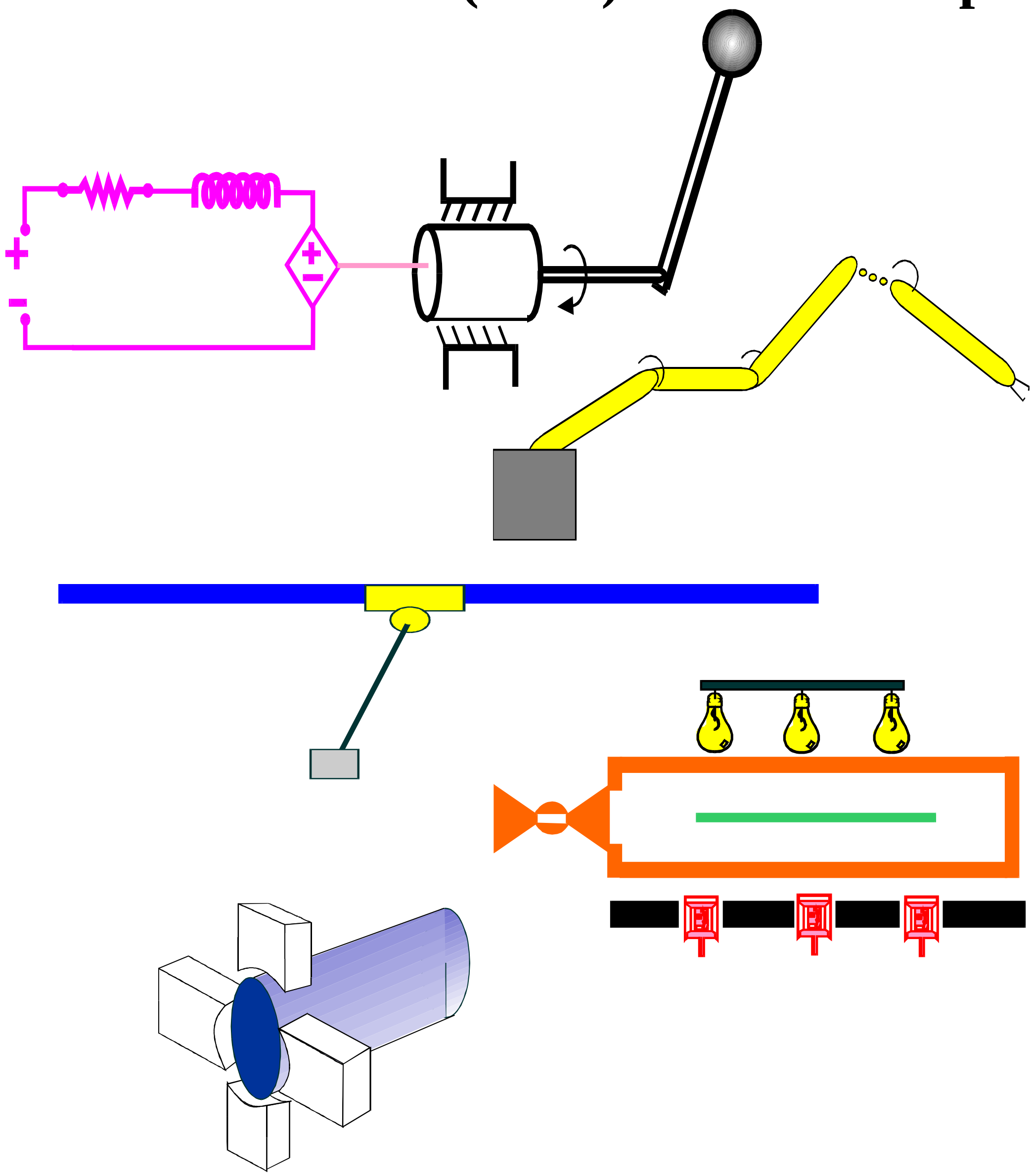
Clemson University
College of Engineering and Science
Control and Robotics (CRB) Technical Report



Number: CU/CRB/3/14/06/#1

Title: Velocity and Structure Estimation of a Moving Object Using a Moving Monocular Camera

Authors: V. K. Chitrakaran, D. M. Dawson, J. Chen, and H. Kannan

Report Documentation Page

Report Documentation Page	Form Approved OMB No. 0704-0188

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**
4. TITLE AND SUBTITLE **Velocity and Structure Estimation of a Moving Object Using a Moving Monocular Camera**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Clemson University,Department of Electrical & Computer Engineering,Clemson,SC,29634-0915**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**9**	

Velocity and Structure Estimation of a Moving Object Using a Moving Monocular Camera

V. K. Chitrakaran†, D. M. Dawson, J. Chen, and H. Kannan

***Abstract*— In this paper, we present the development of a vision-based estimator for simultaneous determination of velocity and structure of an object (i.e., the Euclidean position of its feature points) for the general case where both the object and the camera are moving relative to an inertial frame of reference. The velocity estimation itself requires no explicit kinematic model, while the adaptive structure estimator, synthesized utilizing Lyapunov design methods, is built upon kinematic relationships that rely on homography-based techniques.**

I. Introduction

The application of a camera as a sensor for acquiring the 3D structure of a scene is known as "Structure from Motion (SfM)" in the computer vision literature [17]. The problem usually involves a camera mounted on a moving platform such as a mobile robot whose image data is utilized to map the Euclidean position of *static landmarks* or visual features in the environment. Recent applications of this technique include aerial surveillance and mapping systems such as the work presented in [10] and [12]. A significant extension to the above problem is the case where both the object of interest and the camera are in motion, with potential impact on such applications as vision-based collision avoidance systems for autonomous guidance of multiple vehicles on highways [9]. In this paper, we propose a nonlinear estimation strategy to identify the Euclidean structure and velocity of a *moving object* using a monocular calibrated *moving camera*. The proposed algorithm relies on the availability of a reference image of the target object, captured at a known orientation relative to the camera. For the general case where both the camera and the object are in motion relative to an inertial frame, a single geometric length on the object is assumed to be known, and sensors mounted on the camera are assumed to provide camera velocity information. Typically, linearization based methods such as extended Kalman filtering [1], [5], [17] provide the underlying algorithmic foundation for most SfM results, although the problem of estimating 3D information from 2D images is inherently nonlinear. In this work, we approach the problem using nonlinear system analysis tools. Equations for motion kinematics are first developed in terms of Euclidean and image-space information based on the approach presented in [16]. An integral feedback estimation method introduced in [4] is then employed to identify the linear and angular velocity of the moving object. The estimated velocities facilitate the development of a measurable error system that can be used to formulate a nonlinear least squares adaptive update law for Euclidean structure estimation. Then, the satisfaction of a persistent excitation condition (similar to [2] and others) allows the determination of the coordinates for all the feature points on the object relative to the camera.

Although the requirement of velocity sensors on the camera might seem too restrictive at first, the proposed algorithm does allow for some interesting applications. For example, a camera on-board an aerial reconnaissance vehicle could be employed to track and estimate the dimensions of ground targets (such as moving vehicles) utilizing a single snapshot as the reference image captured, for example, by a satellite. In this case, the camera velocity information can be obtained from the Inertial Navigation system on-board the aircraft. Similarly, a pan-tilt camera could be employed as a passive radar for pose and velocity recovery of a rigid moving object in its field of view. The fact that camera motion is allowed frees the moving object from any restrictive assumptions on its trajectory, as the camera orientation can be continuously updated in order to keep the object within its field of view. A system based on this technique could also be applied in the microscopic domain: a camera equipped with magnifying lenses mapping structure and motion of objects too small to have motion sensors embedded on them.

The remainder of this paper is organized as follows. In Section II, we begin with the development of a geometric model and homography relating the pixel coordinates of visual markers on the object, extracted from the reference image and the continuous sequence of images from the camera. Section III describes the motion kinematics between the camera and the target object. This is followed by the development of estimators for relative velocity and Euclidean position in Sections IV and V, respectively. A Lyapunov stability analysis for Euclidean position estimation is provided in Section V-A. In Appendix II, we illustrate the cases where either the object or the camera is stationary relative to the inertial frame. These special cases add to the range of practical applications of this algorithm, especially since some of assumptions necessary for the general case can be relaxed. Specifically, no velocity sensors on the camera are required for the special cases.

II. Geometric Model

We begin the development of a geometric model relating the camera and the object by introducing some notation.

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.

The authors are with the Dept. of Electrical & Computer Engineering, Clemson University, SC, USA.

†Corresponding Author: Phone/Fax: (864) 656-7708/(864) 656-7218, Email: cvilas@ces.clemson.edu

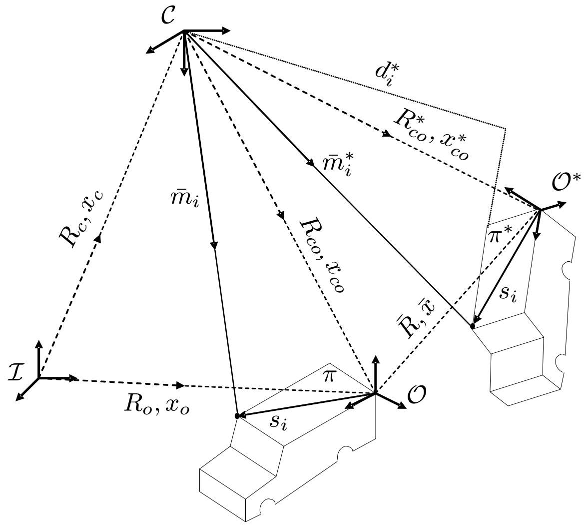


Fig. 1. The geometry of the camera frame and the object frame relative to the inertial frame.

In Figure 1, the inertial, camera and the object frames are denoted by the orthogonal coordinate frames $\mathcal{I}$, $\mathcal{C}$, and $\mathcal{O}$, respectively. For the sake of simplicity, it is assumed that the origin of $\mathcal{C}$ coincides with the optical center of the camera. The vector $x_c(t) \in \mathbb{R}^3$ and the matrix $R_c(t) \in SO(3)$ denote the position and orientation, respectively, of the camera relative to the inertial frame $\mathcal{I}$, and expressed in the coordinates of $\mathcal{I}$, such that $R_c(t)$ defines the mapping $R_c : \mathcal{C} \rightarrow \mathcal{I}$. Similarly, the position and orientation of the object frame relative to the inertial frame are quantified by $x_o(t) \in \mathbb{R}^3$ and $R_o(t) \in SO(3)$, respectively, where $R_o(t)$ is the mapping $R_o : \mathcal{O} \rightarrow \mathcal{I}$. We also define the quantities $x_{co}(t) \in \mathbb{R}^3$ and $R_{co}(t) \in SO(3)$ to be the position and orientation of the object frame relative to the camera frame, and expressed in the camera frame $\mathcal{C}$, such that $R_{co} : \mathcal{O} \rightarrow \mathcal{C}$. The three dimensional structure of a rigid object in the field of view of the camera is described in terms of the 3D location of various visual features on the object relative to the camera frame. If the Euclidean coordinates of the i^{th} feature point O_i on the object is denoted by the constant $s_i \in \mathbb{R}^3$ in the object frame $\mathcal{O}$, and $\bar{m}_i(t) \triangleq \begin{bmatrix} \bar{m}_{ix} & \bar{m}_{iy} & \bar{m}_{iz} \end{bmatrix}^T \in \mathbb{R}^3$ in the camera frame $\mathcal{C}$, then, it can be seen from Figure 1 that,

$$\bar{m}_i = x_{co} + R_{co} s_i. \tag{1}$$

We now develop a geometric relationship that describes how the image coordinates of a feature point on the object change with relative motion between the camera and the object. To this end, we define a *reference* position of the object relative to the camera, denoted by $\mathcal{O}^*$ in Figure 1. At this reference location relative to the camera, the position and orientation of $\mathcal{O}^*$, and the Euclidean coordinates of the feature points relative to the camera frame, are denoted by the constant terms $x_{co}^* \in \mathbb{R}^3$, $R_{co}^* \in SO(3)$ and $\bar{m}_i^* \in \mathbb{R}^3$, respectively. Hence, similar to (1), we have

$$\bar{m}_i^* = x_{co}^* + R_{co}^* s_i. \tag{2}$$

After solving (2) for s_i and then substituting the resulting expression into (1), we have

$$\bar{m}_i = \bar{x} + \bar{R}\bar{m}_i^* \tag{3}$$

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

$$\bar{R} = R_{co}\left(R_{co}^*\right)^T, \qquad \bar{x} = x_{co} - \bar{R}x_{co}^* \ . \tag{4}$$

Let three of the non-collinear feature points on the object, denoted by O_i $\forall i = 1, 2, 3$, define the plane π in frame $\mathcal{O}$, and π^* in $\mathcal{O}^*$. As illustrated in Figure 1, $n^* \in \mathbb{R}^3$ denotes the constant normal to the plane π^* expressed in the coordinates of $\mathcal{C}$, and the constant projections of $\bar{m}_i^*$ along the unit normal n^*, denoted by $d_i^* \in \mathbb{R}$, are given by

$$d_i^* = n^{*T}\bar{m}_i^*. \tag{5}$$

Using (5), it can be easily seen that the relationship in equation (3) can now be expressed as follows

$$\bar{m}_i = \underbrace{\left(\bar{R} + \frac{1}{d_i^*}\bar{x}n^{*T}\right)}_{H}\bar{m}_i^* \tag{6}$$

where $H(t) \in \mathbb{R}^{3\times3}$ denotes a Euclidean homography [14]. To express the above relationship in terms of the measurable image space coordinates of the visual features relative to the camera frame, the normalized Euclidean coordinates $m_i(t), m_i^* \in \mathbb{R}^3$ are defined as follows

$$m_i \triangleq \frac{\bar{m}_i}{\bar{m}_{iz}}, \qquad m_i^* \triangleq \frac{\bar{m}_i^*}{\bar{m}_{iz}^*}. \tag{7}$$

The image coordinates of these feature points expressed relative to $\mathcal{C}$ are denoted by $p_i(t), p_i^* \in \mathbb{R}^3$ as follows

$$p_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T, \qquad p_i^* = \begin{bmatrix} u_i^* & v_i^* & 1 \end{bmatrix}^T \ . \tag{8}$$

The image coordinates and the normalized Euclidean coordinates are related by the pin-hole camera model [15] such that

$$p_i = Am_i, \qquad p_i^* = Am_i^* \tag{9}$$

where $A \in \mathbb{R}^{3\times3}$ is a known, constant, upper triangular and invertible intrinsic camera calibration matrix. From (6) and (9), the relationship between image coordinates of the corresponding feature points in $\mathcal{O}$ and $\mathcal{O}^*$ can be expressed as follows

$$p_i = \underbrace{\frac{\bar{m}_{iz}^*}{\bar{m}_{iz}}}_{\alpha_i} \underbrace{A\left(\bar{R} + \bar{x}_{hi}(n^*)^T\right)A^{-1}}_{G} p_i^* \tag{10}$$

where $\alpha_i \in \mathbb{R}$ denotes the depth ratio, and $\bar{x}_{hi}(t) = \dfrac{\bar{x}(t)}{d_i^*} \in \mathbb{R}^3$ denotes the scaled translation vector. The matrix $G(t) \in \mathbb{R}^{3\times3}$ defined in (10) is a full rank homogeneous collineation matrix defined up to a scale factor [15]. If the structure of the object is planar, all feature points lie on the same plane, and hence the distances d_i^* defined in (5) is the same for all feature points, henceforth, denoted as d^*. In this case, the collineation $G(t)$ is defined up to the same scale factor, and hence, one of its elements can be set to unity without loss of generality. Given a reference image of the object corresponding to $\mathcal{O}^*$, and a continuous stream of images of the object from the camera corresponding to $\mathcal{O}$ at every instant of time, $G(t)$ can be estimated from a set of linear equations (10) obtained from at least four matched feature points $(p_i(t), p_i^*)$ that are coplanar but non-collinear. If the structure of the object is not planar, the Virtual Parallax method described in [15] could be utilized, where three of the non-collinear feature points on the object are utilized to define a virtual plane at the distance d^* from the camera.. An overview of the determination of the collineation matrix $G(t)$ and the depth ratios $\alpha_i(t)$ for both the planar and non-planar cases are also given in [3]. 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 [8], [14], [21]), $H(t)$ can be decomposed into its constituent rotation matrix $\bar{R}(t)$, unit normal vector n^*, scaled translation vector $\bar{x}_h(t) \triangleq \dfrac{\bar{x}(t)}{d^*}$ and the depth ratio $\alpha_i(t)$. It is assumed that the constant rotation matrix R_{co}^* is known. $R_{co}(t)$ can therefore be computed from (4). Hence $R_{co}(t)$, $\bar{R}(t), \bar{x}_h(t)$ 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_{co}^* be known. We consider this to be a mild assumption since the reference image of the object can be acquired offline after placing the object or the camera at some known orientation relative to each other.

III. Kinematics

Let $v_c(t), \omega_c(t) \in \mathbb{R}^3$ denote the translational and rotational velocities of the camera, relative to $\mathcal{I}$ and expressed in the coordinates of $\mathcal{I}$. In the coordinates of $\mathcal{C}$, the same physical quantities relative to $\mathcal{I}$ are denoted as $v_{cc}(t), \omega_{cc}(t) \in \mathbb{R}^3$. Similarly, $v_o(t), \omega_o(t) \in \mathbb{R}^3$ and $v_{oo}(t), \omega_{oo}(t) \in \mathbb{R}^3$ denote the translational and rotational velocities of the object frame, all relative to $\mathcal{I}$, expressed in the coordinates of $\mathcal{I}$ and $\mathcal{O}$, respectively. These velocities are related to each other in the following manner,

$$\begin{aligned} \begin{bmatrix} v_{cc}^T & \omega_{cc}^T \end{bmatrix}^T &= R_c^T \begin{bmatrix} v_c^T & \omega_c^T \end{bmatrix}^T, && (11) \\ \begin{bmatrix} v_{oo}^T & \omega_{oo}^T \end{bmatrix}^T &= R_o^T \begin{bmatrix} v_o^T & \omega_o^T \end{bmatrix}^T. && (12) \end{aligned}$$

To quantify the translation between the coordinate frames $\mathcal{O}^*$ and $\mathcal{O}$, we define $e_v(t) \in \mathbb{R}^3$ in terms of the image coordinates of one of the feature points on the plane π. For notational simplicity, we chose O_1 and hence,

$$e_v \triangleq \begin{bmatrix} u_1 - u_1^* & v_1 - v_1^* & -\ln(\alpha_1) \end{bmatrix}^T. \qquad (13)$$

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. After taking the time derivative of (13), the following translational kinematics can be obtained (see Appendix I for details)

$$\dot{e}_v = \frac{\alpha_1}{\bar{m}_{1z}^*} A_{e1} \left[S(x_{co})\omega_{cc} - R_{co}v_r + R_{co}S(s_1)\omega_r \right] \qquad (14)$$

where $S(\cdot) \in \mathbb{R}^{3\times3}$ denotes the skew-symmetric form of the vector s_1 as defined in [19], and $v_r(t), \omega_r(t) \in \mathbb{R}^3$ are the relative translational and rotational velocities between the camera and the object defined in the following manner

$$v_r = R_o^T \left(v_c - v_o \right), \qquad \omega_r = R_o^T \left(\omega_c - \omega_o \right) . \qquad (15)$$

In (14), $A_{ei}(t) \in \mathbb{R}^{3\times3}$ is a function of the camera intrinsic calibration parameters and image coordinates of the i^{th} feature point as shown below

$$A_{ei} \triangleq A - \begin{bmatrix} 0 & 0 & u_i \\ 0 & 0 & v_i \\ 0 & 0 & 0 \end{bmatrix}. \qquad (16)$$

Similarly, to quantify the rotation between $\mathcal{O}^*$ and $\mathcal{O}$, we define $e_\omega(t) \in \mathbb{R}^3$ using the axis-angle representation [19] as follows

$$e_\omega \triangleq \mu\phi \qquad (17)$$

where $\mu(t) \in \mathbb{R}^3$ represents a unit rotation axis, and $\phi(t) \in \mathbb{R}$ denotes the rotation angle about $\mu(t)$ confined to the region $-\pi < \phi(t) < \pi$, and defined as follows

$$\phi = \cos^{-1}\left(\frac{1}{2}\left(tr(\bar{R}) - 1 \right) \right), \qquad S\left(\mu\right) = \frac{\bar{R} - \bar{R}^T}{2\sin(\phi)} \ . \qquad (18)$$

In (18), the notation $tr\left(\cdot\right)$ denotes the trace of a matrix. After taking the time derivative of (17), the following expression for rotational kinematics is obtained (see Appendix I for details)

$$\dot{e}_\omega = -L_\omega R_{co}\omega_r \qquad (19)$$

where the Jacobian-like term $L_\omega(t) \in \mathbb{R}^{3\times3}$ is given by the following expression

$$L_\omega = I_3 - \frac{\phi}{2} S\left(\mu\right) + \left(1 - \frac{sinc\left(\phi\right)}{sinc^2\left(\dfrac{\phi}{2}\right)} \right) S\left(\mu\right)^2 \qquad (20)$$

and $sinc\left(\phi\right) \triangleq \dfrac{\sin\left(\phi\right)}{\phi}$. From (14) and (19), the kinematics of relative motion between the camera and the object can thus be expressed as follows

$$\dot{e} = Jv + f \qquad (21)$$

where $e(t) \triangleq \begin{bmatrix} e_v^T & e_\omega^T \end{bmatrix}^T \in \mathbb{R}^6$, and $v(t) \triangleq \begin{bmatrix} v_r^T & \omega_r^T \end{bmatrix}^T \in \mathbb{R}^6$. The vector $f(t) \in \mathbb{R}^6$ and the matrix $J(t) \in \mathbb{R}^{6\times6}$ in (21) are defined as follows

$$\begin{aligned} J &= \begin{bmatrix} -\dfrac{\alpha_1}{\bar{m}_{1z}^*} A_{e1} R_{co} & \dfrac{\alpha_1}{\bar{m}_{1z}^*} A_{e1} R_{co} S(s_1) \\ 0_{3\times3} & -L_\omega R_{co} \end{bmatrix}, && (22) \\ f &= \begin{bmatrix} \dfrac{\alpha_1}{\bar{m}_{1z}^*} A_{e1} S(x_{co})\omega_{cc} \\ 0_3 \end{bmatrix} && (23) \end{aligned}$$

where $0_{3\times 3} \in \mathbb{R}^{3\times 3}$ denotes a 3×3 zero matrix, and $0_3 \in \mathbb{R}^3$ denotes a zero vector.

We assume that a single geometric length $s_1 \in \mathbb{R}^3$ between two feature points on the object is known. This allows us to compute x^*_{co} using the following alternative expression

$$x^*_{co} = \left[diag(\bar{\gamma}_1 - \bar{\gamma}_2) + \bar{R}\right]^{-1} \left[diag(\bar{\gamma}_1 - \bar{\gamma}_2)R^*_{co} - R_{co}\right] s_1 \tag{24}$$

where

$$\bar{\gamma}_1 = \left[diag\left(m^*_1\right)\right]^{-1} n^{*T} m^*_1 \bar{x}_h, \tag{25}$$
$$\bar{\gamma}_2 = \left[diag\left(m^*_1\right)\right]^{-1} \frac{m_1}{\alpha_1} \tag{26}$$

and $m_1(t), m^*_1, n^*, \bar{x}_h(t)$ and $\alpha_1(t)$ are all measurable signals. Also, since $m^*_1 = \frac{1}{\bar{m}^*_{1z}}(x^*_{co} + R^*_{co}s_1)$, $\bar{m}^*_{1z}$ and $\bar{m}^*_1$ can both be computed. From (4), (5), and the definition for $\bar{x}_h(t)$, we can compute $x_{co}(t)$. Also, as mentioned previously, it is assumed that the velocities $v_{cc}(t), \omega_{cc}(t)$ are available from sensors on the camera. Hence, each element of $J(t)$ and $f(t)$ given in (22) and (23) are known.

Remark 2: By exploiting the fact that $\mu(t)$ is a unit vector (i.e., $\|\mu\|^2 = 1$), the determinant of $L_\omega(t)$ can be derived as follows [16]

$$\det L_\omega = \frac{1}{sinc^2\left(\dfrac{\theta}{2}\right)}. \tag{27}$$

From (27), it is clear that $L_\omega(t)$ is only singular for multiples of 2π (i.e., out of the assumed workspace). It can also been seen that $\det(A_{e1}R_{co}) \neq 0$. Hence, the matrix $J(t)$ is invertible [11].

IV. Relative Velocity Estimation

In [20], a model-free estimator for asymptotic identification of a velocity signal was presented, utilizing only the measured position signal for estimation. Based on this work, in [4], we presented a detailed analysis of the application of this observer for estimation of the velocity of a moving object in the field of view of a fixed camera. Specifically, designating $\hat{e}(t) \in \mathbb{R}^6$ as the estimate for the kinematic signal $e(t)$, the observer was designed as follows

$$\begin{aligned}\dot{\hat{e}} \triangleq& \int_{t_0}^{t}(K + I_{6\times 6})\tilde{e}(\tau)d\tau + \int_{t_0}^{t}\rho\,\mathrm{sgn}\left(\tilde{e}(\tau)\right)d\tau \\ &+(K + I_{6\times 6})\tilde{e}(t)\end{aligned} \tag{28}$$

where $\tilde{e}(t) \triangleq e(t) - \hat{e}(t) \in \mathbb{R}^6$ is the estimation error signal, $K, \rho \in \mathbb{R}^{6\times 6}$ are positive definite constant diagonal gain matrices, $I_6 \in \mathbb{R}^{6\times 6}$ is the 6×6 identity matrix, t_0 is the initial time, and sgn$(\tilde{e}(t))$ denotes the standard signum function applied to each element of the vector $\tilde{e}(t)$. The above estimator is guaranteed to asymptotically identify the signal $\dot{e}(t)$ provided $\dot{e}(t), \ddot{e}(t), \dddot{e}(t) \in \mathcal{L}_\infty$, and the gain matrix ρ satisfies the inequality $\rho_i \geq \left|\ddot{e}_i\right| + \left|\dddot{e}_i\right|, \forall i = 1, 2, ...6$. It is assumed that the relative velocity, acceleration and jerk between the moving object and the camera are bounded, ***i.e.***, $v(t), \dot{v}(t), \ddot{v}(t) \in \mathcal{L}_\infty$. Given these assumptions, the structure of (22) and (23) allows us to show that the bounds on $\dot{e}(t), \ddot{e}(t)$ and $\dddot{e}(t)$ are satisfied. Hence, based on the analysis in [4], $\dot{\hat{e}}(t) \rightarrow \dot{e}(t)$ as $t \rightarrow \infty$. Since $J(t)$ is known and invertible, the six degree-of-freedom relative velocity between the object and the camera can be identified as follows

$$\hat{v}(t) = J^{-1}(t)\left(\dot{\hat{e}}(t) - f(t)\right) \tag{29}$$

and $\hat{v}(t) \rightarrow v(t)$ as $t \rightarrow \infty$.

Remark 3: The definition of relative velocity in (15) can be expressed in the following equivalent form

$$v_r = R^T_{co}v_{cc} - v_{oo}, \qquad \omega_r = R^T_{co}\omega_{cc} - \omega_{oo}\ . \tag{30}$$

Since we made the assumption that the camera velocity $\begin{bmatrix} v^T_{cc} & \omega^T_{cc}\end{bmatrix}^T$ is known, (28) and (30) allows us to recover an estimate for the object velocity $\begin{bmatrix} v^T_{oo} & \omega^T_{oo}\end{bmatrix}^T$.

V. Euclidean Structure Estimation

To facilitate the development of an estimator for Euclidean coordinates of the feature points on the object (***i.e.***, the vector s_i relative to the object frame $\mathcal{O}$, $\bar{m}_i(t)$ and $\bar{m}^*_i$ relative to the camera frame $\mathcal{C}$ for all i feature points on the object), we first define the extended image coordinates $p_{ei}(t) \in \mathbb{R}^3$ as

$$p_{ei} \triangleq \begin{bmatrix} u_i & v_i & -\ln(\alpha_i)\end{bmatrix}^T \tag{31}$$

From the development in the previous section for translational kinematics, the following expression for time derivative of the (31) can be obtained

$$\begin{aligned}\dot{p}_{ei} =& -\frac{\alpha_i}{\bar{m}^*_{iz}}A_{ei}R_{co}v_r + \frac{\alpha_i}{\bar{m}^*_{iz}}A_{ei}S(x_{co})\omega_{cc} \\ &+\frac{\alpha_i}{\bar{m}^*_{iz}}A_{ei}R_{co}S(s_i)\omega_r \\ =& W_{1i}V_{vw}\theta_i + W_{2i}\left[\theta_i\right]_1\end{aligned} \tag{32}$$

where $W_{1i}(\cdot) \in \mathbb{R}^{3\times 3}, W_{2i}(\cdot) \in \mathbb{R}^3, V_{vw}(t) \in \mathbb{R}^{3\times 4}$, and $\theta_i \in \mathbb{R}^4$ are as follows

$$W_{1i} = -\alpha_i A_{ei}R_{co} \tag{33}$$
$$W_{2i} = \alpha_i A_{ei}S(x_{co})\omega_{cc} \tag{34}$$
$$V_{vw} = \begin{bmatrix} v_r & S(\omega_r)\end{bmatrix} \tag{35}$$
$$\theta_i = \begin{bmatrix} \dfrac{1}{\bar{m}^*_{iz}} & \dfrac{s^T_i}{\bar{m}^*_{iz}}\end{bmatrix}^T. \tag{36}$$

In (32), the notation $\left[\theta_i\right]_1$ denotes the first element in the vector θ_i. Note that in (32), we have linearly parameterized the time derivative of the extended image coordinates in terms of known or measurable quantities $W_{1i}(\cdot)$ and $W_{2i}(\cdot)$ and the unknowns $V_{vw}(t)$ and θ_i. An estimate for $V_{vw}(t)$, denoted by $\hat{V}_{vw}(t)$, is available from re-arranging the vector $\hat{v}(t)$ computed from (29). Our objective in this section is to develop an estimator for the unknown constant θ_i in (36) which will allow us to compute the Euclidean coordinates of the i^{th} feature point on the object. To facilitate this objective, the parameter estimation error $\tilde{\theta}_i(t)$ is defined as follows

$$\tilde{\theta}_i \triangleq \theta_i - \hat{\theta}_i \tag{37}$$

where $\hat{\theta}_i(t) \in \mathbb{R}^4$ is a subsequently designed parameter update signal. Motivated by the subsequent stability analysis, we introduce a measurable filter signal $\zeta_i(t) \in \mathbb{R}^{3\times 4}$, and an un-measurable filter signal $\eta_i(t) \in \mathbb{R}^3$ defined as follows

$$\begin{aligned} \dot{\zeta}_i &= -\beta_i\zeta_i + W_{3i} && (38)\\ \dot{\eta}_i &= -\beta_i\eta_i + W_{1i}\tilde{V}_{vw}\theta_i && (39)\end{aligned}$$

where $\beta_i \in \mathbb{R}$ is a scalar positive gain, and $\tilde{V}_{vw}(t) \triangleq V_{vw}(t) - \hat{V}_{vw}(t) \in \mathbb{R}^{3\times 4}$ is an estimation error signal, and

$$W_{3i} = W_{1i}\hat{V}_{vw} + \left[\begin{array}{cc} W_{2i} & 0_{3\times 3}\end{array}\right]. \qquad (40)$$

The stability analysis presented in the next sub-section motivated the following design for estimates of the extended image coordinates $p_{ei}(t)$, denoted by $\hat{p}_{ei}(t) \in \mathbb{R}^3$, and an adaptive least-squares update law [18] for the Euclidean parameters θ_i, as follows

$$\begin{aligned} \dot{\hat{p}}_{ei} &= \beta_i\tilde{p}_{ei} + \zeta_i\,\dot{\hat{\theta}}_i + W_{1i}\hat{V}_{vw}\hat{\theta}_i + W_{2i}\left[\theta_i\right]_1 && (41)\\ \dot{\hat{\theta}}_i &= L_i\zeta_i^T\tilde{p}_{ei} && (42)\end{aligned}$$

where $\tilde{p}_{ei}(t) \triangleq p_{ei}(t) - \hat{p}_{ei}(t) \in \mathbb{R}^3$ denotes the measurable estimation error signal. $L_i(t) \in \mathbb{R}^{4\times 4}$ is an estimation gain recursively computed as

$$\frac{d}{dt}(L_i^{-1}) = \zeta_i^T\zeta_i \qquad (43)$$

and initialized such that $L_i^{-1}(0) > 0$ to ensure that it is positive definite for all time t as required by the stability analysis. From (32) and (41), the time derivative of the estimation error in extended image coordinates can be computed as follows

$$\dot{\tilde{p}}_{ei} = -\beta_i\tilde{p}_{ei} - \zeta_i\,\dot{\hat{\theta}}_i + W_{1i}\tilde{V}_{vw}\theta_i + W_{3i}\tilde{\theta}_i. \qquad (44)$$

The above equation, and (39) allows us to develop an alternate expression for $\tilde{p}_{ei}(t)$ as follows

$$\tilde{p}_{ei} = \zeta_i\tilde{\theta}_i + \eta_i. \qquad (45)$$

A. Stability Analysis

Theorem 1: The update law defined in (42) ensures that $\tilde{\theta}_i(t) \to 0$ as $t \to \infty$ provided that the following persistent excitation condition [18] holds

$$\gamma_1 I_{4\times 4} \leq \int_{t_0}^{t_0+T} \zeta_i^T(\tau)\zeta_i(\tau)d\tau \leq \gamma_2 I_{4\times 4} \qquad (46)$$

and provided that the gains β_i satisfy the following inequalities

$$\begin{aligned} \beta_i &> k_{1i} + k_{2i}\left\|W_{1i}\right\|_\infty^2 && (47)\\ k_{1i} &> 2 && (48)\end{aligned}$$

where $t_0, \gamma_1, \gamma_2, T, k_{1i}, k_{2i} \in \mathbb{R}$ are positive constants, $I_{4\times 4} \in \mathbb{R}^{4\times 4}$ is the 4×4 identity matrix, and the notation $\|.\|_\infty$ denotes the induced ∞-norm of a matrix [18].

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

$$V \triangleq \frac{1}{2}\tilde{\theta}_i^T L_i^{-1}\tilde{\theta}_i + \frac{1}{2}\eta_i^T\eta_i. \qquad (49)$$

After taking the time derivative of (49), the following expression can be obtained

$$\begin{aligned} \dot{V} &= -\frac{1}{2}\left\|\zeta_i\tilde{\theta}_i\right\|^2 - \tilde{\theta}_i^T\zeta_i^T\eta_i - \beta_i\left\|\eta_i\right\|^2 \\ &\quad +\eta_i^T W_{1i}\tilde{V}_{vw}\theta_i \\ &\leq -\frac{1}{2}\left\|\zeta_i\tilde{\theta}_i\right\|^2 - \beta_i\left\|\eta_i\right\|^2 \\ &\quad +\left\|\theta_i\right\|\left\|W_{1i}\right\|_\infty\left\|\tilde{V}_{vw}\right\|_\infty\left\|\eta_i\right\| \\ &\quad +\left\|\zeta_i\tilde{\theta}_i\right\|\left\|\eta_i\right\| - k_{1i}\left\|\eta_i\right\|^2 + k_{1i}\left\|\eta_i\right\|^2 \\ &\quad +k_{2i}\left\|W_{1i}\right\|_\infty^2\left\|\eta_i\right\|^2 - k_{2i}\left\|W_{1i}\right\|_\infty^2\left\|\eta_i\right\|^2 \qquad (50)\end{aligned}$$

where $k_{1i}, k_{2i} \in \mathbb{R}$ are positive constants as previously mentioned. Further simplification of (50) after utilizing the non-linear damping argument [13] results in the following expression

$$\begin{aligned} \dot{V} &\leq -\left(\frac{1}{2} - \frac{1}{k_{1i}}\right)\left\|\zeta_i\tilde{\theta}_i\right\|^2 \\ &\quad -\left(\beta_i - k_{1i} - k_{2i}\left\|W_{1i}\right\|_\infty^2\right)\left\|\eta_i\right\|^2 \\ &\quad +\frac{1}{k_{2i}}\left\|\theta_i\right\|^2\left\|\tilde{V}_{vw}\right\|_\infty^2. \qquad (51)\end{aligned}$$

The gains β_i, k_{1i} and k_{2i} are selected to ensure that

$$\begin{aligned} \frac{1}{2} - \frac{1}{k_{1i}} &\geq \mu_{1i} > 0 && (52)\\ \beta_i - k_{1i} - k_{2i}\left\|W_{1i}\right\|_\infty^2 &\geq \mu_{2i} > 0 && (53)\end{aligned}$$

where $\mu_{1i}, \mu_{2i} \in \mathbb{R}$ are positive constants. The above gain conditions allow us to further upper-bound the time derivative of (49) in the following manner

$$\dot{V} \leq -\mu_{1i}\left\|\zeta_i\tilde{\theta}_i\right\|^2 - \mu_{2i}\left\|\eta_i\right\|^2 + \frac{1}{k_{2i}}\left\|\theta_i\right\|^2\left\|\tilde{V}_{vw}\right\|_\infty^2. \qquad (54)$$

In the analysis provided in the development of the kinematic estimator in [4], it was shown that a filter signal $r(t) \in \mathbb{R}^6$ defined as $r(t) = \tilde{e}(t) + \dot{\tilde{e}}(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2$. From this result it is easy to show that the signals $\tilde{e}(t)$, $\dot{\tilde{e}}(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2$ [6]. Since $J(t) \in L_\infty$ and invertible, it follows that $\tilde{v}(t) = J^{-1}(t)\,\dot{\tilde{e}}(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2$. Hence it follows that $\left\|\tilde{V}_{vw}(t)\right\|_\infty^2 \in \mathcal{L}_1$ and

$$\int_0^\infty \frac{1}{k_{2i}}\left\|\theta_i(\tau)\right\|^2\left\|\tilde{V}_{vw}(\tau)\right\|_\infty^2 d\tau \leq \varepsilon \qquad (55)$$

where $\varepsilon \in \mathbb{R}$ is a positive constant. From (49), the integral of (54), and (55), it can be concluded that

$$\begin{aligned} &\int_0^\infty\left(\mu_{1i}\left\|\zeta_i(\tau)\tilde{\theta}_i(\tau)\right\|^2 + \mu_{2i}\left\|\eta_i(\tau)\right\|^2\right)d\tau \\ &\leq V(0) - V(\infty) + \varepsilon. \qquad (56)\end{aligned}$$

Hence, $\zeta_i(t)\tilde{\theta}_i(t), \eta_i(t) \in \mathcal{L}_2$. Also, from (56) and the fact that $V(t)$ is non-negative, it can be concluded that $V(t) \leq V(0) + \varepsilon \in \mathcal{L}_\infty$. Therefore, from (49), $\tilde{\theta}_i^T(t)L_i^{-1}(t)\tilde{\theta}_i(t), \eta_i(t) \in \mathcal{L}_\infty$. Based on the assumption that the persistent excitation condition in (46) is satisfied, and $L_i^{-1}(0)$ is chosen to be positive definite, we can use (43) to show that $L_{\bar{J}}^{-1}(t)$ is always positive definite; hence, it must follow that $\dot{\theta}_i(t) \in \mathcal{L}_\infty$. Since $W_{3i}(\cdot)$ in (40) is composed of bounded terms, and is the input to the stable filter in (38), it can be shown that $\zeta_i(t), \dot{\zeta}_i(t) \in \mathcal{L}_\infty$ [6], and consequently, $\zeta_i(t)\theta_i(t) \in \mathcal{L}_\infty$. It follows from (45) that $\tilde{p}_{ei}(t) \in \mathcal{L}_\infty$, and hence, from (42), $\dot{\hat{\theta}}_i(t), \dot{\tilde{\theta}}_i(t) \in \mathcal{L}_\infty$. Based on these arguments, it is easy to see that $\frac{d}{dt}\left(\zeta_i(t)\tilde{\theta}_i(t)\right) \in \mathcal{L}_\infty$. Therefore, $\zeta_i(t)\tilde{\theta}_i(t)$ is uniformly continuous [7], and since we also have that $\zeta_i(t)\tilde{\theta}_i(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2$, we can conclude the following result from Barbalat's Lemma [7]

$$\zeta_i(t)\tilde{\theta}_i(t) \to 0 \text{ as } t \to \infty. \tag{57}$$

If the signal $\zeta_i(t)$ satisfies the persistent excitation condition given in (46), then it can be concluded [18] from (57) that

$$\tilde{\theta}_i(t) \to 0 \text{ as } t \to \infty. \tag{58}$$

□

Remark 4: From (7), (9), (36), (42), and Theorem 1, the constant Euclidean position of all feature points on the object at the reference position, and the time varying Euclidean position of the feature points on the moving object relative to the camera can be computed as follows

$$\hat{\bar{m}}_i^*(t) = \frac{1}{\left[\hat{\theta}_i(t)\right]_1} A^{-1} p_i^* \tag{59}$$

$$\hat{\bar{m}}_i(t) = \frac{1}{\alpha_i(t)\left[\hat{\theta}_i(t)\right]_1} A^{-1} p_i(t). \tag{60}$$

VI. Conclusions

This work developed a technique for the recovery of structure and motion for the general case where both the object of interest as well as the camera are in motion. Homography-based techniques and adaptive estimation theory provided the basis for the estimator design. Future work will include the validation of the proposed estimator through simulations and experiments. The applicability of this work for tasks such as aerial surveillance of moving targets is well motivated since no explicit model describing the object or its motion is required by the estimator.

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] 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.

[3] 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," *Proc. of the 2005 American Control Conference*, pp. 4601-4606, June 2005.

[4] V. K. 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.

[5] 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.

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

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

[8] 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.

[9] J. M. Ferryman, S. J. Maybank, and A. D. Worrall, "Visual Surveillance for Moving Vehicles," *Intl. Journal of Computer Vision*, Vol. 37, No. 2, pp. 187-197, 2000.

[10] T. Fukao, K. Fujitani, and T. Kanade, "An Autonomous Blimp for Surveillance System," *Proc. of the 2003 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, pp. 1820-1825, October 2003.

[11] R. Horn, and C. Johnson, *Matrix Analysis*, Cambridge University Press, ISBN: 0521305861, 1985.

[12] T. Kanade, O. Amidi, Q. Ke, "Real-Time and 3D Vision for Autonomous Small and Micro Air Vehicles," *Proc. of the 2004 IEEE Conf. on Decision and Control*, pp. 1655-1662, Bahamas, 2004.

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

[14] Y. Ma, S. Soatto, J. Košecká, and S. Sastry, *An Invitation to 3D Vision*, Springer-Verlag, ISBN: 0387008934, 2003.

[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, 2000, pp. 79–97.

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

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

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

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

[20] B. Xian, M. S. de Queiroz, D. M. Dawson, and M. L. McIntyre, "A Discontinuous Output Feedback Controller and Velocity Observer for Nonlinear Mechanical Systems," *Automatica*, Vol. 40, No. 4, pp. 695-700, April 2004.

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

Appendix I
Development of Translation and Rotation Kinematics

To facilitate the development of the kinematic equations, the following properties of rotational matrices and skew symmetric matrices are utilized [19]

$$\dot{R}_c = R_c S(\omega_{cc}) \tag{61}$$
$$\dot{R}_c = S(\omega_c) R_c \tag{62}$$
$$\dot{R}_o = S(\omega_o) R_o \tag{63}$$
$$S(u)\zeta = -S(\zeta) u \tag{64}$$
$$S(Ru) = RS(u)R^T \tag{65}$$

where $R(t) \in SO(3)$ is a rotation matrix, $u(t), \zeta(t) \in \mathbb{R}^3$ are some arbitrary vectors, and the rest of the terms were defined previously.

To develop the translational kinematics, the time derivative of (13) is obtained as follows

$$\dot{e}_v = \frac{\alpha_1}{\bar{m}_{iz}^*} A_{e1}\, \dot{\bar{m}}_1 \tag{66}$$

From Figure 1, the position and orientation of the object relative to the camera are as follows

$$x_{co} = R_c^T (x_o - x_c) \tag{67}$$

$$R_{co} = R_c^T R_o. \tag{68}$$

After taking the time derivative of the above equations, and utilizing the properties in (61) to (65), we have

$$\dot{x}_{co} = S(x_{co})\omega_{cc} + R_c^T(v_o - v_c) \tag{69}$$

$$\dot{R}_{co} = R_c^T S(\omega_o) R_o - R_c^T S(\omega_c) R_o. \tag{70}$$

Utilizing (69) and (70), the expression for the time derivative of (1) for $i = 1$ is given by the following expression

$$\dot{\bar{m}}_1 = S(x_{co})\omega_{cc} - R_{co} v_r + R_{co} S(s_1)\omega_r. \tag{71}$$

After substituting (71) in (66), the translational kinematics in (14) can be obtained.

The development of rotation kinematics follows the description given in [16], to which the reader is referred to for details. Only a brief description is given here. We define $\bar{\omega}(t) \in \mathbb{R}^3$ to be the relative rotational velocity between the frames $\mathcal{O}^*$ and $\mathcal{O}$ as observed by the camera. Then it can be shown that [19]

$$\dot{\bar{R}} = S(\bar{\omega})\bar{R}. \tag{72}$$

The open-loop error system for $e_\omega(t)$ is derived based on the following exponential parameterization [19]

$$\bar{R} = I_{3\times3} + S(\mu)\sin(\phi) + 2\sin^2(\frac{\phi}{2})S(\mu)^2 \tag{73}$$

where $I_{3\times3}$ is the 3×3 identity matrix. Using (73) and its time derivative in (72), it can be shown that

$$S(\bar{\omega}) = \sin(\phi)S(\dot{\mu}) + S(\mu)\dot{\phi} + (1 - \cos(\phi))\, S(S(\mu)\dot{\mu}). \tag{74}$$

To facilitate further development, the time derivative of (17) is determined as follows

$$\dot{e}_\omega = \dot{\mu}\phi + \mu\dot{\phi}. \tag{75}$$

By multiplying (75) by $\left(I_{3\times3} + S(\mu)^2\right)$, the following expression can be obtained

$$\left(I_{3\times3} + S(\mu)^2\right)\dot{e}_\omega = \mu\dot{\phi} \tag{76}$$

where we utilized the following properties

$$\mu^T\mu = 1, \qquad \mu^T\dot{\mu} = 0 \tag{77}$$

$$S(\mu)^2 = \mu\mu^T - I_{3\times3} \tag{78}$$

Likewise, by multiplying (75) by $-S(\mu)^2$ and then utilizing (77), the following expression is obtained

$$-S(\mu)^2\dot{e}_\omega = \dot{\mu}\phi. \tag{79}$$

From the expression in (74), the properties given in (64), (75), (76), (79), and the fact that

$$\sin^2(\phi) = \frac{1}{2}(1 - \cos(2\phi)),$$

we cab obtain the following expression

$$\bar{\omega} = L_\omega^{-1}\dot{e}_\omega \tag{80}$$

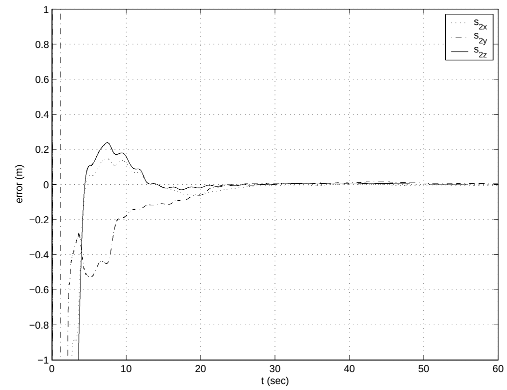


Fig. 2. Estimation error in the coordinates of a feature point on the object (fixed camera, 1% pixel noise added).

where $L_\omega(t)$ is defined in (20). From (72), (4), (68), and the properties in (62) and (63), it can be shown that

$$\bar{\omega} = -R_{co}\omega_r \tag{81}$$

where $\omega_r(t) \in \mathbb{R}^3$ was defined in (15). After multiplying both sides of (80) by $L_\omega(t)$, and utilizing (81), the rotational kinematics given in (19) can be obtained.

Appendix II
Special Cases: A Fixed Camera, or a Fixed Object

In the previous sections, we presented the development for the general case where both the camera and the object are in motion relative to an inertial frame. As described in Sections IV and V, such a treatment imposed a few restrictive assumptions on the system, chief among them being that the persistent excitation condition of (46) was satisfied at all times, the camera velocity $\left[\begin{array}{cc} v_{cc}^T(t) & \omega_{cc}^T(t) \end{array}\right]^T$ was measurable, a single position vector s_1 was known, and the orientation of the object at a reference position relative to the camera (R_{co}^*) was known. The fact that s_1 and R_{co}^* was available enabled us to compute the depth information $\bar{m}_{iz}^*$ and the position vector $x_{co}(t)$, apart from the rotation matrix $R_{co}(t)$. As shown below, for the special cases where either the object or the camera are stationary, some of these requirements can be relaxed, increasing their potential for application in a wider range of real-world scenarios. These cases are derived in detail in [3], and here we present a summary.

1) Fixed Camera: When the camera is fixed relative to the inertial frame, the kinematics expression in (21) is given by

$$\dot{e} = J_{fc} v_{oo} \tag{82}$$

where $J_{fc}(t) \in \mathbb{R}^{6\times6}$ is the following

$$J_{fc} = \begin{bmatrix} \frac{\alpha_1}{\bar{m}_{iz}^*} A_{e1} R_{co} & -\frac{\alpha_1}{\bar{m}_{iz}^*} A_{e1} R_{co} S(s_1) \\ 0_{3\times3} & L_\omega R_{co} \end{bmatrix}. \tag{83}$$

The time derivative of extended image coordinates in (32) reduces to the expression given below

$$\dot{p}_{ei} = W_{1i} V_{vw} \theta_i \tag{84}$$

where $W_{1i}(\cdot) = \alpha_i A_{ei} R_{co} \in \mathbb{R}^{3\times 3}$ and $V_{vw} = \begin{bmatrix} v_{oo} & S(\omega_{oo}) \end{bmatrix} \in \mathbb{R}^{3\times 4}$. All assumptions and requirements are the same as in the general case, except that the estimation of the object velocity and structure relative to the fixed camera requires no additional sensors on the camera (since $\omega_{cc}(t) = 0$), and the computation of the position vector $x_{co}(t)$ is not required. As an example, Figure 2 shows the errors in the estimation of the Euclidean coordinates of one of the features points, in the simulation of a planar object moving in the field of view of a fixed camera.

2) Moving Camera: With the object stationary relative to the inertial frame, the kinematics of the moving camera can be expressed as follows

$$\dot{e} = J_{mc} v_{cc} \tag{85}$$

where $J_{mc}(t) \in \mathbb{R}^{6\times 6}$ is the following Jacobian-like matrix

$$J_{mc} = \begin{bmatrix} -\dfrac{\alpha_1}{\bar{m}_{iz}^*} A_{e1} & A_{e1} S(m_1) \\ 0_{3\times 3} & -L_\omega \end{bmatrix}. \tag{86}$$

In the above expression, notice that all terms are either known apriori, or directly measurable, except the constant depth $\bar{m}_{iz}^*$. If the camera can be moved away from its reference position by a known translation vector $\bar{x}_k \in \mathbb{R}^3$, then $\bar{m}_{iz}^*$ can be computed offline without the knowledge of R_{co}^* and s_1. Decomposition of the Euclidean homography between the normalized Euclidean coordinates of the feature points obtained at the reference position, and at $\bar{x}_k$ away from the reference position, respectively, can yield the scaled translation vector $\dfrac{\bar{x}_k}{d^*} \in \mathbb{R}^3$. Then, it can be seen that[1]

$$\bar{m}_{iz}^* = \frac{d^*}{n^{*T} m_1^*} = \frac{d^*}{n^{*T} A^{-1} p_1^*}. \tag{87}$$

Hence, from the above discussion, we note that it is possible to develop a velocity estimator for the moving camera case without having to know the reference orientation matrix R_{co}^* and any geometric length s_1 on the object. However, as a consequence of the specific approach we have employed in the development of the Euclidean estimator in Section V, we are unable to benefit from this relaxation in the assumptions for velocity estimation. In [3], we formulate the Euclidean estimation in a slightly different manner that allows us to eliminate the requirement of R_{co}^* and s_1; *i.e.*, the time derivative of the extended pixel coordinates in (32) is expressed as

$$\dot{p}_{ei} = W_{1i} v_{cc} \theta_i + W_{2i} \omega_{cc} \tag{88}$$

where $W_{1i}(\cdot) \in \mathbb{R}^{3\times 3}, W_{2i}(\cdot) \in \mathbb{R}^{3\times 3}$, and $\theta_i \in \mathbb{R}$ are defined as follows

$$\begin{aligned} W_{1i} &= -\alpha_i A_{ei} && (89) \\ W_{2i} &= A_{ei} S(m_i) && (90) \\ \theta_i &= \frac{1}{\bar{m}_{iz}^*}. && (91) \end{aligned}$$

A measurable filter signal $\zeta_i(t) \in \mathbb{R}^3$, similar to (38), is now defined in the following manner

$$\dot{\zeta}_i = -\beta_i \zeta_i + W_{1i} \hat{v}_{cc} \tag{92}$$

where $\hat{v}_{cc}(t) \in \mathbb{R}^3$ is the estimate for translational velocity of the camera. In [3], following an approach similar to that presented in Section V, it is shown that the depth parameters θ_i can be estimated using the adaptive least-squares update law of (42), and subsequently, the Euclidean position of all feature points on the object relative to the camera can be estimated. Again, the need for additional on-board sensing is eliminated. Also note that as a consequence of the way the filter signal $\zeta_i(t)$ in (92) is formulated, the persistent excitation of (46) is also simplified, as it now depends only on the translational velocity signal.

[1]Note that for any feature point O_i coplanar with π^*, $\bar{m}_{iz}^*$ could be computed this way.