Identification of a Moving Object's Velocity with a Fixed Camera*

V. K. Chitrakaran†, D. M. Dawson†, W. E. Dixon‡, and J. Chen†

†Department of Electrical & Computer Engineering, Clemson University, Clemson, SC 29634-0915

‡Eng. Science and Tech. Div. - Robotics, Oak Ridge Nat. Lab., P.O. Box 2008, Oak Ridge, TN 37831-6305

E-mail: dixonwe@ornl.gov

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.

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 [7], 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 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 [9]. In [21], a window position predictor for object tracking was utilized. In [11], an observer for estimating the object velocity was utilized; however, a description of the object's kinematics must be known. In [8], the problem of identifying the motion and shape parameters of a planar object undergoing Riccati motion was examined in great detail. In [12], an autoregressive discrete-time model is used to predict the location of features of a moving object. In [1], trajectory filtering and prediction techniques are utilized to track a moving object. Some of the work [25] 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 [2] and [22], object-centered models are utilized to estimate the translation

*This research was supported in part U.S. DOE Office of Biological and Environmental Research (OBER) Environmental Management Sciences Program (EMSP) project ID No. 82797 at ORNL, a subcontract to ORNL by the Florida Department of Citrus, and by U.S. NSF Grant DMI-9457967, ONR Grant N00014-99-1-0589, a DOC Grant, and an ARO Automotive Center Grant.

and the center of rotation of the object. In [26], 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

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

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 [10], 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 [23] 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 [23], 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 (e.g., see [17] and [19]).

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 goaded problem, we fuse homography-based techniques with a Lyapunov synthesized estimator to asymptotically identify the object's unknown velocity[1]. 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 [3] which relies on the camera-in-hand work presented in [14]. The continuous, Lyapunov-based estimation strategy has its roots in an example developed in [20] and the general framework developed in [27]. 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.

The paper is organized in the following manner. In Section 2, the geometric model for the identification problem is presented. In Section 3, we illustrate how a homography-based approach can be used to obtain Euclidean information about the moving object from a single fixed camera. In Section 4, we derive the object's kinematics based on the Euclidean reconstruction of the moving object. In Section 5, we develop the Lyapunov-based estimator for identifying the translational and rotational velocity of the object. Some concluding remarks are made in Section 6.

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

2 Geometric Model

To facilitate the subsequent object velocity identification problem, four target points located on an object denoted by O_i $\forall i = 1, 2, 3, 4$ are considered to be coplanar[2] 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 $\mathcal{I}$

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

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

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

$$d^* = n^{*T} \bar{m}_i^*. \tag{4}$$

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

$$\bar{m}_i = x_f + R s_i \tag{5}$$

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

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

$$\bar{m}_i = \bar{x}_f + \bar{R} \bar{m}_i^* \tag{7}$$

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

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

See Appendix A for further insight into the geometric significance of the new rotational and translational variables. From (4), it is easy to see how the relationship in (7) can now be expressed as follows

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

[2]It should be noted that that if that if four coplanar target points are not available then the subsequent development can exploit the classic eight-points algorithm [15] with no four of the eight target points being coplanar.

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.).*

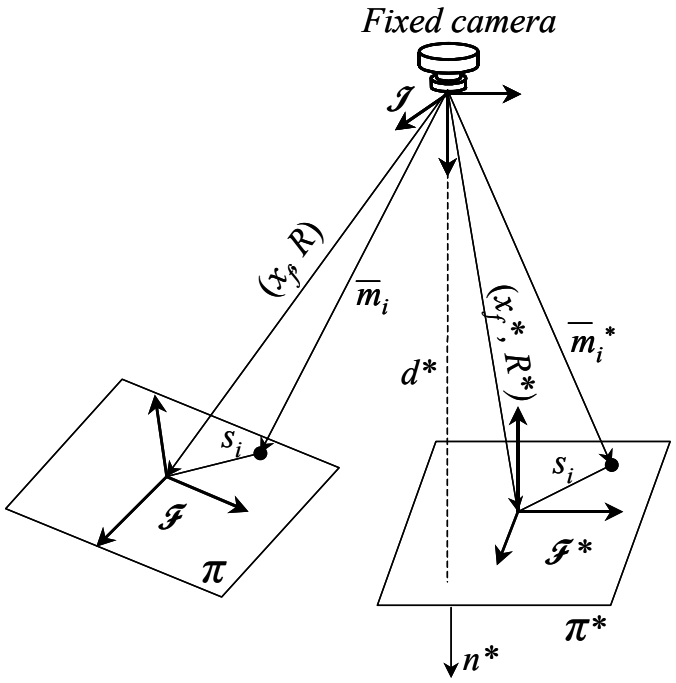

Figure 1: Coordinate frame relationships.

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

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

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

The rotation and translation between the normalized coordinates can now be related through a Euclidean homography, denoted by $H(t) \in \mathbb{R}^{3\times 3}$, as follows

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

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

$$\bar{x}_h = \frac{\bar{x}_f}{d^*}. \tag{13}$$

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

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

The projected pixel coordinates of the target points are related to the normalized task-space coordinates by the following pin-hole lens models [4]

$$p_i = Am_i \qquad p_i^* = Am_i^* \tag{15}$$

where $A \in \mathbb{R}^{3\times3}$ is a known, constant, and invertible intrinsic camera calibration matrix. After substituting (15) into (12), the following relationship can be developed

$$p_i = \alpha_i \underbrace{\left(AHA^{-1}\right)}_{G} p_i^* \tag{16}$$

where $G(t) = [g_{ij}(t)]\ \forall i, j = 1, 2, 3 \in \mathbb{R}^{3\times3}$ 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

$$p_i = \alpha_i g_{33} G_n p_i^* \tag{17}$$

where $G_n(t) \in \mathbb{R}^{3\times3}$ 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 (*e.g.*, see [5, 28]), the 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 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

$$e_v = p_e - p_e^*. \tag{18}$$

In (18), $p_e(t) \in \mathbb{R}^3$ denotes the following extended image coordinates [14] of an image point[3] on π in terms of the inertial coordinate system $\mathcal{I}$

$$p_e = \left[\begin{array}{ccc} u_1 & v_1 & \ln(z_1) \end{array} \right]^T \tag{19}$$

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 $\mathcal{I}$

$$p_e^* = \left[\begin{array}{ccc} u_1^* & v_1^* & \ln(z_1^*) \end{array} \right]^T . \tag{20}$$

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 B for further details)

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

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

$$A_e = A - \left[\begin{array}{ccc} 0 & 0 & u_0 \\ 0 & 0 & v_0 \\ 0 & 0 & 0 \end{array} \right] \tag{22}$$

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

$$L_v = \left[\begin{array}{ccc} 1 & 0 & -\dfrac{x_1}{z_1} \\ 0 & 1 & -\dfrac{y_1}{z_1} \\ 0 & 0 & 1 \end{array} \right] . \tag{23}$$

To develop the rotation kinematics for the object, $e_\omega(t) \in \mathbb{R}^3$ is defined using the angle axis representation [24] to quantify the rotation of $\mathcal{F}$ with respect to the fixed coordinate system $\mathcal{F}^*$ as follows

$$e_\omega \triangleq u(t)\theta(t) . \tag{24}$$

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

$$-\pi < \theta(t) < \pi. \tag{25}$$

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

$$\dot{e}_\omega = L_\omega \omega_e . \tag{26}$$

[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.

In (26), the Jacobian-like matrix $L_\omega(t) \in \mathbb{R}^{3\times3}$ is defined as

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

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

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

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 observer strategy that is robust to these parameters. Since the intrinsic calibration matrix is assumed to be known in this paper, the observer 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 [24], 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 [24]*

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

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

$$0 \leq \theta_p(t) \leq \pi. \tag{29}$$

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

$$u_p\theta_p = u\theta. \tag{30}$$

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 [16]*

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

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 observer that can be used to identify the translational and rotational velocity of an object expressed in $\mathcal{I}$, denoted by $v(t) = \left[\begin{array}{cc} v_e & \omega_e \end{array} \right]^T \in \mathbb{R}^6$. To facilitate this objective, the object kinematics are expressed in the following compact form

$$\dot{e} = Jv \tag{32}$$

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

$$J = \left[\begin{array}{cc} \dfrac{\alpha_1}{z_1^*} A_e L_v & -\dfrac{\alpha_1}{z_1^*} A_e L_v R \left[s_1 \right]_\times R^T \\ 0 & L_\omega \end{array} \right] \tag{33}$$

where (21) and (26) were utilized, and $e(t) = \left[\begin{array}{cc} e_v^T & e_\omega^T \end{array} \right]^T \in \mathbb{R}^6$. 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 $\ddot{e}(t)$, $\dddot{e}(t) \in L_\infty$; hence, we can use the previous assumptions to show that

$$\sum_{i=1}^{6} |\ddot{e}_i(t)| \leq \beta_1, \quad \sum_{i=1}^{6} |\dddot{e}_i(t)| \leq \beta_2 \tag{34}$$

where β_1, $\beta_2 \in \mathbb{R}$ are known positive constants.

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

$$\lim_{t\to\infty} \left\| \tilde{e}(t) \right\|, \ \left\| \dot{\tilde{e}}(t) \right\| = 0 \tag{35}$$

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

$$\tilde{e} = e - \hat{e} \tag{36}$$

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 Observer Development

To facilitate the following analysis, we define a filtered observation error, denoted by $r(t) \in \mathbb{R}^6$, as follows [23]

$$r = \dot{\tilde{e}} + \tilde{e} \ . \tag{37}$$

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

$$\dot{r} = \ddot{e} - \ddot{\hat{e}} + \dot{\tilde{e}}. \tag{38}$$

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

$$\dot{\hat{e}} = \hat{\kappa} \tag{39}$$

where $\hat{\kappa}(t) \in \mathbb{R}^6$ is defined as follows[4]

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

where $K \in \mathbb{R}^{6\times6}$ is a positive constant diagonal gain matrix, $\rho \in \mathbb{R}$ is a positive constant, $I_6 \in \mathbb{R}^{6\times6}$ 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

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

where the time derivative of (40) was utilized.

5.3 Analysis

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

$$\rho > \beta_1 + \beta_2 \tag{42}$$

where β_1 and β_2 are introduced in (34).

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

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

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

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

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

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

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

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

[4]The structure of the observer is motivated by the previous work in [27] that was inspired by an example in [20].

The expression in (46) can be rewritten as follows

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

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

$$\dot{V} \leq -K\|r\|^2 \ , \tag{48}$$

it is clear from (43) that $r(t), \tilde{e}(t),\ P(t) \in L_\infty$ and that $r(t) \in L_2$ [18]. Based on the fact that $r(t) \in L_\infty$, linear analysis techniques [18] can be used to determined 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 [23] can be used to prove that

$$\lim_{t \to \infty} \|r(t)\| = 0 \ . \tag{49}$$

Given (49), linear analysis techniques [18] can be used to determine that the result in (35) is obtained. □

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

$$\lim_{t \to \infty} \dot{\hat{e}}_i(t) = \dot{e}_i(t) \quad \forall i = 1, 2, ...6. \tag{50}$$

Based on (50), the definition in (39) can be used to conclude that

$$\lim_{t \to \infty} \hat{\kappa}_i(t) = \dot{\hat{e}}_i(t) \quad \forall i = 1, 2, ...6. \tag{51}$$

Hence, (32) and (51) can be used to conclude that

$$\lim_{t \to \infty} \hat{\kappa}_i(t) = (Jv)_i \quad \forall i = 1, 2, ...6. \tag{52}$$

From the result in (52), it is clear that the object velocity can be identified provided $J(t)$ of (33) is known and is invertible. Based on the fact that $\alpha_1(t)$, $L_v(t)$, and $L_\omega(t)$ are known signals obtained from the homography decomposition, the assumption that A_e of (22) is known (i.e., the camera is assumed to be calibrated), and the assumption that s_1 is known (i.e., a partial object model is known), it is clear from (33) that 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, the measureable auxiliary signals $\gamma_1(t)$, $\gamma_2(t)$, $\gamma_3 \in \mathbb{R}^3$ are defined as follows

$$\gamma_1 = n^{*T} m_1^* \bar{x}_h \tag{53}$$

$$\gamma_2 = \frac{m_1}{\alpha_1} \tag{54}$$

$$\gamma_3 = m_1^* \ . \tag{55}$$

Based on (4)-(6), (8), (10), (11), and (13), the auxiliary functions in (53)-(55) can be rewritten as follows

$$\gamma_1 = \frac{x_f}{z_1^*} - \bar{R}\frac{x_f^*}{z_1^*} \tag{56}$$

$$\gamma_2 = \frac{x_f}{z_1^*} + R\frac{s_1}{z_1^*} \tag{57}$$

$$\gamma_3 = \frac{x_f^*}{z_1^*} + R^*\frac{s_1}{z_1^*} \ . \tag{58}$$

To facilitate further analysis, the auxiliary variables $\bar{\gamma}_1(t)$, $\bar{\gamma}_2(t) \in \mathbb{R}^3$ are defined as follows as

$$\bar{\gamma}_1 = \left[diag\left(\gamma_3\right)\right]^{-1}\gamma_1 \qquad \bar{\gamma}_2 = \left[diag\left(\gamma_3\right)\right]^{-1}\gamma_2 \tag{59}$$

where the notation $diag\left(\xi\right)$ denotes a diagonal matrix with the i^{th} diagonal element equal to the i^{th} element of a vector ξ. After substituting (56)-(58) into (59) and multiplying through by z_1^*, the following expressions can be obtained

$$diag\left(\bar{\gamma}_1\right)\left[x_f^* + R^* s_1\right] = x_f - \bar{R}x_f^* \tag{60}$$

$$diag\left(\bar{\gamma}_2\right)\left[x_f^* + R^* s_1\right] = x_f + Rs_1. \tag{61}$$

After subtracting (61) from (60), the following expression can be obtained

$$diag(\bar{\gamma}_1 - \bar{\gamma}_2)\left(x_f^* + R^* s_1\right) = -\left(\bar{R}x_f^* + Rs_1\right). \tag{62}$$

The expression in (62) can the be solved for x_f^* in terms of measurable signals as follows

$$x_f^* = \left[diag(\bar{\gamma}_1 - \bar{\gamma}_2) + \bar{R}\right]^{-1}\left[diag(\bar{\gamma}_1 - \bar{\gamma}_2)R^* - R\right]s_1. \tag{63}$$

Since γ_3 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^*. Therefore, all the elements of $J(t)$ can either be determined or are assumed to be known. After utilizing (23), (31), and the facts that z_1^*, $\alpha_1(t) > 0$, the definition given in (33) can be used to conclude that $J(t)$ is invertible. Since $J(t)$ is composed of known terms and is invertible, then (52) can be used to conclude that the object velocity can be identified given the result of (35).

Remark 5 *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 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 objects are that its velocity and its first two time derivatives be bounded, and that a single geometric length between two feature points on the object be known a priori. Future work will concentrate on experimental validation of the proposed estimators as well as its ramifications for other vision-based applications. Specifically, it seems that the proposed estimator might be able to be utilized in a typical camera-in-hand application that requires a robot manipulator end-effector to track a moving object. The applicability of the proposed approach for this type of object tracking applications is well motivated since the estimator does not require an explicit model for describing the movement of the object.

References

[1] P. K. Allen, A. Timcenko, B. Yoshimi, and P. Michelman, "Trajectory Filtering and Prediction for Automated Tracking and Grasping of a Moving Object, *Proceedings IEEE Conference on Robotics and Automation*, Nice, France, 1992, pp. 1850-1856.

[2] T J. Broida and R. Chellappa, "Estimating the Kinematics and Structure of a Rigid Object from a Sequence of Monocular Images, *IEEE Transactions PAMI*, Vol. 13, No. 6, pp. 497-513, 1991.

[3] J. Chen, A. Behal, D. Dawson, and Y. Fang, "2.5D Visual Servoing with a Fixed Camera", *Proc. of the American Control Conference*, June 2003, accepted, to appear.

[4] O. Faugeras, *Three-Dimensional Computer Vision*, The MIT Press, Cambridge Massachusetts, 2001.

[5] O. Faugeras and F. Lustman, "Motion and Structure From Motion in a Piecewise Planar Environment", *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 2, No. 3, pp. 485-508, 1988.

[6] C. A. Felippa, *A Systematic Approach to the Element-Independent Corotational Dynamics of Finite Elements*, Center for Aerospace Structures Document Number CU-CAS-00-03, College of Engineering, University of Colorado, January 2000.

[7] B. K. Ghosh, M. Jankovic, and Y. T. Wu, "Perspective Problems in System Theory and its Application to Machine Vision," *Journal of Mathematical Systems, Estimation, and Control*, Vol. 4, No. 1, pp. 3-38, 1994.

[8] B. K. Ghosh, H. Inaba, and S. Takahashi, "Identification of Riccati Dynamics Under Perspective and Orthographic Observations," *IEEE Transactions on Automatic Control*, Vol. 45, pp. 1267-1278, July 2000.

[9] B. K. Ghosh and E. P. Loucks, "A Realization Theory for Perspective Systems with Application to Parameter Estimation Problems in Machine Vision, *IEEE Transactions on Automatic Control*, Vol. 41, No. 12, pp. 1706-1722, 1996.

[10] B. K. Ghosh, N. Xi, and T. J. Tarn, *Control in Robotics and Automation: Sensor-Based Integration*, pp. 74-75, Academic Press: London, UK, 1999.

[11] K. Hashimoto and T. Noritsugu, "Observer-based Control for Visual Servoing, *Proc. 13th IFAC World Congress*, Vol. F. pp 453- 458, San Francisco, 1996.

[12] A. J. Koivo and N. Houshangi, "Real-Time Vision Feedback for Servoing Robotic Manipulator with Self-Tuning Controller, *IEEE Transactions Systems, Man, and Cybernetics*, Vol 21, No, 1 pp. 134-142, 1991.

[13] E. Malis, "Contributions à la modélisation et à la commande en asservissement visuel", Ph.D. Dissertation, University of Rennes I, IRISA, France, Nov. 1998.

[14] E. Malis, F. Chaumette, and S. Bodet, "2 1/2 D Visual Servoing," *IEEE Transactions on Robotics and Automation*, Vol. 15, No. 2, pp. 238-250, April 1999.

[15] E. Malis and F. Chaumette, "2 1/2 D Visual Servoing with Respect to Unknown Objects Through a New Estimation Scheme of Camera Displacement," *International Journal of Computer Vision*, Vol. 37, No. 1, pp. 79-97, June 2000.

[16] E. Malis and F. Chaumette, "Theoretical Improvements in the Stability Analysis of a New Class of Model-Free Visual Servoing Methods," *IEEE Transactions on Robotics and Automation*, Vol. 18, No. 2, pp. 176-186, April 2002.

[17] W. Messner, R. Horowitz, W.W. Kao, and M. Boals, "A New Adaptive Learning Rule," *IEEE Trans. Automatic Control*, Vol. 36, No. 2, pp. 188-197, Feb. 1991.

[18] M. de Queiroz, D. Dawson, S. Nagarkatti, and F. Zhang, *Lyapunov-based Control of Mechanical Systems*, Birkhauser, New York, 2000.

[19] Z. Qu, *Robust Control of Nonlinear Uncertain Systems*, Wiley, 1998.

[20] Z. Qu and J.-X. Xu, "Model-Based Learning Controls and Their Comparisons Using Lyapunov Direct Method," *Asian Journal of Control*, Vol. 4, No. 1, pp. 99-110, March 2002.

[21] A. A. Rizzi and D. E. Koditchek, "An Active Visual Estimator for Dexterous Manipulation", *Trans. on Robotics and Automation*, Vol. 12, No. 5, pp. 697-713, 1996.

[22] H. Shariat and K. Price, "Motion Estimation with More Than Two Frames", *IEEE Transactions PAMI*, Vol. 12, No. 5, pp. 417-434, 1990.

[23] J. J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice Hall, Inc: Englewood Cliff, NJ, 1991.

[24] M. W. Spong and M. Vidyasagar, *Robot Dynamics and Control*, John Wiley and Sons, Inc: New York, NY, 1989.

[25] R. Y. Tsai and T S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, *IEEE Transactions PAMI*, Vol.6 No. 1, pp. 13-26, 1984.

[26] A. M. Waxman and J. H. Duncan, "Binocular Image Flows: Step Toward Stereo-Motion Fusion", *IEEE Transactions PAMI*, Vol. 8, No. 6, 1986.

[27] B. Xian, D.M. Dawson, M.S. de Queiroz, and J. Chen, "A Continuous Asymptotic Tracking Control Strategy for Uncertain Multi-Input Nonlinear Systems," submitted to the *IEEE Trans. Automatic Control*, Oct. 2002.

[28] Z. Zhang and A. R. Hanson, "Scaled Euclidean 3D Reconstruction Based on Externally Uncalibrated Cameras," *IEEE Symp. on Computer Vision*, pp. 37-42, 1995.

Appendix A

To obtain geometric insight into the structure of $\bar{R}(t)$ and $\bar{x}_f(t)$ defined in (8) can be obtained from Figure 1 by placing a fictitious camera that has a frame $\mathcal{I}^*$attached to its center such that $\mathcal{I}^*$ initially coincides with $\mathcal{I}$. Since $\mathcal{I}$ and $\mathcal{I}^*$ coincide, the relationship between $\mathcal{I}^*$ and $\mathcal{F}^*$ can be denoted by rotational and translational parameters (x_f^*, R^*) as is evident from Figure 1. Without relative translational or rotational motion between $\mathcal{I}^*$ and $\mathcal{F}^*$, the two coordinate frames are moved

until $\mathcal{F}^*$ aligns with $\mathcal{F}$, resulting in Figure 2. It is now evident that the fixed camera problem reduces to a stereo vision problem with the parameters $\left(x_f - R\left(R^*\right)^T x_f^*, R\left(R^*\right)^T\right)$ denoting the translation and rotation between $\mathcal{I}$ and $\mathcal{I}^*$.

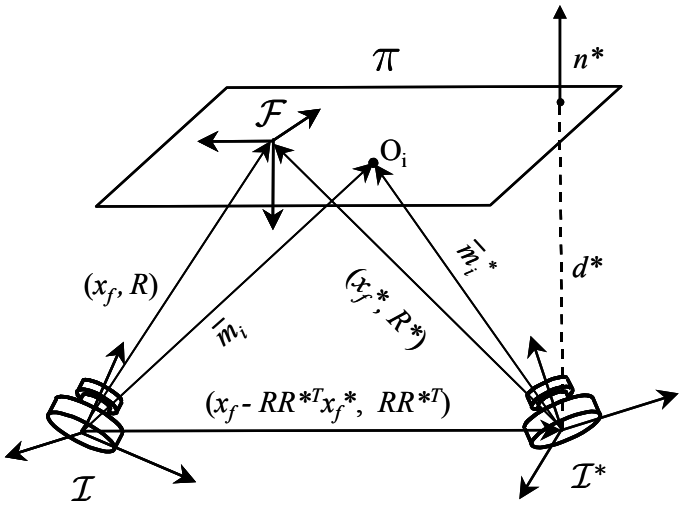

Figure 2: Geometric relationships for $\bar{R}(t)$ and $\bar{x}_f(t)$.

Appendix B

Based on the previous definitions for $\omega_e(t)$ and $R(t)$, the following property can be determined [24]

$$[\omega_e]_\times = \dot{R}R^T. \tag{64}$$

From (8) and (64), the following relationship can be determined

$$[\omega_e]_\times = \dot{\bar{R}}\bar{R}^T. \tag{65}$$

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 [24]

$$\bar{R} = I_3 + \sin\theta\,[u]_\times + 2\sin^2\frac{\theta}{2}[u]_\times^2 \tag{66}$$

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 (65) can be used along with (66) and the time derivative of (66), to obtain the following expression

$$[\omega_e]_\times = \sin\theta\,[\dot{u}]_\times + [u]_\times\dot{\theta} + (1-\cos\theta)\left[[u]_\times\dot{u}\right]_\times \tag{67}$$

where the following properties were utilized [6], [13]

$$\begin{aligned}
[u]_\times\zeta &= -[\zeta]_\times u, && (68)\\
[u]_\times^2 &= uu^T - I_3, && (69)\\
[u]_\times uu^T &= 0 && (70)\\
[u]_\times[\dot{u}]_\times[u]_\times &= 0, && (71)\\
\left[[u]_\times\dot{u}\right]_\times &= [u]_\times[\dot{u}]_\times - [\dot{u}]_\times[u]_\times. && (72)
\end{aligned}$$

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

$$\dot{e}_\omega = \dot{u}\theta + u\dot{\theta}. \tag{73}$$

By multiplying (73) by $\left(I_3 + [u]_\times^2\right)$, the following expression can be obtained

$$\left(I_3 + [u]_\times^2\right) \dot{e}_\omega = u\dot{\theta} \tag{74}$$

where (69) and the following properties were utilized

$$u^T u = 1 \qquad u^T \dot{u} = 0 \ . \tag{75}$$

Likewise, by multiplying (73) by $-[u]_\times^2$ and then utilizing (75) the following expression is obtained

$$-[u]_\times^2 \dot{e}_\omega = \dot{u}\theta. \tag{76}$$

From the expression in (67), the properties given in (68), (73), (74), (76), and the fact that

$$\sin^2\theta = \frac{1}{2}\left(1 - \cos 2\theta\right),$$

can be used to obtain the following expression

$$\omega_e = L_\omega^{-1} \dot{e}_\omega \tag{77}$$

where $L_\omega(t)$ is defined in (27). After multiplying both sides of (77) 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

$$\dot{e}_v = \dot{p}_e - \dot{p}_e^* = \frac{1}{z_1} A_e L_v \dot{\bar{m}}_1. \tag{78}$$

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

$$\dot{\bar{m}}_1 = v_e + [\omega_e]_\times R s_1 \ . \tag{79}$$

After utilizing (68) and the following property [6]

$$[R s_1]_\times = R [s_1]_\times R^T \ , \tag{80}$$

the expression in (79) can be expressed as follows

$$\dot{\bar{m}}_1 = v_e - R [s_1]_\times R^T \omega_e \ . \tag{81}$$

After substituting (81) into (78), the object kinematics given in (21) can be obtained.

Appendix C

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

$$\begin{aligned} \int_{t_0}^{t} L(\sigma)\, d\sigma &= \int_{t_0}^{t} \dot{\tilde{e}}^T(\sigma)\, \ddot{e}(\sigma)\, d\sigma - \rho \int_{t_0}^{t} \dot{\tilde{e}}^T(\sigma) \operatorname{sgn}(\tilde{e}(\sigma)) d\sigma \\ &\quad + \int_{t_0}^{t} \tilde{e}^T(\sigma) \left(\ddot{e}(\sigma) - \rho \operatorname{sgn}(\tilde{e}(\sigma))\right) d\sigma. \end{aligned} \tag{82}$$

After integrating the first integral of (82) by parts and evaluating the second integral, the following expression can be obtained

$$\begin{aligned}\int_{t_0}^{t} L(\sigma)\, d\sigma &= \tilde{e}^T(t)\ddot{e}(t) - \tilde{e}^T(t_0)\ddot{e}(t_0) - \rho \sum_{i=1}^{6} \left|\tilde{e}_i^T(t)\right| + \rho \sum_{i=1}^{6} \left|\tilde{e}_i^T(t_0)\right| \\ &\quad + \int_{t_0}^{t} \tilde{e}^T(\sigma)\left(\ddot{e}(\sigma) - \dddot{e}(\sigma) - \rho \mathrm{sgn}(\tilde{e}(\sigma))\right) d\sigma\ .\end{aligned} \tag{83}$$

The expression in (83) can be upper bounded as follows

$$\begin{aligned}\int_{t_0}^{t} L(\sigma)\, d\sigma &\leq \rho \sum_{i=1}^{6} \left|\tilde{e}_i^T(t_0)\right| - \tilde{e}^T(t_0)\ddot{e}(t_0) + \sum_{i=1}^{6} \left(|\ddot{e}_i(t)| - \rho\right) \left|\tilde{e}_i^T(t)\right| \\ &\quad + \int_{t_0}^{t} \sum_{i=1}^{6} \left|\tilde{e}_i^T(\sigma)\right| \left(|\ddot{e}_i(\sigma)| + |\dddot{e}_i(\sigma)| - \rho\right) d\sigma.\end{aligned} \tag{84}$$

If ρ is chosen according to (42), then (84) can be upper bounded by ζ_b of (45); hence, $P(t) \geq 0$.