

Identification of a Moving Object's Velocity with a Fixed Camera [★]

Vilas K. Chitrakaran ^{a,*}, Darren M. Dawson ^a, Warren E. Dixon ^b, Jian Chen ^a

^a*Department of Electrical & Computer Engineering, Clemson University, Clemson, SC 29634-0915*

^b*Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville, FL 32611*

Abstract

In this paper, a continuous estimator strategy is utilized to asymptotically identify the six degree-of-freedom velocity of a moving object using a single fixed camera. The design of the estimator is facilitated by the fusion of homography-based techniques with Lyapunov design methods. Similar to the stereo vision paradigm, the proposed estimator utilizes different views of the object from a single camera to calculate 3D information from 2D images. In contrast to some of the previous work in this area, no explicit model is used to describe the movement of the object; rather, the estimator is constructed based on bounds on the object's velocity, acceleration, and jerk.

Key words: Homography; Lyapunov; Velocity Observer

1 Introduction

Often in an engineering application, one is tempted to use a camera to determine the velocity of a moving object. However, as stated in Ghosh, Jankovic, and Wu (1994), the use of a camera requires one to interpret the motion of a 3-dimensional (3D) object through 2D images provided by the camera. That is, the primary problem is that 3D information is compressed or nonlinearly transformed into 2D information; hence, techniques or methods must be developed to obtain 3D information despite the fact that only 2D information is available. To address the identification of the object's velocity (i.e., the motion parameters), many researchers have developed various approaches. For example, if a model for the object's motion is known, an observer can be used to estimate the object's velocity Ghosh and Loucks (1996). In Rizzi and Koditchev (1996), a window position predictor for object tracking was utilized. In Hashimoto and Noritsugu (1996), an observer for estimating the object velocity was utilized; however, a description of the ob-

ject's kinematics must be known. In Ghosh, Inaba, and Takahashi (2000), the problem of identifying the motion and shape parameters of a planar object undergoing Riccati motion was examined in great detail. In Koivo and Houshangi (1991), an autoregressive discrete-time model is used to predict the location of features of a moving object. In Allen, Timcenko, Yoshimi, and Michelman (1992), trajectory filtering and prediction techniques are utilized to track a moving object. Some of the work (Tsai and Huang, 1984) involves the use of camera-centered models that compute values for the motion parameters at each new frame to produce the motion of the object. In Broida and Chellappa (1991) and Shariat and Price (1990), object-centered models are utilized to estimate the translation and the center of rotation of the object. In Waxman and Duncan (1986), the motion parameters of an object are determined via a stereo vision approach.

While it is difficult to make broad statements concerning much of the previous work on velocity identification, it does seem that a good amount of effort has been focused on developing system theory-based algorithms to estimate the object's velocity or compensate for the object's velocity as part of a feedforward control scheme. For example, one might assume that object kinematics can be described as follows

$$\dot{x} = Y(x)\phi \quad (1)$$

[★] This paper was not presented at any IFAC meeting.

^{*} Corresponding author. Tel. +1-864-656-7708. Fax +1-864-656-7218.

Email addresses: cvilas@ces.clemson.edu (Vilas K. Chitrakaran), ddawson@ces.clemson.edu (Darren M. Dawson), wdixon@ufl.edu (Warren E. Dixon), jianc@ces.clemson.edu (Jian Chen).

where $x(t)$, $\dot{x}(t)$ denote the object's position vector and object's velocity vector, respectively, $Y(x)$ denotes a known regression matrix, and ϕ denotes an unknown, constant vector. As illustrated in Ghosh, Xi, and Tarn (1999), the object model of (1) can be used to describe many types of object models (e.g., constant-velocity, and cyclic motions). If $x(t)$ is measureable, it is easy to imagine how adaptive control techniques (Slotine and Li, 1991) can be utilized to formulate an adaptive update law that could compensate for unknown effects represented by the parameter ϕ for a typical control problem. In addition, if $x(t)$ is persistently exciting, one might be able to also show that the unknown parameter ϕ could be identified asymptotically. In a similar manner, robust control strategies or learning control strategies could be used to compensate for unknown object kinematics under the standard assumptions for these types of controllers (Messner, Horowitz, Kao, and Boals, 1991; Qu, 1998).

While the above control techniques provide different methods for compensating for unknown object kinematics, these methods do not seem to provide much help with regard to identifying the object's velocity if not much is known about the motion of the object. That is, from a systems theory point of view, one must develop a method of asymptotically identifying a time-varying signal with as little information as possible. This problem is made even more difficult because the sensor being used to gather the information about the object is a camera, and as mentioned before, the use of a camera requires one to interpret the motion of a 3D object from 2D images. To attack this double-goaled problem, we fuse homography-based techniques with a Lyapunov synthesized estimator to asymptotically identify the object's unknown velocity¹. Similar to the stereo vision paradigm, the proposed approach uses different views of the object from a single camera to calculate 3D information from 2D images. The homography-based techniques are based on fixed camera work presented in Chen, Behal, Dawson, and Fang (2003) which relies on the camera-in-hand work presented in Malis, Chaumette, and Bodet (1999). The continuous, Lyapunov-based estimation strategy has its roots in an example developed in Qu and Xu (2002) and the general framework developed in Xian, De Queiroz, and Dawson (2004). The only requirements on the object are that its velocity, acceleration, and jerk be bounded, and that a single geometric length between two feature points on the object be known a priori.

2 Geometric Model

To facilitate the subsequent object velocity identification problem, four target points located on an object

¹ The object's six degree-of-freedom translational and rotational velocity is asymptotically identified.

denoted by $O_i \forall i = 1, 2, 3, 4$ are considered to be coplanar² and not colinear. Based on this assumption, consider a fixed plane, denoted by π^* , that is defined by a reference image of the object. In addition, let π represent the motion of the plane containing the object feature points (see Figure 1). To develop a relationship between the planes, an inertial coordinate system, denoted by \mathcal{I} , is defined where the origin coincides with the center of a fixed camera. The 3D coordinates of the target points on π and π^* can be respectively expressed in terms of \mathcal{I}

$$\bar{m}_i(t) \triangleq \begin{bmatrix} x_i(t) & y_i(t) & z_i(t) \end{bmatrix}^T \quad (2)$$

$$\bar{m}_i^* \triangleq \begin{bmatrix} x_i^* & y_i^* & z_i^* \end{bmatrix}^T \quad (3)$$

under the standard assumption that the distances from the origin of \mathcal{I} to the target points remains positive (i.e., $z_i(t), z_i^* > \varepsilon$ where ε is an arbitrarily small positive constant). Orthogonal coordinate systems \mathcal{F} and \mathcal{F}^* are attached to π and π^* , respectively, where the origin of the coordinate systems coincides with the object (see Figure 1). To relate the coordinate systems, let $R(t)$, $R^* \in SO(3)$ denote the rotation between \mathcal{F} and \mathcal{I} , and \mathcal{F}^* and \mathcal{I} , respectively, and let $x_f(t)$, $x_f^* \in \mathbb{R}^3$ denote the respective translation vectors expressed in the coordinates of \mathcal{I} . As also illustrated in Figure 1, $n^* \in \mathbb{R}^3$ denotes the constant normal to the plane π^* expressed in the coordinates of \mathcal{I} , $s_i \in \mathbb{R}^3$ denotes the constant coordinates of the target points located on the object reference frame, and the constant distance $d^* \in \mathbb{R}$ from \mathcal{I} to \mathcal{F}^* along the unit normal is given by

$$d^* = n^{*T} \bar{m}_i^*. \quad (4)$$

The subsequent development is based on the assumption that the constant coordinates of one target point s_i is known. For simplicity and without loss of generality, we assume that the coordinate s_1 is known (i.e., the subsequent development requires a single geometric length between two feature points on the object be known a priori).

From the geometry between the coordinate frames depicted in Figure 1, the following relationships can be developed

$$\bar{m}_i = x_f + R s_i \quad (5)$$

$$\bar{m}_i^* = x_f^* + R^* s_i. \quad (6)$$

After solving (6) for s_i and then substituting the resulting expression into (5), the following relationships can be obtained

$$\bar{m}_i = \bar{x}_f + \bar{R} \bar{m}_i^* \quad (7)$$

² It should be noted that if four coplanar target points are not available then the subsequent development can exploit the classic eight-points algorithm (Malis and Chaumette, 2000) with no four of the eight target points being coplanar.

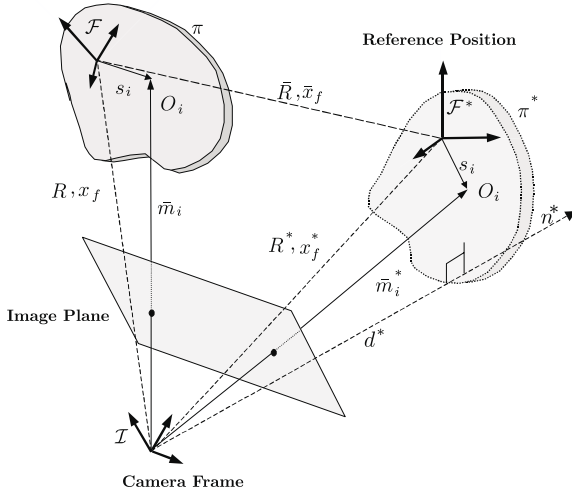


Fig. 1. Coordinate frame relationships.

where $\bar{R}(t) \in SO(3)$ and $\bar{x}_f(t) \in \mathbb{R}^3$ are new rotational and translational variables, respectively, defined as follows

$$\bar{R} = R(R^*)^T \quad \bar{x}_f = x_f - \bar{R}x_f^*. \quad (8)$$

From (4), it is easy to see how the relationship in (7) can now be expressed as follows

$$\bar{m}_i = \left(\bar{R} + \frac{\bar{x}_f}{d^*} n^{*T} \right) \bar{m}_i^*. \quad (9)$$

Remark 1 The subsequent development requires that the constant rotation matrix R^* be known. This is considered to be a mild assumption since the constant rotation matrix R^* can be obtained a priori using various methods (e.g., a second camera, Euclidean measurements, etc.).

3 Euclidean Reconstruction

The relationship given by (9) provides a means for formulating a translation and rotation error between \mathcal{F} and \mathcal{F}^* . Since the Euclidean position of \mathcal{F} and \mathcal{F}^* cannot be directly measured, a method for calculating the position and rotational error using pixel information is developed in this section (i.e., pixel information is the measurable quantity as opposed to $\bar{m}_i(t)$ and \bar{m}_i^*). To this end, the normalized 3D task-space coordinates of the points on π and π^* can be respectively expressed in terms of \mathcal{I} as $m_i(t)$, $m_i^* \in \mathbb{R}^3$, as follows

$$m_i \triangleq \frac{\bar{m}_i}{z_i} = \begin{bmatrix} x_i & y_i & 1 \end{bmatrix}^T \quad (10)$$

$$m_i^* \triangleq \frac{\bar{m}_i^*}{z_i^*} = \begin{bmatrix} x_i^* & y_i^* & 1 \end{bmatrix}^T. \quad (11)$$

The rotation and translation between the normalized coordinates can now be related through a Euclidean ho-

mography, denoted by $H(t) \in \mathbb{R}^{3 \times 3}$, as follows

$$m_i = \underbrace{\frac{z_i^*}{z_i}}_{\alpha_i} \underbrace{\left(\bar{R} + \bar{x}_h(n^*)^T \right)}_H m_i^* \quad (12)$$

where $\alpha_i(t) \in \mathbb{R}$ denotes a depth ratio, and $\bar{x}_h(t) \in \mathbb{R}^3$ denotes a scaled translation vector that is defined as follows

$$\bar{x}_h = \frac{\bar{x}_f}{d^*}. \quad (13)$$

In addition to having a task-space coordinate as described previously, each target point O_i , O_i^* will also have a projected pixel coordinate expressed in terms of \mathcal{I} , denoted by $u_i(t)$, $v_i(t)$, u_i^* , $v_i^* \in \mathbb{R}$, that is respectively defined as an element of $p_i(t)$, $p_i^* \in \mathbb{R}^3$, as follows

$$p_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T \quad p_i^* = \begin{bmatrix} u_i^* & v_i^* & 1 \end{bmatrix}^T. \quad (14)$$

The projected pixel coordinates of the target points are related to the normalized task-space coordinates by the following pin-hole lens models (Faugeras, 2001)

$$p_i = A m_i \quad p_i^* = A m_i^* \quad (15)$$

where $A \in \mathbb{R}^{3 \times 3}$ is a known, constant, and invertible intrinsic camera calibration matrix. After substituting (15) into (12), the following relationship can be developed

$$p_i = \alpha_i \underbrace{(A H A^{-1})}_G p_i^* \quad (16)$$

where $G(t) = [g_{ij}(t)] \forall i, j = 1, 2, 3 \in \mathbb{R}^{3 \times 3}$ denotes a projective homography. After normalizing $G(t)$ by dividing through by the element $g_{33}(t)$, which is assumed to be nonzero without loss of generality, the projective relationship in (16) can be expressed as follows

$$p_i = \alpha_i g_{33} G_n p_i^* \quad (17)$$

where $G_n(t) \in \mathbb{R}^{3 \times 3}$ denotes the normalized projective homography. From (17), a set of 12 linear equations given by the 4 target point pairs $(p_i^*, p_i(t))$ with 3 equations per target pair can be used to determine $G_n(t)$ and $\alpha_i(t)g_{33}(t)$. Based on the fact that the intrinsic calibration matrix A is assumed to be known, (16) and (17) can be used to determine the product $g_{33}(t)H(t)$. By utilizing various techniques (Faugeras and Lustman, 1988; Zhang and Hanson, 1995), the product $g_{33}(t)H(t)$ can be decomposed into rotational and translational components as in (12). Specifically, the scale factor $g_{33}(t)$, the rotation matrix $\bar{R}(t)$, the unit normal vector n^* , and the scaled translation vector denoted by $\bar{x}_h(t)$ can all be computed from the decomposition of the product

$g_{33}(t)H(t)$. Since the product $\alpha_i(t)g_{33}(t)$ can be computed from (17), and $g_{33}(t)$ can be determined through the decomposition of the product $g_{33}(t)H(t)$, the depth ratio $\alpha_i(t)$ can be also be computed. Based on the assumption that R^* is known and the fact that $\bar{R}(t)$ can be computed from the homography decomposition, (8) can be used to compute $R(t)$. Hence, $R(t)$, $\bar{R}(t)$, $\bar{x}_h(t)$, n^* , and the depth ratio $\alpha_i(t)$ are all known signals that can be used in the subsequent estimator design.

4 Object Kinematics

Based on information obtained from the Euclidean reconstruction, the object kinematics are developed in this section. To develop the translation kinematics for the object, $e_v(t) \in \mathbb{R}^3$ is defined to quantify the translation of \mathcal{F} with respect to the fixed coordinate system \mathcal{F}^* as follows

$$e_v = p_e - p_e^* . \quad (18)$$

In (18), $p_e(t) \in \mathbb{R}^3$ denotes the following extended image coordinates (Malis et al., 1999) of an image point ³ on π in terms of the inertial coordinate system \mathcal{I}

$$p_e = \begin{bmatrix} u_1 & v_1 & \ln(z_1) \end{bmatrix}^T \quad (19)$$

where $\ln(\cdot)$ denotes the natural logarithm, and $p_e^* \in \mathbb{R}^3$ denotes the following extended image coordinates of the corresponding image point on π^* in terms of \mathcal{I}

$$p_e^* = \begin{bmatrix} u_1^* & v_1^* & \ln(z_1^*) \end{bmatrix}^T . \quad (20)$$

The first two elements of $e_v(t)$ are directly measured from the images. By exploiting standard properties of the natural logarithm, it is clear that the third element of $e_v(t)$ is equivalent to $-\ln(\alpha_1)$; hence, $e_v(t)$ is a known signal since $\alpha_1(t)$ is computed during the Euclidean reconstruction. After taking the time derivative of (18), the following translational kinematics can be obtained (see Appendix A for further details)

$$\dot{e}_v = \dot{p}_e = \frac{\alpha_1}{z_1^*} A_e L_v [v_e - R[s_1]_\times R^T \omega_e] \quad (21)$$

where $v_e(t), \omega_e(t) \in \mathbb{R}^3$ denote the *unknown* linear and angular velocity of the object expressed in \mathcal{I} , respectively. In (21), $A_e \in \mathbb{R}^{3 \times 3}$ is defined as follows

$$A_e = A - \begin{bmatrix} 0 & 0 & u_0 \\ 0 & 0 & v_0 \\ 0 & 0 & 0 \end{bmatrix} \quad (22)$$

³ Any point O_i on π can be utilized in the subsequent development; however, to reduce the notational complexity, we have elected to select the image point O_1 , and hence, the subscript 1 is utilized in lieu of i in the subsequent development.

where $u_0, v_0 \in \mathbb{R}$ denote the pixel coordinates of the principal point (i.e., the image center that is defined as the frame buffer coordinates of the intersection of the optical axis with the image plane), and the auxiliary Jacobian-like matrix $L_v(t) \in \mathbb{R}^{3 \times 3}$ is defined as

$$L_v = \begin{bmatrix} 1 & 0 & -\frac{x_1}{z_1} \\ 0 & 1 & -\frac{y_1}{z_1} \\ 0 & 0 & 1 \end{bmatrix} . \quad (23)$$

To develop the rotation kinematics for the object, $e_\omega(t) \in \mathbb{R}^3$ is defined using the angle axis representation (Spong and Vidyasagar, 1989) to quantify the rotation of \mathcal{F} with respect to the fixed coordinate system \mathcal{F}^* as follows

$$e_\omega \triangleq u(t)\theta(t) . \quad (24)$$

In (24), $u(t) \in \mathbb{R}^3$ represents a unit rotation axes, and $\theta(t) \in \mathbb{R}$ denotes the rotation angle about $u(t)$ that is assumed to be confined to the following region

$$-\pi < \theta(t) < \pi . \quad (25)$$

After taking the time derivative of (24), the following expression can be obtained (see Appendix A for further details)

$$\dot{e}_\omega = L_\omega \omega_e . \quad (26)$$

In (26), the Jacobian-like matrix $L_\omega(t) \in \mathbb{R}^{3 \times 3}$ is defined as

$$L_\omega = I_3 - \frac{\theta}{2} [u]_\times + \left(1 - \frac{\text{sinc}(\theta)}{\text{sinc}^2\left(\frac{\theta}{2}\right)} \right) [u]_\times^2 \quad (27)$$

where $[u]_\times$ denotes the 3×3 skew-symmetric form of $u(t)$ and

$$\text{sinc}(\theta(t)) \triangleq \frac{\sin \theta(t)}{\theta(t)} .$$

Remark 2 The structure of (18)-(20) is motivated by the fact that developing the object kinematics using partial pixel information clarifies the influence of the camera intrinsic calibration matrix. By observing the influence of the intrinsic calibration parameters, future efforts might be directed at developing an estimator strategy that is robust to these parameters. Since the intrinsic calibration matrix is assumed to be known in this paper, the estimator strategy could also be developed based solely on reconstructed Euclidean information (e.g., $\bar{x}_h(t)$, $\bar{R}(t)$).

Remark 3 As stated in Spong and Vidyasagar (1989), the angle axis representation in (24) is not unique, in the sense that a rotation of $-\theta(t)$ about $-u(t)$ is equal to a rotation of $\theta(t)$ about $u(t)$. A particular solution for

$\theta(t)$ and $u(t)$ can be determined as follows (Spong and Vidyasagar, 1989)

$$\theta_p = \cos^{-1} \left(\frac{1}{2} (tr(\bar{R}) - 1) \right) \quad [u_p]_{\times} = \frac{\bar{R} - \bar{R}^T}{2 \sin(\theta_p)} \quad (28)$$

where the notation $tr(\cdot)$ denotes the trace of a matrix and $[u_p]_{\times}$ denotes the 3×3 skew-symmetric form of $u_p(t)$. From (28), it is clear that

$$0 \leq \theta_p(t) \leq \pi. \quad (29)$$

While (29) is confined to a smaller region than $\theta(t)$ in (25), it is not more restrictive in the sense that

$$u_p \theta_p = u \theta. \quad (30)$$

The constraint in (29) is consistent with the computation of $[u(t)]_{\times}$ in (28) since a clockwise rotation (i.e., $-\pi \leq \theta(t) \leq 0$) is equivalent to a counterclockwise rotation (i.e., $0 \leq \theta(t) \leq \pi$) with the axis of rotation reversed. Hence, based on (30) and the functional structure of the object kinematics, the particular solutions $\theta_p(t)$ and $u_p(t)$ can be used in lieu of $\theta(t)$ and $u(t)$ without loss of generality and without confining $\theta(t)$ to a smaller region. Since, we do not distinguish between rotations that are off by multiples of 2π , all rotational possibilities are considered via the parameterization of (24) along with the computation of (28).

Remark 4 By exploiting the fact that $u(t)$ is a unit vector (i.e., $\|u\|^2 = 1$), the determinant of $L_{\omega}(t)$ can be derived as follows (Malis and Chaumette, 2002)

$$\det L_{\omega} = \frac{1}{\sin^2 \left(\frac{\theta}{2} \right)}. \quad (31)$$

From (31), it is clear that $L_{\omega}(t)$ is only singular for multiples of 2π (i.e., out of the assumed workspace).

5 Velocity Identification

5.1 Objective

The objective in this paper is to develop an estimator that can be used to identify the translational and rotational velocity of an object expressed in \mathcal{I} , denoted by $v(t) = \begin{bmatrix} v_e^T & \omega_e^T \end{bmatrix}^T \in \mathbb{R}^6$. To facilitate this objective, the object kinematics are expressed in the following compact form

$$\dot{e} = Jv \quad (32)$$

where the Jacobian-like matrix $J(t) \in \mathbb{R}^{6 \times 6}$ is defined as

$$J = \begin{bmatrix} \frac{\alpha_1}{z_1^*} A_e L_v - \frac{\alpha_1}{z_1^*} A_e L_v R [s_1]_{\times} R^T \\ 0 & L_{\omega} \end{bmatrix} \quad (33)$$

where (21) and (26) were utilized, and $e(t) = \begin{bmatrix} e_v^T & e_{\omega}^T \end{bmatrix}^T \in \mathbb{R}^6$. The subsequent development is based on the assumption that $v(t)$ of (32) is bounded and is second order differentiable with bounded derivatives. It is also assumed that if $v(t)$ is bounded, then the structure of (32) ensures that $e(t)$ is bounded. From (32), (33), and the previous assumptions, it is clear that if $e(t)$, $v(t) \in L_{\infty}$, then from (32) we can see that $\dot{e}(t) \in L_{\infty}$. We can differentiate (32) to show that $\ddot{e}(t), \ddot{v}(t) \in L_{\infty}$; hence, we can use the previous assumptions to show that

$$|\ddot{e}_i(t)| + |\ddot{v}_i(t)| \leq \beta_i \quad \forall i = 1, 2, \dots, 6. \quad (34)$$

where $\beta_i \in \mathbb{R}$ are known positive constants.

Based on the error system for the object kinematics given in (32) and the inequalities in (34), an estimator is designed in the next section to ensure that

$$\|\tilde{e}(t)\|, \|\dot{\tilde{e}}(t)\| \rightarrow 0 \text{ as } t \rightarrow \infty \quad (35)$$

where the observation error signal $\tilde{e}(t) \in \mathbb{R}^6$ is defined as follows

$$\tilde{e} = e - \hat{e} \quad (36)$$

where $\hat{e}(t) \in \mathbb{R}^6$ denotes a subsequently designed estimate for $e(t)$. Once the result in (35) is obtained, additional development is provided that proves $v(t)$ can be exactly identified.

5.2 Estimator Development

To facilitate the following analysis, we define a filtered observation error, denoted by $r(t) \in \mathbb{R}^6$, as follows (Slotine and Li, 1991)

$$r = \dot{\tilde{e}} + \tilde{e}. \quad (37)$$

After taking the time derivative of (37), the following expression can be obtained

$$\dot{r} = \ddot{\tilde{e}} - \dot{\tilde{e}} + \dot{\tilde{e}}. \quad (38)$$

Based on subsequent analysis, $\hat{e}(t)$ is generated from the following differential expression

$$\dot{\hat{e}} = \hat{\kappa} \quad (39)$$

where $\hat{\kappa}(t) \in \mathbb{R}^6$ is defined as follows⁴

$$\hat{\kappa}(t) = \int_{t_0}^t (K + I_6) \tilde{e}(\tau) d\tau + \int_{t_0}^t \rho \text{sgn}(\tilde{e}) d\tau + (K + I_6) \tilde{e}(t) \quad (40)$$

where $K, \rho \in \mathbb{R}^{6 \times 6}$ are positive definite constant diagonal gain matrices, $I_6 \in \mathbb{R}^{6 \times 6}$ denotes the 6×6 identity matrix, t_0 is the initial time, and the notation $\text{sgn}(\tilde{e})$ denotes the standard signum function applied to each element of the vector $\tilde{e}(t)$. After taking the time derivative of (39) and substituting the resulting expression into (38), the following expression is obtained

$$\dot{r} = \ddot{e} - (K + I_6)r - \rho \text{sgn}(\tilde{e}) + \dot{\tilde{e}} \quad (41)$$

where the time derivative of (40) was utilized.

5.3 Analysis

Theorem 1 *The estimator defined in (39) and (40) can be used to obtain the objective given in (35) provided the elements of the estimator gain ρ is selected as follows*

$$\rho_i > \beta_i \quad \forall i = 1, 2, \dots, 6. \quad (42)$$

where ρ_i is the i^{th} diagonal element in the gain matrix ρ , and β_i are introduced in (34).

Proof: To prove Theorem 1, a non-negative function, denoted by $V(t) \in \mathbb{R}$, is defined as follows

$$V \triangleq \frac{1}{2} r^T r + \frac{1}{2} \tilde{e}^T \tilde{e} + P \quad (43)$$

where the auxiliary function $P(t) \in \mathbb{R}$ is defined as follows

$$P(t) \triangleq \zeta_b - \int_{t_0}^t L(\tau) d\tau \quad (44)$$

where $\zeta_b, L(t) \in \mathbb{R}$ are auxiliary terms defined as follows

$$\zeta_b \triangleq \sum_{i=1}^6 \rho_i |\tilde{e}_i(t_0)| - \tilde{e}^T(t_0) \ddot{e}(t_0) \quad L \triangleq r^T [\ddot{e} - \rho \text{sgn}(\tilde{e})]. \quad (45)$$

In Appendix B, the auxiliary function $P(t)$ introduced in (43) is proven to be non-negative (i.e., $P(t) \geq 0$) provided the sufficient condition given in (42) is satisfied. After taking the time derivative of (43), the following expression is obtained

$$\dot{V} = r^T [\ddot{e} - (K + I_6)r - \rho \text{sgn}(\tilde{e}) + \dot{\tilde{e}}] + \dot{\tilde{e}}^T \tilde{e} - L. \quad (46)$$

⁴ The structure of the estimator is motivated by the previous work in Xian et al. (2004) that was inspired by an example in Qu and Xu (2002).

The expression in (46) can be rewritten as follows

$$\dot{V} = -(K + I_6) \|r\|^2 + \|r\|^2 - \|\tilde{e}\|^2 \quad (47)$$

where (37) and (45) were utilized. After simplifying (47) as follows

$$\dot{V} \leq -K \|r\|^2, \quad (48)$$

it is clear from (43) that $r(t), \tilde{e}(t), P(t) \in L_\infty$ and that $r(t) \in L_2$ (Dixon, Behal, Dawson, and Nagarkatti, 2003). Based on the fact that $r(t) \in L_\infty$, linear analysis techniques (Dixon et al., 2003) can be used to determine that $\dot{\tilde{e}}(t) \in L_\infty$. Since $\tilde{e}(t), \dot{\tilde{e}}(t) \in L_\infty$, (36), (39), and the assumption that $e(t), \dot{e}(t) \in L_\infty$ can be used to determine that $\hat{e}(t), \dot{\hat{e}}(t), \hat{\kappa}(t) \in L_\infty$. Also, since $\ddot{e}(t), r(t), \tilde{e}(t), \dot{\tilde{e}}(t) \in L_\infty$, (41) can be used to determine that $\dot{r}(t) \in L_\infty$. Based on the fact that $r(t) \in L_\infty \cap L_2$ and $\dot{r}(t) \in L_\infty$, Barbalat's Lemma (Slotine and Li, 1991) can be used to prove that

$$\|r(t)\| \rightarrow 0 \text{ as } t \rightarrow \infty. \quad (49)$$

Given (49), linear analysis techniques (Dixon et al., 2003) can be used to determine that the result in (35) is obtained. \square

Remark 5 *Notice from (37), (41) and (44) that the differential equations describing the closed-loop system has a discontinuous right-hand side:*

$$\dot{\tilde{e}} = -\tilde{e} + r \quad (50)$$

$$\dot{r} = \ddot{e} - (K + I_6)r - \rho \text{sgn}(\tilde{e}) + \dot{\tilde{e}} \quad (51)$$

$$\dot{P} = -r^T [\ddot{e} - \rho \text{sgn}(\tilde{e})]. \quad (52)$$

Let $\xi := [\tilde{e}^T \ r^T \ P]^T$ and $g(\xi, t): \mathbb{R}^6 \times \mathbb{R}^6 \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^{13}$ denote the right-hand side of (50)-(52). Since the above proof requires that a solution exist for $\dot{\xi} = g(\xi, t)$, it is important that we comment on the existence of solutions to (50)-(52). To this end, we follow the arguments used (Polycarpou and Ioannou, 1993; Qu, 1998) to discuss the existence of Filippov's generalized solution to (50)-(52). First, note that $g(\xi, t)$ is continuous except in the set $\{(\xi, t) \mid \tilde{e} = 0\}$. Let $\mathcal{G}(\xi, t)$ be a compact, convex, upper semicontinuous set-valued map that embeds the differential equation $\dot{\xi} = g(\xi, t)$ into the differential inclusions $\dot{\xi} \in \mathcal{G}(\xi, t)$. From Theorem 2.7 of Qu (1998), we then know that an absolute continuous solution exists to $\dot{\xi} \in \mathcal{G}(\xi, t)$ that is a generalized solution to $\dot{\xi} = g(\xi, t)$. A common choice for $\mathcal{G}(\xi, t)$ that satisfies the above conditions is the closed convex hull of $g(\xi, t)$ (Polycarpou and Ioannou, 1993; Qu, 1998). A proof that this choice for $\mathcal{G}(\xi, t)$ is upper semicontinuous is given in Gutman (1979).

Theorem 2 *Given the results in (35), the object velocity expressed in \mathcal{I} (i.e., $v(t) = [v_e^T \ \omega_e^T]^T$) can be exactly*

determined provided the result in (35) is obtained and a single geometric length between two feature points (i.e., s_1) is known.

Proof: Given the result in (35), the definition given in (36) can be used to determine that

$$\dot{\hat{e}}_i(t) \rightarrow \dot{e}_i(t) \text{ as } t \rightarrow \infty \quad \forall i = 1, 2, \dots, 6. \quad (53)$$

Based on (53), the definition in (39) can be used to conclude that

$$\hat{\kappa}_i(t) \rightarrow \dot{\hat{e}}_i(t) \text{ as } t \rightarrow \infty \quad \forall i = 1, 2, \dots, 6. \quad (54)$$

Hence, (32) and (54) can be used to conclude that

$$\hat{\kappa}_i(t) \rightarrow (Jv)_i \text{ as } t \rightarrow \infty \quad \forall i = 1, 2, \dots, 6. \quad (55)$$

From the result in (55), it is clear that the object velocity can be identified since $J(t)$ of (33) is known and is invertible. Each element of $J(t)$ is known with the possible exception of the constant depth related parameter z_1^* .

To illustrate that the constant z_1^* is known, consider the following alternate representation for x_f^*

$$x_f^* = [\text{diag}(\bar{\gamma}_1 - \bar{\gamma}_2) + \bar{R}]^{-1} [\text{diag}(\bar{\gamma}_1 - \bar{\gamma}_2)R^* - R] s_1. \quad (56)$$

where

$$\bar{\gamma}_1 = [\text{diag}(m_1^*)]^{-1} n^{*T} m_1^* \bar{x}_h \quad \bar{\gamma}_2 = [\text{diag}(m_1^*)]^{-1} \frac{m_1}{\alpha_1} \quad (57)$$

and

$$m_1^* = \frac{x_f^*}{z_1^*} + R^* \frac{s_1}{z_1^*} \quad (58)$$

Since m_1^* and x_f^* can be computed, and R^* and s_1 are assumed to be known, (58) can be used to solve for z_1^* .

Remark 6 Given that R^* is assumed to be a known rotation matrix, it is straightforward to prove that the object velocity expressed in \mathcal{F} can be determined from the object velocity expressed in \mathcal{I} .

6 Simulation Results

A software based simulator was developed in order to test the performance of the velocity observer. A planar object with four markers for feature points was chosen as the body undergoing motion. The velocity of the object

along each of its 6 degrees of freedom was set to the following

$$v_i = 0.2 \sin(t) \quad \forall i = 1, 2, \dots, 6. \quad (59)$$

The coordinates of the feature point s_1 in the object's coordinate frame \mathcal{F}^* was $[-0.1 \ 0.1 \ 0.01]^T$. The object's reference position x_f^* and orientation R^* relative to the camera were selected as follows

$$x_f^* = \begin{bmatrix} 0 & 0 & 2 \end{bmatrix}^T \quad (60)$$

$$R^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \quad (61)$$

Note that in a practical implementation of this algorithm only R^* needs to be known and x_f^* is not required. Based on calibration parameters from an existing CCD camera, the camera intrinsic calibration matrix A was chosen to be the following

$$A = \begin{bmatrix} 1268.16 & 0 & 257.49 \\ 0 & 1267.50 & 253.10 \\ 0 & 0 & 1 \end{bmatrix}. \quad (62)$$

The simulator operated at the sampling frequency of 1 kHz. The algorithm performed best with estimator gains selected through trial and error as follows

$$K = \text{diag}(300, 300, 200, 15, 15, 15) \quad (63)$$

$$\rho = \text{diag}(100, 100, 10, 1, 1, 1).$$

Figure 2 depicts the velocity error along each of the 6 degrees of freedom as obtained from $J^{-1}\hat{\kappa}$ for the chosen estimator gains. Four feature points were used in this simulation to estimate the normalized homography matrix G_n and the scale factor $\alpha_i g_{33}$.

6.1 Simulation Discussion

While the numerical simulation provides an example of the performance of the velocity estimator, several issues must be considered for a practical implementation. The performance of the algorithm depends on the accuracy of estimation of the homography matrix since the Euclidean space information utilized in the object kinematics is obtained from decomposition of this matrix. Therefore inaccuracies in determining the location of a feature from one frame to the next frame (i.e., feature tracking) will lead to errors in the construction and decomposition of the homography matrix, leading to errors in the velocity estimation. Inaccuracies in determining the feature point coordinates in an image is a similar problem faced in numerous sensor based feedback

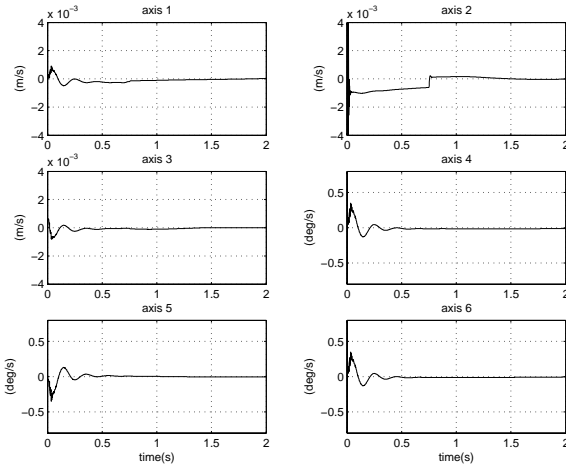


Fig. 2. Velocity estimation error along each of the six degrees of freedom.

applications (e.g., noise associated with a force/torque sensor). Practically, errors related to sensor inaccuracies can often be addressed with an ad hoc filter scheme or other mechanisms (e.g., an intelligent image processing and feature tracking algorithm, redundant feature points and an optimal homography computation algorithm such as Kanatani, Ohta, and Kanazawa (2000)). Despite the discrete nature of the sensor feedback in this and other applications, the algorithm design and analysis is often performed in the continuous time domain. Motivation for this simplification is due to the problematic and open issues related to the development of discrete Lyapunov analysis methods.

For this estimation strategy, the type of feature point used is only relevant to the specific application. For example, artificial markers such as skin tags typically employed in the animation and biomechanics literature could be used, or natural features (doorway frame, corners of an object, etc.) could be used when coupled with an appropriate image processing algorithm. In essence, the performance of the proposed velocity estimator will be improved by using an advanced feature tracking algorithm (i.e., the ability to accurately identify the image-space coordinates of an object feature in a single and in successive frames).

7 Conclusions

In this paper, we presented a continuous estimator strategy that can be utilized to asymptotically identify the six degree-of-freedom velocity of a moving object using a single fixed camera. The design of the estimator is based on a novel fusion of homography-based vision techniques and Lyapunov control design tools. The only requirements on the object are that its velocity and first two time derivatives be bounded, and that a single geometric length between two feature points on the object

be known a priori. Some of the practical applications of this technique are measurement of vibrations in flexible structures and estimation of motion in free-form bodies. It provides all the advantages non-contact measurement techniques can offer, most importantly the ability to work with hostile or delicate structures in a variety of environments. Future work will concentrate on the ramifications of the proposed estimator for other vision based applications.

Acknowledgements

This research was supported in part by U.S. NSF Grant DMI-9457967, ONR Grant N00014-99-1-0589, a DOC Grant, and an ARO Automotive Center Grant at Clemson University, and in part by AFOSR contract number F49620-03-1-0381 at the University of Florida.

A Rotational and Translational Kinematics

Based on the previous definitions for $\omega_e(t)$ and $R(t)$, the following property can be determined (Spong and Vidyasagar, 1989)

$$[\omega_e]_{\times} = \dot{R}R^T. \quad (\text{A.1})$$

From (8) and (A.1), the following relationship can be determined

$$[\omega_e]_{\times} = \dot{\bar{R}}\bar{R}^T. \quad (\text{A.2})$$

While several parameterizations can be used to express $\bar{R}(t)$ in terms of $u(t)$ and $\theta(t)$, the open-loop error system for $e_{\omega}(t)$ is derived based on the following exponential parameterization (Spong and Vidyasagar, 1989)

$$\bar{R} = I_3 + \sin \theta [u]_{\times} + 2 \sin^2 \frac{\theta}{2} [u]_{\times}^2 \quad (\text{A.3})$$

where the notation I_i denotes an $i \times i$ identity matrix, and the notation $[u]_{\times}$ denotes the skew-symmetric matrix form of $u(t)$. To facilitate the development of the open-loop dynamics for $e_{\omega}(t)$, the expression developed in (A.2) can be used along with (A.3) and the time derivative of (A.3), to obtain the following expression

$$[\omega_e]_{\times} = \sin \theta [\dot{u}]_{\times} + [u]_{\times} \dot{\theta} + (1 - \cos \theta) [[u]_{\times} \dot{u}]_{\times} \quad (\text{A.4})$$

where the following properties were utilized (Felippa, 2000; Malis, 1998)

$$[u]_{\times} \zeta = -[\zeta]_{\times} u, \quad (\text{A.5})$$

$$[u]_{\times}^2 = uu^T - I_3, \quad (\text{A.6})$$

$$[u]_{\times} uu^T = 0 \quad (\text{A.7})$$

$$[u]_{\times} [\dot{u}]_{\times} [u]_{\times} = 0, \quad (\text{A.8})$$

$$[[u]_{\times} \dot{u}]_{\times} = [u]_{\times} [\dot{u}]_{\times} - [\dot{u}]_{\times} [u]_{\times}. \quad (\text{A.9})$$

To facilitate further development, the time derivative of (24) is determined as follows

$$\dot{e}_\omega = \dot{u}\theta + u\dot{\theta}. \quad (\text{A.10})$$

By multiplying (A.10) by $(I_3 + [u]_\times^2)$, the following expression can be obtained

$$(I_3 + [u]_\times^2) \dot{e}_\omega = u\dot{\theta} \quad (\text{A.11})$$

where (A.6) and the following properties were utilized

$$u^T u = 1 \quad u^T \dot{u} = 0. \quad (\text{A.12})$$

Likewise, by multiplying (A.10) by $-[u]_\times^2$ and then utilizing (A.12) the following expression is obtained

$$-[u]_\times^2 \dot{e}_\omega = \dot{u}\theta. \quad (\text{A.13})$$

From the expression in (A.4), the properties given in (A.5), (A.10), (A.11), (A.13), and the fact that $\sin^2 \theta = \frac{1}{2}(1 - \cos 2\theta)$, we can obtain the following expression

$$\omega_e = L_\omega^{-1} \dot{e}_\omega \quad (\text{A.14})$$

where $L_\omega(t)$ is defined in (27). After multiplying both sides of (A.14) by $L_\omega(t)$, the object kinematics given in (26) can be obtained.

To develop the open-loop error system for $e_v(t)$, the time derivative of (18) is obtained as follows

$$\dot{e}_v = \dot{p}_e - \dot{p}_e^* = \frac{1}{z_1} A_e L_v \dot{m}_1. \quad (\text{A.15})$$

After taking the time derivative of (5), $\dot{m}_1(t)$ can be determined as follows

$$\dot{m}_1 = v_e + [\omega_e]_\times R s_1. \quad (\text{A.16})$$

After utilizing (A.5) and the following property (Felippa, 2000)

$$[R s_1]_\times = R [s_1]_\times R^T, \quad (\text{A.17})$$

the expression in (A.16) can be expressed as follows

$$\dot{m}_1 = v_e - R [s_1]_\times R^T \omega_e. \quad (\text{A.18})$$

After substituting (A.18) into (A.15), the object kinematics given in (21) can be obtained.

B Proof of Non-negativity of Auxilliary Function $P(t)$

In (43), the function $P(t)$ is required to be non-negative. To prove that $P(t) \geq 0$, the definition of $r(t)$ given in (37) is substituted into the definition of $L(t)$ given in (45), and the result is integrated to obtain the following expression

$$\begin{aligned} \int_{t_0}^t L(\sigma) d\sigma &= \int_{t_0}^t \dot{\tilde{e}}^T(\sigma) \ddot{e}(\sigma) d\sigma \\ &\quad - \int_{t_0}^t \dot{\tilde{e}}^T(\sigma) \rho \text{sgn}(\tilde{e}(\sigma)) d\sigma \\ &\quad + \int_{t_0}^t \tilde{e}^T(\sigma) (\ddot{e}(\sigma) - \rho \text{sgn}(\tilde{e}(\sigma))) d\sigma. \end{aligned} \quad (\text{B.1})$$

After integrating the first integral of (B.1) by parts and evaluating the second integral, the following expression can be obtained

$$\begin{aligned} \int_{t_0}^t L(\sigma) d\sigma &= \tilde{e}^T(t) \ddot{e}(t) - \tilde{e}^T(t_0) \ddot{e}(t_0) - \sum_{i=1}^6 \rho_i |\tilde{e}_i(t)| \\ &\quad + \sum_{i=1}^6 \rho_i |\tilde{e}_i(t_0)| + \int_{t_0}^t \tilde{e}^T(\sigma) \ddot{e}(\sigma) d\sigma \\ &\quad - \int_{t_0}^t \tilde{e}^T(\sigma) (\ddot{e}(\sigma) + \rho \text{sgn}(\tilde{e}(\sigma))) d\sigma. \end{aligned} \quad (\text{B.2})$$

The expression in (B.2) can be upper bounded as follows

$$\begin{aligned} \int_{t_0}^t L(\sigma) d\sigma &\leq \sum_{i=1}^6 \rho_i |\tilde{e}_i(t_0)| - \tilde{e}^T(t_0) \ddot{e}(t_0) \\ &\quad + \sum_{i=1}^6 (|\ddot{e}_i(t)| - \rho_i) |\tilde{e}_i(t)| \\ &\quad + \int_{t_0}^t \sum_{i=1}^6 |\tilde{e}_i(\sigma)| |\ddot{e}_i(\sigma)| d\sigma \\ &\quad + \int_{t_0}^t \sum_{i=1}^6 |\tilde{e}_i(\sigma)| (|\ddot{e}_i(\sigma)| - \rho_i) d\sigma. \end{aligned} \quad (\text{B.3})$$

If ρ is chosen according to (42), then (B.3) can be upper bounded by ζ_b of (45); hence, $P(t) \geq 0$.

References

- Allen, P. K., Timcenko, A., Yoshimi, B., Michelman, P., 1992. Trajectory filtering and prediction for automated tracking and grasping of a moving object. In: Proc. IEEE Conference on Robotics and Automation. Nice, France.

- Broida, T. J., Chellappa, R., 1991. Estimating the kinematics and structure of a rigid object from a sequence of monocular images. *IEEE Transactions on PAMI* 13 (6), 497–513.
- Chen, J., Behal, A., Dawson, D., Fang, Y., 2003. 2.5d visual servoing with a fixed camera. In: *Proc. of the American Control Conference*. Nice, France.
- Dixon, W. E., Behal, A., Dawson, D., Nagarkatti, S., 2003. *Nonlinear Control of Engineering Systems: A Lyapunov-Based Approach*. Birkhäuser Boston, ISBN: 0-8176-4265-X.
- Faugeras, O., 2001. *Three-Dimensional Computer Vision*. The MIT Press, Cambridge Massachusetts.
- Faugeras, O., Lustman, F., 1988. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence* 2 (3), 485–508.
- Felippa, C. A., 2000. A systematic approach to the element-independent corotational dynamics of finite elements. Tech. rep., Center for Aerospace Structures, College of Engineering, University of Colorado., document number CU-CAS-00-03.
- Ghosh, B. K., Inaba, H., Takahashi, S., 2000. Identification of riccati dynamics under perspective and orthographic observations. *IEEE Transactions on Automatic Control* 45, 1267–1278.
- Ghosh, B. K., Jankovic, M., Wu, Y. T., 1994. Perspective problems in system theory and its application to machine vision. *Journal of Mathematical Systems, Estimation, and Control* 4 (1), 3–38.
- Ghosh, B. K., Loucks, E. P., 1996. A realization theory for perspective systems with application to parameter estimation problems in machine vision. *IEEE Transactions on Automatic Control* 41 (12), 1706–1722.
- Ghosh, B. K., Xi, N., Tarn, T. J., 1999. *Control in Robotics and Automation: Sensor-Based Integration*. Academic Press: London, UK.
- Gutman, S., 1979. Uncertain dynamical systems – a lyapunov min-max approach. *IEEE Trans. Automatic Control* 24 (3), 437–443.
- Hashimoto, K., Noritsugu, T., 1996. Observer-based control for visual servoing. In: *Proc. 13th IFAC World Congress*. San Francisco, California.
- Kanatani, K., Ohta, N., Kanazawa, Y., 2000. Optimal homography computation with a reliability measure. *IEICE Transactions on Information and Systems* E83-D(7), 1369–1374.
- Koivo, A. J., Houshang, N., 1991. Real-time vision feedback for servoing robotic manipulator with self-tuning controller. *IEEE Transactions Systems, Man, and Cybernetics* 21 (1), 134–142.
- Malis, E., 1998. *Contributions à la modélisation et à la commande en asservissement visuel*. Ph.D. thesis, University of Rennes I, IRISA, France.
- Malis, E., Chaumette, F., 2000. 2 1/2 d visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *International Journal of Computer Vision* 37 (1), 79–97.
- Malis, E., Chaumette, F., 2002. Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods. *IEEE Trans. on Robotics and Automation* 18 (2), 176–186.
- Malis, E., Chaumette, F., Bodet, S., 1999. 2 1/2 d visual servoing. *IEEE Transactions on Robotics and Automation* 15 (2), 238–250.
- Messner, W., Horowitz, R., Kao, W. W., Boals, M., 1991. A new adaptive learning rule. *IEEE Trans. Automatic Control* 36 (2), 188–197.
- Polycarpou, M. M., Ioannou, P. A., 1993. On the existence and uniqueness of solutions in adaptive control systems. *IEEE Trans. Automatic Control* 38 (3), 474–479.
- Qu, Z., 1998. *Robust Control of Nonlinear Uncertain Systems*. Wiley.
- Qu, Z., Xu, J. X., 2002. Model-based learning controls and their comparisons using lyapunov direct method. *Asian Journal of Control* 4 (1), 99–110.
- Rizzi, A. A., Koditchev, D. E., 1996. An active visual estimator for dexterous manipulation. *IEEE Trans. on Robotics and Automation* 12 (5), 697–713.
- Shariat, H., Price, K., 1990. Motion estimation with more than two frames. *IEEE Transactions PAMI* 12 (5), 417–434.
- Slotine, J. J. E., Li, W., 1991. *Applied Nonlinear Control*. Prentice Hall, Inc, Englewood Cliff, NJ.
- Spong, M. W., Vidyasagar, M., 1989. *Robot Dynamics and Control*. John Wiley and Sons, Inc, New York, NY.
- Tsai, R. Y., Huang, T. S., 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions PAMI* 6 (1), 13–26.
- Waxman, A. M., Duncan, J. H., 1986. Binocular image flows: Step toward stereo-motion fusion. *IEEE Transactions PAMI* 8 (6).
- Xian, B., De Queiroz, M. S., Dawson, D. M., 2004. *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Heidelberg, Germany, Ch. A Continuous Control Mechanism for Uncertain Nonlinear Systems, Optimal Control, Stabilization, and Nonsmooth Analysis, to appear.
- Zhang, Z., Hanson, A. R., 1995. Scaled euclidean 3d reconstruction based on externally uncalibrated cameras. In: *IEEE Symp. on Computer Vision*.