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

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 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 meet the conditions, we give explicit formulas for the ambiguities in the reconstructed scene, motion and calibration. Such a characterization is crucial both for designing robust estimation algorithms (that do not try to recover parameters that cannot be recovered), and for generating novel views of the scene by controlling the vantage point. To this end, we characterize explicitly all the vantage points that give rise to a valid Euclidean reprojection regardless of the ambiguity in the reconstruction. We also characterize vantage points that generate views that are altogether invariant to the ambiguity. All the results are presented using simple notation that involves no tensors nor complex projective geometry, and should be accessible with basic background in linear algebra.

1. Introduction

Reconstructing spatial properties of a scene from a number of images taken by an uncalibrated 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 have led to several important and useful results. The problem is that *conditions for a unique Euclidean reconstruction are almost never satisfied in sequence of images of practical interest*. In fact, they require as a necessary condition that the camera undergoes rotation about at least two independent axes, which is rarely the case both in video processing and in autonomous navigation [14].

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

(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 make sure that the image generated portrays a "valid" Euclidean scene?

On our way to answering these questions, we pause to reflect on the nature of multilinear 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 multilinear constraints carry geometric information on the camera system that is not contained in bilinear ones?

1.1. Relation to previous work

The study of ambiguities in Euclidean reconstruction **(i)** arises naturally in the problem of motion and structure recovery and self-calibration from multiple cameras. There is a vast body of literature on this topic, which cannot be reviewed in the limited space allowed. 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 (see for instance [4, 8, 10, 13, 19, 20, 21] and references therein).

It has long been known that in the absence of any *a priori* information about motion, calibration and scene structure, reconstruction can be performed at least up to a projective transformation [6]. Utilizing additional knowledge about the relationship between geometric entities in the image (e.g., parallelism) one can stratify the different levels of reconstructions from projective all the way to Euclidean [3, 5, 6, 18]. At such a level of generality, the conditions on the uniqueness and existence of solutions are restrictive and the algorithms are computationally costly, often exhibiting local minima [12].

The nature of the constraints among images of the same point in different cameras has been studied extensively, and is known to be multilinear (see for instance [7, 10, 20]). The algebraic dependency among constraints **(iii)** has been established by means of elimination [21] or other algebraic geometric tools [9]. However, an explicit characterization of how the information is encoded in different constraints - which is crucial in the design of robust estimation algorithms - is hard to derive by such means.

Recently, Sturm [19] has proposed a taxonomy of critical motions, that is motions which do not allow a unique reconstruction. However, not only the given taxonomy is by no means intrinsic to Euclidean reconstruction (see [14]), but also no explicit characterization of the ambiguities in the reconstructed shape, motion and calibration has been given. A natural continuation of these efforts involved the analysis of cases where the motion and/or calibration were restricted either to planar or linear motion [2, 18] and techniques were proposed for affine reconstruction or up to one parameter family.

Several techniques have been proposed to synthesize novel views of a reconstructed scene **(ii)**: in [1], trilinear constraints have been exploited to help generate reprojected images for a calibrated camera. In the case of a partially uncalibrated camera, such a method has to face the issues of whether the reprojected image portrays a valid Euclidean scene.

1.2. Outline of this paper and its contributions

As we anticipated in the previous section, the answer to question **(iii)** has been established before on an algebraic footing – the algebraic ideals generated by trilinear and quadrilinear constraints (as polynomials of image coordinates) are necessarily contained in that generated by bilinear ones [9]. However, in order to give a complete account of ambiguities in 3D Euclidean reconstruction (especially for self-calibration and motion recovery), it is crucial to know how the information on the Euclidean configuration of a camera system is encoded in the multilinear constraints. In section 2 we give a novel, complete and rigorous proof that unveils how the information encoded in trilinear and quadrilinear constraints depends on that in bilinear ones. There we also discuss the role of multilinear constraints with regards to singular configurations of points.

The well-known - but conservative - answer to question **(i)** is that structure can at least 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 for the case of constant calibration.[1] There, we give explicit formulas of exact ambiguities in the reconstruction of scene structure, camera motion and calibration with respect to all subgroups of the Euclidean motion. In principle, one should study ambiguities corresponding to all critical configurations as given in [14]. However, it is only the

[1] In fact, even in the case of time-varying calibration, in principle, the best one can do is an affine reconstruction, not just a projective one!

ambiguities that exhibit a *group structure* that are of practical importance in the design of estimation algorithms. In such a case, not only can the analysis be considerably simplified but also clean formulas for all generic ambiguities can be derived. Such formulas are important for 3D reconstruction as well as for synthesizing novel 2D views.

Question **(ii)** is then answered in section 4, where we characterize the complete set of vantage points that generate "valid" images of the scene regardless of generic ambiguities in 3D reconstruction.

These results have great practical significance, because they quantify precisely to what extent scene structure, camera motion and calibration can be estimated in sequences for which many of the techniques available todate do not apply. Furthermore, the analysis clarifies the process of 2D view synthesis from novel viewpoints. In addition to that, we give a novel account of known results on 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 terse as possible. Our tools are borrowed from linear algebra and some differential geometry, although all the results should be accessible without background in the latter. We use the language of (Lie) groups because that allows us to give an explicit characterization of all the ambiguities in a concise and intuitive fashion. Traditional tools involved in the analysis of self-calibration involved complex loci in projective spaces (e.g., the "absolute conic"), which can be hard to grasp for someone not proficient in algebraic geometry.

2. Dependency of multilinear constraints revisited

We model the world as a collection of points in a 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 intrinsic parameters of the camera. Without loss of generality we will re-scale the above equation so that the determinant of A is 1. The set of 3×3 matrices with determinant one is called Special Linear group denoted by $SL(3)$. 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 $SO(3)$; $g(t) = (R(t), p(t))$ belongs to $SE(3)$, 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

$$M_i \doteq (A(t_i)R(t_i), A(t_i)p(t_i)) \in \mathbb{R}^{3\times 4} \tag{2}$$

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

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

which we re-write in a more compact notation as $\mathbf{X}\vec{\lambda} = Mq$. We call $M \in \mathbb{R}^{3n\times 4}$ the *motion matrix* and $\mathbf{X}$ the *image matrix.*

2.1. Constraints on multiple images

Let $\vec{\mathbf{m}}_i \in \mathbb{R}^{3n}, i = 1, \ldots, 4$ denote the four columns of the matrix M and $\vec{\mathbf{X}}_i \in \mathbb{R}^{3n}, i = 1, \ldots, n$ 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 equation:

$$\vec{\mathbf{m}}_1 \wedge \vec{\mathbf{m}}_2 \wedge \vec{\mathbf{m}}_3 \wedge \vec{\mathbf{m}}_4 \wedge \vec{\mathbf{X}}_1 \wedge \cdots \wedge \vec{\mathbf{X}}_n = 0. \tag{3}$$

This constraint, which is multilinear in the measurements $\mathbf{x}(t_i)$ simply expresses the fact that the columns of M and $\mathbf{X}$ are linearly dependent. 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, the coefficients of all the multilinear constraints are minors of the motion matrix M. As it has been shown (see, for instance, Triggs in [20]), constraints involving more than four frames are necessarily dependent on quadrilinear, trilinear and bilinear ones. In this section we go one step further to discuss how trilinear and quadrilinear constraints are dependent on bilinear ones.

When studying the dependency among constraints, one must distinguish between *algebraic* and *geometric* dependency. Roughly speaking, algebraic dependency concerns the conditions that a point in an image must satisfy in order to be the correspondent of a point in another image. Vice versa, geometric dependency is concerned with the information that corresponding points give on the operator that maps one to the other. The two notions are related but not equivalent, and the latter bears important consequences when one is to use the constraints in optimization algorithms to recover structure and calibration. While the geometric dependency of multilinear constraints has been established before under the assumption of constant calibration [10], we give a novel, simple and rigorous proof that is valid under the more general assumption of time-varying calibration.

2.2. Algebraic vs. geometric dependency

To clarify the relation between algebraic and geometric dependency[2], note that in general we can express a multilinear constraint in the form: $\sum_j \alpha_j(M)\beta_j(\mathbf{X}) = 0$ where α_j are some polynomials of entries of M and β_j polynomials of entries of the image coordinates, with M and $\mathbf{X}$ defined as before. α_j's are called the *coefficients* of multilinear constraints. Studying the *algebraic dependency* between constraints then corresponds to fixing the coefficients α_j and asking whether there are some additional constraints among the image coordinates $\mathbf{X}$ generated by three and four views[3]. This problem has been studied many researchers and an elegant answer can be found in [9] by explicitly characterizing the *primary decomposition* of the ideal (in the polynomial ring of image coordinates $\mathbf{x}_i$'s) generated by the bilinear constraints in terms of that generated by trilinear ones or quadrilinear ones.

Geometric dependency, on the other hand, investigates whether, given the image coordinates $\mathbf{X}$, the coefficients α_j corresponding to motion parameters in additional views can give additional information about M. These two different types of dependencies were previously pointed out (see for instance the work of Heyden [10]). For both types of dependencies, the answer is negative, *i.e.,* trilinear and quadrilinear constraints in general are dependent of bilinear ones. We here give a simple but rigorous study of the geometric dependency. The results will also validate the ambiguity analysis given in following sections.

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 manifold and denoted by $G(m, k)$. Therefore, M is an element of $G(9, 4)$. By just re-arranging the three blocks $M_i,\ i = 1, \ldots, 3$ into three pairs,

[2] This subsection is for the benefit of the reader already familiar with existing work on the algebraic dependency among multilinear constraints. The reader who is not at ease with algebraic geometry or unfamiliar with the existing literature can skip this subsection without loss of continuity

[3] In other words, it addresses the dependency among algebraic ideals associated with the three types of multilinear constraints.

(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 $G(6,4)$

$$\begin{array}{rcl} \phi : G(9,4) & \to & G(6,4) \times G(6,4) \times G(6,4) \\ \begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix} & \mapsto & \left(\begin{pmatrix} M_1 \\ M_2 \end{pmatrix}, \begin{pmatrix} M_2 \\ M_3 \end{pmatrix}, \begin{pmatrix} M_1 \\ M_3 \end{pmatrix} \right). \end{array}$$

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. 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:

$$M \sim M' \; if \; \exists \, \lambda_i \in \mathbb{R}^*, M_i = \lambda_i M_i', \quad i = 1, \ldots, n \tag{4}$$

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

We are now ready to prove that coefficients α_j's in trilinear and quadrilinear constraints depend on those in bilinear ones.

Theorem 1 (Geometric dependency) *Given three (or four) views, the coefficients of all bilinear constraints or equivalently the corresponding fundamental matrices uniquely determine the motion matrix M as an element in $G(9,4)/\sim$ (or $G(12,4)/\sim$) given that $Ker(M_i)$'s are linearly independent.*

Proof: It is known that between any pair of images (i,j) the motion matrix: $\begin{pmatrix} M_i \\ M_j \end{pmatrix} \in G(6,4)$, is determined by the corresponding fundamental matrix F_{ij} up to two scalars λ_i, λ_j: $\begin{pmatrix} \lambda_i M_i \\ \lambda_j M_j \end{pmatrix} \in G(6,4), \quad \lambda_j \in \mathbb{R}^*$. Hence for the three view case all we need to prove is that the map:

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

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 $\lambda_i \in \mathbb{R}^*$ and $G_i \in GL(4)$,[5] $i = 1,2,3$. This yields $M_1(\lambda_1 G_1 - G_3) = 0$, $M_2(\lambda_2 G_2 - G_1) = 0$, $M_3(\lambda_3 G_3 - G_2) = 0$. 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:

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

Combining these three equations, we obtain:

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

The matrix on the right hand side of the equation has a non-trivial null space since its column vectors are in the space $span\{Ker(M_1), Ker(M_2), Ker(M_3)\}$ 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

[4] 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.

[5] $GL(4)$ is the general linear group of all non-degenerate 4×4 real matrices.

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 $\lambda_1 G_1 = G_3$, $\lambda_2 G_2 = G_1$, $\lambda_3 G_3 = G_2$. This implies

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

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

In the case of four views, 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. This directly follows from the above proof of the three frame case. ∎

Comment 1 *As a consequence of the theorem, coefficients α_j's in trilinear and quadrilinear constraints are functions of those in bilinear ones. 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. We emphasize that the geometric dependency does* not *imply that two views are sufficient for reconstruction! It claims 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 and camera (see comment 2). For four views, the condition that $Ker(M_i), i = 1, \ldots, 4$ are linearly independent is not necessary. A less conservative condition is that there exist two groups of three frames which satisfy the condition for the three view case.*

Theorem 1 requires that the one-dimensional kernels of the matrices $M_i, i = 1, \ldots, n$ ($n = 3$ or 4) are linearly independent. Note that the kernels of M_i for $i = 1, 2, 3, 4$ are given by $(-p_i^T R_i, 1)^T$, where the vector $R_i^T p_i \in \mathbb{R}^3$ is exactly the position of the i^{th} camera center. Hence the condition of the theorem is satisfied if and only if the centers of projection of the cameras generate 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.

Comment 2 (Critical surfaces and motions) *Although we have shown that the coefficients of multilinear constraints depend on those of bilinear ones, we have assumed that the latter (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 [15], we may obtain up to three ambiguous solutions from the bilinear constraints. This is one of the cases when trilinear and quadrilinear constraints provide useful information. On this topic, see also [16]. Also, when the camera is undergoing a rectilinear motion (i.e., all optical centers are aligned), trilinear constraints provide independent information in addition to bilinear ones. This fact has been pointed out before; see for instance Heyden in [11].*

3. Reconstruction under motion subgroups

The goal of this section is to study all "critical" motion groups that do *not* allow unique reconstruction of structure, motion and calibration. While a *classification* of such critical motions has been presented before (see [14]), we here go well beyond by giving an *explicit characterization* of the ambiguity in the reconstruction for each critical motion. Such an explicit characterization is crucial in deriving the ambiguity in the generation of novel views of a scene, which we study in section 4.

In this section, we characterize the generic ambiguity in the recovery of (a) structure, (b) motion and (c) calibration corresponding to each possible critical motion. A subgroup of $SE(3)$ is called *critical* if the reconstruction is not unique when the motion of the camera is restricted to it. For the purpose of this section, we assume that the calibration matrix A is constant.

3.1. Some preliminaries

So far the only restriction we have imposed on the constant calibration matrix A is that it is non-singular and is normalized as to have $\det(A) = 1$. However, A can only be determined up to an equivalence class of rotations, that is $A \in SL(3)/SO(3)$.[6] For more detail, please see [14]. The unrecoverable rotation in our choice of A simply corresponds to a rotation of the entire camera system. We borrow the following statement directly from [14]:

Theorem 2 (Necessary and sufficient condition for a unique calibration) *For a set of camera motions $\{(R_i, p_i)\}$ where none of the rotation component R_i is of the form $e^{\widehat{u}_i k\pi}$ with $\|u_i\| = 1, k \in \mathbb{Z}$, the camera calibration A as an element in $SL(3)/SO(3)$ is uniquely determined if and only if at least two of the axes u_i's are linearly independent.*

Although the necessity of the independence of the rotation axes has been long known in the literature (see e.g. [13]), the sufficiency is not proven till recent [14]. This theorem states a very important and useful fact: the condition for a unique calibration has nothing to do with translation (as opposed to the results given in [19])! [7] See [14] for the detail. Due to this theorem, all proper continuous subgroups of $SE(3)$ except $SO(3)$ are critical for self-calibration. So the first step in our analysis consists in classifying all continuous Lie subgroups of $SE(3)$ which are critical. It is a well known fact that a complete list of subgroups of $SE(3)$ can be classified by all Lie subalgebras of the Lie algebra $se(3)$ of $SE(3)$ and then exponentiate them. It is then straightfoward to show that each of these subgroups must have the same ambiguity in reconstruction as one in the following list (as we will explain in the comments):

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

Comment 3 *The above list is by no means a complete list of ALL subgroups of $SE(3)$. For example, the "planar orbital motion", i.e., camera moving on a circle with the optical axis always facing the center, is none of the motion in the above list. However, it is can be treated as a special case of the planar motion since, as far as reconstruction is concerned, they obviously have the same generic ambiguities. In order to show that all subgroups have the same ambiguity in reconstruction as one of the above motions, we must go through all the possible Lie subalgebras of $se(3)$. It can be shown that, if a Lie subalgebra has at least 4 dimension and has two independent rotation components, then it must be $se(3)$ itself. Now the only interesting case is some three dimensional Lie subalgebras which, without loss of generality, are generated by elements:*

$$X = \begin{pmatrix} \widehat{e}_1 & u \\ 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} \widehat{e}_2 & v \\ 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} \widehat{e}_3 & w \\ 0 & 0 \end{pmatrix} \tag{5}$$

where e_1, e_2, e_3 are standard basis of $\mathbb{R}^3$ and u, v and w are three vectors in $\mathbb{R}^3$. In order for the Lie algebra generated by X, Y, Z is three dimensional. We must have the vector $\alpha = (u^T, v^T, w^T)^T \in \mathbb{R}^9$ in the null space of the matrix:

$$Q = \begin{pmatrix} \widehat{e}_2 & -\widehat{e}_1 & I \\ I & \widehat{e}_3 & -\widehat{e}_2 \\ -\widehat{e}_3 & I & \widehat{e}_1 \end{pmatrix}. \tag{6}$$

[6] Here take left cosets as elements in the quotient space. A representation of this quotient space is given, for instance, by upper-triangular matrices; such a representation is commonly used in modeling calibration matrices by means of physical parameters of cameras such as focal length, principal point and pixel skew.

[7] This is because we here only consider the *generic ambiguity* in reconstruction, *i.e.,* such ambiguity exists no matter what the camera sees and no matter what the algorithms do.

That is $Q\alpha = 0$. If $\alpha = 0$, then the subgroup generated by the algebra is just the pure rotation group $SO(3)$. If $\alpha \neq 0$, then the subgroup generated contains three indenpendent rotation axes and translation (parallax). For such subgroups, a unique reconstruction is available. That is, they are not critical for reconstruction or have the same ambiguity as the full rigid body motion $SE(3)$.

We are now ready to explore to what extent scene structure, camera motion and calibration can be reconstructed when motion is constrained onto 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: $q(t) = (R(t), p(t))q$. To simplify notation, for any $u \in \mathbb{R}^3$ we define $\widehat{u}$ to be a 3 skew-symmetric matrix such that $\forall v \in \mathbb{R}^3$ the cross product $u \times v = \widehat{u}v$.

3.2. Generic ambiguities in structure, motion and calibration

Translational motion ($\mathbb{R}^3$ and its subgroups). The coordinate transformation between different views is given by $Aq(t) = Aq(t_0) + Ap(t), p(t) \in \mathbb{R}^3$. According to Theorem 2, 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 transformation A. We therefore have the following

Theorem 3 (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.*

Note that this ambiguity corresponds exactly to an affine reconstruction [18].

Rotational motion ($SO(3)$). The action of $SO(3)$ transforms the coordinates in different cameras by $Aq(t) = AR(t)q(t_0), R(t) \in SO(3)$. According to Theorem 2, 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 4 (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.*

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 more detail.

Let $e_1 = (1, 0, 0)^T, e_2 = (0, 1, 0)^T, e_3 = (0, 0, 1)^T \in \mathbb{R}^3$ 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 $SE(2)$.

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 $SL(3)/SO(3)$. According to [14], any possible calibration matrix $A_0 \in SL(3)/SO(3)$ is such that the matrix $S = A_0^{-T} A_0^{-1}$ is in the *symmetric real kernel* ($SRKer$) of the Lyapunov map for all $C = A^{-T}R^TA^T, R \in SO(2)$:

$$L : \mathbb{C}^{3\times3} \to \mathbb{C}^{3\times3}; \quad X \mapsto X - CXC^T. \tag{7}$$

By the choice of e_1, e_2, e_3, the real eigenvector of R is e_3. Imposing $S \in SL(3)$, we obtain $S = A^{-T} D(s) A^{-1}$, where $D(s) \in \mathbb{R}^{3\times 3}$ is a matrix function of s:

$$D(s) = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1/s^2 \end{pmatrix}, \quad s \in \mathbb{R} \setminus \{0\}. \tag{8}$$

Geometrically, this reveals that only metric information within the plane can be recovered while the relative scale between the plane and its normal direction cannot be determined. If we choose an erroneous matrix A_0 from the set of possible solutions for calibration, then $A_0 B = A$ for some matrix $B \in SL(3)$. Since $A_0^{-T} A_0^{-1}$ is necessarily in $SRKer(L)$, we further have that, for some $s \in \mathbb{R}$,

$$A_0^{-T} A_0^{-1} = A^{-T} D(s) A^{-1} \;\Rightarrow\; B^T B = D(s). \tag{9}$$

A solution of (9) 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:

$$G_{SE(2)} = \{D(s) \mid s \in \mathbb{R} \setminus \{0\}\}. \tag{10}$$

Then the solution space of (9) 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 $F = A^{-T} R^T A^T \widehat{p'}$ as follows: we know that $A = A_0 B$ for some $B = HD(s) \in SO(3)G_{SE(2)}$. Then we define $E = A_0^T F A_0$ and note that, for $R = \exp(\widehat{e}_3 \theta)$, we have that $D(s)$ commutes with R *i.e.,* $D(s) R D(s)^{-1} = R$. Then E is an essential matrix since $E = H^{-T} D^{-T}(s) R^T \widehat{p} D^{-1}(s) H^{-1} = HR^T H^T \widehat{HD(s)} p$. 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* (in a plane rotated by H from the original one). 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:

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

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 5 (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 4 *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 $G_{SE(2)}$.*

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 just studied. 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 4 and Theorem 5 we have:

Corollary 1 (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.*

From the above discussion of subgroups of $SE(3)$ we have seen that generic ambiguities exist for many proper subgroup of $SE(3)$. Furthermore, such ambiguities - which have been derived above based only on bilinear constraints, are not resolved by multilinear constraints according to Theorem 1.

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 $D(s)$ given by equation (8). Therefore, after reconstruction we have

$$\tilde{q}(K) = Kq, \quad \tilde{A}(K) = AK^{-1}. \tag{11}$$

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 $\tilde{g}$ and there is no guarantee that it is an image of the original Euclidean scene.

It is only natural, then, to ask *what is the set of vantage points that generate a valid reprojection*, that is an image of the original scene q taken as if the camera A was placed at some vantage point $g(K)$. We discuss this issue in section 4.1. A stronger condition to require is that the reprojection be independent (*invariant*) of the ambiguity K, so that we have $g(K) = \tilde{g}$ regardless of K; we discuss this issue in section 4.2.

4.1. Valid Euclidean reprojection

In order to characterize the vantage points - specified by motions $\tilde{g}$ - that produce a valid reprojection we must find $\tilde{g}$ such that: $\tilde{A}(K)\tilde{g}\tilde{q}(K) = Ag(K)q$ for some $g(K) \in SE(3)$. Since the reprojected image $\tilde{\mathbf{x}}$ is $\tilde{\lambda}\tilde{\mathbf{x}}(K) = \tilde{A}(K)\tilde{g}\tilde{q}(K) = Ag(K)q$, the characterization of all such motions $\tilde{g}$ is given by the following Lie group:

$$R(K) = \{\tilde{g} \in SE(3) \mid K^{-1}\tilde{g}K \subset SE(3)\}. \tag{12}$$

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 6 *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 minimally parameterized by the quotient space $SE(3)/R(K)$.

Comment 5 (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. 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.*

4.2. Invariant reprojection

In order for the view taken from $\tilde{g}$ to be unique, we must have

$$\tilde{\lambda}\tilde{\mathbf{x}} = \tilde{A}(K)\tilde{g}\tilde{q}(K) = AK^{-1}\tilde{g}Kq \tag{13}$$

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 $N(K)$, the so called *normalizer* of K in $SE(3)$. Therefore, all we have to do is to characterize the normalizers for the ambiguity subgroups studied in section 3.

Theorem 7 *The set of viewpoints that are invariant to reprojection is given by the normalizer of the ambiguity subgroup. For each of the motion subgroups analyzed in section 3 the corresponding normalizer 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 motions in every subgroup, the reprojection performed under any viewpoint determined by the groups above is unique.

5. 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 action (on the entire configuration of the camera system). We characterize such subgroups explicitly for all possible motion groups 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. We achieve the goal by using results from two view analysis [14]. This is possible because the coefficients of multilinear constraints are geometrically dependent of those of bilinear constraints. Therefore, the only advantage in considering multilinear constraints is in the presence of singular surfaces and rectilinear motions. Our future research agenda involves the design of optimal algorithms to recover all (and only!) the parameters that can be estimated from the data based upon their generic ambiguities. The reconstruction and reprojection problem studied in this paper is for a constant calibration matrix. We will present generalized results for the time-varying case in future work.

Acknowledgment

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

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? *INRIA Research Report*, No. 3225, July, 1997.

[7] O. Faugeras and T. Papadopoulo. Grassmann-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. 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. Algebraic properties of multilinear constraints. *Mathematical Methods in Applied Sciences*, 20(13):1135–1162, 1997.

[10] A. Heyden, G. Sparr, and K. Åström. Perception and action using multilinear forms. In G. Sommer and J. Koenderink, editors, *Algebraic Frames for the Perception-Action Cycle*, pages 54–65. Springer-Verlag, 1997.

[11] A. Heyden. Reduced multilinear constraints – theory and experiments. In *International Journal of Computer Vision*, vol. 30, no. 2, pages 5-26, 1998.

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

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

[14] Y. Ma, R. Vidal, J. Košecká, and S. Sastry. Camera self-calibration: Geometry and Algorithms. *Submitted to IEEE transactions on PAMI and also see UC Berkeley Technical Report* UCB/ERL No. M99/32, June 1999.

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

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

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

[18] 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.

[19] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. *CVPR*, pages 1100–1105, 1997.

[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.