A Minimal Solution for Camera Calibration Using Independent Pairwise Correspondences

Francisco Vasconcelos[1], João P. Barreto[1], and Edmond Boyer[2]

[1] ISR, University of Coimbra, Portugal
[2] INRIA Grenoble Rhône-Alpes, France

Abstract. We propose a minimal algorithm for fully calibrating a camera from 11 independent pairwise point correspondences with two other calibrated cameras. Unlike previous approaches, our method neither requires triple correspondences, nor prior knowledge about the viewed scene. This algorithm can be used to insert or re-calibrate a new camera into an existing network, without having to interrupt operation. Its main strength comes from the fact that it is often difficult to find triple correspondences in a camera network. This makes our algorithm, for the specified use cases, probably the most suited calibration solution that does not require a calibration target, and hence can be performed without human interaction.

Keywords: Camera Networks; Multiple View Geometry; Camera Calibration

1 Introduction

Camera networks are sets of cameras whose fields of view are usually shared between two or more cameras within a network. They find applications in several domains that are concerned with the capture, the record, and the analysis of dynamic scenes, for instance surveillance and animation modelling applications [15]. Most of these applications require the calibration of cameras in order to perform geometric operations such as reconstruction. One widely proposed approach to achieve this purpose is to use a calibration pattern or rig [18, 2, 20, 19, 14, 4], which is typically an offline procedure, requiring human intervention. However, there are many situations where a simpler and unsupervised scheme is desirable; In particular when adding a camera or modifying its location or characteristics while operating the network.

Without prior 3D information, image correspondences between cameras must be considered. Since camera networks are often sparse, correspondences in 3 or more images can be difficult to obtain, hence preventing the use of traditional calibration tools [11]. In contrast, correspondences between 2 images are more likely to be available by construction of camera networks.

This paper addresses the issue of fully calibrating a camera given independent correspondences with 2 calibrated cameras. The situations particularly targeted are the addition or the modification of a camera in a calibrated network under

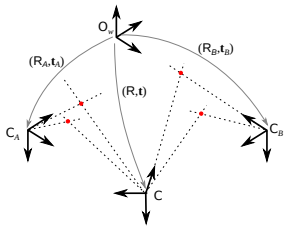


Fig. 1. We consider the problem of fully calibrating the camera $\mathbf{C}$, given pairwise correspondences with two calibrated cameras $\mathbf{C}_A$ and $\mathbf{C}_B$.

operation, which are common situations in practical camera network setups. The literature provides solutions when correspondences with 3 or more cameras are available [11], as well as when there is a mixture of corresponences with 2 and 3 cameras [8]. However, few efforts have been made to solve for the case with only 2-view correspondences. We investigate this issue and derive a minimal solution that requires 11 correspondences to estimate the 11 parameters of the unknown camera.

Our contribution to camera calibration is twofold:

- A minimal solution for the full calibration of a camera given independent correspondences with 2 calibrated cameras.
- A simple and efficient scheme that can incrementally update camera network calibration using already calibrated cameras.

1.1 Notation

Scalars are represented by plain letters, e.g. λ, vectors are indicated by bold symbols, e.g. $\mathbf{t}$, and matrices are denoted by letters in sans serif font, e.g. T. 3D lines are expressed in homogeneous Plucker coordinates, e.g. the 6×1 vector $\mathbf{L}$. The equality up to scale is denoted by $\sim$ in order to be distinguished from the strict equality $=$, and the operator $[\mathbf{v}]_\times$ designates the 3×3 skew symmetric matrix of a 3×1 vector $\mathbf{v}$. We also use matrix superscripts, e. g. $\mathsf{T}^{\{n\}}$, to denote its nth column.

2 Problem Statement

Let's consider two calibrated cameras $\mathbf{C}_A$ and $\mathbf{C}_B$, such that the matrices of intrinsic parameters are K_A and K_B, and the absolute poses are expressed in a world coordinate system $\mathbf{O}_w$ by the rotations matrices R_A and R_B, and the translation vectors $\mathbf{t}_A$ and $\mathbf{t}_B$. Consider an additional camera $\mathbf{C}$ for which both the intrinsic calibration K, and the extrinsic calibration R, $\mathbf{t}$ are unknown. Our article addresses the problem of calibrating this third camera using as input data a set of a image correspondences $(\mathbf{x}^{(i)}, \mathbf{x}_A^{(i)})$ between $\mathbf{C}$ and $\mathbf{C}_A$, and set of

b image correspondences $(\mathbf{x}^{(a+j)}, \mathbf{x}_B^{(j)})$ between $\mathbf{C}$ and $\mathbf{C}_B$ (Fig. 1). We assume that the two sets of pairwise matches are independent, meaning that

$$\mathbf{x}^i \neq \mathbf{x}^{a+j}, \forall_{i=1\ldots a,\, j=1\ldots b}\,.$$

In other words, there are no triplets of correspondences generated by scene points that are simultaneously seen by the three cameras. The absence of triple matches precludes the application of the standard calibration techniques that are described in text books [7, 11]. These approaches typically rely on the recovery of 3D points using the calibrated stereo views and standard triangulation [6]. These 3D points can in turn be used as reference points for the calibration of the 3rd camera [11, 2]. A possible alternative is to build a *measurement matrix* with the image correspondences, and perform projective factorization using the Sturm-Triggs algorithm [17] with a suitable extension for handling missing data [18]. However, this class of methods is meant for problems with multiple cameras and large number of correspondences, and it is unlikely that they will converge to a solution with pairwise correspondences only.

In summary, and to the best of our knowledge, the calibration of a camera using independent pairwise correspondences with two other views has never been addressed in the literature before. We present in the following a minimal solution when 7 or more matches with one of the view are available.

3 Linear system of equations with a minimum number of unknowns

In this section we derive a system of linear equations that has a minimum number of unknowns and fully constrains the camera calibration. The problem is formulated in the context of epipolar geometry between general camera models [12], with one side being the uncalibrated pin-hole camera $\mathbf{C}$, and the other side being the pair of calibrated cameras $\mathbf{C}_A$ and $\mathbf{C}_B$ that can be understood as a particular instance of a non-central imaging device denoted by $\mathbf{C}_A \cup \mathbf{C}_B$. It is shown below that under such configuration the corresponding back-projection lines must satisfy a bilinear relation expressed by a 3×5 matrix, and that the estimation of the epipolar geometry using a DLT-like approach cannot be achieved with less than 14 pairwise matches.

Note that when the intrinsics are known, this problem is a particular case of the pose estimation between calibrated general camera models [12] that has already been solved both linearly [9] and using the minimal number of 6 pairwise correspondences [16].

3.1 Line Incidence Relations

Let $\mathbf{x}_A$ and $\mathbf{x}_B$ be image points in $\mathbf{C}_A$ and $\mathbf{C}_B$. Since the cameras are fully calibrated, the corresponding back-projection lines $\mathbf{L}_A$ and $\mathbf{L}_B$ can be expressed

in the common world reference frame $\mathbf{O}_w$ by a homogeneous Plücker vector

$$\mathbf{L}_{A/B} \sim \begin{pmatrix} \mathbf{d}_{A/B} \\ \mathbf{m}_{A/B} \end{pmatrix},$$

with the 3-vectors $\mathbf{d}_{A/B}$ and $\mathbf{m}_{A/B}$ being respectively the direction and the momentum of the line. In a similar manner, an image point $\mathbf{x}$ in $\mathbf{C}$ gives raise to a back-projection line $\mathbf{L}$ that is represented in the local camera reference frame by

$$\mathbf{L} \sim \begin{pmatrix} \mathbf{d} \\ \mathbf{0} \end{pmatrix},$$

with the direction depending on the matrix of intrinsic parameters K

$$\mathbf{d} \sim \mathsf{K}^{-1}\mathbf{x}. \tag{1}$$

If $\mathbf{x}$ and $\mathbf{x}_{A/B}$ are image correspondences, then the back-projection lines $\mathbf{L}$ and $\mathbf{L}_{A/B}$ must be incident. Given the rigid displacement between the reference frames $\mathbf{O}_w$ and $\mathbf{C}$, and the condition for two lines in Plücker coordinates to intersect, it comes that the following condition must hold

$$\mathbf{L}^{\mathsf{T}} \begin{pmatrix} \mathsf{0} & \mathsf{I} \\ \mathsf{I} & \mathsf{0} \end{pmatrix} \begin{pmatrix} \mathsf{R} & \mathsf{0} \\ [\mathbf{t}]_{\times}\mathsf{R} & \mathsf{R} \end{pmatrix} \mathbf{L}_{A/B} = 0.$$

Since the momentum of $\mathbf{L}$ is always zero, then the above equation can be re-written as

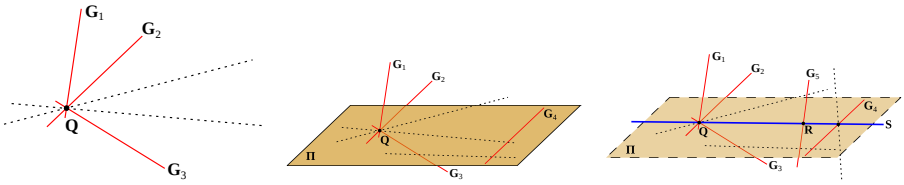
$$\mathbf{d}^{\mathsf{T}} \left([\mathbf{t}]_{\times}\mathsf{R}\ \mathsf{R} \right) \mathbf{L}_{A/B} = 0 \tag{2}$$

Equation 2 is the particular case of the generalized epipolar constraint proposed in [12] when one of the cameras is a conventional pin-hole. However, and similarly to the general case, the bilinear relation between back-projection lines is expressed by a 3×6 matrix that encodes the calibration parameters. Therefore, the linear estimation the 18 entries of the matrix up to a global scale factor still requires a minimum of 17 image correspondences between $\mathbf{C}$ and the camera pair $\mathbf{C}_A \cup \mathbf{C}_B$.

3.2 Analysis using linear line subspaces

In our case the parametrization of equation 2 leads to a linear estimation problem that is sub-determined. This is a situation similar to the degenerate configurations recently reported in [9] in the context of motion estimation using a calibrated multi-camera rig. We use the theory of linear line subspaces [13] to explain the underlying reasons of the sub-determination, and prove that the calibration problem can be formulated in a linear manner using a minimum of 15 parameters

It is well known that a line in 3D represented in Plücker coordinates can be thought as a point in $\mathbb{P}^5$ lying in the so-called *Klein Quadric*. Let's consider a generic hyperplane in $\mathbb{P}^5$ with dimension $N \leq 5$. The hyperplane intersects the



(a) Line Bundle (b) Linear line subspace for $N = 4$ (c) Linear line congruent

Fig. 2. Line subspaces; in each figure, the dotted lines represent possible lines that could result from the linear combination of the generation basis.

quadric in a locus that defines, via Plücker mapping, a certain subset of lines in the original 3D space. This subset is called a *linear line subspace* (LLS) of dimension N, and each line $\mathbf{L}$ in the LLS is in the linear span of N other lines $\mathbf{G}$ with independent Plücker vectors [13].

The lines going through a generic 3D point $\mathbf{Q}$ form a *line bundle* that is often used to model the back-projection rays of a pin-hole camera (Fig. 2(a)). The line bundle is a LLS of dimension $N = 3$, and each line $\mathbf{L}$ going through $\mathbf{Q}$ can be uniquely expressed as the linear combination of any other three non-coplanar lines in the bundle. It is said that these three lines $\{\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3\}$ are a basis for the LLS. Consider now an additional line $\mathbf{G}_4$ that is not in the bundle. In this case the span of $\{\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4\}$ contains a LLS of dimension $N = 4$ that comprises all the lines that go through point $\mathbf{Q}$, and all the lines that lie in the plane $\mathbf{\Pi}$ defined by $\mathbf{Q}$ and $\mathbf{G}_4$ (Fig. 2(b)). Finally, the addition of a fifth line $\mathbf{G}_5$ to the basis gives raise to a LLS of dimension $N = 5$ that is called a *linear line congruent* (LLC) [13]. The LLC includes all the lines tangent to an axis $\mathbf{S}$ that, in the particular case of Fig. 2(c), is defined by the center $\mathbf{Q}$ and the point $\mathbf{R}$ where $\mathbf{G}_5$ meets the plane $\mathbf{\Pi}$.

Let's return to our calibration problem where the generalized camera $\mathbf{C}_A \cup \mathbf{C}_B$ is modeled by the union of two distinct line bundles. The key observation is that every possible back-projection line $\mathbf{L}_{A/B}$ must be tangent to the line going through $\mathbf{C}_A$ and $\mathbf{C}_B$ (the baseline). Thus, and since the lines $\mathbf{L}_{A/B}$ are contained in a LLC, they can be represented in a unique manner as the linear combination of any 5 lines $\mathbf{G}_i$ that intersect the baseline

$$\mathbf{L}_{A/B} \sim \underbrace{\left(\mathbf{G}_1\ \mathbf{G}_2\ \mathbf{G}_3\ \mathbf{G}_4\ \mathbf{G}_5\right)}_{\mathsf{G}} \boldsymbol{\lambda}_{A/B}\,,$$

where G is a 6×5 matrix with full rank, and $\boldsymbol{\lambda}_{A/B}$ is a 5-vector defined up to scale. Replacing in equation 2 yields

$$\mathbf{d}^{\mathsf{T}} \left([\mathbf{t}]_{\times} \mathsf{R}\ \mathsf{R}\right)\ \mathsf{G}\, \boldsymbol{\lambda}_{A/B} = 0 \tag{3}$$

We have just re-written the epipolar constraint of equation 2 as a bilinear relation between the direction $\mathbf{d}$ of the line $\mathbf{L}$ in camera $\mathbf{C}$, and the representation

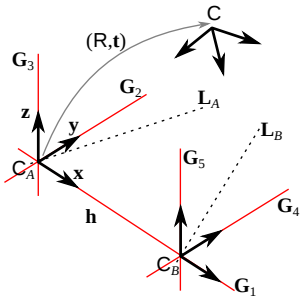


Fig. 3. The space generated by two bundles of lines (the rays of 2 pinhole cameras) can be fully represented as the linear span of $\{\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4, \mathbf{G}_5\}$.

$\boldsymbol{\lambda}_{A/B}$ of the back-projection line $\mathbf{L}_{A/B}$ in the generalized camera $\mathbf{C}_A \cup \mathbf{C}_B$. Since the bilinear relation is now encoded by a 3×5 matrix with 15 entries, then 14 image point correspondences are sufficient for estimating the epipolar geometry in a DLT-like manner. The discussion clearly explains why the 18 parameter formulation of equation 2 is ambiguous [9], and shows that a compact linear formulation of the stated calibration problem must necessarily have 15 parameters because the lowest dimensional linear sub-space containing all the back-projection rays of two pin-holes is a LLC.

3.3 Compact linear formulation

Given the two arbitrary calibrated cameras, it is always possible to perform a change of reference frames for achieving the configuration exhibited in Fig. 3.3. We consider, without any loss of generality, that the world reference frame is aligned with the coordinate system of camera $\mathbf{C}_A$, and that the X-axis is coincident with the baseline defined by the projection centers of the two pin-holes. The local reference frame of the second camera is assumed to have origin in $\mathbf{C}_B$ and to be parallel to the coordinate system of $\mathbf{C}_A$. Under such circumstances the rigid transformation that maps point coordinates from $\mathbf{C}_B$ to $\mathbf{C}_A$ is given by

$$\mathsf{T}_{B \to A} = \begin{pmatrix} \mathtt{I} & \mathbf{h} \\ \mathbf{0} & 1 \end{pmatrix}$$

with $\mathtt{I}$ being the 3×3 identity matrix and $\mathbf{h} = \begin{pmatrix} h & 0 & 0 \end{pmatrix}^{\mathsf{T}}$. Since the axes X, Y, Z of the system of coordinates of $\mathbf{C}_A$, and the axes Y, Z of the reference frame of $\mathbf{C}_B$ are linearly independent lines, then they can be used to establish a basis G for the LLC defined by the baseline. It comes that

$$\mathsf{G} \sim \begin{pmatrix} \mathtt{I} & \mathtt{I}^{\{2,3\}} \\ \mathtt{0} & [\mathbf{h}]_{\times}^{\{2,3\}} \end{pmatrix}$$

with the upper script $\{2, 3\}$ denoting the second and third columns of the matrix.

Let's now consider an image correspondence $(\mathbf{x}, \mathbf{x}_A)$ between $\mathbf{C}$ and $\mathbf{C}_A$. The back-projection of $\mathbf{x}_A$ is a line $\mathbf{L}_A$ with direction $\mathbf{d}_A$ expressed in the reference frame of $\mathbf{C}_A$. Given the basis G above, it comes that $\mathbf{L}_A \sim \mathsf{G}\boldsymbol{\lambda}_A$ with

$$\boldsymbol{\lambda}_A \sim \left(\mathbf{d}_A^\mathsf{T}\ 0\ 0\right)^\mathsf{T}.$$

Replacing in equation 3, and making $\mathbf{d} \sim \mathsf{K}^{-1}\mathbf{x}$, yields

$$\mathbf{x}^\mathsf{T}\,\mathsf{F}_A\,\mathbf{d}_A = 0 \tag{4}$$

with F_A being the standard fundamental matrix between the uncalibrated camera $\mathbf{C}$ and the calibrated view $\mathbf{C}_A$

$$\mathsf{F}_A = \mathsf{K}^{-\mathsf{T}}\,[\mathbf{t}]_\times\,\mathsf{R} \tag{5}$$

Repeating the reasoning for the case of an image correspondence $(\mathbf{x}, \mathbf{x}_B)$ between $\mathbf{C}$ and $\mathbf{C}_B$, it comes that

$$\boldsymbol{\lambda}_B \sim \left(d_{B,1}\ 0\ 0\ d_{B,2}\ d_{B,3}\right)^\mathsf{T}.$$

with $\mathbf{d}_B \sim \left(d_{B,1}\ d_{B,2}\ d_{B,3}\right)^\mathsf{T}$ being the direction of the back-projection line $\mathbf{L}_B$ expressed in the local reference frame of camera $\mathbf{C}_B$. Making $\mathbf{L}_B \sim \mathsf{G}\boldsymbol{\lambda}_B$ in equation 3, and taking into account that the first column of $[\mathbf{h}]_\times$ is a null vector, we obtain that

$$\mathbf{x}^\mathsf{T}\,\mathsf{F}_B\,\mathbf{d}_B = 0 \tag{6}$$

with F_B being the fundamental matrix between $\mathbf{C}$ and $\mathbf{C}_B$ that can be written as

$$\mathsf{F}_B = \mathsf{F}_A + \mathsf{K}^{-1}\mathsf{R}[\mathbf{h}]_\times \tag{7}$$

It follows from the equation above that the first columns of matrices F_A and F_B are always equal ($\mathsf{F}_A^{\{1\}} = \mathsf{F}_B^{\{1\}}$).

Given the image correspondences $(\mathbf{x}^{(i)}, \mathbf{x}_A^{(i)})$, with $i = 1, \ldots a$, and $(\mathbf{x}^{(a+j)}, \mathbf{x}_B^{(j)})$ with $j = 1, \ldots b$, we can determine the line directions $\mathbf{d}_A^{(i)} \sim \mathsf{K}_A^{-1}\mathbf{x}_A^{(i)}$ and $\mathbf{d}_B^{(j)} \sim \mathsf{K}_B^{-1}\mathbf{x}_B^{(j)}$, and establish a system of linear equations based on the bilinear constraints of equations 4 and 6

$$\begin{pmatrix}
x_1^{(1)}\mathbf{d}_A^{(1)\mathsf{T}} & x_2^{(1)}\mathbf{d}_A^{(1)\mathsf{T}} & x_3^{(1)}\mathbf{d}_A^{(1)\mathsf{T}} & \mathbf{0}^\mathsf{T} & \mathbf{0}^\mathsf{T} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_1^{(a)}\mathbf{d}_A^{(a)\mathsf{T}} & x_2^{(a)}\mathbf{d}_A^{(a)\mathsf{T}} & x_3^{(a)}\mathbf{d}_A^{(a)\mathsf{T}} & \mathbf{0}^\mathsf{T} & \mathbf{0}^\mathsf{T} \\
x_1^{(a+1)}\mathbf{d}_B^{(1)\mathsf{T}} & \mathbf{0}^\mathsf{T} & \mathbf{0}^\mathsf{T} & x_2^{(a+1)}\mathbf{d}_B^{(1)\mathsf{T}} & x_3^{(a+1)}\mathbf{d}_B^{(1)\mathsf{T}} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
x_1^{(a+b)}\mathbf{d}_B^{(b)\mathsf{T}} & \mathbf{0}^\mathsf{T} & \mathbf{0}^\mathsf{T} & x_2^{(a+b)}\mathbf{d}_B^{(b)\mathsf{T}} & x_3^{(a+b)}\mathbf{d}_B^{(b)\mathsf{T}}
\end{pmatrix}
\begin{pmatrix}
\mathsf{F}_A^{\{1\}} \\ \mathsf{F}_A^{\{2\}} \\ \mathsf{F}_A^{\{3\}} \\ \mathsf{F}_B^{\{2\}} \\ \mathsf{F}_B^{\{3\}}
\end{pmatrix} = 0 \tag{8}$$

If $a + b \geq 14$ then the fundamental matrices F_A and F_B can be determined up to a common scale factor using a standard DLT approach.

4 A minimal solution for the estimation of F_A and F_B

We have shown that the two fundamental matrices, F_A and F_B, that encode the calibration information K, R, and $\mathbf{t}$, can be determined from a minimum of 14 independent image correspondences. However, the total number of independent unknowns is 11 (5 intrinsic parameters and 6 extrinsic parameters) meaning that the estimation problem can be further constrained. Two of these constrains are rather obvious:

$$\det(\mathsf{F}_A) = 0 \tag{9}$$

$$\det(\mathsf{F}_B) = 0 \tag{10}$$

For the third constraint it must observed that the sum of F_A and F_B is still a fundamental matrix. From equations 5 and 7 it comes after algebraic manipulation that

$$\mathsf{F}_A + \mathsf{F}_B = \mathsf{K}^{-1}[2\mathbf{t} + \mathsf{R}\mathbf{h}]_\times \mathsf{R}\,,$$

which means that the following condition must hold

$$\det(\mathsf{F}_A + \mathsf{F}_B) = 0 \tag{11}$$

The equation above basically enforces the condition that F_A and F_B must be two fundamental matrices encoding the same rotation R.

4.1 Outline of the estimation algorithm

This section outlines an algorithm for estimating F_A and F_B from a minimum number of $a + b = 11$ pairwise correspondences. The solution is found by determining the 4-dimensional null space of the measurement matrix of the the linear system of equation 8, followed by intersecting the span of this null-space with the locus defined by the polynomials constraints of equations 9 to 11. Instead of solving a system of 3 third order polynomials in 3 variables, we explore the sparsity of the measurement matrix and simplify the problem to solving 1 cubic polynomial in 1 variable, and a system of 2 quadratic polynomials in 2 variables. In order for this to be possible 7 of the 11 image matches must be in the same calibrated view ($a = 7$, $b = 4$)[3].

1. Build the linear system of equation 8 from the 11 pairwise correspondences, and determine the 4-dimensional solution space using SVD. The solution space is spanned by the columns of the 15×4 matrix $\mathsf{V}^{\{12\ldots15\}}$ (the last 4 columns of V)
2. The first 9 rows of $\mathsf{V}^{\{12\ldots15\}}$ define always a rank 2 sub-matrix due to the the structure of the linear system and the fact that $a = 7$. Thus, the solution space of F_A is spanned by the two columns of the sub-matrix, $\mathbf{a}$ and $\mathbf{a}'$, that are linearly independent, which enables to write $\mathsf{F}_A(\alpha) = \mathsf{A}' + \alpha\mathsf{A}$ with α being a free parameter.

[3] The current version of the minimal algorithm cannot cope with the situation of $a = 6$ and $b = 5$

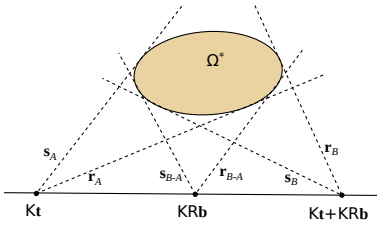


Fig. 4. Conic envelope Ω establishes linear relations $\mathbf{s}^\mathsf{T}\mathsf{K}\mathsf{K}^\mathsf{T}\mathbf{s} = 0$ and $\mathbf{r}^\mathsf{T}\mathsf{K}\mathsf{K}^\mathsf{T}\mathbf{r} = 0$.

3. Compute α by solving the cubic constraint of equation 9 and determine the fundamental matrix F_A.
4. Substitute F_A in the linear system which results in 4 equations in 7 unknowns. The solution space of this system is 3-dimensional and F_B can be written as the linear span $\mathsf{F}_B(\beta_1, \beta_2) = \mathsf{B}'' + \beta_1\mathsf{B}' + \beta_2\mathsf{B})$
5. Substitute F_A and $\mathsf{F}_B(\beta_1, \beta_2)$ in equations 10 and 11. This leads to a bivariate system of 2 quadratic equations that corresponds geometrically to determining the point intersections of two conic curves. Compute β_1 and β_2 by solving the bivariate system [3], and determine the fundamental matrix F_B.

Since the cubic equation of step 3 gives up to 3 discrete solutions, and the bivariate system of quadric equations has at most 4 distinct solutions, then there is a maximum of 12 possible solutions for the pair of fundamental matrices $(\mathsf{F}_A, \mathsf{F}_B)$.

5 Factorization of F_A and F_B

So far we have shown how to estimate the fundamental matrices F_A and F_B from a minimum of 11 pairwise correspondences. In order to solve the calibration problem, F_A and F_B must be factorized into the intrinsic parameters K and the relative pose R, $\mathbf{t}$. The absence of intrinsics in the right side of the fundamental matrices leads to a simplified version of Kruppa's equations [11, 7] that enable the recovery of K in a relatively straightforward manner. This section discusses how this can be achieved. After knowing K, we can compute the essential matrix E_A and apply standard techniques for determining the rotation R and the translation $\mathbf{t}$ up to scale factor [11, 7]. This scale factor can be easily found using the known baseline between $\mathbf{C}_A$ and $\mathbf{C}_B$.

Let's now discuss the extraction of the matrix K. Consider the fundamental matrix F_A that is given in equation 5. After some algebraic manipulations we obtain that

$$\mathsf{F}_A\,\mathsf{F}_A^\mathsf{T} \;\sim\; [\mathsf{K}\mathbf{t}]_\times\,\mathsf{K}\mathsf{K}^\mathsf{T}\,[\mathsf{K}\mathbf{t}]_\times$$

From the result above it follows that, if $\mathbf{y}$ is a point in the projective plane that satisfies

$$\mathbf{y}^\mathsf{T}\,\mathsf{F}_A\,\mathsf{F}_A^\mathsf{T}\,\mathbf{y} \;=\; 0\,,$$

then the line defined by $\mathbf{y}$ and the left epipole of F_A must lie in the conic envelope $\mathsf{K}\mathsf{K}^\mathsf{T}$ that is the dual of the image of the absolute conic (DIAC) [11, 7]. $\mathsf{F}_A\,\mathsf{F}_A^\mathsf{T}$ is a rank 2 symmetric matrix that represents a degenerate conic. Thus, and since this conic consists in two lines $\mathbf{s}_A$, $\mathbf{r}_A$ that intersect in the left epipole, it is easy to conclude that $\mathbf{s}_A$, $\mathbf{r}_A$ must belong to the DIAC. The same reasoning can be applied to the fundamental matrix F_B of equation 7

$$\mathsf{F}_B\,\mathsf{F}_B{}^\mathsf{T} \;\sim\; [\mathsf{K}(\mathsf{R}\mathbf{h}+\mathbf{t})]_\times \mathsf{K}\mathsf{K}^\mathsf{T}[\mathsf{K}(\mathsf{R}\mathbf{h}+\mathbf{t})]_\times \,,$$

and to the matrix $\mathsf{F}_B - \mathsf{F}_A$ that is still rank deficient because the first columns of the two fundamental matrices are equal

$$(\mathsf{F}_B - \mathsf{F}_A)(\mathsf{F}_B - \mathsf{F}_A)^\mathsf{T} \;\sim\; [\mathsf{K}\mathsf{R}\mathbf{h}]_\times \mathsf{K}\mathsf{K}^\mathsf{T}[\mathsf{K}\mathsf{R}\mathbf{h}]_\times \,.$$

Summarizing, and as shown in Fig. 4, the DIAC is fully constrained by the line pairs arising from the rank 2 degenerate conics $\mathsf{F}_A\mathsf{F}_A^\mathsf{T}$, $\mathsf{F}_B\mathsf{F}_B^\mathsf{T}$, and $(\mathsf{F}_B - \mathsf{F}_A)(\mathsf{F}_B - \mathsf{F}_A)^\mathsf{T}$. It is important to refer that, although we have six lines, they only give raise to five independent constraints on the parameters of the DIAC. This is explained by the fact that their pairwise intersections are collinear.

6 Experimental Results

In this section we validate our algorithm, and compare it against a calibration procedure based on a minimum of 6 triplets of correspondences, described in [11] (hereinafter called the 6 point algorithm). We found this approach to be the most common alternative to our algorithm in a real calibration network situation.

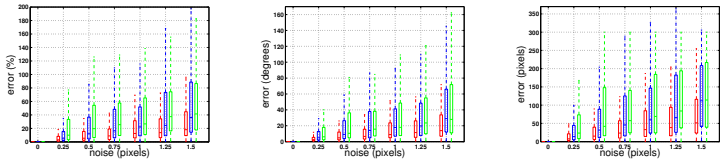
6.1 6 Point Algorithm

Given the cameras C_A, C_B, and C from our problem, suppose that instead of pairwise correspondences we have triplets of correspondences $(\mathbf{x}_A^{(i)}, \mathbf{x}_B^{(i)}, \mathbf{x}^{(i)})$ across the 3 views. We can use the points $\mathbf{x}_A^{(i)}$ and $\mathbf{x}_B^{(i)}$ to obtain a set of 3D points $\mathbf{X}^{(i)}$ by triangulation, and then establish a set of linear equations of the form

$$\mathbf{x}^{(i)} = \mathsf{P}\begin{pmatrix}\mathbf{X}^{(i)}\\ 1\end{pmatrix} \tag{12}$$

where P is a 3×4 up to scale projection matrix, that can be factorized into the calibration parameters R, $\mathbf{t}$, and K with standard methods [7]. As P has 11 free parameters, and each point $\mathbf{X}^{(i)}$ establishes 2 linearly independent constraints, a minimum of 6 triplets are required to compute a calibration. Therefore we need a set of 12 pairwise correspondences, one more than with our proposed 11 correspondences algorithm.

If we compare our algorithm against the 6 Point, with both using their minimum number of correspondences, it is expected that the latter is more stable, since its additional constraint helps to average the noise. In order to make a



(a) Translation error (b) Rotation error (c) Focal length error

Fig. 5. Simulation distribution errors for different levels of noise. The inner box markers are median values, the box limits are the 25th and 75th percentiles, and the dotted lines limits are the minimum and maximum values. **Red:** 6 point; **Blue:** $5\frac{1}{2}$ point; **Green:** 11 correspondences.

fair comparison of stability, we also test a modified version of the 6 point algorithm, that we designate by $5\frac{1}{2}$ point algorithm. The modification consists in discarding one of the 12 linear equations on the matrix P, making it minimally constrained. The goal of this algorithm is only to extract meaningful insights in the simulation environment, and has no practical purpose.

6.2 Synthetic Validation

A simulation environment was built to evaluate the stability of our algorithm, in comparison with the 6 point and the $5\frac{1}{2}$ point algorithms. In each trial we generate 3 pinhole cameras in random poses: two of them are calibrated network nodes (C_A and C_B) and the other one is a new camera that we want to calibrate (C). Then we generate two sets of 3D points: for our algorithm we generate a set of 11 points, such that C_A only detects 7 of them, C_B detects the remaining 4, and C detetects all of them; for the 6 point algorithm we generate 6 points that are all detected by the 3 cameras. By doing this we simulate the conditions for which each algorithm uses its minimum number of correspondences.

The error distribution in terms of translation, rotation and focal length for 100 trials is displayed in Fig. 5, for different levels of noise. Since we have access to the groudtruth calibration parameters R_{GT}, $\mathbf{t}_{GT}$ and K_{GT}, the rotation error is measured as the euler angle of the residual rotation matrix $\mathsf{R}^\mathsf{T}\mathsf{R}_{GT}$, the translation error as $\frac{||\mathbf{t}-\mathbf{t}_{GT}||_2}{||\mathbf{t}_{GT}||_2}$, and the focal length error as $|f_1 - f_{GT}|$, considering that $diag(\mathsf{K}_{GT}) = \begin{pmatrix} af_{GT} & a^{-1}f_{GT} & 1 \end{pmatrix}^\mathsf{T}$.

We can observe that, as expected, our algorithm is clearly more sensitive to noise than the 6 point algorithm, and has a similar stability with the $5\frac{1}{2}$ point algorithm. Although both $5\frac{1}{2}$ point and the 11 correspondences algorithms use the same number of constraints, the former makes more assumptions about the data, since the correspondences must be triple. In theory, this more structured input also favors the stability of the $5\frac{1}{2}$ point algorithm, however, there is no empirical evidence from this simulation that our algorithm has a noticeably inferior performance.

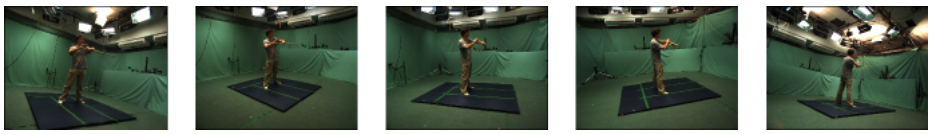


Fig. 6. Selected images from one frame of the *stick* dataset, from the 4d repository [1].

6.3 Validation with Real Data

As we have seen in simulation, the 6 point algorithm seems to be more stable than our 11 correspondences approach, and when we compare it using the same number of linear equations ($5\frac{1}{2}$ point), we obain equiparable results. However, in a real camera network, situations arise in which the overlap between cameras is not enough for the 6 point algorithm to find enough triple correspondences. In these cases our algorithm is the only available solution.

To demonstrate this we use a dataset from the 4d repository [1], containing synchronized images of a room. We selected 5 images from one particular frame (Fig. 6), and used SIFT features [10] to establish point correspondences between the images, and tested their geometric consistency against the groudtruth calibration to select the inlier correspondences. We consider as inliers, for the 11 correspondences algorithm, the pairwise correspondences with epipolar error inferior to 1.5 pixels, and for the 6 point algorithm triplets with reprojection error inferior to 1.5 pixels. As this is an experiment for validation purposes only, we conveniently have access to groundtruth values to exclude outlier measurements. In a practical situation, outliers can be discarded with a robust estimator, such as RANSAC [5].

We tested both algorithms on all 30 possible configurations of 3 views, and present median results for the errors in translation, rotation, and focal length in table 1. In many cases both algorithms fail to find a reasonable solution, and we consider them as failures. They can happen in 2 situations: when there are not enough correspondences to compute a solution (less than 11 pairwise correspondences, or less than 6 triplets); and when the algorithms output unreasonable results (we consider them unreasonable for errors in focal length greater than 100%). The failure events are discarded for computing the median results.

It is clear that in this dataset there are many configurations in which it is impossible to use the 6 point algorithm, as its percentage of failures is considerably higher. In Fig. 7 we depict an example of a configuration in which this happens. In this case there are many consistent pairwise correspondences (blue and green), however, there is only a single consistent triplet (red), making it impossible to use the 6 point algorithm. The fact that there is a wide baseline between the 3 images explains why it is so difficult to find common correspondences across the 3 images, but like in this case, there are many camera networks in which, by construction, there are no nodes with short baselines.

Furthermore, considering only the specific configurations that have enough triplets to use the 6 point algorithm, our solution is considerabily more stable

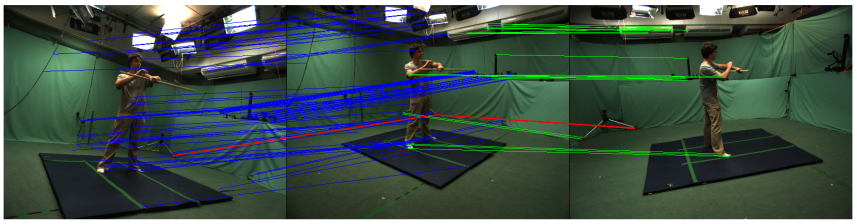


Fig. 7. In this configuration, the left and right images are the calibrated nodes $\mathbf{C}_A$ and $\mathbf{C}_B$, and the center image is the uncalibrated camera $\mathbf{C}$. In blue and green are the pairwise inlier correspondences, and in red the single triple inlier correspondence.

	R (deg)	t (mm)	f (pixels)	failures (%)
6 point	2.67	94.32	19.69	80
11 point	1.90	199.60	46.80	33
11 point*	0.54	58.49	8.58	0

Table 1. Median errors for 30 different calibration attempts, using both the 6 point and our algorithm (11 point). In the third column (signaled by *) we only considered the configurations that resulted in success for the 6 point algorithm.

(third line of table 1). The reason for this is the much higher number of consistent correspondences that are available to our algorithm, which provide more constraints. Note that in the simulation experiment presented previously, it occurs the opposite situation, i. e., since we use the minimum number of correspondences for both methods, the 6 point algorithm uses more constraints than our algorithm.

7 Conclusions

We presented a new algorithm to fully calibrate a camera, given the minimal number of 11 pairwise correspondences with other 2 calibrated cameras. We demonstrated experimentally that in many configurations of conventional camera networks, the absence of triple correspondences makes our algorithm the only approach that can effectivelly compute a solution. Note that we only presented a closed form minimal solution, and that in a real scenario its performance can be significantly improved by robust estimators such as RANSAC and non-linear refinement routines (bundle adjustment). This will constitute future work.

8 Acknowledgements

The authors acknowledge the Portuguese Science Foundation (FCT) that generously funded this work through grants PTDC/EIA-CCO/109120/2008 and SFRH/BD/72323/2010. The authors also acknowledge the French National Research Agency (ANR) and the project MORPHO.

References

1. 4d repository. http://4drepository.inrialpes.fr/pages/home
2. Barreto, J., Daniilidis, K.: Wide area multiple camera calibration and estimation of radial distortion. In: OMNIVIS'2004 - Int. Workshop in Omnidirectional vision, camera networks, and non-conventional cameras (2004)
3. Barreto, J.P.: General central projection systems: Modeling, calibration and visual servoing. Ph.D. thesis, PhD Thesis, University of Coimbra, Coimbra, Portugal (2004)
4. Courchay, J., Dalalyan, A., Keriven, R., Sturm, P.: A global camera network calibration method with linear programming. In: Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (2010)
5. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), 381–395 (1981)
6. Hartley, R., Sturm, P.: Triangulation. Computer Vision and Image Understanding (1997)
7. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge Academic Press (2003)
8. Josephson, K., Byrod, M., Kahl, F., Åström, K.: Image-based localization using hybrid feature correspondences. In: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on. pp. 1 –8 (2007)
9. Kim, J., Hodong, L., Hartley, R.: Motion Estimation for Nonoverlapping Multicamera Rigs: Linear Algebraic and Linf Geometric Solutions. IEEE Trans. in Pattern Analysis and Machine Intelligence 32(6), 1044–1058 (2010)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110 (November 2004)
11. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.: An invitation to 3-D vision: from images to geometric models. Springer (2004)
12. Pless, R.: Using many cameras as one. In: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on (2003)
13. Pottmann, H., Wallner, J.: Computational line geometry. Springer Verlag, Berlin, 1 edn. (2001)
14. Shen, E., Hornsey, R.: Multi-camera network calibration with a non-planar target. Sensors Journal, IEEE 11(10), 2356 –2364 (2011)
15. Starck, J., Hilton, A.: Surface capture for performance-based animation. Computer Graphics and Applications, IEEE 27(3), 21 –31 (2007)
16. Stewénius, H., Nistér, D., Oskarsson, M., Åström, K.: Solutions to minimal generalized relative pose problems. In: Workshop on Omnidirectional Vision. Beijing China (2005)
17. Sturm, P., Triggs, B.: A factorization based algorithm for multi image projective structure and motion. In: European Conference in Computer Vision (1996)
18. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multicamera self-calibration for virtual environments. Presence: Teleoper. Virtual Environ. 14(4), 407–422 (2005)
19. Zaharescu, A., Horaud, R., Ronfard, R., Lefort, L.: Multiple camera calibration using robust perspective factorization. In: 3D Data Processing, Visualization, and Transmission, Third International Symposium on. pp. 504 –511 (2006)
20. Zhao, Z., Liu, Y.: Practical multi-camera calibration algorithm with 1d objects for virtual environments. In: Multimedia and Expo, 2008 IEEE International Conference on. pp. 1197 –1200 (2008)