Homography Based Visual Servo Control with Scene Reconstruction

Anup Parikh[1], Rushikesh Kamalapurkar[1], Hsi-Yuan Chen[1], Warren E. Dixon[1]

***Abstract*—Homography based visual servoing is an approach that blends image based feedback with feedback that is reconstructed from the image to control an autonomous system to move along a desired trajectory. Adaptive control methods have been previously developed by compensating for an unknown parameter (i.e., the depth of a feature) in the dynamics, where persistence of excitation assumptions are used for parameter identification. Rather than assume *persistent* excitation, an augmented adaptive update law that uses recorded data is utilized in this paper to guarantee exponential tracking and parameter identification with only *finite* excitation. By identifying the depth parameter, the structure of the scene can be reconstructed, enabling simultaneous mapping and control.**

I. Introduction

The objective in visual servo control (also known as visual servoing) is to control the motion of an autonomous system to track a desired trajectory which is either known or generated from camera images. Two primary methods of visual servoing are image based (IBVS) and position based (PBVS). In IBVS, the tracking error is quantified based on pixel coordinates of features in the desired and actual camera images. Although the features appear to efficiently converge to their desired locations, the 3D trajectory of the camera generated by IBVS can be inefficient or not physically possible [1]–[4].

To address these shortcomings of IBVS while still only using a single camera, PBVS utilizes known scene geometry to determine the 6 degree of freedom (DOF) position and orientation (pose) of the camera, and the tracking error is quantified by the pose error. However, a drawback to PBVS is that the fully scaled geometry of the scene must be known. Due to the nature of image formation, scaled geometric information is not instantaneously measurable from the camera, and therefore additional calibration or preprocessing may be required before this method can be applied.

An effective middle ground between these two methods is known as homography-based (e.g., [5]–[10]) or 2.5D (e.g., [11]–[14]) visual servoing. In planar environments, the orientation and scaled translation of camera can be determined, up to one unknown constant (i.e., the depth of the homography plane with respect to the initial camera pose), without requiring knowledge of the scene geometry. Robust and adaptive control methods can then be used to compensate for the unknown depth parameter and achieve the control objective (e.g. [5]–[7], [9]). Since the error is quantified based on both the camera pose and a feature position, the resulting 6DOF camera trajectory is more efficient than those produced with IBVS. See [1]–[4] for a more in-depth review of visual servo control.

A known deficiency in adaptive control is the lack of a guarantee of parameter convergence using instantaneous data without persistence of excitation (PE) [15]–[17]. Specifically considering the visual servoing problem, the unknown depth estimate cannot be shown to converge to its true value (e.g., [5]–[7], [9]) without assuming PE (e.g., [18]–[21] or), a condition that cannot be checked *a priori* for nonlinear systems, is difficult to check online, and may require restricting camera motion to satisfy (e.g., [22]–[24]). A recently developed technique known as concurrent learning (e.g., [25], [26]) has been shown to guarantee exponential convergence by augmenting the adaptive update law to utilize recorded data. This method is similar to the familiar concept in linear least squares (LLS) where input-output data is utilized to estimate the parameters of a function [27]. Similar to LLS, the data must be sufficiently rich for the concurrent learning technique to obtain a unique solution. In concurrent learning, the richness condition is satisfied if the system is excited for a *finite* duration of time, and therefore is much easier to satisfy than PE. In addition, the richness condition can easily be verified online by checking the minimum singular value of an $m \times m$ matrix, where m is the number of unknown parameters.

For the problem addressed in this paper, an adaptive update law inspired by the concurrent learning concept is used to identify the depth of the reference plane that induces the homography relationship. A unique obstacle that emerges when trying to apply the concurrent learning strategy to visual servo control is that the uncertainties appear on both sides of the dynamics (i.e., the uncertainties are multiplied by both the state and state derivatives). The dynamics are formulated in this way to avoid numerical issues associated with estimating the inverse of the uncertainties (i.e., the inverse range as opposed to the range), as well as to mitigate the need for additional robust terms to compensate for the uncertainties being multiplied by the control input. In developing a novel adaptive update law to guarantee parameter convergence, the outcome of this work is that

[1]Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville FL 32611-6250, USA Email:{anuppari, rkamalapurkar, hychen, wdixon}@ufl.edu

This research is supported in part by a Task Order contract with the Air Force Research Laboratory, Munitions Directorate at Eglin AFB, US National Geospatial Intelligence Agency Grant HM0177-12-1-0006 and US Department of Agriculture National Robotics Initiative Grant 2013-67021-21074. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsoring agency.

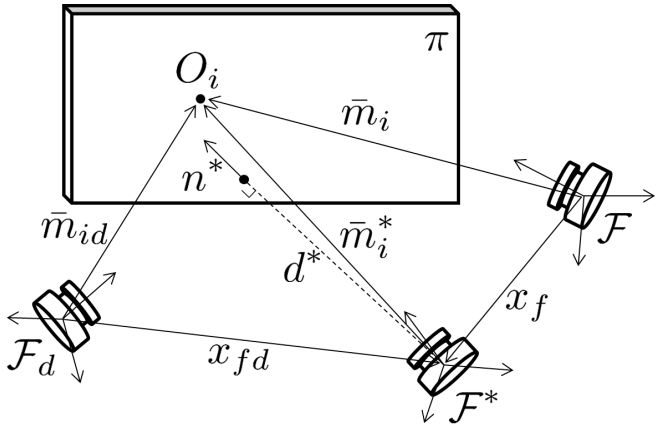

Figure 1. Geometric relationships considered in this paper.

not only is the tracking performance improved (exponential tracking versus asymptotic for traditional gradient based adaptive control without PE), but the structure of the plane (i.e., the 3D coordinates of all features on the plane) is also determined. By reconstructing the scene online, we integrate simultaneous localization and mapping (SLAM) with control.

II. Geometric Model and Euclidean Reconstruction

The camera-in-hand problem is considered in the following development. In this configuration, the camera is attached to a mobile autonomous system while observing a fixed planar object, and the camera velocities themselves are controlled. This work can be extended to the camera-to-hand configuration [7] where the camera is stationary and observes the mobile autonomous system to be controlled. Without loss of generality[1], assume a fixed object in the workspace has at least 4 coplanar but not collinear feature points, and let π denote the plane they lie on. Let $\mathcal{F}$ denote a reference frame attached to the moving camera, with a coordinate system defined as having its origin at the principle point of the camera, Z-axis coincident with the optical axis of the camera and positive outward, and X-axis parallel to the horizontal axis of the image sensor. Let $\mathcal{F}^*$ denote an inertial reference frame, with coordinate system typically taken as coincident with that of $\mathcal{F}$ at the initial time. Therefore, π is fixed with respect to $\mathcal{F}^*$. Let $\mathcal{F}_d$ denote a reference frame attached to the desired pose of the camera, with coordinate system defined in the same manner as that of $\mathcal{F}$. The vectors $\bar{m}_i$, $\bar{m}_i^*$, $\bar{m}_{id} \in \mathbb{R}^3$ denote the position of a point O_i on π with respect to the origin of each of the coordinate systems, as shown in Figure 1, and expressed as

$$\bar{m}_i = \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T$$
$$\bar{m}_i^* = \begin{bmatrix} x_i^* & y_i^* & z_i^* \end{bmatrix}^T$$
$$\bar{m}_{id} = \begin{bmatrix} x_{id} & y_{id} & z_{id} \end{bmatrix}^T$$

[1]The virtual parallax method can be used in cases where there are at least 8 feature points in general position, with no 4 points being coplanar.

where x_i, y_i, $z_i \in \mathbb{R}$, x_i^*, y_i^*, $z_i^* \in \mathbb{R}$ and x_{id}, y_{id}, $z_{id} \in \mathbb{R}$ are the Euclidean coordinates of the feature O_i expressed in the coordinate system of $\mathcal{F}$, $\mathcal{F}^*$ and $\mathcal{F}_d$, respectively. Then the vectors $\bar{m}_i$ and $\bar{m}_{id}$ can be expressed in terms of the constant $\bar{m}_i^*$ as

$$\bar{m}_i = x_f + R\bar{m}_i^* \tag{1}$$

$$\bar{m}_{id} = x_{fd} + R_d\bar{m}_i^* \tag{2}$$

where $x_f \in \mathbb{R}^3$ denotes the vector from the origin of $\mathcal{F}$ to the origin of $\mathcal{F}^*$, expressed in the coordinate system of $\mathcal{F}$, $x_{fd} \in \mathbb{R}^3$ denotes the vector from the origin of $\mathcal{F}_d$ to the origin of $\mathcal{F}^*$, expressed in the coordinate system of $\mathcal{F}_d$ and $R \in SO(3)$ and $R_d \in SO(3)$ denote the orientation of the coordinate system of $\mathcal{F}^*$ relative to $\mathcal{F}$ and $\mathcal{F}_d$, respectively. The constant distance from the origin of $\mathcal{F}^*$ to the plane π, denoted $d^* \in \mathbb{R}$, is given by

$$d^* = n^{*T}\bar{m}_i^* \tag{3}$$

where $n^* \in \mathbb{R}^3$ is the unit vector perpendicular to π and passing through the origin of $\mathcal{F}^*$, expressed in the coordinate system of $\mathcal{F}^*$. Using (3), (1) and (2) can be rewritten as

$$\bar{m}_i = \left(R + \frac{x_f}{d^*}n^{*T}\right)\bar{m}_i^* \tag{4}$$

$$\bar{m}_{id} = \left(R_d + \frac{x_{fd}}{d^*}n^{*T}\right)\bar{m}_i^*. \tag{5}$$

The normalized Euclidean coordinates, denoted m_i, m_i^*, $m_{id} \in \mathbb{R}^3$, are defined as

$$m_i \triangleq \frac{\bar{m}_i}{z_i} = \begin{bmatrix} \frac{x_i}{z_i} & \frac{y_i}{z_i} & 1 \end{bmatrix}^T$$
$$m_i^* \triangleq \frac{\bar{m}_i^*}{z_i^*} = \begin{bmatrix} \frac{x_i^*}{z_i^*} & \frac{y_i^*}{z_i^*} & 1 \end{bmatrix}^T$$
$$m_{id} \triangleq \frac{\bar{m}_{id}}{z_{id}} = \begin{bmatrix} \frac{x_{id}}{z_{id}} & \frac{y_{id}}{z_{id}} & 1 \end{bmatrix}^T$$

which can be used to rewrite (4) and (5) as

$$m_i = \underbrace{\frac{z_i^*}{z_i}}_{\alpha_i}\underbrace{\left(R + \frac{x_f}{d^*}n^{*T}\right)}_{H}m_i^* \tag{6}$$

$$m_{id} = \underbrace{\frac{z_i^*}{z_{id}}}_{\alpha_{id}}\underbrace{\left(R_d + \frac{x_{fd}}{d^*}n^{*T}\right)}_{H_d}m_i^* \tag{7}$$

where α_i, $\alpha_{id} \in \mathbb{R}$ are the time-varying depth ratios, and H, $H_d \in \mathbb{R}^{3\times3}$ denote the Euclidean homography matrices. The pixel coordinates of the projection of the feature onto the imaging sensor attached to $\mathcal{F}$, $\mathcal{F}^*$ and $\mathcal{F}_d$, respectively, are denoted p_i, p_i^*, $p_{id} \in \mathbb{R}^3$ and defined as

$$p_i \triangleq \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T$$
$$p_i^* \triangleq \begin{bmatrix} u_i^* & v_i^* & 1 \end{bmatrix}^T$$
$$p_{id} \triangleq \begin{bmatrix} u_{id} & v_{id} & 1 \end{bmatrix}^T$$

where u_i, v_i, u_i^*, v_i^*, u_{id}, $v_{id} \in \mathbb{R}$. The pinhole camera model relates the pixel coordinates to the normalized Euclidean

coordinates by the known, constant, invertible [28, Chapter 3.3] intrinsic camera calibration matrix $A \in \mathbb{R}^{3\times 3}$ as

$$p_i = Am_i, \quad p_i^* = Am_i^*, \quad p_{id} = Am_{id}.$$

Based on (6) and (7), the homography relationship between the pixel coordinates can be expressed as

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

$$p_{id} = \alpha_{id} \underbrace{\left(AH_dA^{-1}\right)}_{G_d} p_i^*$$

where G, $G_d \in \mathbb{R}^{3\times 3}$ are the projective homography matrices.

The projective homography matrix between the image at $\mathcal{F}^*$ and the images taken from $\mathcal{F}$ and $\mathcal{F}_d$ can be directly computed from the pixel coordinates of the four coplanar feature points [28, Chapter 5.3]. The methods described in [28]–[30] can then be used to decompose G and G_d into the depth ratios, α_i, α_{id}, the relative rotation matrices, R, R_d and the unit normal vector n^*. These signals can then be used to provide feedback in the subsequently designed controller.

III. Control Objective

The objective is to design a controller for the camera velocities such that the 6 DOF trajectory of the camera tracks the trajectory of the (real or virtual) camera attached to $\mathcal{F}_d$. In addition, since the camera only provides 2D information, Euclidean reconstruction of the features on π is desired. Since the full scale translations x_f and x_{fd} are not available from the homography decomposition, the tracking control objective is expressed as the desire to achieve: $R(t) \to R_d(t)$, $m_1(t) \to m_{1d}(t)$ and $z_1(t) \to z_{1d}(t)$ as $t \to \infty$, where feature 1 was selected for notational convenience and without loss of generality. The Euclidean reconstruction objective can be expressed as $\hat{z}_1^* \to z_1^*$ as $t \to \infty$, where $\hat{z}_1^* \in \mathbb{R}$ is the estimate of the depth of the point O_1 relative to $\mathcal{F}^*$. With $\hat{z}_1^*$, the methods described in [31, Section III] and [32, Section III] can be used to reconstruct the depth of the homography plane and subsequently the Euclidean coordinates of all features on π.

To quantify the translation tracking error, let $e_v \in \mathbb{R}^3$ be defined as

$$e_v = p_e - p_{ed} \tag{8}$$

where p_e, $p_{ed} \in \mathbb{R}^3$ are defined as

$$p_e = \begin{bmatrix} u_1 & v_1 & -\ln(\alpha_1) \end{bmatrix}^T$$

$$p_{ed} = \begin{bmatrix} u_{1d} & v_{1d} & -\ln(\alpha_{1d}) \end{bmatrix}^T.$$

The rotation tracking error, $e_\omega \in \mathbb{R}^3$, is defined as

$$e_\omega = \Theta - \Theta_d \tag{9}$$

where Θ, $\Theta_d \in \mathbb{R}^3$ denote the axis-angle representation[2] of R and R_d, respectively, and are defined as

$$\Theta \triangleq u\theta \quad \Theta_d \triangleq u_d\theta_d$$

where u, $u_d \in \mathbb{R}^3$ are unit vectors representing the rotation axes, and θ, $\theta_d \in (-\pi, \pi)$ represent the rotation angles about u and u_d. One method for transforming from a rotation matrix, R, to an axis-angle, u-θ, as shown in [33, Chapter 2.5], is given by

$$\theta = \arccos\left(\frac{1}{2}\left(\text{tr}\left(R\right) - 1\right)\right)$$

$$[u]_\times = \frac{R - R^T}{2\sin(\theta)}$$

where $\text{tr}(\cdot)$ denotes the matrix trace operator, and $[u]_\times$ denotes the skew-symmetric expansion of u. The inverse map from the axis-angle representation to the rotation matrix is given by Rodrigues' rotation formula. Finally, the feature depth estimation error, $\tilde{z}_1^* \in \mathbb{R}$, is defined as

$$\tilde{z}_1^* = z_1^* - \hat{z}_1^*. \tag{10}$$

IV. Control Development

The open loop error dynamics for the orientation states are developed by taking the time derivative of (9), resulting in the following expression [7]:

$$\dot{e}_\omega = -L_\omega \omega_c - \dot{\Theta}_d \tag{11}$$

where $\omega_c \in \mathbb{R}^3$ is the angular velocity of the camera (i.e. the angular velocity of $\mathcal{F}$ with respect to $\mathcal{F}^*$) expressed in the coordinate system of $\mathcal{F}$, and $L_\omega \in \mathbb{R}^{3\times 3}$ is defined as

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

where I_3 denotes the 3x3 identity matrix, and the function $\text{sinc}(\cdot) : \mathbb{R} \to \mathbb{R}$ is defined as

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

The matrix L_ω can be shown to be invertible for all $\theta \in (-\pi, \pi)$ [7]. The open loop dynamics for the translation error can be derived as [7]

$$z_1^* \dot{e}_v = -\alpha_1 L_v v_c + \phi z_1^* \tag{12}$$

where $\phi \in \mathbb{R}^3$ is a time-varying signal defined as $\phi \triangleq L_v [m_1]_\times \omega_c - \dot{p}_{ed}$, $v_c \in \mathbb{R}^3$ is the linear velocity of the camera (i.e., the linear velocity of $\mathcal{F}$ with respect to $\mathcal{F}^*$) expressed in the coordinate system of $\mathcal{F}$, $L_v \in \mathbb{R}^{3\times 3}$ is an invertible matrix defined as

$$L_v \triangleq \left(A - \begin{bmatrix} 0 & 0 & u_0 \\ 0 & 0 & v_0 \\ 0 & 0 & 0 \end{bmatrix}\right) \begin{bmatrix} 1 & 0 & -\frac{x_1(t)}{z_1(t)} \\ 0 & 1 & -\frac{y_1(t)}{z_1(t)} \\ 0 & 0 & 1 \end{bmatrix}$$

[2]Although the axis-angle orientation parametrization is considered here, the learning strategy employed as part of the parameter adaptation law discussed in the next section is also applicable when using the quaternion parameterization [9].

and u_0, $v_0 \in \mathbb{R}$ are the principle point coordinates of the camera, i.e. the pixel coordinates of the intersection of the optical axis of the camera with the image sensor plane. Note that L_v is known and can be calculated from A and m_1, and therefore, $\phi(t)$ is known.

Remark 1. Typically, in tracking control problems, the controller is designed with a feedforward component based on the desired trajectory, with an added restriction that the desired trajectory and its derivatives are continuous. In the visual servoing problem, the desired trajectory, encoded in p_{ed}, u_d and θ_d, can come from camera images, taken at a finite frame rate (e.g., "teach-by-showing"). As described in [7, Remark 3], a sufficiently smooth function can be fit to the desired trajectory such that the derivatives are all known and bounded. In the following development, p_{ed}, u_d, θ_d, $\dot{p}_{ed}$, $\dot{u}_d$ and $\dot{\theta}_d$ are assumed to be continuous, known, and bounded.

Based on the open loop error dynamics in (11) and (12), and the control objective, the inputs to the system, i.e. the camera velocities, are designed as

$$\omega_c = L_\omega^{-1}\left(K_\omega e_\omega - \dot{\Theta}_d\right) \tag{13}$$

$$v_c = \frac{1}{\alpha_1} L_v^{-1}\left[K_v e_v + \phi \hat{z}_1^*\right] \tag{14}$$

where K_ω, $K_v \in \mathbb{R}^{3\times 3}$ are constant, positive definite, diagonal gain matrices. The adaptive update law for the feature depth estimate is inspired by the concurrent learning method in [25] and is given by

$$\begin{aligned}\dot{\hat{z}}_1^* &= \gamma_1 e_v^T \phi \\ &+ \gamma_1\gamma_2 \sum_{k=1}^{N} \left(\dot{e}_v(t_k) - \phi(t_k)\right)^T \left[-\alpha_1(t_k) L_v(t_k) v_c(t_k)\right] \\ &+\gamma_1\gamma_2 \sum_{k=1}^{N} \left(\dot{e}_v(t_k) - \phi(t_k)\right)^T \left[-\left(\dot{e}_v(t_k) - \phi(t_k)\right)\hat{z}_1^*(t)\right]\end{aligned} \tag{15}$$

where γ_1, $\gamma_2 \in \mathbb{R}_{>0}$ are positive constant gains, $N \in \mathbb{Z}^+$is a positive constant, $t_k \in [0, t]$ are time points between the initial time and the current time, and the concurrent learning term (i.e., the second term) in (15) represents saved data. The principal idea behind this design is to utilize recorded input-output data generated by the dynamics in (12) to further improve the parameter estimate. Since the second term uses recorded data, optimal smoothers can be utilized to accurately estimate $\dot{e}_v$ at the past time points t_k in cases where it cannot be directly measured (i.e., optical flow measurements are not available). Therefore, the update law in (15) is implementable. The relationship in (12) and the definition of $\tilde{z}_1^*$ can then be used to rewrite (15) as

$$\dot{\hat{z}}_1^* = \gamma_1 e_v^T \phi + \gamma_1\gamma_2 \sum_{k=1}^{N} \left\|\dot{e}_v(t_k) - \phi(t_k)\right\|^2 \tilde{z}_1^*(t). \tag{16}$$

Substituting the controllers given in (13) and (14) into the open loop error dynamics in (11) and (12) results in

$$\dot{e}_\omega = -K_\omega e_\omega \tag{17}$$

$$z_1^* \dot{e}_v = -K_v e_v + \phi \tilde{z}_1^*. \tag{18}$$

Remark 2. When the controller is initialized, no recorded data is available, and the concurrent learning term is zero. Once a sufficient amount of time has passed (as described in Assumption 1), recorded data can be used to populate the history stack (i.e., the individual summands in the concurrent learning term) in (15). Time points for which recorded data is added to the history stack can be selected arbitrarily as long as the summand in (16) is nonzero and the sum monotonically increases with time; however, as shown in the subsequent stability analysis, the sum should be maximized to maximize the parameter convergence rate.

V. Stability Analysis

Assumption 1. The system is sufficiently excited over a finite duration of time, i.e. $T \triangleq \min\left\{\tau \in [0,\infty) \mid \left\|\dot{e}_v(\tau) - \phi(\tau)\right\|^2 > 0\right\}$ exists and is finite, after which the history stack is nonzero.

Remark 3. The finite excitation condition in Assumption 1 is much easier to satisfy than the PE condition used in previous visual servoing results (e.g., [18]–[20]) since excitation is only required over a finite duration of time, whereas in PE, as the name implies, excitation is required for all time. In addition, the finite excitation condition can be checked online by checking the minimum eigenvalue of the sum of the regressor in (16) (i.e., $\sum_{k=1}^{N} \left\|\dot{e}_v(t_k) - \phi(t_k)\right\|^2$), whereas for PE it is difficult to find the correct time window, T, to check if the integral of the regressor is lower bounded.

Theorem 1. *The controller in (13) and (14), along with the adaptive update law given in (15), ensure the tracking errors and parameter estimation errors are bounded during the interval* $0 \le t \le T$.

Proof: During the interval $0 \le t \le T$, the history stack is empty and the concurrent learning term in (16) is zero. The stability analysis in [7] demonstrates asymptotic stability of the states. This implies $\|\eta(t)\| \le \sqrt{\frac{\beta_2}{\beta_1}}\|\eta(0)\|$, $\forall t \in [0, T]$, where $\eta \triangleq \begin{bmatrix} e_\omega^T & e_v^T & \tilde{z}_1^* \end{bmatrix}^T$. ■

Theorem 2. *The controller in (13) and (14), along with the adaptive update law given in (15), ensure the tracking errors and parameter estimation errors are exponentially driven to the origin, in the sense that there exists positive constants* λ, $C \in \mathbb{R}_{>0}$ *such that*

$$\|\eta(t)\| \le C\|\eta(0)\| e^{-\lambda t}.$$

Proof: Let $V : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ be a positive definite function defined as

$$V(e_\omega, e_v, \tilde{z}_1^*) \triangleq \frac{1}{2} e_\omega^T e_\omega + \frac{z_1^*}{2} e_v^T e_v + \frac{1}{2\gamma_1}\tilde{z}_1^{*2}. \tag{19}$$

The Lyapunov function candidate in (19) can be bounded as

$$\beta_1 \|\eta\|^2 \leq V \leq \beta_2 \|\eta\|^2 \tag{20}$$

where $\beta_1 \triangleq \min\left\{\frac{1}{2}, \frac{1}{2\gamma_1}\right\}$ and $\beta_2 \triangleq \max\left\{\frac{1}{2}, \frac{1}{2\gamma_1}\right\}$. Taking the time derivative of (19) and substituting the expressions in (16)-(18) results in

$$\dot{V} \leq -e_\omega^T K_\omega e_\omega - e_v^T K_v e_v - \lambda_z \gamma_2 \tilde{z}_1^{*2} \tag{21}$$

where $\lambda_z \triangleq \|\dot{e}_v(T) - \phi(T)\|^2 > 0$ (Assumption 1 guarantees $\lambda_z > 0$). The expression in (21) can then be rewritten as

$$\dot{V} \leq -\lambda_1 V \tag{22}$$

where $\lambda_1 \triangleq \min\{\lambda_{min}(K_\omega), \lambda_{min}(K_v), \lambda_z\gamma_2\}$ and $\lambda_{min}(\cdot)$ refers to the minimum eigenvalue of $(\cdot)$. Using the Comparison Lemma [34, Lemma 3.4], the trajectory of V can be upper bounded as

$$V \leq V(e_\omega(T), e_v(T), \tilde{z}_1^*(T))\, e^{-\lambda_1(t-T)}.$$

Using (20), the tracking and estimation errors are upper bounded as

$$\|\eta(t)\| \leq \frac{\beta_2}{\beta_1} \|\eta(T)\| e^{-\frac{1}{2}\lambda_1(t-T)}.$$

From Theorem 1, this can be further bounded as

$$\|\eta(t)\| \leq e^{\frac{1}{2}\lambda_1(T)} \left(\frac{\beta_2}{\beta_1}\right) \|\eta(0)\| e^{-\frac{1}{2}\lambda_1(t)}.$$

■

Remark 4. As mentioned in Section IV, the adaptive update law in (15) utilizes state derivatives (i.e., optical flow measurements). If state derivative measurements are not available, an estimate can be used, and the tracking and parameter estimation errors can be shown to be uniformly ultimately bounded, where the residual error is proportional to the derivative estimation error, as discussed in [25].

From Theorem 2 $\|\eta\| \to 0$ as $t \to \infty$, and therefore $\|e_v\| \to 0$ and $\|e_\omega\| \to 0$. From the definition of e_v in (8), $\|e_v\| \to 0$ implies $p_e \to p_{ed}$, and since the camera calibration matrix, A, is invertible, $m_1 \to m_{1d}$ as $t \to \infty$. Also, $\|e_v\| \to 0$ implies $\frac{\alpha_{1d}}{\alpha_1} \to 1$ and therefore $z_1 \to z_{1d}$. Since $m_1 \to m_{1d}$ and $z_1 \to z_{1d}$, $\bar{m}_i \to \bar{m}_{id}$ and the position tracking objective is achieved. Similarly, $\|e_\omega\| \to 0$ implies $\Theta \to \Theta_d$. Since u is a unit vector and θ has a restricted domain, $\Theta \to \Theta_d$ implies $u \to u_d$ and $\theta \to \theta_d$. Using the Rodrigues' rotation formula, this implies $R \to R_d$, and therefore the orientation tracking objective is achieved.

VI. Conclusion

A novel adaptive controller for homography based visual servoing is developed. The adaptive update law is augmented with a concurrent learning term to identify the depth to the homography plane, and subsequently identify the structure of features on the plane. In addition to parameter convergence, the control design also guarantees exponential tracking performance. The development in this paper provides preliminary results for future efforts focused on integrating feature reconstruction into concurrent SLAM and control.

References

[1] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 651–670, Oct. 1996.

[2] G. Hu, N. Gans, and W. E. Dixon, *Encyclopedia of Complexity and Systems Science*. Springer, 2009, vol. 1, ch. Adaptive Visual Servo Control, pp. 42–63.

[3] F. Chaumette and S. Hutchinson, "Visual servo control part I: Basic approaches," *IEEE Rob. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, 2006.

[4] ——, "Visual servo control part II: Advanced approaches," *IEEE Rob. Autom. Mag.*, vol. 14, no. 1, pp. 109–118, 2006.

[5] Y. Fang, W. E. Dixon, D. M. Dawson, and P. Chawda, "Homography-based visual servo regulation of mobile robots," *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 35, pp. 1041–1050, 2005.

[6] W.-H. Chen and W. X. Zheng, "On improved robust stabilization of uncertain systems with unknown input delay," *Automatica*, vol. 42, no. 6, pp. 1067–1072, 2006.

[7] J. Chen, D. M. Dawson, W. E. Dixon, and A. Behal, "Adaptive homography-based visual servo tracking for fixed and camera-in-hand configurations," *IEEE Trans. Control Syst. Technol.*, vol. 13, pp. 814–825, 2005.

[8] G. Hu, W. Mackunis, N. Gans, W. E. Dixon, J. Chen, A. Behal, and D. Dawson, "Homography-based visual servo control with imperfect camera calibration," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1318–1324, 2009.

[9] G. Hu, N. Gans, N. Fitz-Coy, and W. E. Dixon, "Adaptive homography-based visual servo tracking control via a quaternion formulation," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 1, pp. 128–135, 2010.

[10] G. Hu, N. Gans, and W. E. Dixon, "Quaternion-based visual servo control in the presence of camera calibration error," *Int. J. Robust Nonlinear Control*, vol. 20, no. 5, pp. 489–503, 2010.

[11] E. Malis and F. Chaumette, "2 1/2 D visual servoing with respect to unknown objects through a new estimation scheme of camera displacement," *Int. J. Comput. Vision*, vol. 37, no. 1, pp. 79–97, 2000.

[12] E. Malis, F. Chaumette, and S. Bodet, "2 1/2 D visual servoing," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 238–250, 1999.

[13] F. Chaumette and E. Malis, "2 1/2 D visual servoing: a possible solution to improve image-based and position-based visual servoings," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2000, pp. 630–635.

[14] H. H. Abdelkader, Y. Mezouar, N. Andreff, and P. Martinet, "2 1/2 D visual servoing with central catadioptric cameras," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2005, pp. 3572–3577.

[15] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Upper Saddle River, NJ: Prentice-Hall, 1989.

[16] K. Narendra and A. Annaswamy, *Stable Adaptive Systems*. Prentice-Hall, Inc., 1989.

[17] K. Astrom and B. Wittenmark, *Adaptive Control*. Reading, MA: Addison-Wesley, 1989.

[18] V. Stepanyan and N. Hovakimyan, "Adaptive disturbance rejection controller for visual tracking of a maneuvering target," *J. Guid. Control Dynam.*, vol. 30, no. 4, pp. 1090–1106, 2007.

[19] ——, "Visual tracking of a maneuvering target," *J. Guid. Control Dynam.*, vol. 31, no. 1, pp. 66–80, 2008.

[20] N. Gans, G. Hu, J. Shen, Y. Zhang, and W. E. Dixon, "Adaptive visual servo control to simultaneously stabilize image and pose error," *Mechatron.*, vol. 22, no. 4, pp. 410–422, 2012.

[21] A. De Luca, G. Oriolo, and P. R. Giordano, "Feature depth observation for image-based visual servoing: Theory and experiments," *Int. J. Robot. Res.*, vol. 27, no. 10, pp. 1093–1116, 2008.

[22] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithm for estimating depth from image sequences," *Int. J. Comput. Vision*, vol. 3, pp. 209–236, 1989.

[23] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin, "Structure from controlled motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 5, pp. 492–504, May 1996.

[24] C. E. Smith and N. P. Papanikolopoulos, "Computation of shape through controlled active exploration," in *Proc. Int. Conf. Robot. Autom.*, May 1994.

[25] G. Chowdhary, T. Yucelen, M. Mühlegg, and E. N. Johnson, "Concurrent learning adaptive control of linear systems with exponentially convergent bounds," *Int. J. Adapt. Control Signal Process.*, vol. 27, no. 4, pp. 280–301, 2013.

[26] G. V. Chowdhary and E. N. Johnson, "Theory and flight-test validation of a concurrent-learning adaptive controller," *J. Guid. Control Dynam.*, vol. 34, no. 2, pp. 592–607, March 2011.
[27] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, 2nd ed., ser. CRC Applied Mathematics and Nonlinear Science. Chapman & Hall, 2011.
[28] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision*. Springer, 2004.
[29] O. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment," *Int. J. Pattern Recognit Artif Intell.*, vol. 2, no. 3, pp. 485–508, 1988.
[30] Z. Zhang and A. Hanson, "Scaled Euclidean 3D reconstruction based on externally uncalibrated cameras," in *Proc. IEEE Int. Symp. Comput. Vis.*, Nov. 1995, pp. 37–42.
[31] N. R. Gans, A. Dani, and W. E. Dixon, "Visual servoing to an arbitrary pose with respect to an object given a single known length," in *Proc. Am. Control Conf.*, Seattle, WA, USA, June 2008, pp. 1261–1267.
[32] K. Dupree, N. R. Gans, W. MacKunis, , and W. E. Dixon, "Euclidean calculation of feature points of a rotating satellite: A daisy chaining approach," *AIAA J. Guid. Control Dyn.*, vol. 31, pp. 954–961, 2008.
[33] M. Spong and M. Vidyasagar, *Robot Dynamics and Control*. New York: John Wiley & Sons Inc., 1989.
[34] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.