EUCLIDEAN RECONSTRUCTION AND REPROJECTION UP TO SUBGROUPS

by

Yi Ma, Stefano Soatto, Jana Kosecka and Shankar Sastry

Memorandum No. UCB/ERL M98/65

24 November 1998

COVER

EUCLIDEAN RECONSTRUCTION AND REPROJECTION UP TO SUBGROUPS

by

Yi Ma, Stefano Soatto, Jana Kosecka and Shankar Sastry

Memorandum No. UCB/ERL M98/65
24 November 1998

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Euclidean Reconstruction and Reprojection Up to Subgroups *

Yi Ma† Stefano Soatto‡ Jana Košecká† Shankar Sastry†
† UC Berkeley, Berkeley - CA 94720 ‡ Washington University, St.Louis - MO 63130

November 24, 1998

Abstract

The necessary and sufficient conditions for being able to estimate scene structure, motion and camera calibration from a sequence of images are very rarely satisfied in practice. What exactly, then, can be estimated in sequences of practical importance, when such conditions are not satisfied? In this paper we give a complete answer to this question. For every camera motion that fails to satisfy the conditions for unique reconstruction, we give an explicit characterization of the ambiguity in the reconstructed scene, motion and calibration. When the purpose of the reconstruction is to generate novel views of the scene, we characterize the vantage points that give raise to a valid Euclidean reprojection. We also characterize viewpoints that make the re-projection invariant to the ambiguity.

The key to our findings lies in a powerful result on the dependency of multilinear constraints: we prove that the coefficients of multilinear constraints involving any number of images can be generated from coefficients of bilinear constraints alone. Therefore, all the analysis involving n views can be carried out using two views at a time.

Key words: multilinear constraints, camera self-calibration, structure from motion, reprojection.

1 Introduction

Reconstructing spatial properties of a scene from a number of images taken by an unknown camera is a classical problem in Computer Vision. It is particularly important when the camera used to acquire the images is not available for calibration, as for instance in video post-processing, or when the calibration changes in time, as in vision-based navigation. If we represent the scene by a number of isolated points in three-dimensional space and the imaging process by an ideal perspective projection, the problem can be reduced to a purely geometric one, which has been subject to the intense scrutiny of a number of researchers during the past ten years. Their efforts, which we discuss in more detail in section 1.2, have led to several important and useful results. For instance, we now know that the dependency among two or more images of the same points are described by multilinear constraints, and we have necessary and sufficient conditions for being able to reconstruct the three-dimensional position of the points, the motion of the camera and its calibration (up to a global scale factor). The problem is that *such conditions are almost never satisfied in sequence of images of practical importance.* In fact, they require that the camera undergoes rotation about at least two independent axes, which is rarely the case both in video processing and in autonomous navigation.

In this paper we address the question of *what can be done when the necessary and sufficient conditions for unique reconstruction are* **not** *satisfied.* In particular:

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

(i) For all the motions that do not satisfy the conditions, to what extent can we reconstruct structure, motion and calibration?

(ii) If the goal of the reconstruction is to produce a new view of the scene from a different vantage point, how can we generate images of a "valid" Euclidean scene?

On our way to answering these questions, we pause to reflect on the nature of multi-linear constraints. While constraints involving two images at a time (fundamental constraints) are well understood and involve clean notation and geometric interpretation, muti-linear constraints are more difficult to work with and to interpret. It seems therefore natural to ask the following question

(iii) Do multi-linear constraints carry information that is not contained in bilinear ones?

These are fairly fundamental questions that have been addressed by several researchers, as we discuss in section 1.2. Before doing that, however, we want to highlight the answers given on this paper, so that the reader know what to expect (and what not to expect) from it.

1.1 Outline of the paper and its contributions

The main contributions of this paper can be summarized as follows. Concerning question (iii), this paper proves that the information encoded in trilinear and quadrilinear constraints is dependent on that encoded in bilinear ones. Bilinear constraints, also known as "fundamental" or "epipolar", are therefore the only independent ones and truly deserve their name. This is discussed in section 2. There we also discuss the role of multilinear constraints with regards to singular configurations of points. The well-known, and conservative, answer to question (i) is that structure can always be recovered up to a global projective transformation of the three-dimensional space. However, there is more to be said, as we do in section 3. There, we first classify all the subgroups of the Euclidean group of rigid motions and, for each of them, we give an explicit characterization of the ambiguity in the reconstruction of scene structure and camera motion and calibration. Question (ii) is answered in section 4, where we characterize the set of vantage points that generate "valid" images of the scene. In section 5 we characterize the ambiguities in the reconstruction associated to subgroups of the calibration group. This is important when (part of) the calibration parameters change across images, as for instance in a zooming camera.

These results have great practical significance, because they quantify precisely to what extent scene structure, camera motion and calibration can be estimated in most sequences, for which many of the results available up to date do not apply. In addition to that, we clarify the role of multilinear constraints and their relationship to bilinear ones.

Granted the potential impact on applications, this paper is mainly concerned with theory. We address neither algorithmic issues, nor do we perform experiments of any sort: the validation of our statements is in the proofs. We have tried to keep our notation as slender as possible, avoiding any tensorial notation and projective geometry. Our tools are borrowed from linear algebra and some differential geometry, although all the results should be accessible without background in the latter.

1.2 Relations to previous work

The study of ambiguities naturally arise in the problem of motion and structure recovery and self-calibration from multiple cameras. There is a vast body of literature in this field; some of the earliest contributions are due to Carlsson [4], Faugeras and Mourrain [7], Hartley [8], Triggs [20], Shashua [19, 21], Luong et al. [11]. Here we only comment on some of the work that is most closely related to this paper, while we refer the reader to the literature for more details, references and appropriate credits.

It has long been known that in the absence of any *a priori* information about motion, calibration and scene structure, reconstruction can be performed only up to a projective transformation [6]. Utilizing additional knowledge about the relationship between geometric entities in the image (e.g. line parallelism, absolute conic) one can stratify the different levels of reconstructions from projective all the way to Euclidean [6, 3, 5, 18]. At such a level of generality, the conditions on the uniqueness and existence of solutions have not been established and the algorithms are computationally costly, often exhibiting local minima [10]. Natural continuations of these efforts assumed cases where the motion and/or calibration were restricted either to planar or linear motion [16, 2] and provided techniques for affine reconstruction. In [9], Heyden and Åström consider the skew-frew camera model and use bundle adjustment methods to estimate the camera motion and (time-varying) camera intrinsic parameters simultaneously. In [17], Pollefeys et al. show the effectiveness of a technique for (partial) self-calibration using a nonlinear iteration to estimate focal length and principal point.

In [1], trilinear constraints are exploited to help generate reprojected images for a calibrated camera. In the case of a partially uncalibrated camera, such a method may have to deal with issues such as whether the reprojected image is still "valid". Werman and Shashua [21] studied algebraic relationship among constraints of multiple images using elimination methods; however "informational" dependence among multilinear constraints has not been assessed.

2 Dependency of multilinear constraints

2.1 Notation

We model the world as a collection of points in three-dimensional Euclidean space, which we represent in homogeneous coordinates as $q = (q_1, q_2, q_3, 1)^T \in \mathbb{R}^4$. The perspective projection of the generic point onto the two-dimensional image plane is represented by homogeneous coordinates $\mathbf{x} \in \mathbb{R}^3$ that satisfy

$$\lambda(t)\mathbf{x}(t) = A(t)g(t)q, \quad t \in \mathbb{R} \tag{1}$$

where $\lambda(t) \in \mathbb{R}$ is a scalar parameter related to the distance of the point q from the center of projection and the non-singular matrix $A(t)$ - called "calibration matrix" - describes the geometry of the camera. Without loss of generality we will re-scale the above equation so that the determinant of A is one. The set of 3×3 matrices with determinant one is called Special Linear group and indicated by $A(t) \in SL(3)$. The rigid motion of the camera $g(t)$ is represented by a translation vector $p(t) \in \mathbb{R}^3$ and a rotation matrix $R(t)$, that is an orthogonal matrix with determinant equal to one. Such matrices form a group called Special Orthogonal group and indicated by $R(t) \in SO(3)$, and $g(t) = (p(t), R(t)) \in SE(3)$, the special Euclidean group of rigid motion in $\mathbb{R}^3$. The action of $g(t)$ on the point q is given by $g(t)q = R(t)q + p(t)$. In equation (1) we will assume that $\mathbf{x}(t)$ is measured, while everything else is unknown.

When we consider measurements at n different times, we organize the above equations by defining

$$M_i \doteq (A(t_i)R(t_i), A(t_i)p(t_i)) \in \mathbb{R}^{3\times4}, \quad i = 1, 2, \dots, n, \tag{2}$$

which we will assume to be full-rank, $rank(M_i) = 3$ for $i = 1, \dots, n$, so that

$$\begin{pmatrix} \mathbf{x}(t_1) & 0 & \cdots & 0 \\ 0 & \mathbf{x}(t_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{x}(t_n) \end{pmatrix} \begin{pmatrix} \lambda(t_1) \\ \lambda(t_2) \\ \vdots \\ \lambda(t_n) \end{pmatrix} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{pmatrix} q \tag{3}$$

which we re-write in a more compact notation as

$$\mathbf{X}\vec{\lambda} = Mq \tag{4}$$

we call $M \in \mathbb{R}^{3n \times 4}$ the *motion matrix.*

2.2 Constraints on multiple images

Let $\vec{\mathbf{m}}_i \in \mathbb{R}^{3n}, i = 1, \dots, 4$ denote the four columns of the matrix M and $\vec{\mathbf{X}}_i \in \mathbb{R}^{3n}, i = 1, \dots, n$ be the n columns of the matrix $\mathbf{X}$. Then the coordinates $\mathbf{x}(t_i)$ represent the same point seen from different views only if they satisfy the following wedge product [1] equation:

$$\vec{\mathbf{m}}_1 \wedge \vec{\mathbf{m}}_2 \wedge \vec{\mathbf{m}}_3 \wedge \vec{\mathbf{m}}_4 \wedge \vec{\mathbf{X}}_1 \wedge \cdots \wedge \vec{\mathbf{X}}_n = 0. \tag{5}$$

This constraint, which is multi-linear in the measurements $\mathbf{x}(t_i)$ simply expresses the fact that $\mathbf{X}$ must be contained in the span of M. Constraints involving four images are call *quadrilinear*, constraints involving three images are called *trilinear*, and those involving two images are called *bilinear* or *fundamental.* In general, coefficients of all the multilinear constraints are minors of the motion matrix M. As it has been proven by Triggs [20] *et al.*, constraints involving more than four frames are necessarily dependent of quadrilinear, trilinear and bilinear ones.

In this section we go one step further to prove that the information encoded in trilinear and quadrilinear constraints is dependent on that encoded in bilinear ones.

2.3 Dependency of trilinear and quadrilinear constraints on bilinear ones

Consider the case $n = 3$ and, for the moment, disregard the internal structure of the motion matrix $M \in \mathbb{R}^{9 \times 4}$. Its columns can be interpreted as a basis of a four-dimensional subspace of the nine-dimensional space. The set of k-dimensional subspaces of an m-dimensional space is called a Grassmannian and denoted by $G(m, k)$. Therefore, M is an element of $G(9, 4)$. By just re-arranging the three blocks M_i, $i = 1 \dots 3$ into three pairs, (M_1, M_2), (M_1, M_3) and (M_2, M_3), we define a map ϕ between $G(9, 4)$ and three copies of $G(6, 4)$

$$\phi : G(9,4) \;\rightarrow\; G(6,4) \times G(6,4) \times G(6,4) \tag{6}$$

$$\begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix} \;\mapsto\; \left(\begin{pmatrix} M_1 \\ M_2 \end{pmatrix}, \begin{pmatrix} M_2 \\ M_3 \end{pmatrix}, \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \right). \tag{7}$$

The question of whether trilinear constraints are independent of bilinear ones is tightly related to whether these two representations of the motion matrix M are equivalent, in the sense that they

[1]For the mathematically inclined reader wedge product $\wedge$ is a specialized version of a tensor product over the space of a sum of all alternating k-tensors $\Lambda^k(V^*)$ which preserves the alternating property; k-tensor is a multilinear function $f : V \times V \times \dots V \rightarrow \mathbb{R}$; the alternating tensor is a tensor f which satisfies the following skew-symmetric property $f = (-1)^{sign(\sigma)} f$, where f^σ is a σ-permutation of f, e.g. an alternating 2-tensor satisfies $f(x, y) = -f(y, x)$ for $x, y \in \mathbb{R}^n$. The notation allows to state some of the known results in a compact way and mediates elegant proofs.

span spaces of the same dimension. In order to answer this question we introduce the following lemma.

Lemma 1 *The map $\phi : G(9,4) \to G(6,4)^3$ above is injective if $Ker(M_i) \in \mathbb{R}^4, i = 1,2,3$ are linearly independent.*

Proof: We want to show that if $\phi(M)$ and $\phi(M')$ span the same subspace, then so do M and M'. Now suppose this is not true. If $\phi(M)$ and $\phi(M')$ span the same subspace, there must exist non-singular matrices $G_1, G_2, G_3 \in GL(4)$ [2] such that $\phi(M')_i = \phi(M)_i G_i, \quad i = 1,2,3$. Then $M_1G_1 = M_1G_3 = M_1', \quad M_2G_1 = M_2G_2 = M_2', \quad M_3G_2 = M_3G_3 = M_3'$ and therefore

$$M_1(G_1 - G_3) = 0, \quad M_2(G_1 - G_2) = 0, \quad M_3(G_2 - G_3) = 0. \tag{8}$$

Since $Ker(M_i), i = 1,2,3$ are linearly independent, we have $G_1 = G_2 = G_3$ and therefore M and M' span the same subspace. In other words, they represent the same element in $G(9,4)$. Hence the map ϕ is injective. ∎

Since the coefficients in the multilinear constraints are homogeneous in the entries of each block M_i, the motion matrix M is only determined up to the equivalence relation:

$$M \sim M' \; if \; \exists \, \lambda_i \in \mathbb{R}^*, M_i = \lambda_i M_i', \quad i = 1, \ldots, n \tag{9}$$

where $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$. Thus the motion matrix is only well-defined as an element of the quotient space $G(3n,4)/\sim$ which is of dimension $(11n - 15)$, [3] as it was already noted by Triggs [20].

We are now ready to prove that trilinear coefficients depend on bilinear ones.

Theorem 1 (Dependency of trilinear coefficients) *Given three views of a point ($n = 3$), the coefficients of all bilinear constraints (or equivalently the corresponding fundamental matrices) uniquely determine the motion matrix M (as an element of $G(9,4)/\sim$) given that $Ker(M_i), i = 1,2,3$ are linearly independent.*

Proof: It is known that between any pair of images (i,j) the motion matrix:

$$\begin{pmatrix} M_i \\ M_j \end{pmatrix} \in G(6,4), \tag{10}$$

is determined by the corresponding fundamental matrix F_{ij} up to two scalars λ_i, λ_j:

$$\begin{pmatrix} \lambda_i M_i \\ \lambda_j M_j \end{pmatrix} \in G(6,4), \quad \lambda_j \in \mathbb{R}^*. \tag{11}$$

Hence all we need to prove is that the map:

$$\tilde{\phi} : (G(9,4)/\sim) \;\; \to \;\; (G(6,4)/\sim)^3 \tag{12}$$

is injective. To this end, assume $\tilde{\phi}(M) = \tilde{\phi}(M')$; then we have that, after re-scaling, $\begin{pmatrix} M_1' \\ M_2' \end{pmatrix} = \begin{pmatrix} \lambda_1 M_1 \\ M_2 \end{pmatrix} G_1$, $\begin{pmatrix} M_2' \\ M_3' \end{pmatrix} = \begin{pmatrix} \lambda_2 M_2 \\ M_3 \end{pmatrix} G_2$, $\begin{pmatrix} M_1' \\ M_3' \end{pmatrix} = \begin{pmatrix} M_1 \\ \lambda_3 M_3 \end{pmatrix} G_3$ for some $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}^*$.

[2] $GL(m)$ stands for the general linear group of $m \times m$ non-singular matrices.

[3] The Grassmannian $G(3n,4)$ has dimension $(3n-4)4 = 12n - 16$. The dimension of the quotient space is $n - 1$ smaller since the equivalence relation has $n - 1$ independent scales.

This yields $M_1(\lambda_1 G_1 - G_3) = 0, \quad M_2(\lambda_2 G_2 - G_1) = 0, \quad M_3(\lambda_3 G_3 - G_2) = 0$. Therefore there exist $U_i \in \mathbb{R}^{4\times 4}, i = 1, 2, 3$ with each column of U_i is in $Ker(M_i)$ such that:

$$G_3 - \lambda_1 G_1 = U_1, \quad G_1 - \lambda_2 G_2 = U_2, \quad G_2 - \lambda_3 G_3 = U_3. \tag{13}$$

Combining these three equations, we obtain:

$$(1 - \lambda_1\lambda_2\lambda_3)G_1 = \lambda_2\lambda_3 U_1 + \lambda_2 U_3 + U_2. \tag{14}$$

The matrix on the right hand side of the equation has a non-trivial null-space since its columns are in $span\{Ker(M_1), Ker(M_2), Ker(M_3)\}$ which has dimension three. However, G_1 is non-singular, and therefore it must be $\lambda_1\lambda_2\lambda_3 = 1$. This gives $\lambda_1 G_1 - G_3 = -\lambda_1(\lambda_2 G_2 - G_1) - \lambda_1\lambda_2(\lambda_3 G_3 - G_2)$. That is, the columns of $\lambda_1 G_1 - G_3$ are linear combinations of columns of $\lambda_2 G_2 - G_1$ and $\lambda_3 G_3 - G_2$. But $Ker(M_i), i = 1, 2, 3$ are linearly independent. Thus we have $\lambda_1 G_1 = G_3, \quad \lambda_2 G_2 = G_1, \quad \lambda_3 G_3 = G_2$. This implies

$$\begin{pmatrix} M_1' \\ M_2' \\ M_3' \end{pmatrix} = \begin{pmatrix} \lambda_1 M_1 \\ M_2 \\ \lambda_1\lambda_3 M_3 \end{pmatrix} G_1. \tag{15}$$

which means that M' and M are the same, up to the equivalence relation defined in equation (9). Therefore, they represent the same element in $G(9,4)/\sim$, which means that the map $\tilde{\phi}$ is injective. ∎

Comment 1 *While the above proof shows that the map $\tilde{\phi}$ can be inverted, it does not provide an explicit characterization of the inverse. Such an inverse can in principle be highly non-linear and conditioning issues need to be taken into account in the design of estimation algorithms.*

Comment 2 *The reader familiar with the literature on multilinear geometry may be a bit surprised at our statements, since it is often written that there are three independent trilinear constraints in n views of a point. This only concerns* algebraic dependency, *and is not in contrast with our findings.*

What the literature refers to is the counting of the number of independent generators of the ideal associated with trilinear constraints. In other words, it is the count of the number of independent constraints on the coordinates of the points $\mathbf{x}(t_i)$.

Now, such constraints could be expressed through bilinear constraints, or through trilinear constraints. Our claim states that the coefficients of trilinear constraints are dependent on the coefficients of bilinear ones. Note that our statement concerns coefficients of the constraints, *not the coordinates of the points.*

To the risk of being tedious, we emphasize that we are not *saying that two views are sufficient for reconstruction! We claim that given n views, their geometry is characterized by considering only combinations of pairs of them through bilinear constraints, while trilinear constraints are of help* only *in the case of singular configurations of points (see comment 4).*

In the case of four images, $n = 4$, in order to show that coefficients in quadrilinear constraints also depend on bilinear ones, one only needs to check that the obvious map from $G(12,4)/\sim$ to $(G(9,4)/\sim)^4$ is injective. The proof directly follows that of the three frame case. We therefore state this as a corollary to the above theorem

Corollary 1 (Dependency of quadrilinear coefficients) *In the case of four frames, coefficients in all the bilinear constraints (or equivalently the corresponding fundamental matrices) uniquely determine the motion matrix M as an element in $G(12,4)/\sim$ given that $Ker(M_i), i = 1, \dots, 4$ are linearly independent.*

Mathematically, the theorem states that, under the given conditions, the maps:

$$\tilde{\phi}_1 : (G(12,4)/\sim) \rightarrow (G(9,4)/\sim)^4 \quad (16)$$
$$\tilde{\phi}_2 : (G(12,4)/\sim) \rightarrow (G(6,4)/\sim)^6 \quad (17)$$

obtained by rearranging the blocks of M are both injective.

Comment 3 *In corollary 1, the condition that $Ker(M_i), i = 1, \dots, 4$ are linearly independent is not necessary. A less conservative condition is that there exist two groups of three frames which satisfy the condition given in Theorem 1.*

Both Theorem 1 and Corollary 1 require that the one-dimensional kernels of the matrices $M_i, i = 1, \dots, n$ are linearly independent. The following Lemma gives a geometric interpretation of this condition.

Lemma 2 *The kernels of the matrices $M_i, i = 1, \dots, n$ with $n \leq 4$ are linearly independent if and only if the center of projection of the cameras generates a hyper-plane of dimension $n - 1$. In particular, when $n = 3$, the three camera centers form a triangle, and when $n = 4$, the four camera centers form a tetrahedron.*

Proof: Since $A(t_i)$ is non-singular, it does not change the kernel of M_i. Express the camera positions with respect to the first camera frame, so that $R(t_1) = I_{3\times3}, p(t_1) = 0$ and $M_i = (R_i, p_i), i = 2,3,4$. The kernels of M_i for $i = 1,2,3,4$ are then given by $(-p_i^T R_i, 1)^T$. In this notation, the vector $-R_i^T p_i \in \mathbb{R}^3$ is exactly the position of the i^{th} camera center (with respect to the first frame). ■

Comment 4 (Critical surfaces) *Although we have shown that the coefficients of multilinear constraints depend on those of bilinear ones, we have assumed that the coefficients of bilinear constraints, or the corresponding fundamental matrices, are uniquely determined by the epipolar geometry. However, this is not true when all the points lie on critical surfaces. In this case, as argued by Maybank in [13], we may obtain up to three ambiguous solutions from the bilinear constraints. This is the only case when trilinear and quadrilinear constraints provide useful information. On this topic, see also [14].*

In order to extract calibration and motion parameters from the motion matrix, one needs to study how these parameters are embedded in the motion matrix, *i.e.* study the map from the parameter space to $G(3n,4)/\sim$. Now, knowing that bilinear constraints contain all the useful information on the motion matrix, we only need to study how the calibration and motion parameters are encoded in the fundamental matrix. Using the three frame case as an example, mathematically, we need to check if the map from the parameter space to $(G(6,4)/\sim)^3$ is injective. If such a map is not injective, then calibration and motion can be reconstructed only up to a subset of the parameter space.

Solving the full calibration problem in the case of general $SE(3)$ motion requires solution to Kruppa's equations, which has been shown to be difficult and numerically ill-conditioned [10]. In the following section we will consider a simplified problem, where we restrict the class of motions and study to what extent can one recover calibration parameters.

3 Reconstruction under motion subgroups

The goal of this section is to classify all the motions that do *not* allow unique reconstruction of structure, motion and calibration, and for each of them to characterize the ambiguity in the reconstruction. We first quote some results that will be used repeatedly in this section.

3.1 Preliminaries

So far the only restriction we have imposed on the calibration matrix A is that it is non-singular and is normalized as to have $\det(A) = 1$. However, this is not a suitable representation, for different matrices A correspond to the same projection model. In fact, if we consider any rotation matrix R_0 and let $B = AR_0^{-1}$, then from $BR_0q = BR_0Rq_0 + BR_0p$ we can conclude that a point q, moving under (R, p) and viewed with a calibration matrix A cannot be distinguished from a point R_0q moving under $(R_0RR_0^T, R_0p)$ with calibration matrix B. In other words, A can only be determined up to an equivalence class of rotations[4], that is $A \in SL(3)/SO(3)$. Ma et al. ([12]) have shown that, given a sequence of images of corresponding points, the calibration matrix A is not affected by translation, and is uniquely determined (as an element of $SL(3)/SO(3)$) if and only if the rotation R spans at least two independent directions. Since we often refer to this result, we quote it here as a lemma.

Lemma 3 (Ma-Košecká-Sastry) *Consider n uncalibrated images having (purely rotational) motion matrices $M_i = AR_iA^{-1}, i = 1, \ldots, n$, and call $u^i \in \mathbb{R}^3$ the real eigenvectors of R_i. Then the calibration matrix A is uniquely determined (up to an equivalence class of rotations) if and only if at least two of the eigenvectors $u^i, i = 1, \ldots, n$ are linearly independent.*

The condition on the independence of the rotation axes has been established previously by [11] for the case of affine structure and general class of motions.

3.2 Subgroups of the special Euclidean group

The first step in our analysis consists in classifying all continuous Lie subgroups of $SE(3)$;. It is a well known fact of differential geometry[5] that all such subgroups can be generated by left (or right) invariant distributions, and therefore by proper subspaces of the Lie algebra $se(3)$. Given a proper subspace $\Delta \subset se(3)$, let $\bar{\Delta}$ be the smallest subspace of $se(3)$ containing Δ which is *involutive*, *i.e.* $\forall\ V, W \in \bar{\Delta}$ their Lie bracket $[V, W] = VW - WV$ is also in $\bar{\Delta}$. Then the Lie subgroup generated by Δ is given by: [6]

$$G_\Delta = < g = \exp(V) \mid V \in \bar{\Delta} > . \tag{18}$$

On the other hand, every continuous Lie subgroup can be constructed in this way. Since $se(3)$ is finite dimensional, the number of proper Lie subalgebras is finite, and so is the number of continuous Lie subgroups of $SE(3)$. It is then readily seen that a complete list (up to conjugation) of these

[4] A representation of the quotient space is given, for instance, by upper-triangular matrices; such a representation is is often used in modeling calibration matrices.

[5] The reader who is unfamiliar with differential geometry can skip this paragraph without loss of continuity.

[6] The exponential of a matrix $V \in \mathbb{R}^{3\times3}$ is defined to be: $\exp(V) = I + V + \frac{V^2}{2!} + \frac{V^3}{3!} + \cdots$.

subgroups is:

Translational Motion: $(\mathbb{R}^3, +)$ and its subgroups
Rotational Motion: $(SO(3), \cdot)$ and its subgroup $(SO(2), \cdot)$
Planar Motion: $SE(2)$
Cylindrical Motion: $(SO(2), \cdot) \times (\mathbb{R}, +)$
Planar + Elevation: $SE(2) \times (\mathbb{R}, +)$

We are now ready to explore to what extent scene structure, camera motion and calibration can be reconstructed when motion is constrained to one of the above subgroups. In other words, we will study the *generic* ambiguities of the reconstruction problem. In what follows, we use $q(t) = (q_1(t), q_2(t), q_3(t))^T \in \mathbb{R}^3$ to denote the 3D coordinates of the point $q = (q_1, q_2, q_3, 1)^T \in \mathbb{R}^4$ with respect to the camera frame at time t:

$$q(t) = (R(t), p(t))q. \tag{19}$$

3.3 Translational motion ($\mathbb{R}^3$ and its subgroups)

Pure translational motion is generated by elements of $se(3)$ of the form:

$$\xi = \begin{pmatrix} 0 & 0 & 0 & u_1 \\ 0 & 0 & 0 & u_2 \\ 0 & 0 & 0 & u_3 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad u_1, u_2, u_3 \in \mathbb{R}. \tag{20}$$

The coordinate transformation between different views is given by:

$$Aq(t) = Aq(t_0) + Ap(t), \quad p(t) \in \mathbb{R}^3. \tag{21}$$

It is well known (see for instance [12]) that the calibration $A \in SL(3)$ cannot be recovered from pure translational motion, and therefore the corresponding structure q and translational motion p can be recovered only up to the unknown linear transformation A of the projective coordinates. We therefore have the following

Theorem 2 (Ambiguity under $\mathbb{R}^3$) *Consider an uncalibrated camera described by the calibration matrix $A \in SL(3)$, undergoing purely translational motion $\mathbb{R}^3$ (or any of its nontrivial subgroups) and let B be an arbitrary matrix in $SL(3)$. If the camera motion $p \in \mathbb{R}^3$ and the scene structure $q \in \mathbb{R}^4$ are unknown, then B, $B^{-1}Ap$ and $B^{-1}Aq$ are the only generic ambiguous solutions for the camera calibration, camera motion and the scene structure respectively.*

Comment 5 *Thus the group $SL(3)$ can be viewed as characterizing the generic ambiguity of reconstruction under pure translation, and will therefore be called the "ambiguity subgroup".*

In section 2 we have argued that multilinear constraints do not provide additional information. We verify here that, indeed, multilinear constraints do not reduce the generic ambiguity, since

$$\begin{pmatrix} A & 0 & \mathbf{x}_0 & 0 & 0 \\ A & Ap_1 & 0 & \mathbf{x}_1 & 0 \\ A & Ap_2 & 0 & 0 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} A^{-1}B & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} B & 0 & \mathbf{x}_0 & 0 & 0 \\ B & Ap_1 & 0 & \mathbf{x}_1 & 0 \\ B & Ap_2 & 0 & 0 & \mathbf{x}_2 \end{pmatrix} \tag{22}$$

and therefore the two sides of the equation span the same subspace. Hence trilinear constraints are identical for all the ambiguous solutions. One can easily check that the same is true for quadrilinear constraints.

3.4 Rotational motion ($SO(3)$)

Pure rotation is generated by elements of $se(3)$ of the form:

$$\xi = \begin{pmatrix} 0 & \omega_3 & -\omega_2 & 0 \\ -\omega_3 & 0 & \omega_1 & 0 \\ \omega_2 & -\omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \omega_1, \omega_2, \omega_3 \in \mathbb{R}. \tag{23}$$

If any two of $\omega_1, \omega_2, \omega_3$ are zero, the subgroup $SO(2)$ is generated instead. The action of $SO(3)$ transforms the coordinates in different cameras by:

$$Aq(t) = AR(t)q(t_0), \quad R(t) \in SO(3). \tag{24}$$

According to lemma 3, the calibration A can be recovered uniquely, and so can the rotational motion $R(t) \in SO(3)$. However, it is well known that the depth information of the structure cannot be recovered at all. We summarize these facts into the following:

Theorem 3 (Ambiguity under $SO(3)$) *Consider an uncalibrated camera with calibration matrix $A \in SL(3)$ undergoing purely rotational motion $SO(3)$ and let λ be an arbitrary (positive) scalar. If both the camera motion $R \in SO(3)$ and the scene structure $q \in \mathbb{R}^3$ are unknown, then A, R and $\lambda \cdot q$ are the only generic ambiguous solutions for the camera calibration, camera motion and the scene structure respectively.*

Comment 6 *The multiplicative group $(\mathbb{R}^+, \cdot)$ can be viewed as characterizing the ambiguity of the reconstruction under pure rotation. Note that such a group $(\mathbb{R}^+, \cdot)$ acts independently on each point. More properly, for each point the group consists of all smooth functions $\phi : \mathbb{RP}^2 \to \mathbb{R}^+$.*

As for the case of pure translation, there is no independent constraint among three or more images.

3.5 Planar motion ($SE(2)$)

While the previous two cases were of somewhat academic interest and the theorems portray well-known facts, planar motion arises very often in applications. We will therefore study this case in some detail.

Let $e_1 = (1,0,0)^T \in \mathbb{R}^3$, $e_2 = (0,1,0)^T \in \mathbb{R}^3$ and $e_3 = (0,0,1)^T \in \mathbb{R}^3$ be the standard basis of $\mathbb{R}^3$. Without loss of generality, we may assume the camera motion is on the plane normal to e_3 and is represented by the subgroup $SE(2)$

$$SE(2) = \left\{ \begin{pmatrix} R & p \\ 0 & 1 \end{pmatrix} \middle| R = exp(\hat{e}_3\theta), p = (p_1, p_2, 0)^T, \theta \in \mathbb{R} \right\}. \tag{25}$$

This group can be regarded as generated by elements of $se(3)$ of the form:

$$\xi = \begin{pmatrix} 0 & \omega_3 & 0 & u_1 \\ -\omega_3 & 0 & 0 & u_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad u_1, u_2, \omega_3 \in \mathbb{R}. \tag{26}$$

It is readily seen that the rotation matrix $R = exp(\hat{e}_3\theta)$ has the form:

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \theta \in \mathbb{R}. \tag{27}$$

Let A be the unknown calibration matrix of the camera. As described in section 3.1 we consider A as an element of the quotient space $SL(3)/SO(3)$. We will assume that bilinear constraints allow us to estimate a fundamental matrix

$$F = A^{-T}R^TA^T\widehat{p'} \in \mathbb{R}^{3\times3} \tag{28}$$

where $p' = Ap \in \mathbb{R}^3$. In the particular case of pure rotation, we can instead obtain from the data a matrix of the form:

$$C = A^{-T}R^TA^T. \tag{29}$$

In both cases, the rotation matrix is $R = exp(\hat{e}_3\theta)$, which has eigenvalue/eigenvector pairs given by

$$\begin{array}{ll} \lambda_1 = 1, & u^1 = (0,0,1)^T \in \mathbb{R}^3; \\ \lambda_2 = e^{i\theta}, & u^2 = (i,1,0)^T \in \mathbb{C}^3; \\ \lambda_3 = e^{-i\theta}, & u^3 = (-i,1,0)^T \in \mathbb{C}^3 \end{array} \tag{30}$$

where $i = \sqrt{-1}$. Then, following [12], any possible calibration matrix $A \in SL(3)/SO(3)$ is such that the matrix $S = A^{-T}A^{-1}$ is in the *symmetric real kernel* ($SRKer$) of the Lyapunov map:

$$L : \mathbb{C}^{3\times3} \to \mathbb{C}^{3\times3}; \qquad V \mapsto V - CVC^T. \tag{31}$$

We define the matrices $S_1 = A^{-T}u^1(u^1)^TA^{-1}$, $S_2 = A^{-T}u^2(u^2)^TA^{-1}$, $S_3 = A^{-T}u^3(u^3)^TA^{-1}$ so that S has the general form $S = \beta S_1 + s(S_2 + S_3)$. Substituting (30) and also imposing $S \in SL(3)$, we obtain $S = A^{-T}D(s)A^{-1}$, where $D(s) \in \mathbb{R}^{3\times3}$ is a matrix function of s:

$$D(s) = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1/s^2 \end{pmatrix}, \quad s \in \mathbb{R}\setminus\{0\}. \tag{32}$$

If we choose a matrix $A_0 \in SRKer(L)$, and suppose that $A_0B = A$ for some matrix $B \in SL(3)$, we then have that, for some $s \in \mathbb{R}$,

$$A^{-T}D(s)A^{-1} = A_0^{-T}A_0^{-1} \qquad \Rightarrow B^TB = D(s). \tag{33}$$

A solution of (33) is of the form $B = HD(s)$ with $H \in SO(3)$ and $s \in \mathbb{R}$. Let us define a one-parameter Lie group $G_{SE(2)}$ as:

$$G_{SE(2)} = \{D(s) \mid s \in \mathbb{R}\setminus\{0\}\}. \tag{34}$$

Note that there is a natural diffeomorphism between $G_{SE(2)}$ and $\mathbb{R}\setminus\{0\}$ as a multiplication group. Then the solution space of (33) is given by $SO(3)G_{SE(2)}$. The group $G_{SE(2)}$ can be viewed as a natural representation of ambiguous solutions in the space $SL(3)/SO(3)$.

Once we have a calibration matrix, say A_0, we can extract motion from the fundamental matrix $F = A^{-T}R^TA^T\widehat{p'} \in \mathbb{R}^{3\times3}$ as follows: we know that $A = A_0B$ for some $B = HD(s) \in SO(3)G_{SE(2)}$. Then we define $E = A_0^TFA_0$ and note that, for $R = exp(\hat{e}_3\theta)$, we have $D(s)RD(s)^{-1} = R$. Then E is an essential matrix since $E = H^{-T}D^{-T}(s)R^T\hat{p}D^{-1}(s)H^{-1} = HR^TH^T\widehat{HD(s)p}$. The motion recovered from E is therefore $(HRH^T, HD(s)p) \in SE(3)$, where $(R,p) \in SE(2)$ is the true motion. Note that $(HRH^T, HD(s)p)$ is actually a planar motion. The coordinate transformation in the uncalibrated camera frame is given by $Aq(t) = ARq(t_0) + Ap(t)$. If, instead, the matrix A_0 is chosen to justify the camera calibration, the coordinate transformation becomes:

$$\begin{aligned} & A_0 B q(t) = A_0 B R q(t_0) + A_0 B p(t) \\ \Rightarrow \quad & HD(s)q(t) = HRH^T(HD(s)q(t_0)) + HD(s)p(t). \end{aligned} \tag{35}$$

Therefore, any point q viewed with an uncalibrated camera A undergoing a motion $(R,p) \in SE(2)$ is not distinguishable from the point $HD(s)q$ viewed with an uncalibrated camera $A_0 = AD^{-1}(s)H^T$ undergoing a motion $(HRH^T, HD(s)p) \in SE(2)$. We have therefore proven the following

Theorem 4 (Ambiguity under $SE(2)$) *Consider a camera with unknown calibration matrix $A \in SL(3)$ undergoing planar motion $SE(2)$ and let $B(s) = HD(s)$ with $H \in SO(3)$ and $D(s) \in G_{SE(2)}$. If both the camera motion $(R,p) \in SE(2)$ and the scene structure $q \in \mathbb{R}^3$ are unknown, then $AB^{-1}(s) \in SL(3)$, $(HRH^T, B(s)p) \in SE(2)$ and $B(s)q \in \mathbb{R}^3$ are the only generic ambiguous solutions for the camera calibration, camera motion and scene structure respectively.*

Comment 7 *Note that the role of the matrix $H \in SO(3)$ is just to rotate the overall configuration. Therefore, the only generic ambiguity of the reconstruction is characterized by the one parameter Lie group $G_{SE(2)}$. The planar motion and affine reconstruction case was considered in [2]. Linking it with the results of Luong [11] on self-calibration given affine structure one can come to a similar observation that one of the parameters of the calibration matrix is unconstrained.*

Further note that the above ambiguities are obtained only from bilinear constraints between pairs of images. We now verify, as we expect from section 2, that multilinear constraints do not reduce the ambiguity. In fact, the matrix $D(s)$ commutes with the rotation matrix, so that

$$\begin{pmatrix} A & 0 & \mathbf{x}_0 & 0 & 0 \\ AR_1 & Ap_1 & 0 & \mathbf{x}_1 & 0 \\ AR_2 & Ap_2 & 0 & 0 & \mathbf{x}_2 \end{pmatrix} \begin{pmatrix} B^{-1}(s) & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_0 & 0 & \mathbf{x}_0 & 0 & 0 \\ A_0HR_1H^T & Ap_1 & 0 & \mathbf{x}_1 & 0 \\ A_0HR_2H^T & Ap_2 & 0 & 0 & \mathbf{x}_2 \end{pmatrix} \tag{36}$$

Thus trilinear constraints are identical for all ambiguous solutions. A similar result holds for quadrilinear constraints.

3.6 Subgroups $SO(2)$, $SO(2) \times \mathbb{R}$ and $SE(2) \times \mathbb{R}$

We conclude our discussion on subgroups of $SE(3)$ by studying $SO(2)$, $SO(2) \times \mathbb{R}$ and $SE(2) \times \mathbb{R}$ together. This is because their generic ambiguities are similar to the case of $SE(2)$, which we have studied in the previous section. Consider elements of $se(3)$ of the form

$$\xi = \begin{pmatrix} 0 & \omega_3 & 0 & u_1 \\ -\omega_3 & 0 & 0 & u_2 \\ 0 & 0 & 0 & u_3 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad u_1, u_2, u_3, \omega_3 \in \mathbb{R}. \tag{37}$$

which generate the subgroup $SE(2) \times \mathbb{R}$; if $u_1 = u_2 = 0$, they generate the cylindrical motion $SO(2) \times \mathbb{R}$; if $u_1 = u_2 = u_3 = 0$, they generate $SO(2)$.

Notice that in the discussion of the ambiguity $G_{SE(2)}$, we did not use the fact that the translation p has to satisfy $p_3 = 0$. Therefore, *the generic reconstruction ambiguities of $SO(2)\times\mathbb{R}$ and $SE(2)\times\mathbb{R}$ are exactly the same as that of $SE(2)$*. The only different case is $SO(2)$. It is readily seen that the ambiguity of $SO(2)$ is the "product" of that of $SE(2)$ and that of $SO(3)$ due to the fact $SO(2) = SE(2) \cap SO(3)$. As a consequence of Theorem 3 and Theorem 4 we have:

Corollary 2 (Ambiguity under $SO(2)$) *Consider an uncalibrated camera with calibration matrix $A \in SL(3)$ undergoing a motion in $SO(2)$ and let $B(s) = HD(s)$ with $H \in SO(3)$, $D(s) \in G_{SE(2)}$ and $\lambda \in (R^+, \cdot)$. If both the camera motion $R \in SO(3)$ and the scene structure $q \in \mathbb{R}^3$ are unknown, then $AB^{-1}(s) \in SL(3)$, $HRH^T \in SO(3)$ and $\lambda \cdot B(s)q \in \mathbb{R}^3$ are the only generic ambiguous solutions for the camera calibration, camera motion and scene structure respectively.*

Comment 8 *The generic ambiguities for reconstruction under motions in $SO(2)\times\mathbb{R}$ and $SE(2)\times\mathbb{R}$ are characterized by $G_{SE(2)}$; the ambiguities of $SO(2)$ are given by the product group $G_{SE(2)} \times (R^+, \cdot)$. Furthermore, these ambiguities persist even if multilinear constraints are used for the reconstruction.*

Finally, as we have seen in section 3.1, there is no generic ambiguity in calibration, motion and structure reconstruction under the full group $SE(3)$. From the above discussion of subgroups of $SE(3)$ we have seen that generic ambiguity exists for any proper subgroup of $SE(3)$. Furthermore, such ambiguities - which have been derived based only on bilinear constraints, are not affected by multilinear constraints.

4 Reprojection under partial reconstruction

In the previous section we have seen that, in general, it is possible to reconstruct the calibration matrix A and the scene's structure q only *up to a subgroup* - which we call K, the ambiguity subgroup. For instance, in the case of planar motion, an element in K has the form $D(s)$ where $D(s) \in G_{SE(2)}$ a one-parameter group given by equation (32)[7]. Therefore, after reconstruction we have

$$\tilde{q}(K) = Kq, \quad \tilde{A}(K) = AK^{-1}. \tag{38}$$

Now, suppose one wants to generate a novel view of the scene, $\tilde{\mathbf{x}}$ from a new vantage point, which is specified by a motion $\tilde{g} \in SE(3)$ and must satisfy $\tilde{\lambda}\tilde{\mathbf{x}}(K) = \tilde{A}(K)\tilde{g}\tilde{q}(K)$. In general, the reprojection $\tilde{\mathbf{x}}(K)$ depends both on the ambiguity subgroup K and on the vantage point, determined by $\tilde{g}$. A question that arises naturally is what is the set of vantage points that generate a *valid reprojection*, i.e. an image of the scene q taken with the camera A from *some* vantage point $g(K)$. A stronger condition to require is that the reprojection be *invariant* with respect to the ambiguity K, so that we have $g(K) = \tilde{g}$ independent of K.

4.1 Valid Euclidean reprojection

In order to characterize the vantage points, specified by a motion $\tilde{g}$, that produce a valid reprojection we must

$$\text{find } \tilde{g} \quad \text{such that} \quad \tilde{A}(K)\tilde{g}\tilde{q}(K) = Ag(K)q. \tag{39}$$

for some $g(K) \in SE(3)$. Since the reprojected image $\tilde{\mathbf{x}}$ is

$$\tilde{\lambda}\tilde{\mathbf{x}}(K) = \tilde{A}(K)\tilde{g}\tilde{q}(K) = Ag(K)q \tag{40}$$

[7]The ambiguity of planar motion is generally $HD(s)$ with $H \in SO(3)$. Given that A is triangular, such H can be assumed to be the identity without loss of generality.

the characterization of all such motions $\tilde{g}$ is given by the following Lie group:

$$R(K) = \{\tilde{g} \in SE(3) \mid K^{-1}\tilde{g}K \subset SE(3)\}. \quad (41)$$

We call $R(K)$ the *reprojection group* for a given ambiguity group K. For each of the generic ambiguities we studied in section 3, the corresponding reprojection group is given by the following

Theorem 5 *The reprojection groups corresponding to each of the ambiguity groups K studied in section 3 are given by*

1. $R(K) = (\mathbb{R}^3, +)$ *for* $K = SL(3)$ *(ambiguity of* $(\mathbb{R}^3, +)$*).*
2. $R(K) = SO(2)$ *for* $K = G_{SE(2)} \times (\mathbb{R}^+, \cdot)$ *(ambiguity of* $SO(2)$*).*
3. $R(K) = SE(2) \times \mathbb{R}$ *for* $K = G_{SE(2)}$ *(ambiguity of* $SE(2), SO(2) \times \mathbb{R}, SE(2) \times \mathbb{R}$*).*
4. $R(K) = SE(3)$ *for* $K = I$ *(ambiguity of* $SE(3)$*).*

Even though the reprojected image is, in general, not unique, the family of all such images are still parameterized by the same ambiguity group K. For a motion outside of the group $R(K)$, *i.e.* for a $\tilde{g} \in SE(3) \setminus R(K)$, the action of the ambiguity group K on a reprojected image cannot simply be represented as moving the camera: it will have to be a more general non-Euclidean transformation of the shape of the scene. However, the family of all such non-Euclidean shapes are still parameterized by the quotient space $SE(3)/R(K)$[8].

Comment 9 (Choice of a "basis" for reprojection) *Note that in order to specify the viewpoint it is not just sufficient to choose the motion $\tilde{g}$ for, in general, $g(K) \neq \tilde{g}$. Therefore, an imaginary "visual-effect operator" will have to adjust the viewpoint $g(K)$ acting on the parameters in K. However, the ambiguity subgroups derived in section 3 are one-parameter groups (for the most important cases) and therefore* the choice is restricted to one parameter. *In a projective framework (such as [6]), the user has to specify a projective basis of three-dimensional space, that is 15 parameters. This is usually done by specifying the three-dimensional position of 5 points in space.*

It seems inconvenient that, in general, the image reprojected at a given viewpoint is not unique, and therefore one has to choose one (or more) parameters ad-hoc. It seems therefore natural to ask the question of when the reprojection is not only *valid*, but indeed *unique.*

4.2 Invariant reprojection

In order for the reprojection from the viewpoint identified by $\tilde{g}$ to be unique, we must have

$$\tilde{\lambda}\tilde{\mathbf{x}} = \tilde{A}(K)\tilde{g}\tilde{q}(K) = AK^{-1}\tilde{g}Kq \quad (42)$$

independent of K. Equivalently we must have $K^{-1}\tilde{g}K = \tilde{g}$ where K is the ambiguity generated by the motion on a subgroup G of $SE(3)$. The set of $\tilde{g}$ that satisfy this condition is a group $N(K)$, called the *normalizer* of K in $SE(3)$. Therefore, we have to do is to characterize the normalizers for the ambiguity subgroups studied in section 3.

[8]In general, $R(K)$ is not a normal subgroup of $SE(3)$. But if it is, the quotient $SE(3)/R(K)$ is also a group. That is, all the non-Euclidean shapes that show up in the reprojection can also be characterized by a group, as in the case of the generic ambiguity.

Theorem 6 *The set of viewpoints that are invariant to reprojection is given by the normalizer of the ambiguity subgroup. For each of the motion subgroup analyzed in section 3 the corresponding centralizer of the ambiguity group is given by*

1. $N(K) = I$ *for* $K = SL(3)$ *(ambiguity of* $(\mathbb{R}^3, +)$*).*
2. $N(K) = SO(2)$ *for* $K = G_{SE(2)} \times (\mathbb{R}^+, \cdot)$ *(ambiguity of* $SO(2)$*).*
3. $N(K) = SO(2)$ *for* $K = G_{SE(2)}$ *(ambiguity of* $SE(2), SO(2) \times \mathbb{R}, SE(2) \times \mathbb{R}$*).*
4. $N(K) = SE(3)$ *for* $K = I$ *(ambiguity of* $SE(3)$*).*

For motion in every subgroup, the reprojection performed under any viewpoint determined by the groups above is unique.

5 Reconstruction under calibration subgroups

In the previous section we studied reconstruction under different motion subgroups. In this section we explore reconstruction under different calibration subgroups, i.e. when the calibration is represented by a matrix A that is constrained onto a subgroup of $SL(3)/SO(3)$. However, unlike in previous sections where the calibration matrix was unknown but constant, here we will allow $A(t)$ to change in time, therefore describing a trajectory on subgroups of $SL(3)/SO(3)$.

The ultimate goal is to have a complete taxonomy of the ambiguities in reconstruction with with respect to all possible motion and calibration subgroups. It is readily seen that the smaller the motion subgroup and larger the calibration subgroup, the larger the ambiguity.

5.1 Partial, constant calibration

As we have discussed in 3.1, a natural representation of the space $SL(3)/SO(3)$ is the space of all upper-triangular matrices in $SL(3)$, which we denote with K_3. This representation has an immediate geometric interpretation in terms of parameters such as focal length or the principal point of the system. ¿From a practical standpoint the important cases to consider are: known principal point, skew is not present and known aspect ratio. These cases have been previously explored in the literature, both in the general and restricted motion settings. It has been shown that such restrictions of the parameter space reduces the reconstruction ambiguity. Moons et al. [16] have shown that in the case of no skew, known principal point and aspect ratio = 1, the affine structure can be reconstructed from two views related by pure translation. Same calibration subgroup but general class of motions has been considered by [17] which proposed an algorithm for full self-calibration. Finally, Heyden and Åström [9] proved that Euclidean reconstruction is possible, assuming known aspect ratio and no skew.

5.2 General time-varying calibration

We now consider the case where calibration is allowed to vary arbitrarily. In this case, the motion matrix M may be written in terms of the motion and calibration parameters as

$$M = \begin{pmatrix} A(t_1) & 0 \\ A(t_2)R(t_2) & A(t_2)p(t_2) \\ \vdots & \\ A(t_n)R(t_n) & A(t_n)p(t_n) \end{pmatrix} \in \mathbb{R}^{3n \times 4}, \tag{43}$$

where the camera motion is expressed with respect to the first camera frame. Clearly M has overall $(5+6)n-6 = 11n-6$ unknowns. However, as a motion matrix, M is an element of a space of dimension $11n-15$. Therefore, no matter how large n is, the motion $(R(t_i), p(t_i))$ and calibration $A(t_i)$ can never be uniquely determined from M. This has long been known, see for instance [7]. Therefore, following a path analogous to the one traced in the previous sections, we will consider restrictions of the calibration group K_3.

The bilinear constraints associated to fundamental matrices between time instants t_i and t_j

$$F(t_i,t_j) = \lambda(t_i,t_j)A^{-T}(t_j)R^T(t_i,t_j)A^T(t_i)\widehat{p'}(t_i,t_j) \in \mathbb{R}^{3\times 3} \tag{44}$$

where $(R(t_i,t_j), p(t_i,t_j))$ describes the motion between the i^{th} and j^{th} frames, $p'(t_i,t_j) \in \mathbb{R}^3$ is defined to be $A(t_i)p(t_i,t_j)$ and $\widehat{p'} \in \mathbb{R}^{3\times 3}$ is the skew-symmetric matrix associated with p'. The following equations, known as Kruppa's equations, can be interpreted in a projective geometric framework as specifying the absolute conic (see [15] for details), or they can be interpreted in a differential geometric setting as expressing the invariance of the Riemannian metric of the Euclidean space (see [12] for details). In any case they represent constraints on the calibration matrix A and are given by

$$F(t_i,t_j)^T A(t_j)A^T(t_j)F(t_i,t_j) = \lambda^2(t_i,t_j)\widehat{p'}^T(t_i,t_i)A(t_i)A^T(t_i)\widehat{p'}(t_i,t_j). \tag{45}$$

We define, as before, $S^{-1}(t) = A(t)A^T(t)$. Let $B(t_i,t_j) = A(t_j)A^{-1}(t_i)$, and:

$$\tilde{F}(t_i,t_j) = F(t_i,t_j)B^{-1}(t_i,t_j), \quad \tilde{p}(t_i,t_j) = B(t_i,t_j)p'(t_i,t_j). \tag{46}$$

Notice that for fixed t_j, the matrix $\tilde{F}(t_i,t_j)$ thus defined has the form of a fundamental matrix for constant calibration parameters:

$$\tilde{F}(t_i,t_j) = \lambda(t_i,t_j)A^{-T}(t_j)R^T(t_i,t_j)A^T(t_j)\widehat{\tilde{p}}(t_i,t_j) \in \mathbb{R}^{3\times 3}. \tag{47}$$

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

$$\tilde{F}^T(t_i,t_j)S^{-1}(t_j)\tilde{F}(t_i,t_j) = \lambda^2(t_i,t_j)\widehat{\tilde{p}}^T(t_i,t_j)S^{-1}(t_j)\widehat{\tilde{p}}(t_i,t_j). \tag{48}$$

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

Note that the scale λ is introduced for convenience only. To eliminate it, one can write (45), for instance, as a wedge product form:

$$\left(F^T(t_i,t_j)S^{-1}(t_j)F(t_i,t_j)\right) \wedge \left(\widehat{p'}^T(t_i,t_j)S^{-1}(t_i)\widehat{p'}(t_i,t_j)\right) = 0 \tag{49}$$

where we view a 3×3 matrix as a vector in $\mathbb{R}^9$. Although there seem to be many equations in (49), due to the symmetry of the two wedged matrices and the fact that both F and $\widehat{p'}$ are degenerate, there are in fact only *two* algebraically (or linearly) independent constraints.

Since from equation (49) we cannot estimate all the parameters of the calibration, it is only natural to study its restriction to a subset of the parameter set. In particular, it is possible to study the solution on (orbits on the) subgroups of the calibration space K_3. For the sake of example, we show how this works for a time-varying focal length; a similar exercise can be repeated for a time-varying principal point.

Example: time-varying focal length

Equation (49) is the unsimplified version of Kruppa's equations for time-varying parameters. When the only changing parameter is f, without loss of generality we can specialize the above equation by choosing $S^{-1}(t) = A_0 D^2(t) A_0^T$ and

$$A_0 = \begin{pmatrix} 1 & s & u \\ 0 & k & v \\ 0 & 0 & 1 \end{pmatrix} \quad D(t) = \begin{pmatrix} f(t) & 0 & 0 \\ 0 & f(t) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{50}$$

The set of all such $D(t)$ form a group, denoted by K_f. Each Kruppa equation is quadratic in the 2 unknown parameters $f(t_j)^2$ and $f(t_i)^2$, which for simplicity we call f_1^2 and f_2^2 respectively.

Comment 10 *In (49), the symmetric matrix S is different in the two sides of the equation. Such a matrix can be linked with the coefficients of the absolute conic, in a projective geometric framework. Several previous studies have assumed that S in the two sides is the same, for instance [17], sect. 3. However, this is an approximation, and it holds only if the focal length changes "slowly".*

Suppose that the fundamental matrix between time t_i and t_j is given. For simplicity we call it F and partition it into a four blocks as

$$F = \begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix} \tag{51}$$

where $F_1 \in \mathbb{R}^{2\times 2}$ (and the other blocks are sized consequently). Then, assuming $A_0 = I$ without loss of generality, we can re-write the (un-wedged) Kruppa equations as

$$F^T D_1^2 F = -\lambda^2 \widehat{p} D_2^2 \widehat{p}. \tag{52}$$

Given the structure of $D^2 = \mathrm{diag}\{f^{-2}, f^{-2}, 1\}$ we can
re-write the right-hand side of (52) as

$$f_1^{-2} \begin{bmatrix} F_1^T F_1 & F_1^T F_2 \\ F_2^T F_1 & F_2^T F_2 \end{bmatrix} + \begin{bmatrix} F_3^T F_3 & F_3^T F_4 \\ F_3 F_4 & F_4^2 \end{bmatrix} \doteq f_1^{-2}\phi + \psi \tag{53}$$

and the left-hand side as

$$f_2^{-2} \begin{bmatrix} p_3^2 I & p_3 \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \\ -p_3[p_1 \; - p_2] & p_1^2 + p_2^2 \end{bmatrix} + \begin{bmatrix} -p_2^2 & p_1 p_2 & 0 \\ p_1 p_2 & -p_1^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \doteq \lambda^2 f_2^{-2} U + \lambda^2 V \tag{54}$$

Therefore, (52) can be written as

$$f_1^{-2}\phi + \psi = \lambda^2 f_2^{-2} U + \lambda^2 V. \tag{55}$$

If we call u_i^T the rows of U, and similarly with V, ϕ, ψ, we can eliminate λ^2 from the third rows obtaining

$$\lambda^2 = \frac{f_1^2}{f_2^2}\alpha + \frac{1}{f_2^2}\beta \tag{56}$$

where $\alpha = \frac{\phi_3^T u_3}{u_3^T u_3}$ and $\beta = \frac{\psi_3^T u_3}{u_3^T u_3}$. After substitution, we finally obtain

$$f_1^2 f_2^2(\alpha U - \phi) + f_1^2(\alpha V) + f_2^2(\beta U - \psi) + \beta V = 0 \tag{57}$$

The above equation admits a solution as soon as its (matrix) coefficients span at least three independent directions, which is true under general position conditions. Similar calculations can be performed for the case of a time-varying principal point.

Comment 11 *In the above discussion we have tacitly assumed that* $u_3^T u_3 = (p_1^2+p_2^2)(p_1^2+p_2^2+p_3^2) \neq 0$. *This is equivalent to saying that translation cannot be only along the optical axis, since in that case it is not distinguishable from zooming.*

6 Conclusions

When the necessary and sufficient conditions for a unique reconstruction of scene structure, camera motion and calibration are not satisfied, it is still possible to retrieve a reconstruction up to a global subgroup. We characterize such subgroup explicitly for every possible motion of the camera. The reconstructed structure can then be re-projected to generate novel views of the scene. We characterize the "basis" of the reprojection corresponding to each subgroup, and also the motions that generate a unique reprojection. Finally, we consider a time-varying calibration, which is described as a trajectory on a group, and show how restrictions on certain subgroups can be reconstructed.

We achieve these results using only epipolar (i.e. bilinear) geometry, with no use of tensors nor of projective geometry. This is made possible by the core result of this paper, which proves that the coefficients of multilinear constraints can be derived from the coefficients of bilinear constraints alone. Therefore, the only advantage in considering multilinear constraints is in the presence of singular surfaces.

Our future research agenda involves the design of optimal algorithms to perform all (and only!) the parameters that can be estimated from the data based upon their generic ambiguities. This will involve hierarchical estimation on quotient spaces, and will be carried out using tools from Riemannian geometry.

References

[1] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1034 – 1040, 1997.

[2] P. Beardsley and A. Zisserman. Affine calibration of mobile vehicles. In *Europe-China Workshop on Geometric Modeling and Invariants for Computer Vision*, 1995.

[3] B. Boufama, R. Mohr, and F. Veillon. Euclidean constraints for uncalibrated reconstruction. In *ICCV*, pages 466–470, Berlin, Germany, 1993.

[4] S. Carlsson. Multiple image invariance using the double algebra. In *Applications of invariance in computer vision*, 1994.

[5] S. Christy and R. Horaud. Euclidean shape and motion from multiple perspective views via affine iterations. *IEEE PAMI*, 18(11):1098–104, November 1996.

[6] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *ECCV*, pages 563–578. Springer-Verlag, 1992.

[7] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondence between n images. In *Proc. of the IEEE workshop of representation of visual scenes*, 1995.

[8] R. Hartley. Lines and points in three views; and integrated approach. In *Proc. of the Image Understanding Workshop*, 1994.

[9] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *CVPR*, 1997.

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

[11] Q.-T. Luong and T. Vieville. Canonical representations for the geometries of multiple projective views. *ECCV*, pages 589–599, 1994.

[12] Y. Ma, J. Košecká, and S. Sastry. A mathematical theory of camera self-calibration. *UC Berkeley Technical Report UCB/ERL*, 1998.

[13] S. Maybank. *Theory of Reconstruction from Image Motion.* Springer-Verlag, 1993.

[14] S. Maybank and A. Shashua. Ambiguity in reconstruction from images of six points. In *ICCV*, pages 703–8, Bombay, India, 1988.

[15] S. J. Maybank and O. D Faugeras. A theory of self-calibration of a moving camera. *IJCV*, 8(2):123–152, Aug 1992.

[16] T. Moon, L. Van Gool, M. Van Dients, and E. Pauwels. Affine reconstruction from perspective pairs obtained by a translating camera. In J. L. Mundy and A. Zisserman, editors, *Applications of Invariance in Computer Vision*, pages 297–316, 1993.

[17] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. of the IEEE ICCV*, 1998.

[18] M. Pollefeys, L. Van Gool, and M. Proesmans. Euclidean 3d reconstruction from image sequences with variable focal lengths. In *Proc. of the ECCV*, 1996.

[19] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.

[20] B. Triggs. The geometry of projective reconstruction I: Matching constraints and the joint image. *International Journal of Computer Vision*, to appear.

[21] M. Werman and A. Shashua. The study of 3D-from-2D using elimination. In *Proc. of Europe-China Workshop on Geometrical Modeling and Invariants for Computer Vision*, pages 94–101, Xi'an, China, 1995.