

CDS

TECHNICAL MEMORANDUM NO. CIT-CDS 94-001

January 17, 1994

revised March 8, 1994

“Observability/Identifiability of Rigid Motion under Perspective Projection”

Stefano Soatto and Pietro Perona

Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125

Observability/Identifiability of Rigid Motion under Perspective Projection

Stefano Soatto and Pietro Perona *

California Institute of Technology 116-81-Pasadena, CA 91125

soatto@systems.caltech.edu

Keywords: dynamic vision, motion estimation, nonlinear observability, implicit systems, exterior differential systems, differential algebraic equations

Abstract

The “visual motion” problem consists of estimating the motion of an object viewed under projection. In this paper we address the feasibility of such a problem.

We will show that the model which *defines* the visual motion problem for feature points in the euclidean 3D space lacks of both linear and local (weak) observability. The locally observable manifold is covered with three levels of lie differentiations. Indeed, by imposing metric constraints on the state-space, it is possible to reduce the set of indistinguishable states.

We will then analyze a model for visual motion estimation in terms of *identification of an Exterior Differential System*, with the parameters living on a topological manifold, called the “*essential manifold*”, which includes explicitly in its definition the forementioned metric constraints. We will show that rigid motion is globally observable/identifiable under perspective projection with zero level of lie differentiation under some general position conditions. Such conditions hold when the viewer does not move on a quadric surface containing all the visible points.

1 Introduction

Animals face everyday tasks which require the ability to estimate the relative motion between them and the objects populating the environment (or the environment itself), such as walking, avoiding obstacles, grasping objects etc. . Only recently, dynamic estimation and control techniques have given encouraging results for designing automatic systems which mimic such abilities [13, 12, 20, 21, 3, 26].

If we restrict our attention to motions inside a “static scene”, then the rigid motion constraint and the perspective projection map *define* a nonlinear dynamical model. Motion estimation may be formalized in terms of parameter identification and/or state estimation of such a model. Traditionally the estimation task has been performed using an Extended Kalman Filter (EKF) [31, 8, 29].

In this paper we address the feasibility of estimating the motion of a rigid object from perspective observations using a dynamic model.

*Research funded by the California Institute of Technology, ONR grant N00014-93-1-0990 and an AT&T Foundation Special Purpose grant. This work is registered as CDS Technical Report CIT-CDS 94-001, California Institute of Technology, January 1994 – revised February 1994. Submitted to the invited session on “Dynamic Vision, System Theoretical Methods and Control Applications” at the 33rd IEEE conf. on Decision and Control, Florida, 1994.

A crucial issue in dynamic estimation/identification is the observability of the model, or the identifiability of its parameters. We will see that the model which “defines” the visual motion problem for feature points in the euclidean 3D space is neither linearly observable nor locally weakly observable. It is possible, as we will see, to reduce the set of locally indistinguishable states by imposing metric constraints on the state space; however, the model suffers some structural limitations which make the local-linearization based methods purely conditioned and not robust enough to be used in real world applications.

Rigid motion is indeed globally observable from perspective projections, once the problem is formulated in the appropriate topological space. In this paper we analyze a new formulation for motion estimation [47] in terms of *identification of an Exterior Differential System* [7] *with the parameters living on a topological manifold*, called the “essential manifold”. Using some results from the computational vision literature, we show that this model is globally observable without any lie-differentiation under general position conditions. Such conditions are met when the object and the path of the center of projection cannot be imbedded in a quadric surface [15, 37], and can be verified using a simple rank test.

1.1 Existing literature and relations to previous work

The use of dynamic observers for estimating scene structure and/or motion dates back to the eighties [16, 6, 36, 23]. These works assumed that either the viewer’s motion or the shape of the object was known. More recent works deal with the estimation of both structure and motion recursively from perspective projections [24, 5, 42, 1, 49] using an EKF. The schemes are based upon minor variations of the same model, and none of them addresses the issue of its observability. There is also a vast literature on stereo-motion and on batch schemes for recovering structure and motion; for an extensive review see [14] and references therein.

Our work is somehow orthogonal to [17, 11], in what they assess the feasibility of structure estimation for known motion. We study instead the problem of motion estimation for unknown structure. Once motion is known, structure is linearly observable from the rigid motion model. Therefore if we estimate motion, and the estimates are properly weighted through the second order statistics of the estimation error, any “structure from motion” module incorporating motion error (for example [42, 49]) may be used for estimating scene structure reliably. Note that motion and structure play interchangeable roles: when either one is known, the other can be uniquely determined. However, motion can be reconstructed independent of structure, and structure estimates play a role only in disambiguating multiple solutions and in propagating scale information across time.

The model described in the last part of this paper is inspired by [34], and the results used for proving its observability/identifiability are from [37, 15].

1.2 Organization of the paper

In section 2 we define and formalize the visual motion problem, first in full generality, and then restricted to the case in which the scene is described by a number of “feature points” in the euclidean space. In this case the motion problem is *defined* by the rigid motion constraint and the perspective projection map.

In section 3 we show some alternative formulations of motion estimation in terms of inversion/estimation or identification/estimation of a nonlinear dynamical model. Motivated by the limitations of such models, we reformulate the problem in terms of state estimation of a nonlinear model defined on a linear state-space, raising the issue of observability.

In section 4 we address the linear observability and local (weak) observability of the model which *defines* the motion problem in the case of feature points in the euclidean 3D space. We show that it is neither linearly observable nor locally observable. We show how to reduce the set of indistinguishable states by imposing metric constraints on the state space. However, the local observability codistribution reaches full rank after three levels of bracketing, indicating that the model is hardly observable.

In section 5 we will describe a formulation of motion estimation as *identification of a systems in exterior differential form, with the parameters living in a topological manifold*, which we call the “essential manifold”. The model is globally observable with no level of differentiation once general position conditions hold. Such conditions are characterized using a simple rank test.

2 Visual motion and structure estimation

So far we have discussed about scene structure and motion estimation without referring to a specific scene (or object) representation. In a systems-theoretic framework, the structure of an object is encoded by a finite number of “feature points” on a differentiable manifold. These could be the coefficients of polynomial curves or splines fitting the contours of the object, or parameters describing locally its surface or, in the simplest case, points in the euclidean 3D space. In such a case the “structure manifold” is simply \mathbb{R}^3 , and the features correspond to salient regions of the object, as for example visible corners. In what follows we assume we are given the “correspondence” of projection of feature points across time, i.e. at each time instant we measure the projection of each feature point onto the image plane (or retina) and we know which point corresponds to which across time (for a review of some available methods for solving the correspondence problem see [2]).

Since we are interested in relative motion, it is equivalent to assume that the viewer is moving inside a static scene, carrying along his reference frame (viewer-centered representation) or that the scene is a rigid object which is moving in front of the viewer together with its reference (object-centered representation).

Viewer-centered representation The viewer maintains a local coordinatization of the feature space, which changes in time as he moves inside the scene. Meanwhile he perceives the projection of the scene onto its sensor (retina or CCD), which corresponds to measurements of a time invariant projection of the feature manifold in the motion-dependent (viewer-based) coordinatization.

If the viewer moves inside a *static* scene, his motion between two time instants is described as a point in $SE(3)$, the group of rigid transformations of \mathbb{R}^3 [39]. In this case we may imagine the viewer-centered coordinates as a family of diffeomorphisms parametrized by points of $SE(3)$. If the viewer moves with constant velocity, motion is represented by a single point, otherwise he describes a whole path in $SE(3)$.

Object-centered representation Alternatively we may imagine a “static” coordinate parametrization of the feature space and the viewer measuring the output of a projective map which is time-varying according to the motion of the scene. If the scene is moving rigidly, the viewer measures the output of a family of maps from the feature manifold onto an appropriate real-projective space, parametrized by points in $SE(3)$.

These two viewpoints, object-centered and viewer-centered, are substantially equivalent, since they may be transformed one into the other via a motion dependent diffeomorphism; however, each has its own advantages, depending on the application.

In both cases the visual motion problem consists in estimating the path in $SE(3)$ describing the relative motion between the viewer and the object.

2.1 Formalization of the model

Let us call N the feature n -manifold, $p \in N$ a feature point; M is the motion manifold, $q \in M$ represents motion,

$$\begin{aligned}\pi : N &\rightarrow \mathbb{R}P^{n-1} \\ p &\mapsto y = \pi(p)\end{aligned}$$

represents a perspective projection. Let \tilde{N} and \tilde{p} denote the local coordinates correspondents of N and p . We suppose there exists a set of C^∞ compatible charts on N :

$$\begin{aligned}\phi : N &\rightarrow \tilde{N} \\ p &\mapsto \tilde{p} = \phi(p).\end{aligned}$$

The visual motion problem is formalized in the **viewer-centered** representation by a family of maps on \tilde{N} parametrized by $q \in M$:

$$\begin{cases} \tilde{p}(t+1) = f(\tilde{p}(t), q) ; \tilde{p} \in \tilde{N}, q \in M \\ y = \pi(\tilde{p}) \end{cases}$$

where f is the map encoding rigid motion; $f(\tilde{p}, q) = \phi_q(\tilde{p})$, and

$$\begin{aligned}\phi_q : N \times M &\rightarrow \tilde{N} \\ (p, q) &\mapsto \tilde{p} = \phi_q(p) ; q \in M\end{aligned}$$

is the viewer-centered local-coordinates chart of N , which varies in time as the viewer moves with $q(t)$. Alternatively motion may be represented by its velocity: if $q \in M = SE(3)$, $\tilde{M} = T_e SE(3) \doteq se(3)$ is the lie algebra of twists corresponding to rigid motions. In this case we formalize the visual motion problem using a family of vector fields on \tilde{N} :

$$\begin{cases} \dot{\tilde{p}} = f(\tilde{p}, \tilde{q}) ; \tilde{p} \in \tilde{N}, \tilde{q} \in \tilde{M} \\ y = \pi(\tilde{p}). \end{cases}$$

The **object-centered** representation corresponds to the model

$$\begin{cases} p(t+1) = p(t) \\ y = \pi_q(p) \end{cases}$$

where

$$\begin{aligned}\pi_q : N \times M &\rightarrow \mathbb{R}P^{n-1} \\ (p, q) &\mapsto \pi \circ \phi^{-1} \circ f(\phi(p), q).\end{aligned}$$

Point-based visual motion

Let us now consider a simple paradigm, in which $p \in M = \mathbb{R}^3$ is a salient point in the scene, $\mathbf{X} \doteq [X, Y, Z]^T$ are its coordinates with respect to an orthonormal reference frame centered in the pupil of the viewer, with the Z axis pointing forward and X, Y arranged as to form a right-handed frame. Let $\tilde{q} = [V, \Omega]^T \in se(3)$ represent the canonical screw coordinates of the rigid motion of the viewer (body velocity [39]). As the viewer moves, each point describes a vector field on \mathbb{R}^3 ; in the viewer-centered representation, $f(\tilde{p}, \tilde{q})$ is simply

$$f(\mathbf{X}, V(t), \Omega(t)) = \Omega(t) \wedge \mathbf{X} + V(t).$$

If we represent motion between two time instants t and $t + \tau$, and velocity is held constant between the two samples, we have

$$f(\mathbf{X}, T(t, \tau), R(t, \tau)) = R(t, \tau)\mathbf{X} + T(t, \tau)$$

where (T, R) are related to (V, Ω) via [39]

$$R(t, \tau) \doteq e^{(\Omega(t) \wedge) \tau} \tag{1}$$

$$T(t, \tau) \doteq \frac{1}{\|\Omega(t)\|} \left[\left(I - e^{(\Omega(t) \wedge) \tau} \right) \Omega(t) \wedge + \Omega(t) \Omega(t)^T \tau \right] V(t). \tag{2}$$

In what follows we will assume $\tau = 1$ (constant sampling rate).

The map π is the trivial association of each $p \neq 0$ with its projective coordinates as an element of $\mathbb{R}P^2$:

$$\begin{aligned} \pi : \mathbb{R}^3 - \{0\} &\rightarrow \mathbb{R}P^2 \\ p &\mapsto [p]. \end{aligned}$$

In summary, if we represent the scene structure using points in the euclidean 3D space, the visual motion problem is *defined* by the constraints of rigid motion and perspective projection. For instance, in the viewer-centered instantaneous representation we have

$$\begin{cases} \dot{\mathbf{X}} = \Omega \wedge \mathbf{X} + V ; \mathbf{X}(0) = \mathbf{X}_0 \\ \mathbf{x} = \pi(\mathbf{X}), \end{cases}$$

where $\mathbf{x} = [x, y, 1]^T$ are the coordinates of the projection of the point \mathbf{X} onto the image plane.

3 Systems-theoretic characterization of visual motion estimation

We have seen that, if the scene is represented by a set of n points in 3D space moving rigidly with respect to the viewer, the visual motion problem is *defined* by the rigidity constraint and the perspective projection equations. If \mathbf{X}_i are the coordinates of the i^{th} point and \mathbf{x}_i the corresponding projection, we may write

$$\begin{cases} \dot{\mathbf{X}}_i = \Omega \wedge \mathbf{X}_i + V & \mathbf{X}(0) = \mathbf{X}_0 \\ \mathbf{x}_i = \pi(\mathbf{X}_i) + \nu_i & \nu_i \in \mathcal{N}(0, R_i) \end{cases} \quad \forall i = 1 : n \tag{3}$$

where ν_i represents an error in measuring the position of the projection of the point i . Solving the visual motion problem consists in estimating \mathbf{X}_i, V and Ω for all the visible points i , i.e.

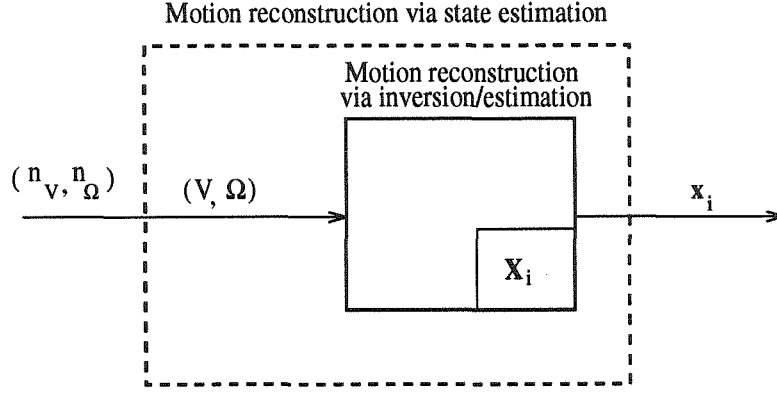


Figure 1: Interpretation of rigid motion reconstruction as inversion/estimation or identification/estimation of a dynamic model. The problem is transformed into a state estimation task

reconstructing both the input and the initial state of the above system from its noisy output (see fig. 1). Alternatively motion may be viewed as a vector of unknown parameters in the model (3) which have to be identified.

In the present section we show that it is possible to invert/identify the above system and solve for motion in a least-squares fashion; the solution is analogous to that found in [22]. However, the structure of the inverse system is intrinsically instantaneous, since the original model is driftless, and hence it does not exploit the benefits of dynamic observers.

3.1 Motion reconstruction via inversion/identification of a nonlinear model

Motion estimation may be viewed as an inversion problem for the model (3) when the initial state (structure) is unknown. It is well known that under certain conditions on the relative degree, it is possible to invert a nonlinear system [28]. In order to do that, we compute lie derivatives of the output along the state vector fields until the components of the input appear. If the coupling matrix is nonsingular, we may invert it and reconstruct the input of the system from bracket combinations of its output.

In our case the model is *driftless*, and therefore all the components of the input appear at the first level of differentiation:

$$\dot{\mathbf{x}}(t) = \left[\frac{1}{Z} \mathcal{A}(\mathbf{x}) \quad \mathcal{B}(\mathbf{x}) \right] \begin{bmatrix} V(t) \\ \Omega(t) \end{bmatrix}$$

where

$$\mathcal{A} \doteq \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \end{bmatrix} \quad \mathcal{B} \doteq \begin{bmatrix} -xy & 1+x^2 & -y \\ -1-y^2 & xy & x \end{bmatrix}. \quad (4)$$

If we observe enough points, we have an overdetermined system which we may solve for the motion parameters in a least-squares sense. Call $\mathcal{C}_i \doteq \left[\frac{1}{Z_i} \mathcal{A}_i \quad \mathcal{B}_i \right]$, we have

$$\begin{bmatrix} \hat{V}(t) \\ \hat{\Omega}(t) \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathcal{C}_i \\ \vdots \end{bmatrix}^\dagger \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} \doteq \mathcal{C}^\dagger \dot{\mathbf{x}}$$

where the symbol \dagger denotes the pseudo-inverse. Note that \mathcal{C}_i depends on the depth of the point, Z_i , which we do not know.

In order to reconstruct the initial depth it is necessary to *observe* it. Dynamic observers are in essence computing differentiations of the output until the matrix which couples the initial condition and the derivatives of the output (observability matrix or observability codistribution) has full rank. In our case, however, *both the input and the initial state appear at the same level of differentiation*, since the model is driftless. Therefore we may hope to recover either motion or depth with such a technique, but not both.

It is still possible, however, to exploit the above constraint to recover rigid motion by solving a nonlinear optimization problem constrained to a linear subspace. Such a problem may be formulated as the identification of an Exterior Differential System, as it is done in [46]. See [22, 46] for more details on this formulation.

3.2 Motion reconstruction as state estimation

Because the model described in the previous section has no drift dynamics, left-inversion/state-estimation reduces to a static (instantaneous) procedure and hence it does not exploit the noise rejection properties of dynamic observers.

One possible way to proceed, based on the above considerations, is to *introduce some dynamics* into the problem, using “dynamic extension”. Instead of considering motion as the input of the system, we consider as input its time derivative, and insert motion into the state dynamics.

If we view motion estimation as identification of the model (3), this “dynamic extension” corresponds to transforming the identification problem into a state estimation task. This has been done in the literature of recursive identification since the early sixties (see for example [44, 10, 30] and references therein). In essence the parameters to be estimated are inserted into the state dynamics and a Kalman filter is used as a parameter estimator (see figure 1).

In both interpretations we are led to an augmented model with the input being inserted into the state dynamics:

$$\begin{cases} \dot{\mathbf{X}}_i = \Omega \wedge \mathbf{X}_i + V & \mathbf{X}_i(0) = \mathbf{X}_{i0} \\ \dot{V} = f_V(V, \nu_V) & V(0) = V_0 \\ \dot{\Omega} = f_\Omega(\Omega, \nu_\Omega) & \Omega(0) = \Omega_0 \\ \mathbf{x}_i = \pi(\mathbf{X}_i) + \nu_i \end{cases} \quad \forall i = 1 : n. \quad (5)$$

Since we do not know f_V and f_Ω , the visual motion problem may be formulated as an “unknown-input/state estimation” problem. However, we may have some qualitative information about f_V, f_Ω which we want to exploit, for example a dynamical model, when the camera is mounted on a moving vehicle.

In absence of such information, f_V and f_Ω may describe a statistical model. The simplest case is $f_V = f_\Omega \cong 0$, which corresponds to constant velocity (or “small acceleration”). A model often used is brownian motion: $f_V = \nu_V, f_\Omega = \nu_\Omega$ are white, zero-mean gaussian noises¹ whose variances are to be considered as tuning parameters. Note that the random walk model allows us to deal with time-varying parameters (when the variation is slow compared to the sampling period) and also the positive definite variance of the model error helps preventing saturation, when a Kalman filter is used to perform the estimation task.

A crucial issue in state estimation using observers is of course *observability* of the model, which we address in the following section.

¹This notation is incorrect, since white noise is not integrable; indeed the notation is customary in statistical estimation, and we will adopt it here.

4 Perspective local observability of rigid motion

In this section we study the local observability of the model (3). We show that, in the case of constant velocity (or small acceleration), it is not locally observable. However, by enforcing metric constraints on the state space, it is possible to reduce the set of indistinguishable states. Some definitions and standard results on local observability may be found in appendix A.

4.1 Linear observability

We consider the linearization of the model (5): define $A \doteq \frac{\partial f}{\partial x} C \doteq \frac{\partial \pi}{\partial x}$ and \mathcal{A} as in (4). Suppose for simplicity $n = 1$; it is immediate to see that

$$\begin{aligned} A &= \begin{bmatrix} (\Omega \wedge) & I_3 & -(\mathbf{X} \wedge) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ C &= \left[\frac{1}{Z} \mathcal{A} \quad 0 \quad 0 \right] \\ CA^i &= \frac{1}{Z} \mathcal{A} \left[(\Omega \wedge)^i \quad (\Omega \wedge)^{i-1} \quad -(\Omega \wedge)^{i-1} (\mathbf{X} \wedge) \right]. \end{aligned}$$

The observability matrix for the linearized system is hence

$$O = \begin{bmatrix} \frac{1}{Z} \mathcal{A} & 0 & 0 \\ \frac{1}{Z} \mathcal{A} (\Omega \wedge) & \frac{1}{Z} \mathcal{A} & -\frac{1}{Z} \mathcal{A} (\mathbf{X} \wedge) \\ \vdots & \vdots & \vdots \\ \frac{1}{Z} \mathcal{A} (\Omega \wedge)^8 & \frac{1}{Z} \mathcal{A} (\Omega \wedge)^7 & -\frac{1}{Z} \mathcal{A} (\Omega \wedge)^7 (\mathbf{X} \wedge) \end{bmatrix}$$

The above reduces, for planar motion, to the matrix

$$\begin{bmatrix} \frac{1}{Z} & -\frac{X}{Z^2} & 0 & 0 & 0 \\ -\frac{X}{Z^2} \Omega_Y & -\frac{1}{Z} \Omega_Y & -1 - \frac{X}{Z^2} & \frac{1}{Z} & -\frac{X}{Z^2} \\ -\frac{1}{Z} \Omega_Y^2 & \frac{X}{Z^2} \Omega_Y^2 & 0 & -\frac{X}{Z^2} \Omega_Y & -\frac{1}{Z} \Omega_Y \\ \frac{X}{Z^2} \Omega_Y^3 & \frac{1}{Z} \Omega_Y^3 & \Omega_Y^2 + \frac{X^2}{Z^2} \Omega_Y^2 & -\frac{1}{Z} \Omega_Y^2 & \frac{X}{Z^2} \Omega_Y^2 \\ \frac{1}{Z} \Omega_Y^3 & -\frac{X}{Z^2} \Omega_Y^3 & 0 & \frac{X}{Z^2} \Omega_Y^3 & \frac{1}{Z} \Omega_Y^3 \end{bmatrix}$$

which is easily seen to have rank 3, since the last two rows are a scalar combination of rows 2 and 3 (and so are the subsequent rows). With a similar procedure the full matrix O is shown to have rank 5, in face of a state space of dimension 9. The linearized system is therefore not observable, and we say that the original model is not linearly observable.

4.2 Local observability

The local observability space is defined as the set of the output functions and all their possible lie derivatives along vector fields in the accessibility algebra [28]. Under slow acceleration conditions, the vector field in (3) is autonomous, and therefore the observability space is spanned by $\{\pi, L_f \pi, \dots, L_f^k \pi \dots\}$ where f denotes the state vector field. The observability codistribution is spanned by $d\mathcal{O} \doteq \{dh \mid h \in \mathcal{O}\}$. The state manifold is \mathbb{R}^9 ; the rank of the observability codistribution reaches its maximum of 8 after three levels of lie differentiation.

The null space of the observability codistribution, in case of nonzero forward translation, is:

$$\text{Null}([dh, dL_f h, dL_f^2 h, dL_f^3 h]) = \text{Span} \left[\begin{array}{cccccc} \frac{X}{V_Z} & \frac{Y}{V_Z} & \frac{Z}{V_Z} & \frac{V_X}{V_Z} & \frac{V_Y}{V_Z} & 1 \ 0 \ 0 \ 0 \end{array} \right].$$

In the case of zero forward translation, after four levels of differentiation we have a null space spanned by

$$\left[\frac{X}{V_Y} \quad \frac{Y}{V_Y} \quad \frac{Z}{V_Y} \quad \frac{V_X}{V_Y} \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \right].$$

In case of only lateral translation, the null space of the observability codistribution is spanned by

$$\left[\frac{X}{V_X} \quad \frac{Y}{V_X} \quad \frac{Z}{V_X} \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \right]$$

when translation is along X , while for translation along Y we have

$$\text{Null}([dh, dL_f h, \dots, dL_f^3 h]) = \left[\frac{X}{V_Y} \quad \frac{Y}{V_Y} \quad \frac{Z}{V_Y} \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \right].$$

In the case of pure rotation, a basis of the null space of the observability codistribution is obviously

$$\left[\frac{X}{Z} \quad \frac{Y}{Z} \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \right],$$

and all the points having the same projective coordinates are indistinguishable. In the case of nonzero forward translation, the set of states which are indistinguishable from $[\mathbf{X}_0 \ V_0 \ \Omega_0]^T$ is

$$I \left(\begin{bmatrix} \mathbf{X}_0 \\ V_0 \\ \Omega_0 \end{bmatrix} \right) = \left\{ \begin{bmatrix} \mathbf{X}_0 \frac{V_{03}+s}{V_{03}} \\ V_0 \frac{V_{03}+s}{V_{03}} \\ \Omega_0 \end{bmatrix} \mid s \in \mathbb{R} \right\},$$

which is a linear variety of dimension one, and similarly for the other cases.

4.3 Global scale ambiguity: metric constraints on the state manifold

Consider the solution $\mathbf{X}_i(t, \mathbf{X}_{i0}, V_0, \Omega_0)$ of (5) starting from the initial condition $\mathbf{X}_{i0}, V_0, \Omega_0$:

$$\mathbf{X}_i(t) = \begin{cases} e^{(\Omega_0 \wedge)t} \mathbf{X}_{i0} + \mathcal{T}(\Omega_0, t) V_0 & \text{if } \|\Omega_0\| \neq 0 \\ V_0 t & \text{otherwise} \end{cases}$$

where

$$\mathcal{T}(\Omega, t) \doteq \frac{1}{\|\Omega\|} \left[(I - e^{(\Omega \wedge)t}) \Omega \wedge + \Omega \Omega^T t \right].$$

It is immediate to see that $\mathbf{X}_i(t, \alpha \mathbf{X}_{i0}, \alpha V_0, \Omega_0) = \alpha \mathbf{X}_i(t, \mathbf{X}_{i0}, V_0, \Omega_0) \ \forall \ \mathbf{X}_{i0}, V_0, \Omega_0, t, \alpha$. Since for central projection we have $\mathbf{x}_i(t) = \pi(\mathbf{X}_i) = \pi(\alpha \mathbf{X}_i)$, we conclude that any initial condition $\alpha_1 \mathbf{X}_{i0}, \alpha_1 V_0, \Omega_0$ is indistinguishable from $\alpha_2 \mathbf{X}_{i0}, \alpha_2 V_0, \Omega_0$, for any possible $\alpha_1, \alpha_2 \in \mathbb{R}$.

This one-dimensional unobservable space is very familiar, as we experience that an object moving in front of us produces the same impression of a similar one which is “twice as big, twice as far, and moving twice as fast”. However, we may impose norm constraints on the visible objects or on the translational velocity in order to get rid of the scale factor ambiguity. For example, if we impose $\|V_0\| = 1$, two initial conditions are indistinguishable only if $\alpha_1 = \pm \alpha_2$.

There is still some information which is hidden by the model (5): we know that, if an object is visible, it must be in front of the observer, i.e. $Z_i \geq 0 \ \forall \ i$. Moreover, no points are allowed to lie on the focal plane $Z = 0$ (plane at infinity), and therefore $\alpha_1 = \alpha_2$.

If we apply such metric constraints to the locally unobservable codistribution, we can reduce the set of indistinguishable states to the trivial set. However, an appropriate model should include such constraints *explicitly* into the state manifold. This may be done at the price of transforming the state from the linear space \mathbb{R}^9 to the differentiable manifold with boundary $\mathbb{R}^2 \times H^1 \times \mathbb{S}^2 \times \mathbb{R}^3$ [4].

We now summarize some of the limitations of the model (5):

- The model is not locally observable. Metric constraints which make the model observable are not explicitly encoded in the state representation.
- Three levels of lie bracketing are needed to cover the observable part of the state space. Indeed, we know that it is possible to estimate motion and structure from the first derivative of the projection of the points (optical flow) [34, 27, 15].
- The model has the property of being “block diagonal” with respect to the structure, so that the states corresponding to different points are independent. Indeed it is strongly intuitive that *the more points are visible, the better the perception of motion ought to be.*

5 Global observability: motion estimation as identification of an Exterior Differential System

In this section we describe an alternative model for formulating the visual motion problem which has been presented in [47] and is related to a motion representation first introduced by Longuet-Higgins in [34].

Motion estimation is viewed as a problem of identifying a system in Exterior Differential Form [7] with parameters on a topological manifold, called the “essential manifold” [47]. We show that the model is globally observable/identifiable with zero level of differentiation for any number of visible points. When more points are available, the redundancy may be exploited in order to reduce the effect of the measurement noise.

5.1 The “essential model”

Consider a point in 3D space, with coordinates $\mathbf{X}_i(t)$ in the viewer’s reference. Let $\mathbf{X}_i(t + \tau)$ be the coordinates after a rigid motion of the viewer (T, R) , of which (V, Ω) are the canonical screw coordinates [39] as in equations (1,2)².

It is immediate to see that $\mathbf{X}_i(t)$, $\mathbf{X}_i(t + \tau)$ and T are coplanar, and hence their triple product is zero. Once expressed in a common reference, for example the viewer’s at time t , the coplanarity constraint becomes [34]

$$\mathbf{X}_i^T(t + \tau)RT \wedge \mathbf{X}_i(t) = 0 \quad \forall i = 1 : n.$$

Note that the same relationship holds for $\mathbf{x}_i(t)$ and $\mathbf{x}_i(t + \tau)$, since they represent the projective coordinates of $\mathbf{X}_i(t)$ and $\mathbf{X}_i(t + \tau)$; $T \wedge \in so(3)$ is a skew symmetric matrix. After defining the *essential matrix* as $\mathbf{Q} \doteq R(T \wedge)$, the *essential constraint* becomes

$$\mathbf{x}_i^T(t + \tau)\mathbf{Q}\mathbf{x}_i(t) = 0 \quad \forall i = 1 : n. \quad (6)$$

Since there is an arbitrary scale factor in the above equality, we impose $\|\mathbf{Q}\| = \|T\| = 1$. It has been shown [34, 15, 50] that this constraint is not only a consequence of a rigid motion, but it also characterizes it, in the sense that a sufficient number of constraints (6) allows us to characterize rigid motion [34]. The essential matrix was first introduced by Longuet-Higgins [34], together with a quasi-linear batch technique for estimating structure and motion from two views and more than 8 visible points. His technique was then extended and developed in [50, 19, 15]. In [43]

²Note that T in this section differs from the one used in the previous section. Rigid motion is represented here as $\mathbf{X}(t + \tau) = R(\mathbf{X}(t) - T)$, for consistency with the notation of [34].

the algorithm is implemented using three pipelined Singular Value Decompositions. Finally [47] proposes a recursive version of the motion estimation technique based on the essential constraint. The essential matrices are points of the space

$$\tilde{E} \doteq \{RS | R \in SO(3), S = T\wedge \in so(3) \mid \|T\| = 1\}$$

which has the structure of an algebraic variety [15]. We now show that a slight modification of E is a topological manifold of class at least C^0 .

Theorem 5.1 *Let $d_{x,x'}(\mathbf{Q})$ be the triangulation function³, which gives the depth of a point from its motion \mathbf{Q} and its projective coordinates \mathbf{x}, \mathbf{x}' . Then $E \doteq \tilde{E} \cap d_{x,x'}^{-1}(\mathbb{R}_+^2)$ is a topological manifold of class at least C_0 .*

Proof:

E inherits the topology from \mathbb{R}^9 . Consider the map

$$\begin{aligned} \Phi : E &\rightarrow \mathbf{S}^2 \times \mathbb{R}^3 \sim \mathbb{R}^5 \\ \mathbf{Q} &\mapsto \begin{bmatrix} T \\ \Omega \end{bmatrix} = \begin{bmatrix} \pm \mathbf{V}_{.3} \\ \mathbf{U}R_Z(\pm \frac{\pi}{2})\mathbf{V}^T \end{bmatrix} \end{aligned} \quad (7)$$

where \mathbf{U}, \mathbf{V} are defined by the Singular Value Decomposition (SVD) [18] of $\mathbf{Q} = \mathbf{U}\Sigma\mathbf{V}^T$, $\mathbf{V}_{.3}$ denotes the third column of \mathbf{V} and $R_Z(\frac{\pi}{2})$ is a rotation of $\frac{\pi}{2}$ about the Z axis. As usual Ω is the rotation 3-vector corresponding to the 3×3 rotation matrix $\mathbf{U}R_Z(\frac{\pi}{2})\mathbf{V}^T$ and is obtained using the Rodrigues' formulae [39], which give in fact a local representation of $SO(3)$ ⁴. T is represented in spherical coordinates. Note that the map Φ defines the local coordinates of the essential manifold modulo a sign in the direction of translation and in the rotation angle of R_Z , therefore the map Φ associates to each element of the essential manifold 4 distinct points in local coordinates. This ambiguity can be resolved by imposing the “positive depth constraint”, i.e. that each visible point lies in front of the observer [34, 35]. Consider one of the four local counterparts of $\mathbf{Q} \in E$, and the function $d_{\mathbf{x},\mathbf{x}'} : E \rightarrow \mathbb{R}^{1+1}$ defined by

$$d_{\mathbf{x},\mathbf{x}'}(\mathbf{Q}) = [Z, Z']^T \quad (8)$$

with $Z = \frac{\langle \mathbf{n}^i, \mathbf{m}^i \rangle}{\|\mathbf{m}^i\|^2} \quad \forall i = 1, \dots, n, \quad \mathbf{m}^i = (R\mathbf{x}^i) \wedge \mathbf{x}'^i$ and $\mathbf{n}^i = (RT) \wedge \mathbf{x}'^i$, which gives depth of each point as a function of the projection and the motion parameters⁵. Note that it is locally smooth away from zero translation. Now redefine the essential space as

$$E \doteq E \cap d_{x,x'}^{-1}(\mathbb{R}_+^2)^n = \{\mathbf{Q} = RS | R \in SO(3), S \doteq T\wedge \in so(3), \|T\| = 1, d_{x^i, x'^i}(\mathbf{Q}) > 0 \forall i = 1..n\}$$

where \mathbb{R}_+ is the positive open half space of \mathbb{R} , $d_{x,x'}^{-1}$ denotes the preimage of $d_{x,x'}$. Consider Φ restricted to E . It follows from the properties of the SVD that Φ is continuous, and furthermore it is bijective. It can be shown (see appendix B) that $\mathbf{Q} \in E \Leftrightarrow \Sigma = \text{diag}\{1 \ 1 \ 0\}$ and hence the

³See equation (8) in the proof for an instance of realization of the triangulation function.

⁴We use a hybrid representation, with T being the discrete translation between two successive samples, while Ω the instantaneous rotational velocity which corresponds to R . Note that (V, Ω) and (T, R) are related via the exponential map of equations (1,2).

⁵Note that when the triangulation is computed for a large number of points, a “Total Least Squares” solution for Z is most appropriate [18]

subspaces $\langle \mathbf{V}_1, \mathbf{V}_2 \rangle$ and $\langle \mathbf{U}_1, \mathbf{U}_2 \rangle$ are allowed to switch. This happens, however, without affecting continuity of T and Ω . The inverse map is simply

$$\begin{aligned} \Phi^{-1} : \mathbf{S}^2 \times \mathbb{R}^3 &\rightarrow E \\ \begin{bmatrix} T \\ \Omega \end{bmatrix} &\mapsto e^{(\Omega \wedge)}(T \wedge). \end{aligned}$$

which is smooth. Hence E is a topological manifold of class at least C_0 .

Q.E.D.

Remark 5.1 *Note that the essential manifold is defined independently of the existence of observed points, as a compact representation of $SE(3)$, the former being embedded in a linear space of smaller dimension. However, in order E to qualify as a manifold, we need to take into account at least one visible point. This is done by imposing the “positive depth constraint”, i.e. imposing that the visible points have positive depth in both the coordinate frames at time t and $t + \tau$.*

If we let $\tau \rightarrow 0$ in the above construction, under the small acceleration assumption we may write the essential constraint as

$$\begin{cases} \dot{\mathbf{x}}_i^T \mathbf{Q} \mathbf{x}_i = 0 & \mathbf{Q} \in E \\ y_i = \mathbf{x}_i + \nu_i, \end{cases} \quad \forall i = 1 : n$$

which represents a system in exterior differential form [7]: $f(x)dx = 0$, with $f(x) = (\mathbf{Q}^T x)^T$. Hence motion estimation may be viewed as *identification of a linear exterior differential system, with parameters on the essential manifold*. Note that observability/identifiability of the above model is independent of the euclidean structure, for it depends solely on the projective coordinates of the points. Note also that the role of structure (depth) is to allow us choosing one of the four branches of the local coordinates chart. This needs to be done only at one time step, and then propagated across time.

5.2 Observability for $N \geq 8$ points

Since the essential constraint is linear in \mathbf{Q} , it is possible to write it using the (improper) notation

$$\chi(\mathbf{x}, \dot{\mathbf{x}})\mathbf{Q} = 0$$

where χ is a $n \times 9$ matrix and \mathbf{Q} is interpreted as a nine-vector obtained by stacking the columns of \mathbf{Q} on top of each other. The generic row of χ is $[x_1 x'_1 \ x_2 x'_1 \ x'_1 \ x_1 x'_2 \ x_2 x'_2 \ x'_2 \ x_1 \ x_2 \ 1]$. We write $\mathbf{x} \doteq [x_1 \dots x_n]^T$.

Following the track of the previous sections, we will assume small acceleration or a statistical model for motion which, lifted to the essential manifold, results in a statistical model for \mathbf{Q} . The resulting model has the form

$$\begin{cases} \dot{\mathbf{Q}} = f(\mathbf{Q}, \nu_Q) \\ \chi(\mathbf{x}, \dot{\mathbf{x}})\mathbf{Q} = \nu_\chi \end{cases} \quad \mathbf{Q} \in E$$

where f is either zero or some statistical model; ν_Q and ν_χ are noise processes which can be characterized, as discussed in [48]. In [47], two recursive schemes are proposed for solving the estimation problem: one is based upon an Implicit Extended Kalman Filter (IEKF) in the local coordinates of the essential manifold, the other is based upon a linear update on the linear embedding space \mathbb{R}^9 , followed by a projection onto the essential manifold.

Now consider χ : if it has rank 8, then there exists a unique \mathbf{Q} which spans its null space modulo a sign, since we have imposed a constraint on its norm. This generates four distinct points in the local coordinates which reduce to a single solution once the positive depth constraint is imposed. Once this is done at one step, we choose a branch of the local coordinates map (which then becomes a local homeomorphism) and stick with it for the subsequent time steps [34, 50, 15].

We are led naturally to the following definition:

Definition 5.1 We say the points \mathbf{x} are in general position $\Leftrightarrow \text{rank}(\chi(\mathbf{x}, \dot{\mathbf{x}})) = 8$.

Note that the general position condition also depends on motion. The following are tautologies which come as a direct consequence of the definitions:

Claim 5.1 The essential model is observable/identifiable modulo a sign under general position conditions.

Claim 5.2 If an essential model is in general position then it is possible to reconstruct the motion (V, Ω) of the viewer modulo four solutions. The solution is unique once the positive depth constraint is imposed at one time instant.

We still have to address the issue of the conditions under which the matrix χ has full rank. Furthermore we need to deal with the case of less than 8 visible points, since it automatically excludes general position conditions.

5.3 Observability with less than 8 points

When less than 8 points are visible, it is not possible to achieve the above sufficient conditions for motion observability. Suppose that, at time $t + \tau_i$, the matrix $\chi(t + \tau_i)$ has a null space of dimension k_i . If the viewer moves with constant velocity, then we may write

$$\begin{aligned} \chi(t)\mathbf{Q}(t) &= 0 \\ \chi(t + \tau_1)\mathbf{Q}(t + \tau_1) &= \chi(t + \tau_1)\mathbf{Q}(t) = 0 \\ &\vdots = \vdots \\ \chi(t + \tau_p)\mathbf{Q}(t + \tau_p) &= \chi(t + \tau_p)\mathbf{Q}(t) = 0 \end{aligned}$$

until $k_0 + k_1 + \dots + k_p = 8$. If this happens, we may go back to the previous case and restate the sufficient conditions for motion observability for the extended matrix

$$\bar{\chi}_p \doteq \begin{bmatrix} \chi(t) \\ \chi(t + \tau_1) \\ \vdots \\ \chi(t + \tau_p) \end{bmatrix}.$$

Of course this may not happen. As a consequence of the above observations, we redefine general position as follows:

Definition 5.2 We say an essential model is in general position (GP) when either there are more than 8 visible points and χ has rank 8, or there exists a time instant τ_p such that the extended matrix $\bar{\chi}_p$ has rank 8.

Then we may restate the sufficient conditions for observability as a new tautology

Theorem 5.2 If an essential model is in general position, then it is globally observable.

5.4 General position: rank condition for global observability of rigid motion

We are now interested in writing explicitly the general positions condition and promote the previous tautologies to true claims. This is done using results in [35, 15, 50] for the case of more than 8 points. The claim, extended to our general position condition, may be stated as:

Theorem 5.3 *An essential model is in general position \iff there does not exist a (proper) quadric surface⁶ in \mathbb{R}^3 which contains all the visible points and the path of the center of projection.*

Remark 5.2 *We report here a proof given by Mennucci [38] for the case of more than 8 visible points. See also [35]. The general case is obtained by substituting $\mathbf{x}_{t+\tau_i}$ for \mathbf{x}_i . Note that the quadric surface is a thin set in the 3D euclidean space. The measurement noise in the projected coordinates is sufficient to set the model in general position. Note also that $T \neq 0$ plays a critical role in achieving global observability, while Ω (or R) is uninfliuent.*

Proof:

Let $T \neq 0$. Consider the points to be fixed in an intermediate reference system, where their coordinates are (\mathbf{X}_i'') such that $\mathbf{X}_i' = R(\mathbf{X}_i'' - T)$, $\mathbf{X}_i = R^T(\mathbf{X}_i'' + T)$; then $\mathbf{X}_i'^T \mathbf{Q} \mathbf{X}_i = 0 \quad 1 \leq i \leq n$, and the same holds for \mathbf{x} in place of \mathbf{X} .

By substitution we get

$$\mathbf{x}_i'^T \mathbf{Q} \mathbf{x}_i = [R(\mathbf{x}_i'' - T)]^T \mathbf{Q} R^T(\mathbf{x}_i'' + T) = (\mathbf{x}_i'' - T)^T R^T \mathbf{Q} R^T(\mathbf{x}_i'' + T) = 0 \quad 1 \leq i \leq n \quad (9)$$

We may change the variable in this equation to be $\mathbf{Q}' = R^T \mathbf{Q} R^T$; since R is invertible, this would not change any of the considerations below. Note that this implies that there is no dependency on rotation; we will therefore assume $R = I$ without loss of generality.

Equation (9) becomes

$$(\mathbf{x}_i'' - T) \mathbf{Q} (\mathbf{x}_i'' + T) = \mathbf{x}_i''^T \mathbf{Q} \mathbf{x}_i'' - T^T \mathbf{Q} \mathbf{x}_i'' + \mathbf{x}_i''^T \mathbf{Q} T - T^T \mathbf{Q} T = 0 \quad 1 \leq i \leq n \quad (10)$$

Call $\langle \mathbf{Q} \rangle \doteq \{\mathbf{Q} \in \mathbb{R}^{3 \times 3} \mid (\mathbf{x}_i'' - T)^T \mathbf{Q} (\mathbf{x}_i'' + T) = 0, \quad 1 \leq i \leq n\}$; $\langle \mathbf{Q} \rangle$ is a vector subspace of $\mathbb{R}^{3 \times 3}$, and the fact that there is only one solution is equivalent to saying that the dimension of $\langle \mathbf{Q} \rangle$ is one; indeed, $\dim(\langle \mathbf{Q} \rangle)$ is always bigger or equal than one, since it contains the matrix $T \wedge$, as can be seen by direct substitution in eq. (10).

Suppose that the equation (9) holds for a matrix M , and decompose it in the symmetric and antisymmetric part $A = \frac{M - M^T}{2}$, $S = \frac{M + M^T}{2}$, then

$$\mathbf{x}_i''^T S \mathbf{x}_i'' - 2T^T A \mathbf{x}_i'' - T^T S T = 0 \quad 1 \leq i \leq n.$$

Consider the set $\langle V \rangle \doteq \{x \in \mathbb{R}^3 \mid x^T S x - 2T^T A x - T^T S T = 0\}$. This set always contains the two points T and $-T$, the centers of projection (as a simple computation shows).

Suppose there is no (proper) quadric surface containing the points \mathbf{x}_i'' ; then it must be that $V = \mathbb{R}^3$, that means that $S = 0$ and $T^T A = 0$; this means that M is necessarily a multiple of $T \wedge = \mathbf{Q}$, so we get that $\dim(\langle \mathbf{Q} \rangle) = 1$.

Viceversa, suppose that the symmetric part S of M is nonzero or that $T^T A \neq 0$: then the set $\langle V \rangle$ is a quadric surface that contains the points \mathbf{x}_i'' (by definition), and it contains the points T and $-T$; the latest are the two centers of projection (if the symmetric part $S = 0$, then the set $\{x \in \mathbb{R}^3 \mid T^T A x = 0\}$ is a plane, that is anyway a quadric surface). Q.E.D.

⁶A quadric surface is a set $\{x \in \mathbb{R}^{3 \times 3} \mid x^t A x + b^t x + c = 0\}$ where A is a 3×3 matrix, b is a 3-vector and c is a scalar. It is proper if it is a proper subset of \mathbb{R}^3 .

6 Conclusions

We have analyzed the observability of rigid motion under projection. The model which defines the problem for feature points in the euclidean 3D space lacks of local observability. The observable manifold is covered with three levels of lie differentiation. The problem is indeed observable once formulated in the appropriate topological space.

We have studied a formulation of visual motion estimation in terms of identification of an Exterior Differential System with parameters on the essential manifold [47]. The model is globally observable/identifiable with zero level of bracketing; when more points are available, redundancy may be exploited to reduce the effect of measurement noise.

Acknowledgements

The authors wish to thank Prof. Ruggero Frezza and Prof. Giorgio Picci for their constant support, advice and directions.

The discussions with Prof. Richard Murray, Michiel van Nieuwstadt and Andrea Mennucci were very helpful in understanding the geometric structure of the problem. The suggestions of Prof. J. K. Åström, Prof. John Doyle, Prof. Manfred Morari and Prof. Shankar Sastry were also very valuable.

References

- [1] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. *Proc. CVPR*, New York, 1993.
- [2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. RPL-TR 9107, Queen's University Kingston, Ontario, Robotics and perception laboratory, 1992. Also in *Proc. CVPR 1992*, pp 236-242.
- [3] A. Blake, M. Taylor, and A. Cox. Grasping visual symmetry. *Proc. of the ICCV*, 1993.
- [4] W. Boothby. *Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, 1986.
- [5] T. Broida and R. Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular frames. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991.
- [6] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE trans. PAMI*, Jan. 1986.
- [7] Bryant, Chern, Goldberg, and Goldsmith. *Exterior Differential Systems*. Mathematical Research Institute. Springer Verlag, 1992.
- [8] R.S. Bucy. Non-linear filtering theory. *IEEE Trans. A.C. AC-10*, 198, 1965.
- [9] Ke Wang Chen. Observability and invertibility of nonlinear systems: a differential algebraic approach. Technical report, PhD Thesis–Linköping University–Sweden, 1991.
- [10] H. Cox. On the estimation of state variables and parameters from noisy dynamical systems. *IEEE Trans. A.C. AC-9*, 5-12, 1964.
- [11] W. Dayawansa, B. Ghosh, C. Martin, and X. Wang. A necessary and sufficient condition for the perspective observability problem. -, 1993.
- [12] E. D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241–261, 1988.
- [13] E. D. Dickmanns and V. Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1:223–240, 1988.
- [14] O. Faugeras. *Three dimensional vision, a geometric viewpoint*. MIT Press, 1993.
- [15] O. D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *Int. J. of Computer Vision*, 1990.
- [16] D.B. Gennery. Tracking known 3-dimensional object. In *Proc. AAAI 2nd Natl. Conf. Artif. Intell.*, pages 13–17, Pittsburg, PA, 1982.
- [17] B. Ghosh, M Jankovic, and Y. Wu. Perspective problems in systems theory and its application in machine vision. *Submitted to Journal of Math. Systems, Est. and Control*, 1993.
- [18] G. Golub and C. Van Loan. *Matrix computations*. Johns Hopkins University Press, 2 edition, 1989.

- [19] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd Europ. Conf. Comput. Vision, G. Sandini (Ed.), LNCS-Series Vol. 588, Springer-Verlag, 1992.*
- [20] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura. Image-based dynamic visual servo for a hand-eye manipulator. In Kodama Kimura, editor, *Recent advances in mathematical theory of systems, control, networks, and signal processing II*, pages 609–614. Proceedings of the international symposium of MTNS, Mita Press, 1991.
- [21] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura. Manipulator control with image-based visual servo. In *IEEE intl' Conference on Robotics and Automation*, pages 2267–2272, 1991.
- [22] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion i: algorithm and implementation. *Int. J. Comp. Vision vol. 7 (2)*, 1992.
- [23] J. Heel. Dynamic motion vision. *Proc. of the DARPA image understanding workshop*, 1989.
- [24] J. Heel. Direct estimation of structure and motion from multiple frames. *AI Memo 1190, MIT AI Lab*, March 1990.
- [25] R. Hermann and A. J. Krener. Nonlinear controllability and observability. *IEEE Trans. Aut. Contr. AC-22 pp. 728-740*, 1977.
- [26] C.C. Ho and N.H. McClamrock. Autonomous spacecraft docking using a computer vision systm. *Proc. of the 31st CDC – Tucson, AZ*, 1992.
- [27] B.K.P. Horn. Relative orientation. *Int. J. of Computer Vision*, 4:59–78, 1990.
- [28] A. Isidori. *Nonlinear Control Systems*. Springer Verlag, 1989.
- [29] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [30] J.K. Åström and P. Eykhoff. System identification, a survey. *Automatica, vol. 7*, 1971.
- [31] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.
- [32] A. J. Krener and A. Isidori. Linearization by output injection and nonlinear observers. *Systems and Control Letters vol. 3*, 1983.
- [33] W. Lee and K. Nam. Observer design for autonomous discrete-time nonlinear systems. *Systems and Control Letters vol. 17*, 1991.
- [34] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
- [35] H.C. Longuet-Higgins. Configurations that defeat the eight-point algorithm. *Mental processes: studies in cognitive science, MIT press*, 1987.
- [36] L. Matthies, R. Szelisky, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.
- [37] S. Maybank. *Theory of reconstruction from image motion*. Springer Verlag, 1992.

- [38] Andrea C. Mennucci. Recovering ego motion: an alternative proof for the configurations that defeat the eight-point longuet-higgins' algorithm. *CDS memo, California Institute of Technology*, Jan. 1994.
- [39] R.M. Murray, Z. Li, and S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. Preprint, 1993.
- [40] H. Nijmeijer. Observability of autonomous discrete time nonlinear systems. *Int. J. Control* vol. 36 (5), 1982.
- [41] H. Nijmeijer and A.J. Van Der Shaft. *Nonlinear Dynamical Control Systems*. Springer Verlag, 1990.
- [42] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.
- [43] P. Perona and S. Soatto. Motion and structure from 2 perspective views of p points – algorithm and error analysis. *Technical Report CIT/CNS 23-93 – California Institute of Technology*, Oct. 1992.
- [44] R.E.Kopp and R.J. Orford. Linear regression applied to system identification for adaptive control systems. *AIAA Journal* , 2300-2306, 1963.
- [45] A. J. Van Der Shaft. Observability and controllability for smooth nonlinear systems. *SIAM J. Control and Optim.* vol. 20 (3), 1982.
- [46] S. Soatto. Subspace methods for recovering rigid motion realized by identifying exterior differential systems. *Technical Report CIT-CDS 94-005, California Institute of Technology*, 1994.
- [47] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *Proc. of the ECCV 94 – To appear in “Lecture Notes in Computer Sciences”, Springer Verlag*, May 1994.
- [48] S. Soatto, R. Frezza, P. Perona, and G. Picci. Dynamic estimation of rigid motion from perspective views via recursive identification of exterior differential systems with parameters on a topological manifold. *Technical Report CIT-CDS 94-004 , California Institute of Technology*, Feb. 1994.
- [49] S. Soatto, P. Perona, R. Frezza, and G. Picci. Recursive motion and structure estimation with complete error characterization. In *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recogn.*, pages 428–433, New York, June 1993.
- [50] J. Weng, T. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(5):451–476, 1989.

A Notation

In this section we introduce some notation, referring to [40, 45, 33, 32, 25], for the system:

$$\begin{cases} \dot{x} = f(x, u) ; x(t_0) = x_0 \\ y = h(x) \end{cases} \quad (*)$$

where $x \in N \subset \mathbb{R}^n$, some n-dimensional manifold, $u \in M \subset \mathbb{R}^m$ and $y \in P \subset \mathbb{R}^p$; it is assumed that f and h are smooth functions. The set of admissible inputs is described as $\mathcal{U} \doteq \{u : \mathbb{R}^+ \rightarrow P \subset \mathbb{R}^p\}$ such that

1. \mathcal{U} is closed under concatenation
2. f describes a family of vector fields parametrized by $\bar{u} \in P$.
3. u are piecewise constant functions which are piecewise continuous from the right:

$$u(t) \doteq \{\bar{u}_i \text{ for } t \in I_i = [t_1 + \dots + t_{i-1}, t_1 + \dots + t_i) \mid \bar{u}_i \in P \subset \mathbb{R}^p, \forall i\}.$$

We call $f_i \doteq f(x, \bar{u}_i)$; in the time interval I_i the system evolves along the integral curve of f_i . The above assumptions may be partially released; however, they are general enough for our purposes. In studying the visual motion problem, we will be mostly concerned with the autonomous case: $u(t) = 0 \forall t$.

Definition A.1 x_1 and x_2 are said to be indistinguishable (and denoted with $x_1 I x_2$) $\Leftrightarrow \forall u \in \mathcal{U}$, $h(\phi_t(x_1, u)) = h(\phi_t(x_2, u)) \quad \forall t \geq 0$.

$I(x) \doteq \{x_i \mid x_i I x ; x \in N\}$ is the set of states which are indistinguishable from x .

Definition A.2 (*) is completely observable (C-O) at x $\Leftrightarrow I(x) = \{x\}$.

(*) is completely observable \Leftrightarrow it is C-O at $x \forall x \in N$.

Definition A.3 Given an open set $U \subset N$, x_1 and x_2 are said to be U-indistinguishable (and denoted with $x_1 I^U x_2$) $\Leftrightarrow \{\phi_t(x_1, u) \in U, \phi_t(x_2, u) \in U \forall t \in [t_0, t_1]\} \Rightarrow h(\phi_t(x_1, u)) = h(\phi_t(x_2, u)) \quad \forall t \in [t_0, t_1]$.

$I^U(x) \doteq \{x_i \mid x_i I^U x\}$ is the set of states which are U-indistinguishable from x .

Definition A.4 (*) is said to be locally weakly observable (L-W-O) at x $\Leftrightarrow \exists U, x \in U \mid \forall V \subset U, x \in V, I^V(x) = \{x\}$.

(*) is said to be locally weakly observable \Leftrightarrow it is L-W-O at $x \forall x \in N$.

Definition A.5 The observability space \mathcal{O} for (*) is defined to be the smallest subspace of $C^\infty(N)$ which contains the functions $\{h_1 \dots h_p\}$ and is invariant under lie differentiation along vector fields in $\tau \doteq \{f_i = f(x, \bar{u}_i)\}$.

Definition A.6 The observability codistribution is defined as

$$d\mathcal{O} \doteq \{d\lambda \mid \lambda \in \mathcal{O}\}$$

The observability codistribution is the smallest codistribution which is invariant for (*) and contains the forms dh . It can be shown that the definition does not change if we allow the vector fields in τ to belong to the accessibility algebra, which consists of repeated lie brackets of vector fields in τ .

Definition A.7 A system is said to satisfy the observability rank condition (ORC) at $p \stackrel{*}{\Leftrightarrow} \dim(d\mathcal{O})_p = n$.

Remark A.1 The ORC can be stated in terms of exterior differential systems. In fact we may interpret the observability codistribution as a Pfaffian system [7]

$$d\mathcal{O} \doteq dh + dL_{\bar{f}}h + \dots + dL_{\bar{f}}^{(n-1)}h$$

where $\bar{f} \doteq f(\cdot, \bar{u})$, n is the dimension of the state-space manifold N . The observability rank condition may be state as:

Definition A.8 .

The system $(*)$ satisfies the observability rank condition at $p \stackrel{*}{\Leftrightarrow} d\mathcal{O}_p = T_p^*N$

Theorem A.1 If $\dim(\mathcal{O}) = n$ at p , then $(*)$ is locally weakly observable in a neighborhood of p .

Proof:

see [41, 28, 9] This condition is not necessary [9]; however, the following result holds:

Theorem A.2 If \mathcal{O} has constant dimension and the system $(*)$ is locally weakly observable, then $\text{rank}(\mathcal{O}) = n$.

B Characterization of the essential space

The essential space has been defined in section 5; in this appendix we show a simple characterization which is due to Faugeras and Maybank [15, 37].

Theorem B.1 .

Let $Q = U\Sigma V^T$ be the SVD of an element of $\mathbb{R}^{3 \times 3}$. Then

$$Q \in E \Leftrightarrow \Sigma = \Sigma_0 = \text{diag}\{\lambda \ \lambda \ 0\} \mid \lambda \in \mathbb{R}^+.$$

Proof:

(\Rightarrow) let $Q = RS \mid R \in SO(3), S \in so(3)$; $\sigma(Q)$, the set of singular values of Q , is such that $\sigma(Q) = \sqrt{\sigma(QQ^T)}$. Next observe that $QQ^T = RSS^TR^T = SS^T = -S^2$. Also $\forall S \in so(3) \exists ! T \mid S = (T \wedge)$, and the singular values of S^2 are $\{0, \|T\|^2, \|T\|^2\}$. Hence if $Q \in E$, it has two equal singular values and a zero singular value.

(\Leftarrow) let $Q = U\Sigma_0V^T$ for some orthonormal U, V and for some λ . Let furthermore $R_Z(\frac{\pi}{2})$ be a rotation of $\frac{\pi}{2}$ about the Z axis, then

$$Q = U\Sigma_0V^T = UR_Z(\frac{\pi}{2})^T V^T V R_Z(\frac{\pi}{2}) \Sigma_0 V^T.$$

Now call $R \doteq UR_Z(\frac{\pi}{2})^T V^T$ and $S \doteq VR_Z(\frac{\pi}{2}) \Sigma_0 V^T$; it is immediate to see that $RR^T = R^T R = I_3$ and $S^T = -S$, hence the claim. Q.E.D.