A MATHEMATICAL THEORY OF CAMERA SELF-CALIBRATION

by

Yi Ma, Jana Kosecka and Shankar Sastry

Memorandum No. UCB/ERL M98/64

20 October 1998

COVER

A MATHEMATICAL THEORY OF CAMERA SELF-CALIBRATION

by

Yi Ma, Jana Kosecka and Shankar Sastry

Memorandum No. UCB/ERL M98/64

20 October 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A Mathematical Theory of Camera Self-Calibration *

Yi Ma Jana Košecká Shankar Sastry

Electronics Research Laboratory
University of California at Berkeley
Berkeley, CA 94720-1774
{mayi, janka, sastry}@eecs.berkeley.edu

July 20, 1998

Abstract

In this paper, a mathematical theory of camera self-calibration is developed from a differential geometry viewpoint and no projective geometry is assumed. The problem of camera self-calibration is shown to be equivalent to the problem of recovering an unknown (Riemannian) metric of an appropriate space. An intrinsic geometric interpretation is thus revealed for camera "intrinsic" parameters: the intrinsic parameter space can be characterized as the quotient space $SL(3)/SO(3)$. Complete lists of geometric invariants associated to an uncalibrated camera are given. The (dual) absolute conic is shown to be a special (co)invariant generated by the lists. The self-calibration problem is then studied in both discrete and differential settings. In the discrete case, the Kruppa equation is derived from a projective geometry free approach. In the differential case, it is shown that the intrinsic parameter space is reduced to the space of singular values of the intrinsic parameter matrix. Self-calibrations associated to different camera motions are analyzed and their relations with the Kruppa equation are clearly revealed. In particular, necessary and sufficient conditions for a unique calibration are given in the case of pure rotation. Analysis of algorithms associated with these theories will be presented in a sequal to this paper.

Key words: camera self-calibration, epipolar constraint, fundamental matrix, the Kruppa equation, Euclidean invariants, Lie groups and Lie algebras.

Introduction

The problem of camera self-calibration refers to the problem of obtaining intrinsic parameters of a camera using only information from image measurements, without any *a priori* knowledge about the motion between frames and the structure of the observed scene. The general calibration problem is motivated by a variety of applications using vision as a sensor which requires the knowledge of a full Euclidean structure of the environment, which is possible only when the intrinsic parameters of the camera are known.

Both theoretical studies as well as practical algorithms have recently received an increased interest in the computer vision community. The appeal of a successful solution to the camera

*Computer vision research report. This work is supported by ARO under the MURI grant DAAH04-96-1-0341.

self-calibration problem, lies in the elimination of the need for an external calibration object [21] as well as the possibility of on-line calibration of time-varying internal parameters of the camera. The latter feature is of importance for active vision systems.

The original problem of determining whether the image measurements "only" are sufficient for obtaining the information about intrinsic parameters of the camera has been answered in the computer vision context by Faugeras and Maybank in [18]. The approach and solution utilized invariant properties of the image of the absolute conic. Since the absolute conic is invariant under Euclidean transformations (i.e. its parameterization is independent of the position of the camera) and depends only on the camera intrinsic parameters. The recovery of the image of the absolute conic is then equivalent to the recovery of the intrinsic calibration matrix. The constraints on the absolute conic captured by the epipolar transformation are expressed by the so called Kruppa equation.

One class of approaches to self-calibration directly utilizes the Kruppa equations, which provided quadratic constraints in conics parameters. Each epipolar constraint provided two such equations, requiring the total of three frames, for solution of all the unknowns. Proposed solution to the Kruppa equations using homotopy continuation was quite computationally expensive and required a good accuracy of the measurements [18]. An alternative iterative scheme was proposed by [10]. Another class of approaches for the intrinsic parameters instead of directly using Kruppa's equations, solves for the entire projection matrices which are compatible with the structure of the scene [8].

In spite of the fact that the basic formulation of the appropriate constraints is in place and there are many success stories [23] which apply the proposed algorithms mostly for 3D-reconstruction problems, to our knowledge, there is no complete analysis of the necessary and sufficient conditions for unique solution of the calibration problem. This often leads to situations where the algorithms are applied in ill-conditioned settings or where a unique solution is not obtainable.

The derivation of the Kruppa equations was mainly developed in a projective geometry framework and its understanding required good intuition of the projective geometric entities (with the exception of [6]). This derivation is quite involved and the development appears to be rather unnatural since, both the constraints captured by Kruppa's equations and the image of (dual) absolute conic are in fact directly linked to the invariants of the group of Euclidean transformations. We provide an alternative derivation of Kruppa equations, which in addition to being concise and elegant, also provides an intrinsic methods for deriving the conditions for uniqueness of the self-calibration problem. Further, it can also be derived in the same way the Kruppa equation for the case when the camera intrisic parameters are time-varying.

A clear feature of our approach is to tackle the sophisticated Kruppa equation through a study of several important special cases, such as the pure rotation case previously investigated by Hartley [8]. Not only connections between these special cases with the Kruppa equation is clearly revealed, but also more detailed results about these cases themselves are carefully presented in a unified fashion. For the first time, similarities and differences between the discrete and differential (or continuous) cases are also clearly discovered.

The outline of the paper is as following. The geometric model of an uncalibrated camera is given in Section 1. Section 2 reveals the intrinsic geometric meanings of the camera's intrinsic parameters. Complete lists of geometric invariants associated to an uncalibrated camera are given in Section 3, including an explaination for the (dual) absolute conics in terms of these invariants. Section 4 and 5 provide a geometric characterization of the space of fundamental matrices. The main theory of camera self-calibration is developed in Section 6 where conditions for unique calibration and linear schemes are studied in detail. Section 7 and Section 8 extend the results to the differential (continuous) and time-varying cases, respectively.

The theoretical contributions of our paper are as follows:

1. A differential geometric framework is proposed for the study of camera self-calibration. Geometric characterization of the space of fundamental matrices.

2. Derivation of the Kruppa equation and the time-varying Kruppa equation as invariants of the appropriate group of transformations. It is shown that the Kruppa equation becomes degenerate in the differential (continuous) case.

3. A clear proof of sufficient and necessary conditions for unique calibration for a class of trivial motions (pure translation/rotation or constant translation/rotation, etc.) and outline of the associated *linear* schemes for camera self-calibration.

4. Necessary conditions for a unique solution of the Kruppa equation and its relations with camera motion.

1 Uncalibrated camera motion and projection model

We begin with introducing the mathematical model of an uncalibrated camera in a three dimensional Euclidean space.

Consider that a camera is set in a three dimensional Euclidean space M. Then M is isometric to $\mathbb{R}^3$. This isometry equips M with a global coordinate chart and for a point q in M, it is assigned a three dimensional coordinate

$$q = (q_1, q_2, q_3)^T \in \mathbb{R}^3. \tag{1}$$

Sometimes it is convenient to represent the point $q \in M$ in homogeneous coordinates as:

$$\underline{q} = (q_1, q_2, q_3, 1)^T \in \mathbb{R}^4. \tag{2}$$

In this way, M is viewed as a submanifold embedded in $\mathbb{R}^4$. To differentiate the notation, we will use underlined symbol ($\underline{q}$ v.s. q) for the homogeneous representation. Let T_qM be the set of all vectors (in a Euclidean space, a vector is defined to be the difference between two points) in M with the starting point q (*i.e.* T_qM is the tangent space of M at q). Then any vector $u \in T_qM$ in its homogeneous representation has the form:

$$\underline{u} = (u_1, u_2, u_3, 0)^T \in \mathbb{R}^4. \tag{3}$$

So as a vector space T_qM is isomorphic to $\mathbb{R}^3$. A non-redundant representation of the same vector $u \in T_qM$ is just:

$$u = (u_1, u_2, u_3)^T \in \mathbb{R}^3. \tag{4}$$

The Euclidean metric g on M is then simply given by:

$$g_q(u, v) = u^T v, \quad \forall u, v \in T_qM, \quad \forall q \in M. \tag{5}$$

Sometime we use the pair (M, g) to emphasize that M is a manifold with a preassigned (Riemannian) metric g.

The isometry (diffeomorphism which preserves metric) group of M is the so called Euclidean group $E(3)$. The motion of the camera is usually modeled as the subgroup of $E(3)$ which preserves the orientation of the space M, *i.e.* the special Euclidean group $SE(3)$. $SE(3)$ can be represented in homogeneous coordinates as:

$$SE(3) = \left\{ \begin{pmatrix} R & p \\ 0 & 1 \end{pmatrix} \middle| \, p \in \mathbb{R}^3, R \in SO(3) \right\} \subset \mathbb{R}^{4\times 4} \tag{6}$$

where $SO(3)$ is the space of 3×3 rotation matrices (unitary matrices with determinant $+1$). We know the isotropy group of M at a point q is the orthogonal group $O(3)$. $SO(3)$ is just the subgroup of $O(3)$ which is the connected component of the identity I. Given an element $h \in SE(3)$ and a point $q \in M$, h maps the coordinates of q to new ones. In the homogeneous representation, these new coordinates are given by $h\underline{q}$.

A curve $h(t) \in SE(3), t \in \mathbb{F}$ is used to represent the translation and rotation of the camera coordinate frame F_t at time t relative to its initial coordinate frame F_{t_0} at time t_0. By abuse of notation, the group element $h(t)$ serves both as a specification of the configuration of the camera and as a transformation taking the coordinates of a point in the F_{t_0} frame to that of the same point in the F_t frame. Clearly, $h(t)$ is uniquely determined by its rotational part $R(t) \in SO(3)$ and translational part $p(t) \in \mathbb{R}^3$. Sometimes we denote $h(t)$ by $(R(t), p(t))$ as a shorthand. Let $\underline{q}(t) = (q(t)^T, 1)^T \in \mathbb{R}^4$ be the homogeneous coordinates of a point $q \in M$ with respect to the camera coordinate frame at time $t \in \mathbb{R}$. Then the coordinate transformation is given by:

$$\underline{q}(t) = h(t)\underline{q}(t_0). \tag{7}$$

In three dimensional representation, the above coordinate transformation is simply equivalent to:

$$q(t) = R(t)q(t_0) + p(t). \tag{8}$$

We assume that the camera coordinate frame is chosen such that the optical center of the camera, denoted by o, is the same as the origin of the frame. Define the image of a point $q \in M$ to be the vector $\mathbf{x} \in T_oM$ which is determined by o and the intersection of the half ray $\{o + \lambda \cdot u \mid u = q - o, \lambda \in \mathbb{R}^+\}$ with a pre-specified image surface (for example, a unit sphere or a plane). Then both the spherical projection and perspective projection fit into this type of imaging model. For a point $q \in M$ with coordinates $\underline{q} = (q_1, q_2, q_3, 1)^T \in \mathbb{R}^4$, since the optical center o always have the coordinates $(0, 0, 0, 1)^T \in \mathbb{R}^4$, the vector $u = q - o \in T_oM$ is then given by $u = (q_1, q_2, q_3)^T \in \mathbb{R}^3$. Define the projection matrix $P \in \mathbb{R}^{3\times 4}$:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{9}$$

Then the projection matrix P gives a map from the space M to T_oM:

$$\begin{aligned} P : M &\rightarrow T_oM & (10) \\ q &\mapsto u = P\underline{q}. & (11) \end{aligned}$$

According to the definition, the image $\mathbf{x}$ of the point q differs from the vector $u = P\underline{q}$ by an arbitrary positive scale, which depends on the pre-specified image surface. In general, the relation between $q \in M$ and its image $\mathbf{x}$ is therefore given by:

$$\lambda \mathbf{x} = P\underline{q} \tag{12}$$

for some $\lambda \in \mathbb{R}^+$. The unknown scalar λ encodes the depth information of q and we call λ the **scale** of the point q with respect to the image $\mathbf{x}$. The equation (12) characterizes the mathematical model of an ideal calibrated camera. For a study of calibrated camera, one may refer to Ma, Kosecka and Sastry [12, 11].

In this paper, we are going to study uncalibrated camera. By an uncalibrated camera, we mean that the image received by the camera is distorted by an unknown linear transformation. This linear transformation is usually assumed to be invertible. Mathematically, this linear transformation is an isomorphism ϕ of the vector space $T_o M$:

$$\begin{aligned} \phi : T_o M &\rightarrow T_o M \\ u &\mapsto Au, \end{aligned}$$

where $A \in \mathbb{R}^{3\times 3}$ is an invertible matrix representing the linear map ϕ. The actually received image $\mathbf{x}$ is then determined by the intersection of the image surface and the ray $\{o + \lambda \cdot u\}$ where

$$u = AP\underline{q}.$$

Without loss of generality, we may assume that A has determinant 1, *i.e.* A is an element in $SL(3)$ (the Lie group consisting of all invertible 3×3 real matrices with determinant 1). For a representative image $\mathbf{x} \in \mathbb{R}^{\text{˘}}$ of q, we have the relation:

$$\lambda \mathbf{x} = AP\underline{q} \tag{13}$$

for some scale $\lambda \in \mathbb{R}^+$. The equation (13) then characterizes the mathematical model of an uncalibrated camera.

Comments 1 *In the computer vision literature, it is assumed that the matrix A is of the following form:*

$$A = \begin{pmatrix} s_x & s_\theta & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{14}$$

The parameters in the matrix A are called "intrinsic parameters" associated to the uncalibrated camera. Note that such an A is not in $SL(3)$ and does not form a group either. We will soon see that, this choice is equivalent to ours (in some sense).

If we know the linear transformation ϕ, *i.e.* the matrix A, then the problems associated to an uncalibrated camera can be reduced to those of a calibrated camera. So one important problem we need to study about an uncalibrated camera is: knowing the image $\mathbf{x}$, to what extent one may recover the unknown linear transformation ϕ or the matrix A, and how. This is the so-called **camera self-calibration problem**.

2 Intrinsic geometric interpretation for camera intrinsic parameters

Before trying to solve the camera self-calibration problem, we first need to know some geometric properties of an uncalibrated camera. In this section, differential geometric properties of an uncalibrated camera will be explicitly revealed: the study of an uncalibrated camera is equivalent to the study of a calibrated camera in a (Euclidean) space with an unknown metric. Further, the problem of recovering the linear transformation matrix A is mathematically equivalent to recovering this unknown metric. Consequently, the camera intrinsic parameters given in (14) can be intrinsically characterized as the space $SL(3)/SO(3)$. Some elementary Riemannian geometry notion will be used here (to maintain the generality of the geometry). For good references of Riemannian geometry, one may refer to [2, 9, 20].

Let M' be another Euclidean space (isometric to $\mathbb{R}^3$) with a Euclidean structure induced as follows. Consider a map from M' to M:

$$\begin{aligned} \psi : M' &\rightarrow M \\ q' &\mapsto q = A^{-1}q' \end{aligned}$$

where q' and q are 3 dimensional coordinates of the points $q' \in M'$ and $\psi(q') \in M$ respectively.The differential of the map ψ at a point $q' \in M'$ is just the push-forward map:

$$\begin{aligned} \psi_* : T_{q'}M' &\rightarrow T_{\psi(q')}M \\ u &\mapsto A^{-1}u. \end{aligned}$$

Then the metric g on M induces a metric on M' as the pull-back $\psi^*(g)$, which is explicitly given by:

$$\psi^*(g)_{q'}(u,v) = g_{\psi(q')}(\psi_*(u), \psi_*(v)) = u^T A^{-T} A^{-1} v, \quad \forall u, v \in T_{q'}M', \quad \forall q' \in M'. \tag{15}$$

We define the symmetric matrix $S \in \mathbb{R}^{3\times3}$ associated to the matrix A as:

$$S = A^{-T}A^{-1}. \tag{16}$$

Then the metric $\psi^*(g)$ on the space M' is determined by the matrix S. Let $\mathbb{K} \subset SL(3)$ be the subgroup of $SL(3)$ which consists of all upper-triangle matrices. That is, any matrix $A \in \mathbb{K}$ has the form:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}. \tag{17}$$

Note that if A is upper-triangular, so is A^{-1}. Clearly, there is a one-to-one correspondence between $\mathbb{K}$ and the set of all upper-triangular matrices of the form given in (14); also the equation (16) gives a finite-to-one correspondence between $\mathbb{K}$ and the set of all 3×3 symmetric matrices with determinant 1. Usually, only one of the upper-triangular matrices corresponding to the same symmetric matrix is physically possible. Thus, if the matrix A of the uncalibrated camera does have the form given by (14), the camera self-calibration problem is equivalent to the problem of recovering the matrix S, *i.e.* the metric $\psi^*(g)$ of the space M'.

Now let us consider the case that the uncalibrated camera is characterized by an arbitrary matrix $A \in SL(3)$. A has the QR-decomposition:

$$A = QR, \quad Q \in \mathbb{K}, R \in SO(3). \tag{18}$$

Then $A^{-1} = R^T Q^{-1}$ and the associated symmetric matrix $S = A^{-T}A^{-1} = Q^{-T}Q^{-1}$. In general, if $A = BR$ with $A, B \in SL(3), R \in SO(3)$ and S_A and S_B are associated symmetric matrices of A and B respectively, then $S_A = S_B$. In this case, we say that matrices A and B are equivalent. The quotient space $SL(3)/SO(3)$ will be called the **intrinsic parameter space.** It gives an "intrinsic-indeed" interpretation for the camera intrinsic parameters given in (14). This will be explained in more detail in the rest of this section.

Without knowing camera motion and scene structure, the matrix $A \in SL(3)$ can only be recovered up to an equivalence class $[A] \in SL(3)/SO(3)$. To see this, suppose $B \in SL(3)$ is another matrix in the same equivalence class as A. Then $A = BR_0$ for some $R_0 \in SO(3)$. The coordinate transformation (8) yields:

$$Aq(t) = ARq(t_0) + Ap(t) \quad \Leftrightarrow \quad BR_0q(t) = BR_0R(t)R_0^T R_0q(t_0) + BR_0p(t). \tag{19}$$

Notice that the conjugation:

$$\begin{array}{rcl} \mathrm{Ad}_{R_0} : SO(3) & \to & SO(3) \\ R & \mapsto & R_0RR_0^T \end{array}$$

is a group homomorphism. Then there is no way to tell an uncalibrated camera with transformation matrix A taking the motion $(R(t), p(t))$ and observing the point $q \in M$ from another uncalibrated camera with transformation matrix B taking the motion $(R_0R(t)R_0^T, R_0p(t))$ and observing the point $R_0q \in M$. We will soon see that this property will naturally show up in the fundamental matrix (to be introduced soon) when we study epipolar constraint.

Therefore, without knowing camera motion and scene structure, the matrix A associated with an uncalibrated camera can only be recovered up to an equivalence class $[A]$ in the space $SL(3)/SO(3)$. The subgroup $\mathbb{K}$ of all upper-triangular matrices in $SL(3)$ is one representation of such a space, as is the space of 3×3 symmetric matrices with determinant 1. Thus, $SL(3)/SO(3)$ does provide an intrinsic geometric interpretation for the unknown camera parameters. In general, the problem of camera self-calibration is then equivalent to the problem of recovering the symmetric matrix $S = A^{-T}A^{-1}$, *i.e.* the metric of the space M', from which the upper-triangle representation of the intrinsic parameters can be easily obtained.

The space M' essentially is also a Euclidean space. But with respect to the chosen coordinate charts, the metric form $\psi^*(g)$ is unknown. From (8), the coordinate transformation in the space M' is given by:

$$Aq(t) = AR(t)q(t_0) + Ap(t) \quad \Leftrightarrow \quad q'(t) = AR(t)A^{-1}q'(t_0) + p'(t) \tag{20}$$

where $q' = Aq$ and $p' = Ap$. In homogeneous coordinates, the transformation group on M' is given by:

$$G = \left\{ \left(\begin{array}{cc} ARA^{-1} & p' \\ 0 & 1 \end{array} \right) \middle| \, p' \in \mathbb{R}^3, R \in SO(3) \right\} \subset \mathbb{R}^{4\times 4} \tag{21}$$

It is direct to check that the metric $\psi^*(g)$ is invariant under the action of G. Thus G is a subgroup of the isometry group[1] of M'. If the motion of a (calibrated) camera in the space M' is given by $h'(t) \in G, t \in \mathbb{R}$, the homogeneous coordinates of a point $q' \in M'$ satisfy:

$$\underline{q}'(t) = h'(t)\underline{q}'(t_0). \tag{22}$$

From the previous section, the image of the point q' with respect to a calibrated camera is given by:

$$\lambda \mathbf{x} = P\underline{q}'. \tag{23}$$

It is then direct to check that this image is the same as the image of $q = \psi(q') \in M$ with respect to the uncalibrated camera, *i.e.* we have:

$$\lambda \mathbf{x} = AP\underline{q}. \tag{24}$$

From this property, the problem of camera self-calibration is indeed equivalent to the problem of recovering the unknown (Riemannian) metric of a proper space assuming a calibrated camera. Since isometric transformation (group) of the space M' preserves its metric, invariants preserved by such transformation are therefore keys to recover the unknown metric. The next section is about to give a complete account of these invariants.

3 Geometric invariants associated to uncalibrated camera

Although the explicit form of the metric of the space M' is unknown, we know M' is isomorphic to the Euclidean space M through the isomorphism $\psi : M' \to M$. Thus the invariants of M' under its isometry group G are equivalent to the invariants of M under the Euclidean group.

The complete list of Euclidean invariants is given by the following theorem:

Theorem 1 (Euclidean invariants) *For a n dimensional vector space V, a complete list of basic invariants of the group $SO(n)$ consists of (1) the inner product $g(u,v) = u^T v$ of two vectors $u, v \in V$ and (2) the determinant* $\det[u^1, \ldots, u^n]$ *of n vectors $u^1, \ldots, u^n \in V$.*

See Weyl [22] for a proof of this theorem and see Ma, Kosecka and Sastry [11] for a more detailed discussion about applications of this theorem in structure reconstruction. From the theorem, the set of all Euclidean invariants is the $\mathbb{R}$-algebra generated by these two types of basic invariants. In the uncalibrated camera case, applying this theorem to the three dimensional space M', we have:

Corollary 1 (Invariants of uncalibrated camera) *For the space M', a complete list of basic invariants of the isometry group G consists of (1) the inner product $\psi^*(g)(u,v) = u^T A^{-T} A^{-1} v$ of two vectors $u, v \in TM'$ and (2) the determinant* $\det[A^{-1}u^1, A^{-1}u^2, A^{-1}u^3]$ *of three vectors $u^1, u^2, u^3 \in TM'$.*

Then the set of invariants associated to an uncalibrated camera is the $\mathbb{R}$-algebra generated by these two types of basic invariants. Since

$$\det[A^{-1}u^1, A^{-1}u^2, A^{-1}u^2] \quad = \quad \det(A^{-1}) \cdot \det[u^1, u^2, u^3],$$

[1] Isometry group of a space M is all transformations which preserve metric (or distance).

it follows that the invariant $\det[A^{-1}u^1, A^{-1}u^2, A^{-1}u^3]$ is independent of the matrix A. Therefore the determinant type invariant is useless for recovering the unknown matrix A and only the inner product type invariant can be helpful.

For any n-dimensional vector space V, its **dual space** $V^\vee$ is defined to be the vector space of all linear functions on V. An element in $V^\vee$ is called a **covector**. If $e^i, i = 1, \ldots, n$ are a basis for V, then the set of linear functions $e_j, j = 1, \ldots, n$ defined as:

$$e_j(e^i) = \delta_{ij} \tag{25}$$

form a (dual) basis for the dual space $V^\vee$. The **pairing** between V and its dual $V^\vee$ is defined to be the bilinear map:

$$\begin{aligned} <\cdot,\cdot>: V^\vee \times V &\to \mathbb{R} && (26)\\ (u, v) &\mapsto u(v). && (27) \end{aligned}$$

If we use the coordinate vector $u = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n$ to represent a covector $u = \sum_{j=1}^n \alpha_j e_j \in V^\vee, \alpha_j \in \mathbb{R}$, and similarly, $v = (\beta_1, \ldots, \beta_n)^T \in \mathbb{R}^n$ to represent $v = \sum_{i=1}^n \beta_i e^i \in V, \beta_i \in \mathbb{R}$ (note that we use column vector convention for both vectors and covectors in this paper), then with respect to the chosen bases the pairing is given by:

$$< u, v >= u^T v.$$

For a linear transformation $f : V \to V$, its dual is defined to be the linear transformation $f^\vee : V^\vee \to V^\vee$ which preserves the pairing:

$$< u, f(v) >=< f^\vee(u), v >, \quad \forall u \in V^\vee, v \in V. \tag{28}$$

Let $A \in \mathbb{R}^{n\times n}$ be the matrix representing f with respect to the basis $e^i, i = 1, \ldots, n$. Since:

$$< u, f(v) >= u^T A v = (A^T u)^T v, \tag{29}$$

it follows that the dual $f^\vee$ is represented by A^T with respect to the (dual) basis $e_j, j = 1, \ldots, n$.

The invariants given in Corollary 1 are invariants of the vector space $TM' \cong \mathbb{R}^3$ under the action of the isotropy subgroup $ASO(3)A^{-1}$ of G on M' (here we identify an element in $ASO(3)A^{-1}$ with its differential map). As we know from above, this group action induces a (dual) action on the dual space of TM', denoted by T^*M'. This dual action can then be represented by the group $A^{-T}SO(3)A^T$ since

$$(ARA^{-1})^T = A^{-T}R^TA^T \quad \in A^{-T}SO(3)A^T$$

for all $R \in SO(3)$. We call invariants associated to the dual group action (on the covectors) as **coinvariants**. As we will soon see, the Kruppa equation can be viewed as such coinvariants. Consequently we have:

Corollary 2 (Coinvariants of uncalibrated camera) *For the space M', a complete list of basic coinvariants of the isometry group G consists of (1) the induced inner product $\xi^T AA^T \eta$ of two covectors $\xi, \eta \in T^*M'$ and (2) the determinant $\det[\xi_1, \xi_2, \xi_3]$ of three covectors $\xi_1, \xi_2, \xi_3 \in T^*M'$.*

Note that in the above we use the convention that vectors are always enumerated by superscript and covectors by subscript. One may also refer to Weyl [22] or Goodman and Wallach [5] for a detailed study of polynomial invariants of classical groups – Corollary 1 and 2 can then be deduced from the First Fundamental Theorem of groups $G \subset GL(V)$ preserving a non-degenerate (symmetric) form (see [5]). Note that the induced inner product on T^*M' is given by the symmetric matrix $S^{-1} = AA^T$, the inverse of $S = A^{-T}A^{-1}$. As we will soon see, coinvariants naturally show up in the recovery of S^{-1} from fundamental matrices.

Next we want to show that the **absolute conic** (or the dual absolute conic) is actually a special invariant generated by inner product type invariants (or coinvariants). In the projective geometry approach, the absolute conic plays an important role in camera self-calibration.

In order to give a rigorous definition of the absolute conic, we needs to introduce the space $\mathbb{CP}^3$, the three dimensional complex projective space[2]. Let $\underline{q} = (q_1, q_2, q_3, q_4)^T \in \mathbb{C}^4$ be the homogeneous representation of a point q in $\mathbb{CP}^3$. Then the absolute conic, denoted by Ω, is defined to be the set of points in $\mathbb{CP}^3$ satisfying:

$$q_1^2 + q_2^2 + q_3^2 = 0, \quad q_4 = 0 \tag{30}$$

Note that this set is invariant under the complex Euclidean group:

$$E(3, \mathbb{C}) = \left\{ \left(\begin{array}{cc} R & p \\ 0 & 1 \end{array} \right) \middle| \, p \in \mathbb{C}^3, R \in U(3) \right\} \subset \mathbb{C}^{4\times 4} \tag{31}$$

where $U(3)$ is the space of all (complex) 3×3 unitary matrices. The special Euclidean group $SE(3)$ is just a subgroup of $E(3, \mathbb{C})$ hence the absolute conic is invariant under $SE(3)$ as well.

For any $\underline{q} = (q_1, q_2, q_3, q_4)^T \in \Omega$, suppose

$$q_j = u_j + iv_j, \quad u_j, v_j \in \mathbb{R}, \quad j = 1, \ldots, 4 \tag{32}$$

where $i = \sqrt{-1}$. Since $u_4 = v_4 = 0$, we obtain a pair of vectors $\underline{u} = (u_1, u_2, u_2, 0)^T$ and $\underline{v} = (v_1, v_2, v_3, 0)^T$ of the 3 dimensional (real) Euclidean space M (in homogeneous representation). From (30), these two vectors satisfy:

$$u^T u = v^T v, \quad u^T v = 0 \tag{33}$$

On the other hand, any pair of vectors $u, v \in TM$ which satisfy the above conditions (*i.e.* u and v are orthogonal to each other and have the same length) define a point on the absolute conic Ω. Therefore, the absolute conic Ω is the same as the set of all pairs of such vectors. Since all the inner product type quantities in (33) are invariant under the Euclidean group $SE(3)$, the absolute conic Ω is simply generated by these basic invariants.

In the uncalibrated camera case, if we let $S = A^{-T}A^{-1}$ and $\underline{q}' = (q_1, q_2, q_3, q_4)^T \in \mathbb{C}^4$, the corresponding absolute conic (30) is then given by:

$$(q_1, q_2, q_3)S(q_1, q_2, q_3)^T = 0, \quad q_4 = 0. \tag{34}$$

Therefore, the camera self-calibration problem is also equivalent to the problem of recovering this absolute conic (for example see Maybank [17]). It is direct to check that this absolute conic is

[2] $\mathbb{CP}^3$ is the space of all one dimensional (complex) subspaces in $\mathbb{C}^4$, *i.e.* the quotient space $\mathbb{C}^4 / \sim$ where the equivalence relation $\sim$ is: $(z_1, z_2, z_3, z_4)^T \sim (z \cdot z_1, z \cdot z_2, z \cdot z_3, z \cdot z_4)^T$ for all $z \neq 0$.

generated by basic invariants given in Corollary 1. Define the dual absolute conic $\Omega^\vee$ to be the set of points in $\mathbb{CP}^3$ satisfying:

$$(q_1, q_2, q_3)S^{-1}(q_1, q_2, q_3)^T = 0, \quad q_4 = 0. \tag{35}$$

Similarly, one can show that it is generated by the inner product type coinvariants given in Corollary 2.

4 Epipolar geometry

Before we can apply the invariant theory given in the previous section to the problem of camera self-calibration, we first need to know what quantities we can directly obtain from images and what type of geometric entities they are. This section and Section 5 are going to show that fundamental matrices which can be estimated from the epipolar constraint are in fact covectors – Section 6 then shows that their associated coinvariants directly give the Kruppa equation.

The epipolar (or Longuet-Higgins) constraint plays an important role in the study of geometry of calibrated camera. In this section, we study its uncalibrated version. First, we introduce some notation. For a three dimensional vector $p = (p_1, p_2, p_3)^T \in \mathbb{R}^3$, we define the skew-symmetric matrix $\hat{p} \in \mathbb{R}^{3\times 3}$ associated to p as:

$$\hat{p} = \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix}. \tag{36}$$

Then for another vector $l \in \mathbb{R}^3$, the cross-product $p \times l$ is equal to $\hat{p}l$.

From (20), for a point $q' \in M'$ we have

$$\begin{aligned} & q'(t) = AR(t)A^{-1}q'(t_0) + p'(t) \quad \Rightarrow \quad p'(t) \times q'(t) = p'(t) \times AR(t)A^{-1}q'(t_0) \\ \Rightarrow \quad & q'(t_0)^T A^{-T} R(t)^T A^T \widehat{p'(t)} q'(t) = 0. \end{aligned} \tag{37}$$

Let $\mathbf{x}_1 \in \mathbb{R}^3$ and $\mathbf{x}_2 \in \mathbb{R}^3$ be images of q' at time t_0 and t respectively, *i.e.* there exist $\lambda_1, \lambda_2 \in \mathbb{R}^+$ such that $\lambda_1 \mathbf{x}_1 = q'(t_0)$ and $\lambda_2 \mathbf{x}_2 = q'(t)$. To simplify the notation, we will drop the time dependence from the motion $(AR(t)A^{-1}, p'(t))$ and simply denote it by (ARA^{-1}, p'). Then (37) yields:

$$\mathbf{x}_1^T A^{-T} R^T A^T \widehat{p'} \mathbf{x}_2 = 0. \tag{38}$$

Note that in the above equation the matrix

$$F_1 = A^{-T} R^T A^T \widehat{p'} \quad \in \mathbb{R}^{3\times 3} \tag{39}$$

is of particular interest – it contains useful information about camera intrinsic parameters as well as the motion of camera.

Recall that the motion (ARA^{-1}, p') in the space M' is equivalent to the motion (R, p) in the space M, with $p = A^{-1}p'$. Also from (20), we have

$$\begin{aligned} & A^{-1}q'(t) = R(t)A^{-1}q'(t_0) + p(t) \quad \Rightarrow \quad p(t) \times A^{-1}q'(t) = p(t) \times R(t)A^{-1}q'(t_0) \\ \Rightarrow \quad & q'(t_0)^T A^{-T} R(t)^T \widehat{p(t)} A^{-1} q'(t) = 0 \end{aligned} \tag{40}$$

We then have a second form for the constraint given in (38):

$$\mathbf{x}_1^T A^{-T} R^T \hat{p} A^{-1} \mathbf{x}_2 = 0. \tag{41}$$

The matrix

$$F_2 = A^{-T} R^T \hat{p} A^{-1} \quad \in \mathbb{R}^{3\times 3} \tag{42}$$

is called the **fundamental matrix** in the computer vision literature. In fact, the two constraints (38) and (41) are equivalent and they are both called the **epipolar constraint**. We prove this by showing that the two matrices F_1 and F_2 are actually equal.

Lemma 1 *If $p \in \mathbb{R}^3$ and $A \in SL(3)$, then $A^T \hat{p} A = \widehat{A^{-1}p}$.*

Proof: Since both $A^T\widehat{(\cdot)}A$ and $\widehat{A^{-1}(\cdot)}$ are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^{3\times 3}$, using the fact that $\det(A) = 1$, one may directly verify that these two linear maps are equal on the bases: $(1,0,0)^T, (0,1,0)^T$ or $(0,0,1)^T$. ■

According to the lemma, we have:

$$F_2 = A^{-T} R^T \hat{p} A^{-1} = A^{-T} R^T A^T A^{-T} \hat{p} A^{-1} = A^{-T} R^T A^T \widehat{p'} = F_1. \tag{43}$$

We then can denote F_1 and F_2 by the same name F. Define the space of fundamental matrices associated to $A \in SL(3)$ as:

$$\mathcal{F} = \{A^{-T} R^T \hat{p} A^{-1} \mid R \in SO(3), p \in \mathbb{R}^3\}. \tag{44}$$

The space $\mathcal{F}$ is also called **fundamental space**.

In the preceding section, we have known that if two matrices A and B are in the same equivalence class of $SL(3)/SO(3)$, we are not able to tell them apart only from images. We may assume $B = AR_0$ for some $R_0 \in SO(3)$. Then with the same camera motion (R, p), the fundamental matrix associated with B is:

$$B^{-T} R^T \hat{p} B^{-1} = A^{-T} R_0 R^T \hat{p} R_0^T A^{-1} = A^{-T} (R_0 R^T R_0^T) \widehat{R_0 p} A^{-1}. \tag{45}$$

As we noticed, the essential matrix $R^T \hat{p}$ is simply replaced by another essential matrix $(R_0 R^T R_0^T)\widehat{R_0 p}$. Therefore, without knowing the camera motion, from only the fundamental matrix, one cannot tell camera B from camera A.

5 Geometric characterization of the space of fundamental matrices

In this section, we give a geometric characterization of the space of fundamental matrices. It will be shown that this space can be naturally identified with the cotangent bundle of the matrix Lie group $ASO(3)A^{-1}$, therefore, fundamental matrices by their nature can be viewed as covectors. This characterization is quite different from the conventional way of characterizing fundamental matrices as a degenerate matrix which represents the epipolar map between two image planes (for example see [10]), but it directly connects a fundamental matrix with its Kruppa equation, as we will soon see in Section 6.

We define a metric g on the space $\mathbb{R}^{3\times3}$ as:

$$g(B,C) = \operatorname{tr}(B^T SC), \quad \forall B, C \in \mathbb{R}^{3\times3} \tag{46}$$

where $S = A^{-T}A^{-1}$. It is direct to check that so defined g is indeed a metric. This metric may be used to identify the space $\mathbb{R}^{3\times3}$ with its dual $(\mathbb{R}^{3\times3})^\vee$ (the space of linear functions on $\mathbb{R}^{3\times3}$). In other words, under this identification, given a matrix $C \in \mathbb{R}^{3\times3}$, we may identify it as a member in the dual space $(\mathbb{R}^{3\times3})^\vee$ through:

$$\begin{aligned} f : \mathbb{R}^{3\times3} &\to (\mathbb{R}^{3\times3})^\vee \\ C &\mapsto C^\vee = g(\cdot, C). \end{aligned}$$

From the metric definition (46), $C^\vee$ can be represented in the matrix form as $C^\vee = SC$. Since S is non-degenerate, the map f is an isomorphism and it induces a metric on the dual space as follows:

$$g^\vee(B^\vee, C^\vee) = g(B,C) = \operatorname{tr}((B^\vee)^T S^{-1} C^\vee). \tag{47}$$

A tangent vector of the Lie group $ASO(3)A^{-1}$ has the form $AR^T\hat{p}A^{-1} \in \mathbb{R}^{3\times3}$ where $R \in SO(3)$ and $p \in \mathbb{R}^3$. By restricting this metric to the tangent space of $ASO(3)A^{-1}$, *i.e.* $T(ASO(3)A^{-1})$, the metric g induces a metric on the Lie group $ASO(3)A^{-1}$:

$$g(AR^T\hat{p}_1A^{-1}, AR^T\hat{p}_2A^{-1}) = g(A\hat{p}_1A^{-1}, A\hat{p}_2A^{-1}). \tag{48}$$

The equality shows that this induced metric on the Lie group $ASO(3)A^{-1}$ is left invariant.

The cotangent vector corresponding to the tangent vector $AR^T\hat{p}A^{-1} \in T(ASO(3)A^{-1})$ is given by:

$$(AR^T\hat{p}A^{-1})^\vee = SAR^T\hat{p}A^{-1} = A^{-T}R^T\hat{p}A^{-1}. \tag{49}$$

Note that the matrix $A^{-T}R^T\hat{p}A^{-1}$ is the exact form of a fundamental matrix. Therefore, the space of all fundamental matrices can be interpreted as the cotangent space of the Lie group $ASO(3)A^{-1}$, *i.e.* $T^*(ASO(3)A^{-1})$. There is an induced metric on the cotangent bundle:

$$g^\vee(A^{-T}R^T\hat{p}_1A^{-1}, A^{-T}R^T\hat{p}_2A^{-1}) = g^\vee(\widehat{p_1'}, \widehat{p_2'}) \tag{50}$$

where $p_1' = Ap_1$ and $p_2' = Ap_2$. Since a fundamental matrix can only be determined up to scale, we may consider the unit cotangent bundle $T_1^*(ASO(3)A^{-1})$. Define the space of normalized fundamental matrices to be:

$$\mathcal{F}_1 = \{A^{-T}R^T\hat{p}A^{-1} \mid R \in SO(3), p \in \mathbb{R}^3, g^\vee(\widehat{Ap}, \widehat{Ap}) = 1\}. \tag{51}$$

The space $\mathcal{F}_1$ is also called **normalized fundamental space**. The relation between the normalized fundamental space $\mathcal{F}_1$ and the unit cotangent bundle $T_1^*(ASO(3)A^{-1})$ is given by:

Theorem 2 *The unit cotangent bundle $T_1^*(ASO(3)A^{-1})$ is a double covering of the normalized fundamental space $\mathcal{F}_1$.*

The proof essentially follows from the fact that the unit tangent bundle $T_1(SO(3))$ is a double covering of the normalized essential space (see Ma, Kosecka and Sastry [12]). For a fixed matrix $A \in SL(3)$, the normalized fundamental space $\mathcal{F}_1$ is a five dimensional connected compact manifold embedded in $\mathbb{R}^{3\times3}$.

Comments 2 *Usually the eight point algorithm can still be used to estimate the fundamental matrix. However, the matrix directly obtained from solving the LLSE problem may not be exactly in the fundamental space or the normalized fundamental space.*

6 Camera self-calibration

After all the preparation in geometry, we are now ready to investigate possible schemes for recovering the unknown intrinsic parameter matrix A, or equivalently, the symmetric matrix $S = A^{-T}A^{-1}$.

6.1 The Kruppa equation

We first assume that both the rotation R and translation p are non-trivial, *i.e.* $R \neq I$ and $p \neq 0$ hence the epipolar constraint (38) is not degenerate and the fundamental matrix can be estimated. The camera self-calibration problem is then reduced to recovering the symmetric matrix S from fundamental matrices, *i.e.* recovering $S = A^{-T}A^{-1}$ from matrices of the form $F = A^{-T}R^T\hat{p}A^{-1}$. It turns out that it is easier to use the other form of the fundamental matrix $F = A^{-T}R^TA^T\widehat{p'}$ with $p' = Ap$. From the fundamental matrix the epipole vector p' can be directly computed as the null space of F. Without loss of generality, we may assume $\|p'\| = 1$. The corresponding fundamental matrix F is then called a **unit fundamental matrix** (to be separated from the normalized fundamental matrix). In this section, all vectors (by their nature) are covectors hence will be denoted with subscripts – but we always use column vector convention to represent them unless otherwise stated. Suppose the standard basis of $\mathbb{R}^3$ is:

$$e_1 = (1,0,0)^T, \quad e_2 = (0,1,0)^T, \quad e_3 = (0,0,1)^T \quad \in \mathbb{R}^3. \tag{52}$$

Now pick any rotation matrix $R_0 \in SO(3)$ such that $R_0p' = e_3$. Using Lemma 1, we have:

$$\widehat{p'} = R_0^T\hat{e}_3R_0. \tag{53}$$

Define matrix $D \in \mathbb{R}^{3\times3}$ to be

$$D = R_0FR_0^T = (R_0A)^{-T}R^T(R_0A)^T\hat{e}_3. \tag{54}$$

Then D has the form $D = (d_1, d_2, 0)$ with $d_1, d_2 \in \mathbb{R}^3$ as the first and second column vectors of D. From the definition of D we have:

$$d_1 = (R_0A)^{-T}R^T(R_0A)^Te_2, \quad d_2 = -(R_0A)^{-T}R^T(R_0A)^Te_1. \tag{55}$$

Define matrix $K = R_0A \in SL(3)$. Note that (55) is in the form of a transformation on covectors that we discussed in Section 3. According to Corollary 2, coinvariants of the group $KSO(3)K^{-1}$ (*i.e.* the invariants of the dual group $K^{-T}SO(3)K^T$) give:

$$\begin{aligned}(d_1)^TKK^Td_1 &= (e_2)^TKK^Te_2,\\ (d_2)^TKK^Td_2 &= (e_1)^TKK^Te_1,\\ (d_1)^TKK^Td_2 &= -(e_2)^TKK^Te_1.\end{aligned} \tag{56}$$

Note that $KK^T = R_0AA^TR_0^T = R_0S^{-1}R_0^T$ where as usual $S = A^{-T}A^{-1}$. If we know K^TK, the symmetric matrix S can be calculated from the chosen R_0. By defining covectors $\xi_1, \xi_2, \eta_1, \eta_2 \in \mathbb{R}^3$ as:

$$\xi_1 = R_0^T d_1, \quad \xi_2 = R_0^T d_2; \quad \eta_1 = -R_0^T e_1, \quad \eta_2 = R_0^T e_2,$$

then (56) directly gives constraints on S^{-1} as:

$$\begin{aligned} \xi_1^T S^{-1}\xi_1 &= \eta_2^T S^{-1}\eta_2, \\ \xi_2^T S^{-1}\xi_2 &= \eta_1^T S^{-1}\eta_1, \\ \xi_1^T S^{-1}\xi_2 &= \eta_1^T S^{-1}\eta_2. \end{aligned} \tag{57}$$

We thus obtain three homogeneous constraints on the matrix S^{-1}, the inverse of the matrix S. These constraints can be used to compute S^{-1} hence S.

The above derivation is based on the assumption that the fundamental matrix F is unit, *i.e.* $\|p'\| = 1$. However, since the epipolar constraint is homogeneous in the fundamental matrix F, it can only be determined up to an arbitrary scale. Suppose λ is the length of the vector $p' \in \mathbb{R}^3$ in $F = A^{-T}R^TA^T\widehat{p'}$. Consequently, the vectors d_1 and d_2 are also scaled by the same λ, as are ξ_1 and ξ_2. Then the ratio between the left and right hand side quantities in each equation of (57) is equal to λ^2. This gives two independent constraints on S^{-1}, the so called **Kruppa equation:**

$$\lambda^2 = \frac{\xi_1^T S^{-1}\xi_1}{\eta_2^T S^{-1}\eta_2} = \frac{\xi_2^T S^{-1}\xi_2}{\eta_1^T S^{-1}\eta_1} = \frac{\xi_1^T S^{-1}\xi_2}{\eta_1^T S^{-1}\eta_2}. \tag{58}$$

Alternative means of obtaining the Kruppa equation is by utilizing algebraic relationships between projective geometric quantities [18] or via SVD characterization of F [6]. Here we obtain the same equation from a quite different approach. Equation (58) further reveals the geometric meaning of the Kruppa ratio: it is the square of the length of the vector p' in the fundamental matrix F. Note that the above approach of deriving Kruppa equation does not have to use the singular value decomposition (SVD) of F, hence, computationally, it is less costly. Each fundamental matrix provides two (Kruppa) constraints on S^{-1}. Since the symmetric matrix S has six degrees of freedom, in general at least three fundamental matrices are needed to uniquely determine S.

The above derivation of Kruppa equation is straight forward but the expression (58) depends on a particular rotation matrix R_0 that one chooses – note that the choice of R_0 is not unique. In fact, there is an even simpler way to get a equivalent expression for the Kruppa equation in a matrix form. Given a unit fundamental matrix $F = A^{-T}R^TA^T\widehat{p'}$, note that the element $A^{-T}R^TA^T \in A^{-T}SO(3)A^T$ acts on each column of the skew matrix $\widehat{p'}$. It is then natural to view the fundamental matrix F as an cotangent vector (of the group $ASO(3)A^{-1}$) with appropriate coinvariants associated to it. Applying Corollary 2, one directly gets the matrix equation:

$$F^T S^{-1} F = \widehat{p'}^T S^{-1}\widehat{p'}. \tag{59}$$

We call this equation the **normalized matrix Kruppa equation**. It is readily seen that this equation is equivalent to (57). If F is not unit and is scaled by $\lambda \in \mathbb{R}$, *i.e.* $F = \lambda A^{-T}R^TA^T\widehat{p'}$, we then have the **matrix Kruppa equation:**

$$F^T S^{-1} F = \lambda^2\widehat{p'}^T S^{-1}\widehat{p'}. \tag{60}$$

This equation is equivalent to the scalar Kruppa equation (58) and is independent of the choice of the rotation matrix R_0. If we view a 3×3 matrix as a vector in $\mathbb{R}^9$, (60) simply says that the two vectors $F^TS^{-1}F$ and $\widehat{p'}^T S^{-1}\widehat{p'}$ are linearly dependent. This fact can be expressed in a more compact form:

$$\left(F^TS^{-1}F\right) \wedge \left(\widehat{p'}^T S^{-1}\widehat{p'}\right) = 0 \tag{61}$$

where the wedge product is the standard one of the vector space $\mathbb{R}^9$. This equation is homogeneous in the fundamental matrix F and the vector p' hence F and p' do not have to be unit. However, this equation may introduce extra solutions for S^{-1} because it does not impose the condition that the quantity λ^2 has to be positive.

Algebraic properties of Kruppa equation were first studied by Maybank and Faugeras in [18]. However, conditions on possible dependences among Kruppa equations obtained from different fundamental matrices were not clearly given, at least not given in a geometrically intuitive form. Therefore it is hard to tell whether a set of Kruppa equations give a unique solution for calibration. Since the Kruppa equation is highly nonlinear in S^{-1}, most Kruppa equation based algorithms suffer from being computationally expensive and having multiple local minimums. These reasons have motivated us to study the geometric nature of this equation and hope to gain a better understanding and obtain simpler methods for camera self-calibration. On our way to do so, we first study camera self-calibration for some special (and simple) camera motions and then show that so-obtained results also help us to understand the general case and the Kruppa equation better.

6.2 Cases with pure translation or pure rotation

We first consider the case that there is only pure translation p and the rotation component R is always equal to the identity I. In this case, the fundamental matrix F has the form $A^{-T}\hat{p}A^{-1}$. According to Lemma 1,

$$A^{-T}\hat{p}A^{-1} = \widehat{Ap}. \tag{62}$$

So Ap can be directly recovered. But no matter how many such fundamental matrices are given, one can never recover A without knowing the actual translational motion p. Therefore, **rotational motion is absolutely necessary for camera self-calibration.**

Another special case is that there is only pure rotation and no translation. This case has been thoroughly studied in the literature [8], but no proof of the necessity and sufficiency of the conditions for a unique calibration has ever been given. We here give a clear answer to it.

In the pure rotation case, the fundamental matrix F is not well-defined hence the Kruppa equation based approach cannot be used here – but this does not necessarily say that these two cases are not deeply related. In stead, a matrix of the form $ARA^{-1} \in ASO(3)A^{-1}$ can be directly estimated from no less than four image correspondences between two images. The problem of estimating such a matrix was mentioned by Hartley [8]. Here, to be self-contained, we explicitly give out this linear scheme. In the pure rotation case, corresponding image pairs $(\mathbf{x}_1^j, \mathbf{x}_2^j), j = 1, \ldots, 4$ satisfy:

$$\lambda_2^j \mathbf{x}_2^j = ARA^{-1}\lambda_1^j \mathbf{x}_1^j, \quad j = 1, \ldots, 4 \tag{63}$$

for some scales $\lambda_1^j, \lambda_2^j, j = 1, \ldots, 4$. Then we have eight linear constraints on the matrix ARA^{-1} as:

$$\widehat{\mathbf{x}}_2^j ARA^{-1}\mathbf{x}_1^j = 0 \tag{64}$$

for $j = 1, \ldots, 4$. In general these linear equations uniquely determine the matrix ARA^{-1}.

Information about the matrix A is therefore encoded in the conjugate group $ASO(3)A^{-1}$ of $SO(3)$. It will be useful to understand the relation between the two groups: $SO(3)$ and $ASO(3)A^{-1}$. In particular, we need to study the problem: given an element, say matrix $C \in \mathbb{R}^{3\times3}$, in the group $ASO(3)A^{-1}$, how much does it tell us about the matrix A? Since $C \in ASO(3)A^{-1}$, there exists a matrix $R \in SO(3)$ such that $C = ARA^{-1}$. As usual, let $S = A^{-T}A^{-1}$, we have:

$$S - C^T SC = 0. \tag{65}$$

That is, S has to be in the kernel of the linear map:

$$\begin{aligned} L : \mathbb{C}^{3\times3} &\to \mathbb{C}^{3\times3} \\ X &\mapsto X - C^T XC. \end{aligned} \tag{66}$$

Note that this is a Lyapunov map. According to Callier and Desoer [4], it has eigenvalues $1 - \lambda_i^* \lambda_j, 1 \leq i, j \leq 3$ where $\lambda_i, i = 1, 2, 3$ are eigenvalues of the matrix C.

Suppose the rotation matrix R has eigenvalues $1, \alpha, \bar{\alpha} \in \mathbb{C}$ with $\alpha\bar{\alpha} = 1$ and corresponding right eigenvectors $u, v, \bar{v} \in \mathbb{C}^3$. [3] Then the matrix C has the same eigenvalues and corresponding right eigenvectors become $A^{-T}u, A^{-T}v, A^{-T}\bar{v} \in \mathbb{C}^3$. Then the matrix S is in the 3 dimensional subspace:

$$\mathrm{Ker}(L) = \mathrm{span}\{S_1 = A^{-T}uu^*A^{-1}, S_2 = A^{-T}vv^*A^{-1}, S_3 = A^{-T}\bar{v}\bar{v}^*A^{-1}\} \quad \subset \mathbb{C}^{3\times3}. \tag{67}$$

This is the kernel of the linear map L. Since $R \neq I$, S_1 is real, $S_2 = \bar{S}_3$ and S_1, S_2, S_3 are linearly independent. For a real symmetric solution of S, it must have the form $S = \beta S_1 + \gamma(S_2 + S_3)$ with $\beta, \gamma \in \mathbb{R}$. The solution space for symmetric real S is only two dimensional[4]. We call this two dimensional space as the **symmetric real kernel** of the map L, denoted as $\mathrm{SRKer}(L)$. Summarizing the above we obtain:

Lemma 2 *Given a matrix $C = ARA^{-1}$ in the matrix group $ASO(3)A^{-1}$ with the rotation matrix R not equal to the identity matrix I, the symmetric real kernel associated with the Lyapunov map $L : X - C^T XC$ is of 2 dimension.*

Since the symmetric real kernel associated with one matrix $C \in ASO(3)A^{-1}$ is only two dimensional, one more effective constraint on S will be able to uniquely determine it, for example see Hartley [7, 8]. However, we are more interested in uniquely determining S from elements in $ASO(3)A^{-1}$. Suppose we know n elements $C_j, j = 1, \ldots n$ in the group $ASO(3)A^{-1}$. Then S has to be in the (symmetric real) kernels of all the linear maps:

$$\begin{aligned} L_j : \mathbb{C}^{3\times3} &\to \mathbb{C}^{3\times3}, \qquad j = 1, \ldots, n \\ X &\mapsto X - C_j^T XC_j. \end{aligned} \tag{68}$$

That is $S \in \mathrm{SRKer}(L_1) \cap \ldots \cap \mathrm{SRKer}(L_n)$.

Before we give a useful sufficient condition for such S to be unique, we state a lemma from linear algebra.

[3] Here we still use column vector convention to represent right eigenvectors. Therefore, we have $u^*C = u^*, v^*C = \alpha v^*$ where $(\cdot)^*$ means the Hermitian transpose.

[4] L has a three dimensional real kernel but one dimension is $i(S_2 - S_3)$ which is skew-symmetric.

Lemma 3 *If $u, v, w \in \mathbb{R}^3$ are linearly independent, then the matrices uu^T, vv^T and ww^T are linearly independent.*

Proof: For non-zero $\lambda_1, \lambda_2 \in \mathbb{R}$, the matrix $\lambda_1 uu^T + \lambda_2 vv^T$ always has rank 2 while the matrix ww^T has rank 1. ■

Lemma 4 *Given matrices $C_j = AR_jA^{-1} \in ASO(3)A^{-1}, j = 1, \ldots, n$ with $R_j \neq I$ and given the real right (hence left) eigenvectors $u^j \in \mathbb{R}^3$ of R_j (i.e. the principal axis of the rotation matrix R_j). If three of the n principal axes $u^j, j = 1, \ldots, n$ are linearly independent, then there is a unique real symmetric matrix $S \in SL(3)$ satisfying $S = C_j^T S C_j, j = 1, \ldots, n$ hence $S = A^{-T}A^{-1}$.*

Proof: We may assume u^1, u^2, u^3 are linearly independent. Then according to Lemma 3 the three matrices $u^1(u^1)^T, u^2(u^2)^T, u^3(u^3)^T$ are linearly independent. Then

$$\text{span}\{A^{-T}u^1(u^1)^TA^{-1}, A^{-T}u^2(u^2)^TA^{-1}, A^{-T}u^3(u^3)^TA^{-1}\}$$

has three dimensions. Thus $\text{SRKer}(L_3)$ is not fully contained in $\text{SRKer}(L_1) \cap \text{SRKer}(L_2)$ hence their intersection $\text{SRKer}(L_1) \cap \text{SRKer}(L_2) \cap \text{SRKer}(L_3)$ has at most one dimension. This guarantees that S has a unique solution. ■

If we study the condition more carefully, we can actually obtain the following necessary and sufficient condition.

Theorem 3 (Sufficient and necessary condition of unique calibration) *Given matrices $C_j = AR_jA^{-1} \in ASO(3)A^{-1}, j = 1, \ldots, n$ with $R_j \neq I$ and given the real right (hence left) eigenvectors $u^j \in \mathbb{R}^3$ of R_j (i.e. the principal axis of the rotation matrix R_j). The real symmetric matrix $S = A^{-T}A^{-1} \in SL(3)$ is uniquely determined if and only if at least* two *of the n principal axes $u^j, j = 1, \ldots, n$ are linearly independent.*

Proof: The necessity is obvious: if two rotation matrices R_i and R_j have the same axis, they have the same eigenvectors hence $\text{SRKer}(L_i) = \text{SRKer}(L_j)$. We now only need to prove the sufficiency. We may assume u^1 and u^2 are linearly independent and both are unit vectors. Define matrices $R_{n+1} = R_1R_2$ and $C_{n+1} \in \mathbb{R}^{3\times3}$:

$$C_{n+1} = C_1C_2 = AR_1A^{-1}AR_2A^{-1} = AR_{n+1}A^{-1}. \tag{69}$$

The rotation matrix R_1 must has the form $\exp(\widehat{u^1}\theta_1)$ for some $\theta_1 \in \mathbb{R}$ and similarly $R_2 = \exp(\widehat{u^2}\theta_2)$ for some $\theta_2 \in \mathbb{R}$. Then the axis $u^{n+1} \in \mathbb{R}^3$ of the rotation matrix R_{n+1} is given by (see Murray, Li and Sastry [19]):

$$u^{n+1} = \cos(\theta_1/2)\sin(\theta_2/2) \cdot u^2 + \cos(\theta_2/2)\sin(\theta_1/2) \cdot u^1 + \sin(\theta_1/2)\sin(\theta_2/2) \cdot (u^1 \times u^2). \tag{70}$$

Note u^{n+1} may not be of unit length here. But it is linearly independent of u^1 and u^2 if both θ_1 and θ_2 are not zero. This then reduces to the case of Lemma 4. ■

According to this theorem, the simplest way to calibrate an uncalibrated camera is to rotate it about two different axes. The self-calibration algorithm in this case will be completely linear and a unique solution is also guaranteed.

Comments 3 *Note that Kruppa equation gives direct constraints on S^{-1} instead of S. In the pure rotation case, we have:*

$$S = C^T S C \quad \Leftrightarrow \quad S^{-1} = C^{-1} S^{-1} C^{-T}. \tag{71}$$

Thus, we also obtain direct constraints on S^{-1}. This provides a possibility of combining these two types of constraints in a unified framework for estimating S^{-1}.

6.3 Cases with constant translation or constant rotation

By the case of constant translation, we mean that the translation $p(t)$ or $p'(t)$ always has a constant direction and the axis of rotation $R(t)$ is arbitrary; by the case with constant rotation, we mean that the axis of the rotation $R(t)$ is constant and the translation direction is arbitrary. Note that these two cases are different from the cases with pure translation or pure rotation.

We first study the case with constant rotation. The fundamental matrices obtained in this case have the form:

$$F_i = A^{-T} R_i^T A^T \widehat{p'_i}, \quad i = 1, \ldots, n \tag{72}$$

where $R \in SO(3)$ and $p'_i \in \mathbb{R}^3, i = 1, \ldots, n$. Note that the columns of matrices F_i's are linear combinations of columns of the matrices $A^{-T} R_i^T A^T, 1 \leq i \leq n$, or equivalently the rows of the matrices $C_i = A R_i A^{-1}, 1 \leq i \leq n$. Since R_i's have the same axis, according to Theorem 3, such matrices are not sufficient to determine the camera calibration. Therefore, **not only rotational motion but also variation in rotational motion is necessary for camera self-calibration.**

We next study the case with constant translation – the translation vector p' always has a fixed direction. That is, the fundamental matrices obtained in this case are supposed to have the common form:

$$F_i = A^{-T} R_i^T A^T \widehat{p'}, \quad i = 1, \ldots, n \tag{73}$$

where $R_i \in SO(3), i = 1, \ldots, n$ and $p' \in \mathbb{R}^3$. Now suppose we have three consecutive images I_1, I_2, I_3. The fundamental matrices estimated between the i^{th} and j^{th} images are denoted as $F_{ij}, 1 \leq i < j \leq 3$. Then under the constant translation assumption, they all have the form:

$$F_{ij} = A^{-T} R_{ij}^T A^T \widehat{p'}, \quad 1 \leq i < j \leq 3 \tag{74}$$

and we may assume $\|p'\| = 1$. Pick any rotation matrix R_0 such that $R_0 p' = e_3$. Then:

$$D_{ij} = R_0 F_{ij} R_0^T = (R_0 A)^{-T} R_{ij} (R_0 A)^T \widehat{e_3}, \quad 1 \leq i < j \leq 3. \tag{75}$$

Define matrices $C_{ij} \in \mathbb{R}^{3\times3}$:

$$C_{ij} = (R_0 A)^{-T} R_{ij} (R_0 A)^T, \quad 1 \leq i < j \leq 3. \tag{76}$$

It is readily seen that the first two columns of C_{ij} can be directly obtained from D_{ij}. We also have relations:

$$C_{12} \cdot C_{23} = C_{13}. \tag{77}$$

Let us denote the k^{th} column of the matrix C_{ij} as $c_{ij}^k \in \mathbb{R}^3, 1 \leq k \leq 3, 1 \leq i < j \leq 3$. Then even if D_{ij}'s are estimated up to arbitrary scales, from (77) we still have:

$$(c_{12}^1, c_{12}^2, c_{12}^3) \cdot (c_{23}^1, c_{23}^2) = \lambda_{13}(c_{13}^1, c_{13}^2) \tag{78}$$

where $\lambda_{13} \in \mathbb{R}$. If D_{ij}'s are already unit fundamental matrices, then $\lambda_{13} = 1$. From the above linear equations the unknown columns c_{12}^3 and scalars λ_{13} can be uniquely determined. Since we have more equations than unknowns, the numerical solution can be taken as the associated LLSE estimate. We can therefore obtain C_{12} up to a scale. Knowing that it is in $SL(3)$, we can further normalize it and obtain the matrix $C_{12} \in (R_0A)^{-T}SO(3)(R_0A)^T$. We thus obtain one element $C_1 = R_0^T C_{12} R_0 = A^{-T} R_{12} A^T$ in $A^{-T}SO(3)A^T$. If we have another three such image frames, we can obtain another element, say C_2, in $A^{-T}SO(3)A^T$. According to Theorem 3, if the rotational components in these two elements have different principal axes, the symmetric matrix $S^{-1} = AA^T$ is uniquely determined.

Although the above approach requires two groups of three images with constant translation vectors, the algorithm is purely linear and a unique solution is guaranteed. Most importantly, it provides an alternative way to the nonlinear Kruppa equation in the case that the camera has both rotational and translational motion. In practice, the above approach can be approximately used as long as the variation of the direction of translation is relatively slower, comparing to the variation of the direction of the rotation.

Example 1 *In order to obtain a group of image frames with constant (relative) translation p', one needs to impose some constraints on the motion of the camera. In the case that the camera is attached to a mobile robot with kinematics given by:*

$$\dot{h} = h \begin{pmatrix} \hat{\omega} & v \\ 0 & 1 \end{pmatrix} \tag{79}$$

where $h \in SE(3)$ and $\omega, v \in \mathbb{R}^3$. In order for $p'(t)$ to be constant, one may simply choose

$$\omega = \alpha_1 u, \quad v = \alpha_2 u \tag{80}$$

where $u \in \mathbb{R}^3$ is a fixed unit vector and $\alpha_1, \alpha_2 \in \mathbb{R}$ are constants. In this scheme, one does not have to know ω and v, and this constraint can be imposed by a low level controller which only works whenever camera self-calibration is needed.

Comments 4 *As we have seen above, translation causes problems for camera self-calibration while rotation always serves as a positive factor. Interestingly, the situation is quite the opposite in the case of reconstructing structure from motion where, as well-known, it is impossible to recover 3D structure only from rotation and translation is absolutely necessary.*

6.4 General case

Now let us go back to the general case and study the matrix Kruppa equation (59) and (60). Given a unit fundamental matrix $F = A^{-T}R^TA^T\widehat{p'}$ with p' of unit length, let $C = A^{-T}R^TA^T$, define the linear map $\sigma : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ as:

$$\sigma : X \mapsto C^TXC - X, \quad X \in \mathbb{R}^{3\times3}, \tag{81}$$

and define the linear map $\tau : \mathbb{R}^{3\times3} \rightarrow \mathbb{R}^{3\times3}$ as:

$$\tau : Y \mapsto \widehat{p'}^T Y \widehat{p'}, \quad Y \in \mathbb{R}^{3\times3}. \tag{82}$$

The solution S^{-1} of the normalized Kruppa equation (59) is then in the (symmetric real) kernel of the composition map:

$$\tau \circ \sigma : \quad \mathbb{R}^{3\times3} \xrightarrow{\sigma} \mathbb{R}^{3\times3} \xrightarrow{\tau} \mathbb{R}^{3\times3}. \tag{83}$$

This interpretation of Kruppa equation decomposes effects of the rotational and translational parts of the motion: if there is no translational motion, *i.e.* there is no map τ, the symmetric kernel of σ is just a two dimensional subspace according to Theorem 3; if the translational motion is non-zero, the kernel is enlarged by composing the map τ. According to (57), in general, the symmetric real kernel of the composition map $\tau \circ \sigma$ is three dimensional.

Because of the unknown scale λ, the solution space for the matrix Kruppa equation (60) is even larger than that of the normalized one (59). It will be helpful to know under what conditions the matrix Kruppa equation may have the same solution as the normalized one.

Lemma 5 *Given a fundamental matrix $F = A^{-T}R^T A^T \widehat{p'}$ with $p' = Ap$, a real symmetric matrix $X \in \mathbb{R}^{3\times3}$ is a solution of $F^T X F = \lambda^2 \widehat{p'}^T X \widehat{p'}$ if and only if $Y = A^{-1} X A^{-T}$ is a solution of $E^T Y E = \lambda^2 \hat{p}^T Y \hat{p}$ with $E = R^T \hat{p}$.*

The proof is trivial. This simple lemma however states a very important fact: given a set of fundamental matrices $F_i = A^{-T} R_i^T A^T \widehat{p'_i}$ with $p'_i = Ap_i, i = 1, \ldots, n$, there is a one-to-one correspondence between the set of solutions of the Kruppa equations:

$$F_i^T X F_i = \lambda_i^2 \widehat{p'_i}^T X \widehat{p'_i}, \quad i = 1, \ldots, n. \tag{84}$$

and the set of solutions of the equations:

$$E_i^T Y E_i = \lambda_i^2 \hat{p}_i^T Y \hat{p}_i, \quad i = 1, \ldots, n \tag{85}$$

where $E_i = R_i^T \hat{p}_i$ are essential matrices associated to the given fundamental matrices. Note that these essential matrices are determined only by the camera motion, therefore **conditions of uniqueness of the solution of Kruppa equations should only depend on the camera motion.**

What do we gain from this observation? It immediately gives us some suggestion for possible camera motions which may make the use of the Kruppa equation for self-calibration simpler.

Theorem 4 *Given a camera motion $(R, p) \in SE(3)$, if the axis of the rotation R is perpendicular to the translational motion vector p, then the matrix Kruppa equation:*

$$\hat{p}^T R Y R^T \hat{p} = \lambda^2 \hat{p}^T Y \hat{p} \tag{86}$$

has the same solutions of non-degenerate real symmetric Y as the normalized matrix Kruppa equation:

$$\hat{p}^T R Y R^T \hat{p} = \hat{p}^T Y \hat{p}. \tag{87}$$

Proof: Since the columns of $\hat{p}$ span the subspace which is perpendicular to the vector p, the eigenvector u of R is in this subspace. Thus we have:

$$u^T R Y R^T u = \lambda^2 u^T Y u \quad \Rightarrow \quad u^T Y u = \lambda^2 u^T Y u. \tag{88}$$

Hence $\lambda^2 = 1$ if Y is non-degenerate. ∎

Under the conditions given by the theorem, then there would be no solution for λ in the Kruppa equation (60) besides the true scale of the fundamental matrix. Then each fundamental matrix can be immediately made unit by picking any (non-degenerate) solution of the associated Kruppa equation. Once the fundamental matrices are unit, the problem of solving the calibration matrix S^{-1} from $n \geq 3$ normalized matrix Kruppa equations becomes a simple **linear** one!

Comments 5 *Interestingly, in the case of human eyes, such conditions are closely satisfied: the main rotation of human eyes and head are yaw and pitch which have axes perpendicular to the direction of walking. As the theorem suggests, self-calibration in this situation is much easier than roll motion is allowed. Similar cases can also often be found in vision-guided navigation systems.*

We here summarize some of the conditions on (R_i, p_i)'s such that equations in (85) may give a unique solution for calibration. From the study of the constant rotation case, we know R_i's cannot share a common principal axis otherwise the solutions would be a one-parameter family. Consequently, one can actually show that full calibration is not possible if the camera motion is restricted to any proper subgroup of $SE(3)$ [14]. Also, due to $\exp(\hat{p}_i\pi)\hat{p}_i = -\hat{p}_i$, in order that equations given in (85) are not trivial, it is required that R_i is not equal to $\exp(\hat{p}_i\pi)$. However, these are just necessary conditions, and it remains an open problem that for given three fundamental matrices under what (appropriate) conditions on the motion $(R_i, p_i), i = 1, 2, 3$ the associated three Kruppa equations will have a unique (or finite number of) solution(s). Although it is claimed in [18] that three is the minimum number of Kruppa equations needed, such conditions were not clearly studied.

7 Differential case

So far, we have understood camera self-calibration when the motion of the camera is discrete – positions of the camera are specified as discrete points in $SE(3)$. In this section, we study its differential (or continuous) version. Define the angular velocity $\hat{\omega} = \dot{R}(t)R^T(t) \in so(3)$ and linear velocity $v = -\hat{\omega}p(t) + \dot{p}(t) \in \mathbb{R}^3$ and. Let $v' = Av \in \mathbb{R}^3, \omega' = A\omega \in \mathbb{R}^3$. Differentiating the equation (20) with respect to time t, we obtain:

$$\dot{r} = A\hat{\omega}A^{-1}r + v' \tag{89}$$

where, to simplify the notation, we use r to replace the original notation $q' \in M'$.

7.1 General motion case

By the general case we mean that both the angular and linear velocities ω and v are non-zero. Note that $r = \lambda \mathbf{x}$ yields $\dot{r} = \dot{\lambda}\mathbf{x} + \lambda\dot{\mathbf{x}}$. Then (89) gives:

$$\begin{aligned} &\dot{r} = A\hat{\omega}A^{-1}r + v' \quad \Rightarrow \quad (v' + \mathbf{x}) \times \dot{r} = (v' + \mathbf{x}) \times A\hat{\omega}A^{-1}r \\ \Rightarrow \quad &\dot{\mathbf{x}}^T A^{-T}\hat{v}A^{-1}\mathbf{x} + \mathbf{x}^T A^{-T}\hat{\omega}\hat{v}A^{-1}\mathbf{x} = 0. \end{aligned} \tag{90}$$

The last equation is called the **differential epipolar constraint.** Let $s \in \mathbb{R}^{3\times3}$ to be $s = \frac{1}{2}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})$. Define the **differential fundamental matrix** $F' \in \mathbb{R}^{6\times3}$ to be:

$$F' = \begin{pmatrix} A^{-T}\hat{v}A^{-1} \\ A^{-T}sA^{-1} \end{pmatrix}. \tag{91}$$

F' can therefore be estimated from as few as eight optical flows $(\mathbf{x}, \dot{\mathbf{x}})$ from (90) (see Ma, Kosecka and Sastry [12]).

Note that $\widehat{v'} = A^{-T}\hat{v}A^{-1}$ and $\widehat{\omega'} = A^{-T}\hat{\omega}A^{-1}$. Applying Lemma 1 repeatedly, we obtain

$$A^{-T}sA^{-1} = \frac{1}{2}A^{-T}(\hat{\omega}\hat{v} + \hat{v}\hat{\omega})A^{-1} = \frac{1}{2}(A^{-T}\hat{\omega}A^{T}\widehat{v'} + \widehat{v'}A\hat{\omega}A^{-1}) = \frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'}). \tag{92}$$

Then the differential epipolar constraint (90) is equivalent to:

$$\dot{\mathbf{x}}^T\widehat{v'}\mathbf{x} + \mathbf{x}^T\frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'})\mathbf{x} = 0. \tag{93}$$

Suppose $S^{-1} = BB^T$ for another $B \in SL(3)$, then $A = BR_0$ for some $R_0 \in SO(3)$. We have:

$$\begin{aligned} & \dot{\mathbf{x}}^T\widehat{v'}\mathbf{x} + \mathbf{x}^T\frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'})\mathbf{x} = 0 \\ \Leftrightarrow\quad & \dot{\mathbf{x}}^T\widehat{v'}\mathbf{x} + \mathbf{x}^T\frac{1}{2}(\widehat{\omega'}BB^T\widehat{v'} + \widehat{v'}BB^T\widehat{\omega'})\mathbf{x} = 0 \\ \Leftrightarrow\quad & \dot{\mathbf{x}}^TB^{-T}\widehat{R_0v}B^{-1}\mathbf{x} + \mathbf{x}^TB^{-T}\widehat{R_0\omega}\widehat{R_0v}B^{-1}\mathbf{x} = 0. \end{aligned} \tag{94}$$

Comparing to (90), one cannot tell the camera A with motion (ω, v) from the camera B with motion $(R_0\omega, R_0v)$. Thus, like the discrete case, without knowing the camera motion the calibration can only be recovered in the space $SL(3)/SO(3)$, *i.e.* only the symmetric matrix S^{-1} hence S can be recovered.

However, unlike the discrete case, the matrix S cannot be fully recovered in the differential case. Since $S^{-1} = AA^T$ is a symmetric matrix, it can be diagnalized as:

$$S^{-1} = R_1^T\Sigma R_1, \quad R_1 \in SO(3) \tag{95}$$

where $\Sigma = \mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3\}$. Then let $\omega'' = R_1\omega'$ and $v'' = R_1v'$. Applying Lemma 1, we have:

$$\begin{aligned} \widehat{v'} &= R_1^T\widehat{v''}R_1 \\ \frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'}) &= R_1^T\frac{1}{2}(\widehat{\omega''}\Sigma\widehat{v''} + \widehat{v''}\Sigma\widehat{\omega''})R_1. \end{aligned} \tag{96}$$

Thus the differential epipolar constraint (90) is also equivalent to:

$$(R_1\dot{\mathbf{x}})^T\widehat{v''}(R_1\mathbf{x}) + (R_1\mathbf{x})^T\frac{1}{2}(\widehat{\omega''}\Sigma\widehat{v''} + \widehat{v''}\Sigma\widehat{\omega''})(R_1\mathbf{x}) = 0. \tag{97}$$

From this equation, one can see that there is no way to tell a camera A with $AA^T = R_1^T\Sigma R_1$ from a camera $B = R_1A$. Therefore, only the diagonal matrix Σ can be recovered as camera parameters since both the scene structure and camera motion are unknown.

Note that Σ is in $SL(3)$ hence $\sigma_1\sigma_2\sigma_3 = 1$. The singular values only have two degrees of freedom. Hence we have:

Theorem 5 *Consider an uncalibrated camera with an unknown calibration matrix $A \in SL(3)$. Then only the eigenvalues of AA^T can be recovered from the bilinear differential epipolar constraint.*

If we define that two matrices in $SL(3)$ are equivalent if and only if they have the same singular values. The intrinsic parameter space is then reduced to the space $SL(3)/\sim$ where $\sim$ represents this equivalence relation. The fact that only two camera parameters can be recovered was known to Brooks, Chojnacki and Baumela [3]. They have also shown how to do calibration for certain matrices A with only two unknown parameters. But the intuitive geometric reason was hidden in their algebraic geometry arguments.

Comments 6 *It is a little surprising to see that the discrete and differential cases are different for the first time, especially knowing that in the calibrated case these two cases have exactly parallel theories. We believe that this has something to do with the map:*

$$\begin{aligned} \gamma_A : \mathbb{R}^{3\times3} &\rightarrow \mathbb{R}^{3\times3} \\ B &\mapsto ABA^T \end{aligned}$$

where A is an arbitrary matrix in $\mathbb{R}^{3\times3}$. Let $so(3)$ be the Lie algebra of $SO(3)$. The restricted map $\gamma_A\,|_{so(3)}$ is an endomorphism while $\gamma_A\,|_{SO(3)}$ is not. Consider $\gamma_A\,|_{so(3)}$ to be the first order approximation of $\gamma_A\,|_{SO(3)}$. Then the information about the camera matrix A will not fully show up until the second order term of the map γ_A. This also explains why in the discrete case the (Kruppa) constraints that we can get for A must be nonlinear.

Comments 7 *From the above discussion, if one only uses the (bilinear) differential epipolar constraint, at most two intrinsic parameters of the calibration matrix A can be recovered. However, it is still possible that the full information about A can be recovered from multilinear constraints on the higher order derivatives of optical flow. A complete list of such constraints are given in Ma, Kosecka and Sastry [11] or Astrom [1].*

7.2 Cases with pure translation or pure rotation

Since full calibration is not always possible in the general case, we need to know if it is possible in some special cases.

First we consider the case that the camera does pure translation, *i.e.* the angular velocity ω is always zero. In this case, the differential fundamental matrix F' has the form

$$F' = \begin{pmatrix} A^{-T}\hat{v}A^{-1} \\ 0 \end{pmatrix} = \begin{pmatrix} \widehat{v'} \\ 0 \end{pmatrix}. \tag{98}$$

Only the vector $v' = Av$ is recovered. As in the discrete case, there is no way to recover the matrix A from it without knowing the actual linear velocity v. Therefore, **rotational motion is absolutely necessary for camera self-calibration in the differential case.**

Another special case is when there is only rotational motion, *i.e.* the linear velocity v is always zero. In this case the differential fundamental matrix is no longer well defined. However from the equation (89) we have:

$$\begin{aligned} \dot{r} = A\hat{\omega}A^{-1}r \quad &\Rightarrow \quad \dot{\lambda}\mathbf{x} + \lambda\dot{\mathbf{x}} = A\hat{\omega}\lambda A^{-1}\mathbf{x} \\ \Rightarrow \quad \hat{\mathbf{x}}\dot{\mathbf{x}} &= \hat{\mathbf{x}}A\hat{\omega}A^{-1}\mathbf{x}. \end{aligned} \tag{99}$$

This equation gives two independent constraints on the matrix $A\hat{\omega}A^{-1}$. Given as few as four optical flow measurements $(\mathbf{x}, \dot{\mathbf{x}})$, one may uniquely determine the matrix $A\hat{\omega}A^{-1}$ with ω normalized, *i.e.* $\|\omega\| = 1$. Such normalization is possible because $A\hat{\omega}A^{-1}$ has the same eigenvalues as $\hat{\omega}$. Then the calibration problem becomes how to recover $S = A^{-T}A^{-1}$ or $S^{-1} = AA^T$ from matrices of the form $A\hat{\omega}A^{-1}$. We may assume ω is always normalized. Notice that this problem is exactly a differential version of the discrete pure rotation case.

Let $C' = A\hat{\omega}A^{-1} \in \mathbb{R}^{3\times 3}$. Then we have:

$$SC' = A^{-T}\hat{\omega}A^{-1} = \widehat{\omega'} \tag{100}$$

where $\omega' = A\omega$. Thus $SC' = -(SC')^T$, *i.e.* $SC' + (C')^T S = 0$. That is, S has to be in the kernel of the linear map:

$$\begin{aligned} L' : \mathbb{C}^{3\times 3} &\rightarrow \mathbb{C}^{3\times 3} \\ X &\mapsto (C')^T X + XC' \end{aligned} \tag{101}$$

This is also a Lyapunov map. If $\omega \neq 0$, the eigenvalues of $\hat{\omega}$ have the form $0, i\alpha, -i\alpha$ with $\alpha \in \mathbb{R}$. Let the corresponding eigenvectors are $\omega, u, \bar{u} \in \mathbb{C}^3$. According to Callier and Desoer [4], the null space of the map L' has three dimensions and is given by:

$$\mathrm{Ker}(L') = \mathrm{span}\{S_1 = A^{-T}\omega\omega^* A^{-1}, S_2 = A^{-T}uu^* A^{-1}, S_3 = A^{-T}\bar{u}\bar{u}^* A^{-1}\}. \tag{102}$$

As in the discrete case, the symmetric real S is of the form $S = \beta S_1 + \gamma(S_2 + S_3)$, *i.e.* the symmetric real kernel of L' is only two dimensional. We denote this space as $\mathrm{SRKer}(L')$. We thus have:

Lemma 6 *Given a matrix $C' = A\hat{\omega}A^{-1}$ with $\omega \in S^2$, the symmetric real kernel associated with the Lyapunov map $L' : (C')^T X - XC'$ is of 2 dimension.*

Similarly to the proof of the discrete case, we also obtain:

Lemma 7 *Given matrices $C'_j = A\hat{\omega}_j A^{-1} \in \mathbb{R}^{3\times 3}, j = 1, \ldots, n$ with $\|\omega_j\| = 1$. If three of the n vectors $\omega_j, j = 1, \ldots, n$ are linearly independent, then there is a unique real symmetric matrix $S \in SL(3)$ satisfying $(C'_j)^T S + SC'_j = 0, j = 1, \ldots, n$ hence $S = A^{-T}A^{-1}$.*

Following this lemma, we further have:

Theorem 6 (Sufficient and necessary condition of unique calibration) *Given matrices $C'_j = A\hat{\omega}_j A^{-1} \in \mathbb{R}^{3\times 3}, j = 1, \ldots, n$ with $\|\omega_j\| = 1$. The real symmetric matrix $S = A^{-T}A^{-1} \in SL(3)$ is uniquely determined if and only if at least* **two** *of the n vectors $\omega_j, j = 1, \ldots, n$ are linearly independent.*

Proof: We may assume ω_1 and ω_2 are linearly independent. As the discrete case, we construct a third matrix of the form $A\hat{\omega}A^{-1}$. We may define the matrix to be $C'_{n+1} \in \mathbb{R}^{3\times 3}$:

$$C'_{n+1} = A\hat{\omega}_1 A^{-1} A\hat{\omega}_2 A^{-1} - A\hat{\omega}_2 A^{-1} A\hat{\omega}_1 A^{-1} = A[\hat{\omega}_1, \hat{\omega}_2]A^{-1} \tag{103}$$

where the bracket operator $[\cdot,\cdot]$ is the Lie bracket on the Lie algebra $so(3)$ of the Lie group $SO(3)$:

$$[\hat{\omega}_1, \hat{\omega}_2] = \hat{\omega}_1\hat{\omega}_2 - \hat{\omega}_2\hat{\omega}_1 = \hat{\omega}_{n+1} \tag{104}$$

where $\omega_{n+1} = \omega_1 \times \omega_2$. One may refer to Murray, Li and Sastry [19] for a more detailed discussion on the Lie algebra $so(3)$. Clearly, ω_{n+1} is linearly independent of ω_1, ω_2. This reduces to the case of Lemma 7. ∎

8 Time-varying case

So far we have only been dealing with the case when all intrinsic parameters of the camera are fixed (or time-invariant). In this section, we study briefly the case when intrinsic parameters of the camera are time-varying and show how the theory of the time-invariant case may help us to understand the time-varying problem. In particular, various forms of the Kruppa equation in the time-varying case will be derived.

Suppose the motion of the camera between two times $t_i \in \mathbb{R}$ and $t_j \in \mathbb{R}$ is

$$(R(t_i, t_j), p(t_i, t_j)), \quad R(t_i, t_j) \in SO(3), p(t_i, t_j) \in \mathbb{R}^3,$$

that is:

$$q(t_i) = R(t_i, t_j)q(t_j) + p(t_i, t_j). \tag{105}$$

Suppose the calibration matrix at time t_i is $A(t_i) \in SL(3)$ and at time t_j is $A(t_j) \in SL(3)$. Then we have the time-varying version of the epipolar constraint:

$$\mathbf{x}(t_j)^T A^{-T}(t_j) R^T(t_i, t_j) A^T(t_i) \widehat{p'}(t_i, t_j) \mathbf{x}(t_i) = 0 \tag{106}$$

where $p'(t_i, t_j) = A(t_i)p(t_i, t_j)$ and $\mathbf{x}(t_i)$ and $\mathbf{x}(t_j)$ are images of the point q at time t_i and t_j respectively. Then we can estimate the fundamental matrix

$$F(t_i, t_j) = \lambda(t_i, t_j) A^{-T}(t_j) R^T(t_i, t_j) A^T(t_i) \widehat{p'}(t_i, t_j) \ \in \mathbb{R}^{3\times3} \tag{107}$$

from the epipolar constraint, where $\lambda(t_i, t_j) \in \mathbb{R}$ is an unknown scalar since we let $p'(t_i, t_j)$ be of unit length. Then we have tne **time-varying Kruppa equation:**

$$F(t_i, t_j)^T A(t_j) A^T(t_j) F(t_i, t_j) = \lambda^2(t_i, t_j) \widehat{p'}^T(t_i, t_j) A(t_i) A^T(t_i) \widehat{p'}(t_i, t_j). \tag{108}$$

Define $S^{-1}(t) = A(t)A^T(t)$. The time-varying Kruppa equation becomes:

$$F(t_i, t_j)^T S^{-1}(t_j) F(t_i, t_j) = \lambda^2(t_i, t_j) \widehat{p'}^T(t_i, t_j) S^{-1}(t_i) \widehat{p'}(t_i, t_j). \tag{109}$$

As in the time-invariant case, we also have the wedge product form:

$$\left(F^T(t_i, t_j) S^{-1}(t_j) F(t_i, t_j)\right) \wedge \left(\widehat{p'}^T(t_i, t_j) S^{-1}(t_i) \widehat{p'}(t_i, t_j)\right) = 0 \tag{110}$$

where we view a 3×3 matrix as a vector in $\mathbb{R}^9$ and the wedge product between two vectors in $\mathbb{R}^9$ is then defined as usual. An interesting feature of the wedge product form is that it is bilinear in the two matrices $S^{-1}(t_i)$ and $S^{-1}(t_j)$. Therefore, knowing $S^{-1}(t_{i-k}), k = 1, 2, 3$, $S^{-1}(t_i)$ can be estimated linearly from the equations:

$$\left(F^T(t_i, t_{i-k}) S^{-1}(t_{i-k}) F(t_i, t_{i-k})\right) \wedge \left(\widehat{p'}^T(t_i, t_{i-k}) S^{-1}(t_i) \widehat{p'}(t_i, t_{i-k})\right) = 0, \quad k = 1, 2, 3. \tag{111}$$

This suggests a *linear recursive scheme* for estimating time-varying calibration. It requires paring the current frame with the previous three frames with known calibrations. Experiments show that such linear scheme is *not* numerically stable and is very sensitive to errors. However, to directly solve the time-varying Kruppa equation is a highly nonlinear problem and it will be our future research to search for a fast and stable algorithm.

As in the time-invariant case, we need to know conditions on a unique solution of the time-varying Kruppa equation. Let $B(t_i, t_j) = A(t_j)A^{-1}(t_i)$, and:

$$\tilde{F}(t_i, t_j) = F(t_i, t_j)B^{-1}(t_i, t_j), \quad \tilde{p}(t_i, t_j) = B(t_i, t_j)p'(t_i, t_j). \tag{112}$$

Note that for fixed t_i, the so defined $\tilde{F}(t_i, t_j)$ has the form of a fundamental matrix in the time-invariant case:

$$\tilde{F}(t_i, t_j) = \lambda(t_i, t_j)A^{-T}(t_j)R^T(t_i, t_j)A^T(t_j)\widehat{\tilde{p}}(t_i, t_j) \in \mathbb{R}^{3\times3}. \tag{113}$$

Hence, knowing $A(t_i)$, the symmetric matrix $S^{-1}(t_j)$ is a solution of the Kruppa equations associated to all such time-invariant fundamental matrices for $t_i \neq t_j$:

$$\tilde{F}^T(t_i, t_j)S^{-1}(t_j)\tilde{F}(t_i, t_j) = \lambda^2(t_i, t_j)\widehat{\tilde{p}}^T(t_i, t_j)S^{-1}(t_j)\widehat{\tilde{p}}(t_i, t_j). \tag{114}$$

This set of equations is obtained as if the calibration is fixed as $S^{-1}(t_j)$. Conditions on the uniqueness of $S^{-1}(t_j)$ can then be studied as in the time-invariant case, which, as we have known, only depend on the relative motions between the j^{th} and i^{th} frames for all $t_i \neq t_j$.

9 Discussions and future work

In this paper, we have proposed a geometric approach for the study of camera self-calibration. The intrinsic geometric meanings of fundamental matrices and the Kruppa equation (and the time-varying Kruppa equation) are discovered in a unified geometric framework, so are their relations with various results concerning camera self-calibration with respect to special camera motions (such as the pure rotation case). Not only it is shown that rotation about two different axes is necessary and sufficient for a unique calibration, but also the relationship between rotation/translation and the Kruppa equation is clearly explained.

As in several of our other papers [12, 13, 11], we investigate differential case as the limit of the discrete case. For camera self-calibration, although essential similarities still exist between these two cases, there is no differential version of the Kruppa equation – the one we have will be a degenerate one which can only determine (at most) two intrinsic parameters of the camera. This also explains the nonlinearity of the Kruppa equation.

Although it has been pointed out in this paper that conditions on the uniqueness (or finiteness) of solutions of the Kruppa equations only depend on camera motion, sufficient conditions are not yet given for general case – one may refer to [14] for a partial answer about unique calibration as well as structure reconstruction from fundamental matrices. On the other hand, special camera motions are studied in this paper for understanding calibration only. In [14], subgroups of $SE(3)$ will be systematically studied for recovering motion, structure and calibration simultaneously.

It is very interesting to see in this paper that Lyapunov maps show up in the proof of necessary and sufficient conditions of unique calibration. Properties of the Lie group $SO(3)$ and its Lie algebra $so(3)$ also play important roles in the proof. These are subjects well studied in system theory or robotics. Their presence in the theory of self-calibration suggests that multi-view geometry, system theory and robotics share a common mathematical ground.

Although in this paper the self-calibration theory is only developed for the Euclidean case, most theorems can be easily generalized to a larger class of Riemannian manifolds (for example see Ma and Sastry [15]). In fact, it may be shown that in general, multi-view geometry is about studying certain intrinsic geometric properties of certain Lie groups (isometry groups of the corresponding spaces). Most of the important objects encountered in multi-view geometry can then be interpreted intrinsically. In a three dimensional Euclidean space for example, the associated Lie group is $SE(3)$ and the associated isotropy subgroup $SO(3)$. Multi-view geometry in this space is then about the study of the quotient space $SE(3)/SO(3)$. For an axiomatic formulation of multi-view geometry based on Lie groups, one may refer to Ma, Shakernia, Kosecka and Sastry [16]. In this way, one may convert most problems in multi-view geometry to some pure differential geometry problems.

Skeletons of algorithms can already be derived from the theories presented in this paper. We will give a more detailed analysis in another paper, where numerical and statistical issues will be systematically addressed for the problem of solving the time-invariant and time-varying Kruppa equations.

References

[1] Kalle Åström. *Invariancy Metheods for Points, Curves and Surfaces in Computational Vision.* PhD thesis, Department of Mathematics, Lund University, 1996.

[2] William M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry.* Academic Press, second edition, 1986.

[3] M.J. Brooks, W. Chojnacki, and L. Baumela. Determining the ego-motion of an uncalibrated camera from instantaneous optical flow. *Journal of Optical Society of America A (Optics, Image Science and Vision*, 14(10):2670–7, 1997.

[4] Frank M. Callier and Charles A. Desoer. *Linear System Theory.* Springer Texts in Electrical Engineering. Springer-Verlag, 1991.

[5] Roe Goodman and Nolan R. Wallach. *Representations and Invariants of the Classical Groups.* Cambridge University Press, 1998.

[6] R.I. Hartley. Kruppa's equations derived from the fundamental matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):133–135, February 1997.

[7] Richard I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proceeding of European Conference on Computer Vision*, pages 471–8, 1994.

[8] Richard I. Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1):5–23, 1997.

[9] Shoshichi Kobayashi and Tatsumi Nomizu. *Foundations of Differential Geometry: Volume I and Volume II.* John Wiley & Sons, Inc., 1996.

[10] Quang-Tuan Luong and Olivier Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *IJCV*, 1994. toappear.

[11] Yi Ma, Jana Kosecka, and Shankar Sastry. Euclidean structure and motion from image sequences. In *UC Berkeley Memorandum No. UCB/ERL M98/38*, 1998.

[12] Yi Ma, Jana Kosecka, and Shankar Sastry. Motion recovery from image sequences: Discrete viewpoint vs. differential viewpoint. In *Proceeding of European Conference on Computer Vision, or UC Berkeley Memorandum No. UCB/ERL M98/11*, 1998.

[13] Yi Ma, Jana Kosecka, and Shankar Sastry. Optimal motion from image sequences: A Riemannian viewpoint. In *submitted to IEEE Conference on Decision and Control*, 1998.

[14] Yi Ma, Jana Kosecka, Stefano Soatto, and Shankar Sastry. Uncalibrated reconstruction up to relevant subgroups. In *UC Berkeley Memorandum UCB/ERL*, 1998.

[15] Yi Ma and Shankar Sastry. Vision theory in spaces of constant curvature. In *UC Berkeley Memorandum No. UCB/ERL M98/36*, 1998.

[16] Yi Ma, Omid Shakernia, Jana Kosecka, and Shankar Sastry. A Lie theoretic approach to motion and structure in computer vision. In *Proceedings of the 14th World Congress of IFAC*, 1999.

[17] Stephen Maybank. *Theory of Reconstruction from Image Motion.* Springer Series in Information Sciences. Springer-Verlag, 1993.

[18] Stephen J. Maybank and Oliver D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[19] Richard M. Murray, Zexiang Li, and Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation.* CRC Press, 1993.

[20] Michael Spivak. *A Comprehensive Introduction to Differential Geometry: Volume II.* Publish or Perish, Inc., second edition, 1979.

[21] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings, CVPR '86 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, June 22–26, 1986)*, IEEE Publ.86CH2290-5, pages 364–374. IEEE, 1986.

[22] Hermann Weyl. *The Classical Groups: Their Invariants and Representations.* Princeton University Press, 1946.

[23] ECCV Workshop. *3D structure from Multiple Images of Large-Scale Environments.* in connection with ECCV'98, 1998.