CAMERA SELF-CALIBRATION: GEOMETRY AND ALGORITHMS

by

Yi Ma, Rene Vidal, Jana Kosecka and Shankar Sastry

Memorandum No. UCB/ERL M99/32

6 June 1999

CAMERA SELF-CALIBRATION: GEOMETRY AND ALGORITHMS

by

Yi Ma, Rene Vidal, Jana Kosecka and Shankar Sastry

Memorandum No. UCB/ERL M99/32

6 June 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Camera Self-Calibration: Geometry and Algorithms *

Yi Ma René Vidal Jana Košecká Shankar Sastry

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

June 6, 1999

Abstract

In this paper, a geometric theory of camera self-calibration is developed. The problem of camera self-calibration is shown to be equivalent to the problem of recovering an unknown (Riemannian) metric of an appropriate space. This observation leads to a new account of the necessary and sufficient condition for a unique calibration. Based on this understanding, we obtain a new and complete critical motion analysis without introducing a projective space. A complete list of geometric invariants associated to an uncalibrated camera is given. Due to a new characterization of fundamental matrices, the Kruppa equations are re-derived and directly associated to the basic (co)invariants of the uncalibrated camera. We study general questions about the solvability of the Kruppa equations and show that, in some special cases, the Kruppa equations can be renormalized so as to allow for linear self-calibration algorithms. A further study of these cases not only reveals generic difficulties in conventional self-calibration methods based on the nonlinear Kruppa equations, but also clarifies some incorrect results in the literature about the solutions of the Kruppa equations. Since Kruppa equations do not provide sufficient constraints on camera calibration, in this paper we give a complete account of exactly what is missing in Kruppa equations. Our results clearly resolve the discrepancy between the Kruppa equations and the necessary and sufficient condition for a unique calibration. Self-calibration for the differential case is also studied in the same geometric framework. It is shown that the intrinsic parameter space is reduced to the space of singular values of the intrinsic parameter matrix if only differential epipolar constraints are used. Simulation results are presented for evaluation of the performance of the proposed linear algorithms.

Key words: geometry of uncalibrated camera, camera self-calibration, invariants of uncalibrated camera, epipolar geometry, fundamental matrix, the Kruppa equations, Kruppa equation renormalization.

1 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

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

about the motion between frames and the structure of the observed scene. The general calibration problem is motivated by a variety of applications in mobile robot navigation and control using on-board computer vision system as a motion sensor. Many navigation or control tasks, such as target tracking, obstacle avoidance or map building, require the knowledge of both the camera (or the object) motion and 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 of camera self-calibration have recently received an increased interest in the computer vision and robotics community. The appeal of a successful solution to the camera self-calibration problem lies in the elimination of the need for an external calibration object [27] as well as the possibility of on-line calibration of time-varying internal parameters of the camera. The latter feature is of great importance for active vision systems. The majority of the camera self-calibration in the computer vision literature have been derived in a projective geometry framework. Here, we redevelop the theory in a differential geometric framework which enables not only new perspectives and algorithms but also a resolution of some mistreated problems in self-calibration.

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 [22]. The proposed approach and solution utilize invariant properties of the image of the so called absolute conic. Since the absolute conic is invariant under Euclidean transformations (*i.e.*, its representation 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 camera intrinsic parameter matrix. The constraints on the absolute conic are captured by the so called Kruppa equations derived by Kruppa in 1913.

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 [8]). This derivation is quite involved and the development appears to be rather unnatural since, both the constraints captured by Kruppa equations and the image of (dual) absolute conic are in fact directly linked to the invariants of the group of Euclidean transformation (rather than projective transformation). We here provide an alternative derivation of Kruppa equations, which in addition to being concise and elegant, also provides an intrinsic geometric interpretation of the so called fundamental matrices and its associated Kruppa equations. Such an interpretation is crucial for designing intrinsic optimization schemes for solving the problem (for example, see [17]).

In spite of the fact that the basic formulation of appropriate constraints, such as the Kruppa equations, is in place and there are many successful applications [29], to our knowledge, there is not yet a clear understanding of the geometry of an uncalibrated camera, and there is no complete analysis of the necessary and sufficient condition for a unique solution of the self-calibration problem. This often leads to situations where the algorithms are applied in ill-conditioned settings or where a unique solution is not even obtainable. The differential geometric approach we take in this paper will allow us to fully understand the intrinsic geometric characterization of an uncalibrated camera and it will easily lead to a clear answer to the questions:

> (i) *What is the necessary and sufficient condition for a unique solution of camera self-calibration? Do Kruppa equations provide sufficient conditions on the camera intrinsic parameters?*

The first question has been previously studied by [25]. However the analysis is incorrect since it makes a wrong assumption that one can at best recover the structure up to an arbitrary projective transformation from uncalibrated images [10]. Therefore, the results given in [25] are incorrect and have led to a misleading characterization of the necessary and sufficient condition for a unique solution of self-calibration (see Section 2.3 and 3.6 for a more detailed account). In this paper, we will give the necessary and sufficient condition in a very clear and compact form. Our results imply that, in principle, one can recover 3D Euclidean motion and structure up to a one parameter family from two uncalibrated images, as opposed to an arbitrary projective transformation [10]. Answer to the second question is unfortunately no, as counter examples have been given in the literature (e.g. [26]). Here we will give a complete account of exactly what is missing in the Kruppa equations. As we will see, there exist solutions of the Kruppa equations which do not allow any Euclidean reconstruction of the camera motion and scene structure. After excluding these solutions, solving Kruppa equations is then equivalent to the necessary and sufficient condition for a unique self-calibration.

One class of approaches to the design of self-calibration algorithms instead of directly using the Kruppa equations, solves for the entire projection matrices which are compatible with the camera motion and structure of the scene [9]. Such methods suffer severely from numerous local minima. Another class of approaches, as we have mentioned, directly utilizes the Kruppa equations which provide quadratic constraints in the camera intrinsic parameters. The so called epipolar constraint between a pair of images provides 2 such constraints, hence it usually requires the total of 3 pairs of images for a unique solution of all the 5 unknown parameters. The solution proposed to solve the Kruppa equations in the literature using homotopy continuation is quite computationally expensive and requires a good accuracy of the measurements [22]. Some alternative schemes have been explored in [12, 30]. It has been well-known that, in presence of noise, these Kruppa equation based approaches do not usually provide good recovery of the camera calibration [3]. Thus, it is important to answer:

> **(ii)** *Under what conditions can the Kruppa equations become degenerate or ill-conditioned? When such conditions are satisfied, how do the self-calibration algorithms need to be modified?*

The answer to the former question is rather unfortunate: for camera motions such that the rotation axis is parallel or perpendicular to the translation, the Kruppa equations are degenerate (in the sense that constriants provided are dependent); most practical image sequences are in fact taken through motions close to these two types. This explains why conventional approaches to self-calibration based on the (nonlinear) Kruppa equations usually fail when being applied to real image sequences. However, we further show in this paper that when such motions occur, the corresponding Kruppa equations can be "renormalized" and become linear. This gives us opportunities to design linear self-calibration algorithms besides the pure rotation case [9]. Our study also clarifies some incorrect analysis and results that exist in the literature regarding the solutions of the Kruppa equations [30]. This is discussed in Section 3.5.2.

It has been known that it is possible to develop a parallel set of theory and algorithms for recovering camera motion and scene structure for the discrete and differential (or continuous) cases [16, 17]. We therefore ask:

> **(iii)** *Whether there is a parallel theory and a set of algorithms of self-calibration for the discrete and differential cases?*

The answer is unfortunately no, as was previously pointed out by [4]. Due to certain degeneracy of the differential epipolar constraint, it is in general impossible to obtain a full calibration from it while, for the discrete case, full information of camera calibration is already available from the epipolar constraint only. In this paper, similarities and differences between the discrete and differential cases are unified in the same geometric framework.

Paper Outline. Section 2 studies the geometry of an uncalibrated camera system. It gives an intrinsic geometric interpretation of the camera self-calibration problem. The necessary and sufficient condition for a unique calibration follows from this interpretation. Section 3 studies practical schemes for camera self-calibration. As a theoretical foundation for the design of self-calibration algorithms, geometric invariants associated to an uncalibrated camera are studied in detail. In particular, we show that the (dual) absolute conics are generated by these basic invariants. Based on invariant theory, we provide a geometric characterization of the space of fundamental matrices. This characterization naturally associates the Kruppa equations with basic invariants of the uncalibrated camera. We then study several important cases which allow for linear self-calibration algorithms. These cases also reveal difficulties in the conventional Kruppa equation based approaches. Section 4 provides a brief study of the differential case, as a comparison to the theory of the discrete case. Some preliminary experiments of proposed algorithms are presented in Section 5.

2 Geometry for the Uncalibrated Camera

2.1 Uncalibrated Camera Motion and Projection Model

We begin by 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 symbols (e.g. $\underline{q}$ vs. 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). Arbitrary vectors $u \in T_qM$ in homogeneous representation have 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 (metric preserving diffeomorphism) 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:

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

where $SO(3)$ is the space of 3×3 rotation matrices (orthogonal matrices with determinant $+1$). We know the isotropy group of M leaving a point q fixed is the orthogonal group $O(3)$. $SO(3)$ is the subgroup of $O(3)$ which is the connected component containing 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 $h(t) \in SE(3), t \in \mathbb{R}$, 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 those 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 $\mathbb{R}^3$, the above is simply:

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

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 has 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 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 $\mathbf{x}$. For perspective projection $\lambda = q_3$; for spherical projection $\lambda = \|q\|$. The equation (12) characterizes the mathematical model of an ideal calibrated camera. Figure 1 illustrates the images of a point q with the camera at two different locations. For a study of calibrated camera, one may refer to [21, 15, 17, 20].

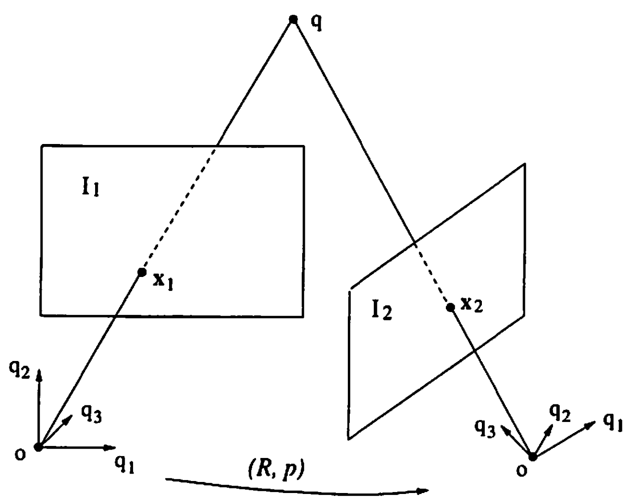

Figure 1: Two projections $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ of a 3D point q from two vantage points. The relative Euclidean transformation is given by $(R, p) \in SE(3)$.

In this paper, we are going to study an uncalibrated camera. By an uncalibrated camera, we mean that the image received by the camera is distorted by an unknown linear transformation.[1] This linear transformation is usually assumed to be invertible. Mathematically, this linear transformation is an isomorphism ϕ of the vector space T_oM:

$$\begin{aligned} \phi : T_oM &\to T_oM \\ u &\mapsto Au, \end{aligned}$$

where $A \in \mathbb{R}^{3\times 3}$ is an invertible matrix representing the linear map ϕ. We will refer to it as the **calibration matrix**[2] of an uncalibrated camera. 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 the image

[1]Although nonlinear transformations have also been studied in the literature, linear transformations give a very good model of the physical parameters of a camera.

[2]"Calibration matrix", "intrinsic parameter matrix" and "intrinsic parameters" are different names of the same thing.

$\mathbf{x} \in \mathbb{R}^3$ 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, as illustrated in Figure 2.

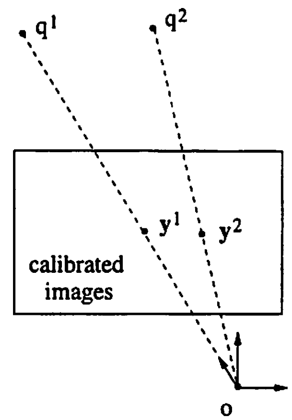

Figure 2: The actually received uncalibrated images $\mathbf{x}^1, \mathbf{x}^2 \in \mathbb{R}^3$ of two 3D points q^1 and q^2. We here use $\mathbf{y}^1, \mathbf{y}^2 \in \mathbb{R}^3$ to represent the calibrated images (with respect to a normalized coordinate system). The linear map ϕ expresses the transformation between the calibrated and uncalibrated images.

Comments 1 *The calibration matrix A is frequently assumed in the literature 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 of the matrix A are called "intrinsic parameters" associated with the camera. Note that such an A is not necessarily in $SL(3)$. As we will soon see, this choice is practically equivalent to ours. Moreover, viewing camera calibration as an (unknown) isomorphism on T_oM makes it quite natural to generalize the theory for the Euclidean space to any other Riemannian space (see [18]).

If we know the linear transformation ϕ, *i.e.*, the calibration matrix A, then the problems associated to an uncalibrated camera can be reduced to those of a calibrated camera, which have been well understood. The central goal of this paper is hence to study the

> **Camera Self-calibration Problem:** *from only the image measurements* $\mathbf{x}$ *of a cloud of 3D points taken by an uncalibrated camera at different vantage points, to what extent can we recover the unknown camera calibration, i.e., the linear transformation ϕ or the matrix A, and how?*

2.2 Intrinsic Geometric Interpretation for Camera Calibration

Before trying to solve the camera self-calibration problem, we first need to know some geometric properties of an uncalibrated camera: we will see that the study of an uncalibrated camera is equivalent to that of a calibrated camera in a (Euclidean) space with an unknown metric. Further, the problem of recovering the calibration matrix A is mathematically equivalent to that of recovering this unknown metric. Consequently, the camera intrinsic parameters given in (14) can be geometrically characterized as the space $SL(3)/SO(3)$. Some elementary Riemannian geometry notation will be used here. For good references on Riemannian geometry, we refer the reader to [2, 11, 24].

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 $S \in \mathbb{R}^{3\times 3}$ 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-triangular 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 (by the Cholesky factorization). 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:

$$Aq(t) = ARq(t_0) + Ap(t) \quad \Leftrightarrow \quad BR_0q(t) = BR_0R(t)R_0^TR_0q(t_0) + BR_0p(t). \tag{19}$$

Notice that the conjugation:

$$\begin{array}{rcl} \mathrm{Ad}_r : SE(3) & \rightarrow & SE(3) \\ h & \mapsto & rhr^{-1} \end{array}$$

is a group homomorphism where $r = \begin{pmatrix} R_0 & 0 \\ 0 & 1 \end{pmatrix}$. Then there is no way to tell an uncalibrated camera with calibration matrix A undergoing the motion $(R(t), p(t))$ and observing the point $q \in M$ from another uncalibrated camera with calibration matrix B undergoing 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) 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-triangular representation of the intrinsic parameters can be easily obtained from Cholesky factorization.

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:

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

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[3] 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:

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

[3]The isometry group of a space M is the set of all transformations which preserve metric (or distance).

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 M' assuming a calibrated camera.

2.3 Necessary and Sufficient Condition for Unique Calibration

It is of great importance to know under what conditions the unknown camera calibration A (as an element in $SL(3)/SO(3)$), or the metric S of M' can be uniquely recovered. Mathematically, we can interpret the isometry group G of M' as a representation of the Euclidean group $SE(3)$ induced by the map ψ. Any element $h \in SE(3)$ is hence represented by a corresponding element $h' \in G$ given by the conjugation:

$$h' \;=\; aha^{-1} \tag{25}$$

where $a = \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{4\times4}$. In the following, unless otherwise stated, we always use a to denote this 4×4 matrix associated with $A \in SL(3)$. Another useful notation we introduce here is that, for an arbitrary vector $p = (p_1, p_2, p_3)^T \in \mathbb{R}^3$, we define the skew-symmetric matrix $\widehat{p} \in \mathbb{R}^{3\times3}$ associated to p as:

$$\widehat{p} = \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix} \in so(3). \tag{26}$$

Then for another vector $l \in \mathbb{R}^3$, the cross-product $p \times l$ is equal to $\widehat{p}l$. This notation will be frequently used in the rest of this paper.

Suppose that the camera motion is given by a subset W of $SE(3)$. If there exists an $A \in SL(3)/SO(3)$ different from I such that $W' = aWa^{-1}$ is also a subset of $SE(3)$, then there is no way one can tell a calibrated camera undergoing motion W' from an uncalibrated camera with calibration matrix A undergoing motion W. On the other hand, if, for a given $W \subset SE(3)$, there is no $A \in SL(3)/SO(3)$ such that $aWa^{-1} \subset SE(3)$, then the only possible case is that the camera is calibrated and undergoing motion W. This leads to the following definition:

Definition 1 (Critical motion) *A set of camera motion $W \subset SE(3)$ is called* **critical** *for self-calibration if and only if there exists an $A \in SL(3)/SO(3)$ different from I such that aWa^{-1} is also a subset of $SE(3)$.*

Comments 2 *If a camera with a calibration matrix A undergoing motion W is confused with B undergoing motion W', it is equivalent to a camera $B^{-1}A$ undergoing motion W being confused with a calibrated camera (of calibration matrix I) undergoing motion W'. Thus the above definition* *does not result in any loss of generality. One must note that this definition has little to do with fundamental matrix or Kruppa equations. A clear relationship between this definition and Kruppa equations will be given in Section 3.6 when we study self-calibration algorithms.*

Now, finding the necessary and sufficient condition for a unique calibration is equivalent to characterizing all critical motions. Note that for

$$h = \begin{pmatrix} R & p \\ 0 & 1 \end{pmatrix} \quad \in W \tag{27}$$

we have

$$h' = \begin{pmatrix} ARA^{-1} & Ap \\ 0 & 1 \end{pmatrix}. \tag{28}$$

h' is in $SE(3)$ if and only if ARA^{-1} is an element in $SO(3)$. We then have

$$ARA^{-1}(A^{-T}R^TA^T) = I \quad \Leftrightarrow \quad RXR^T = X \tag{29}$$

where $X = A^{-1}A^{-T}$. Thus X has to be in the **symmetric real kernel** of the Lyapunov map:

$$\begin{aligned} L : \mathbb{C}^{3\times 3} &\rightarrow \mathbb{C}^{3\times 3} \\ X &\mapsto X - RXR^T. \end{aligned} \tag{30}$$

We will denote this kernel as SRKer(L). According to Callier and Desoer [5], the map L 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 R. Without loss of generality, the rotation matrix R has eigenvalues $1, \alpha, \bar{\alpha} \in \mathbb{C}$ and corresponding right eigenvectors $u, v, \bar{v} \in \mathbb{C}^3$. Then the (complex) kernel of L is given by:

$$\text{Ker}(L) \;=\; \text{span}\{X_1 = uu^*, X_2 = vv^*, X_3 = \bar{v}\bar{v}^*\} \quad \subset \mathbb{C}^{3\times 3} \tag{31}$$

where, for a vector $v \in \mathbb{C}^3$, $\bar{v}$ is its conjugate and v^* is its conjugate transpose. We assume here R is neither the identity matrix I or a 180^o rotation, *i.e.*, R is not of the form $e^{\hat{u}k\pi}$ for some $k \in \mathbb{Z}$ and some $u \in \mathbb{R}^3$ of unit length. Then only X_1 is real and $X_2 = \bar{X}_3$ are complex, and L has a three dimensional real kernel but one dimension is spanned by $i(X_2 - X_3)$ which is skew-symmetric (here $i = \sqrt{-1}$). Therefore, the solution space for a symmetric real X is 2 dimensional and must have the form $X = \beta X_1 + \gamma(X_2 + X_3)$ with $\beta, \gamma \in \mathbb{R}$. Summarizing the above we obtain:

Lemma 1 *Given a rotation matrix R not of the form $e^{\hat{u}k\pi}$ for some $k \in \mathbb{Z}$ and some $u \in \mathbb{R}^3$ of unit length, the symmetric real kernel associated with the Lyapunov map $L : X \mapsto X - RXR^T$ is 2 dimensional. If R is of the form $e^{\hat{u}k\pi}$, then SRKer(L) is 4 dimensional if k is odd and 6 dimensional if k is even.*

Note that the case when the rotation is 180^o has no practical significance in real situations, since no image correspondences are available in this case. Thus, from now on we may assume that

Assumption 1 *All rotations that we consider for the camera self-calibration problem are strictly less than 180^o unless otherwise stated.*

As a direct consequence of Lemma 1, any $W \subset SE(3)$ with only one element is critical. Suppose, instead, there are n elements $h_j, j = 1, \ldots n$ in W. For aWa^{-1} to be in $SE(3)$, $X = A^{-1}A^{-T}$ has to be in the intersection of symmetric real kernels of all the linear maps:

$$\begin{array}{rcll} L_j : \mathbb{C}^{3\times 3} & \to & \mathbb{C}^{3\times 3}, & j = 1, \ldots, n \\ X & \mapsto & X - R_j X R_j^T. & \end{array} \tag{32}$$

That is $X \in \cap_{j=1}^{n} \mathrm{SRKer}(L_j)$.

Theorem 1 (Necessary and sufficient condition for a unique calibration) *A motion subset $W \subset SE(3)$ is not critical for self-calibration if and only if there are at least* **two** *elements of W whose rotation axes are linearly independent.*

Proof: The necessity is obvious: if two rotation matrices R_1 and R_2 have the same axis, they have the same eigenvectors hence $\mathrm{SRKer}(L_1) = \mathrm{SRKer}(L_2)$ where $L_i : X \mapsto X - R_i X R_i^T, i = 1, 2$. We now only need to prove the sufficiency. We may assume u_1 and u_2 are the two rotation axes of R_1 and R_2 respectively and are linearly independent. Since, by assumption 1, both R_1 and R_2 considered are not 180^o rotation, both $\mathrm{SRKer}(L_1)$ and $\mathrm{SRKer}(L_2)$ are 2 dimensional. Since u_1 and u_2 are linearly independent, the matrices $u_1 u_1^T$ and $u_2 u_2^T$ are linearly independent and are in $\mathrm{SRKer}(L_1)$ and $\mathrm{SRKer}(L_2)$ respectively. Thus $\mathrm{SRKer}(L_1)$ is not fully contained in $\mathrm{SRKer}(L_2)$ hence their intersection $\mathrm{SRKer}(L_1) \cap \mathrm{SRKer}(L_2)$ has at most 1 dimension. Thus $X = I$ for $X \in SL(3)$. ■

From the above discussion, criticality of a motion set only depends on its rotation components.

Corollary 1 (Critical Lie subgroups) *Any proper Lie subgroup of $SE(3)$ except $SO(3)$ is a critical subgroup for self-calibration.*

According to this corollary, if the motion of the camera falls into any of the Lie subgroups of $SE(3)$, unique self-calibration is impossible. For a more detailed analysis of to what extend we can still recover camera calibration, motion and scene structure with respect to any of the Lie subgroups of $SE(3)$, one may refer to [20].

Although it has little practical importance, in order to make the theory complete, we also give the results of self-calibration in presence of rotation of 180^o (for simplicity, we here do not give the proof). Combined with Theorem 1, they give necessary and sufficient conditions for a unique calibration in the most general case.

Remark 1 *Suppose $R_i = e^{\widehat{u}_i \theta_i}, i = 1, 2$ are elements in W. u_i are vectors of unit length. Let L_i be the Lyapunov map associated to R_i. Then we have the following cases:*

$$\begin{array}{rcl} u_1^T u_2 = 0, \ \ |\theta_1| = |\theta_2| = \pi & \Rightarrow & SRKer(L_1) \cap SRKer(L_2) = span\{I, \ u_1 u_1^T, \ u_2 u_2^T\}, \\ 0 < |u_1^T u_2| < 1, |\theta_1| = |\theta_2| = \pi & \Rightarrow & SRKer(L_1) \cap SRKer(L_2) = span\{I, \widehat{u}_2 u_1 u_1^T \widehat{u}_2\}, \\ u_1^T u_2 = 0, \ \ |\theta_1| = \pi, \ 0 < |\theta_2| < \pi & \Rightarrow & SRKer(L_1) \cap SRKer(L_2) = span\{I, u_2 u_2^T\}, \\ 0 < |u_1^T u_2| < 1, |\theta_1| = \pi, 0 < |\theta_2| < \pi & \Rightarrow & SRKer(L_1) \cap SRKer(L_2) = span\{I\}. \end{array}$$

Comments 3 *Definition 1 and (the derivation of) Theorem 1 apparently contradict the results given in [25]. The analysis in [25] is based on a stratification scheme of "projective to affine and* *then to Euclidean". As far as the self-calibration problem is concerned, assuming such a scheme has led to a circular logic: the goal of the study is exactly to discover to what extent one can recover structure from uncalibrated images. In fact, the very first step of this stratification scheme is already wrong: the possibility of self-calibration suggests that one can certainly do better than an "arbitrary" projective reconstruction. Since the assumption of projective reconstruction is wrong, there is no need for a discussion of absolute conics not in the plane at infinite. The list of "critical motion sequences" given in [25] is, therefore, not a correct characterization of all critical camera motions which do not guarantee a unique camera self-calibration. It is by no means intrinsic to the self-calibration problem.*[4] *Later in this paper (see Section 3.6), in a language of fundamental matrices and Kruppa equations, we will give a clear characterization of solutions of valid Euclidean reconstruction from (two) uncalibrated images. It will not only justify again Definition 1 but also suggest a new "stratification" scenario.*

3 Camera Self-calibration Algorithms

Although Theorem 1 has established conditions under which a unique self-calibration is guaranteed, the proof does not yet provide any algorithm for recovering the unknown calibration A or the metric S of the space M'. This will be the task for the rest of the paper. That is, we will be looking for algorithms which allow us to recover information of calibration from certain image measurements.

3.1 Geometric Invariants Associated to Uncalibrated 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. This section will give a complete account of these invariants. 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 in one-to-one correspondence to the invariants of M under the Euclidean group. The complete list of Euclidean invariants is given by the following theorem:

Theorem 2 (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 [28] for a proof of this theorem and see [14] 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 2 (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 $u^1, u^2, u^3 \in TM'$.*

[4] For example, the class 5 motion is certainly not critical at all.

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

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

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:

$$e_j(e^i) = \delta_{ij} \tag{33}$$

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:

$$\begin{aligned} <\cdot,\cdot>: V^\vee \times V \quad &\rightarrow \quad \mathbb{R} &(34)\\ (u, v) \quad &\mapsto \quad u(v). &(35) \end{aligned}$$

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

$$< u, f(v) >=< f^\vee(u), v >, \quad \forall u \in V^\vee, v \in V. \tag{36}$$

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:

$$< u, f(v) >= u^T A v = (A^T u)^T v, \tag{37}$$

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 2 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 with 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 3 (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 enumerated by superscript and covectors by subscript. One may also refer to Weyl [28] or Goodman and Wallach [6] for a detailed study of polynomial invariants of classical groups - Corollary 2 and 3 can then be deduced from the First Fundamental Theorem of groups $G \subset GL(V)$ preserving a non-degenerate (symmetric) form (see [6]). 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 need to introduce the space $\mathbb{CP}^3$, the three dimensional complex projective space[5]. 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:

$$q_1^2 + q_2^2 + q_3^2 = 0, \quad q_4 = 0 \tag{38}$$

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

$$E(3, \mathbb{C}) = \left\{ \begin{pmatrix} R & p \\ 0 & 1 \end{pmatrix} \middle| \, p \in \mathbb{C}^3, R \in U(3) \right\} \subset \mathbb{C}^{4\times4} \tag{39}$$

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

$$q_j = u_j + iv_j, \quad u_j, v_j \in \mathbb{R}, \quad j = 1, \dots, 4 \tag{40}$$

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 (38), these two vectors satisfy:

$$u^T u = v^T v, \quad u^T v = 0 \tag{41}$$

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 (41) 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 (38) is then given by:

$$(q_1, q_2, q_3)S(q_1, q_2, q_3)^T = 0, \quad q_4 = 0. \tag{42}$$

Therefore, the camera self-calibration problem is also equivalent to the problem of recovering this absolute conic (for example see Maybank [21]). It is direct to check that this absolute conic is

[5] $\mathbb{CP}^3$ 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 2. Define the dual absolute conic $\Omega^\vee$ to be the set of points in $\mathbb{CP}^3$ satisfying:

$$(q_1, q_2, q_3)S^{-1}(q_1, q_2, q_3)^T = 0, \quad q_4 = 0. \tag{43}$$

Similarly, one can show that it is generated by the inner product type coinvariants given in Corollary 3.

3.2 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. Section 3.2 and 3.3 are going to show that fundamental matrices estimated from the epipolar constraint are in fact covectors. Section 3.4 shows that their associated coinvariants directly give the Kruppa equations.

The epipolar (or Longuet-Higgins) constraint plays an important role in the study of the geometry of calibrated cameras. In this section, we study its uncalibrated version. From (20), for a point $q' \in M'$ we have:

$$\begin{aligned} & q'(t) = AR(t)A^{-1}q'(t_0) + p'(t) \quad \Rightarrow \quad p'(t) \times q'(t) = p'(t) \times AR(t)A^{-1}q'(t_0) \\ \Rightarrow \quad & q'(t_0)^T A^{-T} R(t)^T A^T \widehat{p'(t)} q'(t) = 0. \end{aligned} \tag{44}$$

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 (44) yields:

$$\mathbf{x}_1^T A^{-T} R^T A^T \widehat{p'} \mathbf{x}_2 = 0. \tag{45}$$

Note that in the above equation the matrix:

$$F_1 = A^{-T} R^T A^T \widehat{p'} \quad \in \mathbb{R}^{3\times 3} \tag{46}$$

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:

$$\begin{aligned} & A^{-1}q'(t) = R(t)A^{-1}q'(t_0) + p(t) \quad \Rightarrow \quad p(t) \times A^{-1}q'(t) = p(t) \times R(t)A^{-1}q'(t_0) \\ \Rightarrow \quad & q'(t_0)^T A^{-T} R(t)^T \widehat{p(t)} A^{-1} q'(t) = 0 \end{aligned} \tag{47}$$

We then have a second form for the constraint given in (45):

$$\mathbf{x}_1^T A^{-T} R^T \widehat{p} A^{-1} \mathbf{x}_2 = 0. \tag{48}$$

The matrix

$$F_2 = A^{-T} R^T \widehat{p} A^{-1} \quad \in \mathbb{R}^{3\times 3} \tag{49}$$

is called the **fundamental matrix** in the computer vision literature. When $A = I$, the fundamental matrix simply becomes $R^T\widehat{p}$ which is called the **essential matrix** in the literature and plays a very important role in motion recovery [21]. In fact, the two constraints (45) and (48) 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 2 *If $p \in \mathbb{R}^3$ and $A \in SL(3)$, then $A^T\widehat{p}A = \widehat{A^{-1}p}$.*

Proof: Since both $A^T\widehat{(\cdot)}A$ and $\widehat{A^{-1}(\cdot)}$ are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^{3\times3}$, 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 $(0,0,1)^T$. ∎

This simple lemma will be frequently used throughout the paper. By this lemma, we have:

$$F_2 = A^{-T}R^T\widehat{p}A^{-1} = A^{-T}R^TA^TA^{-T}\widehat{p}A^{-1} = A^{-T}R^TA^T\widehat{p'} = F_1. \tag{50}$$

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:

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

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

In the preceding section, we have shown 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:

$$B^{-T}R^T\widehat{p}B^{-1} = A^{-T}R_0R^T\widehat{p}R_0^TA^{-1} = A^{-T}(R_0R^TR_0^T)\widehat{R_0p}A^{-1}. \tag{52}$$

As we noticed, the essential matrix $R^T\widehat{p}$ is simply replaced by another essential matrix $(R_0R^TR_0^T)\widehat{R_0p}$. Therefore, without knowing the camera motion, from only the fundamental matrix, one cannot tell camera B from camera A.

3.3 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 [12]), but it directly connects a fundamental matrix with its Kruppa equation, as we will soon see in Section 3.4.

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

$$g(B,C) \;=\; \mathrm{tr}(B^TSC), \quad \forall B,C \in \mathbb{R}^{3\times3} \tag{53}$$

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 (53), $C^\vee$ 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:

$$g^\vee(B^\vee, C^\vee) = g(B, C) = \operatorname{tr}((B^\vee)^T S^{-1} C^\vee). \tag{54}$$

A tangent vector of the Lie group $ASO(3)A^{-1}$ has the form $AR^T\widehat{p}A^{-1} \in \mathbb{R}^{3\times3}$ 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 $ASO(3)A^{-1}$:

$$g(AR^T\widehat{p}_1A^{-1}, AR^T\widehat{p}_2A^{-1}) = g(A\widehat{p}_1A^{-1}, A\widehat{p}_2A^{-1}). \tag{55}$$

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 $AR^T\widehat{p}A^{-1} \in T(ASO(3)A^{-1})$ is given by:

$$(AR^T\widehat{p}A^{-1})^\vee = SAR^T\widehat{p}A^{-1} = A^{-T}R^T\widehat{p}A^{-1}. \tag{56}$$

Note that the matrix $A^{-T}R^T\widehat{p}A^{-1}$ 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:

$$g^\vee(A^{-T}R^T\widehat{p}_1A^{-1}, A^{-T}R^T\widehat{p}_2A^{-1}) = g^\vee(\widehat{p'_1}, \widehat{p'_2}) \tag{57}$$

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 unit fundamental matrices to be:

$$\mathcal{F}_1 = \{A^{-T}R^T\widehat{p}A^{-1} \mid R \in SO(3), p \in \mathbb{R}^3, g^\vee(\widehat{Ap}, \widehat{Ap}) = 1\}. \tag{58}$$

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

Theorem 3 (Geometric characterization of fundamental space) *The unit cotangent bundle $T_1^*(ASO(3)A^{-1})$ is a double covering of the unit 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 space of (normalized) essential matrices (see[16]). 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 4 *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.*

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 $S = A^{-T}A^{-1}$.

3.4 The Kruppa Equations

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 (45) 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 $F = A^{-T}R^T\widehat{p}A^{-1}$. 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 **normalized fundamental matrix** (to be separated from the unit 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:

$$e_1 = (1,0,0)^T, \quad e_2 = (0,1,0)^T, \quad e_3 = (0,0,1)^T \quad \in \mathbb{R}^3. \tag{59}$$

Now pick any rotation matrix $R_0 \in SO(3)$ such that $R_0p' = e_3$. Using Lemma 2, we have:

$$\widehat{p'} = R_0^T\widehat{e_3}R_0. \tag{60}$$

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

$$D = R_0FR_0^T = (R_0A)^{-T}R^T(R_0A)^T\widehat{e_3}. \tag{61}$$

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:

$$d_1 = (R_0A)^{-T}R^T(R_0A)^Te_2, \quad d_2 = -(R_0A)^{-T}R^T(R_0A)^Te_1. \tag{62}$$

Define matrix $K = R_0A \in SL(3)$. Note that (62) is in the form of a transformation on covectors that we discussed in Section 3.1. According to Corollary 3, coinvariants of the group $KSO(3)K^{-1}$ (*i.e.*, the invariants of the dual group $K^{-T}SO(3)K^T$) give:

$$\begin{array}{rcl} (d_1)^TKK^Td_1 & = & (e_2)^TKK^Te_2, \\ (d_2)^TKK^Td_2 & = & (e_1)^TKK^Te_1, \\ (d_1)^TKK^Td_2 & = & -(e_2)^TKK^Te_1. \end{array} \tag{63}$$

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:

$$\xi_1 = R_0^Td_1, \quad \xi_2 = R_0^Td_2; \quad \eta_1 = -R_0^Te_1, \quad \eta_2 = R_0^Te_2,$$

then (63) directly gives constraints on S^{-1} as:

$$\begin{array}{rcl} \xi_1^TS^{-1}\xi_1 & = & \eta_2^TS^{-1}\eta_2, \\ \xi_2^TS^{-1}\xi_2 & = & \eta_1^TS^{-1}\eta_1, \\ \xi_1^TS^{-1}\xi_2 & = & \eta_1^TS^{-1}\eta_2. \end{array} \tag{64}$$

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 normalized, *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 (64) is equal to λ^2. This gives two independent constraints on S^{-1}, the so called **Kruppa equations** (due to their first discovery by Kruppa in 1913):

$$\lambda^2 = \frac{\xi_1^T S^{-1} \xi_1}{\eta_2^T S^{-1} \eta_2} = \frac{\xi_2^T S^{-1} \xi_2}{\eta_1^T S^{-1} \eta_1} = \frac{\xi_1^T S^{-1} \xi_2}{\eta_1^T S^{-1} \eta_2}. \tag{65}$$

Alternative means of obtaining the Kruppa equations are by utilizing algebraic relationships between projective geometric quantities [22] or via SVD characterization of F [8]. Here we obtain the same equations from a quite different approach. Equation (65) 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. In general, each fundamental matrix provides at most two algebraic constraints on S^{-1}. Since the symmetric matrix S has five degrees of freedom, in general at least three fundamental matrices are needed to uniquely determine S.

Comments 5 *One must be aware that solving Kruppa equations for camera calibration is not equivalent to the camera self-calibration problem in the sense that there may exist solutions of Kruppa equations which are not solutions of a "valid" self-calibration. Given a non-critical set of camera motions, the associated Kruppa equations do not necessarily give enough constraints to solve for the calibration matrix A. See Section 3.6 for a complete account.*

The above derivation of Kruppa equations is straightforward, but the expression (65) depends on a particular rotation matrix R_0 that one chooses – note that the choice of R_0 is not unique. However, there is an even simpler way to get an equivalent expression for the Kruppa equations in a matrix form. Given a normalized 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 3, one directly gets the matrix equation:

$$F^T S^{-1} F = \widehat{p'}^T S^{-1} \widehat{p'}. \tag{66}$$

We call this equation the **normalized matrix Kruppa equation**. It is readily seen that this equation is equivalent to (64). If F is not normalized 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**:

$$F^T S^{-1} F = \lambda^2 \widehat{p'}^T S^{-1} \widehat{p'}. \tag{67}$$

This equation is equivalent to the scalar version given by (65) and is independent of the choice of the rotation matrix R_0.

3.5 Solving the Kruppa Equations

Algebraic properties of Kruppa equations have been previously studied in [22, 30]. However, conditions on dependences among Kruppa equations obtained from a fundamental matrix have not been described systematically. Therefore it is hard to tell in practice whether a given set of Kruppa equations provide a unique solution for calibration. As we will soon see in this section, for a very rich class of camera motions (which commonly occur in many practical applications), the Kruppa equations become degenerate. Moreover, since the (matrix) Kruppa equations (65) or (67) are highly nonlinear in S^{-1}, most self-calibration algorithms which are based on directly solving the Kruppa equations suffer from being computationally expensive or having multiple local minima [13, 3]. These reasons have motivated us to study the geometric nature of Kruppa equations and to gain a better understanding of the difficulties commonly encountered in camera self-calibration. Our attempts to resolve these difficulties will lead to simplified algorithms for self-calibration. These algorithms are linear and better conditioned for a specific class of camera motion.

Given a fundamental matrix $F = A^{-T}R^TA^T\widehat{p'}$ with p' of unit length, the normalized matrix Kruppa equation (66) can be rewritten in the following way:

$$\widehat{p'}^T(S^{-1} - ARA^{-1}S^{-1}A^{-T}R^TA^T)\widehat{p'} = 0. \tag{68}$$

According to this form, if we define $C = A^{-T}R^TA^T$, a linear (Lyapunov) map $\sigma : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ as $\sigma : X \mapsto X - C^TXC$, and a linear map $\tau : \mathbb{R}^{3\times3} \to \mathbb{R}^{3\times3}$ as $\tau : Y \mapsto \widehat{p'}^T Y\widehat{p'}$, then the solution S^{-1} of equation (68) is exactly the (symmetric real) kernel of the composition map:

$$\tau \circ \sigma : \quad \mathbb{R}^{3\times3} \xrightarrow{\sigma} \mathbb{R}^{3\times3} \xrightarrow{\tau} \mathbb{R}^{3\times3}. \tag{69}$$

This interpretation of Kruppa equations clearly decomposes effects of the rotational and translational parts of the motion: if there is no translation *i.e.*, $p = 0$, then there is no map τ; if the translation is non-zero, the kernel is enlarged by composing the map τ. In general, the symmetric real kernel of the composition map $\tau \circ \sigma$ is 3 dimensional – while the kernel of σ is only 2 dimensional, due to Lemma 1.

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

The proof is trivial. This simple lemma, however, states a very important fact: given a set of fundamental matrices $F_i = A^{-T}R_i^TA^T\widehat{p_i'}$ 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:

$$F_i^TXF_i = \lambda_i^2\widehat{p_i'}^TX\widehat{p_i'}, \quad i = 1, \ldots, n. \tag{70}$$

and the set of solutions of the equations:

$$E_i^TYE_i = \lambda_i^2\widehat{p}_i^TY\widehat{p}_i, \quad i = 1, \ldots, n \tag{71}$$

where $E_i = R_i^T\widehat{p}_i$ 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 only depend on the camera motion.

3.5.1 Pure Rotation Case

From the decomposition of the (normalized) Kruppa equation given by (69), it is natural to first study the kernel of the map σ. This corresponds to the case that the camera undergoes pure rotational motion. This case has been previously studied in the vision literature [9]. Here, we give a brief review of the main results associated with this case.

In the pure rotation case, corresponding image pairs $(\mathbf{x}_1^j, \mathbf{x}_2^j), j = 1, \ldots, n$ satisfy:

$$\lambda_2^j \mathbf{x}_2^j = ARA^{-1}\lambda_1^j \mathbf{x}_1^j, \tag{72}$$

for some scales $\lambda_1^j, \lambda_2^j, j = 1, \ldots, n$. Then the image correspondences satisfy a degenerate version of epipolar constraint:

$$\widehat{\mathbf{x}}_2^j ARA^{-1}\mathbf{x}_1^j = 0 \tag{73}$$

for $j = 1, \ldots, n$. From these linear equations, in general, 4 pair of image correspondences 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 $C \in \mathbb{R}^{3\times 3}$, 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:

$$S - C^T SC = 0. \tag{74}$$

That is, S has to be in the symmetric real kernel of the Lyapunov map:

$$\begin{aligned} L : \mathbb{C}^{3\times 3} &\rightarrow \mathbb{C}^{3\times 3} \\ X &\mapsto X - C^T XC. \end{aligned} \tag{75}$$

As a direct corollary to Theorem 1, we have:

Corollary 4 *Given two matrices $C_j = AR_jA^{-1} \in ASO(3)A^{-1}, j = 1,2$ where $R_j = e^{\widehat{u}_j\theta_j}$ with θ_j's not equal to $k\pi, k \in \mathbb{Z}$, then $SRKer(L_1) \cap SRKer(L_2)$ is 1 dimensional if and only if u_1 and u_2 are linearly independent.*

According to this corollary, 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.

3.5.2 Fundamental Matrix and Kruppa Equation Renormalization

In a more general situation when the translational motion is present, the problem of solving Kruppa equations is no longer as easy as in the pure rotation case. From the derivation of the Kruppa equations (65) or (67), we observe that the reason that they are nonlinear is because we usually do not know the scale λ. It will be helpful to know under what conditions the matrix Kruppa equation has the same solution as the normalized one. Here we will study two special cases for which we are able to know directly what the missing λ is for the fundamental matrix. Therefore, the fundamental matrix can be **renormalized** and we can solve the camera calibration from the normalized matrix Kruppa equations (which are linear). These two cases are when the rotation axis is parallel or perpendicular to the translation. That is, if the motion is represented by $(R, p) \in SE(3)$ and $u \in \mathbb{R}^3$ is the axis of R, then the two cases are when u is parallel or perpendicular to p. As we will soon see, these two cases are of great theoretical importance: not only the calibration algorithms become linear, but also they reveal certain subtleties of the Kruppa equations and explain when the nonlinear Kruppa equations are most likely to become ill-conditioned.

Theorem 4 (Kruppa equation renormalization) *Consider a camera motion $(R,p) \in SE(3)$ where $R = e^{\widehat{u}\theta}$ with u of unit length. If $0 < \theta < \pi$ and u is parallel or perpendicular to p, then the matrix Kruppa equation: $\widehat{p}^T RYR^T\widehat{p} = \lambda^2\widehat{p}^T Y\widehat{p}$ has the same positive definite solutions of Y as the normalized matrix Kruppa equation: $\widehat{p}^T RYR^T\widehat{p} = \widehat{p}^T Y\widehat{p}$.*

Proof: For the parallel case, let $x \in \mathbb{R}^3$ be a vector of unit length in the plane spanned by column vectors of $\widehat{p}$. All such x are on a unit circle. There exists $x_0 \in \mathbb{R}^3$ on the circle such that $x_0^T Y x_0$ is maximum. We then have $x_0^T RYR^T x_0 = \lambda^2 x_0^T Y x_0$, hence $\lambda^2 \leq 1$. Similarly, if we pick x_0 such that $x_0^T Y x_0$ is minimum, we have $\lambda^2 \geq 1$. Therefore, $\lambda^2 = 1$. For the perpendicular case, since the columns of $\widehat{p}$ span the subspace which is perpendicular to the vector p, the eigenvector u of R is in this subspace. Thus we have: $u^T RYR^T u = \lambda^2 u^T Y u \Rightarrow u^T Y u = \lambda^2 u^T Y u$. Hence $\lambda^2 = 1$ if Y is positive definite. ∎

Under the conditions given by the theorem, there is no solution for λ in the Kruppa equation (67) besides the true scale of the fundamental matrix. The following lemma allows to directly compute this λ for such a fundamental matrix:

Lemma 4 *Given a fundamental matrix $F = \lambda A^{-T} R^T A^T \widehat{p'}$ with $\|p'\| = 1$, if $p = A^{-1}p'$ is parallel to the axis of R, then λ^2 is $\|F\widehat{p'}F^T\|$,[6] and if p is perpendicular to the axis of R, then λ is one of the two non-zero eigenvalues of the matrix $F\widehat{p'}^T$.*

Proof: First we prove for the parallel case. It is straightforward to check that, in general, $F\widehat{p'}F^T = \lambda^2 \widehat{AR^Tp}$. Since now the axis of R is parallel to p, we have $F\widehat{p'}F^T = \lambda^2\widehat{p'}$. For the perpendicular case, let $u \in \mathbb{R}^3$ be the axis of R and let $p = A^{-1}p'$. By assumption p is perpendicular to u. Then there exists $v \in \mathbb{R}^3$ such that $u = \widehat{p}A^{-1}v$. Then it is direct to check that $\widehat{p'}v$ is the eigenvector of $F\widehat{p'}^T$ corresponding to the eigenvalue λ. ∎

Then each fundamental matrix can be immediately normalized by dividing by the scale λ. Once the fundamental matrices are normalized, the problem of solving the calibration matrix S^{-1} from normalized matrix Kruppa equations becomes a simple **linear** one! A normalized matrix Kruppa equation in general imposes 3 linearly independent constraints given by (64) on the unknown calibration. However, this is no longer the case for the special motions that we are considering here.

Lemma 5 *Consider a camera motion $(R,p) \in SE(3)$ where $R = e^{\widehat{u}\theta}$ with u of unit length. If $0 < \theta < \pi$ and u is parallel or perpendicular to p, then the normalized matrix Kruppa equation: $\widehat{p}^T RYR^T\widehat{p} = \widehat{p}^T Y\widehat{p}$ imposes only 2 linearly independent constraints on the symmetric matrix Y.*

Proof: For the parallel case, by restricting Y to the plane spanned by the column vectors of $\widehat{p}$, it is a symmetric matrix $\tilde{Y}$ in $\mathbb{R}^{2\times 2}$. The rotation matrix $R \in SO(3)$ restricted to this plane is a rotation $\tilde{R} \in SO(2)$. The normalized matrix Kruppa equation is then equivalent to:

$$\tilde{Y} - \tilde{R}\tilde{Y}\tilde{R}^T = 0. \tag{76}$$

Since $0 < \theta < \pi$, this equation imposes exactly 2 constraints on the 3 dimensional space of 2×2 real symmetric matrices. The identity $I_{2\times 2}$ is the only solution. Hence the normalized Kruppa equation imposes exactly 2 linearly independent constraints on Y.

[6] Here $\|\cdot\|$ represents the 2-norm.

For the perpendicular case, since u in the plane spanned by the column vectors of $\widehat{p}$, there exist $v \in \mathbb{R}^3$ such that (u, v) form an orthonormal basis of the plane. Then the normalized matrix Kruppa equation is equivalent to:

$$\widehat{p}^T RYR^T \widehat{p} = \widehat{p}^T Y \widehat{p} \quad \Leftrightarrow \quad (u,v)^T RYR^T (u,v) = (u,v)^T Y (u,v). \tag{77}$$

Since $R^T u = u$, the above equation is equivalent to:

$$v^T RYu = v^T Yu, \quad v^T RYR^T v = v^T Yv. \tag{78}$$

These are the only two constraints given by the normalized Kruppa equation. ∎

According to this lemma, although we can renormalize the fundamental matrix in the case when rotation axis and translation are parallel or perpendicular, we only get two independent constraints from the resulting (normalized) Kruppa equation corresponding to a single fundamental matrix. Hence, for these motions in general, we need 3 such fundamental matrices to uniquely determine the unknown calibration. Consequently, if we do not renormalize the fundamental matrix in this case and directly use the Kruppa equations given by (65) to solve for calibration, the two nonlinear equations given by (65) are in fact algebraically dependent! Therefore, one can only get 1 instead of 2 constraints on the unknown calibration S^{-1} from one fundamental matrix. Although, mathematically, such motions are only a zero-measured subset of $SE(3)$, they are very commonly encountered in real applications: most images sequences are in fact taken through planar or orbital motions whose rotation axis and translation are unfortunately perpendicular to each other, or through the so called screw motions whose rotation axis and translation are parallel! This observation may explain why self-calibration based on directly solving the Kruppa equations (65) is most likely to be ill-conditioned when being applied to real image sequences taken under such motions. Our analysis, however, shows that such motions do not always mean trouble: once the fundamental matrix or Kruppa equations are renormalized, linear equations can be obtained.

Comments 6 *Interestingly, for a walking human, the main rotation of the eyes and the head is yaw and pitch whose axes are perpendicular to the direction of walking. As the theorem suggests, self-calibration in this situation is linear hence more robust to noise. Similar cases can also often be found in vision-guided navigation systems, on-board planar mobile robots. The screw motion, on the other hand, shows up very frequently in motion of aerial mobile robots such as an autonomous helicopter.*

Comments 7 *From Lemma 5, we can see that some of the analysis and results given in [30] are incorrect. Claims of Lemma 5 are direct counter examples to the claims of Propositions B.5 hence B.9 in [30]: the solutions of the normalized Kruppa equations when the translation is parallel or perpendicular to the rotation axis is always 4 dimensional as opposed to 2 or 3 respectively. If one allows the rotation to be* $180°$, *solutions of the normalized Kruppa equations are even more complicated. For example, we know* $e^{\widehat{u}\pi}\widehat{p} = -\widehat{p}$ *if* u *is of unit length and parallel to* p *(see [15]). Therefore, if* $R = e^{\widehat{u}\pi}$, *the corresponding normalized Kruppa equation is completely degenerate and imposes no constraint on the solution of calibration.*

Remark 2 *Although Lemma 4 claims that for the perpendicular case* λ *is one of the two non-zero eigenvalues of* $F\widehat{p'}^T$, *unfortunately, there is no way to tell which one is the right one – simulations show that it could be either the larger or smaller one. Therefore, in algorithm, for given (more than*

3) fundamental matrices, one needs to consider all possible combinations.[7] *According to Theorem 4, in the noise-free case, only one of the solutions can be positive definite, which corresponds to the the true calibration.*

3.6 Beyond the Kruppa Equations?

The proof of Lemma 4 suggests another constraint that can be derived from the fundamental matrix $F = \lambda A^{-T}R^T A^T\widehat{p'}$ with $\|p'\| = 1$. Since $F\widehat{p'}F^T = \lambda^2\widehat{AR^Tp}$, we can obtain the vector $\lambda^2 AR^T p$ which is also equal to $\lambda^2 AR^T A^{-1}p'$. Let us call this vector $\alpha \in \mathbb{R}^3$. Then it is obvious that the following constraint for $S = A^{-T}A^{-1}$ is satisfied:

$$\alpha^T S\alpha = \lambda^4 p'^T Sp'. \tag{79}$$

Notice that this is a constraint on S, not like the Kruppa equations which are constraints on S^{-1}. Combining the Kruppa equations given in (65) we have:

$$\lambda^2 = \frac{\xi_1^T S^{-1}\xi_1}{\eta_2^T S^{-1}\eta_2} = \frac{\xi_2^T S^{-1}\xi_2}{\eta_1^T S^{-1}\eta_1} = \frac{\xi_1^T S^{-1}\xi_2}{\eta_1^T S^{-1}\eta_2} = \sqrt{\frac{\alpha^T S\alpha}{p'^T Sp'}}. \tag{80}$$

Is the last equation algebraically independent of the two Kruppa equations? Although it seems to be quite different from the Kruppa equations, it is in fact dependent on them.[8] The necessary and sufficient condition for a unique camera calibration given by Theorem 1 claims that two general motions with rotation along different axes already determine the calibration. However, it seems that every effort of looking for the third constraint on S from fundamental matrix only has failed. We hence need to know what information is missing in the Kruppa equations.

Another constraint on the calibration actually lies in the fact that not all S which satisfy the Kruppa equations may give valid Euclidean reconstructions of both the camera motion and scene structure. Suppose that a camera motion is $(R, p) \in SE(3)$ and its associated essential matrix is $E = R^T\widehat{p}$. If there exist $Y = A^{-1}A^{-T}$ such that: $E^TYE = \lambda^2\widehat{p}^TY\widehat{p}$ for some $\lambda \in \mathbb{R}$. Then the matrix $F = A^{-T}EA^{-1} = A^{-T}R^TA^T\widehat{p'}$ is also an essential matrix with $p' = Ap$. That is, there exists $\tilde{R} \in SO(3)$ such that $F = \tilde{R}^T\widehat{p'}$ (see [21] for an account of properties of essential matrices). Under the new calibration A, the coordinate transformation becomes:

$$\lambda_2 A\mathbf{x}_2 = \lambda_1 ARA^{-1}(A\mathbf{x}_1) + p'.$$

Since $F = \tilde{R}^T\widehat{p'}$, we have $ARA^{-1} = \tilde{R} + p'v^T$ for some $v \in \mathbb{R}^3$. We then have:

$$\lambda_2 A\mathbf{x}_2 = \lambda_1\tilde{R}(A\mathbf{x}_1) + \lambda_1 p'v^T(A\mathbf{x}_1) + p'.$$

Let $\beta = \lambda_1 v^T(A\mathbf{x}_1) \in \mathbb{R}$ and we have:

$$\lambda_2 A\mathbf{x}_2 = \lambda_1\tilde{R}(A\mathbf{x}_1) + \beta p' + p'. \quad \Rightarrow \quad \lambda_1\widehat{A\mathbf{x}_2}\tilde{R}(A\mathbf{x}_1) = (\beta + 1)\widehat{p'}A\mathbf{x}_2.$$

Now we prove by contradiction that $v \neq 0$ is not possible for a Euclidean reconstruction. Suppose that $v \neq 0$ and let $N \subset \mathbb{R}^3$ to be the plane $\{q \in \mathbb{R}^3 | v^Tq = -1\}$. Then for all $\lambda_1 A\mathbf{x}_1 \in N$, we have $\beta = -1$ hence:

$$\widehat{A\mathbf{x}_2}\tilde{R}(A\mathbf{x}_1) = 0$$

[7] There are 2^n combinations for n fundamental matrices.

[8] We have shown this numerically.

for all points on the plane N. If the motion $(\tilde{R}, p')$ allows a valid Euclidean reconstruction. Then, we have:

$$\gamma_2 A\mathbf{x}_2 = \gamma_1 \tilde{R} A\mathbf{x}_1 + p'$$

for some positive scales $\gamma_1, \gamma_2 \in \mathbb{R}^+$. Combining the two equations, we have:

$$\widehat{p'}\tilde{R}(A\mathbf{x}_1) = 0$$

for all $A\mathbf{x}_1 \in N$. This is impossible since the null space of $\widehat{p'}\tilde{R}$ is only 1 dimensional. So the relationship cannot hold for all points on the plane N. Therefore we conclude that v has to be zero! We have in fact proved the following theorem:

Theorem 5 (Beyond Kruppa Equations) *Given a camera with calibration matrix I and motion (R, p), among all the solutions $Y = A^{-1}A^{-T}$ of the Kruppa equation $E^T Y E = \lambda^2 \widehat{p}^T Y \widehat{p}$ associated to $E = R^T\widehat{p}$, only those which guarantee $ARA^{-1} \in SO(3)$ may provide a valid Euclidean reconstruction of both camera motion and scene structure. Any other solutions always push some plane in $\mathbb{R}^3$ to the distance at infinite.*

Comments 8 *According to Theorem 5, from two uncalibrated images, we, in principle, can recover the camera calibration, motion and 3D structure up to a one parameter family, as opposed to an arbitrary projective transformation claimed by [10]. For a more detailed discussion of this family of solutions, we refer the reader to [20].*

Theorem 5 explains why we get only two constraints from one fundamental matrix even in the cases when the Kruppa equations can be renormalized – the extra one is imposed by the structure reconstruction. The theorem also resolves the discrepancy between the Kruppa equations and the necessary and sufficient condition for a unique calibration: the Kruppa equations, although convenient to use, do not provide sufficient conditions for a valid calibration which allows a Euclidean reconstruction of both the camera motion and scene structure. In fact Theorem 5 is simply another way of justifying Definition 1 and the proof of Theorem 1 in the language of fundamental (or essential) matrix and Kruppa equation. It claims again that, as far as a Euclidean reconstruction is concerned, it is only the rotational motion that determines the condition for a unique calibration, as opposed to the results given in [25]. However, the conclusion given in Theorem 5 is very hard to harness in algorithms. For example, in order to exclude invalid solutions, one needs feature points on the plane N. It is not yet clear what we can do if such feature points are not available. This remains a subject of our future research.

4 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 $\widehat{\omega} = \dot{R}(t)R^T(t) \in so(3)$ and linear velocity $v = -\widehat{\omega}p(t) + \dot{p}(t) \in \mathbb{R}^3$ 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:

$$\dot{r} = A\widehat{\omega}A^{-1}r + v' \qquad (81)$$

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

4.1 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 (81) gives:

$$\begin{aligned} & \dot{r} = A\widehat{\omega}A^{-1}r + v' \quad \Rightarrow \quad (v' + \mathbf{x}) \times \dot{r} = (v' + \mathbf{x}) \times A\widehat{\omega}A^{-1}r \\ \Rightarrow \quad & \dot{\mathbf{x}}^T A^{-T}\widehat{v}A^{-1}\mathbf{x} + \mathbf{x}^T A^{-T}\widehat{\omega}\widehat{v}A^{-1}\mathbf{x} = 0. \end{aligned} \tag{82}$$

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

$$F' = \begin{pmatrix} A^{-T}\widehat{v}A^{-1} \\ A^{-T}sA^{-1} \end{pmatrix}. \tag{83}$$

F' can therefore be estimated from as few as eight optical flows $(\mathbf{x}, \dot{\mathbf{x}})$ from (82) (see [16]).

Note that $\widehat{v'} = A^{-T}\widehat{v}A^{-1}$ and $\widehat{\omega'} = A^{-T}\widehat{\omega}A^{-1}$. Applying Lemma 2 repeatedly, we obtain

$$A^{-T}sA^{-1} = \frac{1}{2}A^{-T}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})A^{-1} = \frac{1}{2}(A^{-T}\widehat{\omega}A^T\widehat{v'} + \widehat{v'}A\widehat{\omega}A^{-1}) = \frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'}). \tag{84}$$

Then the differential epipolar constraint (82) is equivalent to:

$$\dot{\mathbf{x}}^T\widehat{v'}\mathbf{x} + \mathbf{x}^T\frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'})\mathbf{x} = 0. \tag{85}$$

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

$$\begin{aligned} & \dot{\mathbf{x}}^T\widehat{v'}\mathbf{x} + \mathbf{x}^T\frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'})\mathbf{x} = 0 \\ \Leftrightarrow \quad & \dot{\mathbf{x}}^T\widehat{v'}\mathbf{x} + \mathbf{x}^T\frac{1}{2}(\widehat{\omega'}BB^T\widehat{v'} + \widehat{v'}BB^T\widehat{\omega'})\mathbf{x} = 0 \\ \Leftrightarrow \quad & \dot{\mathbf{x}}^TB^{-T}\widehat{R_0v}B^{-1}\mathbf{x} + \mathbf{x}^TB^{-T}\widehat{R_0\omega}\widehat{R_0v}B^{-1}\mathbf{x} = 0. \end{aligned} \tag{86}$$

Comparing to (82), 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 diagonalized as:

$$S^{-1} = R_1^T\Sigma R_1, \quad R_1 \in SO(3) \tag{87}$$

where $\Sigma = \text{diag}\{\sigma_1, \sigma_2, \sigma_3\}$. Then let $\omega'' = R_1\omega'$ and $v'' = R_1v'$. Applying Lemma 2, we have:

$$\begin{aligned} \widehat{v'} &= R_1^T\widehat{v''}R_1 \\ \frac{1}{2}(\widehat{\omega'}S^{-1}\widehat{v'} + \widehat{v'}S^{-1}\widehat{\omega'}) &= R_1^T\frac{1}{2}(\widehat{\omega''}\Sigma\widehat{v''} + \widehat{v''}\Sigma\widehat{\omega''})R_1. \end{aligned} \tag{88}$$

Thus the differential epipolar constraint (82) is also equivalent to:

$$(R_1\dot{\mathbf{x}})^T\widehat{v''}(R_1\mathbf{x}) + (R_1\mathbf{x})^T\frac{1}{2}(\widehat{\omega''}\Sigma\widehat{v''} + \widehat{v''}\Sigma\widehat{\omega''})(R_1\mathbf{x}) = 0. \tag{89}$$

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 $B = R_1 A$. 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 6 *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 *et al.* [4]. 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 arguments.

Comments 9 *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 almost exactly parallel sets of theory and algorithms. We believe that this has to do with the map:*

$$\begin{array}{rcl} \gamma_A : \mathbb{R}^{3\times 3} & \rightarrow & \mathbb{R}^{3\times 3} \\ B & \mapsto & ABA^T \end{array}$$

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

Comments 10 *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 [14] or [1].*

4.2 Pure Rotation Case

Since full calibration is not possible in the general case when translation is present, we need to know if it is possible in some special case. The only case left 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 (81) we have:

$$\begin{array}{rl} & \dot{r} = A\hat{\omega}A^{-1}r \quad \Rightarrow \quad \dot{\lambda}\mathbf{x} + \lambda\dot{\mathbf{x}} = A\hat{\omega}\lambda A^{-1}\mathbf{x} \\ \Rightarrow & \hat{\mathbf{x}}\dot{\mathbf{x}} = \hat{\mathbf{x}}A\hat{\omega}A^{-1}\mathbf{x}. \end{array} \tag{90}$$

This is a degenerate version of the differential epipolar constraint and it gives two independent constraints on the matrix $A\hat{\omega}A^{-1}$ for each $(\mathbf{x}, \dot{\mathbf{x}})$. Given $n \geq 4$ optical flow measurements $\{(\mathbf{x}_i, \dot{\mathbf{x}}_i)\}_{i=1}^n$, one may uniquely determine the matrix $A\hat{\omega}A^{-1}$ by solving a linear equation:

$$Mc = b \tag{91}$$

where $M \in \mathbb{R}^{2n\times 9}$ is a matrix function of $\{(\mathbf{x}_i, \dot{\mathbf{x}}_i)\}_{i=1}^n$, $b \in \mathbb{R}^9$ is a vector function of $\widehat{\mathbf{x}}_i\dot{\mathbf{x}}_i$'s and $c \in \mathbb{R}^9$ is the 9 entries of $A\widehat{\omega}A^{-1}$. The solution is given by the following is lemma:

Lemma 6 *If $\omega = 0$, then $A\widehat{\omega}A^{-1} = C_p - \gamma I$ where $C_p \in \mathbb{R}^{3\times 3}$ is the matrix corresponding to the least square solution of the equation $Mc = b$ and γ is the unique real eigenvalue of C_p.*

The proof is straightforward. Then the self-calibration problem becomes how to recover $S = A^{-T}A^{-1}$ or $S^{-1} = AA^T$ from matrices of the form $A\widehat{\omega}A^{-1}$. Without loss of generality, we may assume ω is of unit length. Notice that this problem is exactly a differential version of the discrete pure rotation case.

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

$$SC' = A^{-T}\widehat{\omega}A^{-1} = \widehat{\omega'} \tag{92}$$

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 Lyapunov map:

$$\begin{aligned} L' : \mathbb{C}^{3\times 3} &\rightarrow \mathbb{C}^{3\times 3} \\ X &\mapsto (C')^T X + XC' \end{aligned} \tag{93}$$

If $\omega \neq 0$, the eigenvalues of $\widehat{\omega}$ 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 [5], the null space of the map L' has three dimensions and is given by:

$$\mathrm{Ker}(L') = \mathrm{span}\{S_1 = A^{-T}\omega\omega^* A^{-1}, S_2 = A^{-T}uu^* A^{-1}, S_3 = A^{-T}\bar{u}\bar{u}^* A^{-1}\}. \tag{94}$$

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

Similar to the discrete case we have:

Theorem 7 *Given matrices $C'_j = A\widehat{\omega}_j A^{-1} \in \mathbb{R}^{3\times 3}, j = 1, \ldots, n$ 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.*

5 Simulation Results

In this section, we test the performance of the proposed algorithms through different experiments. The error measure between the actual calibration matrix A and the estimated calibration matrix $\tilde{A}$ was chosen to be:

$$error = \frac{\|A - \tilde{A}\|}{\|A\|} \times 100\%$$

Table 1 shows the simulation parameters used in the experiments.[9] The calibration matrix A is

Table 1: Simulation parameters

Parameter	Unit	Value
Number of trials		100
Number of points		20
Number of frames		3-4
Field of view	degrees	90
Depth variation	u.f.l.	100 - 400
Image size	pixels	500 × 500

simply the transformation from the original 2×2 (in unit of focal length) image to the 500×500 pixel image. For these parameters, the true A should be:

$$A = \begin{pmatrix} 250 & 0 & 250 \\ 0 & 250 & 250 \\ 0 & 0 & 1 \end{pmatrix}.$$

The ratio of the magnitude of translation and rotation, or simply the T/R ratio, is compared at the center of the random cloud (scattered in the truncated pyramid specified by the given field of view and depth variation). For all simulations, the number of trials is 100.

Pure rotation case: Figures 3, 4 and 5 show the experiments performed in the pure rotation case. The axes of rotation are X and Y for Figures 3 and 5, and X and Z for Figure 4. The amount of rotation is 20°. The perfect data was corrupted with zero-mean Gaussian noise with standard deviation σ varying from 0 to 5 pixels. In Figures 3 and 4 it can be observed that the algorithm performs very well in the presence of noise, reaching errors of less than 6% for a noise level of 5 pixels. Figure 5 shows the effect of the amount of translation. This experiment is aimed to test the robustness of the pure rotation algorithm with respect to translation. The T/R ratio was varied from 0 to 0.5 and the noise level was set to 2 pixels. It can be observed that the algorithm is not robust with respect to the amount of translation.

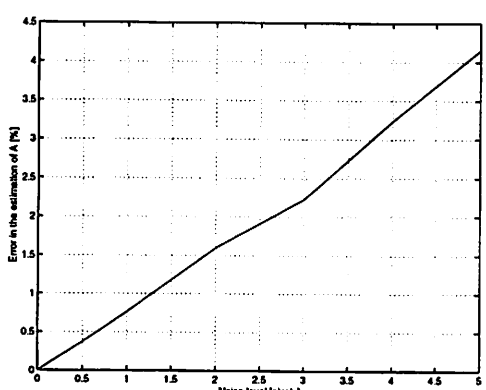

Figure 3: Pure rotation algorithm. Rotation axes X-Y.

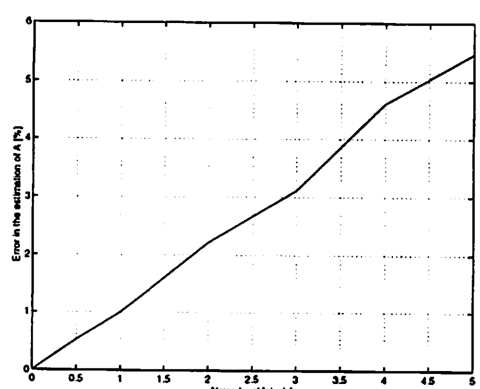

Figure 4: Pure rotation algorithm. Rotation axes X-Z.

[9]u.f.l. stands for unit of focal length.

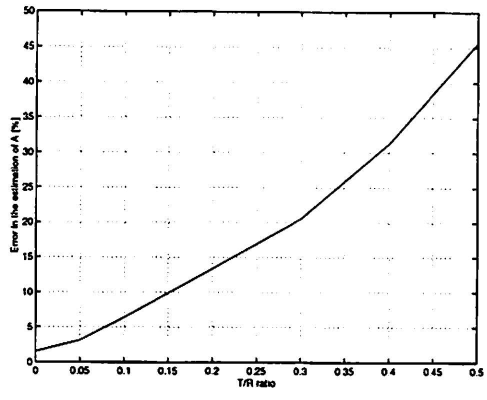

Figure 5: Rotation axes X-Y, $\sigma = 2$.

Translation parallel to rotation axis: Figures 6 and 7 show the experiments performed when translation is parallel to the axis of rotation.[10] The non-isotropic normalization procedure proposed by Hartley [7] and statistically justified by Mühlich and Mester [23] was used to estimate the fundamental matrix. Figure 6 shows the effect of noise in the estimation of the calibration matrix for $T/R = 1$ and a rotation of $\theta = 20^o$ between consecutive frames. It can be seen that the normalization procedure improves the estimation of the calibration matrix, but the improvement is not significant. This result is consistent with that of [23], since the effect of normalization is more important for large noise levels. On the other hand, the performance of the algorithm is not as good as that of the pure rotation case, but still an error of 5% is reached for a noise level of 2 pixels. Figure 7 shows the effect of the angle of rotation in the estimation of the calibration matrix for a noise level of 2 pixels. It can be concluded that a minimum angle of rotation between consecutive frames is required for the algorithm to succeed.

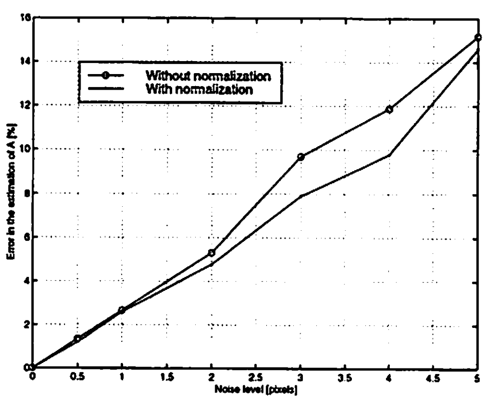

Figure 6: Rotation parallel to translation case. $\theta = 20^o$. Rotation/Translation axes: XX-YY-ZZ, T/R ratio = 1.

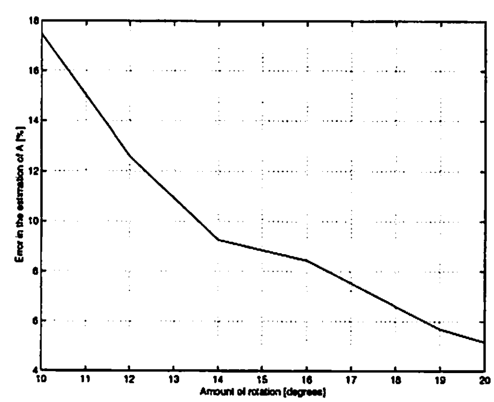

Figure 7: Rotation parallel to translation case. $\sigma = 2$. Rotation/Translation axes: XX-YY-ZZ, T/R ratio = 1.

Translation perpendicular to rotation axis: Figures 8 and 9 show the experiments performed when translation is perpendicular to the axis of rotation. It can be observed that this algorithm

[10]For specifying the Rotation/Translation axes, we simply use symbols such as "XY-YY-ZZ" which means: for the first pair of images the relative motion is rotation along X and translation along Y; for the second pair both rotation and translation are along Y; and for the third pair both rotation and translation are along Z.

is much more sensitive to noise. The noise has to be less than 0.5 pixels in order to get an error of 5%. Experimentally it was found that Kruppa equations are very sensitive to the normalization of the fundamental matrix F and that the eigenvalues λ_1 and λ_2 of $F\hat{p'}^T$ are close to each other. Therefore in the presence of noise, the estimation of those eigenvalues might be ill conditioned (even complex eigenvalues are obtained) and so is the solution of Kruppa equations. Another experimental problem is that more than one non-degenerate solution to Kruppa equations can be found. This is because, when taking all possible combinations of eigenvalues of $F\hat{p'}^T$ in order to normalize F, the smallest eigenvalue of the linear map associated to "incorrect" Kruppa equations can be very small. Besides, the eigenvector associated to this eigenvalue can eventually give a non-degenerate matrix. Thus in the presence of noise, you can not distinguish between the correct and one of these incorrect solutions. The results presented here correspond to the best match (to the ground truth) when more than one solution is found. Finally it is important to note that large motions can significantly improve the performance of the algorithm. Figure 9 shows the error in the estimation of the calibration matrix for a rotation of 30^o. It can be observed that the results are comparable to that of the parallel case with a rotation of 20^o.

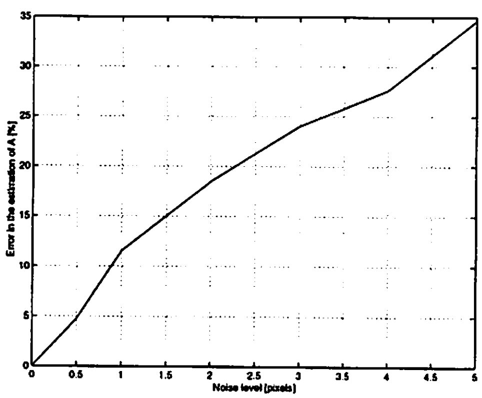

Figure 8: Rotation orthogonal to translation case. $\theta = 20^o$. Rotation/Translation axes: XY-YZ-ZX, T/R ratio = 1.

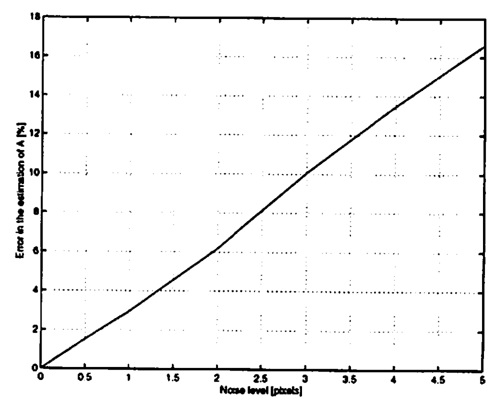

Figure 9: Rotation orthogonal to translation case. $\theta = 30^o$. Rotation/Translation axes: XY-YZ-ZX, T/R ratio = 1.

Robustness: In order to check how robust the algorithms are with respect to the angle ϕ between the rotation axis and translation, we run them with ϕ varying from 0^o to 90^o. The noise level is 2 pixels, amount of rotation is always 20^o and the T/R ratio is 1. Translation and rotation axes are given by Figure 10. Surprisingly, as we can see from the results given in Figure 11, for the range $0^o \leq \phi \leq 50^o$, both algorithms give pretty close estimates. This is because, for this range of angle, numerically the eigenvalues of the matrix $F\widehat{p'}^T$ are complex and their norm is very close to the norm of the matrix $F\widehat{p'}F^T$. Therefore, the computed renormalization scale λ from both algorithms is very close, as is the calibration estimate. For $\phi > 50^o$, the eigenvalues of $F\widehat{p'}^T$ become real and the performance of the two algorithms is no longer the same.

6 Discussions and Future Work

In this paper, we have proposed a geometric approach for the study of camera self-calibration. Based on a new geometric interpretation of the camera calibration, we give a clear account of

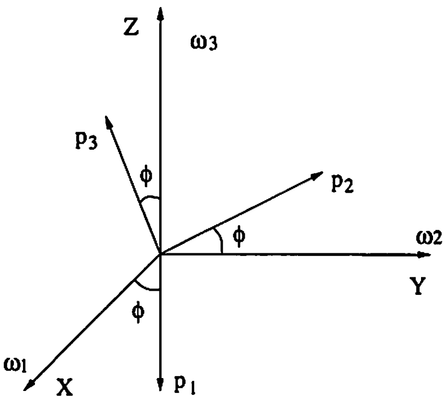

Figure 10: The relation of the three rotation axes $\omega_1, \omega_2, \omega_3$ and three translations p_1, p_2, p_3.

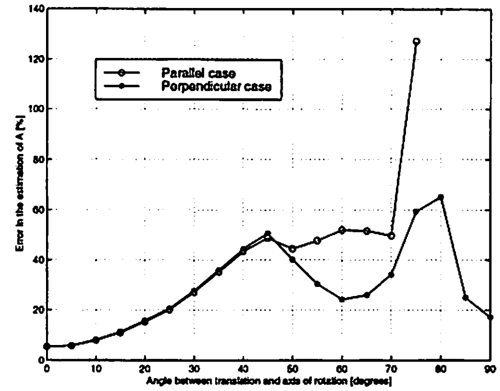

Figure 11: Estimation error in calibration w.r.t. different angle ϕ. Noise level $\sigma = 2$. Rotation and translation axes are shown by the figure to the left. Rotation amount is always 20^o and T/R ratio is 1.

the necessary and sufficient condition for a unique calibration and clarify some misunderstanding existing in the computer vision literature. By the study of the cases when the rotation axis is parallel or perpendicular to the translation, we discovered generic difficulties in the conventional self-calibration schemes based on the nonlinear Kruppa equations. The results not only clarify some incorrect claims in the literature but also provide brand new linear algorithms for self-calibration. As in several of our other papers [16, 15, 14], we have investigated 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 differential epipolar constraint is a degenerate one which can only determine (at most) two intrinsic parameters of the camera. This also explains the nonlinearity of the Kruppa equations in the discrete case.

In [20], subgroups of $SE(3)$ are systematically studied for the purpose of recovering camera motion, calibration and scene structure altogether. Roughly speaking, the ambiguities in the recovery can be represented as certain groups acting on the overall configuration of the camera system. In this paper, we have assumed that the calibration matrix A (or the intrinsic parameters) is always constant. The results obtained here certainly help to study the case when A is time-varying (such as changing focal length) and the associated Kruppa equations become time-varying. Since, in real applications, the camera is usually pre-calibrated and only some of the camera intrinsic parameters may be unknown or time-varying, such as the focal length, a detailed study of the geometry for such a camera system is also of great theoretical and practical importance. In this paper, only basic simulations have been presented. We will give a more detailed report of the performance of the proposed algorithms through more extensive simulations and experiments on real image sequences.

Although the self-calibration theory has only been developed for the Euclidean case, most theorems can be easily generalized to a larger class of Riemannian manifolds (for example see [18]). In fact, it can 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 [19].

7 Acknowledgment

We like to thank Stefano Soatto (of Washington University) for useful discussions and collaboration on some relevant topics on camera self-calibration.

References

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

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

[3] S. Bougnoux. From projective to euclidean space under any pratical situation, a criticism of self-calibration. In *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, pages 790–796, 1998.

[4] 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.

[5] F. M. Callier and C. A. Desoer. *Linear System Theory.* Springer Texts in Electrical Engineering. Springer-Verlag, 1991.

[6] R. Goodman and N. R. Wallach. *Representations and Invariants of the Classical Groups.* Cambridge University Press, 1998.

[7] R. I. Hartley. In defence of the 8-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6), 1997.

[8] 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.

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

[10] R. I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *ICCV*, pages 761–4, 1992.

[11] S. Kobayashi and T. Nomizu. *Foundations of Differential Geometry: Volume I and Volume II.* John Wiley & Sons, Inc., 1996.

[12] Q.-T. Luong and O. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *IJCV*, 1994. toappear.

[13] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *IJCV*, 22(3):261–89, 1997.

[14] Y. Ma, J. Košecká, and S. Sastry. Euclidean structure and motion from image sequences. *UC Berkeley Memorandum No. UCB/ERL M98/38*, 1998.

[15] Y. Ma, J. Košecká, and S. Sastry. Linear differential algorithm for motion recovery: A geometric approach. *submitted to International Journal of Computer Vision*, 1998.

[16] Y. Ma, J. Košecká, and S. 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.

[17] Y. Ma, J. Košecká, and S. Sastry. Optimization criteria and geometric algorithms for motion and structure estimation. *submitted to International Journal of Computer Vision*, 1999.

[18] Y. Ma and S. Sastry. Vision theory in spaces of constant curvature. *UC Berkeley Memorandum No. UCB/ERL M98/36*, 1998.

[19] Y. Ma, O. Shakernia, J. Košecká, and S. Sastry. A Lie theoretic approach to motion and structure in computer vision. In *Proceedings of the 14th World Congress of IFAC*, 1999.

[20] Y. Ma, S. Soatto, J. Košecká, and S. Sastry. Uncalibrated reconstruction up to relevant subgroups. *UC Berkeley Memorandum No. UCB/ERL M98/65 and to appear in Proceedings of ICCV*, 1999.

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

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

[23] M. Mühlich and R. Mester. The role of total least squares in motion analysis. In *Proceedings of European Conference on Computer Vision*, pages 305–321, 1998.

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

[25] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, pages 1100–1105, 1997.

[26] B. Triggs. Autocalibration from planar scenes. In *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, 1998.

[27] 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.

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

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

[30] C. Zeller and O. Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. *Research Report 2793, INRIA, France*, 1996.