2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Euclidean Position Estimation of Features on a Moving Object Using a Single Camera: A Lyapunov-Based Approach[1]

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

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

‡Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville, FL 32611

E-mail: cvilas, ddawson, jianc@ces.clemson.edu; wdixon@ufl.edu

Abstract

In this paper, an adaptive nonlinear estimator is developed to identify the Euclidean coordinates of feature points on a moving object using a single fixed camera. No explicit model is used to describe the movement of the object. Homography-based techniques are used in the development of the object kinematics, while Lyapunov design methods are utilized in the synthesis of the adaptive estimator. Simulation results are included to demonstrate the performance of the estimator.

1 Introduction

The recovery of Euclidean coordinates of feature points on a moving object from a sequence of images is a mainstream research problem with significant potential impact for applications such as autonomous vehicle/robotic guidance, navigation, path planning and control. It bears a close resemblance to the classical problem in computer vision, known as "Structure from Motion (SFM)", which is the determination of 3D structure of a scene from its 2D projections on a moving camera. Although the problem is inherently nonlinear, typical SFM results are based on linearization based methods such as extended Kalman filtering [1, 6, 18]. In recent publications, some researchers have recast the problem as state estimation of a continuous-time perspective dynamic system, and have employed nonlinear system analysis tools in the development of state observers that identify motion and structure parameters [13, 14]. To summarize, these papers show that if the velocity of the moving object (or camera) is known, and satisfy certain obervability conditions, an estimator for the unknown Euclidean position of the feature points can be developed. In [3], an observer for the estimation of camera motion was presented based on perspective observations of a single feature point from the (single) moving camera. The observer development was based on sliding mode and adaptive control techniques, and it was shown that upon satisfaction of a persistent excitation condition [20], the rotational velocity could be fully recovered, and the translational velocity could be recovered upto a scale factor. The depth ambiguity attributed to the unknown scale factor was resolved by resorting to stereo vision. The afore-mentioned approach requires that a model for object motion be known.

In this paper, we present a unique nonlinear estimation strategy to simultaneously estimate the velocity and structure of a moving object using a single camera. Roughly speaking, satisfaction of a persistent excitation condition (similar to [3] and others) allows the determination of the inertial coordinates for all the feature points on the object. A homography-based approach is utilized to develop the object kinematics in terms of reconstructed Euclidean information and image-space information for the fixed camera system. The development of object kinematics relies on the work presented in [2], and requires a priori knowledge of a single geometric length between two feature points on the object. A novel nonlinear integral feedback estimation method developed in our previous efforts [5] is then employed to identify the linear and angular velocity of the moving object. Identifying the velocities of the object facilitates the development of a measurable error system that can be used to formulate a nonlinear least squares adaptive update law. A Lyapunov-based analysis is then presented that indicates if a persistent excitation condition is satisfied then the time-varying Euclidean coordinates of each feature point can be determined.

While the problem of estimating the motion and Euclidean position of features on a moving object is addressed in this paper by using a fixed camera system, the development can also be recast for the camera-in-hand problem where a moving camera observes stationary objects. That is, by recasting the problem for the camera-in-hand, the development in this paper can also be used to address the Simultaneous Localization and Mapping (SLAM) problem [8], where the information gathered from a moving camera is utilized to estimate both the motion of the camera (and hence, the relative position of the vehicle/robot) as well as position of static features in the environment.

2 Geometric Model

In order to develop a geometric relationship between the fixed camera and the moving object, we define an orthogonal coordinate frame, denoted by $\mathcal{F}$, attached to the object and

[1]This work was supported in part by two DOC Grants, an ARO Automotive Center Grant, a DOE Contract, a Honda Corporation Grant, and a DARPA Contract.

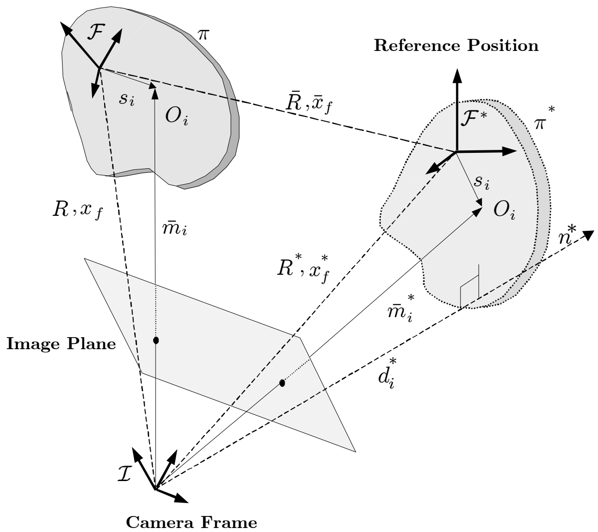

Figure 1: Geometric relationships between a fixed camera and the current and reference positions of a moving object in its field of view.

an inertial coordinate frame, denoted by $\mathcal{I}$, whose origin coincides with the optical center of the fixed camera (see Figure 1). Let the 3D coordinates of the i^{th} feature point on the object be denoted as the constant $s_i \in \mathbb{R}^3$ relative to the object reference frame $\mathcal{F}$, and $\bar{m}_i(t) \in \mathbb{R}^3$ relative to the inertial coordinate system $\mathcal{I}$, such that

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

It is assumed that the object is always in the field of view of the camera, and hence the distances from the origin of $\mathcal{I}$ to all the feature points remain positive (i.e., $z_i(t) > \varepsilon$, where ε is an arbitrarily small positive constant). To relate the coordinate systems, let $R(t) \in SO(3)$ and $x_f(t) \in \mathbb{R}^3$ denote the rotation and translation, respectively, between $\mathcal{F}$ and $\mathcal{I}$. Also, let three of the non-collinear feature points on the object, denoted by O_i $\forall i = 1, 2, 3$, define the plane π shown in Figure 1. Now consider the object to be at some fixed reference position and orientation, denoted by $\mathcal{F}^*$, as defined by a reference image of the object. We can similarly define the constant terms $\bar{m}_i^*$, R^* and x_f^*, and the plane π^* for the object at the reference position. From the geometry between the coordinate frames depicted in Figure 1, the following relationships can be developed

$$\begin{aligned} \bar{m}_i &= x_f + Rs_i & (2)\\ \bar{m}_i^* &= x_f^* + R^* s_i\ . & (3) \end{aligned}$$

After solving (3) for s_i and then substituting the resulting expression into (2), we have

$$\bar{m}_i = \bar{x}_f + \bar{R}\bar{m}_i^* \tag{4}$$

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

$$\bar{R} = R\,(R^*)^T \qquad \bar{x}_f = x_f - \bar{R}x_f^*. \tag{5}$$

It is evident from (5) that $\bar{R}(t)$ and $\bar{x}_f(t)$ quantify the rotation and translation, respectively, between the frames $\mathcal{F}$ and $\mathcal{F}^*$. As also illustrated in Figure 1, $n^* \in \mathbb{R}^3$ denotes the constant normal to the plane π^* expressed in the coordinates of $\mathcal{I}$, and the constant projections of $\bar{m}_i^*$ along the unit normal n^*, denoted by $d_i^* \in \mathbb{R}$ are given by

$$d_i^* = n^{*T}\bar{m}_i^*. \tag{6}$$

Using (6), it can be easily seen that the relationship in equation (4) can now be expressed as follows

$$\bar{m}_i = \underbrace{\left(\bar{R} + \frac{\bar{x}_f}{d_i^*}n^{*T}\right)}_{H}\bar{m}_i^* \tag{7}$$

where $H(t) \in \mathbb{R}^{3\times3}$ denotes a Euclidean homography [11].

Since a video camera is our sensing device, we must develop a geometric relationship between the 3D world in which the moving object resides and its 2D projection in the image plane of the camera. To this end, we define normalized Euclidean coordinates, denoted by $m_i(t), m_i^* \in \mathbb{R}^3$ for the feature points as follows

$$m_i \triangleq \frac{\bar{m}_i}{z_i} \qquad m_i^* \triangleq \frac{\bar{m}_i^*}{z_i^*}. \tag{8}$$

As seen by the camera, each of these feature points have projected pixel coordinates, denoted by $p_i(t), p_i^* \in \mathbb{R}^3$, expressed relative to $\mathcal{I}$ as follows

$$p_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T \qquad p_i^* = \begin{bmatrix} u_i^* & v_i^* & 1 \end{bmatrix}^T. \tag{9}$$

The projected pixel coordinates of the feature points are related to the normalized Euclidean coordinates by the pinhole model of [10] such that

$$p_i = Am_i \qquad p_i^* = Am_i^* \tag{10}$$

where $A \in \mathbb{R}^{3\times3}$ is a known, constant, upper triangular and invertible intrinsic camera calibration matrix [17]. From (7), (8) and (10), the relationship between image coordinates of the corresponding feature points in $\mathcal{F}$ and $\mathcal{F}^*$ can be expressed as follows

$$p_i = \underbrace{\frac{z_i^*}{z_i}}_{\alpha_i}\ \underbrace{A\left(\bar{R} + \bar{x}_{hi}(n^*)^T\right)A^{-1}}_{G}p_i^* \tag{11}$$

where $\alpha_i \in \mathbb{R}$ denotes the depth ratio, and $\bar{x}_{hi}(t) = \dfrac{\bar{x}_f(t)}{d_i^*} \in \mathbb{R}^3$ denotes the scaled translation vector. The matrix $G(t) \in \mathbb{R}^{3\times3}$ defined in (11) is a full rank homogeneous collineation matrix defined upto a scale factor [17]. If the structure of the moving object is planar, all feature points lie on the same plane, and hence the distances d_i^* defined in (6) is the same for all feature points, henceforth denoted as d^*. In this case, the collineation $G(t)$ is defined upto the same scale factor, and hence, one of its elements can be set to unity without loss of generality. $G(t)$ can then be estimated from a set of linear equations (11) obtained from at least four corresponding feature points that are coplanar but non-collinear. If the structure of the object is not planar, the

Virtual Parallax method described in [17] could be utilized. An overview of the determination of the collineation matrix $G(t)$ and the depth ratios $\alpha_i(t)$ using both the methods are given in Appendix A. Based on the fact that the intrinsic camera calibration A is known apriori, we can then determine the Euclidean homography $H(t)$. By utilizing various techniques (see algorithms in [11, 22]), $H(t)$ can be decomposed into its constituent rotation matrix $\bar{R}(t)$, unit normal vector n^*, scaled translation vector $\bar{x}_h(t) \triangleq \frac{\bar{x}_f(t)}{d^*}$ and the depth ratio $\alpha_i(t)$. It is assumed that the constant rotation matrix R^* is known. $R(t)$ can therefore be computed from (5). Hence $R(t), \bar{R}(t), \bar{x}_h(t)$ and $\alpha_i(t)$ are known signals that can be used in the subsequent analysis.

Remark 1 *The subsequent development requires that the constant rotation matrix R^* be known.*

3 Object Kinematics

To quantify the translation of $\mathcal{F}$ relative to the fixed coordinate system $\mathcal{F}^*$, we define $e_v(t) \in \mathbb{R}^3$ in terms of the image coordinates of the feature point O_1 as follows

$$e_v \triangleq \begin{bmatrix} u_1 - u_1^* & v_1 - v_1^* & -\ln(\alpha_1) \end{bmatrix}^T. \qquad (12)$$

In (12) and in the subsequent development, any point O_i on π could have been utilized; however, to reduce the notational complexity, we have elected to select the feature point O_1. The signal $e_v(t)$ is measurable since the first two elements of the vector are obtained from the images and the last element is available from known signals as discussed in the previous section. Following the development in [5], the translational kinematics can be obtained as follows

$$\dot{e}_v = \frac{\alpha_1}{z_1^*} A_{e1} R \left[v_e - [s_1]_\times \omega_e \right] \qquad (13)$$

where the notation $[s_1]_\times$ denotes the 3×3 skew symmetric form of s_1, $v_e(t)$, $\omega_e(t) \in \mathbb{R}^3$ denote the *unknown* linear and angular velocity of the object expressed in the local coordinate frame $\mathcal{F}$, respectively, and $A_{ei}(t) \in \mathbb{R}^{3\times3}$ is a function of the camera intrinsic calibration parameters and image coordinates of the i^{th} feature point as shown below

$$A_{ei} \triangleq A - \begin{bmatrix} 0 & 0 & u_i \\ 0 & 0 & v_i \\ 0 & 0 & 0 \end{bmatrix}. \qquad (14)$$

Similarly, to quantify the rotation of $\mathcal{F}$ relative to $\mathcal{F}^*$, we define $e_w(t) \in \mathbb{R}^3$ using the axis-angle representation [21] as follows

$$e_\omega \triangleq u\phi \qquad (15)$$

where $u(t) \in \mathbb{R}^3$ represents a unit rotation axis, and $\phi(t) \in \mathbb{R}$ denotes the rotation angle about $u(t)$ that is assumed to be confined to the region $-\pi < \phi(t) < \pi$. After taking the time derivative of (15), the following expression can be obtained (see [5] for further details)

$$\dot{e}_\omega = L_\omega R \omega_e \ . \qquad (16)$$

In (16), the Jacobian-like matrix $L_\omega(t) \in \mathbb{R}^{3\times3}$ is defined as

$$L_\omega \triangleq I_3 - \frac{\phi}{2}[u]_\times + \left(1 - \frac{sinc(\phi)}{sinc^2\left(\frac{\phi}{2}\right)} \right) [u]_\times^2 \qquad (17)$$

where $[u]_\times$ denotes the 3×3 skew-symmetric form of $u(t)$, $I_3 \in \mathbb{R}^{3\times3}$ is the 3×3 identity matrix, and

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

From (13) and (16), the kinematics of the object under motion can be expressed as

$$\dot{e} = Jv \qquad (18)$$

where $e(t) \triangleq \begin{bmatrix} e_v^T & e_\omega^T \end{bmatrix}^T \in \mathbb{R}^6$, $v(t) \triangleq \begin{bmatrix} v_e^T & \omega_e^T \end{bmatrix}^T \in \mathbb{R}^6$, and $J(t) \in \mathbb{R}^{6\times6}$ is a Jacobian-like matrix defined as

$$J = \begin{bmatrix} \frac{\alpha_1}{z_1^*} A_{e1} R & -\frac{\alpha_1}{z_1^*} A_{e1} R [s_1]_\times \\ 0_3 & L_\omega R \end{bmatrix} \qquad (19)$$

where $0_3 \in \mathbb{R}^{3\times3}$ denotes a zero matrix.

Remark 2 *In the subsequent analysis, it is assumed that a single geometric length $s_1 \in \mathbb{R}^3$ between two feature points is known. With this assumption, each element of $J(t)$ is known with the possible exception of the constant $z_1^* \in \mathbb{R}$. The reader is referred to [5] where it is shown that z_1^* can also be computed given s_1.*

Remark 3 *It is assumed that the object never leaves the field of view of the camera; hence, from (12) and (15), $e(t) \in \mathcal{L}_\infty$. It is also assumed that the object velocity, acceleration and jerk are bounded, i.e., $v(t), \dot{v}(t), \ddot{v}(t) \in \mathcal{L}_\infty$; hence the structure of (18) allows us to show that $\dot{e}(t), \ddot{e}(t), \dddot{e}(t) \in \mathcal{L}_\infty$.*

4 Identification of Velocity

In [5], an estimator was developed for online asymptotic identification of the signal $\dot{e}(t)$. Designating $\hat{e}(t)$ as the estimate for $e(t)$, the estimator was designed as follows

$$\begin{aligned} \dot{\hat{e}} \quad \triangleq \quad & \int_{t_0}^{t} (K + I_6)\tilde{e}(\tau) d\tau + \int_{t_0}^{t} \rho \operatorname{sgn}(\tilde{e}(\tau)) d\tau \\ & + (K + I_6)\tilde{e}(t) \end{aligned} \qquad (20)$$

where $\tilde{e}(t) \triangleq e(t) - \hat{e}(t) \in \mathbb{R}^6$ is the estimation error for the signal $e(t)$, $K, \rho \in \mathbb{R}^{6\times6}$ are positive definite constant diagonal gain matrices, $I_6 \in \mathbb{R}^{6\times6}$ is the 6×6 identity matrix, t_0 is the initial time, and $\operatorname{sgn}(\tilde{e}(t))$ denotes the standard signum function applied to each element of the vector $\tilde{e}(t)$. The reader is referred to [5] and the references therein for analysis pertaining to the development of the above estimator. In essense, it was shown in [5] that the above estimator asymptotically identifies the signal $\dot{e}(t)$ (*i.e.*, $\dot{\hat{e}}(t) \rightarrow \dot{e}(t)$

as $t \to \infty$) provided the following inequality is satisifed for each diagonal element ρ_i of the gain matrix ρ,

$$\rho_i \geq \left|\ddot{e}_i\right| + \left|\dddot{e}_i\right| \quad \forall i = 1, 2, ...6. \tag{21}$$

Since $J(t)$ is known and invertible, the six degree-of-freedom velocity of the moving object can be identified as follows

$$\hat{v}(t) = J^{-1}(t)\ \dot{\hat{e}}\ (t), \text{ and hence } \hat{v}(t) \to v(t) \text{ as } t \to \infty. \tag{22}$$

5 Euclidean Reconstruction of Feature Points

The central theme of this paper is the identification of Euclidean coordinates of the feature points on a moving object (*i.e.,* the vector s_i relative to the object frame $\mathcal{F}$, $\bar{m}_i(t)$ and $\bar{m}_i^*$ relative to the camera frame $\mathcal{I}$ for all i feature points on the object). To facilitate the development of the estimator, we first define the extended image coordinates, denoted by $p_{ei}(t) \in \mathbb{R}^3$, for any feature point O_i as follows

$$p_{ei} \triangleq \begin{bmatrix} u_i & v_i & -\ln(\alpha_i) \end{bmatrix}^T. \tag{23}$$

Following the development of translational kinematics in (13), it can be shown that the time derivative of (23) is given by

$$\begin{aligned} \dot{p}_{ei} &= \frac{\alpha_i}{z_i^*} A_{ei} R \left[v_e + [\omega_e]_\times s_i \right] \\ &= W_i V_{vw} \theta_i \end{aligned} \tag{24}$$

where $W_i(.) \in \mathbb{R}^{3\times3}, V_{vw}(t) \in \mathbb{R}^{3\times4}$ and $\theta_i \in \mathbb{R}^4$ are defined as follows

$$W_i \triangleq \alpha_i A_{ei} R \tag{25}$$

$$V_{vw} \triangleq \begin{bmatrix} v_e & [\omega_e]_\times \end{bmatrix} \tag{26}$$

$$\theta_i \triangleq \begin{bmatrix} \dfrac{1}{z_i^*} & \dfrac{s_i}{z_i^*}^T \end{bmatrix}^T. \tag{27}$$

The elements of $W_i(.)$ are known and bounded, and an estimate of $V_{vw}(t)$, denoted by $\hat{V}_{vw}(t)$, is available by appropriately re-ordering the vector $\hat{v}(t)$ given in (22).

Our objective is to identify the unknown constant θ_i in (24). To facilitate this objective, we define a parameter estimation error, denoted by $\tilde{\theta}_i(t) \in \mathbb{R}^4$, as follows

$$\tilde{\theta}_i(t) \triangleq \theta_i - \hat{\theta}_i(t) \tag{28}$$

where $\hat{\theta}_i(t) \in \mathbb{R}^4$ is a subsequently designed parameter update signal. We also introduce a measurable filter signal $W_{fi}(t) \in \mathbb{R}^{3\times4}$, and a non-measurable filter signal $\eta_i(t) \in \mathbb{R}^3$ defined as follows

$$\dot{W}_{fi} = -\beta_i W_{fi} + W_i \hat{V}_{vw} \tag{29}$$

$$\dot{\eta}_i = -\beta_i \eta_i + W_i \tilde{V}_{vw} \theta_i \tag{30}$$

where $\beta_i \in \mathbb{R}$ is a scalar positive gain, and $\tilde{V}_{vw}(t) \triangleq V_{vw}(t) - \hat{V}_{vw}(t) \in \mathbb{R}^{3\times4}$ is an estimation error signal.

Motivated by the subsequent stability analysis, we design the following estimate, denoted by $\hat{p}_{ei}(t) \in \mathbb{R}^3$, for the extended image coordinates,

$$\dot{\hat{p}}_{ei} = \beta_i \tilde{p}_{ei} + W_{fi}\ \dot{\hat{\theta}}_i + W_i \hat{V}_{vw} \hat{\theta}_i \tag{31}$$

where $\tilde{p}_{ei}(t) \triangleq p_{ei}(t) - \hat{p}_{ei}(t) \in \mathbb{R}^3$ denotes the measurable estimation error signal for the extended image coordinates of the feature points. The time derivative of this estimation error signal is computed from (24) and (31) as follows

$$\dot{\tilde{p}}_{ei} = -\beta_i \tilde{p}_{ei} - W_{fi}\ \dot{\hat{\theta}}_i + W_i \tilde{V}_{vw} \theta_i + W_i \hat{V}_{vw} \tilde{\theta}_i. \tag{32}$$

From (30) and (32), it can be shown that

$$\tilde{p}_{ei} = W_{fi} \tilde{\theta}_i + \eta_i. \tag{33}$$

Based on the subsequent analysis, we select the following least-squares update law [20] for $\hat{\theta}_i(t)$

$$\dot{\hat{\theta}}_i = L_i W_{fi}^T \tilde{p}_{ei} \tag{34}$$

where $L_i(t) \in \mathbb{R}^{4\times4}$ is an estimation gain that is recursively computed as follows

$$\frac{d}{dt}(L_i^{-1}) = W_{fi}^T W_{fi}. \tag{35}$$

Remark 4 *In the subsequent analysis, it is required that $L_i^{-1}(0)$ in (35) be positive definite. This requirement can be easily satisfied by selecting the appropriate non-zero initial values.*

Remark 5 *In the analysis provided in [5], it was shown that a filter signal $r(t) \in \mathbb{R}^6$ defined as $r(t) = \tilde{e}(t) + \dot{\tilde{e}}\ (t) \in L_\infty \cap L_2$. From this result it is easy to show that the signals $\tilde{e}(t)$, $\dot{\tilde{e}}\ (t) \in L_2$ [7]. Since $J(t) \in L_\infty$ and invertible, it follows that $J^{-1}(t)\ \dot{\tilde{e}}\ (t) \in L_2$. Hence $\tilde{v}(t) \triangleq v(t) - \hat{v}(t) \in L_2$, and it is easy to show that $\left\|\tilde{V}_{vw}(t)\right\|_\infty^2 \in L_1$, where the notation $\|.\|_\infty$ denotes the induced ∞-norm of a matrix [15].*

5.1 Analysis

Theorem 1 *The update law defined in (34) ensures that $\tilde{\theta}_i(t) \to 0$ as $t \to \infty$ provided that the following persistent excitation condition [20] holds*

$$\gamma_1 I_4 \leq \int_{t_0}^{t_0+T} W_{fi}^T(\tau) W_{fi}(\tau) d\tau \leq \gamma_2 I_4 \tag{36}$$

and provided that the gains β_i satisfy the following inequality

$$\beta_i > k_{1i} + k_{2i} \left\|W_i\right\|_\infty^2 \tag{37}$$

where $t_0, \gamma_1, \gamma_2, T, k_{1i}, k_{2i} \in \mathbb{R}$ are positive constants, $I_4 \in \mathbb{R}^{4\times4}$ is the 4×4 identity matrix, the notation $\|.\|_\infty$ denotes the induced ∞-norm of a matrix [15] and k_{1i} must be selected such that

$$k_{1i} > 2. \tag{38}$$

Proof: Let $V(t) \in \mathbb{R}$ denote a non-negative scalar function defined as follows

$$V \triangleq \frac{1}{2}\tilde{\theta}_i^T L_i^{-1} \tilde{\theta}_i + \frac{1}{2}\eta_i^T \eta_i. \tag{39}$$

After taking the time derivative of (39), the following expression can be obtained

$$\begin{aligned}\dot{V} \leq\; & -\frac{1}{2}\left\|W_{fi}\tilde{\theta}_i\right\|^2 - \beta_i \left\|\eta_i\right\|^2 \\ & + \left\|\theta_i\right\| \left\|W_i\right\|_\infty \left\|\tilde{V}_{vw}\right\|_\infty \left\|\eta_i\right\| \\ & + \left\|W_{fi}\tilde{\theta}_i\right\| \left\|\eta_i\right\| - k_{1i}\left\|\eta_i\right\|^2 + k_{1i}\left\|\eta_i\right\|^2 \\ & + k_{2i}\left\|W_i\right\|_\infty^2 \left\|\eta_i\right\|^2 - k_{2i}\left\|W_i\right\|_\infty^2 \left\|\eta_i\right\|^2 \end{aligned} \tag{40}$$

After utilizing the nonlinear damping argument [16], we can simplify (40) further as follows

$$\begin{aligned}\dot{V} \leq\; & -\left(\frac{1}{2}-\frac{1}{k_{1i}}\right)\left\|W_{fi}\tilde{\theta}_i\right\|^2 \\ & -\left(\beta_i - k_{1i} - k_{2i}\left\|W_i\right\|_\infty^2\right)\left\|\eta\right\|^2 \\ & + \frac{1}{k_{2i}}\left\|\theta_i\right\|^2 \left\|\tilde{V}_{vw}\right\|_\infty^2 \end{aligned} \tag{41}$$

where $k_{1i}, k_{2i} \in \mathbb{R}$ are positive constants as previously mentioned. The gains k_{1i}, k_{2i}, and β_i must be selected to ensure that

$$\frac{1}{2}-\frac{1}{k_{1i}} \geq \mu_{1i} > 0 \tag{42}$$

$$\beta_i - k_{1i} - k_{2i}\left\|W_i\right\|_\infty^2 \geq \mu_{2i} > 0 \tag{43}$$

where $\mu_{1i}, \mu_{2i} \in \mathbb{R}$ are positive constants. The gain conditions given by (42) and (43) allow us to formulate the conditions given by (37) and (38), as well as allowing us to further upper bound the time derivative of (39) as follows

$$\dot{V} \leq -\mu_{1i}\left\|W_{fi}\tilde{\theta}_i\right\|^2 - \mu_{2i}\left\|\eta_i\right\|^2 + \frac{1}{k_{2i}}\left\|\theta_i\right\|^2\left\|\tilde{V}_{vw}\right\|_\infty^2. \tag{44}$$

From the discussion given in Remark 5, we can see that the last term in (44) is L_1, hence,

$$\int_0^\infty \frac{1}{k_{2i}}\left\|\theta_i(\tau)\right\|^2 \left\|\tilde{V}_{vw}(\tau)\right\|_\infty^2 d\tau \leq \varepsilon \tag{45}$$

where $\varepsilon \in \mathbb{R}$ is a positive constant. From (39), (44) and (45), we can conclude that

$$\begin{aligned} & \int_0^\infty \left(\mu_{1i}\left\|W_{fi}(\tau)\tilde{\theta}_i(\tau)\right\|^2 + \mu_{2i}\left\|\eta_i(\tau)\right\|^2\right) d\tau \\ \leq\; & V(0) - V(\infty) + \varepsilon. \end{aligned} \tag{46}$$

It can be concluded from (46) that $W_{fi}(t)\tilde{\theta}_i(t), \eta_i(t) \in L_2$. From (46) and the fact that $V(t)$ is non-negative, it can be concluded that $V(t) \leq V(0) + \varepsilon$ for any t, and hence $V(t) \in L_\infty$. Therefore, from (39), $\eta_i(t) \in L_\infty$ and $\tilde{\theta}_i^T(t)L_i^{-1}(t)\tilde{\theta}_i(t) \in L_\infty$. Since $L_i^{-1}(0)$ is positive definite, and the persistent excitation condition in (36) is assumed to be satisfied, we can use (35) to show that $L_i^{-1}(t)$ is always positive definite; hence, it must follow that $\tilde{\theta}_i(t) \in L_\infty$. Since $\hat{v}(t) \in L_\infty$ as shown in [5], it follows from (26) that $\hat{V}_{vw}(t) \in L_\infty$. Hence from (29), and the fact that $W_i(.)$ defined in (25) are composed of bounded terms, $W_{fi}(t), \dot{W}_{fi}(t) \in L_\infty$ [7], and consequently, $W_{fi}(t)\tilde{\theta}_i(t) \in L_\infty$. Therefore, from (33), we can see that $\tilde{p}_{ei}(t) \in L_\infty$. It follows from (34) that $\dot{\hat{\theta}}_i\,(t) \in L_\infty$, and hence $\dot{\tilde{\theta}}_i\,(t) \in L_\infty$. From the fact that $\dot{W}_{fi}(t), \dot{\tilde{\theta}}_i(t) \in L_\infty$, it is easy to show that $\frac{d}{dt}\left(W_{fi}(t)\tilde{\theta}_i(t)\right) \in L_\infty$. Hence, $W_{fi}(t)\tilde{\theta}_i(t)$ is uniformly continuous [9]. Since we also have that $W_{fi}(t)\tilde{\theta}_i(t) \in L_2$, we can conclude that [9]

$$W_{fi}(t)\tilde{\theta}_i(t) \rightarrow 0 \text{ as } t \rightarrow \infty. \tag{47}$$

As shown in Appendix C, if the signal $W_{fi}(t)$ satisifies the persistent excitation condition [20] given in (36), then it can be concluded from (47) that $\tilde{\theta}_i(t) \rightarrow 0$ as $t \rightarrow \infty$.□

Remark 6 *It can be shown that the output $W_{fi}(t)$ of the filter defined in (29) is persistently exciting if the input $W_i(t)\hat{V}_{vw}^T(t)$ to the filter is persistently exciting [19]. Hence, the condition in (36) is satisfied if*

$$\gamma_3 I_4 \leq \int_{t_0}^{t_0+T} \underbrace{\hat{V}_{vw}^T(\tau)W_i^T(\tau)W_i(\tau)\hat{V}_{vw}(\tau)}_{W} d\tau \leq \gamma_4 I_4 \tag{48}$$

where $\gamma_3, \gamma_4 \in \mathbb{R}$ are positive constants. It can be shown upon expansion of the integrand $W(t) \in \mathbb{R}^{4\times 4}$ of (48) that even if only one of the components of translational velocity is non-zero, the first element of $\hat{\theta}_i(t)$, i.e. $\frac{1}{z_i^*}$, will converge to the correct value. Unfortunately, it seems that no inference can be made about the relationship between convergence of the three remaining elements of $\hat{\theta}_i(t)$ and the rotational velocity of the object.

Remark 7 *As stated in the previous remarks, the estimation of object velocity requires the knowledge of the constant rotation matrix $R^* \in \mathbb{R}^{3\times 3}$ and a single geometric length $s_1 \in \mathbb{R}^3$ on the object. Then, utilizing (8), (10), (27) and (34), the estimates for $\bar{m}_i^*$ and $\bar{m}_i$, denoted by $\hat{\bar{m}}_i^*(t)$, $\hat{\bar{m}}_i(t) \in \mathbb{R}^3$, respectively, can be obtained as follows*

$$\hat{\bar{m}}_i^*(t) = \frac{1}{\left[\hat{\theta}_i(t)\right]_1} A^{-1} p_i^* \tag{49}$$

$$\hat{\bar{m}}_i(t) = \frac{\alpha_i(t)}{\left[\hat{\theta}_i(t)\right]_1} A^{-1} p_i(t) \tag{50}$$

where the term in the denominator denotes the first element of the vector $\hat{\theta}_i(t)$.

6 Simulation Results

The adaptive estimation algorithm described in Section 5 was built on top of an existing simulator that was previously utilized to verify the performance of the veclocity estimator of Section 4 and described in detail in [5]. We selected a planar object with four feature points initially 2 meters away along the axis of the camera as the body undergoing motion. The velocity of the object along each of the six degrees of freedom was set to $0.2\sin(t)$. The coordinates of the object feature points in the object's coordinate frame $\mathcal{F}$ were arbitrarily chosen to be the following

$$\begin{aligned} s_1 &= \begin{bmatrix} 1.0 & 0.5 & 0.1 \end{bmatrix}^T \\ s_2 &= \begin{bmatrix} 1.2 & -0.75 & 0.1 \end{bmatrix}^T \\ s_3 &= \begin{bmatrix} 0.0 & -1.0 & 0.1 \end{bmatrix}^T \\ s_4 &= \begin{bmatrix} -1.0 & 0.5 & 0.1 \end{bmatrix}^T. \end{aligned} \tag{51}$$

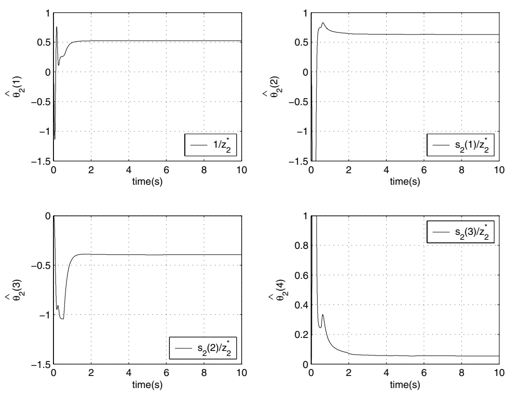

Figure 2: Estimated parameters for the second feature point on the simulated moving object.

The object's reference orientation R^* relative to the camera were selected as $diag(1, -1, -1)$. The simulator operated at the sampling frequency of 1 kHz. The estimator gain β_i was set to 20 for all i feature points. It was observed that the estimates $\hat{\theta}_i(t)$ converged to the correct values within a span of few seconds. As an example, Figure 2 depicts the convergence of $\hat{\theta}_2(t)$ used to compute the constant depth z_2^* relative to camera frame $\mathcal{I}$ and the Euclidean coordinates of the second feature point s_2 relative to the object frame $\mathcal{F}$.

7 Conclusions

This paper presented an adaptive nonlinear estimator to identify the Euclidean coordinates of feature points on an object under motion using a single camera. The only requirements on the object are that its velocity and first two time derivatives must be bounded, the orientation of the object at reference position relative to the camera, and the Euclidean coordinates of a single feature point relative to its coordinate frame must be known. Lyapunov-based system analysis methods and homography-based vision techniques were used in the development of this alternative approach to the classical problem of estimating structure from motion.

References

[1] T. J. Broida, S. Chandrashekhar, and R. Chellappa, "Recursive 3-D Motion Estimation From a Monocular Image Sequence," *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 26, No. 4, 1990.

[2] J. Chen, A. Behal, D. Dawson, and Y. Fang, "2.5D Visual Servoing with a Fixed Camera," *Proc. of the American Control Conference*, pp. 3442-3447, 2003.

[3] X. Chen, and H. Kano, "State Observer for a Class of Nonlinear Systems and Its Application to Machine Vision," *IEEE Transactions on Automatic Control*, Vol. 49, No. 11, 2004.

[4] V. K. Chitrakaran, D. M. Dawson, J. Chen, and W. E. Dixon, "Euclidean Position Estimation of Features on a Moving Object Using a Single Camera: A Lyapunov-based Approach," Clemson University CRB Technical Report, CU/CRB/9/9/04/#2, http://www.ces.clemson.edu/ece/crb/publictn/tr.htm, Sept., 2004.

[5] V. Chitrakaran, D. M. Dawson, W. E. Dixon, and J. Chen, "Identification of a Moving Object's Velocity with a Fixed Camera," *Automatica*, Vol. 41, No. 3, pp. 553-562, March 2005.

[6] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from Motion Causally Integrated Over Time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, pp.523-535, April 2002.

[7] C. A. Desoer, and M. Vidyasagar, *Feedback Systems: Input-Output Properties*, Academic Press, 1975.

[8] G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A Solution to the Simultaneous Localisation and Map Building (SLAM) Problem," *IEEE Trans. in Robotics and Automation*, Vol. 17, No. 3, pp. 229-241, 2001.

[9] W. E. Dixon, A. Behal, D. M. Dawson, and S. Nagarkatti, *Nonlinear Control of Engineering Systems: A Lyapunov-Based Approach*, Birkhäuser Boston, 2003.

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

[11] 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.

[12] Richard I. Hartley, "In Defense of the Eight-Point Algorithm," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 6, pp. 580-593, 1997.

[13] X. Hu, and T. Ersson, "Active State Estimation of Nonlinear Systems," *Automatica*, Vol. 40, pp. 2075-2082, 2004.

[14] M. Jankovic, and B. K. Ghosh, "Visually Guided Ranging from Observations of Points, Lines and Curves via an Identifier Based Nonlinear Observer," *Systems and Control Letters*, Vol. 25, pp. 63-73, 1995.

[15] H.K. Khalil, *Nonlinear Systems*, third edition, Prentice Hall, 2002.

[16] M. Krstić, I. Kanellakopoulos, and P. Kokotović, *Nonlinear and Adaptive Control Design*, New York, NY: John Wiley and Sons, 1995.

[17] 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, 2000.

[18] J. Oliensis, "A Critique of Structure From Motion Algorithms," *Computer Vision and Image Understanding*, Vol. 80, No. 2, pp. 172-214, 2000.

[19] S. Sastry, and M. Bodson, Adaptive Control: Stability, Convergence, and Robustness, Prentice Hall, Inc: Englewood Cliffs, NJ, 1989.

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

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

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

See [4] for appendices A, B and C.