

# POSE ESTIMATION OF AUTONOMOUS VEHICLES USING VISUAL INFORMATION: A MINIMUM-ENERGY ESTIMATOR APPROACH<sup>\*</sup>

António Pedro Aguiar      João Pedro Hespanha

*Department of Electrical and Computer Engineering,  
Univ. of California, Santa Barbara, CA 93106-9560  
{aguiar,hespanha}@ece.ucsb.edu*

**Abstract:** This paper addresses the pose estimation problem for autonomous vehicles that use a monocular charged-coupled-device (CCD) camera mounted on-board that observes the apparent motion of stationary points. We formulate the problem in the framework of state estimation of a state-affine system with multiple perspective outputs. Resorting to dynamic programming, we derive a minimum-energy estimator which produces an estimate of the state that is “most compatible” with the dynamics, in the sense that it requires the least amount of noise energy to explain the measured outputs. In our formulation we take directly into account that the measurements arrive at discrete-time instants, are time-delayed, and may not be complete. In this way, we can deal with usual problems in vision systems such as noise as well as latency and intermittency of observations. The convergence of the proposed observer system is analyzed and simulations results are presented and discussed.

**Keywords:** Visual servo control; Observers for nonlinear systems; Estimation; Robotics

## 1. INTRODUCTION

A fundamental problem in mobile robotics is the determination of the position and orientation with respect to an inertial coordinate system. A promising solution to this problem is to utilize a camera mounted on the robot that observes the apparent motion on the image of stationary points. The linear and angular velocities of the camera can be assumed known in its own coordinate system (possibly with errors due to noise) but not in the inertial coordinate system. This is quite reasonable in mobile robotics where the

motion of the camera is determined by the applied control signals. The problem of estimating the position and orientation of a camera mounted on a rigid body from the apparent motion of point features has a long tradition in the computer vision literature (cf., e.g., [Matthies, Kanade and Szeliski, 1989; Jankovic and Ghosh, 1995; Soatto, Frezza and Perona, 1996; Kaminer, Pascoal, Kang and Yakimenko, 2001; Chiuso, Favaro, Jin and Soatto, 2002; Rehlinger and Ghosh, 2003] and references therein). Interesting algorithms are the ones that are filtering-like or iterative that continuously improve the estimates as more data (i.e., images) are acquired and that are robust with respect to measurement noise. Soatto et al. [1996] formulates the visual motion estimation problem in terms of identification of nonlinear implicit systems with parameters on a topological man-

---

<sup>\*</sup> This material is based upon work supported by the National Science Foundation under Grant No. ECS-0093762. The work of A. Pedro Aguiar was supported by a Pos-Doc Fellowship PRAXIS XXI from the Portuguese Foundation of Science and Technology.

ifold and propose a dynamic solution either in the local coordinates or in the embedding space of the parameter manifold. In [Rehbinder and Ghosh, 2003], rigid-body pose estimation using inertial sensors and a monocular camera is considered. A local convergent observer where the states evolve on  $SO(3)$  (the rotation estimation is decoupled from the position estimation) is proposed. In the area of wheeled mobile robots, Ma, Kosecka and Sastry [1999] address the problem of tracking an arbitrarily shaped continuous ground curve by formulating it as controlling the shape of the curve in the image plane. Observability of the curve dynamics is studied and an extended Kalman filter is proposed to dynamically estimate the image quantities needed for feedback control from the actual noisy images. An application for landing an unmanned air vehicle using vision in the control loop is described in [Shakernia, Ma, Koo and Sastry, 1999]. In [Kaminer et al., 2001], based on measurements provided by airborne vision and inertial sensors, the problem of navigation system design for autonomous aircraft landing is addressed. The authors cast the problem in a linear parametrically varying framework and solve it using tools that borrows from the theory of linear matrix inequalities. These results are extended in [Hespanha, Yakimenko, Kaminer and Pascoal, 2001] to deal with the so-called out-of-frame events.

Inspired by the (single output) perspective systems introduced by Ghosh, Jankovic and Wu [1994], in Section 2 we formulate the problem of estimating the position and orientation of a controlled rigid body using measurements from an on-board charged-coupled-device CCD camera as a state-estimation problem of a perspective system. The reader is referred to [Ghosh et al., 1994; Ghosh and Loucks, 1995; Takahashi and Ghosh, 2001] for several other examples of perspective systems in the context of motion and shape estimation.

We propose a minimum-energy estimator that produces an estimate for the state of the perspective system that is “most compatible” with the system’s dynamics and measured outputs. In particular, the optimal state estimate  $\hat{x}$  at time  $t$  is defined to be the value for the state that is compatible with the observations collected up to time  $t$  and the dynamics of the system for the “smallest” possible measurement noise and disturbances, with “smallest” understood in an integral-square sense. This type of estimators were first proposed by Mortensen [1968] and further refined by Hijab [1980]. Game theoretical versions of these estimators were proposed by McEneaney [1998]. It was recently shown by Krener [2003] that this type of estimators is globally convergent when the system is observable for *every* input. In [Hespanha, 2002], it was shown that for pro-

jective systems with multiple inputs, convergence can be obtained under less restrictive observability assumptions. In [Aguiar and Hespanha, 2003a], we improve upon the results in [Hespanha, 2002] by incorporating quadratic state-constraints in the minimum-energy formulation. State constraints allow one to take into account that some elements of the state must lie in given manifolds. In the context of rigid body motion, typically part of the state is a rotation matrix that is known to lie in  $SO(3)$ . This can be expressed by quadratic constraints. In [Aguiar and Hespanha, 2003a], the state estimates were used to close the loop and control a mobile robot to a desired position, defined with respect to visual landmarks. State-constraints for systems with perspective outputs also appeared in [Takahashi and Ghosh, 2001] in the context of motion estimation using a CCD camera and a laser range finder, where the measurements from camera and range finder were related by an algebraic constraint.

One of the main novelties of this paper is that we explicitly address the fact that the measurements arrive at discrete-time instants, are time-delayed, and may not be complete. In this way, we can deal with usual problems associated to vision systems such as noise, latency, and intermittency. Resorting to dynamic programming, in Section 3 we derive a minimum-energy estimator. The state-estimator has the desired property that, under suitable observability assumptions, the state-estimate converges asymptotically to the true value of the state in the absence of noise and disturbance. In the presence of noise, the estimate remains bounded away from the true value of the state (cf. Section 4). We can therefore use this state-estimator to design output-feedback controllers by using the estimated state to drive state-feedback controllers. In Section 5 we illustrate the performance of the proposed estimator through computer simulation.

Due to space limitations, all the proofs are omitted. These can be found in [Aguiar and Hespanha, 2003b].

## 2. PROBLEM STATEMENT

Consider a continuous-time system described by

$$\dot{x} = A(u)x + b(u) + G(u)\mathbf{d}, \quad (1)$$

$$\alpha_j y_j = C_j(u)x + d_j(u) + \mathbf{n}_j, \quad (2)$$

$j \in \mathcal{I} := \{1, 2, \dots, N\}$ , where  $x \in \mathbb{R}^n$  denotes the state of the system,  $u \in \mathbb{R}^{n_u}$  its input,  $y_j \in \mathbb{R}^{m_j}$  its  $j$ th perspective output,  $\mathbf{d} \in \mathbb{R}^{n_d}$  an input disturbance that cannot be measured, and  $\mathbf{n}_j \in \mathbb{R}^{m_j}$  measurement noise affecting the  $j$ th output. Each  $\alpha_j \in \mathbb{R}$ ,  $j \in \mathcal{I}$  denotes a scalar that is determined by a normalization constraint such as

$$\|y_j\| = 1 \quad \text{or} \quad v'_j y_j = 1, \quad (3)$$

where the  $v_j \in \mathbb{R}^{m_j}$  denote constant vectors. We call (1)–(2) a *state-affine system with multiple perspective outputs*, or for short simply a *system with perspective outputs*.

As it is shown in Appendix, this system describes the kinematic model of a rigid-body whose outputs are the homogeneous image coordinates of  $N$  fixed points provided by an on-board camera.

Suppose now that we acquire the measurements only at discrete times  $t'_i$ ,  $i = 0, 1, \dots, k$ , with  $t'_0 < t'_1 < \dots < t'_k$ , and we only have access to them after a time-delay  $\tau_i$ . Then, from (2) we have

$$\begin{aligned} \alpha_j(t_i)\mathbf{y}_j(t_i) &= \alpha_j(t'_i)y_j(t'_i) \\ &= C_j(u(t'_i))x(t'_i) + d_j(u(t'_i)) + \mathbf{n}_j(t'_i) \end{aligned}$$

where  $\mathbf{y}$  denotes the discrete time-delay observed variable, and  $t_i = t'_i + \tau_i$ . Furthermore, suppose that the measurements may not be complete, that is,  $j \in \mathcal{I}_i$ ,  $i = 0, 1, \dots, k$ , where  $\mathcal{I}_i \subseteq \mathcal{I}$  is an index set that typically is equal to  $\mathcal{I}$  if we receive all the measurements at the same time  $t_i$ . Then,

$$\begin{aligned} \alpha_j(t_i)\mathbf{y}_j(t_i) &= C_j(u(t_i - \tau_i))x(t_i - \tau_i) \\ &\quad + d_j(u(t_i - \tau_i)) + \mathbf{n}_j(t_i - \tau_i), \end{aligned} \quad (4)$$

$j \in \mathcal{I}_i$ ,  $i = 0, 1, \dots, k$ . The problem under consideration is to design an observer, formulated as a deterministic optimization problem, which estimates the continuous-time state vector  $x(t)$  governed by equation (1), given the discrete time-delay measurements  $\mathbf{y}(t_i)$  expressed by the output equation (4).

Before we formulate the optimization problem, observe first from (1) that  $x(t_i)$  satisfies

$$\begin{aligned} x(t_i) &= \Phi(t_i, t_i - \tau_i)x(t_i - \tau_i) \\ &\quad + \int_{t_i - \tau_i}^{t_i} \Phi(t_i, \sigma) [b(u(\sigma)) + G(u(\sigma))\mathbf{d}(\sigma)] d\sigma, \end{aligned}$$

where  $\Phi(t, t_0)$  is the transition matrix of system (1) satisfying the differential equation  $\dot{\Phi} = A(u)\Phi$ . Therefore,

$$\begin{aligned} x(t_i - \tau_i) &= \Phi^{-1}(t_i, t_i - \tau_i)x(t_i) - \Phi^{-1}(t_i, t_i - \tau_i) \\ &\quad \int_{t_i - \tau_i}^{t_i} \Phi(t_i, \sigma) [b(u(\sigma)) + G(u(\sigma))\mathbf{d}(\sigma)] d\sigma. \end{aligned}$$

Replacing this equation in (4) we get for  $j \in \mathcal{I}_i$ ,  $i = 0, 1, \dots, k$

$$\alpha_j(t_i)\mathbf{y}_j(t_i) = \bar{C}_j(u)x(t_i) + \bar{d}_j(u) + \bar{\mathbf{n}}_j(t_i), \quad (5)$$

where

$$\begin{aligned} \bar{C}_j(u) &:= C_j(u(t_i - \tau_i))\Phi(t_i - \tau_i, t_i), \\ \bar{d}_j(u) &:= -\bar{C}_j(u) \int_{t_i - \tau_i}^{t_i} \Phi(t_i, \sigma)b(u(\sigma)) d\sigma + d_j(u(t_i - \tau_i)), \\ \bar{\mathbf{n}}_j(t_i) &:= -\bar{C}_j(u) \int_{t_i - \tau_i}^{t_i} \Phi(t_i, \sigma)G(u(\sigma))\mathbf{d}(\sigma) d\sigma + \mathbf{n}_j(t_i - \tau_i). \end{aligned}$$

The optimization problem can now be stated as follows:

*Problem 1.* Consider the continuous-time state equation (1) together with the discrete-time perspective output equation (5). Given an input  $u$

defined on an interval  $[0, t)$ , and measured outputs  $\mathbf{y}_j(t_i)$ , with  $i = 0, 1, \dots, k$ ,  $t_0 := 0$ ,  $t_k \leq t$ , and  $k := \arg \min_k (t - t_k)$ , compute the estimate  $\hat{x}(t)$  of the state at time  $t \geq 0$  defined as

$$\hat{x}(t) := \arg \min_{z \in \mathbb{R}^n} J(z, t), \quad (6)$$

where

$$\begin{aligned} J(z; t) &:= \min_{\substack{\mathbf{d}: [0, t), \bar{\mathbf{n}}_j(t_i), \alpha_j(t_i) \\ i=0, 1, \dots, k}} \left\{ (x(0) - \hat{x}_0)' P_0 (x(0) - \hat{x}_0) \right. \\ &\quad + \int_0^t \|\mathbf{d}(\sigma)\|^2 d\sigma + \sum_{i=0}^k \sum_{j \in \mathcal{I}_i} \|\bar{\mathbf{n}}_j(t_i)\|^2 : \\ &\quad x(t) = z, \quad \dot{x} = A(u)x + b(u) + G(u)\mathbf{d}, \\ &\quad \left. \alpha_j(t_i)y_j(t_i) = \bar{C}_j(u)x(t_i) + \bar{d}_j(u) + \bar{\mathbf{n}}_j(t_i) \right\}, \end{aligned} \quad (7)$$

$\mathcal{I}_i \subseteq \mathcal{I}$ ,  $P_0 > 0$  and  $\hat{x}_0$  encodes a-priori information about the state.

The estimate  $\hat{x}(t)$  can be interpreted as the value for which the measured outputs can be made compatible with the system dynamics (1) and (5) for the “smallest” possible noise  $\bar{\mathbf{n}}_j$  and disturbance  $\mathbf{d}$ . Notice also that this formulation includes the case when we do not have all the measurements at the same time  $t_i$ . In that case,  $\mathcal{I}_i$  is a strict subset of  $\mathcal{I}$ .

### 3. THE OBSERVER EQUATIONS

In this section we present the observer equations that were derived by resorting to dynamic programming. In that framework, the function  $J(z; t)$ ,  $z \in \mathbb{R}^n$ ,  $t \geq 0$  is viewed as a cost-to-go. In what follows, given a signal  $x$  with a jump at time  $t$ , we denote by  $x(t^-)$  the limit of  $x(\tau)$  as  $\tau \uparrow t$  from below, *i.e.*,  $x(t^-) := \lim_{\tau \uparrow t} x(\tau)$ . Without loss of generality we consider that  $x$  is continuous from above at every point, *i.e.*,  $x(t) = \lim_{\tau \downarrow t} x(\tau)$ . The following result solves Problem 1.

*Theorem 1.* The estimate  $\hat{x}(t)$  of the state at time  $t \geq 0$  defined by (6) and (7) can be computed as a solution to the *impulse system* described by the initial conditions

$$t_0 = 0, \quad P(t_0) = P_0, \quad \hat{x}(t_0) = \hat{x}_0, \quad (8)$$

the dynamic equations for  $t_i \leq t < t_{i+1}$ ,  $i = 0, 1, \dots, k$

$$\begin{aligned} \dot{P}(t) &= -P(t)A(u) - A(u)'P(t) \\ &\quad - P(t)G(u)G(u)'P(t), \quad P(t_i) = P_i \end{aligned} \quad (9)$$

$$\dot{\hat{x}}(t) = A(u)\hat{x}(t) + b(u), \quad \hat{x}(t_i) = \hat{x}_i \quad (10)$$

and the impulse equations for  $t = t_{i+1}$ ,  $i = 0, 1, \dots, k-1$

$$P(t_{i+1}) = P(t_{i+1}^-) + W(t_{i+1}), \quad (11)$$

$$\begin{aligned} \hat{x}(t_{i+1}) &= \hat{x}(t_{i+1}^-) \\ &\quad - P(t_{i+1})^{-1} [W(t_{i+1})\hat{x}(t_{i+1}^-) + w(t_{i+1})] \end{aligned} \quad (12)$$

where

$$W(t_{i+1}) := \sum_{j \in \mathcal{I}_{i+1}} \bar{C}_j'(u) \left( I - \frac{\mathbf{y}_j(t_{i+1}) \mathbf{y}_j(t_{i+1})'}{\|\mathbf{y}_j(t_{i+1})\|^2} \right) \bar{C}_j(u),$$

$$w(t_{i+1}) := \sum_{j \in \mathcal{I}_{i+1}} \bar{C}_j'(u) \left( I - \frac{\mathbf{y}_j(t_{i+1}) \mathbf{y}_j(t_{i+1})'}{\|\mathbf{y}_j(t_{i+1})\|^2} \right) \bar{d}_j(u).$$

Furthermore, the cost function  $J(z; t)$  defined in (7) can be written as

$$J(z; t) = (z - \hat{x}(t))' P(t) (z - \hat{x}(t)) + c(t), \quad (13)$$

where

$$c(t) = \begin{cases} 0, & t_i \leq t < t_{i+1} \\ c(t_i), & t_i \leq t < t_{i+1} \\ -(P(t^-) \hat{x}(t^-) + \hat{x}(t^-)' P(t^-) x(t^-) + c(t^-)) \\ -w(t)' [P(t^-) + W(t)]^{-T} (P(t^-) \hat{x}(t^-) - w(t)) \\ + \sum_{j \in \mathcal{I}_{i+1}} \bar{d}_j \left( I - \frac{\mathbf{y}_j \mathbf{y}_j'}{\|\mathbf{y}_j\|^2} \right) \bar{d}_j, & t = t_{i+1} \end{cases}$$

for all  $i = 0, 1, \dots, k-1$ .

*Proof.* See [Aguiar and Hespanha, 2003b].  $\square$

*Remark 1.* From equation (10) we can conclude, as it was expected, that since in the time interval  $(t_k, t_{k+1})$  there is no additional information, the dynamic equation of the observer for  $\hat{x}(t)$  is a replica of the state dynamic equation of system (1). Notice also that for  $t = t_{i+1}$ ,  $i = 0, 1, \dots, k-1$ , the state-estimation equation (12) can be rewritten as

$$\hat{x}(t_{i+1}) = \hat{x}(t_{i+1}^-) + P(t_{i+1})^{-1} \sum_{j \in \mathcal{I}_{i+1}} \bar{C}_j'(u) \left( \hat{\alpha}_j \mathbf{y}_j(t_{i+1}) - \bar{C}_j(u) \hat{x}(t_{i+1}^-) - \bar{d}_j \right),$$

$$\hat{\alpha}_j = \frac{\mathbf{y}_j'(t_{i+1}) (\bar{C}_j(u) \hat{x}(t_{i+1}^-) + \bar{d}_j)}{\|\mathbf{y}_j(t_{i+1})\|^2}$$

which emphasizes the parallel between (12) and a Kalman filter for linear systems.

#### 4. ESTIMATOR CONVERGENCE

In this section we investigate under what conditions the state estimate  $\hat{x}$  provided by Theorem 1 converges to the true state  $x$  of the perspective system. The following technical assumptions are needed:

*Assumption 1.* There exist positive constants  $\delta$ ,  $\Delta \in (0, \infty)$  such that  $\delta I \leq G(u)G(u)' \leq \Delta I$ ,  $\forall_u \in \mathbb{R}^{n_u}$

*Assumption 2.* Let  $\text{Num}(t, \sigma)$  denote the number of discontinuities of the observer system in the open interval  $(\sigma, t)$ . There exist finite positive constants  $\tau_D$  and  $N_0$ , for which the following condition holds:

$$\text{Num}(t, \sigma) \leq N_0 + \frac{t - \sigma}{\tau_D}.$$

The constant  $\tau_D$  is called the *average dwell-time* and  $N_0$  the *chatter bound*.

Assumption 1 is a mild assumption and essentially guarantees that  $G(u)$  is bounded and full-row rank, “uniformly” over all possible inputs.

Assumption 2 roughly speaking guarantees that the observer system occasionally can have consecutive discontinuities separated by less than  $\tau_d$ , but the average interval between consecutive discontinuities is no less than  $\tau_D$ . This type of condition typically arises in the context of logic-based switching control (cf., e.g., [Hespanha and Morse, 1999] and references therein).

The following result establishes the convergence of the state estimate.

*Theorem 2.* Assuming that the solution to the process (1)-(2) exists globally, the solution to the impulse state estimator (8)-(12) also exists globally. Moreover, when Assumptions 1-2 hold and  $P^{-1}$  remains uniformly bounded, there exist positive constants  $c$ ,  $r < 1$ ,  $\gamma_d$ ,  $\gamma_1, \dots, \gamma_N$  such that

$$\|\tilde{x}(t_k)\| \leq cr^k \|\tilde{x}(0)\| + \gamma_d \sup_{\tau \in (0, t_k)} \|\mathbf{d}(\tau)\|$$

$$+ \sum_{j=1}^N \gamma_j \sup_{\tau \in (0, t_k)} \|\bar{\mathbf{n}}_j(\tau)\|, \quad t_k > 0$$

where  $\tilde{x}(t) := \hat{x}(t) - x(t)$  denotes the state estimation error.

*Proof.* See [Aguiar and Hespanha, 2003b].  $\square$

Some condition on the observability of (1)-(2) would be expected to achieve convergence of the estimated state  $\hat{x}$  to the process state  $x$ . In Theorem 2 this condition appears in the form of the requirement that  $P^{-1}$  remains bounded. The following result provides a condition under which this happens.

*Lemma 1.* The matrix  $P$  remains lower bounded by a positive value  $\delta$ , i.e.,  $P(t) \geq \delta I > 0$  along trajectories of the system (1), (5), and the state-estimator (8)-(12), provided that there exist positive constants  $\mathbf{N}$ ,  $\epsilon$  such that the following persistence of excitation condition

$$\frac{1}{\mathbf{N}} \sum_{j=0}^{\mathbf{N}} \Phi(t_{i+j}, t_i)' W(t_{i+j}) \Phi(t_{i+j}, t_i) \geq \epsilon I > 0, \quad (14)$$

$i = 0, 1, \dots, k$ , holds, where  $\Phi(t, \tau)$  denotes the state transition matrix of  $\dot{z} = A(u)z$ .

*Proof.* See [Aguiar and Hespanha, 2003b].  $\square$

Combining Theorem 2 and Lemma 1 we obtain the following:

*Corollary 1.* When Assumptions 1 and 2 hold, and there exist constants  $\mathbf{N}$ ,  $\epsilon$  such that the persistence of excitation condition (14) holds, the state-estimate  $\hat{x}$  converges to the state  $x$  in the absence of disturbance input and measurement noise. When the disturbance and noise are bounded but nonzero,  $\hat{x}$  may not converge to  $x$  but remains bounded away from it.

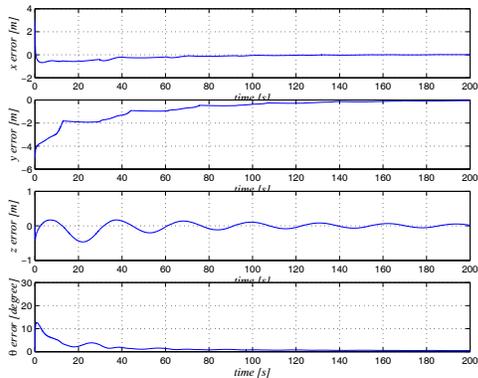


Fig. 1. Time evolution of the estimation errors in position and orientation.

## 5. SIMULATION RESULTS

This section illustrates the performance of the proposed estimator through computer simulation. The objective is to estimate the position and orientation of a unicycle mobile robot that starts at position  $x_0 = -2.0\text{ m}$ ,  $y_0 = -5.0\text{ m}$  with orientation  $\theta_0 = 0\text{ rad}$  and follows a circular path with linear velocity  $v = 0.3\text{ m/s}$  and angular velocity  $\omega = 0.2\text{ rad/s}$ . A pin-hole camera is fixed with the robot with its optical axis aligned with the linear velocity. The vision sampling interval is  $T = 0.4\text{ s}$  and the time-delay is  $\tau = 0.2\text{ s}$ . The estimator was initialized with  $\hat{x}_0 = -5.0\text{ m}$ ,  $\hat{y}_0 = 0\text{ m}$ , and  $\hat{R} = I$ . The measurements given by the camera were corrupted with additive Gaussian noise with standard deviation equal to roughly 5% of the measurements. The visual landmarks are four corners ( $N = 4$ ) of a square that belongs to the plane perpendicular to the inertial  $x$  –  $axis$  and contains the origin. Notice that since the robot is describing a circular trajectory, the visual landmarks will periodically be out of the camera’s field of view (negative depth). While this happens, the estimator does not receive visual measurements.

Fig. 1 displays the time evolution of the estimation errors. It can be seen that the estimated pose tend to a small neighborhood of the true value. Observe also, as expected, that the estimation errors only reduce significantly when the visual landmarks are in the camera’s field of view. This can be checked by comparing Fig. 1 with Fig. 2 which shows the time evolution of the minimum and maximum singular values of  $P$ , respectively, and also indicates through the time evolution of variable  $\sigma$  when the estimator is receiving ( $\sigma = 0$ ) or not ( $\sigma = -1$ ) measurements from the camera.

## 6. CONCLUSIONS

We considered the problem of estimating the position and orientation of autonomous vehicles

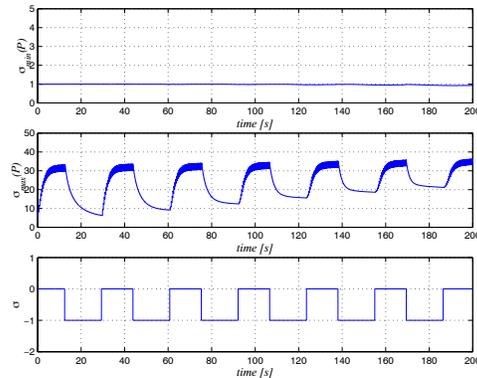


Fig. 2. Time evolution of the minimum and maximum singular values of  $P$ , respectively; and the variable  $\sigma$  which indicates when the estimator is receiving ( $\sigma = 0$ ) or not ( $\sigma = -1$ ) measurements from the camera.

that use an onboard monocular charged-coupled-device (CCD) camera that observes the apparent motion of stationary points. We designed a dynamical impulsive system that produces an estimate of the state that is “most compatible” with the dynamics, in the sense that it requires the least amount of noise energy to explain the measured output. We explicitly take into account that the measurements arrive at discrete-time instants, are time-delayed, and may not be complete. In this way, we can deal with measurement noise, latency, and intermittency. The convergence of the proposed estimator system was analyzed and illustrated through computer simulation. Future work will address experimental validation of these results.

## APPENDIX

Consider a coordinate frame  $\{b\}$  attached to a rigid body that moves with respect to an inertial frame  $\{i\}$ . We denote<sup>1</sup> by  $(p_{ib}, R_{ib}) \in \text{SE}(3)$  the configuration of the frame  $\{b\}$  with respect to  $\{i\}$ , and  $(v_{ib}^b, \Omega_{ib}^b) \in \text{se}(3)$  the twist that defines the velocity of frame  $\{b\}$  with respect to  $\{i\}$ , expressed in the frame  $\{b\}$ . Suppose that a camera attached to the body frame  $\{b\}$  sees  $N$  points  $Q_1, Q_2, \dots, Q_N$ . Denoting by  $y_j \in \mathbb{R}^3$  the homogeneous image coordinates provided by the camera of the point  $Q_j$ , the dynamics of the system can be described by the following system with  $N$  perspective outputs: (see [Hespanha, 2002] for details)

$$\dot{q}_1^b = -\Omega_{ib}^b q_1^b + R_{ib}'^i q_1^i - v_{ib}^b, \quad (15)$$

$$\dot{R}_{ib}'^b = -\Omega_{ib}^b R_{ib}'^b, \quad (16)$$

$$\alpha_j y_j = F(p_{cb} + R_{cb} q_1^b + R_{cb} R_{ib}'^i (q_j^i - q_1^i)), \quad (17)$$

<sup>1</sup> We denote by  $\text{SE}(3)$  the Cartesian product of  $\mathbb{R}^3$  with the group  $\text{SO}(3)$  of  $3 \times 3$  rotation matrices; and by  $\text{se}(3)$  the Cartesian product of  $\mathbb{R}^3$  with the space  $\text{so}(3)$  of  $3 \times 3$  skew-symmetric matrices (cf., e.g., [Murray, Li and Sastry, 1994]).

$\forall j \in \{1, 2, \dots, N\}$ , where  $(p_{cb}, R_{cb}) \in \text{SE}(3)$  denotes the configuration of the frame  $\{b\}$  with respect to the camera's frame  $\{c\}$ , and  $F$  an upper triangular matrix with the camera's intrinsic parameters, of the form

$$\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ 0 & f_{22} & f_{21} \\ 0 & 0 & 1 \end{bmatrix},$$

where each  $f_{ij}$  denotes a scalar. For ideal pin-hole cameras  $f_{12} = f_{13} = f_{21} = 0$  and  $f_{11} = f_{22} = f$ , where  $f$  is the camera's focal length. For the general case, we refer the reader *e.g.* to [Faugeras, 1993]. Note that  $F$  and  $(p_{cb}, R_{cb})$  can be time-varying in case the camera is allowed to zoom or pan and tilt, which is often needed to get good visual information. The normalization constraints (3) are given by

$$[0 \ 0 \ 1] y_j = 1, \quad \forall j \in \{1, 2, \dots, N\}.$$

To proceed we need the following notation: Given a  $m \times n$ -matrix  $M$ , we denote by  $\text{stack}(M)$  the  $mn$ -vector obtained from stacking the columns of  $M$  one on top of each other, with the first column on top. Given two matrices  $M_i \in \mathbb{R}^{m_i \times n_i}$ ,  $i \in \{1, 2\}$  we denote by  $M_1 \otimes M_2 \in \mathbb{R}^{m_1 m_2 \times n_1 n_2}$  the Kronecker product of  $M_1$  by  $M_2$ . Using the fact that given three matrices with appropriate dimensions  $\text{stack}(AXB) = (B' \otimes A) \text{stack}(X)$  [Horn and Johnson, 1994], we can re-write (15)–(17) as follows:

$$\begin{aligned} \dot{q}_1^b &= -\Omega_{ib}^b q_1^b + (I_{3 \times 3} \otimes \dot{q}_1^i) \text{stack}(R_{ib}) - v_{ib}^b, \\ \text{stack}(\dot{R}_{ib}) &= (-\Omega_{ib}^b \otimes I_{3 \times 3}) \text{stack}(R_{ib}), \\ \alpha_j y_j &= F p_{cb} + F R_{cb} q_1^b \\ &\quad + (F R_{cb} \otimes (q_j^i - q_1^i)') \text{stack}(R_{ib}). \end{aligned}$$

Thus, defining

$$x := \begin{bmatrix} q_1^b \\ \text{stack}(R_{ib}) \end{bmatrix} \in \mathbb{R}^{12},$$

and

$$\begin{aligned} A &:= \begin{bmatrix} -\Omega_{ib}^b & I_{3 \times 3} \otimes \dot{q}_1^i \\ 0 & -\Omega_{ib}^b \otimes I_{3 \times 3} \end{bmatrix}, & b &:= \begin{bmatrix} -v_{ib}^b \\ 0 \end{bmatrix}, \\ C_j &:= [F R_{cb} \quad F R_{cb} \otimes (q_j^i - q_1^i)'], & d_j &:= F p_{cb}, \end{aligned}$$

it can be seen that the system (15)–(17) belongs to the class of systems of the form (1)–(2). Once we compute estimates  $\hat{R}_{ib}$  and  $\hat{q}_1^b$  for  $R_{ib}$  and  $q_1^b$ , respectively, we can also estimate  $p_{ib}$  using

$$\hat{p}_{ib} = \hat{q}_1^i - \hat{R}_{ib} \hat{q}_1^b.$$

## REFERENCES

- Aguiar, A. P. and Hespanha, J. P. [2003a]. Minimum-energy state estimation for systems with perspective outputs and state constraints, *Proc. of the 42nd Conf. on Decision and Contr.*
- Aguiar, A. P. and Hespanha, J. P. [2003b]. Pose estimation of mobile robots using visual information: A minimum-energy estimator approach, *Technical report*, University of California, Santa Barbara.
- Chiussi, A., Favaro, P., Jin, H. and Soatto, S. [2002]. Structure from motion causally integrated over time. To appear in *IEEE Trans. on Pattern Anal. Mach. Intell.*
- Faugeras, O. [1993]. *Three-Dimensional Computer Vision: a Geometric Viewpoint*, MIT Press, Cambridge, MA.
- Ghosh, B. K., Jankovic, M. and Wu, Y. T. [1994]. Perspective problems in system theory and its application in machine vision, *J. Mathematical Syst. Estimation Contr.* **4**(1): 3–38.
- Ghosh, B. K. and Loucks, E. P. [1995]. A perspective theory for motion and shape estimation in machine vision, *SIAM J. Contr. Optimization* **33**(5): 1530–1559.
- Hespanha, J. P. [2002]. State estimation and control for systems with perspective outputs, *Proc. of the 41st Conf. on Decision and Contr.*
- Hespanha, J. P. and Morse, A. S. [1999]. Stability of switched systems with average dwell-time, *Proc. of the 38th Conf. on Decision and Contr.*
- Hespanha, J., Yakimenko, O., Kaminer, I. and Pascoal, A. [2001]. Linear parametrically varying systems with brief instabilities: an application to integrated vision/IMU navigation, *Proc. of the 40th Conf. on Decision and Contr.*
- Hijab, O. J. [1980]. *Minimum Energy Estimation*, PhD thesis, University of California, Berkeley.
- Horn, R. A. and Johnson, C. R. [1994]. *Topics in Matrix Analysis*, Cambridge University Press., Cambridge.
- Jankovic, M. and Ghosh [1995]. Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer, *Syst. & Contr. Lett.* **25**: 63–73.
- Kaminer, I., Pascoal, A. M., Kang, W. and Yakimenko, O. [2001]. Integrated vision/inertial navigation systems design using nonlinear filtering, *IEEE Trans. Aerospace and Electronic Syst.* **37**(1): 158–172.
- Krener, A. J. [2003]. The convergence of the minimum energy estimator, in W. Kang (ed.), *New Trends in Nonlinear Dynamics and Control, and Their Applications*, Lecture Notes in Control and Information Sciences, Springer-Verlag. To appear.
- Ma, Y., Kosecka, J. and Sastry, S. [1999]. Vision guided navigation for a nonholonomic mobile robot, *IEEE Trans. Robot. Automat.* **15**(3): 521–536.
- Matthies, L., Kanade, T. and Szeliski, R. [1989]. Kalman filter-based algorithms for estimating depth from image sequences, *Int. J. of Comput. Vision* **3**: 209–236.
- McEneaney, W. M. [1998]. Robust/H-infinity filtering for nonlinear systems, *Syst. & Contr. Lett.* (33): 315–325.
- Mortensen, R. E. [1968]. Maximum likelihood recursive nonlinear filtering, *J. Opt. Theory and Applications* **2**: 386–394.
- Murray, R. M., Li, Z. and Sastry, S. S. [1994]. *A Mathematical Introduction to Robotic Manipulation*, CDC Press, Boca Raton, Florida.
- Rehbinder, H. and Ghosh, B. K. [2003]. Pose estimation using line-based dynamic vision and inertial sensors, *IEEE Trans. on Automat. Contr.* **48**(2): 186–199.
- Shakernia, O., Ma, Y., Koo, T. and Sastry, S. [1999]. Landing an unmanned air vehicle: vision based motion estimation and nonlinear control, *Asian J. Contr.* **1**(3): 128–145.
- Soatto, S., Frezza, R. and Perona, P. [1996]. Motion estimation via dynamic vision, *IEEE Trans. on Automat. Contr.* **41**(3): 393–413.
- Takahashi, S. and Ghosh, B. K. [2001]. Motion and shape parameters identification with vision and range, *Proc. of the 2001 Amer. Contr. Conf.*, Vol. 6, pp. 4626–4631.