Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

Motion Estimation via Dynamic Vision

Stefano Soatto† *Pietro Perona*†‡ *Ruggero Frezza*‡ *Giorgio Picci*‡

† California Institute of Technology 116-81, Pasadena–CA 91125 · soatto@systems.caltech.edu
‡ Università di Padova, Dipartimento di Elettronica ed Informatica, 35131 Padova–Italy

Abstract

Estimating the three dimensional motion of an object from a sequence of projections is of paramount importance in a variety of applications in control and robotics, such as autonomous navigation, manipulation, servo, tracking, docking, planning, surveillance. Although "visual motion estimation" is an old problem (the first formulations date back to the beginning of the century), only recently tools from control and estimation theory have hinted at acceptable solutions. Moreover, the problem raises a number of issues of system theoretic interest, such as nonlinear estimation and identification on topological manifolds and observability in a projective geometric framework. In this paper we analyze a formulation of the visual motion estimation problem in terms of identification of nonlinear implicit systems with parameters on the so-called "essential manifold"; the estimation is performed either in the local coordinates or in the embedding space of the parameter manifold.

1. Introduction

"Understanding" the geometry and kinematics of the environment is a basic requirement for humans to successfully accomplish tasks such as walking, driving, recognizing and grasping objects. Although the first formulations of the visual motion problem date back to the beginning of the century [16, 33], only within recent years tools from control and estimation theory have been applied [1, 3, 7, 13, 15, 17, 23, 26, 29, 31], with rather encouraging results in traditionally difficult applications, such as autonomous vehicle navigation [8, 9, 10, 11], visual based tracking and servo [5, 12, 19, 20], visual based manipulation [2, 12, 19], docking [10, 18], visual-based planning [6], active sensing [32]. "Vision in the loop" raises new and interesting problems of system theoretic flavor, ranging from distributed filtering and processing of large amounts of sensory data to the analysis and control of new classes of dynamical systems. Crucial issues in the use of vision as a sensor in control systems are, for example, nonlinear observability and identifiability in a projective geometric framework, and estimation and control on topological manifolds. In the last decade a variety of schemes has been proposed for reconstructing recursively structure for known motion [23], motion for known structure [3, 13, 14] or both structure and motion [1, 17, 26, 29, 31]. In this paper we highlight some limitations of the model employed and study a new formalization in terms of identification of a nonlinear implicit model.

2. The "essential space" representation of rigid motion

A rigid motion may be represented as a point in the Lie group $SE(3)$, which can be embedded in the linear space $\mathcal{GL}(4)$ (and hence exploit the matrix product as composition rule) and is locally diffeomorphic to $\mathbf{R}^6$ via the exponential coordinates. We now discuss an alternative "compact" matrix representation of rigid motion, which can be embedded in a linear space of smaller dimensions. Such a representation is derived from the so-called "essential matrices" introduced by Longuet-Higgins [21].

Consider a point $(T, R) \in SE(3)$, then $T\wedge \in so(3)$ is a skew-symmetric matrix. Now define the space of *"essential matrices"* as the subset of $\mathbf{R}^{3\times3}$

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

Despite the loss of the group structure, the essential space has many interesting geometrical properties: it is an algebraic variety in $\mathbf{RP}^8$ [24] and, if slightly modified, shows the structure of a topological manifold [27] that proves crucial in the solution of the visual motion estimation problem. Note that E may also be identified with $TSO(3)$, the tangent bundle of the Lie group of rotation matrices [30]. This proves that the essential space is indeed a differentiable manifold of dimension 6. The following theorem, due to Huang and Faugeras and reported by Maybank [24], gives a simple characterizing property of the essential space.

Theorem 2.1 *(Huang and Faugeras, 1989)*
Let $\mathbf{Q} = \mathbf{U}\Sigma\mathbf{V}^T$ *be the Singular Value Decomposition (SVD) of a matrix in* $\mathbf{R}^{3\times3}$. *Then* $\mathbf{Q} \in E \Leftrightarrow \Sigma = \Sigma_0 = diag\{\lambda\,\lambda\,0\} \mid \lambda \in \mathbf{R}^+$.

2.1. Local coordinates of the essential space

Consider the map $\Phi : E \rightarrow \mathbf{R}^3 \times SO(3)$ defined by

$$\mathbf{Q} \mapsto \begin{bmatrix} T \\ e^{\Omega\wedge} \end{bmatrix} = \begin{bmatrix} \pm\|T\|\mathbf{V}_{.3} \\ \mathbf{U}R_Z(\pm\frac{\pi}{2})\mathbf{V}^T \end{bmatrix} \quad (2)$$

where $\mathbf{U}, \mathbf{V}$ are defined by the Singular Value Decomposition (SVD) 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. Note that the map Φ defines the local coordinates of the essential space modulo two signs, therefore the map Φ associates to each element of the essential space four distinct points in local coordinates. This ambiguity may be resolved in the context of the visual motion estimation problem by imposing the *"positive depth constraint"*, which means that each visible point lies in front of the viewer. In a case like this we will be able to identify a unique local coordinates homeomorphism and, therefore, specify a local coordinate chart for the essential manifold. As a consequence to theorem 2.1, the basis components of the subspaces $< \mathbf{V}_{.1}, \mathbf{V}_{.2} >$ and $< \mathbf{U}_{.1}, \mathbf{U}_{.2} >$ have one degree of freedom and therefore are allowed to switch order. This happens, however, without affecting continuity of T and Ω. The inverse map is simply $\Phi^{-1}(T, \Omega) = e^{(\Omega\wedge)}(T\wedge)$.

2.2. Projection onto the essential space

Theorem 2.1 suggests a simple "projection" of a generic 3×3 matrix onto the essential space: let us define $pr_{<E>} : \mathbf{R}^{3\times 3} \rightarrow E$ via

$$M \mapsto pr_{<E>}(M) = \mathbf{U}\,\mathrm{diag}\{\lambda, \lambda, 0\}\,\mathbf{V}^T \quad (3)$$

where $\mathbf{U}, \mathbf{V}$ are defined by the SVD of $M = \mathbf{U}\,\mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3\}\,\mathbf{V}^T$, and $\lambda \doteq \frac{\sigma_1+\sigma_2}{2}$. It follows from the properties of the SVD that $pr_{<E>}(M)$ minimizes the Frobenius distance from M to the essential space [24].

3. Motion estimation as recursive identification of a nonlinear implicit model

Consider a rigid object, described by the position of a set of feature points in 3D space. We call $\mathbf{X} = [\,X \;\; Y \;\; Z\,]^T \in \mathbf{R}^3$ the coordinates of a generic point with respect to an orthonormal reference frame centered in the pupil of the viewer, with Z along the optical axis and X, Y parallel to the image plane and arranged as to form a right-handed frame. The rigid motion of the object between two time instants relative to the camera is described as a point $(T, R) \in SE(3)$ which acts on $\mathbf{R}^3$ via[1]

$$\mathbf{X}(t+1) = R(t)\left(\mathbf{X}(t) - T(t)\right).$$

What we are able to measure is the **perspective projection** of the point features onto the image plane, which for

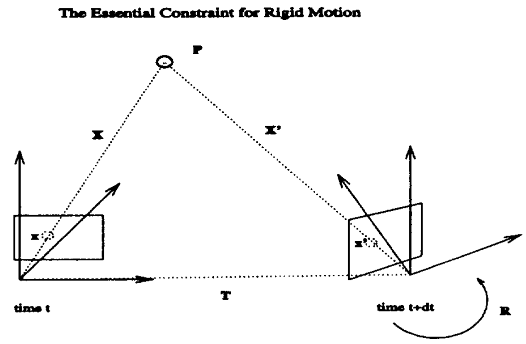

Figure 1: *The coplanarity constraint*

simplicity we represent as the real projective plane. The projection map associates to each $\mathbf{X} \neq 0$ its projective coordinates $\mathbf{x} = [\,\frac{X}{Z} \;\; \frac{Y}{Z} \;\; 1\,]^T \doteq [x\; y\; 1]^T$ as an element of $\mathbf{RP}^2$. We usually measure $\mathbf{x}$ up to some error:

$$\mathbf{y} = \mathbf{x} + n \qquad n \in \mathcal{N}(0, R_n).$$

When a rigid object is moving between two time instants, the coordinates $\mathbf{X}$ of a point at time t, their correspondent $\mathbf{X}'$ at time $t+1$, and the translation vector T are coplanar (fig. 1). Their triple product is therefore zero. This is true of course also for $\mathbf{x}$, $\mathbf{x}'$ and T. When expressed with respect to a common reference frame, for example that at time t, we may write the triple product as

$$\mathbf{x}'^T_i R(T \wedge \mathbf{x}_i) = 0 \qquad \forall\, i = 1 : N. \quad (4)$$

Let us define $\mathbf{Q} \doteq R(T\wedge)$, so that the above constraint, which we now call the "essential constraint", becomes

$$\mathbf{x}'^T_i \mathbf{Q}\mathbf{x}_i = 0 \qquad \forall i = 1 \ldots N. \quad (5)$$

Estimating motion corresponds to identifying the model

$$\begin{cases} (\mathbf{Q}\mathbf{x}_i)^T\mathbf{x}'_i = 0 & \mathbf{Q} \in E \\ \mathbf{y}_i = \mathbf{x}_i + n_i & \forall i = 1 \ldots N, \;\; n_i \in \mathcal{N}(0, R_{n_i}) \end{cases}$$

which is in the form of an Exterior Differential System [4]. Since the constraint (5) is linear in $\mathbf{Q}$, we use the improper notation (after dropping the point index i)

$$\chi(t)\mathbf{Q}(t) = \chi(\mathbf{x}'(t), \mathbf{x}(t))\mathbf{Q}(t) = 0$$

where χ is an $N \times 9$ matrix combining $\mathbf{x}_i, \mathbf{x}'_i$, and $\mathbf{Q}$ is interpreted as a nine-dimensional vector. The generic row of χ has the form $[xx',\; yx',\; x',\; xy',\; yy',\; y',\; x,\; y,\; 1\,]$.

3.1. Choosing the local coordinates for the essential space

The map Φ introduced in eq. (2) defines the local coordinates of the essential space modulo a sign in the direction

[1] We have chosen this convention in favor to the usual $\mathbf{X}(t+1) = R(t)\mathbf{X}(t) + T(t)$ for compatibility with [21].

of translation and in the rotation angle of R_Z, therefore the map Φ associates to each element of the essential space 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 viewer [21, 22]. Consider one of the four local counterparts of $\mathbf{Q} \in E$, and the function $d_{\mathbf{x},\mathbf{x}'} : E \rightarrow \mathbf{R}^{1+1}$ defined by

$$d_{\mathbf{x},\mathbf{x}'}(\mathbf{Q}) = [Z, Z']^T \tag{6}$$

with $Z = \frac{<\mathbf{n}^i, \mathbf{m}^i>}{\|\mathbf{m}^i\|^2} \quad \forall i = 1 \ldots N, \quad \mathbf{m}^i = (R\mathbf{x}^i) \wedge \mathbf{x}'^i$ and $\mathbf{n}^i = (RT) \wedge \mathbf{x}'^i$, which gives the depth of each point as a function of the projection and the motion parameters. Note that it is locally smooth away from zero translation. Therefore, we may use Φ as a local coordinate chart for the following set, which we call the *"normalized essential manifold"* $\mathbf{E} \doteq E \cap d^{-1}_{\mathbf{x},\mathbf{x}'}(\mathbf{R}^2_+)^N$ or equivalently

$$\{\mathbf{Q} = RS | R \in SO(3), S \in so(3), d_{\mathbf{x}^i,\mathbf{x}'^i}(\mathbf{Q}) > 0 \; \forall i\}$$

where $\mathbf{R}_+$ is the positive open half space of $\mathbf{R}$, and $d^{-1}_{\mathbf{x},\mathbf{x}'}$ denotes the preimage of $d_{\mathbf{x},\mathbf{x}'}$. In fact, consider Φ restricted to $\mathbf{E}$. It follows from the properties of the SVD that Φ is continuous and, furthermore, bijective.

3.2. The "essential filter"

Since the essential constraint is an homogeneous algebraic equation, and hence defined only up to a scale factor, it is customary to set the norm of translation to be unitary; this can be done as long as translation is not zero. It can be shown that there is no loss of generality in this assumption [29]. In fact, due to the noise in the measurements, there will always be a translation compatible with the observations (in least-squares sense). The scheme automatically scales translation and inverse depth. We assume therefore $\|\mathbf{Q}\|_2 = \|T\| = 1$. At each time instant we have a set of N constraints in the form

$$\chi(\mathbf{x}'(t), \mathbf{x}(t))\mathbf{Q}(t) = 0,$$

therefore, $\mathbf{Q}$ lies at the intersection between the essential manifold and the linear variety $\chi^{-1}(0)$.

As time progresses, the point $\mathbf{Q}(t)$, corresponding to the actual motion, describes a trajectory on $\mathbf{E}$ (and a corresponding one in local coordinates) according to

$$\mathbf{Q}(t+1) \doteq \mathbf{Q}(t) + n_{\mathbf{Q}}(t).$$

The last equation is indeed just a *definition* of the right-hand side, as we do not know $n_{\mathbf{Q}}(t)$. For now, we will consider the previous equation to be a discrete time dynamical model for $\mathbf{Q}$ on the essential manifold, with $n_{\mathbf{Q}}$ acting as *unknown* input. If we accompany it with the essential constraint, we get

$$\begin{cases} \mathbf{Q}(t+1) = \mathbf{Q}(t) + n_{\mathbf{Q}}(t) & \mathbf{Q} \in E \\ 0 = \chi(\mathbf{y}'(t), \mathbf{y}(t))\mathbf{Q}(t) & \\ \mathbf{y}_i = \mathbf{x}_i + n_i & \forall i = 1 \ldots N. \end{cases} \tag{7}$$

Now the visual motion problem is defined as the estimation of the state of the above model, which is defined on the essential manifold. It can be seen that the system is "linear" (both the state equation and the essential constraint are linear in $\mathbf{Q}$). $\mathbf{E}$, however, is not a linear space. We will see how to solve the estimation task in section 4. The observability/identifiability of the essential models is addressed in [27]. It is proven that the model is globally observable under general position conditions. Such conditions are satisfied if the viewer's path and the visible objects cannot be embedded in a (proper) quadric surface of $\mathbf{R}^3$.

4. Solving the estimation task

At this point we are ready to address the problem of recursively estimating motion from an image sequence. There are two approaches that may be derived naturally from the above formulation.

The first approach we describe consists of composing the equations (7) with the local coordinate chart Φ, ending up with a *nonlinear* dynamical model for motion in $\mathbf{R}^5$. At this point we have to make some assumptions about motion: since we do not have any dynamical model, we will assume a statistical model. In particular, we will assume that motion is a *first order random walk in* $\mathbf{R}^5$ (see fig. 2 top). The problem then is to estimate the state of a nonlinear system driven by white, zero-mean Gaussian noise (see fig. 2 bottom).

In the second approach we change the model for motion: in particular we assume motion to be a *first order random walk in* $\mathbf{R}^9$ *projected onto the essential manifold* (see fig. 2 top). We will see that this leads to a method for estimating motion via solving at each step a *linear estimation* problem in the linear embedding space and then "projecting" the estimate onto the essential manifold (see fig. 2 bottom).

It is very important to understand that these are modeling assumptions about motion which can be validated only a posteriori.

4.1. Estimation in local coordinates

Consider transferring the system (7) into local coordinates $\xi = [T \; \Omega]^T$ where $[T, R] \doteq \Phi(\mathbf{Q})$, $\Omega = Log(R)$ [25] and T is expressed in spherical coordinates of radius one. Then the system in local coordinates becomes

$$\begin{cases} \xi(t+1) = \xi(t) + n_\xi(t) \; ; \; \xi(t_0) = \xi_0 \\ 0 = \chi(\mathbf{y}(t), \mathbf{y}'(t))\mathbf{Q}(\xi(t)) + \tilde{n}(t). \end{cases} \tag{8}$$

We said we modeled motion as a first order random walk: $n_\xi(t) \in \mathcal{N}(0, R_\xi)$ for some R_ξ which is referred to as the variance of the model error. While the above assumption is somewhat arbitrary and can be validated only a posteriori, it is often safe to assume that the noise in the measurements $\mathbf{y}(t), \mathbf{y}'(t)$ are white zero-mean Gaussian processes

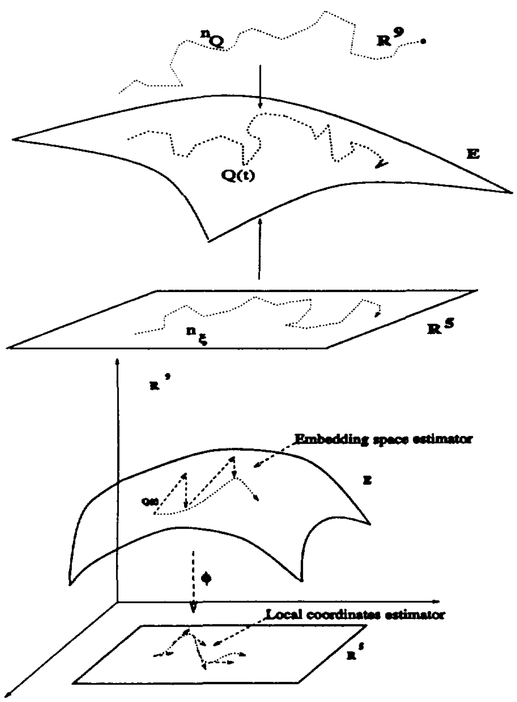

Figure 2: *(Top) Model of motion as a random walk in* $\mathbf{R}^5$ *lifted to the manifold or as a random walk in* $\mathbf{R}^9$ *projected onto the manifold. (Bottom) Estimation on the Essential Space.*

with variance R_n. The estimation scheme for the model above, which takes into account the correlation of the error $\tilde{n}$, is derived in [28]. A simplified version is obtained by approximating $\tilde{n}$ with a white process (note that $\tilde{n}$ is correlated only within one time step). The resulting scheme is based upon an Implicit Extended Kalman Filter (IEKF). We summarize here the equations of the estimator. Call $C \doteq \left(\frac{\partial \chi \mathbf{Q}}{\partial \xi}\right)$ and $D \doteq \left(\frac{\partial \chi \mathbf{Q}}{\partial \mathbf{X}}\right)$, then

Prediction step:

$$\begin{aligned} \hat{\xi}(t+1|t) &= \hat{\xi}(t|t)\,;\ \hat{\xi}(0|0) = \xi_0 & (9)\\ P(t+1|t) &= P(t|t) + R_\xi\,;\ P(0|0) = P_0 & (10) \end{aligned}$$

Update step:

$$\begin{aligned} \hat{\xi}(t+1|t+1) &= \hat{\xi}(t+1|t) - \\ &\quad -L(t+1)\chi(t+1)\mathbf{Q}(\hat{\xi}(t+1|t)) & (11)\\ P(t+1|t+1) &= \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + \\ &\quad +L(t+1)R_{\tilde{n}}(t+1)L^T(t+1) & (12) \end{aligned}$$

Gain:

$$\begin{aligned} L(t+1) &= P(t+1|t)C^T(t+1)\Lambda^{-1}(t+1) & (13)\\ \Lambda(t+1) &= C(t+1)P(t+1|t)C^T(t+1) + \\ &\quad + R_{\tilde{n}}(t+1) & (14)\\ \Gamma(t+1) &= I - L(t+1)C(t+1) & (15)\\ R_{\tilde{n}}(t+1) &= D(t+1)R_n D^T(t+1) & (16) \end{aligned}$$

4.2. Estimation in the embedding space

Suppose that motion, instead of being a random walk in $\mathbf{R}^5$, is represented in the essential manifold as the "projection" of a random walk through $\mathbf{R}^9$ (see fig. 2 top).

We define the operator $\oplus$ that takes two elements in $\mathbf{R}^{3\times3}$, sums them and then projects the result onto the essential manifold:

$$\begin{aligned} \oplus : \mathbf{R}^{3\times3} \times \mathbf{R}^{3\times3} &\rightarrow E \\ M_1, M_2 &\mapsto \mathbf{Q} = pr_{<E>}(M_1 + M_2), \end{aligned}$$

where the symbol "+" is the usual sum in $\mathbf{R}^{3\times3}$. With the above definitions our model for motion simply becomes

$$\mathbf{Q}(t+1) = \mathbf{Q}(t) \oplus n_{\mathbf{Q}}(t), \qquad (17)$$

where $n_{\mathbf{Q}}(t) \in \mathcal{N}(0, R_{n_{\mathbf{Q}}})$ is a white zero-mean Gaussian noise in $\mathbf{R}^9$. If we couple the above equation with (7) we again have a dynamical model on a Euclidean space (in our case $\mathbf{R}^9$) driven by white noise. The Essential Estimator is the least variance filter built for the above model, and corresponds to a linear Kalman filter update in the embedding space, followed by a projection onto the essential manifold. In principle, the gain could be precomputed offline for each possible configuration of motion and feature positions.

Prediction step:

$$\begin{aligned} \hat{\mathbf{Q}}(t+1|t) &= \hat{\mathbf{Q}}(t|t)\,;\ \hat{\mathbf{Q}}(0|0) = \mathbf{Q}_0 & (18)\\ P(t+1|t) &= P(t|t) + R_{\mathbf{Q}}\,;\ P(0|0) = P_0 & (19) \end{aligned}$$

Update step:

$$\begin{aligned} \hat{\mathbf{Q}}(t+1|t+1) &= \hat{\mathbf{Q}}(t+1|t) \oplus \\ &\quad \oplus L(t+1)\chi(t)\hat{\mathbf{Q}}(t+1|t) & (20)\\ P(t+1|t+1) &= \Gamma(t+1)P(t+1|t)\Gamma^T(t+1) + \\ &\quad +L(t+1)R_{\tilde{n}}(t+1)L^T(t+1) & (21) \end{aligned}$$

Gain:

$$\begin{aligned} L(t+1) &= -P(t+1|t)\chi(t)\Lambda^{-1}(t+1) & (22)\\ \Lambda(t+1) &= \chi(t)P(t+1|t)\chi^T(t) + \\ &\quad + R_{\tilde{n}}(t+1) & (23)\\ \Gamma(t+1) &= I - L(t+1)\chi(t) & (24)\\ R_{\tilde{n}}(t+1) &= D(t+1)R_n D^T(t+1) & (25) \end{aligned}$$

5. Experiments

We have tested the described algorithms on a variety of motion and structure configurations for real and synthetic image sequences with variable numbers of feature points. The interested reader may find a complete set of experiments in [28]. For reasons of space we report here a comparison between the two schemes on a simulation experiment, which consists of views of a cloud of points under a discontinuous motion with singular regions.

A cloud of 20 points is distributed uniformly in a square of side 1m whose centroid is placed 1.5m ahead of the viewer. These points are projected onto an image plane of 512×512 pixels, and realize a field of view of approximately 45^o. Gaussian noise with 1 pixel Std has been added to the measurements. The cloud undergoes piecewise constant-velocity motion, with a central region of pure rotation about the optical center. In the following table we report the steady-state error when the filters have reached convergence. The filter on the embedding space requires half floating-point operations per iteration than the filter in local coordinates (40 Kflops/iteration); however, it takes up to 4 times more steps to converge (40 steps). Therefore we have compared the two filters on two different windows (10 for the the filter in local coordinates, 20 for the filter in the embedding space) when they have reached steady state in estimating a constant velocity motion.

Scheme	T_X	T_Y
Local	M: .0002 Std:.0004	M:-.0015 Std: .0048
Embed	M: 3.9754E-5 Std: .0001	M: .0017 Std: .0013

Scheme	T_Z	Ω_Z
Local	M: .0002 Std: .0004	M:-.0002 Std:.0008
Embed	M: .0002 Std: .0001	M: -1.611E-5 Std: .0004

Scheme	Ω_X	Ω_Y
Local	M:.0008 Std: .0022	M:.0002 Std:.0002
Embed	M:-.0008 Std: .0004	M: 3.9949E-6 Std: .0002

6. Conclusions

The problem of estimating three dimensional motion from a sequence of images can be naturally set in the framework of dynamic estimation and identification. Under the assumption of a static scene, the rigid motion constraint and the perspective projection map define a nonlinear dynamical model, and estimating motion is equivalent to observing its state. It has been proven that such a model is not locally weakly observable, unless metric constraints are imposed on the state manifolds [27].

We have analyzed a new formulation based upon the representation of motion through the "essential matrices", introduced by Longuet-Higgins [21]. Motion estimation is equivalent to the identification of a nonlinear implicit model with parameters on a topological manifold. We have studied an algorithm which solves the identification task by estimating the state of a model defined on the parameter manifold. The estimation is performed either in the local coordinates or in the embedding space of the parameter manifold.

Acknowledgements

We wish to thank Prof. J.K. Åström for his discussions on implicit Kalman filtering, Prof. Richard Murray and Prof. Shankar Sastry for their observations and useful suggestions. Also discussions with Michiel van Nieuwstadt and Andrea Mennucci were helpful. This research has been funded by the California Institute of Technology, a scholarship from the University of Padova, a fellowship from the "A. Gini" Foundation, an AT&T Foundation Special Purpose grant, ONR grant N0014-93-1-0990, grant ASI-RS-103 from the Italian Space Agency and the National Young Investigator Award (P.P.). A long version of this paper has been submitted to the IEEE Transactions on Automatic Control.

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] A. Blake, M. Taylor, and A. Cox. Grasping visual symmetry. *Proc. of the ICCV*, 1993.

[3] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE trans. PAMI*, Jan. 1986.

[4] Bryant, Chern, Goldberg, and Goldsmith. *Exterior Differential Systems*. Mathematical Research Institute. Springer Verlag, 1992.

[5] F. Chaumette and A. Santos. Tracking a moving object by visual servoing. *Proc. of the 12th IFAC World Congr. Vol. 9 pp. 409–414*, 1993.

[6] R. Curwen, A. Blake, and A. Zisserman. Real-time visual tracking for surveillance and path planning. *Proc. of the ECCV*, 1992.

[7] W. Dayawansa, B. Ghosh, C. Martin, and X. Wang. A necessary and sufficient condition for the perspective observability problem. –, 1993.

[8] E. D. Dickmanns and Th. Christians. Relative 3d-state estimation for autonomous visual guidance of road vehicles. In *Intelligent autonomous system 2 (IAS-2)*, Amsterdam, 11-14 December 1989.

[9] E. D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241–261, 1988.

[10] E. D. Dickmanns and V. Graefe. Applications of dynamic monocular machine vision. *Machine Vision and Applications*, 1:241–261, 1988.

[11] E. D. Dickmanns and V. Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1:223–240, 1988.

[12] K. Hashimoto T. Kimoto T. Ebine and H. Kimura. Image-based dynamic visual servo for a hand-eye manipulator. *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.

[13] D.B. Gennery. Tracking known 3-dimensional object. In *Proc. AAAI 2nd Natl. Conf. Artif. Intell.*, pages 13–17, Pittsburg, PA, 1982.

[14] D.B. Gennery. Visual tracking of known 3-dimensional object. *Int. J. of Computer Vision*, 7(3):243–270, 1992.

[15] B. Ghosh, M Jankovic, and Y. Wu. Perspective problems in systems theory and its application in machine vision. *Journal of Math. Systems, Est. and Control*, 1994.

[16] E.J. Gibson, J. J. Gibson, O. W. Smith, and H. Flock. Motion parallax as a determinant of perceived depth. *J. Exp. Psych. Vol.45*, 1959.

[17] J. Heel. Direct estimation of structure and motion from multiple frames. *AI Memo 1190, MIT AI Lab*, March 1990.

[18] C.C. Ho and N.H. McClamrock. Autonomous spacecraft docking using a computer vision systm. *Proc. of the 31st CDC – Tucson, AZ*, 1992.

[19] K. Hashimoto T. Kimoto T. Ebine H. Kimura. Manipulator control with image-based visual servo. *IEEE intl' Conference on Robotics and Automation, pages 2267–2272*, 1991.

[20] Ming Lei and Bijoy K. Ghosh. A new nonlinear feedback controller for visually-guided robotic motion tracking. *Proc. of the ECC*, 1993.

[21] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[22] H.C. Longuet-Higgins. Configurations that defeat the eight-point algorithm. *Mental processes: studies in cognitive science, MIT press*, 1987.

[23] L. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.

[24] S. Maybank. *Theory of reconstruction from image motion*. Springer Verlag, 1992.

[25] R.M. Murray, Z. Li, and S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

[26] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.

[27] S. Soatto. Observability/identifiability of rigid motion under perspective. *Submitted to Automatica. Registered as technical Report CIT-CDS 94-001, California Institute of Technology. Reduced version in the proceeding of the 33rd Conf. on Decision and Control. Also available through the Worldwide Web Mosaic* (`http://avalon.caltech.edu/cds/techreports/`), 1994.

[28] S. Soatto, R. Frezza, and P. Perona. Motion estimation via dynamic vision. *Submitted to the IEEE Trans. on Automatic Control. Registered as Technical Report CIT-CDS-94-004, California Institute of Technology. Available through the Worldwide Web Mosaic* (`http://avalon.caltech.edu/cds/techreports/`), 1994.

[29] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *In "Computer Vision ECCV 94, Lecture Notes in Computer Sciences vol. 801", Springer Verlag*, May 1994.

[30] S. Soatto and P. Perona. Structure-independent visual motion control on the essential manifold. *Technical Report CIT-CDS-94-013, California Institute of Technology. Short version in Proc. of the IFAC Symposium on Robot Control, Capri–Italy*, 1994.

[31] 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.

[32] M. Swain and M. Stricker (editors). Promising directions in active vision. Technical Report T.R. CS 91-27, University of Chicago, November 1991. Written by the attendees of the NSF Active Vision Workshop – August 5-7 1991.

[33] Hermann von Helmholtz. *Treatise on Physiological Optics*. 1910.