Theory and Experiments towards Complete Generic Calibration

Srikumar Ramalingam, Peter Sturm, Suresh K. Lodha

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Theory and Experiments towards Complete Generic Calibration

Srikumar Ramalingam — Peter Sturm — Suresh K. Lodha

N° 5562

April 2005

Thème COG

ISSN 0249-6399 ISRN INRIA/RR--5562--FR+ENG

Theory and Experiments towards Complete Generic Calibration

Srikumar Ramalingam* , Peter Sturm† , Suresh K. Lodha‡

Thème COG — Systèmes cognitifs
Projet Movi

Rapport de recherche n° 5562 — April 2005 — 22 pages

Abstract: We consider the problem of calibrating a highly generic imaging model, that consists of a non-parametric association of a projection ray in 3D to every pixel in an image. Previous calibration approaches for this model do not seem to be directly applicable for cameras with large fields of view. In this paper, we describe a complete calibration approach that should in principle be able to handle any camera that can be described by the generic imaging model. Several contributions are provided. We show how to use the images of multiple calibration grids simultaneously, thus increasing the accuracy. This allows to calibrate an image region where these grids overlap. We then show how to calibrate the whole image, by iteratively estimating the pose of calibration grids and computing projection rays. A bundle adjustment procedure is also proposed, and several practical issues are discussed. The proposed approach has been applied on a wide variety of central and non-central cameras including fisheye lens, catadioptric cameras with spherical and hyperbolic mirrors, and multi-camera setups. We also consider the question if non-central models are more appropriate for certain cameras than central models.

Key-words: Calibration, camera model, distortion, stereo, omnidirectional, catadioptric

* srikumar@cse.ucsc.edu – S. Ramalingam is with the University of California at Santa Cruz.
† Peter.Sturm@inrialpes.fr – P. Sturm is with the MOVI project team.
‡ lodha@cse.ucsc.edu – S.K. Lodha is with the University of California at Santa Cruz.

Théorie et algorithmes : vers un calibrage générique complet

Résumé : Nous considérons le problème du calibrage d'un modèle de caméra extrêmement générique qui consiste d'une association non-paramétrique de rayons de projection à chaque pixel dans une image. Les approches de calibrage précédentes pour ce modèle ne semblent pas être directement applicables pour des caméras avec de grands champs de vue. Dans ce papier, nous décrivons une approche de calibrage complète qui en principe devrait être capable de gérer toute caméra qui peut décrite par le modèle générique. Plusieurs contributions sont fournies. D'abord nous montrons comment utiliser plusieurs images de mires de calibrage simultanément, pour obtenir une bonne précision. Cela permet de calibrer la région de l'image qui est couverte par toutes ces mires. Ensuite, nous montrons comment calibrer l'image entière, en alternant des calculs de pose pour des mires et des calculs de rayons de projection. Une procédure d'ajustement de faisceaux est également proposée et différents aspects d'ordre pratique sont discutés. L'approche proposée a été appliquée sur une large variété de caméras centrales et non-centrales : des objectifs de type "fish-eye", des caméras catadioptriques (avec miroir sphérique ou hyperbolique) et des systèmes multi-caméra. Nous considérons aussi la question quand un modèle non-central est plus appropriés pour une caméra qu'un modèle central.

Mots-clés : Calibrage, modèle de caméra, distorsion, stéréo, omnidirectionnel, catadioptrique

1 Introduction

This paper is about camera calibration. We adopt a general non-parametric imaging model that consists in associating one projection ray to each individual pixel. By projection ray we refer to the 3D (half-) line along which the light travels that falls onto the pixel (here, we neglect point spread and the finite spatial extent of a pixel). Rays may be completely unconstrained, i.e. they may not intersect in a single point, in which case the camera is called *non-central*. This general model has been used in various works [9, 15, 17, 18, 19, 21, 22, 25, 26, 28, 29], and is best described in [9], where other properties than geometric ones are also considered.

By adopting this model, one may formulate "black-box calibration" and provide algorithms that allow to calibrate any camera (see figure 1 for examples), be it of pinhole type (with or without optical distortions), catadioptric [2, 13], pushbroom [10], or some other acquisition system [3, 18, 22]. Such calibration algorithms have been proposed in [5, 8, 9, 25]. In this paper, we adopt the approach of [25] which allows to perform calibration from several images of a calibration grid, without having to know the motion between the images. There, calibration is solved by estimating some tensor that relates the different images: two images for central cameras and three for non-central ones. A drawback of that approach is that only these minimum numbers of views can be handled, i.e. the calibration approach can not benefit from multiple images. However, an accurate calibration of the general imaging

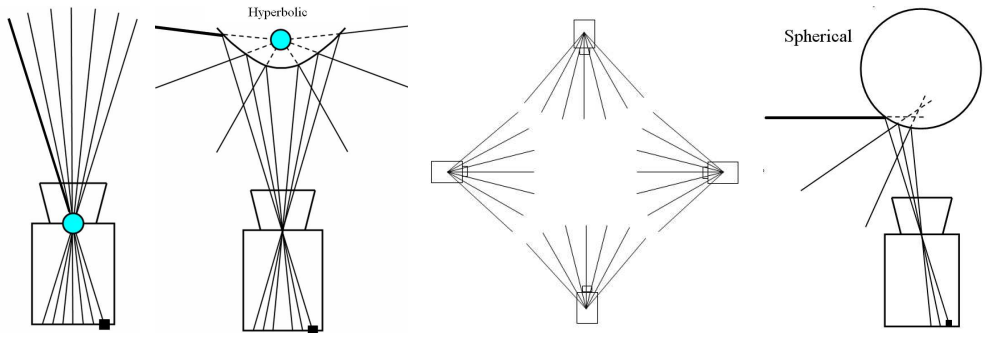


Figure 1: Examples of generic imaging model. Note that 3D rays do not necessarily pass through their associated image pixels. a) pinhole central camera, b) central catadioptric system with hyperbolic mirror, c) multi-camera system with four pinhole cameras (interpreted here as a single, non-central camera), d) non-central catadioptric system with spherical mirror.

model will likely require many more than the minimum number of images, due to the large number of unknowns (one projection ray per pixel).

In this paper, we propose a similar calibration method, but which uses multiple images of calibration grids simultaneously. This allows to calibrate the image region where the projections of the calibration grids overlap. To calibrate the complete image with this method, one would need a calibration grid of appropriate dimensions and shape; especially for omnidirectional cameras (fisheye, catadioptric, etc), this will be cumbersome to produce and handle. We thus also propose methods for concatenating the information contained in multiple images in case there is no global overlap (see examples in figure 2).

The overall proposed calibration approach works as follows. An initial calibration is performed from images of calibration grids that present a sufficient overlap. We then incorporate one image after the other, each time the image having the largest overlap with the already calibrated image region. We show how to compute the pose of the associated calibration grid. Then, given the pose, one may compute projection rays for previously uncalibrated pixels, thus enlarging the calibrated image region. This process is iterated until all images have been used. We also propose a bundle adjustment algorithm for the general imaging model; this can be used at any stage of the procedure, e.g. after computing the pose of a new calibration grid or of course at the end of our procedure.

This procedure and the underlying algorithms are developed for both, non-central and central models, although the central case is described in more detail in this paper. Besides developing algorithms, we are also interested in the question if for certain cameras it is worth going to a full non-central model or not, cf. also [1, 14].

This report complements [20]. It is organized as follows. The calibration approach is described in section 2 and some variants are proposed in section 3. Practical issues are discussed in section 4. Experimental results on a variety of cameras are presented in section 5, followed by conclusions in section 6.

2 Complete Calibration

We first provide an overview of complete generic camera calibration. We take several images of a calibration grid such as to cover the entire image region. Then, matching between image pixels and points on the calibration grids is performed. From such matches, we then compute the pose of each of these grids in a common coordinate system. After this pose computation, a 3D projection ray is computed for each pixel, as follows. For all grid points matching a given pixel, we compute their 3D coordinates (via the pose of the grids). The pixel's projection ray is then simply computed by fitting a straight line to the associated grid points.

For a non-central camera, two grid points per pixel are of course required. If the camera is (assumed to be) central however, a single grid point is enough: as will be seen later, the above stage of pose computation also comprises the estimation of the camera's optical center (in the same coordinate frame as the grids' pose). Thus, we compute projection rays by

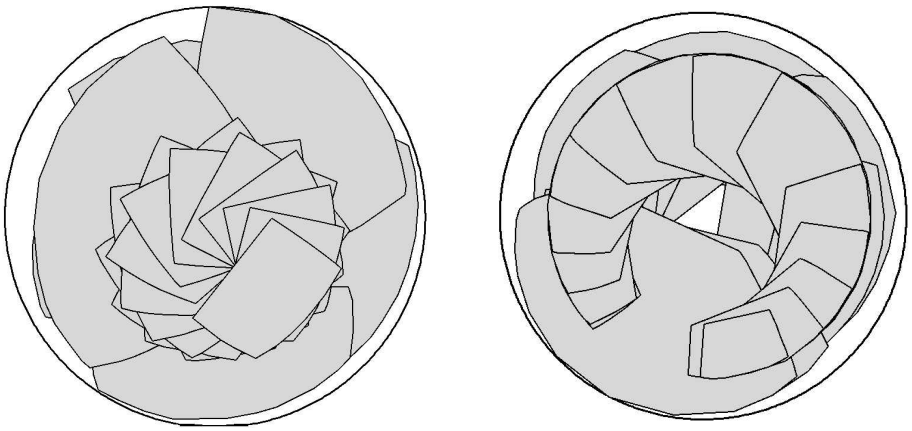


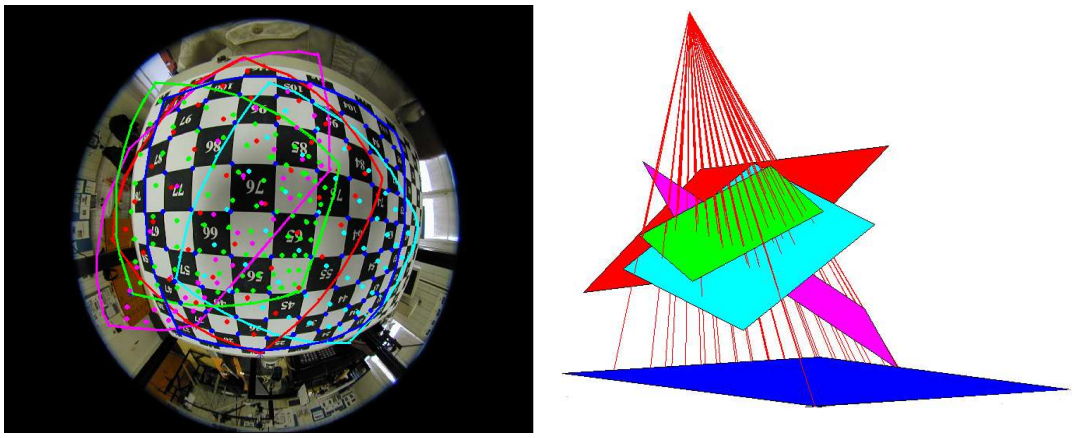
Figure 2: Examples of complete calibration. Left: 23 overlapping calibration grids, used in calibrating a fisheye. Right: 24 overlapping calibration grids used in calibrating a spherical catadioptric camera.

fitting lines to 3D points, but which are constrained to contain the optical center. Naturally, a single point per pixel is sufficient at this stage.

In the following, we describe different parts of our approach in more detail. In this section, we describe the case of central cameras. For conciseness, the non-central case is described more briefly in section 3.1. First, we show how to use the images of multiple grids simultaneously, to compute grid pose and the optical center. As mentioned above, this only allows to use matches for pixels that lie inside the projections of all grids considered. It is then shown how to compute the pose of additional grids; this serves to iteratively calibrate the whole image region, using all grids, even if their projections have no common overlap region. Refinement of calibration after each step, through bundle adjustment, is then discussed in section 2.3.

2.1 Calibration using Multiple Grids

As mentioned above, our motivation here is to devise 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. Let B_i denote the image region associated with i_{th} calibration grid, for $i = 1, \cdots, n$. We first calibrate the region $\cup_{i=2,\cdots,n}(B_1 \cap B_i)$, that is union of all pairs of intersecting grid regions formed with the first grid and all others that have a sufficient overlap with the first one. During this phase, we obtain the poses, that is the orientation and position of these grids w.r.t. a common coordinate system. Further below, we then show how to extend this calibration to the whole region $\cup_{i=0}^{n} B_i$, that is, the union of all grids.



(a) (b)

Figure 3: a) An omnidirectional image taken by fisheye lens and the region of calibration grids occupied in 4 other images (shown using convex hulls of grid points). b) We show the 5 calibrated grid positions, which are used to compute the 3D rays.

We now outline the theory behind calibration using multiple grids. Consider one pixel and its associated grid points, with homogeneous coordinates $Q^k = (Q_1^k, Q_2^k, Q_3^k, Q_4^k)^T$, for grids $k = 1, \ldots, n$. In the following, we consider planar calibration grids, and thus suppose that $Q_3^k = 0$. The use of non-planar calibration grids is actually simpler algebraically, but is less practical, we neglect it in this paper.

Let the unknown grid poses be represented by rotation matrices R^k and translation vectors t^k, such that the point Q^k, given in local grid coordinates, is mapped to global coordinates via

$$\begin{pmatrix} R^k & t^k \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} Q_1^k \\ Q_2^k \\ 0 \\ Q_4^k \end{pmatrix} \tag{1}$$

Furthermore, let $O = (O_1, O_2, O_3, 1)$ be the coordinates of the camera's optical center. As global coordinate system, we adopt, without loss of generality, the reference frame of the first grid, i.e. $R^1 = I$ and $t^1 = 0$.

We now show how to estimate the unknown grid poses and the optical center. This is based on the following collinearity constraint: with the correct poses, the grid points

associated with one pixel, after mapping into the global coordinate system via (1), must be collinear, and in addition, collinear with the optical center. This is because all these points must lie on the pixel's projection ray, i.e. a straight line. Algebraically, this collinearity constraint can be formulated as follows. Consider the matrix containing the coordinates of the collinear points:

$$\begin{pmatrix} O_1 & Q_1^1 & R_{11}^2 Q_1^2 + R_{12}^2 Q_2^2 + t_1^2 Q_4^2 & \cdots \\ O_2 & Q_2^1 & R_{21}^2 Q_1^2 + R_{22}^2 Q_2^2 + t_2^2 Q_4^2 & \cdots \\ O_3 & 0 & R_{31}^2 Q_1^2 + R_{32}^2 Q_2^2 + t_3^2 Q_4^2 & \cdots \\ 1 & Q_4^1 & Q_4^2 & \cdots \end{pmatrix}$$

The collinearity of these points implies that this $4 \times (n+1)$ matrix must be of rank smaller than 3. Consequently, the determinants of all its 3×3 submatrices must vanish.

The vanishing determinants of 3×3 submatrices give equations linking calibration point coordinates and the unknowns (camera poses and optical center). On using the first column (optical center) and two other columns with Q^j and Q^k to form a submatrix, we get bilinear equations in terms of calibration point coordinates Q^j and Q^k. Hence, we may write the equations in the form:

$$Q^{jT} M_{3\times 3} Q^k = 0 \tag{2}$$

This matrix M (a bifocal matching tensor), depends on camera pose and optical center, in a way specific to which 3×3 submatrix is underlying the equation. Using the above equation, we try to estimate such tensors from available correspondences. Since a 3×3 submatrix can be obtained by removing one row and $n-2$ columns at a time, we have $4 \times \binom{n+1}{3}$ possible matching tensors M. However, using simulations we observed that not all of these give unique solutions, i.e. not all these tensors can be estimated uniquely from point matches. Let $T_{ijk;i'j'k'}$ represent the tensor corresponding to the submatrix with rows (i,j,k) and columns (i',j',k'). In the following, we use $2 \times (n-1)$ constraints of the form $T_{x34;12y}, (x = 1,2), (y = 3, \cdots, n)$ for calibration, i.e. constraints combining the optical center and the first grid, with the other grids. For these tensors, the constraint equation (2) takes the following form: $\sum_{i=1}^{i=9} C_i^y V_i^y = 0$ and $\sum_{i=1}^{i=9} C_i^y W_i^y = 0$ for $T_{134;12y}$ and $T_{234;12y}$ respectively. Here, $C_i^y = Q_j^1 Q_k^y$, for appropriate indices j, as shown in Table 1.

V_i^y and W_i^y are computed up to scale using least squares. Note that they share some coefficients (e.g. $R_{3,1}^y$), hence they can be estimated up to the same scale factor, λ_y. We perform this step for $(n-1)$ constraints by choosing $y = 3, \cdots, n$. We now combine all the coupled variables contained in the different tensors, to obtain the following system which links the motion variables of all the grids.

$$\begin{bmatrix} H_{6\times 2}^2 & J_{6\times 6} & .. & 0_{6\times 6} \\ .. & .. & .. & .. \\ H_{6\times 2}^n & 0_{6\times 6} & .. & J_{6\times 6} \end{bmatrix} \begin{bmatrix} -O_1 \\ -O_2 \\ X_{6\times 1}^2 \\ .. \\ X_{6\times 1}^n \end{bmatrix} = \begin{bmatrix} Y_{6\times 1}^2 \\ .. \\ Y_{6\times 1}^n \end{bmatrix}, \tag{3}$$

i	j	k	V_i^y	W_i^y
1	1	1	0	$R_{3,1}^y$
2	1	2	0	$R_{3,2}^y$
3	1	4	0	$-O_3 + t_3^y$
4	2	1	$R_{3,1}^y$	0
5	2	2	$R_{3,2}^y$	0
6	2	4	$-O_3 + t_3^y$	0
7	4	1	$-O_2R_{3,1}^y + O_3R_{2,1}^y$	$-O_1R_{3,1}^y + O_3R_{1,1}^y$
8	4	2	$-O_2R_{3,2}^y + O_3R_{2,2}^y$	$-O_1R_{3,2}^y + O_3R_{1,2}^y$
9	4	4	$-O_2t_3^y + O_3t_2^y$	$-O_1t_3^y + O_3t_1^y$

Table 1: Coupled variables in tensors $T_{134;12y}$ and $T_{234;12y}$ for a central camera.

$$H^i = \begin{bmatrix} 0 & V_4^i \\ 0 & V_5^i \\ 0 & V_6^i \\ V_4^i & 0 \\ V_5^i & 0 \\ V_6^i & 0 \end{bmatrix}, J = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$X^i = \begin{bmatrix} \lambda_i O_3(t_1^i - O_1) \\ \lambda_i O_3(t_2^i - O_2) \\ \lambda_i O_3 R_{1,1}^i \\ \lambda_i O_3 R_{1,2}^i \\ \lambda_i O_3 R_{2,1}^i \\ \lambda_i O_3 R_{2,2}^i \end{bmatrix}, Y^i = \begin{bmatrix} V_7^i \\ V_8^i \\ V_9^i \\ W_7^i \\ W_8^i \\ W_9^i \end{bmatrix}$$

We rewrite Equation (3) as follows:

$$A_{6(n-1)\times(2+6(n-1))} Z_{2+6(n-1)} = Y_{6(n-1)}$$

Since A is of rank $6(n-1)$, we obtain the $(2+6(n-1))$ variables ($Z's$) in a subspace spanned by three vectors expressed with linear factors l_1 and l_2. We use orthogonality constraints on rotation matrices to obtain unique solutions for $Z's$. Specifically, we use three quadratic equations for each grid:

$$\begin{aligned} R_{11}^y R_{12}^y + R_{21}^y R_{22}^y + R_{31}^y R_{32}^y &= 0 \\ (R_{11}^y)^2 + (R_{21}^y)^2 + (R_{31}^y)^2 &= 0 \\ (R_{12}^y)^2 + (R_{22}^y)^2 + (R_{32}^y)^2 &= 0 \end{aligned}$$

On substituting the expressions for the rotational components and solving linear systems with l_1,l_2,O,λ_y and other coupled variables, we finally obtain the solutions for l_1 and l_2.

Using the formulas for $Z's$ given above, it is possible to compute the motion variables uniquely except for a sign ambiguity in n variables. This arises due to ambiguity in the position of each of these grids leading to 2^n possible solutions. As shown in Figure 2, each grid can lie on either side of the center. In the case of a pinhole camera we can resolve this ambiguity by applying the constraint that the grids must lie on one side of the center. However this constraint becomes difficult to apply for omnidirectional cameras where the grids essentially get distributed around the center. We thus apply the following procedure. First we arbitrarily choose the sign for the first grid's position vector. Then we identify the correct location of each of the other grids by minimizing its distance with an already fixed grid, with which it has some overlap. This is easily achieved because we always collect the images of calibration grids in succession and not in a completely random order. Having determined the pose of grids and the optical center, we now compute projection rays for all pixels that have at least one matching point in one of the grids used here. In Figure 3 we show the calibration using five grids simultaneously. Figures 3(a) and (b) show the convex hulls of the calibration grid points and the estimated projection rays respectively.

2.2 Pose Estimation of Additional Grids

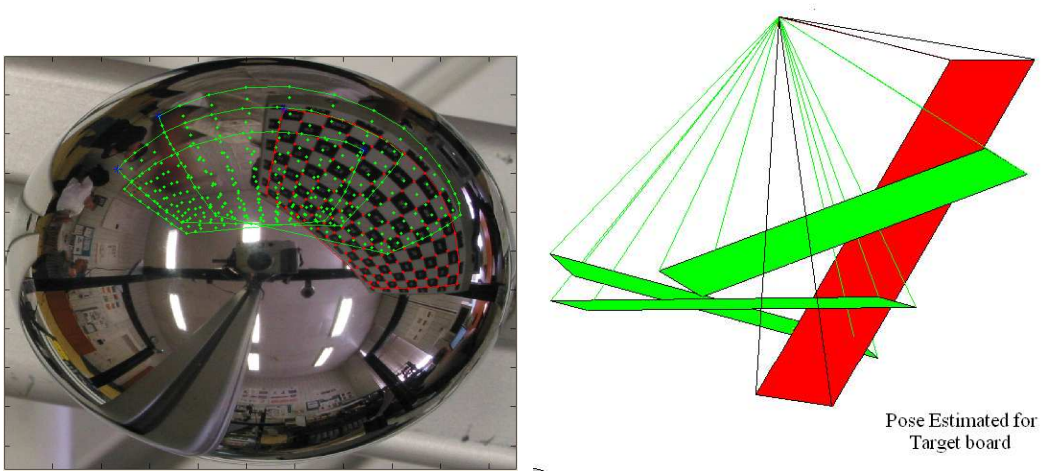
We suppose here that a partial calibration of the camera has been performed with the method of the previous section. As was mentioned, this in general only allows to calibrate a sub-region of the image (unless some first grid covers the whole image). Furthermore, in the previous step, only grids were used whose projection in the image had some overlap with one of the grids ("the first grid"). In order to make the calibration complete, we now show how to include additional grids, which do not have any overlap with the first grid, but with some of the others.

Let C_k denote the region obtained by simultaneous calibration with k grids. We now estimate the pose of the $k+1_{th}$ grid by taking into account the grid points falling within the region $C_k \cap B_{k+1}$. Using this estimated pose, the calibration region is then extended to $C_{k+1} = C_k \cup B_{k+1} = \cup_{i=0}^{n} B_i$. By estimating one grid at a time, we can iteratively calibrate the whole region occupied by the union of all grids. Figure 2 shows examples of the such calibrated region, for a fish-eye and a spherical catadioptric system.

We now explain the pose estimation procedure. Consider three calibrated rays passing through the grid whose pose is to be computed. Let us parameterize generic 3D points on these rays as $A_i + \lambda_i B_i$. Knowledge of the distances d_{ij} between the associated grid points will be used to compute depths λ_i, by solving:

$$|A_i + \lambda_i B_i - A_j - \lambda_j B_j|^2 = d_{ij}^2 \; for \; (i,j) = (l,m,n)$$

Now we have three quadratic equations with three unknowns. We use a polynomial solver to get the theoretical maximum of 8 solutions (they come in pairs of sign-reflected depths). The correct pose is identified by using one or two additional points as in [11]. As mentioned in the previous section, we get two valid solutions for the pose, with are related to one another by a reflection in the optical center. We resolve this ambiguity as discussed in



(a) (b)

Figure 4: Pose Estimation of a new grid using existing calibration. a) We show the new grid, whose pose has to be estimated, and the existing calibration region (using convex hulls of the grid points). b) We show the poses of 3 previously calibrated grids and the estimated pose of the new grid.

section 2.1. This pose estimation procedure is embedded in a RANSAC algorithm, to obtain a robust and stable pose estimation from the rays in the overlapping region. In Figure 4 we show the pose estimation of a new board using existing partial calibration, which was obtained using three calibration grids.

2.3 Bundle Adjustment

We use bundle adjustment [27] to refine the pose of all grids (except for the first one) and the projection rays. Bundle adjustment can be applied at any stage of our approach; we apply it after the initial calibration using multiple (cf. section 2.1), for refining the pose of each additional grid (cf. section 2.2), as well as at the end of the whole calibration. We have developed bundle adjustment algorithms for both central and noncentral scenarios as given below.

2.3.1 Central Scenario:

All the projection rays pass through the optical center. Thus any ray can be expressed using the optical center an a direction vector. The optical center is already computed using

calibration. In addition to the knowledge of the optical center we need to know one more point to compute the direction of a specific ray. When we have more than one grid points lying on the ray we use the centroid of these grid points to compute the direction. This is because a ray passing through a set of points also passes through their centroid. The overall error, E, to be minimized is given below.

$$E = \sum_{i=1}^{n} \sum_{j} (C + \lambda_{ij} D_i - [R_j T_j] P_{ij})^2$$

- C is the camera center(since it is a central camera),
- D_i is the direction of the i_{th} ray starting at C,
- λ_{ij} is a parameter to select specific point on the i_{th} ray,
- P_{ij} is the point on the j_{th} grid lying on the i_{th} ray,
- $[R_j, T_j]$ represents the motion between the representative grid (the first grid or the grid which covers the largest area in the image) and the j_{th} grid.

It is important to note that the grid points from all the grids may not lie on any given ray. Since D_i is a direction vector it has only two degrees of freedom. We use the orthonormal representation, introduced in [4].

$$D_i = U \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

where U, which is an 3×3 rotation matrix, can be updated as shown below.

$$U \leftarrow U R(\theta) \; with \; R(\theta) = R_x(\theta_1) R_y(\theta_2),$$

where $R_x(\theta_1)$ and $R_y(\theta_2)$ are orthonormal rotation matrices representing 3D rotations around x and y axis respectively. When we use the orthonormal representations algebraic constraints are automatically taken into account and the system becomes well-conditioned.

2.3.2 Non-central Scenario:

There is no single optical center for noncentral cameras. Thus we need a separate point and a direction for every projection ray. Two grid points are sufficient to compute to define a ray. In the case of more than two grid points we use least squares approach to compute the optimal ray. Let C and D represent a point on the ray and its direction respectively. Let us have n points Q_i on the ray. We have to minimize the following expression to compute D and λ_i is a parameter corresponding to the closest point on the ray to a given point Q_i.

$$\min_{\lambda_i D} \sum_{i}^{n} |C + \lambda_i D - Q_i|^2$$

Let us replace $Q_i - C$ by V_i. We apply the constraint $D^T D = 1$. We have $\lambda_i = V_i^T D$.

$$\min_{\lambda_i D} \sum_i^n |C + \lambda_i D - Q_i|^2 = \min_D \sum_i^n ((V_i^T D)D - V_i)^T (V_i^T D)D - V_i)$$

$$= \min_D \sum ((V_i^T D)^2 D^T D - 2(V_i^T D)^2 + V_i^T V_i) = \min_D \sum (V_i^T D)^2 + \sum V_i^T V_i$$

$$= \min_D \sum V_i^T V_i - D^T (\sum V_i V_i^T) D = \min_D \sum V_i^T V_i - D^T S D$$

The minimum has to be computed under the condition $D^T D = 1$. Let us look at the eigenvalues m_i of S.

$$\min_{|D|=1} (-D^T S D) = -D^T (S - m_i I) D$$

Thus the eigenvector corresponding to the largest eigenvalue, scaled to norm 1, will be the least squares solution for vector D. We use orthonormal representation for the rays with minimal parameters. We can represent a ray with a minimum of four parameters. We use two parameters to represent the direction D_i of the ray as in the previous case. Let C_i be the point closest to the origin on i_{th} ray. We represent C_i using two parameters from D and two additional parameters (s and θ_3) as shown below.

$$C_i = sV \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

where V, which is a 3×3 rotation matrix and it can be updated as given below.

$$V \leftarrow UR(\theta) \; with \; R(\theta) = R_x(\theta_1) R_y(\theta_2) R_z(\theta_3),$$

where $R_x(\theta_1)$, $R_y(\theta_2)$ and $R_z(\theta_3)$ are orthonormal rotation matrices representing 3D rotations around x and y axis respectively. Also $R_x(\theta_1)$ and $R_y(\theta_2)$ are the same rotation matrices used in the updation of D_i. Thus we have only $(\theta_1, \theta_2, \theta_3, s)$ as the only four parameters in the updation of a ray in noncentral camera. The cost function to be minimized is given below.

$$E = \sum_{i=1}^{n} \sum_j (sV \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \lambda_{ij} D_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - [R_j T_j] P_{ij})^2$$

3 Variants

3.1 Non-Central Cameras

In the non-central case, collinearity constraints require 3 or more grid points per pixel, instead of 2 for central cameras (where the optical center, though unknown, is taken into

account). We use the same notations as in section 2.1. For a non-central camera, we first calibrate the region $\cup_{\{i=3,\cdots,n\}}(B_1 \cap B_2 \cap B_i)$, i.e. the union of intersections of every possible triplet formed with the first two grids and the others. This is done, as above, using all these grids simultaneously. The calibration is then extended to the whole region $\cup_{\{i,j\}}(B_i \cap B_j)$, that is the union of all pairwise intersections of grid regions (this is because the computation of projection rays requires at least 2 grid points per pixel).

We now summarize the calibration procedure, analogously to section 2.1. We have no optical center here, so do consider the following $4 \times n$ matrix of collinear points:

$$\begin{pmatrix} Q_1^1 & R_{11}^2 Q_1^2 + R_{12}^2 Q_2^2 + t_1^2 Q_4^2 & \cdots \\ Q_2^1 & R_{21}^2 Q_1^2 + R_{22}^2 Q_2^2 + t_2^2 Q_4^2 & \cdots \\ 0 & R_{31}^2 Q_1^2 + R_{32}^2 Q_2^2 + t_3^2 Q_4^2 & \cdots \\ Q_4^1 & Q_4^2 & \cdots \end{pmatrix}$$

Similarly to the central case we can apply the collinearity constraint by equating the determinant of every 3×3 submatrix to zero. We take the first three columns and construct three tensors as given in Table 2.

In section 5.2 in [24] we have given an algorithm to calibrate a general 3D camera (noncentral) using two tensors. We slightly modify this algorithm by using three tensors instead of two sensors. Then we stack the linear system with 23 coupled variables as given in Table 2. We observed that by stacking the constraints from all the three tensors together, we directly obtain the coupled variables from both the tensors in the same scale. This modification improved the numerical stability of the algorithm. The estimation of the motion vectors from the coupled variables is already explained in [24]. We verified our algorithm using simulations and real experiments (multicamera setup in section 5).

Now let us look at the extension of this work to multiple views. Using simulations we found, as in the central case, that not all of these cases provide unique solutions. In contrast to the central case, where we used the center and the first grid to build a system linking all the motion variables, we here use the first and second grid to build the system. Thus we have $3 \times (n-2)$ possible tensors, represented by $T_{3jk;12y}, (\{j,k\} \in \{1,2,4\}), (y = 3, \cdots, n)$.

The next step of the calibration chain, pose estimation and computation of further projection rays, is also slightly different compared to central cameras (cf. section 2.2). Here, the calibration region is extended to $C_{k+1} = \cup_{\{i,j\}}(B_i \cap B_j)$, i.e. it contains all pixels that are matched to at least 2 grid points. As for the actual pose estimation, it can be formulated in the same way as for central cameras, but may lead to a set of 8 solutions that does not contain reflected pairs [6, 16]. Disambiguation can be carried out in the same manner as above though, using additional points besides the 3 used for the minimal pose routine.

3.2 Slightly Non-Central Cameras

In order to obtain a unique solution from our non-central calibration algorithm (cf. section 3.1), the camera must be sufficiently non-central. However, especially for omnidirectional cameras, the initial calibration step only allows to calibrate an image sub-region. In

	Coupled motion pars	T^1	T^2	T^4
1	R'_{31}	$Q_2Q'_1Q''_4$	$Q_1Q'_1Q''_4$	0
2	R'_{32}	$Q_2Q'_2Q''_4$	$Q_1Q'_2Q''_4$	0
3	R''_{31}	$-Q_2Q'_4Q''_1$	$-Q_1Q'_4Q''_1$	0
4	R''_{32}	$-Q_2Q'_4Q''_2$	$-Q_1Q'_4Q''_2$	0
5	$t'_3 - t''_3$	$Q_2Q'_4Q''_4$	$Q_1Q'_4Q''_4$	0
6	$R'_{11}R''_{31} - R'_{31}R''_{11}$	0	$Q_4Q'_1Q''_1$	$-Q_2Q'_1Q''_1$
7	$R'_{11}R''_{32} - R'_{31}R''_{12}$	0	$Q_4Q'_1Q''_2$	$-Q_2Q'_1Q''_2$
8	$R'_{12}R''_{31} - R'_{32}R''_{11}$	0	$Q_4Q'_2Q''_1$	$-Q_2Q'_2Q''_1$
9	$R'_{12}R''_{32} - R'_{32}R''_{12}$	0	$Q_4Q'_2Q''_2$	$-Q_2Q'_2Q''_2$
10	$R'_{21}R''_{31} - R'_{31}R''_{21}$	$Q_4Q'_1Q''_1$	0	$Q_1Q'_1Q''_1$
11	$R'_{21}R''_{32} - R'_{31}R''_{22}$	$Q_4Q'_1Q''_2$	0	$Q_1Q'_1Q''_2$
12	$R'_{22}R''_{31} - R'_{32}R''_{21}$	$Q_4Q'_2Q''_1$	0	$Q_1Q'_2Q''_1$
13	$R'_{22}R''_{32} - R'_{32}R''_{22}$	$Q_4Q'_2Q''_2$	0	$Q_1Q'_2Q''_2$
14	$R'_{11}t''_3 - R'_{31}t''_1$	0	$Q_4Q'_1Q''_4$	$-Q_2Q'_1Q''_4$
15	$R'_{12}t''_3 - R'_{32}t''_1$	0	$Q_4Q'_2Q''_4$	$-Q_2Q'_2Q''_4$
16	$R'_{21}t''_3 - R'_{31}t''_2$	$Q_4Q'_1Q''_4$	0	$Q_1Q'_1Q''_4$
17	$R'_{22}t''_3 - R'_{32}t''_2$	$Q_4Q'_2Q''_4$	0	$Q_1Q'_2Q''_4$
18	$R''_{11}t'_3 - R''_{31}t'_1$	0	$-Q_4Q'_4Q''_1$	$Q_2Q'_4Q''_1$
19	$R''_{12}t'_3 - R''_{32}t'_1$	0	$-Q_4Q'_4Q''_2$	$Q_2Q'_4Q''_2$
20	$R''_{21}t'_3 - R''_{31}t'_2$	$-Q_4Q'_4Q''_1$	0	$-Q_1Q'_4Q''_1$
21	$R''_{22}t'_3 - R''_{32}t'_2$	$-Q_4Q'_4Q''_2$	0	$-Q_1Q'_4Q''_2$
22	$t'_1t''_3 - t'_3t''_1$	0	$Q_4Q'_4Q''_4$	$-Q_2Q'_4Q''_4$
23	$t'_2t''_3 - t'_3t''_2$	$Q_4Q'_4Q''_4$	0	$Q_1Q'_4Q''_4$

Table 2: Coupled coefficients of tensors $T1$, $T2$ and $T4$.

the presence of noise in corner extraction and interpolation, the rays associated with pixels in that region, are usually not sufficiently non-central (with the exception of e.g. a multi-camera setup, which is highly non-central). For slightly non-central cameras like fisheye, spherical or hyperbolic catadioptric cameras, we thus start by running the central version of the initial calibration method. Typically we use four to five images simultaneously to calibrate an image region and then use pose estimation to add other images and cover the rest of the image. Then, we relax the central assumption; projection rays are first computed from grid points, without enforcing them to pass through an optical center. After this, a non-central bundle adjustment is performed.

3.3 Selecting the Best Camera Model

We observed that most cameras can be calibrated using our generic calibration algorithms (central or non-central), as shown in Table 3. Several issues have to be discussed though.

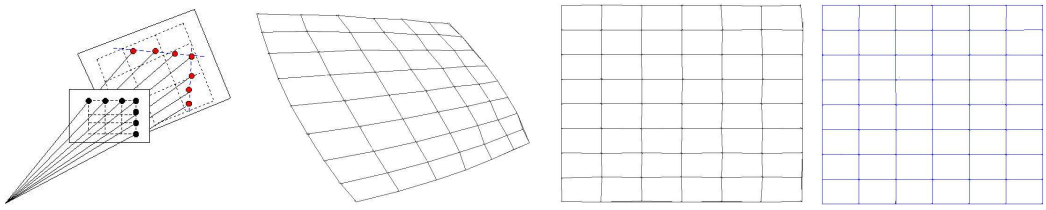
First, the non-central camera model encompasses the central one of course. However, the non-central calibration algorithm of section 3.1, can not be used as such to calibrate a central camera: data (pixel-to-grid correspondences) coming from a central camera, will lead to a higher rank-deficiency in the linear solution of the tensors, causing an incorrect calibration (although residuals will be lower). However, we may, by analyzing the rank of the underlying equation system, detect this problem and maybe even classify the camera as being central and then apply the appropriate calibration algorithm. More generally speaking, this is a model selection problem, and the rank-analysis or any other solution will allow to build a truly complete black box calibration system.

To this end, we have to take into account a few intermediate camera models that may be encountered in practice. One such case is the class of cameras for which there exists a single line that cuts all projection rays (we call them *axial cameras*). Examples are the classical two-camera stereo systems (the mentioned line is the baseline joining the two optical centers) and certain non-central catadioptric cameras, e.g. all catadioptric cameras with a spherical mirror. A yet more special class of cameras are so-called crossed-slits cameras [7], which encompass pushbroom cameras [10]. We are currently specializing our calibration approach to these additional general classes of camera types. Overall, it seems that these 4 classes (central, axial, crossed-slits, fully non-central) and their associated calibration algorithms, maybe with a few additional classes, should be sufficient to calibrate most cameras.

Besides considering these general camera types, we may also discuss the choice between parametric and non-parametric models for a given camera. Parametric models enable faster, simpler and compact algorithms for many applications like perspective image synthesis. In any case we see a potential advantage with our non-parametric generic calibration algorithm. It not only allows to calibrate any camera system by treating it as a black box, it also provides the ability to easily obtain a parametric calibration once the model for the camera is known. We also believe that by using the non-parametric data, which is available as the mapping between pixels and the corresponding rays, we would be having enough and sufficient information to enable an easy and accurate parametric calibration. Every parametric calibration will just be a model-fitting problem, which can be solved as a non-linear optimization problem starting with the good initial solution obtained using generic calibration.

4 Practical Issues

Now we discuss some important practical issues we faced during experiments. The first important issue is the design of the calibration grid. We found that grids with circular targets provide stable calibration compared to checkerboard patterns. In the case of circular targets we can for example use the point spread function to compute their centers. When the images are not focused (which often happens with omnidirectional cameras), the corners of the checkerboard pattern are difficult to estimate. This was found to be a problem e.g. for the calibration of a multi-camera setup, whereas circular targets led to a stable calibration.



(a) (b) (c) (d)

Figure 5: Application of collinearity constraints (see text). a) interpolated grid points of a calibration grid, b) grid in a fisheye image, c) perspectively synthesized grid using calibration, obtained *without* the collinearity constraint, d) perspectively synthesized grid, using calibration, obtained *with* the collinearity constraint.

A second issue is the orientation of calibration grids. Usage of grids with very different orientations and positions is important for stable calibration. One way to easily achieve this is to use calibration grids of different sizes and to put them at different distances from the camera (together with sufficient orientation differences).

The third practical issue is related to the dense image-to-grid matching where we match the 2D image features to 3D grid coordinates. By using a combination of *local* 4-point homography based prediction, local collinearity and orthogonality constraints, we start from four features (circular targets or corners), located at the corners of a square, and incrementally extend the matching along all directions. This automatic approach worked successfully for all pinhole and fisheye images as well as several images obtained using hyperbolic and spherical catadioptric systems. However we also had to use manual input and correction for some images.

The last issue is concerned with a required interpolation process: for every grid point in the first image we compute the interpolated points in the other grids' coordinate systems (since for other grids, the extracted targets or corners do not lie on the same pixels in general). Noise in the feature extraction, the nature of the camera model, etc. usually introduce errors in the interpolation. Thus we use a global collinearity constraint in the case of central cameras and a local collinearity constraint in the case of non-central ones during the process of interpolation. We observed a significant improvement in the stability of the tensors. This is illustrated by the example in Figure 5, which shows the perspective view synthesis (or, distortion correction) for the image of a grid that was not used in the calibration process.

5 Experiments and Results

We have calibrated a wide variety of cameras (both central and non-central) as shown in Table 3. Results are first discussed for several "slightly non-central" cameras, and then for a multi-camera system.

Slightly non-central cameras: central vs. non-central models. For three cameras (a fisheye, a hyperbolic and a spherical catadioptric system, see sample images in Figure 6), we applied both, central calibration and the procedure explained in section 3.2, going from central to non-central. Table 3 shows that the bundle adjustment's residual errors for central and non-central calibration, are very close to one another for the fisheye and hyperbolic catadioptric cameras. This suggests that for the cameras used in the experiments, the central model is appropriate. As for the spherical catadioptric camera, the non-central model has a significantly lower residual, which may suggest that a non-central model is better here.

To further investigate this issue we performed another evaluation. A calibration grid was put on a turntable, and images were acquired for different turntable positions. We are thus able to quantitatively evaluate the calibration, by measuring how close the recovered grid pose corresponds to a turntable sequence. Individual grid points move on a circle in 3D; we thus compute a least squares circle fit to the 3D positions given by the estimated grid pose. The recovered grid poses are shown in Figures 7(a) and (b). Figure 7(c) shows the extension of a line in the grid's coordinate system, for the different poses. Due to the turntable motion, these should envelope a quadric close to a cone, which indeed is the case. A complete quantitative analysis is difficult, but we evaluated how close the trajectories of individual grid points are to being circular (as they should be, due to the turntable motion). The least-squares circle fit for one of the grid points, from its 14 recovered positions, is shown in Figure 7(d). Table 4 shows the RMS errors of circle fits (again, relative to scene size, and given in percent). We note that the non-central model provides a significantly better reconstruction than the central one for the spherical catadioptric camera, which thus confirms the above observation. For the fisheye, the non-central calibration also performs better, but not as significantly. As for the hyperbolic catadioptric camera, the central model gives a better reconstruction though. This can probably be explained as follows. Inspite potential imprecisions in the camera setup, the camera seems to be sufficiently close to a central one, so that the non-central model leads to overfitting. Consequently, although the bundle adjustment's residual is lower than for the central model (which always has to be the case), it gives "predictions" (here, pose or motion estimation) which are unreliable.

Calibration of a multi-camera system. A multi-camera network can be considered as a single generic imaging system. As shown in Figure 8 (left), we used a system of three (approximately pinhole) cameras to capture three images each of a calibration grid. We virtually concatenated the images from the individual cameras and computed all projection rays and the three grid poses in a single reference frame (see Figure 8 (right)), using the non-central algorithm described in section 3.1.

Camera	Images	Rays	Points	RMS
Pinhole (C)	3	217	651	0.04
Fisheye (C)	23	508	2314	0.12
(NC)	23	342	1712	0.10
Sphere (C)	24	380	1441	2.94
(NC)	24	447	1726	0.37
Hyperbolic (C)	24	293	1020	0.40
(NC)	24	190	821	0.34
Multi-Cam (NC)	3	1156	3468	0.69
Eye+Pinhole (C)	3	29	57	0.98

Table 3: Bundle adjustment statistics for different cameras. (C) and (NC) refer to central and non-central calibration respectively, and RMS is the root-mean-square residual error of the bundle adjustment (ray-point distances). It is given in percent, relative to the overall size of the scene (largest pairwise distance between points on calibration grids).

Camera	Grids	Central	Non-Central
Fisheye	14	0.64	0.49
Spherical	19	2.40	1.60
Hyperbolic	12	0.81	1.17

Table 4: RMS error for circle fits to grid points, for turntable sequences (see text).

(a) (b) (c)

Figure 6: sample images for a) hyperbolic catadioptric camera b) spherical catadioptric camera and c) fisheye

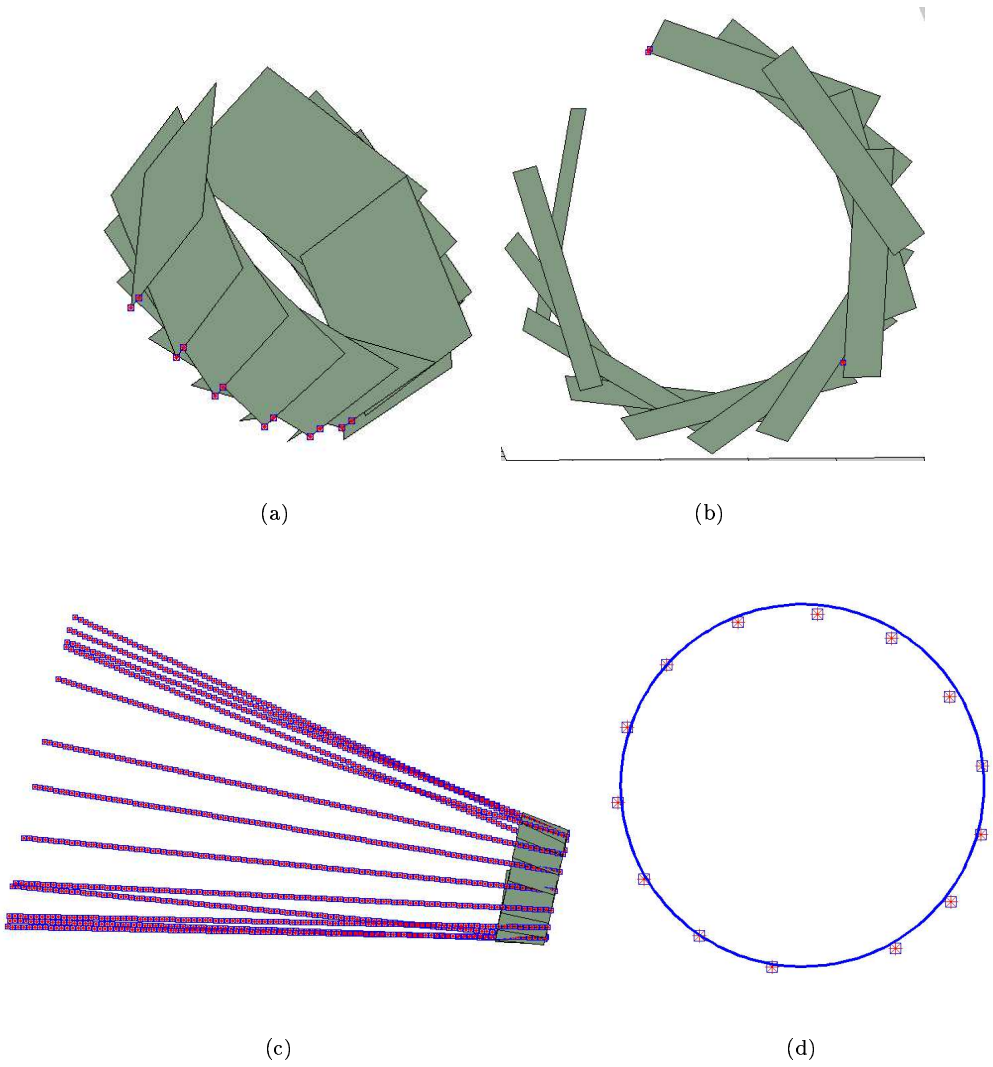


Figure 7: Experiment on pose estimation. (a) and (b) Estimated poses of calibration grid in 14 positions. (c) Extensions of a line on the calibration grid, in all 14 positions. (d) Least squares circle fit to the estimated positions of one grid point.

In order to evaluate the calibration, we compared results with those obtained by plane-based calibration [23, 30], that used the knowledge that the three cameras are pinholes. In

both, our multi-camera calibration, and plane-based calibration, the first grid was used to fix the global coordinate system. We can thus compare the estimated poses of the other two grids for the two methods. This is done for both, the rotational and translational parts of the pose. As for rotation, we measure the angle (in radians) of the relative rotation between the rotation matrices given by the two methods, see columns R_i in Table 5). As for translation, we measure the distance between the estimated 3D positions of the grids' centers of gravity (columns t_i in Table 5) expressed in percent, relative to the scene size. Here, plane-based calibration is done separately for each camera, leading to the three rows of Table 5.

From the non-central multi-camera calibration, we also estimate the positions of the three optical centers, by clustering the projection rays and computing least squares point fits to them. The column "Center" of Table 5 shows the distances between optical centers (expressed in percent and relative to the scene size) computed using this approach and plane-based calibration. The discrepancies are low, suggesting that the non-central calibration of a multi-camera setup is indeed feasible.

Another experiment we carried out was to calibrate a small region of the exotic catadioptric system formed with an eye as mirror, see Figure 9. We are currently working on applying central and non-central models to such a system.

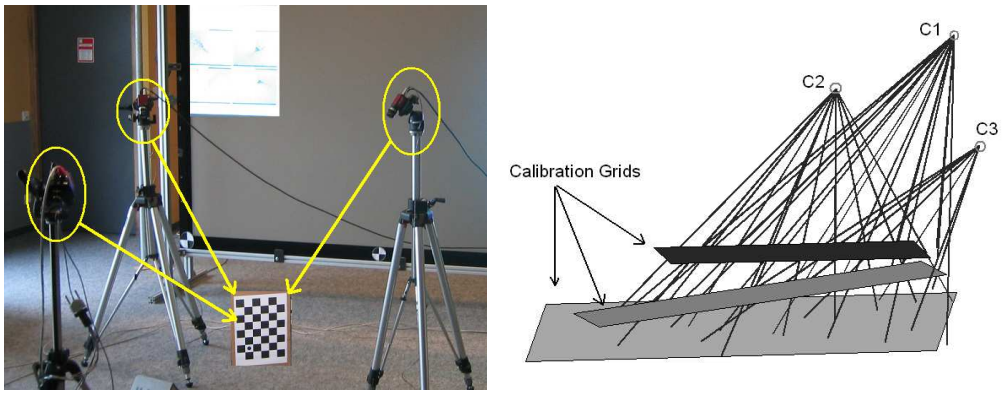


Figure 8: Multi-camera setup consisting of 3 cameras (left). Recovered projection rays and grid poses (right).

Camera	R_2	R_3	t_2	t_3	Center
1	0.0117	0.0359	0.56	3.04	2.78
2	0.0149	0.0085	0.44	2.80	2.17
3	0.0088	0.0249	0.53	2.59	1.16

Table 5: Evaluation of non-central multi-camera calibration relative to plane-based calibration. See text for more details.

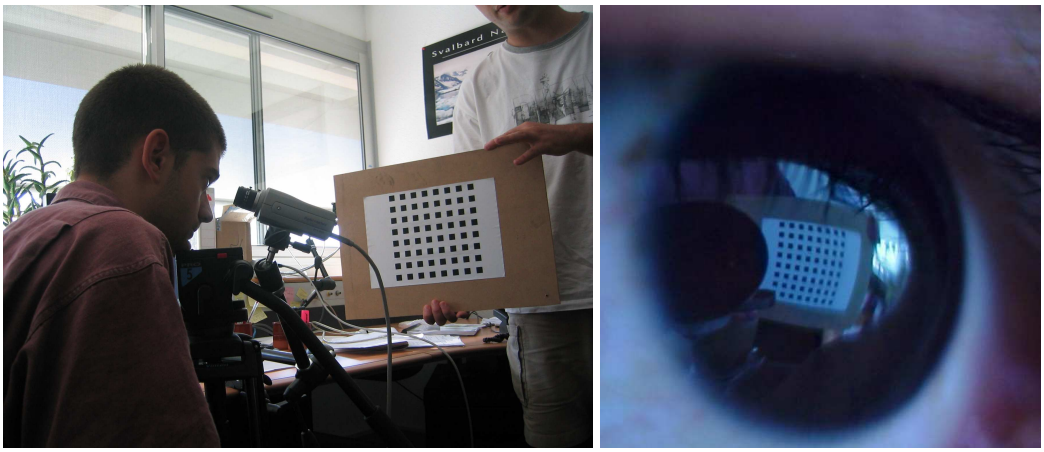


Figure 9: Catadioptric system with an eye as mirror: setup (left) and an acquired image (right).

6 Summary and Conclusions

We have proposed a non-parametric, generic calibration approach and shown its feasibility by calibrating a wide variety of cameras. One of the important issues is in the identification of appropriate models, central or non-central, for slightly non-central cameras. For understanding complex cameras or mirror surfaces, where mathematical modeling might be more demanding, generic calibration can be used as a black box tool to first obtain the projection rays. The nature of these projection rays can be experimented further to identify the right parametric model.

References

[1] D. Aliaga. Accurate Catadioptric Calibration for Real-size Pose Estimation of Room-size Environments, *ICCV*, 2001.

[2] S. Baker and S. Nayar. A theory of catadioptric image formation. *ICCV*, 1998.

[3] H. Bakstein and T. Pajdla. An overview of non-central cameras. *Computer Vision Winter Workshop*, Ljubljana, Slovenia, 2001.

[4] A. Bartoli. On the non-linear optimization of projective motion using minimal parameters. *ECCV*, 2002.

[5] G. Champleboux, S. Lavallée, P. Sautot and P. Cinquin. Accurate Calibration of Cameras and Range Imaging Sensors: the NPBS Method. *ICRA*, 1992.

[6] C.-S. Chen and W.-Y. Chang. On Pose Recovery for Generalized Visual Sensors. *PAMI*, 2004.

[7] Doron Feldman, Tomas Pajdla and Daphna Weinshall. On the Epipolar Geometry of the Crossed-Slits Projection. *ICCV*, 2003.

[8] K.D. Gremban, C.E. Thorpe, and T. Kanade. Geometric Camera Calibration using Systems of Linear Equations. *ICRA*, 1988.

[9] M.D. Grossberg and S.K. Nayar. A general imaging model and a method for finding its parameters. *ICCV*, 2001.

[10] R. Gupta and R.I. Hartley. Linear Pushbroom Cameras. *PAMI* 1997.

[11] R.M. Haralick, C.N. Lee, K. Ottenberg, and M. Nolle. Review and analysis of solutions of the three point perspective pose estimation problem. *IJCV*, 1994.

[12] C. Harris and M. Stephens. A Combined Corner and Edge Detector. *Alvey Vision Conference*, 1988.

[13] R.A. Hicks and R. Bajcsy. Catadioptric Sensors that Approximate Wide-angle Perspective Projections. *CVPR*, 2000.

[14] B. Micusik and T. Pajdla. Autocalibration and 3D Reconstruction with Non-central Catadioptric Cameras. *CVPR*, 2004.

[15] J. Neumann, C. Fermüller, and Y. Aloimonos. Polydioptric Camera Design and 3D Motion Estimation. *CVPR*, 2003.

[16] D. Nistér. A Minimal Solution to the Generalized 3-Point Pose Problem. *CVPR*, 2004.

[17] T. Pajdla. Stereo with oblique cameras. *IJCV*, 2002.

[18] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic Stereo Imaging. *PAMI*, 2001.

[19] R. Pless. Using Many Cameras as One. In *CVPR*, 2003.

[20] S. Ramalingam, P. Sturm, and S.K. Lodha. Towards Complete Generic Camera Calibration. In *CVPR*, 2005.

[21] S. Seitz and J. Kim. The Space of All Stereo Images. *IJCV*, 2002.

[22] H.Y. Shum and L.W. He. Rendering with Concentric Mosaics. *SIGGRAPH*, 1999.

[23] P. Sturm and S. Maybank. On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications. *CVPR*, 1999.

[24] P. Sturm and S. Ramalingam. A generic calibration concept-theory and algorithms. Research Report 5058, INRIA, 2003.

[25] P. Sturm and S. Ramalingam. A generic concept for camera calibration. *ECCV*, 2004.

[26] R. Swaminathan, M.D. Grossberg, and S.K. Nayar. A perspective on distortions. *CVPR*, 2003.

[27] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle Adjustment, A modern synthesis. *Workshop on Vision Algorithms: Theory and Practice*, 2000.

[28] Y. Wexler, A.W. Fitzgibbon, and A. Zisserman. Learning epipolar geometry from image sequences. *CVPR*, 2003.

[29] D.N. Wood, A. Finkelstein, J.F. Hughes, C.E. Thayer, and D.H. Salesin. Multiperspective Panoramas for Cel Animation. *SIGGRAPH*, 1997.

[30] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, 2000.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399