Euclidean Position Estimation of Features on an Object Using a Single Camera: A Lyapunov-Based Approach

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

Abstract

In this paper[1], 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. An extension of this development to the dual case of camera-in-hand is also presented. The performance of the estimator is demonstrated by simulation and experimental results.

V. K. Chitrakaran, D. M. Dawson and J. Chen are with Clemson University. W. E. Dixon is with the University of Florida.

[1]A preliminary version of this paper appeared in [7].

Euclidean Position Estimation of Features on an Object Using a Single Camera: A Lyapunov-Based Approach

I. 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 [2], [9], [24]. In recent publications, some researchers have recast the problem as the 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 [17], [18]. To summarize, these papers show that if the velocity of the moving object (or camera) is known, and satisfies certain obervability conditions, an estimator for the unknown Euclidean position of the feature points can be developed. In [5], 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 [27], the rotational velocity could be fully recovered; furthermore, 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 the 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 [5] 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 [3] and [22], 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 [8] 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 [11], 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.

II. 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 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

$$\bar{m}_i \triangleq \left[\begin{array}{ccc} x_i & y_i & z_i \end{array} \right]^T. \tag{1}$$

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

$$\bar{m}_i = x_f + R s_i \tag{2}$$

$$\bar{m}_i^* = x_f^* + R^* s_i \ . \tag{3}$$

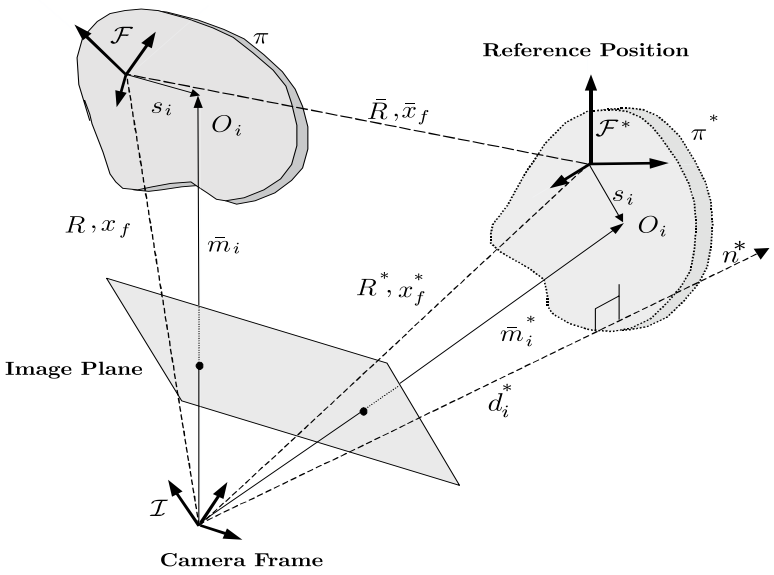

Fig. 1. Geometric relationships between a fixed camera and the current and reference positions of a moving object in its field of view.

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

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

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 $H(t) \in \mathbb{R}^{3\times 3}$ denotes a Euclidean homography [14].

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 pin-hole model of [13] such that

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

where $A \in \mathbb{R}^{3\times 3}$ is a known, constant, upper triangular and invertible intrinsic camera calibration matrix [23]. 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 $G(t) \in \mathbb{R}^{3\times 3}$ defined in (11) is a full rank homogeneous collineation matrix defined upto a scale factor [23]. 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 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 [23] 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 I. 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 [14], [29]), $H(t)$ can be decomposed into its constituent rotation matrix $\bar{R}(t)$, unit normal vector n^*, scaled translation vector $\bar{x}_h(t) \triangleq \dfrac{\bar{x}_f(t)}{d^*}$ 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.

III. 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

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

where $\ln(\cdot) \in \mathbb{R}$ denotes the natural logarithm. 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 [8], the translational kinematics can be obtained as follows

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

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

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

Similarly, to quantify the rotation of $\mathcal{F}$ relative to $\mathcal{F}^*$, we define $e_\omega(t) \in \mathbb{R}^3$ using the axis-angle representation [28] as follows

$$e_\omega \triangleq u\phi \tag{15}$$

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 [8] for further details)

$$\dot{e}_\omega = L_\omega R \omega_e \ . \tag{16}$$

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

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

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

$$\dot{e} = Jv \tag{18}$$

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

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

where $0_3 \in \mathbb{R}^{3\times 3}$ 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 [8] 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), $e(t) \in L_\infty$. It is also assumed that the object velocity, acceleration and jerk are bounded, *i.e.*, $v(t), \dot{v}(t), \ddot{v}(t) \in L_\infty$; hence the structure of (18) allows us to show that $\dot{e}(t), \ddot{e}(t), \dddot{e}(t) \in L_\infty$.

IV. Identification of Velocity

In [8], 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

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

where $\tilde{e}(t) \triangleq e(t) - \hat{e}(t) \in \mathbb{R}^6$ is the estimation error for the signal $e(t)$, and $K, \rho \in \mathbb{R}^{6\times 6}$, 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 $\text{sgn}(\tilde{e}(t))$ denotes the standard signum function applied to each element of the vector $\tilde{e}(t)$. The reader is referred to [8] and the references therein for analysis pertaining to the development of the above estimator. In essense, it was shown in [8] that the above estimator asymptotically identifies the signal $\dot{e}(t)$ provided the following inequality is satisifed for each diagonal element ρ_i of the gain matrix ρ,

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

Hence, $\dot{\hat{e}}_i(t) \to \dot{e}_i(t)$ as $t \to \infty, \forall i = 1, 2, ...6$. Since $J(t)$ is known and invertible, the six degree-of-freedom velocity of the moving object can be identified as follows

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

V. 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

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

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

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

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

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

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 signal, 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 $W_{fi}(t) \in \mathbb{R}^{3\times 4}$, 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 $\tilde{V}_{vw}(t) \triangleq V_{vw}(t) - \hat{V}_{vw}(t) \in \mathbb{R}^{3\times 4}$ is an estimation error signal.

Motivated by the subsequent stability analysis, we design the following estimator, 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 [27] for $\hat{\theta}_i(t)$

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

where $L_i(t) \in \mathbb{R}^{4\times 4}$ 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 [8], 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 $\tilde{e}(t)$, $\dot{\tilde{e}}\ (t) \in L_2$ [10]. 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 [19].

A. 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 [27] 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, $I_4 \in \mathbb{R}^{4\times 4}$ is the 4×4 identity matrix, the notation $\|.\|_\infty$ denotes the induced ∞-norm of a matrix [19], 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

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

After utilizing the nonlinear damping argument [20], we can simplify (40) further as follows

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

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

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

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

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

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

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 [8], 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, $W_{fi}(t), \dot{W}_{fi}(t) \in L_\infty$ [10], 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 [12]. Since we also have that $W_{fi}(t)\tilde{\theta}_i(t) \in L_2$, we can conclude that [12]

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

As shown in Appendix III, if the signal $W_{fi}(t)$ satisifies the persistent excitation condition [27] given in (36), then it can be concluded from (47) that

$$\tilde{\theta}_i(t) \to 0 \text{ as } t \to \infty. \tag{48}$$

□

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 [25]. Hence, the condition in (36) is satisfied if

$$\gamma_3 I_4 \leq \quad \int_{t_0}^{t_0+T} \underbrace{\hat{V}_{vw}^T(\tau)W_i^T(\tau)W_i(\tau)\hat{V}_{vw}(\tau)}_{W}\, d\tau \quad \leq \gamma_4 I_4 \tag{49}$$

where $\gamma_3, \gamma_4 \in \mathbb{R}$ are positive constants. It can be shown upon expansion of the integrand $W(t) \in \mathbb{R}^{4\times4}$ of (49) 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. It should be noted that the translational velocity of the object has no bearing on the convergence of the remaining three elements of $\hat{\theta}_i(t)$, and 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 $R^* \in \mathbb{R}^{3\times3}$ and a single geometric length $s_1 \in \mathbb{R}^3$ on the object. After utilizing (8), (10), (27) and (34), the estimates for $\bar{m}_i^*$, denoted by $\hat{\bar{m}}_i^*(t) \in \mathbb{R}^3$, can be obtained as follows

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

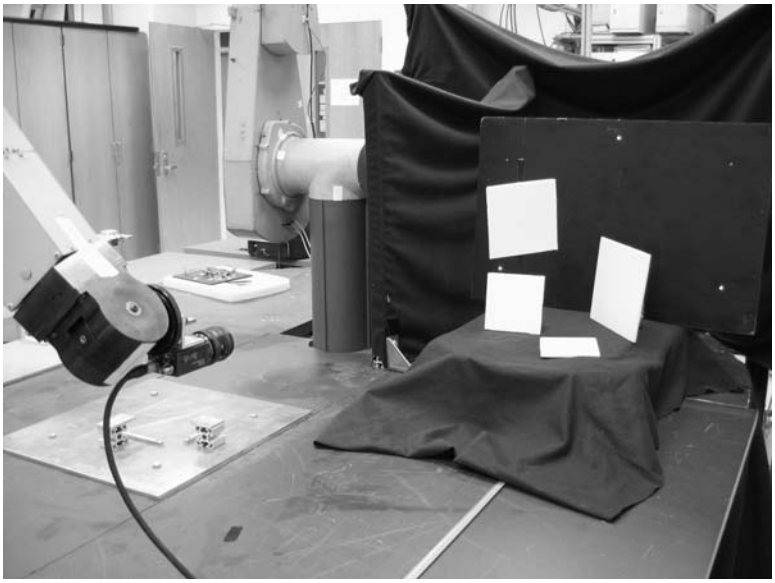

Fig. 2. The experimental test-bed.

where the term in the denominator denotes the first element of the vector $\hat{\theta}_i(t)$. Similarly, the estimates for the time varying Euclidean position of the feature points on the object relative to the camera frame, denoted by $\hat{\bar{m}}_i(t) \in \mathbb{R}^3$, can be calculated as follows

$$\hat{\bar{m}}_i(t) = \frac{1}{\alpha_i(t) \left[\hat{\theta}_i(t)\right]_1} A^{-1} p_i(t). \tag{51}$$

VI. Simulations and Experimental Results

A practical implementation of the estimator described in this paper consists of at least four sub-components: (a) hardware for image acquisition, (b) implementation of an algorithm to extract and track feature points between image frames, (c) the implementation of the depth estimation algorithm itself, and (d) a method to display the reconstructed scene (3D models, depth maps, etc.). This section presents simulation results for the fixed camera case described in the body of the paper, and the camera-in-hand camera case developed as an extension in Appendix II. Experimental results are provided for the camera-in-hand system.

Experimental testbed: As shown in Figure 2, the camera-in-hand system was a calibrated monochrome CCD camera (Sony XC-ST50) mounted on the end-effector of a Puma 560 industrial robotic manipulator whose end-effector was commanded via a PC to move along a smooth trajectory. The camera was interfaced to a second PC dedicated to image processing, and equipped with an Imagenation PXC200AF framegrabber board capable of acquiring images in real time (30 fps) over the PCI bus. A 20 Hz digital signal from the robot control PC triggered the framegrabber to acquire images. The same trigger signal also recorded the actual robot end-effector velocity (which is same as the camera velocity) into a file, which was utilized as ground truth to validate the performance of the estimator in identifying the camera velocity. The computational complexity of feature tracking and depth estimator algorithms made it unfeasible to acquire and process images at the frame rate of the camera. Hence the sequence of images acquired from the camera were encoded into a video file (AVI format, utilizing the lossless huffman codec for compression, where possible) to be processed offline later.

Feature tracking: The image sequences were at least a minute long, and due to thousands of images that must be processed in every sequence, an automatic feature detector and tracker was necessary. We utilized an implementation of the Kanade-Lucas-Tomasi feature tracking algorithm [21] available at [1] for detection and tracking of feature points from one frame to next. The libavformat and libavcodec libraries from the FFmpeg project [15] were utilized to extract individual frames from the video files. The output of the feature point tracking stage was a data file containing image space trajectories of all successfully tracked feature points from the video sequence. This data served as the input to the depth estimation algorithm.

Depth Estimator: The adaptive estimation algorithm described in Section V was implemented in C++ and ran at a sampling frequency of 1 kHz to guarantee sufficient accuracy from the numerical integrators in the estimator. A linear interpolator, followed by 2nd order low pass filtering was used to interpolate 20 Hz image data to 1 kHz required as input to the estimator.

Visualization: A 3D graphical display based on the OpenGL API was developed in order to render the reconstructed scene. A surface model of objects in the scene can be created by generating a 'mesh' from the reconstructed feature points. Using OpenGL libraries, it is easy to develop a display that allows the user to view the objects in the scene from any vantage point by virtually navigating around the objects in a scene using an input device such as a computer mouse pointer.

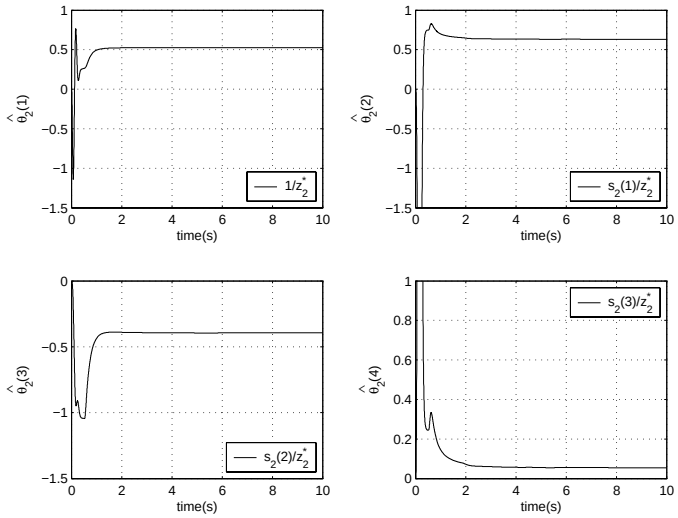

Fig. 3. The estimated parameters for the second feature point on the moving object (fixed camera simulation).

TABLE I

ESTIMATED DIMENSIONS FROM THE SCENE (CAMERA-IN-HAND EXPERIMENT).

Object	Actual dim. (cm.)	Estimated dim. (cm.)
Length I	23.6	23.0
Length II	39.7	39.7
Length III	1.0	1.04
Length IV	13.0	13.2
Length V	100.0	99.0
Length VI	19.8	19.6
Length VII	30.3	30.0

Simulation Results: For simulations, the image acquisition hardware and the image processing step were both replaced with a software component that generated feature trajectories utilizing object kinematics. For the fixed camera simulation, we selected a planar object with four feature points initially 2 meters away along the optical axis of the camera as the body undergoing motion. The velocity of the object along each of the six degrees of freedom were 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

$$\begin{aligned} s_1 &= [\ 1.0 \quad 0.5 \quad 0.1\ ]^T \\ s_2 &= [\ 1.2 \quad -0.75 \quad 0.1\ ]^T \\ s_3 &= [\ 0.0 \quad -1.0 \quad 0.1\ ]^T \\ s_4 &= [\ -1.0 \quad 0.5 \quad 0.1\ ]^T. \end{aligned} \tag{52}$$

The object's reference orientation R^* relative to the camera was selected as $diag(1,-1,-1)$, where $diag(.)$ denotes a diagonal matrix with arguments as the diagonal entries. The estimator gain β_i was set to 20 for all i feature points. As an example, Figure 3 depicts the convergence of $\hat{\theta}_2(t)$ from which the Euclidean coordinates of the second feature point s_2 could be computed as shown in (27). Similar graphs were obtained for convergence of the estimates for the remaining feature points. In the simulation for the camera-in-hand system, we selected a non-planar stationary object with 12 feature points. As shown in the development in Appendix II, the camera-in-hand system directly estimates the inverse of depth ($\frac{1}{z_i^*}$) for all i feature points relative to the camera at its reference position. Figure 4 shows the convergence of the inverse depth estimates, and Figure 5 shows the estimation errors.

Experimental Results: An experimental scene with a doll-house as shown in Figure 6 was utilized to verify the practical performance of the camera-in-hand Euclidean estimator. Figure 7 shows the reconstructed wire-frame view of the doll-house in the scene, displayed using the OpenGL based viewer. Table I shows a comparison between the actual and the estimated lengths in the scene. The time evolution of the inverse depth estimates are shown in Figure 8.

Discussion: The simulation and experimental results in Figures 3, 4, 5 and Table I clearly demonstrate good performance of the estimator. All estimates converge to their expected values in a span of a few seconds. The only difference between the simulations and the experiment is the source of input pixel data. As mentioned previously, in simulations, we generated image space feature point trajectories based on rigid body kinematics and known dimensions of the object. In the case of experiments,

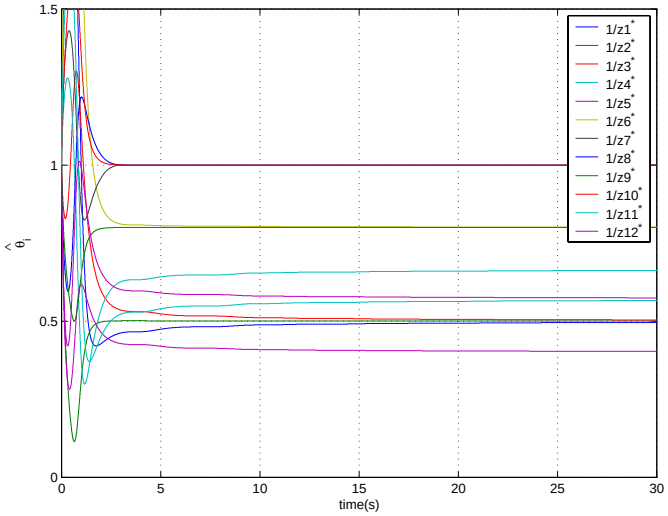

Fig. 4. Inverse depth estimates for all feature points on the stationary object (camera-in-hand simulation).

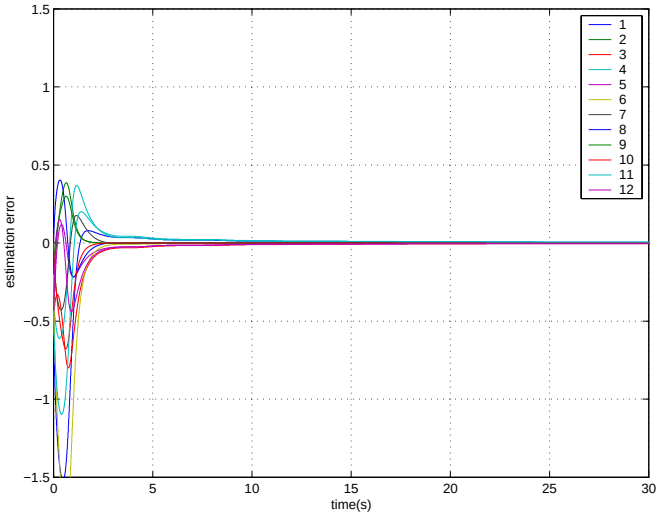

Fig. 5. Error in depth estimation reduce to zero with time (camera-in-hand simulation).

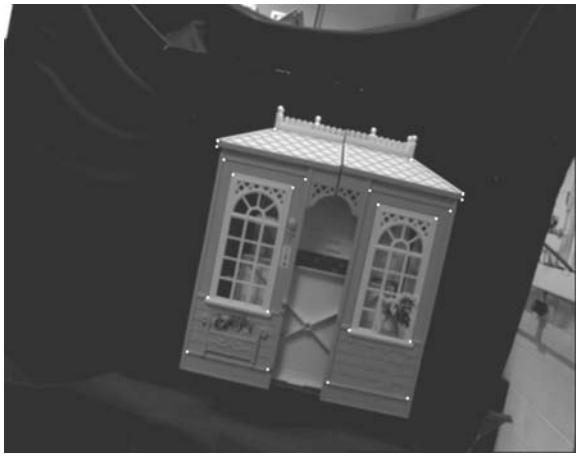

Fig. 6. One of the frames from the doll-house video sequence. The white dots overlayed on the image are the tracked feature points.

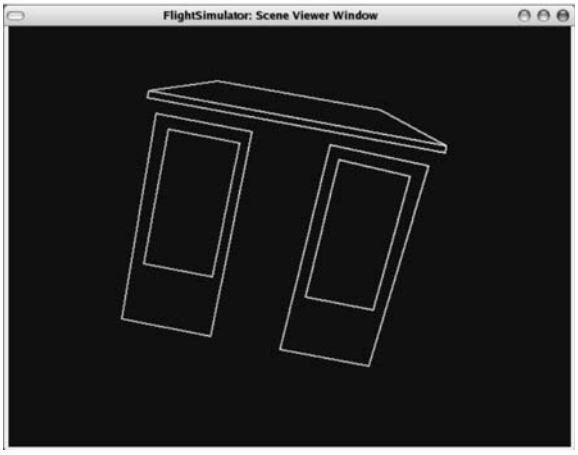

Fig. 7. A wire-frame reconstruction of the doll-house scene (camera-in-hand experiment).

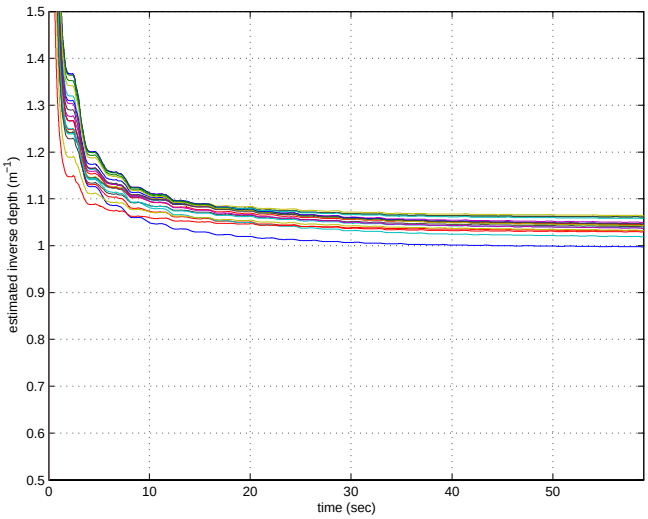

Fig. 8. The time evolution of inverse depth estimates from the camera-in-hand experiment.

the image space trajectories were obtained from the feature tracker that was run on the image sequence previously recorded from the camera. One of the most challenging aspects in a real-world implementation of the estimator is accurate feature tracking. Features that can be tracked well by a tracking algorithm may not adequately describe the geometrical structure of the object, and vice-versa. Quality of feature tracking also depends on factors such as lighting, shadows, magnitude of inter-frame motion, and texturedness, to name a few. Additionally, any compression (such as MPEG) employed in the input video sequence will introduce additional artifacts that can potentially degrade performance of the estimator. There is no one common solution to this problem, and the parameters of the tracker must be tuned to the specific image dataset in hand to obtain best results. See [26] for a discussion on issues related to feature selection and tracking. Apart from the accuracy in feature tracking that directly effects the accuracy in online estimation of the homography matrix relating corresponding feature points, the performance of the estimator also depends on accurate camera calibration. Since the implementation is the same for simulations and the experiment, the errors in estimation (Table I) can mostly be attributed to the jitter in the output of the feature tracker.

VII. 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. 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. Simulation and experimental results demonstrated the performance of the estimator. We have recently extended this technique to the general case where both the camera and the objects of interest in its field of view may be in motion. The reader is referred to [6] for

more details. The applicability of this new development to tasks such as aerial surveillance of moving targets is well motivated since no explicit model describing the object motion is required by the estimator.

APPENDIX I
CALCULATION OF HOMOGRAPHY

A. Homography from Coplanar Feature Points

In this section, we present a method to estimate the collineation $G(t)$ by solving a set of linear equations (11) obtained from at least four corresponding feature points that are coplanar but non-collinear. Based on the arguments in [16], a transformation is applied to the projective coordinates of the corresponding feature points to improve the accuracy in the estimation of $G(t)$. The transformation matrices, denoted by $P(t)$, $P^* \in \mathbb{R}^{3\times 3}$, are defined in terms of the projective coordinates of three of the coplanar non-collinear feature points as follows

$$P \triangleq \left[\begin{array}{ccc} p_1 & p_2 & p_3 \end{array} \right] \qquad P^* \triangleq \left[\begin{array}{ccc} p_1^* & p_2^* & p_3^* \end{array} \right]. \tag{53}$$

From (11) and (53), it is easy to show that

$$P\tilde{G} = GP^* \tag{54}$$

where

$$\begin{aligned} \tilde{G} &= P^{-1}GP^* \\ &= diag\left(\alpha_1^{-1},\, \alpha_2^{-1},\, \alpha_3^{-1}\right) \\ &\triangleq diag\left(\tilde{g}_1,\, \tilde{g}_2,\, \tilde{g}_3\right). \end{aligned} \tag{55}$$

In (55), $diag(.)$ denotes a diagonal matrix with arguments as the diagonal entries. Utilizing (55), the relationship in (11) can now be expressed in terms of $\tilde{G}(t)$ as follows

$$q_i = \alpha_i \tilde{G} q_i^* \tag{56}$$

where

$$q_i = P^{-1}p_i \tag{57}$$

$$q_i^* = P^{*^{-1}}p_i^* \tag{58}$$

define the new transformed projective coordinates. Note that the transformation normalizes the projective coordinates, and it is easy to show that $\left[\begin{array}{ccc} q_1 & q_2 & q_3 \end{array} \right] = \left[\begin{array}{ccc} q_1^* & q_2^* & q_3^* \end{array} \right] = I_3 \in \mathbb{R}^{3\times 3}$ where I_3 is the 3×3 identity matrix. Given the transformed image coordinates of a fourth matching pair of feature points $q_4(t) \triangleq \left[\begin{array}{ccc} q_{4u}(t) & q_{4v}(t) & q_{4w}(t) \end{array} \right]^T \in \mathbb{R}^3$ and $q_4^* \triangleq \left[\begin{array}{ccc} q_{4u}^* & q_{4v}^* & q_{4w}^* \end{array} \right]^T \in \mathbb{R}^3$, we have

$$q_4 = \alpha_4 \tilde{G} q_4^*, \tag{59}$$

where it can be shown that

$$\begin{aligned} q_{4u} &= \frac{q_{4w}}{q_{4w}^*} q_{4u}^* \frac{\alpha_3}{\alpha_1} \\ q_{4v} &= \frac{q_{4w}}{q_{4w}^*} q_{4v}^* \frac{\alpha_3}{\alpha_2} \\ \frac{q_{4w}}{q_{4w}^*} &= \frac{\alpha_4}{\alpha_3}. \end{aligned} \tag{60}$$

The above set of equations can be solved for $\dfrac{\alpha_4(t)}{\alpha_3(t)}$ and $\alpha_3(t)\tilde{G}(t) = diag\left(\dfrac{\alpha_3(t)}{\alpha_1(t)}, \dfrac{\alpha_3(t)}{\alpha_2(t)}, 1\right)$. Since the camera intrinsic calibration matrix is assumed to be known, we can obtain the scaled Euclidean homography which was defined in (7) as $\alpha_3(t)H(t) = \alpha_3(t)A^{-1}G(t)A$. As noted before, $H(t)$ can be decomposed into its constituent rotation matrix $\bar{R}(t)$, unit normal vector n^*, scaled translation vector $\bar{x}_h(t) \triangleq \dfrac{\bar{x}_f(t)}{d^*}$ and the depth ratio $\alpha_3(t)$. With the knowledge of $\alpha_3(t)$ and $\dfrac{\alpha_4(t)}{\alpha_3(t)}$, we can calculate the depth ratios α_1, α_2, α_3, and α_4 for all of the feature points.

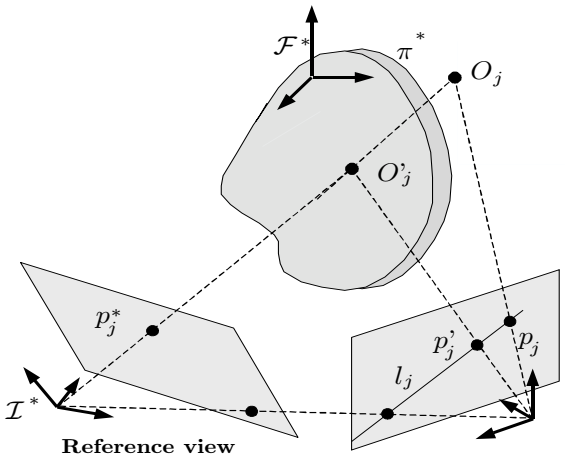

Fig. 9. Virtual parallax.

B. Virtual Parallax Method

In general, all feature points of interest on the moving object may not lie on a plane. In such a case, based on the development in [23], any three feature points on the object may be selected to define the plane π^* shown in Figure 9. All feature points O_i on a plane satisfy (11). Consider a feature point O_j on the object that is not on the plane π^*. Let us define a virtual feature point O'_j,on π^*, defined at the point of intersection of the vector from the optical center of the camera to O_j and the plane π^*. Let p_j^* be the projective image coordinates of the point O_j (and O'_j) on the image plane when the object is at the reference position denoted by $\mathcal{F}^*$. As shown in Figure 9, when the object is viewed from a different pose, resulting from either a motion of the object or a motion of the camera, the actual feature point O_j and the virtual feature point O'_j projects to $p_j(t)$ and $p'_j(t)$, respectively, on the image plane of the camera. For any feature point O_j, both $p_j(t)$ and $p'_j(t)$ lie on the same epipolar line l_j [23] that is given by

$$l_j = p_j \times p'_j \tag{61}$$

where $\times$ denotes the cross product of the two vectors. Since the projective image coordinates of corresponding coplanar feature points satisfy (11), we have

$$l_j = p_j \times G p_j^*. \tag{62}$$

Based on the constraint that all epipolar lines meet at the epipole [23], we can select a set of any three non-coplanar feature points such that the epipolar lines satisfy the following constraint

$$\left| \begin{array}{ccc} l_j & l_k & l_l \end{array} \right| = 0 \tag{63}$$

$$\textit{i.e.,} \left| \begin{array}{ccc} p_j \times G p_j^* & p_k \times G p_k^* & p_l \times G p_l^* \end{array} \right| = 0. \tag{64}$$

The transformation matrices, denoted by $P(t), P^* \in \mathbb{R}^{3\times 3}$, and defined in (53), are constructed using the image coordinates of the three coplanar feature points selected to define the plane π^*. After coordinate transformations defined in (57) and (58), the epipolar constraint of (64) now becomes

$$\left| \begin{array}{ccc} q_i \times \tilde{G} q_i^* & q_j \times \tilde{G} q_j^* & q_k \times \tilde{G} q_k^* \end{array} \right| = 0 \tag{65}$$

where $\tilde{G}(t) \in \mathbb{R}^{3\times 3}$ is defined in (55). As shown in [23], the set of homogeneous equations in (65) can be written in the form

$$C_{jkl} \bar{X} = 0 \tag{66}$$

where $\bar{X} = \left[\tilde{g}_1^2\tilde{g}_2,\ \tilde{g}_1\tilde{g}_2^2,\ \tilde{g}_1^2\tilde{g}_3,\ \tilde{g}_2^2\tilde{g}_3,\ \tilde{g}_1\tilde{g}_3^2,\ \tilde{g}_2\tilde{g}_3^2,\ \tilde{g}_1\tilde{g}_2\tilde{g}_3\right]^T \in \mathbb{R}^7$, and the matrix $C_{jkl} \in \mathbb{R}^{m\times 7}$ is of dimension $m \times 7$ where $m = \frac{n!}{6(n-3)!}$ and n is the number of epipolar lines (*i.e.*, one for image coordinates of each feature point). Hence, apart from three coplanar feature points that define the transformation matrices in (53), we will require atleast five additional feature points (*i.e.*, $n = 5$) in order to solve the set of equations given in (66). As shown in [23], we can then calculate $\tilde{G}(t)$, and hence, subsequently calculate α_1, α_2, α_3, $\bar{R}$ and n^* as previously explained.

C. Calculation of Depth Ratios for Non-coplanar Feature Points:

From (4) and (8), it can be easily shown that

$$m_j = \frac{z_j^*}{z_j}\left(\frac{\bar{x}_f}{z_j^*} + \bar{R} m_j^*\right). \tag{67}$$

After multiplying both sides of the equation with the skew-symmetric form of $\bar{x}_h(t)$, denoted by $[\bar{x}_h(t)]_\times \in \mathbb{R}^{3\times 3}$, we have [23]

$$\begin{aligned}[\bar{x}_h]_\times m_j &= \alpha_j \left([\bar{x}_h]_\times \frac{\bar{x}_f}{z_j^*} + [\bar{x}_h]_\times \bar{R} m_j^* \right) \\ &= \alpha_j [\bar{x}_h]_\times \bar{R} m_j^*. \end{aligned} \tag{68}$$

The signal $\bar{x}_h(t)$ is directly obtained from the decomposition of Euclidean homography matrix $H(t)$. Hence, the depth ratios for feature points O_j not lying on the plane π^* can be computed as follows

$$\alpha_j = \frac{\left\| [\bar{x}_h]_\times m_j \right\|}{\left\| [\bar{x}_h]_\times \bar{R} m_j^* \right\|}. \tag{69}$$

APPENDIX II
EXTENSION TO CAMERA-IN-HAND

A significant extension to the fixed camera system is the case where the camera can move relative to the object. For example, as shown in Figure 10, a camera could be mounted on the end-effector of a robot and used to scan an object in its workspace to determine its structure, as well as determine the robot's position. Let three feature points on the object, denoted by O_1, O_2 and O_3 define the plane π^* in Figure 10. Based on the development in [4], the signal $e(t) \in \mathbb{R}^6$ defined previously in (12) and (15) now quantifies the motion of the camera relative to its reference position. The time derivative of $e(t)$ can be expressed as follows

$$\dot{e} = J_c v \tag{70}$$

where $J_c(t) \in \mathbb{R}^{6\times 6}$ is given by

$$J_c = \begin{bmatrix} -\dfrac{\alpha_1}{z_1^*} A_{e1} & A_{e1} [m_1]_\times \\ 0_3 & -L_\omega \end{bmatrix} \tag{71}$$

where $0_3 \in \mathbb{R}^{3\times 3}$ is a zero matrix, $L_\omega(t) \in \mathbb{R}^{3\times 3}$ has exactly the same form as for the fixed camera case in (17), and $v(t) \triangleq \begin{bmatrix} v_c^T & \omega_c^T \end{bmatrix}^T \in \mathbb{R}^6$ now denotes the velocity of the camera expressed relative to $\mathcal{I}$.

With the exception of the term $z_1^* \in \mathbb{R}$, all other terms in (71) are either measurable or known a priori. If the camera can be moved away from its reference position by a known translation vector $\bar{x}_{fk} \in \mathbb{R}^3$, then z_1^* can be computed offline. Decomposition of the Euclidean homography between the normalized Euclidean coordinates of the feature points obtained at the reference position, and at $\bar{x}_{fk}$ away from the reference position, respectively, can be used to calculate the scaled translation vector $\frac{\bar{x}_{fk}}{d^*} \in \mathbb{R}^3$, where $d^* \in \mathbb{R}$ is the distance from the initial camera position, denoted by $\mathcal{I}^*$, to the plane π^*. Then, it can be seen that [2]

$$z_1^* = \frac{d^*}{n^{*T} m_1^*} = \frac{d^*}{n^{*T} A^{-1} p_1^*}. \tag{72}$$

From (70) and (71), we can show that for any feature point O_i

$$\begin{aligned}\dot{p}_{ei} &= -\frac{\alpha_i}{z_i^*} A_{ei} v_c + A_{ei} [m_i]_\times \omega_c \\ &= W_{1i} v_c \theta_i + W_{2i} \omega_c \end{aligned} \tag{73}$$

where $p_{ei}(t) \in \mathbb{R}^3$ was defined previously in (23), and $W_{1i}(.) \in \mathbb{R}^{3\times 3}$, $W_{2i}(t) \in \mathbb{R}^{3\times 3}$ and $\theta_i \in \mathbb{R}$ are given as follows

$$W_{1i} = -\alpha_i A_{ei} \tag{74}$$

$$W_{2i} = A_{ei} [m_i]_\times \tag{75}$$

$$\theta_i = \frac{1}{z_i^*}. \tag{76}$$

The matrices $W_{1i}(.)$ and $W_{2i}(t)$ are measurable and bounded. The objective is to identify the unknown constant θ_i in (76). With the knowledge of z_i^* for all feature points on the object, their Euclidean coordinates $\bar{m}_i^*$ relative to the camera reference position, denoted by $\mathcal{I}^*$, can be computed from (8) through (10). As before, we define the mismatch between the actual signal θ_i and the estimated signal $\hat{\theta}_i(t)$ as $\tilde{\theta}(t) \in \mathbb{R}$ where $\tilde{\theta}(t) \triangleq \theta - \hat{\theta}(t)$.

[2]Note that for any feature point O_i coplanar with π^*, z_i^* could be computed in this fashion.

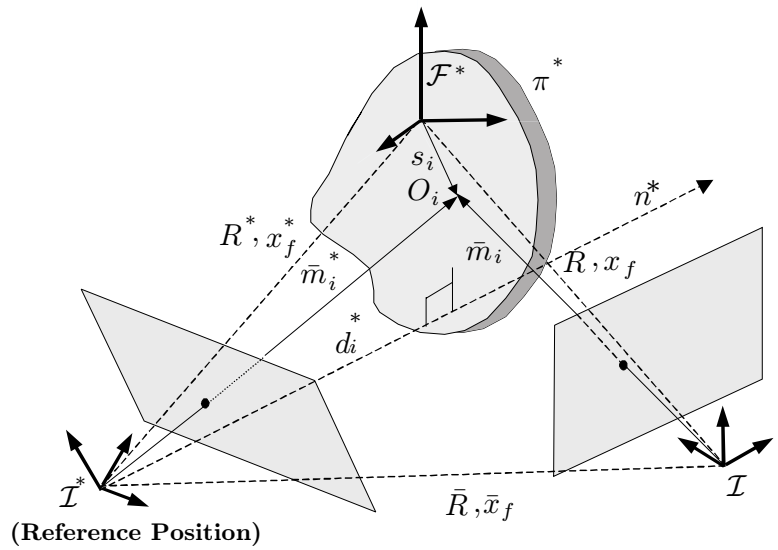

Fig. 10. Coordinate frame relationships for the camera-in-hand case.

To facilitate our objective, we introduce a measurable filter signal $\zeta_i(t) \in \mathbb{R}^3$ and two non-measurable filter signals $\kappa_i(t), \eta_i(t) \in \mathbb{R}^3$ defined as follows

$$\dot{\zeta}_i = -\beta_i \zeta_i + W_{1i} \hat{v}_c \tag{77}$$

$$\dot{\kappa}_i = -\beta_i \kappa_i + W_{1i} \tilde{v}_c \theta_i \tag{78}$$

$$\dot{\eta}_i = -\beta_i \eta_i + W_{2i} \tilde{\omega}_c \tag{79}$$

where $\beta_i \in \mathbb{R}$ is a scalar positive gain, $\hat{v}_c(t), \hat{\omega}_c(t) \in \mathbb{R}^3$ are the estimates for the translational and rotational velocity, respectively, obtained from the velocity observer in Section IV, $\tilde{v}_c(t) \triangleq v_c(t) - \hat{v}_c(t)$ is the mismatch in estimated translational velocity, and $\tilde{\omega}_c(t) \triangleq \omega_c(t) - \hat{\omega}_c(t)$ is the mismatch in estimated rotational velocity. Note that the structure of the velocity observer and the proof of its convergence is exactly identical to the fixed camera case. Likewise, we design the following estimator for the $p_{ei}(t)$

$$\dot{\hat{p}}_{ei} = \beta_i \tilde{p}_{ei} + \zeta_i \dot{\hat{\theta}}_i + W_{1i} \hat{v}_c \hat{\theta}_i + W_{2i} \hat{\omega}_c. \tag{80}$$

From (73) and (80), we have

$$\dot{\tilde{p}}_{ei} = -\beta_i \tilde{p}_{ei} - \zeta_i \dot{\hat{\theta}}_i + W_{1i} \tilde{v}_c \theta_i + W_{1i} \hat{v}_c \tilde{\theta}_i + W_{2i} \tilde{\omega}_c. \tag{81}$$

From (81), (78) and (79), it can be shown that

$$\tilde{p}_{ei} = \zeta_i \tilde{\theta}_i + \kappa_i + \eta_i. \tag{82}$$

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

$$\dot{\hat{\theta}}_i = L_i \zeta_i^T \tilde{p}_{ei} \tag{83}$$

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

$$\frac{d}{dt}(L_i^{-1}) = \zeta_i^T \zeta_i \tag{84}$$

and initialized such that $L_i^{-1}(0) > 0$.

Theorem 2: The update law defined in (83) ensures that $\tilde{\theta}_i \to 0$ as $t \to \infty$ provided that the following persistent excitation condition [27] holds

$$\gamma_5 \leq \int_{t_0}^{t_0+T} \zeta_i^T(\tau) \zeta_i(\tau) d\tau \leq \gamma_6 \tag{85}$$

and provided that the gains β_i satisfy the following inequalities

$$\beta_i > k_{3i} + k_{4i} \left\| W_{1i} \right\|_\infty^2 \tag{86}$$

$$\beta_i > k_{5i} + k_{6i} \left\| W_{2i} \right\|_\infty^2 \tag{87}$$

where $t_0, \gamma_5, \gamma_6, T, k_{3i}, k_{4i}, k_{5i}, k_{6i} \in \mathbb{R}$ are positive constants, the notation $\|.\|_\infty$ denotes the induced ∞-norm of a matrix [19], and k_{3i}, k_{5i} are selected such that

$$\frac{1}{k_{3i}} + \frac{1}{k_{5i}} < \frac{1}{2}. \tag{88}$$

Proof: Similar to the analysis for the fixed camera case, a non-negative function denoted by $V(t) \in \mathbb{R}$ is defined as follows

$$V \triangleq \frac{1}{2}\tilde{\theta}_i^T L_i^{-1} \tilde{\theta}_i + \frac{1}{2}\kappa_i^T \kappa_i + \frac{1}{2}\eta_i^T \eta_i. \tag{89}$$

After taking the time derivative of (89), the following expression can be obtained

$$\begin{aligned} \dot{V} &= -\frac{1}{2}\left\|\tilde{\theta}_i\right\|^2 \|\zeta_i\|^2 - \beta_i \|\kappa_i\|^2 - \beta_i \|\eta_i\|^2 \\ &\quad -\tilde{\theta}_i^T \zeta_i^T \kappa_i - \tilde{\theta}_i^T \zeta_i^T \eta_i + \kappa_i^T W_{1i} \tilde{v}_c \theta_i \\ &\quad +\eta_i^T W_{2i} \tilde{\omega}_c \end{aligned} \tag{90}$$

where (83), (84), (78), (79) and (82) were utilized. Upon further mathematical manipulation of (90), we have,

$$\begin{aligned} \dot{V} &\leq -\frac{1}{2}\left\|\tilde{\theta}_i\right\|^2 \|\zeta_i\|^2 \\ &\quad -\left(\beta_i - k_{3i} - k_{4i}\|W_{1i}\|_\infty^2\right)\|\kappa_i\|^2 \\ &\quad -\left(\beta_i - k_{5i} - k_{6i}\|W_{2i}\|_\infty^2\right)\|\eta_i\|^2 \\ &\quad +\left\|\tilde{\theta}_i\right\|\|\zeta_i\|\|\kappa_i\| - k_{3i}\|\kappa_i\|^2 \\ &\quad +\|\theta_i\|\|\tilde{v}_c\|\|W_{1i}\|_\infty\|\kappa_i\| - k_{4i}\|W_{1i}\|_\infty^2\|\kappa_i\|^2 \\ &\quad +\|\zeta_i\|\left\|\tilde{\theta}_i\right\|\|\eta_i\| - k_{5i}\|\eta_i\|^2 \\ &\quad +\|\tilde{\omega}_c\|\|W_{2i}\|_\infty\|\eta_i\| - k_{6i}\|W_{2i}\|_\infty^2\|\eta_i\|^2 \\ &\leq -\left(\frac{1}{2} - \frac{1}{k_{3i}} - \frac{1}{k_{5i}}\right)\left\|\tilde{\theta}_i\right\|^2\|\zeta_i\|^2 \\ &\quad -\left(\beta_i - k_{3i} - k_{4i}\|W_{1i}\|_\infty^2\right)\|\kappa_i\|^2 \\ &\quad -\left(\beta_i - k_{5i} - k_{6i}\|W_{2i}\|_\infty^2\right)\|\eta_i\|^2 \\ &\quad +\frac{1}{k_{4i}}\|\theta_i\|^2\|\tilde{v}_c\|^2 + \frac{1}{k_{6i}}\|\tilde{\omega}_c\|^2 \end{aligned} \tag{91}$$

where $k_{3i}, k_{4i}, k_{5i}, k_{6i} \in \mathbb{R}$ are positive constants as previously mentioned. The gain constants are selected to ensure that

$$\frac{1}{2} - \frac{1}{k_{3i}} - \frac{1}{k_{5i}} \geq \mu_{3i} > 0 \tag{92}$$

$$\beta_i - k_{3i} - k_{4i}\|W_{1i}\|_\infty^2 \geq \mu_{4i} > 0 \tag{93}$$

$$\beta_i - k_{5i} - k_{6i}\|W_{2i}\|_\infty^2 \geq \mu_{5i} > 0 \tag{94}$$

where $\mu_{3i}, \mu_{4i}, \mu_{5i} \in \mathbb{R}$ are positive constants. The gain conditions given by (92), (93) and (94) allow us to further upper bound the time derivative of (89) as follows

$$\begin{aligned} \dot{V} &\leq -\mu_{3i}\left|\tilde{\theta}_i\right|^2\|\zeta_i\|^2 - \mu_{4i}\|\kappa_i\|^2 - \mu_{5i}\|\eta_i\|^2 \\ &\quad +\frac{1}{k_{4i}}|\theta_i|^2\|\tilde{v}_c\|^2 + \frac{1}{k_{6i}}\|\tilde{\omega}_c\|^2 \\ &\leq -\mu_{3i}\left|\tilde{\theta}_i\right|^2\|\zeta_i\|^2 - \mu_{4i}\|\kappa_i\|^2 - \mu_{5i}\|\eta_i\|^2 \\ &\quad +\mu_{6i}\|\tilde{v}\|^2 \end{aligned} \tag{95}$$

where $\mu_{6i} = \max\left\{\frac{|\theta_i|^2}{k_{4i}}, \frac{1}{k_{6i}}\right\} \in \mathbb{R}$. Following the argument in fixed camera case, $\tilde{v}(t) \in L_2$; hence, we have

$$\int_{t_0}^{t} \mu_{6i}\|\tilde{v}(\tau)\|^2 d\tau \leq \varepsilon \tag{96}$$

where $\varepsilon \in \mathbb{R}$ is a positive constant. From (89), (95) and (96), we can conclude that

$$\begin{aligned} & \int_{t_0}^{t} \left(\mu_{3i} \left| \tilde{\theta}_i(\tau) \right|^2 \|\zeta_i(\tau)\|^2 \right. \\ & \left. + \mu_{4i} \|\kappa_i(\tau)\|^2 + \mu_{5i} \|\eta_i(\tau)\|^2 \right) d\tau \\ \leq \quad & V(0) - V(\infty) + \varepsilon. \end{aligned} \tag{97}$$

From (97), it is clear that $\zeta_i(t)\tilde{\theta}_i(t), \kappa_i(t), \eta_i(t) \in L_2$. Applying the same signal chasing arguments as in the fixed camera case, it can be shown that $\tilde{\theta}_i(t), \kappa_i(t), \eta_i(t) \in L_\infty$. It can also be shown that $\dot{\tilde{\theta}}_i(t), \zeta_i(t), \dot{\zeta}_i(t) \in L_\infty$ and therefore $\frac{d}{dt}\zeta_i(t)\tilde{\theta}_i(t) \in L_\infty$. Hence $\zeta_i(t)\tilde{\theta}_i(t)$ is uniformly continuous [12], and since $\zeta_i(t)\tilde{\theta}_i(t) \in L_\infty$, we have [12]

$$\zeta_i(t)\tilde{\theta}_i(t) \to 0 \text{ as } t \to \infty. \tag{98}$$

Applying the same argument as in the fixed camera case, convergence of $\hat{\theta}_i(t)$ to true parameters is guaranteed (*i.e.*, $\tilde{\theta}_i(t) \to 0$ as $t \to \infty$), if the signal $\zeta_i(t)$ satisfies the persistent excitation condition in (85). □

Remark 8: As in the case of fixed camera (see remark 6), it can be shown that the persistent excitation condition of (85) is satisfied if the translational velocity of the camera is non-zero.

Remark 9: Utilizing (8), (10) and the update law in (83), the estimates for Euclidean coordinates of all i feature points on the object relative to the camera at reference position, denoted by $\hat{\bar{m}}_i^*(t) \in \mathbb{R}^3$, can be determined as follows

$$\hat{\bar{m}}_i^*(t) = \frac{1}{\hat{\theta}_i(t)} A^{-1} p_i^*. \tag{99}$$

Remark 10: The knowledge of z_1^* allows us to resolve the scale factor ambiguity inherent to the Euclidean reconstruction algorithm. However, z_1^* may not always be available, or may not be measurable using the technique described previously in (72) due to practical considerations (for e.g., we may only have a video sequence available as input). With a minor modification to the estimator design, the scale ambiguity can be resolved, and Euclidean coordinates of feature points recovered, if the Euclidean distance between two of the many feature points in the scene is available. Assuming z_1^* is not directly measurable, terms in the velocity kinematics of (70) can be re-arranged in the following manner

$$\dot{e} = \bar{J}_c \bar{v} \tag{100}$$

where the Jacobian $\bar{J}_c(t) \in \mathbb{R}^{6\times 6}$ and a scaled velocity vector $\bar{v}(t) \in \mathbb{R}^6$ are defined as follows

$$\bar{J}_c = \begin{bmatrix} -\alpha_1 A_{e1} & A_{e1} [m_1]_\times \\ 0_3 & -L_\omega \end{bmatrix} \tag{101}$$

$$\bar{v} = \begin{bmatrix} \bar{v}_c^T & \omega_c^T \end{bmatrix}^T, \text{where } \bar{v}_c = \frac{v_c^T}{z_1^*}. \tag{102}$$

The velocity observer of Section IV can now be utilized to identify the scaled velocity $\bar{v}(t)$. Likewise, the time derivative of the extended image coordinates presented in (73) can be re-written in terms of the scaled velocity as follows

$$\dot{p}_{ei} = W_{1i} \bar{v}_c \bar{\theta}_i + W_{2i} \omega_c \tag{103}$$

$$\text{where } \bar{\theta}_i = \frac{z_1^*}{z_i^*}. \tag{104}$$

The rest of the development is identical. Hence, it is clear from (104) that the adaptive estimator of (83) now identifies the depth of all feature points scaled by a contant scalar z_1^*. If the Euclidean distance between any two features i and j ($i \neq j$) is known, then from (83), (99) and (104) it is clear that

$$\lim_{t\to\infty} \left\| \hat{\bar{m}}_i^*(t) - \hat{\bar{m}}_j^*(t) \right\| = \frac{1}{z_1^*} \left\| \bar{m}_i^* - \bar{m}_j^* \right\|. \tag{105}$$

Hence, the unknown scalar z_1^* can be recovered and the Euclidean coordinates of all feature points obtained.

Remark 11: Terms from the rotation matrix $R(t)$ are not present in (71), and therefore, unlike the fixed camera case, the estimate of the velocity of the camera, denoted by $\hat{v}(t)$, can be computed without knowledge of the constant rotation matrix R^*.

Remark 12: As mentioned in Appendix I, decomposition of the Euclidean homography gives the rotation matrix $\bar{R}$, the normal vector n^* and the scaled translation vector $\frac{\bar{x}_f(t)}{d^*}$. Since d^* can now be computed from any feature point on the plane π^* using (6), we can now compute $\bar{x}_f(t)$. Hence, the six degree-of-freedom position of the moving camera relative to its reference position can be computed online.

APPENDIX III

For the sake of clarity in the subsequent analysis, let $\Omega_i(t_0,t) \in \mathbb{R}^{4\times 4}$ be defined as follows

$$\Omega_i(t_0,t) = \int_{t_0}^{t} W_{fi}^T(\tau) W_{fi}(\tau) d\tau \tag{106}$$

where $W_{fi}(t) \in \mathbb{R}^{3\times 4}$ was previously defined in (29). Consider the following expression

$$\begin{aligned}
&\int_{t_0}^{t} \tilde{\theta}_i^T(\tau)\Omega_i(t_0,\tau)\frac{d\tilde{\theta}_i(\tau)}{d\tau}d\tau \\
= \;& \tilde{\theta}_i^T(\tau)\Omega_i(t_0,\tau)\theta_i(\tau)\Big|_{t_0}^{t} \\
&- \int_{t_0}^{t} \frac{d}{d\tau}\left(\tilde{\theta}_i^T(\tau)\Omega_i(t_0,\tau)\right)\tilde{\theta}_i(\tau)d\tau \\
= \;& \tilde{\theta}_i^T(t)\Omega_i(t_0,t)\tilde{\theta}_i(t) - \int_{t_0}^{t} \tilde{\theta}_i^T(\tau)\Omega_i(t_0,\tau)\frac{d\tilde{\theta}_i(\tau)}{d\tau}d\tau \\
&- \int_{t_0}^{t} \tilde{\theta}_i^T(\tau) W_{fi}^T(\tau) W_{fi}(\tau)\tilde{\theta}_i(\tau)d\tau
\end{aligned} \tag{107}$$

where we utilized (106) and the fact that $\Omega(t_0,t_0) = 0$. After re-arranging (107), we have the following expression

$$\begin{aligned}
&\tilde{\theta}_i^T(t)\Omega_i(t_0,t)\tilde{\theta}_i(t) \\
= \;& 2\int_{t_0}^{t} \tilde{\theta}_i^T(\tau)\Omega_i(t_0,\tau)\frac{d\tilde{\theta}_i(\tau)}{d\tau}d\tau \\
&+ \int_{t_0}^{t} \tilde{\theta}_i^T(\tau) W_{fi}^T(\tau) W_{fi}(\tau)\tilde{\theta}_i(\tau)d\tau.
\end{aligned} \tag{108}$$

To facilitate further analysis, we now state the following lemma [19]

Lemma 1: If $f(t)$ is a uniformly continuous function [12], then $\lim_{t\to\infty} f(t) = 0$ if and only if

$$\lim_{t\to\infty} \int_{t}^{t+t'} f(\tau)d\tau = 0 \tag{109}$$

for any positive constant $t' \in \mathbb{R}$.

After substituting for $t = t_0 + T$ in (108), where $T \in \mathbb{R}$ is a positive constant, and applying the limit on both sides of the equation, we have

$$\begin{aligned}
&\lim_{t_0\to\infty} \tilde{\theta}_i^T(t_0+T)\Omega_i(t_0,t_0+T)\tilde{\theta}_i(t_0+T) \\
= \;& \lim_{t_0\to\infty} \left(2\int_{t_0}^{t_0+T} \tilde{\theta}_i^T(\tau)\Omega_i(t_0,\tau)\frac{d\tilde{\theta}_i(\tau)}{d\tau}d\tau \right. \\
&\left. + \int_{t_0}^{t_0+T} \tilde{\theta}_i^T(\tau) W_{fi}^T(\tau) W_{fi}(\tau)\tilde{\theta}_i(\tau)d\tau \right).
\end{aligned} \tag{110}$$

We now examine the terms in the first integral of (110). From the proof of Theorem 1, $\tilde{\theta}_i(t) \in L_\infty$, and from (36) and (106), $\Omega_i(t_0,t_0+T) \in L_\infty$. It was also proved that $W_{fi}(t)\tilde{\theta}_i(t), \eta_i(t) \in L_\infty \cap L_2$. Hence, from (33), $\lim_{t\to\infty}\tilde{p}_{ei}(t) = 0$ and consequently from (28) and (34), $\lim_{t\to\infty} \dot{\tilde{\theta}}_i\,(t) = 0$. Hence, after utilizing Lemma 1, the first integral in (110) vanishes upon evaluation. From (47) and Lemma 1, the second integral in (110) vanishes as well. Therefore, we have

$$\lim_{t_0\to\infty} \tilde{\theta}_i^T(t_0+T)\Omega_i(t_0,t_0+T)\tilde{\theta}_i(t_0+T) = 0. \tag{111}$$

Since $\Omega_i(t_0,t) \geq \gamma_1 I_4$ from (36) for any t_0, we can conclude from (111) that

$$\tilde{\theta}_i(t) \to 0 \text{ as } t \to \infty. \tag{112}$$

ACKNOWLEDGMENT

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.

REFERENCES

[1] S. Birchfield, KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker, http://www.ces.clemson.edu/~stb/klt.

[2] 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.

[3] 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.

[4] J. Chen, W. E. Dixon, D. M. Dawson, and V. Chitrakaran, "Navigation Function Based Visual Servo Control," *Proc. of the 2005 IEEE American Control Conference*, pp. 3682-3687, Portland, USA, 2005.

[5] 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.

[6] V. Chitrakaran, D. M. Dawson, J. Chen, and H. Kannan, "Velocity and Structure Estimation of a Moving Object Using a Moving Monocular Camera," *2006 American Control Conference*, accepted, to appear.

[7] 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 American Control Conference*, pp. 4601-4606, June 2005.

[8] 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.

[9] 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.

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

[11] 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.

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

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

[14] 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.

[15] FFmpeg Project, http://ffmpeg.sourceforge.net/index.php.

[16] R. 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.

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

[18] 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.

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

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

[21] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *International Joint Conference on Artificial Intelligence*, pp. 674-679, 1981.

[22] 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, 1999.

[23] 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.

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

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

[26] J. Shi, and C. Tomasi, "Good Features to Track," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600, 1994.

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

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

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