

Structure from Motion: a new look from the point of view of invariant theory. *

Pierre-Louis Bazin¹

Mireille Boutin²

¹ Department of Engineering,
² Division of Applied Mathematics,
Brown University
Providence RI 02912, USA.

July 2, 2003

Abstract

We present a novel, simple formulation of the problem of 3D object reconstruction from images. In this formulation, the object is seen as lying at the intersection of the projection of orbits of custom built Lie groups actions. The group parameters correspond to unknown, irrelevant quantities such as the camera orientation, the depth parameters of the object with respect to the camera and the focal length. We then use an algorithmic method based on moving frames *à la* Fels-Olver to obtain a fundamental set of invariants of these groups actions. The invariants are used to define a set of equations determining the 3D object, thus providing a mathematical formulation of the problem where the irrelevant parameters do not appear.

1 Introduction

This paper has two goals. Its first goal is to illustrate the potential of using the formalism of invariant theory in certain applications. This potential is, at this point, rather unexploited and we hope to set a trend with these results. Its second goal is to provide new insights on the problem of structure from motion through a novel formulation in terms of group actions.

The problem of structure from motion is rather old and well-studied. It consists in reconstructing an object from a set of pictures of this object (e.g. a movie). In this paper, we consider the case of objects represented by an ordered set of points in \mathbb{R}^3 and assume that the camera parameters (position and orientation of the camera, focal length) are unknowns.

*This work was supported by NSF grants KDI BCS-9980091 and 0074276.

The concept of invariance is of major importance in modern geometry. In the field of computer vision, invariants of classical groups have been used in the design of methods of object recognition and reconstruction for more than a decade (see [12, 13]). In particular, the invariants of the projective and affine transformation groups have been widely used [16, 21, 18]. However, invariant theory can also deal with a variety of other group actions such as the ones we encounter in the problem of structure from motion.

For our purposes, invariants are defined as real-valued functions on a manifold M which remain unchanged under the action (denoted by $*$) of a group G on M . The case of Lie group actions is particularly interesting. (The reader unfamiliar with the concept of Lie group actions may refer to [8] for an introduction.) When the Lie group action satisfies certain conditions, there exists a finite set of *fundamental* invariants I_1, \dots, I_N which are local coordinates for the quotient space M/G . In other words, for any point $z \in M$ there exists a neighborhood U of z which can be written as $U = U_1 \times U_2$ where U_1 is coordinatized by the value of the invariants, and U_2 is coordinatized by some (or all) of the group parameters. A modern theory of moving frames recently developed by Fels and Olver [5, 6] provides us with a systematic way to obtain a set of fundamental invariants for any (regular) Lie group action.

One interest of being able to obtain a set of fundamental invariants in a systematic manner is the following. Many problems involve unknown, irrelevant parameters. Oftentimes, one needs to solve for these irrelevant parameters only because they are involved in the intermediate steps of the solution process, although they do not appear in the final solution. When the irrelevant parameters can be seen as group parameters transforming the other unknowns of the problem, using the coordinates provided by the invariants is a simple way to eliminate them. In these circumstances, the moving frame method is used as a computational method to eliminate unwanted unknowns in a set of equations.

For example, suppose that given the values of $x_1, x_2, y_1, y_2 \in \mathbb{R}$ one is interested in finding the values of the unknowns z_1 and $z_2 \in \mathbb{R}$. Assume that there exists parameters u and $v \in \mathbb{R}$ for which a set of equations of the type

$$\left. \begin{aligned} z_1 &= f_1(x_1, x_2, u) \\ z_2 &= f_2(x_1, x_2, u) \\ z_1 &= g_1(y_1, y_2, v) \\ z_2 &= g_2(y_1, y_2, v) \end{aligned} \right\} \quad (1)$$

holds. Since we are not interested in the values of u and v , it would be desirable to eliminate these two variables from (1). Suppose that the functions f_1 and f_2 correspond to an action of \mathbb{R} on \mathbb{R}^2 parameterized by u ,

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = u * \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

If this group action satisfies certain conditions (to be explained in Section 3), then there exists an invariant $I : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that the equation

$$I(x_1, x_2) = I(z_1, z_2)$$

can be used in place of the first two equations of (1). Similarly, if g_1 and g_2 correspond to an action of \mathbb{R} on \mathbb{R}^2 parameterized by v which satisfies the correct conditions, then we can find another invariant $J : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that the equation

$$J(y_1, y_2) = J(z_1, z_2)$$

can be used in place of the last two equations of (1). We thus see that the solution (z_1, z_2) of our problem lies at the intersection of two orbits: the orbit through (x_1, x_2) under the group action defined by f_1 and f_2 and the orbit through (y_1, y_2) under the group action defined by g_1 and g_2 .

The problem of structure from motion is one that involves many irrelevant unknowns. For example, the camera parameters used for taking each picture are irrelevant since we are merely interested in the structure of the object. It turns out that many of the irrelevant unknowns of this problem can be seen as group parameters acting on the other unknowns and eliminated using the moving frame method.

We begin our exposition in Section 2 with a summary of the relevant theoretical aspects of the theory of moving frames and its application to the computation of (joint) invariants. In Section 3, we consider the problem of structure from motion when the pictures are taken using a pinpoint camera (perspective projections). We reformulate the problem in terms of a Lie group action and use a fundamental set of invariants of this group action to get rid of most of the irrelevant parameters. (As will be explained, it is impossible to get rid of all of them.) In the case of a fixed-focus camera, the invariants describe the object with non-linear equations involving the object points and the camera centers: the camera orientation and the depth parameters of the object with respect to the camera are included in the group parameters and thus eliminated from the problem. Eliminating the camera orientation parameters in the case of a simultaneously translating and rotating camera is known to be difficult, unless vanishing points are identified. By contrast, the invariant framework easily takes care of this problem. In addition, the invariant formulation leads to a simple test to identify camera motions that are pure rotations.

The non-linear equations given by the invariants can be solved with regular bundle adjustment techniques [20] where the non-linear system of equations is viewed as an optimization problem in the unknown parameters. Efficient optimization techniques like the Levenberg-Marquardt algorithm are available to solve such problems. After presenting a practical example of 3D object reconstruction using the invariant framework and such an optimization technique, we finish Section 3 by showing how to deal with the case of a variable focus in a similar fashion.

Of course, other approaches than this *direct* one exist for structure from motion. For example, one can describe the perspective projection in projective space where it is expressed by simple matrix equations. With projective coordinates, projective relations and constraints such as the epipolar constraint or the trifocal tensor can be used to recover both the underlying shape and camera motion from a set of multi-linear equations [9]. In Section 4, we apply our invariant-based method to a formulation using projective coordinates. We obtain a set of invariants involving the projective coordinates of the object points and the camera centers which are, in fact, much simpler than the invariants of the Euclidean approach used in Section 3.

Other existing approaches use an affine approximation of the projection in order to simplify the problem. This leads to factorization techniques [19] based on the SVD decomposition. In Section 5, we apply our method to the case of orthographic projections and obtain a set of linear invariants involving only the object points and the directions of the normals to the camera planes. Our equations can be solved with similar factorization techniques but involve fewer and simpler parameters than the ones in [19].

Our invariant-based approach actually applies to other types of cameras than the ones discussed here. For instance, the model used in Section 3 works for any central projection, whether or not the image points lie on a plane. We hope that the sample cases presented here will convince the reader of the usefulness and versatility of this group theoretic approach to eliminating unknowns and inspire new applications of invariants.

2 Definitions and Theoretical Foundations

Let M be an m -dimensional smooth (Hausdorff) manifold and G be an r -dimensional Lie group. Denote by e the identity in G . Let $*$: $G \times M \rightarrow M$ be an action of G on M , i.e. a map $(g, z) \mapsto g * z$ such that $e * z = z$, for all $z \in M$ and $(gh) * z = g * (h * z)$, for all $z \in M$ and all $g, h \in G$.

Definition 2.1. An *invariant* is a function $I : M \rightarrow \mathbb{R}$ which remains unchanged under the action of the group. In other words,

$$I(g * z) = I(z), \text{ for all } z \in M \text{ and all } g \in G.$$

A *local invariant* is a function $I : U \rightarrow \mathbb{R}$, for some open subset $U \subset M$, such that

$$I(g * z) = I(z), \text{ for all } z \in U \text{ and all } g \in G \text{ s.t. } g * z \in U.$$

Definition 2.2. We say that G acts *semi-regularly* on M if all orbits have the same dimension. If, in addition, any point $p_0 \in M$ is surrounded by an arbitrarily small neighborhood whose intersection with the orbit through p_0 is connected, then we say that G acts *regularly*.

The following theorem, due to Frobenius [7], is of central importance to our approach. A proof can be found in [14]. It provides us with a simple way of characterizing the orbits using invariants.

Theorem 2.3 (Frobenius Theorem). *If G acts on an open set $O \subset M$ semi-regularly with s -dimensional orbits, then $\forall p_0 \in O$ there exist $m - s$ functionally independent local invariants I_1, \dots, I_{m-s} defined on a neighborhood U of p_0 such that any other local invariant H defined near p_0 is a function $H = f(I_1, \dots, I_{m-s})$. If G acts regularly on O , then we can choose I_1, \dots, I_{m-s} to be global invariants on O . In that case, two points $p_1, p_2 \in O$ are in the same orbit relative to G if and only if $I_i(p_1) = I_i(p_2)$, for all $i = 1, \dots, m - s$.*

By *functional independence* of the (smooth) functions I_1, \dots, I_{m-s} on an open set O , we simply mean that the Jacobian matrix of I_1, \dots, I_{m-s} has maximal rank $m - s$ on an open and dense subset of O . The set $\{I_1, \dots, I_{m-s}\}$ is often called a *complete fundamental set of invariants on O* . Note that a complete fundamental set of invariants is not unique.

As we shall consider actions on multiple points, we are interested in the case where $M = \mathcal{V} \times \mathcal{V} \times \dots \times \mathcal{V}$ (n -times) $=: \mathcal{V}^{\times(n)}$ is the Cartesian product of n copies of a manifold \mathcal{V} .

Definition 2.4. We say that G acts *diagonally* on $\mathcal{V}^{\times(n)}$ if there exists an action \cdot of G on \mathcal{V} such that for any $g \in G$, for any $n \in \mathbb{N}$ and any $z_1, \dots, z_n \in \mathcal{V}$, the action $g * (z_1, \dots, z_n)$ can be written as

$$g * (z_1, \dots, z_n) = (g \cdot z_1, \dots, g \cdot z_n).$$

The group actions we will define for our object-camera systems are not diagonal actions. However, for each of these action, there is a normal subgroup H of G (the subgroup generating the translations along the rays of light) such that G/H acts diagonally. So for all practical purposes, we shall ultimately have to deal with diagonal group actions.

In our approach to structure from motion, invariants are used to obtain equations that must be satisfied by the object and the camera. The more invariants we have, the more equations need to be satisfied. We need enough equations to completely determine the object. Observe that the dimension of the orbit is bounded by the dimension of the group. So in the case of a diagonal action, taking more and more copies of \mathcal{V} (i.e. more and more points) allows for the existence of as many invariants as necessary. The question that remains is: how can we obtain an expression for these invariants? Thanks to a new formulation of Cartan's theory of moving frames [4, 5, 6], this problem can be solved in an algorithmic fashion. We now summarize some of the relevant aspects of the theory of moving frames, including how moving frames can be used as a tool to obtain a complete set of fundamental invariants.

Definition 2.5. A (*right*) *moving frame* is a map $\rho : M \rightarrow G$ which is (right) equivariant, i.e. $\rho(g * z) = \rho(z)g^{-1}$, for all $g \in G$ and $z \in M$.

Unfortunately moving frames do not exist for all group actions.

Theorem 2.6. A moving frame exists if and only if the action of the group action satisfies

$$\{g \in G \mid \exists z \in M, g * z = z\} = \{e\},$$

where e denotes the identity in G . This property is called *freeness of the group action*.

Demanding freeness of the group action is very strong. It appears that, in order to be able to deal with the generic cases, we need to relax this condition a little bit.

Definition 2.7. A *local moving frame* is a map $\rho : M \rightarrow G$ such that $\rho(g * z) = \rho(z)g^{-1}$, for all $g \in N_e$, a neighborhood of the identity $e \in G$, and all $z \in M$.

Theorem 2.8. *A local moving frame exists if and only if there exists a neighborhood N_e of the identity in $e \in G$ such that*

$$\{g \in N_e \mid \exists z \in M, g * z = z\} = \{e\},$$

or equivalently, if and only if for all $z \in M$, the dimension of the orbit through z is equal to r , the dimension of G . This property is called local freeness of the group action.

We are now interested in determining a condition on the action of G on M which guarantees that the diagonal action will be locally free on a sufficiently large number of copies of M .

Definition 2.9. We say that G acts on M *effectively* if

$$\{g \in G \mid g * p = p, \text{ for all } p \in M\} = \{e\}.$$

We say that G acts on M *locally effectively* if

$$\{g \in G \mid g * p = p, \text{ for all } p \in M\} \text{ is a discrete subgroup of } G.$$

Many groups do not act effectively. However, given G acting non effectively on M , we can consider $\tilde{G} = G/G_M$, where $G_M = \{g \in G \mid g * z = z, \forall z \in M\}$, which acts in essentially the same way as G except that it acts effectively. Unfortunately, effectiveness is not sufficient to guarantee that the diagonal action eventually becomes locally free.

Definition 2.10. We say that G acts *effectively on subsets of M* if, for any open subset $U \subset M$,

$$\{g \in G \mid g * p = p, \text{ for all } p \in U\} = \{e\}.$$

We say that G acts *locally effectively on subsets of M* if, for any open subset $U \subset M$,

$$\{g \in G \mid g * p = p, \text{ for all } p \in U\} \text{ is a discrete subgroup of } G.$$

Observe that effectiveness on subsets implies effectiveness. The converse, of course, holds for all analytic group actions. However, this is not true in general (see [2] for a counterexample).

Theorem 2.11. [2] *If a group G acts on a manifold \mathcal{V} locally effectively on subsets, then there exists $n \in \mathbb{N}^+$ such that the induced diagonal action of G on $\mathcal{V}^{\times(n)}$ is locally free on an open and dense subset of $\mathcal{V}^{\times(n)}$. This is equivalent to saying that the orbit dimension is equal to the dimension of G on this open and dense subset. We denote by n_0 the minimal integer for which this is true.*

This means that any group action that is effective on subsets (e.g. any analytic group action, once the subgroup acting trivially is modded out) will be locally free on a sufficiently large number of copies of the manifold and a local moving frame will exist on this product.

We now explain how to construct a (local) moving frame and to obtain a complete fundamental set of invariants. A more detailed exposition can be found in [14, Chapter 8]. Let $g = (g_1, \dots, g_r)$ be local coordinates for G in a neighborhood of the identity. Suppose that G acts regularly on M . For simplicity, let us assume in addition that the orbits of G have the same dimension r as G itself. In other words, we are assuming that the action is locally free. Shortly after, we will explain how to deal with the case of merely regular actions using a simple variation of the following algorithm.

- Step 1: Write down the group transformation equations $\bar{x} = g * x$ explicitly.

$$\begin{cases} \bar{x}_1 &= f_1(g_1, \dots, g_r, x_1, \dots, x_m), \\ &\vdots \\ \bar{x}_m &= f_m(g_1, \dots, g_r, x_1, \dots, x_m). \end{cases}$$

- Step 2: Choose constants $c_1, \dots, c_r \in \mathbb{R}$ and set r of the transformed coordinates equal to those constants. For simplicity, we relabel the coordinates and write

$$\begin{cases} f_1(g_1, \dots, g_r, x_1, \dots, x_m) &= c_1, \\ &\vdots \\ f_r(g_1, \dots, g_r, x_1, \dots, x_m) &= c_r. \end{cases} \quad (2)$$

These equations are called the *normalization equations*.

- Step 3: Solve the normalization equations for $g = (g_1, \dots, g_r)$. The solution $g = \rho(x)$ is a moving frame.
- Step 4: Compute the action of the moving frame on the remaining coordinates. The set of resulting functions

$$\begin{cases} \bar{x}_{r+1}|_{g=\rho(x)} &= I_1(x_1, \dots, x_m), \\ &\vdots \\ \bar{x}_m|_{g=\rho(x)} &= I_{m-s}(x_1, \dots, x_m). \end{cases}$$

is a complete fundamental set of local invariants.

The choice of constants in Step 2 is somewhat arbitrary: we are free to choose any numbers for which a solution to the normalization equations exists, provided that these constants define a cross-section (i.e. provided that the normalization equations define a submanifold which is transversal to the orbits). To simplify the solution process, it is usually a good idea to choose as many constants as possible to be zero.

If the action is not free but merely regular, we can still find a system of functionally independent local invariants. We proceed as follows. Let s be the dimension of the orbits of G ($s < r$). We solve the s equations $f_1(g, x) = c_1, \dots, f_s(g, x) = c_s$ for s of the group parameters and replace them in the remaining equations $\bar{x}_{s+1} = f_{s+1}(g, x), \dots, \bar{x}_m = f_m(g, x)$ to get the $m - s$ invariants. The other group parameters g_{s+1}, \dots, g_r will not appear in the final expressions. This procedure is called a *partial moving frame normalization method*.

Equipped with these tools, obtaining invariants becomes a simple systematic procedure. We can thus feel free to consider any Lie group action imaginable and try to obtain its invariants. As we have seen, in theory, the invariants can always be found provided that the group action is locally effective on subsets (which we can always arrange in the case of analytic group actions). Of course, in practice, computational difficulties can be encountered in explicitly determining the invariants. Fortunately, the invariants are easily computed for the problem of structure from motion.

3 The Case of a Perspective Camera.

Let us assume that we are given t sets of n ordered points $p_1^\tau, \dots, p_n^\tau \in \mathbb{R}^2$, $\tau = 1, \dots, t$, which represent t pictures of a 3D unknown object made of n ordered points $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_n \in \mathbb{R}^3$. We would like to determine the points $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_n$ from these pictures. One possible way to try to solve this problem would be to define an equivalence relation between all the possible pictures of an object and to find functions depending on the picture points which are constant on each equivalence class. Characteristics of the object could be inferred from these functions, regardless of the camera position relative to the object. To define such an equivalence class, one could try to use the orbits of a group action; in other words, one could look for *invariants* of a group action that is transitive on the set of pictures of any given object, i.e. *view invariants*.

Unfortunately, as is commonly known in the vision community, *view invariants do not exist for 3D point sets of arbitrary size (in general position)*. One can still build invariants for specific objects (for instance, planar sets of points, pencils of lines, etc.) but not for arbitrary shapes. The problem is that the set of pictures of any object intersects with the set of pictures of other objects. Observe that if a view invariant I takes a constant value c for all pictures of an object O , then I is also equal to c on the set of pictures of any object whose set of pictures intersect with the set of pictures of O . One can actually show [3] that any equivalence relation between all the pictures of each objects defines a unique equivalence class on the space of pictures and thus, any view invariant is trivial. From a group point of view, this means that any group action that is transitive on the set of pictures of any object must be transitive on the set of *all* pictures.

Another way to try to solve this problem would be to use the reverse approach: try to characterize all the possible objects corresponding to a given picture. It would be useful to find functions which are constant on equivalence classes that include all the objects corresponding to a given picture. However, it is easy to see that such equivalence classes of objects are in one to one correspondence with the equivalence classes of pictures discussed above, and thus that only one such equivalence class can exist. There are therefore no *object invariants*. Nevertheless, given a view of an object, one *can* infer information about the object, so there must be a way to overcome this difficulty.

The trick is to define an equivalence relation on a higher dimensional space, sort of lifting the set of objects of different pictures to different “*heights*” in the extra dimensions. For this, we can use the three extra dimensions provided by the camera center position. More precisely, we can construct a Lie group action on the object points and

the camera center which summarizes what is unknown about the object-camera system given a picture of an object and knowing the mechanism used by the camera. Invariants of this group action prove to be sufficient for solving the problem of recovering the object coordinates in \mathbb{R}^3 .

So let us think for a moment about the process of taking a picture. This process involves, first of all, the placement of a camera in space. Then, particles of light start from each point of the object and travel on a straight line in the direction of the camera center, leaving their trace on a film, i.e. on the intersection of the picture plane and the travel lines. So to the picture-camera system placed somewhere in \mathbb{R}^3 , there corresponds a set of n straight lines in \mathbb{R}^3 representing the paths of light going from the object to the camera center (sometimes referred to as "ray bundle").

This process can be seen as a result of the action of group composed of an action of the special Euclidean group $SE(3)$ (i.e. the group of rotations and translations in \mathbb{R}^3) and an action of \mathbb{R}^n on the camera center and the image points (in 3D). Here is how.

Given is a 2D image depicting n points $p_1, \dots, p_n \in \mathbb{R}^2$. We assume this picture was taken by a camera with fixed internal parameters. These parameters can be calibrated beforehand, so that the focal length is $\mathcal{F} = 1^1$ and the 2D image coordinates match the 3D coordinates as defined below. We embed the picture-camera system in \mathbb{R}^3 by setting the camera center to be $\tilde{p}_0 = (0, 0, 0)$ and the picture points \tilde{p}_i 's to be $\tilde{p}_i = p_i \times \mathcal{F}$. This is, in general, not the actual position in which the picture was taken. However, there exists a rigid transformation $g \in SE(3)$ acting diagonally on $(\mathbb{R}^3)^{\times n}$ such that $g * (\tilde{p}_0, \tilde{p}_1, \dots, \tilde{p}_n) = (\mathfrak{P}_0, \mathfrak{P}_1, \dots, \mathfrak{P}_n)$ corresponds to the actual position of the picture-camera system at the moment where the picture was taken. Once the picture points are in this position, we know that we can move each of them independently along each ray of light so to go back to its source on the object. This way, the picture can be mapped onto the object.

In summary, the mapping is given by the transformation

$$\bar{P}_0 = RP_0 + T \quad (3)$$

$$\bar{P}_i = R(P_i + \lambda_i(P_i - P_0)) + T, \text{ for } i = 1, \dots, n, \quad (4)$$

with $R \in SO(3)$ a rotation, $T \in \mathbb{R}^3$ a translation and $\lambda_i \in \mathbb{R}$, a factor of depth, applied to $P_0 = \tilde{p}_0$ and $P_i = \tilde{p}_i$ for $i = 1, \dots, n$. As one can check, this mapping is actually a group action: we have an action of \mathbb{R}^n (parameterized by the λ 's) commuting with an action of $SE(3)$ (parameterized by rotations R and translations T). Therefore, this defines an action of the $(6 + n)$ -dimensional Lie group $SE(3) \times \mathbb{R}^n$, on the $(3n + 3)$ -dimensional manifold $(\mathbb{R}^3)^{\times (n+1)}$.

We would like to determine where \mathfrak{P}_0 and the \mathcal{O}_i 's lie. Given a picture, it is of course impossible to determine the camera center and object points $(\mathfrak{P}_0, \mathcal{O}_1, \dots, \mathcal{O}_n)$. But we know to which orbit under the action of $SE(3) \times \mathbb{R}^n$ they belong, since they belong to the same orbit as the (embedding of the) picture-camera system!

Assuming that the picture points are distinct, then the group action is regular and the orbits are 6-dimensional, for $n = 1$, and $(6 + n)$ -dimensional as soon as $n \geq 2$. Therefore by Theorem 2.3, there are $2n - 3$ fundamental invariants whenever $n \geq 2$

¹The value is arbitrary. It simply fixes the overall scale of the 3D reconstruction.

and these invariants can be used to characterize the orbits. We follow the steps of the moving frame normalization method to obtain them. We set

$$\begin{aligned}\bar{P}_0 &= (0, 0, 0)^T, \\ (0, 1, 0)\bar{P}_1 &= 0, \\ (0, 0, 1)\bar{P}_1 &= 0, \\ (0, 0, 1)\bar{P}_2 &= 0, \\ \text{and } (1, 0, 0) \cdot \bar{P}_i &= 1, \text{ for all } i = 1, \dots, n.\end{aligned}$$

Solving for the group parameters, we obtain

$$\begin{aligned}T &= -RP_0, \\ R &= R_1 R_2 R_3, \\ R_1 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{f}{\sqrt{f^2+g^2}} & \frac{g}{\sqrt{f^2+g^2}} \\ 0 & -\frac{g}{\sqrt{f^2+g^2}} & \frac{f}{\sqrt{f^2+g^2}} \end{pmatrix}, \\ R_2 &= \begin{pmatrix} \frac{\sqrt{x_1^2+y_1^2}}{\sqrt{x_1^2+y_1^2+z_1^2}} & 0 & \frac{z_1}{\sqrt{x_1^2+y_1^2+z_1^2}} \\ 0 & 1 & 0 \\ \frac{-z_1}{\sqrt{x_1^2+y_1^2+z_1^2}} & 0 & \frac{\sqrt{x_1^2+y_1^2}}{\sqrt{x_1^2+y_1^2+z_1^2}} \end{pmatrix}, \\ R_3 &= \begin{pmatrix} \frac{x_1}{\sqrt{x_1^2+y_1^2}} & \frac{y_1}{\sqrt{x_1^2+y_1^2}} & 0 \\ -\frac{y_1}{\sqrt{x_1^2+y_1^2}} & \frac{x_1}{\sqrt{x_1^2+y_1^2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \\ \lambda_i &= \frac{1}{(R(P_i - P_0))_x} - 1,\end{aligned} \tag{5}$$

where $f = \frac{-y_1 x_2 + x_1 y_2}{\sqrt{x_1^2+y_1^2}}$, $g = \frac{z_2(x_1^2+y_1^2) - z_1(x_1 x_2 + y_1 y_2)}{\sqrt{x_1^2+y_1^2}\sqrt{x_1^2+y_1^2+z_1^2}}$, $(x_1, y_1, z_1)^T = P_1 - P_0$ and $(x_2, y_2, z_2)^T = P_2 - P_0$. These group parameters define a moving frame (MF). Replacing the moving frame into the transformation equations, we get:

$$\begin{aligned}\bar{P}_0|_{MF} &= (0, 0, 0)^T, \\ \bar{P}_1|_{MF} &= (1, 0, 0)^T, \\ \bar{P}_2|_{MF} &= \begin{pmatrix} 1 \\ \frac{f\sqrt{x_1^2+y_1^2+z_1^2}(x_1 y_2 - x_2 y_1) + g[z_2(x_1^2+y_1^2) - z_1(x_1 x_2 + y_1 y_2)]}{(x_1 x_2 + y_1 y_2 + z_1 z_2)\sqrt{x_1^2+y_1^2}\sqrt{f^2+g^2}} \\ 0 \end{pmatrix}, \\ \bar{P}_i|_{MF} &= \begin{pmatrix} 1 \\ \frac{f\sqrt{x_1^2+y_1^2+z_1^2}(x_1 y_i - x_i y_1) + g[z_i(x_1^2+y_1^2) - z_1(x_1 x_i + y_1 y_i)]}{(x_1 x_i + y_1 y_i + z_1 z_i)\sqrt{x_1^2+y_1^2}\sqrt{f^2+g^2}} \\ \frac{g\sqrt{x_1^2+y_1^2+z_1^2}(x_i y_1 - x_1 y_i) + f[z_i(x_1^2+y_1^2) - z_1(x_1 x_i + y_1 y_i)]}{(x_1 x_i + y_1 y_i + z_1 z_i)\sqrt{x_1^2+y_1^2}\sqrt{f^2+g^2}} \end{pmatrix}.\end{aligned}$$

for all $i = 3, \dots, n$, where $(x_i, y_i, z_i) = P_i - P_0$. Each component of these vectors is an invariant of the group action.

These expressions have an easy geometric interpretation. Observe that $\sqrt{f^2 + g^2} = \frac{\|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|}{\sqrt{x_1^2 + y_1^2 + z_1^2}}$. So we can rewrite the above system as:

$$\begin{aligned}\bar{P}_0|_{MF} &= (0, 0, 0)^T, \\ \bar{P}_1|_{MF} &= (1, 0, 0)^T, \\ \bar{P}_2|_{MF} &= \begin{pmatrix} 1 \\ \frac{\|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|}{(\mathbf{x}_1, \mathbf{y}_1, z_1) \cdot (\mathbf{x}_2, \mathbf{y}_2, z_2)} \\ 0 \end{pmatrix}, \\ \bar{P}_i|_{MF} &= \begin{pmatrix} 1 \\ \frac{[(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_i, \mathbf{y}_i, z_i)] \cdot [(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)]}{[(\mathbf{x}_1, \mathbf{y}_1, z_1) \cdot (\mathbf{x}_i, \mathbf{y}_i, z_i)] \|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|} \\ \frac{(\mathbf{x}_i, \mathbf{y}_i, z_i) \cdot [(\mathbf{x}_2, \mathbf{y}_2, z_2) \times (\mathbf{x}_1, \mathbf{y}_1, z_1)] \|(\mathbf{x}_1, \mathbf{y}_1, z_1)\|}{[(\mathbf{x}_1, \mathbf{y}_1, z_1) \cdot (\mathbf{x}_i, \mathbf{y}_i, z_i)] \|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|} \end{pmatrix},\end{aligned}$$

where \cdot represents the scalar product between two vectors.

We now see that the components of $\bar{P}_2|_{MF}$ and $\bar{P}_i|_{MF}$ are sine or cosine of angles between the directions spanned by $\bar{P}_1 P_0$, $\bar{P}_2 P_0$, $\bar{P}_i P_0$ and the directions orthogonal to them. These are clearly invariant by translation, rotation, and motion along the projection lines. As a fundamental set, we simply pick the only $2n - 3$ non-constant invariants:

$$\begin{aligned}I_2 &= \frac{\|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|}{(\mathbf{x}_1, \mathbf{y}_1, z_1) \cdot (\mathbf{x}_2, \mathbf{y}_2, z_2)} \\ I_i &= \frac{[(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_i, \mathbf{y}_i, z_i)] \cdot [(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)]}{[(\mathbf{x}_1, \mathbf{y}_1, z_1) \cdot (\mathbf{x}_i, \mathbf{y}_i, z_i)] \|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|} \\ J_i &= \frac{(\mathbf{x}_i, \mathbf{y}_i, z_i) \cdot [(\mathbf{x}_2, \mathbf{y}_2, z_2) \times (\mathbf{x}_1, \mathbf{y}_1, z_1)] \|(\mathbf{x}_1, \mathbf{y}_1, z_1)\|}{[(\mathbf{x}_1, \mathbf{y}_1, z_1) \cdot (\mathbf{x}_i, \mathbf{y}_i, z_i)] \|(\mathbf{x}_1, \mathbf{y}_1, z_1) \times (\mathbf{x}_2, \mathbf{y}_2, z_2)\|}\end{aligned}$$

for $i = 3, \dots, n$.

Each picture taken defines a point in $\mathbb{R}^3 \times (\mathbb{R}^3)^{\times(n)}$ and therefore determines an orbit of our group action. Each orbit is characterized by the set of $2n - 3$ equations given by the invariants. More precisely, indexing the pictures with the discrete parameter $\tau = 1, \dots, t$, we have

$$\begin{aligned}I_i(P_0^\tau, P_1, \dots, P_n) &= \alpha_i^\tau, \text{ for } i = 2, \dots, n, \\ J_j(P_0^\tau, P_1, \dots, P_n) &= \beta_j^\tau, \text{ for } j = 3, \dots, n.\end{aligned}$$

for appropriate constants α_i^τ 's and β_j^τ 's. These constants are prescribed by the pictures: since the picture-camera system itself belongs to the orbits, we have

$$\begin{aligned}\alpha_i^\tau &= I_i(\tilde{p}_0^\tau, \tilde{p}_1^\tau, \dots, \tilde{p}_n^\tau) \\ \beta_j^\tau &= J_j(\tilde{p}_0^\tau, \tilde{p}_1^\tau, \dots, \tilde{p}_n^\tau)\end{aligned}$$

We are interested in solving the equations

$$\begin{aligned} I_i(\mathfrak{P}_0^\tau, \mathcal{O}_1, \dots, \mathcal{O}_n) &= \alpha_i^\tau, \text{ for } i = 2, \dots, n \\ J_j^\tau(\mathfrak{P}_0^\tau, \mathcal{O}_1, \dots, \mathcal{O}_n) &= \beta_j^\tau, \text{ for } j = 3, \dots, n. \end{aligned}$$

for $\tau = 1, \dots, t$. We have $(2n - 3)t$ (non-linear) equations with $3n + 3t$ unknowns, the solution of which is determined up to a rotation and translation of the 3D camera-object system as a whole, which can fix arbitrarily, thus eliminating six variables². For $n > 3$ and $t \geq \frac{3n-6}{2n-6}$, the number of equations is greater than the number of unknowns so we can try to solve them.

Experiments with real video images have been performed (see Fig.1) using a sequential non-linear optimization technique based on the Levenberg-Marquardt algorithm [15]. The points used as picture points $(\tilde{p}_0^\tau, \tilde{p}_1^\tau, \dots, \tilde{p}_n^\tau)$ are the endpoints of lines and rectangles drawn on Fig.1-a. These points have been obtained in successive images with a simple tracking procedure [1]. We computed the values of all the invariants $(\alpha_i^\tau, \beta_j^\tau)$ using these points and solved the $(2n - 3)t$ non-linear equations for the unknowns $(\mathfrak{P}_0^\tau, \mathcal{O}_1, \dots, \mathcal{O}_n)$. The solution gave us the reconstructed 3D object, namely the set of lines and rectangles defined by $(\mathcal{O}_1, \dots, \mathcal{O}_n)$. Although the bottom and left side elements are not perfectly replaced due to noise in the input picture points, the reconstructed object is visually correct in any view. In particular, there is no global distortion as one would fear in the case of projective reconstructions. The computations take only a few minutes.

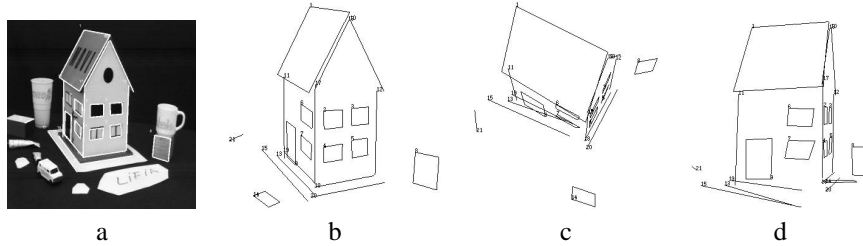


Figure 1: A reconstruction example: a) the first of six images, with line and rectangle features drawn, b) a similar view of the 3D reconstructed features, c) a top view, d) a side view of the reconstruction.

Observe that our camera-system does not take into account the angle of the camera; the orientation of the image plane was only included in the group parameters (not on the space acted on) and thus factored out of the problem in the invariant formulation. Besides the advantage of not having to solve for this unwanted unknown, we also obtain the following lemma.

²However, we should keep in mind that the choice of these variables will affect the numerical resolution[10].

Lemma 3.1. *The motion of the camera between two pictures is a pure rotation (i.e. a rotation around the center of projection P_0) if and only if the values of the invariants $\{I_i, J_j | i = 2, \dots, n, j = 3, \dots, n\}$ evaluated on any corresponding points in the two views are equal.*

Proof. Invariance of our invariants under pure rotations is obvious from the construction of the invariants.

To prove that equality of our invariants evaluated on all corresponding points guarantees that the camera motion is a pure rotation, observe that the first invariant I_2 is the tangent of the angle between the lines $\overline{\mathfrak{P}_0\mathcal{O}_1}$ and $\overline{\mathfrak{P}_0\mathcal{O}_2}$. Its value remains constant for fixed $\mathcal{O}_1, \mathcal{O}_2$ only if \mathfrak{P}_0 moves along a circle around the $\overline{\mathcal{O}_1\mathcal{O}_2}$ axis. This holds for all possible choices of \mathcal{O}_1 and \mathcal{O}_2 so the camera center must lie somewhere on the intersection of a set of circles, which can be taken to intersect at merely one point to guarantee that the camera center does not move. \square

Tomasi and Kanade [19] identify two problems related to using the traditional *direct* approaches to structure from motion in a noisy context. First there is the fact that, when the camera motion is small, the effects of a camera rotation and translation are hard to distinguish. Secondly, obtaining the shape by comparing depths is sensitive to noise, since the depth can be considerably larger than the dimensions of the shape. Note that both these difficulties are bypassed by our approach, since the depths and the rotation of the camera do not appear in our equations. We also provide a new formulation in Euclidean space that involves quantities independent of any choice of world coordinate system, removing the so-called *gauge problem*.

Unfortunately, the non-linearity of the invariants is a serious drawback. As one can tell from the orbit structure of this group action, this is inherent to the problem as formulated. One might want to ask whether there exist coordinates which lead to simpler expressions for the invariants. For example, we could use an inductive moving frame construction [11] in order to obtain invariants of one subgroup of $SE(3) \times \mathbb{R}^n$, say either $SE(3)$ or \mathbb{R}^n , and use these invariants as new coordinates on which the remaining group coordinates are acting. The resulting invariants are actually much simpler in these coordinates. However it turns out that *each* of these new coordinates would involve the camera center. One would thus need to define a new set of coordinates for *each* picture so there would always be more unknowns coordinates than equations. These invariants are thus not useful for recovering the structure from a set of pictures, although they could be good tools for analyzing the motion of an object taken by a fixed or even rotating camera.

Another way of obtaining simpler equations is to work in projective coordinates. However as a trade off, projective coordinates require using more variables. We will discuss this other method in the next section. But first, we generalize the Euclidean coordinate approach to the case of a variable focal length.

3.1 Letting the focal distance vary

It is a bit more complicated to set up the group transformation equations in the case where the focal length is allowed to vary from one image to the next. One way to do

this is the following. Consider P_M , the closest point to P_0 on the image plane, i.e. the embedding of the middle point of the picture. Changing the focal distance corresponds to transforming P_M into a point P'_M with a real parameter α according to the rule

$$P'_M = P_M + \alpha(P_M - P_0).$$

The induced action on the P_i 's can be taken as a rotation about the center of camera P_0 which preserves the distance to the camera center. More precisely, each P_i is moved to a new point P'_i in such a way that $\|P'_i - P_0\| = \|P_i - P_0\|$ and that its transformed picture point is $\tilde{p}'_i = \tilde{p}_i + \alpha(P_M - P_0)$. Since the picture point is given by $p_i = P_0 + \frac{P_i - P_0}{(P_i - P_0) \cdot (P_M - P_0)}$, we find that

$$P'_i = P_0 + \|P_i - P_0\| \frac{\frac{P_i - P_0}{(P_i - P_0) \cdot (P_M - P_0)} + \alpha(P_M - P_0)}{\left\| \frac{P_i - P_0}{(P_i - P_0) \cdot (P_M - P_0)} + \alpha(P_M - P_0) \right\|}.$$

Combining with translations of the P_i 's along the line $P_i P_0$ together with rotations and translations of the line arrangement as a whole, we get the following $(n + 7)$ -dimensional Lie group action:

$$\begin{aligned} \bar{P}_0 &= RP_0 + T, \\ \bar{P}_M &= R(P_M + \alpha(P_0 - P_M)) + T, \\ \bar{P}_i &= R \left(P_0 + (1 + \lambda_i) \|P_i - P_0\| \frac{\frac{P_i - P_0}{(P_i - P_0) \cdot (P_M - P_0)} + \alpha(P_M - P_0)}{\left\| \frac{P_i - P_0}{(P_i - P_0) \cdot (P_M - P_0)} + \alpha(P_M - P_0) \right\|} \right) + T. \end{aligned}$$

Observe that the change of focal length parameterized by α commutes with the action of each λ_i on P_i , because it preserves the norm $\|P_i - P_0\|$.

We apply the previous moving frame normalization technique by setting:

$$\begin{aligned} \bar{P}_0 &= (0, 0, 0)^T, \\ \bar{P}_M &= (1, 0, 0)^T, \\ \bar{P}_1 \cdot (0, 0, 1) &= 0, \\ \text{and } \bar{P}_i \cdot (1, 0, 0) &= 1, \text{ for all } i = 1, \dots, n. \end{aligned}$$

The corresponding group parameters are similar to Eq. (5) for R and T , except that $(x_1, y_1, z_1)^T$ and $(x_2, y_2, z_2)^T$ must be replaced by $(u_1, v_1, w_1)^T = P_M - P_0$ and $(u_2, v_2, w_2)^T = P_1 - P_0$. The other parameters are:

$$\begin{aligned} \alpha &= \frac{1}{\|P_M - P_0\|} - 1, \\ \lambda_i &= \frac{\left\| \frac{P_i - P_0}{(P_i - P_0) \cdot (P_M - P_0)} + \alpha(P_M - P_0) \right\|}{\|P_i - P_0\| \|P_M - P_0\| + \frac{\|P_i - P_0\|}{\|P_M - P_0\|}} - 1, \text{ for all } i = 1, \dots, n. \end{aligned}$$

Replacing these group parameters into the equations for the \bar{P}_i 's, we obtain the follow-

ing complete fundamental set of invariants:

$$\begin{aligned}
I_1 &= \frac{\|(P_M - P_0) \times (P_1 - P_0)\|}{(P_1 - P_0) \cdot (P_M - P_0) (1 + \|P_M - P_0\| - \|P_M - P_0\|^2)}, \\
I_i &= \frac{(P_M - P_0) \times (P_i - P_0) \cdot (P_M - P_0) \times (P_1 - P_0)}{\|(P_M - P_0) \times (P_1 - P_0)\| (P_i - P_0) \cdot (P_M - P_0) (1 + \|P_M - P_0\| - \|P_M - P_0\|^2)}, \\
J_i &= \frac{(P_i - P_0) \cdot [(P_1 - P_0) \times (P_M - P_0)] \|P_M - P_0\|}{\|(P_M - P_0) \times (P_1 - P_0)\| (P_i - P_0) \cdot (P_M - P_0) (1 + \|P_M - P_0\| - \|P_M - P_0\|^2)},
\end{aligned}$$

for $i = 2, \dots, n$.

Observe that solving for P_M and P_0 implies solving for the focal length. Therefore the focal length has not been removed in this formulation. Having to solve for the focal length is undesirable since it can induce numerical instabilities. We can actually completely include the focal length in the group parameters by letting $v = \frac{P_M - P_0}{\|P_M - P_0\|}$ and $\gamma = \alpha \|P_M - P_0\|^2$. Then γ can be seen as a new group parameter $\gamma \in \mathbb{R}_{\neq -1}$ in the group action given by

$$\begin{aligned}
\bar{P}_0 &= RP_0 + T, \\
\bar{v} &= Rv, \\
\bar{P}_i &= R \left(P_0 + (1 + \lambda_i) \|P_i - P_0\| \frac{P_i - P_0 + (\gamma(P_i - P_0) \cdot v)v}{\|P_i - P_0 + (\gamma(P_i - P_0) \cdot v)v\|} \right) + T.
\end{aligned}$$

One can check that the group action parameterized by γ is compatible with the group structure of $\mathbb{R}_{\neq -1}$ with group multiplication \circ given by

$$\gamma_1 \circ \gamma_2 = \gamma_1 + \gamma_2 + \gamma_1 \gamma_2, \text{ for all } \gamma_1, \gamma_2 \in \mathbb{R}_{\neq -1}.$$

We have an action of the $(7 + n)$ -dimensional Lie group $SE(3) \times \mathbb{R}_{\neq -1} \times \mathbb{R}^n$ on a $(3n + 5)$ -dimensional space; there are $2n - 2$ fundamental invariants. To obtain these invariants, we start by setting $\bar{P}_0 = (0, 0, 0)^T$ and solve for T . We then set $\bar{v} = (1, 0, 0)^T$ and the third component of \bar{P}_1 to zero and solve for the rotation matrix R . We skip the details of these computations since they are very similar to the previous cases. Replacing these group parameters into the other transformation equations, we obtain

$$\begin{aligned}
\bar{P}_i &= (1 + \lambda_i) \|P_i - P_0\| \begin{pmatrix} \frac{F_i \cdot v}{\frac{(v \times F_i) \cdot (v \times F_1)}{\|v \times F_1\|}} \\ \frac{F_i \cdot (v \times F_1)}{\|v \times F_1\|} \end{pmatrix}, \text{ for } i = 2, \dots, n, \\
\bar{P}_1 &= (1 + \lambda_1) \|P_1 - P_0\| \begin{pmatrix} F_1 \cdot v \\ \|v \times F_1\| \\ 0 \end{pmatrix},
\end{aligned}$$

where F_i represents the fraction

$$F_i = \frac{P_i - P_0 + (\gamma(P_i - P_0) \cdot v)v}{\|P_i - P_0 + (\gamma(P_i - P_0) \cdot v)v\|}, \text{ for } i = 1, \dots, n.$$

We then set the first component of each \bar{P}_i to one, for $i = 1, \dots, n$, and solve for the λ_i 's. (For this, we need to assume that $(P_i - P_0) \cdot v \neq 0$.) Replacing these λ_i 's into the transformation equations, we get

$$\begin{aligned}\bar{P}_i &= \begin{pmatrix} 1 \\ \frac{(v \times F_i) \cdot (v \times F_1)}{\|v \times F_1\| \|F_i \cdot v\|} \\ \frac{F_i \cdot (v \times F_1)}{\|v \times F_1\| \|F_i \cdot v\|} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{(v \times (P_i - P_0)) \cdot (v \times (P_1 - P_0))}{\|v \times (P_1 - P_0)\| \|(1+\gamma)(P_i - P_0) \cdot v\|} \\ \frac{(P_i - P_0) \cdot (v \times (P_1 - P_0))}{\|(v \times (P_1 - P_0))\| \|(1+\gamma)(P_i - P_0) \cdot v\|} \end{pmatrix}, \text{ for } i = 2, \dots, n, \\ \bar{P}_1 &= \begin{pmatrix} 1 \\ \frac{\|v \times F_1\|}{F_1 \cdot v} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{\|v \times (P_1 - P_0)\|}{(1+\gamma)(P_1 - P_0) \cdot v} \\ 0 \end{pmatrix}.\end{aligned}$$

Provided that $P_1 - P_0$ is not parallel to v , we can finish obtaining the moving frame by setting the second component of \bar{P}_1 to one and solving for γ . Replacing this γ into the two non-constant components of each P_i , we obtain the following two invariants:

$$\begin{aligned}I_i &= \frac{v \cdot (P_1 - P_0)(v \times (P_i - P_0)) \cdot (v \times (P_1 - P_0))}{v \cdot (P_i - P_0)\|v \times (P_1 - P_0)\|^2}, \\ J_i &= \frac{v \cdot (P_1 - P_0)(P_i - P_0) \cdot (v \times (P_1 - P_0))}{v \cdot (P_i - P_0)\|v \times (P_1 - P_0)\|^2},\end{aligned}$$

for $i = 2, \dots, n$. In a similar way as for the case of a fixed focal length, when the number of pictures t and the number of object points n is large enough, these invariants provide a number of equations which are, in most cases, sufficient to solve for P_1, \dots, P_n and P_0^τ, v^τ , for $\tau = 1, \dots, t$.

4 Using Projective Coordinates

The projective space \mathbb{P}^3 is $\{(x, y, z, w) \in \mathbb{R}^4 \setminus \{(0, 0, 0, 0)\}\}$ modulo multiplication by a scalar multiple in $\mathbb{R} \setminus \{0\}$. If $(x, y, z, w) \in \{\mathbb{R}^4 \setminus \{(0, 0, 0, 0)\}\}$, then the coset of (x, y, z, w) in \mathbb{P}^3 is denoted by $(x : y : z : w)$. Consider the chart U of \mathbb{P}^3 defined by $\{(x : y : z : w) \in \mathbb{P}^3 | w \neq 0\}$. The map $\phi : U \rightarrow \mathbb{R}^3$ defined by

$$\phi(x : y : z : w) = \left(\frac{x}{w}, \frac{y}{w}, \frac{z}{w}\right)$$

provides coordinates for the chart U . In other words, Euclidean coordinates of \mathbb{R}^3 are local coordinates for a piece of \mathbb{P}^3 . One way to obtain simpler invariants is to work directly in $\mathbb{R}^4 \setminus \{(0, 0, 0, 0)\}$ (i.e. in projective coordinates) by choosing representatives (x_0, y_0, z_0, w_0) and (x_i, y_i, z_i, w_i) in $\{(x, y, z, w) \in \mathbb{R}^4 | w \neq 0\}$ for the camera center and the object points respectively.

Consider the following action of $SE(3) \times \mathbb{R}^n$ on $n + 1$ copies of \mathbb{R}^4 :

$$\begin{aligned} \begin{pmatrix} \bar{x}_0 \\ \bar{y}_0 \\ \bar{z}_0 \\ \bar{w}_0 \end{pmatrix} &= \begin{pmatrix} & R & T \\ 0 & 0 & 0 \\ & & 1 \end{pmatrix} \begin{pmatrix} \bar{x}_0 \\ \bar{y}_0 \\ \bar{z}_0 \\ \bar{w}_0 \end{pmatrix}, \\ \begin{pmatrix} \bar{x}_i \\ \bar{y}_i \\ \bar{z}_i \\ \bar{w}_i \end{pmatrix} &= \begin{pmatrix} & R & T \\ 0 & 0 & 0 \\ & & 1 \end{pmatrix} \left[\begin{pmatrix} x_i \\ y_i \\ z_i \\ w_i \end{pmatrix} + \lambda_i \left(\begin{pmatrix} x_i \\ y_i \\ z_i \\ w_i \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ w_0 \end{pmatrix} \right) \right], \\ &\text{for } i = 1, \dots, n, \end{aligned}$$

where $R \in SO(3)$ is a 3×3 rotation matrix, $T \in \mathbb{R}^3$ represents a translation and the $\lambda_i \in \mathbb{R}$ are the depth parameters. Since lines through the origin are mapped to lines through the origin, this induces an action on \mathbb{P}^3 . In local coordinates ϕ for the chart U , this corresponds exactly to the action defined by Equations 3 and 4. Assuming that the object points are pairwise distinct (and, of course, also distinct from the camera center), then as soon as the number of object points $n \geq 2$, the orbits are $(6 + n)$ -dimensional both on $(\mathbb{R}^4)^{\times(n+1)}$ and on $(\mathbb{P}^3)^{\times(n+1)}$. There are therefore $3n - 2$ fundamental invariants in projective coordinates.

To obtain a fundamental set of invariants, we start by setting $\omega_i = 1$, for all i 's, and $\bar{P}_0 = (0, 0, 0)^T$. Provided that all $w_i \neq w_0$, the corresponding group parameters are $\lambda_i = \frac{1-w_i}{w_i-w_0}$ and $T = -\frac{1}{w_0}R$. For simplicity, we let $\beta_i = \frac{1-w_i}{w_i-w_0}$. We then set the second and third components of \bar{P}_1 to zero and the third component of \bar{P}_2 to zero and solve for the rotation matrix R to finish obtaining the moving frame. Replacing the moving frame into the transformation equations, we obtain:

$$\begin{aligned} (\bar{x}_0, \bar{y}_0, \bar{z}_0, \bar{w}_0)|_{MF} &= (0, 0, 0, w_0), \\ (\bar{x}_1, \bar{y}_1, \bar{z}_1, \bar{w}_1)|_{MF} &= (\|Q_1\|, 0, 0, 1), \\ (\bar{x}_2, \bar{y}_2, \bar{z}_2, \bar{w}_2)|_{MF} &= \left(\frac{Q_2 \cdot Q_1}{\|Q_1\|}, \frac{\|Q_2 \times Q_1\|}{\|Q_1\|}, 0, 1 \right), \\ (\bar{x}_i, \bar{y}_i, \bar{z}_i, \bar{w}_i)|_{MF} &= \begin{pmatrix} \frac{Q_i \cdot Q_1}{\|Q_1\|} \\ \frac{(Q_1 \times Q_i) \cdot (Q_1 \times Q_2)}{\|Q_1\| \|Q_1 \times Q_2\|} \\ \frac{Q_i \cdot (Q_1 \times Q_2)}{\|Q_1 \times Q_2\|} \\ 1 \end{pmatrix}^T, \\ &\text{for } i = 3, \dots, n, \end{aligned}$$

where \cdot represents the scalar product between two vectors and $Q_i = (1 + \beta_i)P_i - (\beta_i + \frac{1}{w_0})P_0$. As a complete set of fundamental invariants, we can take, in addition to the coordinate w_0 , the functions

$$\begin{aligned} H_i &= Q_i \cdot Q_1, \text{ for } i = 1, \dots, n, \\ I_i &= (Q_1 \times Q_i) \cdot (Q_1 \times Q_2), \text{ for } i = 2, \dots, n, \\ J_i &= Q_i \cdot (Q_1 \times Q_2), \text{ for } i = 3, \dots, n. \end{aligned}$$

The invariants are thus functions of the object points P_i 's, the β_i 's, and the camera center projective coordinates P_0 and w_0 . For every picture, the parameter w_0 can be fixed arbitrarily to any value other than zero. By choosing w_0 to be the same for every picture then all β_i 's are the same for every picture. Given t pictures, the unknowns are thus the P_i 's and β_i 's, for $i = 1, \dots, n$ and P_0^τ , for $\tau = 1, \dots, t$, while the invariants provide $t(3n - 3)$ equations. With enough points and enough pictures, we obtain more equations than unknowns.

5 The Case of an Orthographic Camera

The orthographic camera is an approximation of the perspective camera. In this model, we assume that the camera center lies at infinity and so the rays of light are parallel to each other.

Let $v = (v_x, v_y, v_z)^T$ be the unit direction vector of the rays of light and let P_1, \dots, P_n represent the object points in \mathbb{R}^3 . Any picture of the object provides some information about the structure of the object. What remains unknown is the orientation and the position of the camera at the moment when the picture was taken, as well as the distance from the camera plane to each object point. The following action of $SE(3) \times \mathbb{R}^n$ on $\{(v, P_1, \dots, P_n) \in (\mathbb{R}^3) \times (n+1)\}$ such that $|v| = 1\}$ summarizes what is unknown about the object given a picture.

$$\begin{aligned}\bar{v} &= Rv \\ \bar{P}_i &= R(P_i + \lambda_i v) + T, \text{ for } i = 1, \dots, n,\end{aligned}$$

where $R \in SO(3)$ is a rotation matrix, $T \in \mathbb{R}^3$ represents a translation and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ are the depth parameters. More precisely, given a picture p_1, \dots, p_n the orbit passing through

$$\begin{aligned}P_i &= (p_i, 0), \text{ for } i = 1, \dots, n \\ v &= (0, 0, 1)\end{aligned}$$

(i.e. the embedding of the picture in \mathbb{R}^3) under this group action corresponds to all possible 3D objects that could have been used to take this picture.

To obtain the invariants of this group action, we set $\bar{P}_1 = (0, 0, 0)^T$, the first component of each \bar{P}_i to zero, the second and third components of \bar{P}_1 to zero, and the third component of \bar{P}_1 to zero. We use the partial moving frame normalization method and, in order to obtain a moving frame, solve for all parameters except λ_1 , which does not appear in the final expressions. Replacing this moving frame in the group transforma-

tion equations, we obtain

$$\begin{aligned}
\bar{v}|_{MF} &= (1, 0, 0)^T, \\
\bar{P}_1|_{MF} &= (0, 0, 0)^T, \\
\bar{P}_2|_{MF} &= \begin{pmatrix} 1 \\ \|(P_2 - P_1) - [(P_2 - P_1) \cdot v]v\| \\ 0 \end{pmatrix}, \\
\bar{P}_i|_{MF} &= \begin{pmatrix} 1 \\ (P_i - P_1) \cdot \frac{(P_2 - P_1) - [(P_2 - P_1) \cdot v]v}{\|(P_2 - P_1) - [(P_2 - P_1) \cdot v]v\|} \\ (P_i - P_1) \cdot \left(v \times \frac{(P_2 - P_1) - [(P_2 - P_1) \cdot v]v}{\|(P_2 - P_1) - [(P_2 - P_1) \cdot v]v\|} \right) \end{pmatrix}, \text{ for } i = 1, \dots, n.
\end{aligned}$$

A fundamental set of invariants is given by the non-constant components of these vectors. Observe that some of these expressions are fractions. However, their denominator is actually one of the invariants of the fundamental set. We can thus simply get rid of the denominator and take the following functions as our fundamental set of invariants:

$$\begin{aligned}
I_2 &= \|(P_2 - P_1) - [(P_2 - P_1) \cdot v]v\|, \\
I_i &= (P_i - P_1) \cdot [(P_2 - P_1) - [(P_2 - P_1) \cdot v]v], \text{ for } i = 3, \dots, n, \\
J_i &= (P_i - P_1) \cdot [v \times (P_2 - P_1)] \text{ for } i = 3, \dots, n.
\end{aligned}$$

Given pictures $p_1^\tau, \dots, p_n^\tau \in \mathbb{R}^2$, we let v^τ for $\tau = 1, \dots, n$ be the direction vectors of the rays of light of the camera. The object points P_1, \dots, P_n and direction vectors v^τ thus satisfy the matrix equation

$$\begin{pmatrix} (P_2 - P_1) - [(P_2 - P_1) \cdot v^\tau]v^\tau \\ v^\tau \times (P_2 - P_1) \end{pmatrix} (P_2 - P_1, P_3 - P_1, \dots, P_n - P_1) = \begin{pmatrix} \alpha_2^\tau & \alpha_3^\tau & \dots & \alpha_n^\tau \\ \beta_2^\tau & \beta_3^\tau & \dots & \beta_n^\tau \end{pmatrix},$$

where the α 's and β 's are constants prescribed by the pictures. For solving these equations, we can make a change of variable and let the two entries of the leftmost matrix be two unknown parameters m_1^τ and m_2^τ subject to the condition $|m_1^\tau| + |m_2^\tau| \leq |P_2 - P_1|$. We obtain a factorization equation like the one introduced by Tomasi and Kanade [19], with a different formulation. Note that our system involves only the normal to the camera plane i.e. two parameters, while theirs involve all three parameters specifying the orientation of the camera.

6 Conclusion

This paper presented applications of a systematic technique invented by Fels and Olver for building invariants of a Lie group action. We started by summarizing this technique. We then showed how to formulate the problem of structure from motion in three

different settings (Euclidean coordinates, projective coordinates and orthographic projections) in terms of Lie group actions. In each setting, the group parameters included unknown, unwanted parameters of the problem. These parameters were removed from the equations by reformulating the problem using invariants of these group actions.

The orbit structure and the invariants of the group action provide interesting insights on the geometry of the projections. They also provide a formalization for similar results obtained more empirically [17]. For solving the structure from motion problem, our results relate to several well-known techniques but eliminates additional unnecessary parameters, sometimes difficult to remove otherwise. Further work is now ongoing to make such invariant systems more computationally attractive for various practical applications.

Acknowledgments

Both authors would like to thank David Cooper for support, encouragement and stimulating discussions as well as David Mumford and Jean Ponce for their useful comments.

References

- [1] P. L. Bazin and J. M. Vézien. Tracking geometric primitives in video streams. In *Proceedings of the Fourth Irish Machine Vision and Image Processing Conference*, pages 43–50, Belfast, September 2000.
- [2] M. Boutin. On orbit dimensions under a simultaneous Lie group action on n copies of a manifold. *J. Lie Theory*, 12(1):191–203, 2002.
- [3] J. B. Burns, R. S. Weiss, and E. M. Riseman. The non-existence of general-case view-invariants. In *Geometric invariance in computer vision*, pages 120–131. MIT Press, 1992.
- [4] É. Cartan. *Leçons sur la géométrie projective complexe. La théorie des groupes finis et continus et la géométrie différentielle traitées par la méthode du repère mobile. Leçons sur la théorie des espaces à connexion projective.* Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics]. Éditions Jacques Gabay, Sceaux, 1992. Reprint of the editions of 1931, 1937 and 1937.
- [5] M. Fels and P. J. Olver. Moving coframes. I. A practical algorithm. *Acta Appl. Math.*, 51(2):161–213, 1998.
- [6] M. Fels and P. J. Olver. Moving coframes. II. Regularization and theoretical foundations. *Acta Appl. Math.*, 55(2):127–208, 1999.
- [7] G. Frobenius. Über das Pfaff'sche Problem. *J. Reine Angew. Math.*, 82:230–315, 1877.
- [8] V. V. Gorbatsevich, A. L. Onishchik, and E. B. Vinberg. *Foundations of Lie theory and Lie transformation groups.* Springer-Verlag, Berlin, 1997. Translated from the Russian by A. Kozłowski, Reprint of the 1993 translation [*Lie groups*

- and Lie algebras. I, Encyclopaedia Math. Sci., 20, Springer, Berlin, 1993; MR 95f:22001].
- [9] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, 2001. With a foreword by Olivier Faugeras.
 - [10] K. Kanatani. Gauge-based reliability analysis of 3-d reconstruction from two uncalibrated perspective views. In *Proceedings of the Fifteenth International Conference on Pattern Recognition*, volume 1, pages 76–79, Barcelona, September 2000.
 - [11] I. A. Kogan. Inductive construction of moving frames. In *The geometrical study of differential equations (Washington, DC, 2000)*, volume 285 of *Contemp. Math.*, pages 157–170. Amer. Math. Soc., Providence, RI, 2001.
 - [12] J. L. Mundy and A. Zisserman, editors. *Geometric invariance in computer vision*. Artificial Intelligence. MIT Press, Cambridge, MA, 1992.
 - [13] J. L. Mundy, A. Zisserman, and D. Forsyth, editors. *Workshop on Applications of Invariance in Computer Vision*, volume 825 of *LNCS*. Springer, Ponta Delgada, 1994.
 - [14] P. J. Olver. *Classical invariant theory*, volume 44 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999.
 - [15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge University Press, Cambridge, second edition, 1992. The art of scientific computing.
 - [16] L. Quan. Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1), 1995.
 - [17] C. Rother and S. Carlson. Linear multi view reconstruction and camera recovery. In *Proceedings of the International Conference on Computer Vision, ICCV'01*, Vancouver, July 2001.
 - [18] G. Sparr. Euclidean and affine structure/motion for uncalibrated cameras from affine shape and subsidiary information. In *Proceedings of the SMILE workshop*, pages 187–207, Freiburg, June 1998.
 - [19] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. Comp. Vision*, 9(2), 1992.
 - [20] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. In *Proceedings of Vision Algorithms: Theory and Practice*, Corfu, September 1999.
 - [21] I. Weiss. 3D curve reconstruction from uncalibrated cameras. *Technical report*, University of Maryland, 1996.