *ECCV*, 2008.
- [38] S. Thirithala and M. Pollefeys, "Radial multi-focal tensors - applications to omnidirectional camera calibration," *IJCV*, 2012.
- [39] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *IROS*, 2006.
- [40] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *ICRA*, 2007.
- [41] P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, and J. Barreto, "Camera models and fundamental concepts used in geometric computer vision," *Foundations and Trends in Computer Graphics and Vision*, 2011.
- [42] S. Ramalingam, P. Sturm, and S. Lodha, "Towards complete generic camera calibration," in *CVPR*, 2005.
- [43] S. Ramalingam, S. Lodha, and P. Sturm, "A generic structure-from-motion framework," *CVIU*, 2006.
- [44] A. Shashua and L. Wolf, "Homography tensors: On algebraic entities that represent three views of static or moving planar points," in *ECCV*, 2000.
- [45] P. Sturm and S. Maybank, "On plane-based camera calibration," in *CVPR*, 1999, pp. 432–437.
- [46] Z. Zhang, "A flexible new technique for camera calibration," *PAMI*, vol. 22, no. 11, pp. 1330–1334, 2000.

- [47] S. Ramalingam, "Generic imaging models: Calibration and 3d reconstruction algorithms," Ph.D. dissertation, PhD Thesis, Grenoble-INP, 2007.
- [48] A. Agrawal, Y. Taguchi, and S. Ramalingam, "Analytical forward projection for axial non-central dioptric and catadioptric cameras," in *ECCV*, 2010.
- [49] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari, "A theory of multi-layer flat refractive geometry," in *CVPR*, 2012.
- [50] R. Gupta and R. Hartley, "Linear pushbroom cameras," *PAMI*, 1997.
- [51] D. Feldman, T. Pajdla, and D. Weinshall, "On the epipolar geometry of the crossed-slits projection," in *ICCV*, 2003.



**Srikumar Ramalingam** is a senior principal research scientist at Mitsubishi Electric Research Lab (MERL) since 2008. He received a Marie Curie VisionTrain grant to pursue his doctoral studies in computer science and applied mathematics at INRIA Rhone Alpes (France). He received his PhD in 2007 under the guidance of Dr. Peter Sturm from INPG (National Polytechnical Institute of Grenoble, France). He received several awards such as INPG best thesis prize, AFRIF thesis honorable mention, R&D 100, and

RSS best paper finalist. His research interests are in computer vision, machine learning, robotics, and autonomous driving.



**Peter Sturm** obtained a PhD degree from Grenoble-INP in 1997. His PhD thesis was awarded the SPECIF award (given to one French PhD thesis in Computer Science per year). After a two-year post-doc at Reading University, he joined Inria on a permanent research position. Since 2006, he is Directeur de Recherche (the Inria equivalent of Professor). In 2009/10 he spent a sabbatical at CAMP, TU Munich. Since 2015, he is Deputy Scientific Director of Inria, in charge of the domain Perception,

Cognition, and Interaction.