An Exponential Class of Model-Free Visual Servoing Controllers in the Presence of Uncertain Camera Calibration*

Y. Fang,[1] W. E. Dixon,[2] D. M. Dawson,[1] and J. Chen[1]
[1]Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634-0915
[2]Engineering Science and Technology Division, Robotics and Energetic Machines Group,
Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831-6305

E-mail: dixonwe@ornl.gov, Telephone: (865) 574-9025

Keywords: Homography, Lyapunov Analysis, Fixed Camera, and Target Tracking

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-96OR22464. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

To appear at the IEEE Conference on Decision and Control, December 9-12, 2003, Maui, HA.

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

An Exponential Class of Model-Free Visual Servoing Controllers in the Presence of Uncertain Camera Calibration*

Y. Fang,† W. E. Dixon,‡ D. M. Dawson,† and J. Chen†

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

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

email: dixonwe@ornl.gov

Abstract: In recent papers [14], [16], a new class of model-free (i.e., the 3-dimensional task-space model of the object is unknown) visual servoing methods was proposed that are based on the estimation of the relative camera orientation between two views of an object. By utilizing homography-based techniques, the control problem is decoupled by separating the rotation and translation components. A single controller is used to control the rotation component, and the class members consist of various translation controllers. Each of the current class members has been proven to yield asymptotic regulation in the presence of uncertainty in the intrinsic and extrinsic calibration parameters. New control development and stability analysis techniques are crafted in this paper to develop a new translation controller that yields exponential rotation and translation regulation in the presence of uncertainty in the intrinsic and extrinsic calibration parameters. Extensions to this research can be used to yield exponential regulation by the other translation controllers in the asymptotic class presented in [14].

I. Introduction

Motivated by the significant impact that may be realized by enabling robotic systems with the ability to perform tasks based on a sense of perception, a myriad of research has been directed at vision related issues. One issue that has limited the robustness of vision-based robotic control systems is the lack of depth information since the image-space is a 2-dimensional (2D) projection of the 3D task-space. To compensate for the lack of depth information two mainstream approaches have been developed. The first approach requires a 3D task-space model of the object so that the depth can be estimated from the distance between image feature points. The second approach requires a stereo-based camera configuration; however, the requirement for two cameras increases the cost and computational and power requirements of the system while reducing the overall system reliability. Another issue that has limited the robustness of vision-based robotic control systems is the potential for corrupt sensor data due to the lack of exact camera calibration. Specifically, based on the fact that the camera output is in the image-space and robot controllers are computed in terms of the task-space (joint space), an optic model is often employed to relate image-space data to the task-space. To relate the image-space to the task-space, both intrinsic and extrinsic parameters[1] of the optic model are required. If these parameters are not exactly known, then performance degradation and potential unpredictable response from the system may occur. Motivated by the desire to incorporate robustness to these parameters, several adaptive and robust controllers have been designed (e.g., see [7], [12], [13], [18]). Unfortunately, much of the previous work either constrains the visual servoing problem to a planar case or relies on one of the aforementioned methods to estimate the object depth.

Due to advances in computer vision, a new class of monocular visual servo controllers has been recently developed by Malis and Chaumette in [14] that only requires the relative information between a desired (reference) image and the current image. Moreover, the stability of these controllers can be proven despite the lack of exact knowledge of the camera calibration parameters. To achieve these advancements, the model-free class of controllers exploits the relative information between a desired image and the current image to construct a Euclidean homography that can be used to decouple the rotation and translation components of the visual servo problem. This decoupling strategy has been recently exploited to develop a series of results. For example, in a series of papers by Malis and Chaumette (e.g., [1], [2], [15], and [16]) various kinematic control strategies (coined 2.5D visual servo controllers) exploit information from the task-space (obtained through a projective Euclidean reconstruction from the image data) to regulate the rotation error system, while information from the 2D image-space is utilized to control the translation error system. In [6], Deguchi developed two algorithms to decouple the rotation and translation components using a homography and an epipolar condition. Specifically, Deguchi decomposes the translation and rotation components through a homography and states that the 2.5D controller given in [2] can be utilized. As an alternate method, Deguchi also develops a kinematic controller in [6] that utilizes task-space information to regulate the translation error and image-space information to regulate the rotation error. More recently, Corke and Hutchinson [5] also developed a hybrid image-based visual servoing scheme that decouples rotation and translation components about the z-axis from the remaining degrees of freedom. Motivated by the desire to actively compensate for the aforementioned depth information, [3] developed an adaptive kinematic controller to ensure uniformly ultimately bounded (UUB) set-point regulation of the image space errors while compensating for the unknown depth information, provided conditions on the translational velocity and the bounds on uncertain depth parameters are satisfied. In [4], Conticelli et al. proposed a 3D depth estimation procedure that exploits a prediction error provided a positive definite condition on the interaction matrix is satisfied. In [17], Taylor et al. developed a kinematic controller that utilizes a constant, best-guess

*This research was supported in part by the U.S. DOE Office of Biological and Environmental Research (OBER) Environmental Management Sciences Program (EMSP) projects ID No. 82794 and ID No. 82797 at ORNL, by a subcontract to ORNL by the Florida Department of Citrus through the University of Florida, and by U.S. NSF Grant DMI-9457967, ONR Grant N00014-99-1-0589, a DOC Grant, and an ARO Automotive Center Grant.

[1]The camera calibration parameters are composed of the so-called intrinsic parameters (i.e., image center, camera scale factors, and camera magnification factor) and extrinsic parameters (i.e., camera position and orientation).

estimate of the calibration parameters to achieve local set-point regulation; although, several conditions on the rotation and calibration matrix are required. In [9], Fang et al. recently developed a 2.5D visual servo controller to asymptotically regulate a manipulator end-effector by exploiting Lyapunov-based techniques to develop an adaptive update law that compensated for an unknown depth parameter. Built on the results of [9], Fang et al. designed a homography-based visual servo controller in [10] that asymptotically regulates the position of a wheeled mobile robot despite nonholonomic constraints and parametric uncertainty in the depth parameter. Although the results in [9] and [10] were achieved despite unknown depth information, the intrinsic and extrinsic camera parameters were required to be known; hence, motivation exists to develop controllers that are robust to uncertain intrinsic and extrinsic camera parameters.

In this paper, we extend the class of model-free controllers in [14] to include controllers that yield exponential translation (as opposed to the asymptotic results in [14]). That is, based on the same rotation controller as the previous asymptotic controllers, new control development and stability analysis techniques are crafted in this paper to develop a new translation controller that yields an exponential result. For completeness and to provide foundation for subsequent development, we first develop a closed-loop error system and stability theorem for the rotation controller that is developed in [14]. We then develop a new hybrid translation controller in the presence of uncertainty in the intrinsic and extrinsic camera calibration parameters. In contrast to the asymptotic results developed in [14], the controller is proven to yield exponential stability results. Specifically, the authors of [14] relied on linearization methods (e.g., Theorem 2 of [14]) or perturbation-based analysis methods to conclude local or practically global asymptotic stability. The term practically global is used in lieu of global since the result is not valid for the singular point associated with the angle of rotation or for nonpositive values for the depth from the camera to the target object. The results in this paper are developed by a nonlinear Lyapunov-based approach and formal stability proofs can be developed to prove practically global exponential rotation and translation regulation.

II. Model Development

A. Camera Model

Consider two orthogonal coordinate systems, denoted by F and F^*, where F is attached to a camera that is held by the robot end-effector, and F^* is a fixed coordinate system that represents the constant, desired position and orientation of F. Also consider a reference plane π that is defined by four[2] target points O_i $\forall i = 1, 2, 3, 4$ where the actual and desired 3D coordinates of O_i expressed in terms of F and F^* are denoted by $X_i(t), Y_i(t), Z_i(t) \in R$ and $X_i^*, Y_i^*, Z_i^* \in R$, respectively, and are defined as elements of $\bar{m}_i(t)$, $\bar{m}_i^* \in R^3$ as follows (see Figure 1)

$$\bar{m}_i = \begin{bmatrix} X_i & Y_i & Z_i \end{bmatrix}^T \tag{1}$$

$$\bar{m}_i^* = \begin{bmatrix} X_i^* & Y_i^* & Z_i^* \end{bmatrix}^T. \tag{2}$$

Since the task-space is projected onto the image-space, normalized coordinates, denoted by $m_i(t)$, m_i^*, of the targets points $\bar{m}_i(t)$ and $\bar{m}_i^*$, respectively, can be defined as follows

$$m_i = \frac{\bar{m}_i}{Z_i} = \begin{bmatrix} \frac{X_i}{Z_i} & \frac{Y_i}{Z_i} & 1 \end{bmatrix}^T \tag{3}$$

$$m_i^* = \frac{\bar{m}_i^*}{Z_i^*} = \begin{bmatrix} \frac{X_i^*}{Z_i^*} & \frac{Y_i^*}{Z_i^*} & 1 \end{bmatrix}^T \tag{4}$$

where the standard assumption is made that $Z_i(t) > 0$ and $Z_i^* > 0$.

In addition to having a task-space coordinate, each target point will also have a projected pixel coordinate expressed in terms of F denoted by $u_i(t), v_i(t) \in R$, which are defined as elements of $p_i(t)$ as follows

$$p_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T \tag{5}$$

where the projected pixel coordinates of the target points are related to the normalized task-space coordinates by the following global invertible transformation

$$p_i = A m_i \tag{6}$$

where $A \in R^{3\times 3}$ is a known, constant, and invertible intrinsic camera calibration matrix that is explicitly defined as [15]

$$A = \begin{bmatrix} f k_u & -f k_u \cot\phi & u_0 \\ 0 & \frac{f k_v}{\sin\phi} & v_0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

In (7), the constant parameters u_0, $v_0 \in R$ denote the pixel coordinates of the principal point (i.e., the image center that is defined as the frame buffer coordinates of the intersection of the optical axis with the image plane), k_u, $k_v \in R$ represent camera scaling factors, $\phi \in R$ is the angle between the camera axes, and $f \in R$ denotes the camera focal length. Similarly, the constant, desired pixel coordinates expressed in terms of F^* denoted by $u_i^*, v_i^* \in R$, are defined as elements of p_i^* as follows

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

and can be related to the normalized coordinates m_i^* by the following relationship

$$p_i^* = A m_i^*. \tag{9}$$

From (6) and (9), it is clear that the normalized task-space coordinates of a feature point can be determined from the pixel coordinates of the point. However, this relationship requires the intrinsic calibration parameters to be exactly known. Since the intrinsic camera calibration matrix A is difficult to exactly determine in practice, the computed normalized coordinates are actually estimates, denoted by $\hat{m}_i(t), \hat{m}_i^* \in R^3$, of the true values. These estimates can be expressed as follows [14]

$$\hat{m}_i = \hat{A}^{-1} p_i = \tilde{A} m_i \tag{10}$$

$$\hat{m}_i^* = \hat{A}^{-1} p_i^* = \tilde{A} m_i^* \tag{11}$$

where $\hat{A} \in R^{3\times 3}$ denotes a best-guess estimate of the intrinsic camera calibration matrix A, and the calibration error matrix $\tilde{A} \in R^{3\times 3}$ is defined as follows

$$\tilde{A} = \hat{A}^{-1} A = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} \\ 0 & \tilde{A}_{22} & \tilde{A}_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

where $\tilde{A}_{11}$, $\tilde{A}_{12}$, $\tilde{A}_{13}$, $\tilde{A}_{22}$, $\tilde{A}_{23} \in R$ denote unknown intrinsic calibration mismatch constants.

[2] In general, only 3 points are required to define a plane, however, in the subsequent analysis, 4 target points located on the plane π are required.

B. Homography Development

Based on geometric relationships between the desired image of a target and the current target image (see Figure 1), $m_i(t)$ and m_i^* can be related as follows [9] [15]

$$m_i = \alpha_i H m_i^* \tag{13}$$

where $\alpha_i(t) \in R$ is an unknown scaling factor defined as

$$\alpha_i = \frac{Z_i^*}{Z_i}, \tag{14}$$

and $H(t) \in R^{3\times 3}$ denotes the following Euclidean homography

$$H = R + x_h n^{*T}. \tag{15}$$

In (15), $R(t) \in SO(3)$ denotes the rotation from the desired task-space coordinates to the actual task-space coordinates of the camera, $n^* \in R^3$ denotes the constant unit normal from F^* to π, and $x_h(t) \in R^3$ is related to the actual translation vector from F to F^*, denoted by $x_f(t) \in R^3$, as follows

$$x_f = x_h d^* \tag{16}$$

where $d^* \in R$ denotes a constant, unknown distance from F^* to π. Since $m_i(t)$ and m_i^* can not be exactly determined, the estimates in (10) and (11) can be substituted into (13) to obtain the following relationship

$$\hat{m}_i = \alpha_i \hat{H} \hat{m}_i^* \tag{17}$$

where $\hat{H}(t) \in R^{3\times 3}$ denotes the estimated Euclidean homography [14]

$$\hat{H} = \tilde{A} H \tilde{A}^{-1}. \tag{18}$$

Since $\hat{m}_i(t)$ and $\hat{m}_i^*$ can be determined from (10) and (11), a set of 12 linear equations can be developed from the 4 image point pairs, and (17) can be used to solve for $\hat{H}(t)$ (see [9] for additional details regarding the set of linear equations). Provided additional information is available (e.g., at least 4 vanishing points), various techniques can be used to decompose $\hat{H}(t)$ to obtain the estimated rotation and translation components as follows

$$\hat{H} = \hat{R} + \hat{x}_h \hat{n}^{*T} \tag{19}$$

where $\hat{R}(t) \in R^{3\times 3}$ is related to $R(t)$ as follows

$$\hat{R} = \tilde{A} R \tilde{A}^{-1}, \tag{20}$$

and $\hat{x}_h(t) \in R^3$, $\hat{n}^{*T} \in R^3$ denote the estimate of $x_h(t)$ and n^*, respectively, and are defined as follows

$$\hat{x}_h = \gamma \tilde{A} x_h \qquad \hat{n}^* = \frac{1}{\gamma} \tilde{A}^{-T} n^* \tag{21}$$

where $\gamma \in R$ denotes the following positive constant

$$\gamma = \left\| \tilde{A}^{-T} n^* \right\|. \tag{22}$$

Remark 1: Vanishing points are points on the plane at infinity. As the reference plane π approaches infinity, the scaling term d^* also approaches infinity, and $x_h(t)$, $\hat{x}_h(t)$ approach zero. Hence, (19) can be used to conclude that $\hat{H}(t) = \hat{R}(t)$ on the plane at infinity, and the four vanishing point pairs can be used along with (17) to determine $\hat{R}(t)$. Once $\hat{R}(t)$ has been determined, then the original four image point pairs can be used to determine $\hat{x}_h(t)$ and $\hat{n}^*(t)$.

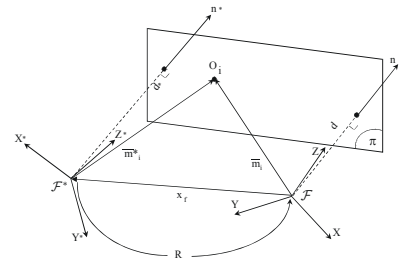

Fig. 1. Motion and structure parameters

C. Control Objective

The objective of this new class of controllers is to ensure that the position/orientation of the camera coordinate frame F is regulated to the desired position/orientation F^*. The camera is mounted on the end-effector of a robot manipulator. Hence, to control the position/orientation of F, a relationship is required to relate the linear and angular camera velocities to the linear and angular velocities of the robot end-effector (i.e., the actual kinematic control input signals). This relationship is dependent on the extrinsic calibration parameters related to the position and orientation of the camera with respect to the end-effector. Specifically, the relationship between the linear and angular velocity of the camera with respect to the end-effector can be determined as follows [14]

$$\begin{bmatrix} v_c \\ \omega_c \end{bmatrix} = \begin{bmatrix} R_r & [t_r]_\times R_r \\ 0 & R_r \end{bmatrix} \begin{bmatrix} v_r \\ \omega_r \end{bmatrix} \tag{23}$$

where $v_c(t)$, $\omega_c(t) \in R^3$ denote the linear and angular velocity of the camera, respectively, while $v_r(t)$, $\omega_r(t) \in R^3$ represent the respective linear velocity and angular velocity of the end-effector, $R_r \in SO(3)$ is the unknown constant rotation between camera and end-effector frames, and $t_r \in R^3$ denotes the unknown constant translation between camera and end-effector frames (R_r and t_r consist of the so-called camera extrinsic matrix).

Based on the development given in Section II-B, it can be shown that the control objective is achieved if the Euclidean homography $H(t)$ approaches the identity matrix. Mathematically, it can be shown that if

$$R(t) \to I_3, \tag{24}$$

and one target point is regulated to its desired location in the sense that

$$\bar{m}_i(t) \to \bar{m}_i^* \tag{25}$$

then the Euclidean homography approaches the identity matrix as follows

$$H(t) \to I_3. \tag{26}$$

In the subsequent analysis for the rotation controller, the objective is to force the angle of rotation to zero. If the angle of rotation is zero, then the objective in (24) will be met. In the analysis for the subsequent hybrid controllers, the objective is to prove that $m_i(t) \to m_i^*$ and that $Z_i(t) \to Z_i^*$. If $m_i(t) \to m_i^*$ and $Z_i(t) \to Z_i^*$, then (1)-(4) can be used to conclude that the objective in (25) will be met. Provided these objectives can be met with an exponential rate of convergence, then the main objective in (26) will be satisfied exponentially fast.

III. Rotation Control

To quantify the rotation mismatch between F and F^* (i.e., $R(t)$ given in (15)), a rotation error-like signal, denoted by $e_\omega(t) \in R^3$, is defined by the angle axis representation as follows [16]

$$e_\omega = u\theta \tag{27}$$

where $u(t) \in R^3$ represents a unit rotation axis, and $\theta(t) \in R$ denotes the rotation about $u(t)$ that is assumed to be confined to the following region [16]

$$-\pi < \theta(t) < \pi. \tag{28}$$

The parameterization $u(t)\theta(t)$ is related to the rotation matrix $R(t)$ by the following expression

$$R = I_3 + \sin\theta\,[u]_\times + 2\sin^2\frac{\theta}{2}[u]_\times^2 \tag{29}$$

where the notation $[u]_\times$ denotes the 3×3 skew-symmetric matrix associated with $u(t)$. After some mathematical development, the open-loop error dynamics for $e_\omega(t)$ can be expressed as follows [16]

$$\dot{e}_\omega = -L_\omega R_r \omega_r \tag{30}$$

where $L_\omega(t) \in R^{3\times 3}$ is defined as follows

$$L_\omega = I_3 - \frac{\theta}{2}[u]_\times + \left(1 - \frac{sinc(\theta)}{sinc^2\left(\frac{\theta}{2}\right)}\right)[u]_\times^2. \tag{31}$$

Since the rotation matrix $R(t)$ and the rotation error $e_\omega(t)$ defined in (27) are unmeasurable, an estimated rotation error $\hat{e}_\omega(t) \in R^3$ is defined as follows

$$\hat{e}_\omega = \hat{u}\hat{\theta} \tag{32}$$

where $\hat{u}(t) \in R^3$, $\hat{\theta}(t) \in R$ represent the estimate of $u(t)$ and $\theta(t)$, respectively. Based on (20), it is clear that $\hat{R}(t)$ is similar to $R(t)$. By exploiting the properties of similar matrices (i.e., similar matrices have the same trace and eigenvalues), the following estimates can be determined [14]

$$\hat{\theta} = \theta \qquad \hat{u} = \mu\tilde{A}u \tag{33}$$

where $\mu \in R$ denotes the following unknown positive constant

$$\mu = \frac{1}{\left\|\tilde{A}u\right\|}. \tag{34}$$

After substituting (33) into (32), $\hat{e}_\omega(t)$ can be expressed in terms of $e_\omega(t)$ as follows

$$\hat{e}_\omega = \mu\tilde{A}e_\omega \tag{35}$$

where (27) has been utilized. Given the open-loop rotation error dynamics in (30), the control input $\omega_r(t)$ can be designed as follows

$$\omega_r = \lambda_w \hat{R}_r^T \hat{e}_\omega \tag{36}$$

where $\lambda_\omega \in R$ denotes a positive control gain, and $\hat{R}_r \in R^{3\times 3}$ denotes a constant best-guess estimate of R_r. After substituting (36) into (30) for $\omega_r(t)$, the following closed-loop dynamics can be obtained [14]

$$\dot{e}_\omega = -\lambda_w \mu L_\omega \tilde{R}_r \tilde{A} e_\omega \tag{37}$$

where (35) has been utilized, and the rotation estimate error matrix $\tilde{R}_r \in R^{3\times 3}$ is defined as follows

$$\tilde{R}_r = R_r \hat{R}_r^T. \tag{38}$$

Remark 2: The angle axis representation in (27) is not unique, in the sense that a rotation of $-\theta(t)$ about $-u(t)$ is equal to a rotation of $\theta(t)$ about $u(t)$. A particular solution for $\theta(t)$ and $u(t)$ can be determined as follows

$$\theta_p = \cos^{-1}\left(\frac{1}{2}\left(\mathrm{tr}\left(\bar{R}\right) - 1\right)\right) \qquad [u_p]_\times = \frac{\bar{R} - \bar{R}^T}{2\sin(\theta_p)} \tag{39}$$

where the notation $\mathrm{tr}(\cdot)$ denotes the trace of a matrix and $[u_p]_\times$ denotes the 3×3 skew-symmetric form of $u_p(t)$. From (39), it is clear that

$$0 \le \theta_p(t) \le \pi. \tag{40}$$

While (40) is confined to a smaller region than $\theta(t)$ in (28), it is not more restrictive in the sense that

$$u_p\theta_p = u\theta. \tag{41}$$

The constraint in (40) is consistent with the computation of $[u(t)]_\times$ in (39) since a clockwise rotation (i.e., $-\pi \le \theta(t) \le 0$) is equivalent to a counterclockwise rotation (i.e., $0 \le \theta(t) \le \pi$) with the axis of rotation reversed. Hence, based on (41) and the functional structure of the object kinematics, the particular solutions $\theta_p(t)$ and $u_p(t)$ can be used in lieu of $\theta(t)$ and $u(t)$ without loss of generality and without confining $\theta(t)$ to a smaller region. Since we do not distinguish between rotations that are off by multiples of 2π, all rotational possibilities are considered via the parameterization of (27) along with the computation of (39).

Theorem 1: The kinematic control input given in (36) ensures that $e_\omega(t)$ defined in (27) is exponentially regulated in the sense that

$$\|e_\omega(t)\| \le \|e_\omega(0)\| \exp(-\lambda_w \mu \beta_0 t), \tag{42}$$

provided the following inequality is satisfied

$$x^T\left(\tilde{R}_r\tilde{A}\right)x \ge \beta_0\|x\|^2 \quad \text{for } \forall x \in \mathbb{R}^3 \tag{43}$$

where

$$x^T\left(\tilde{R}_r\tilde{A}\right)x = x^T\left(\tilde{R}_r\tilde{A}\right)^T x = x^T\left(\frac{\tilde{R}_r\tilde{A} + \left(\tilde{R}_r\tilde{A}\right)^T}{2}\right)x \tag{44}$$

for $\forall x \in R^3$, and $\beta_0 \in R$ denotes the following minimum eigenvalue

$$\beta_0 = \lambda_{\min}\left\{\frac{\tilde{R}_r\tilde{A} + \left(\tilde{R}_r\tilde{A}\right)^T}{2}\right\}. \tag{45}$$

Proof: Details available upon request, also see [14].

IV. New Translation Control Class

As stated previously, the contribution of this paper is to extend the class of model-free controllers developed in [14] to include a new set of translation controllers that yield exponential stability results (as opposed to the asymptotic results presented in [14]). In [14], three different sets of translation controllers are developed including a hybrid controller, an alternate hybrid controller, and a model-free position-based controller. In this paper, we develop a hybrid translation controller. Extensions to the alternate hybrid controller and a model-free position-based controller can also be developed (details available upon request).

A. Hybrid Translation Control

A.1 Control Design

To quantify the translation mismatch between the actual and desired 3D Euclidean camera position, a hybrid[3] translation error signal, denoted by $e_v(t) \in R^3$, is defined as follows

$$e_v = m_e - m_e^* \tag{46}$$

where $m_e(t) \in R^3$ denotes the extended coordinates of an image point on π in terms of F and is defined as follows[4]

$$\begin{aligned} m_e &= \begin{bmatrix} m_{e1}(t) & m_{e2}(t) & m_{e3}(t) \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{X_1}{Z_1} & \frac{Y_1}{Z_1} & \ln(Z_1) \end{bmatrix}^T \end{aligned} \tag{47}$$

and $m_e^* \in R^3$ denotes the extended coordinates of the corresponding desired image point on π in terms of F^* as follows

$$\begin{aligned} m_e^* &= \begin{bmatrix} m_{e1}^* & m_{e2}^* & m_{e3}^* \end{bmatrix}^T \\ &= \begin{bmatrix} \frac{X_1^*}{Z_1^*} & \frac{Y_1^*}{Z_1^*} & \ln(Z_1^*) \end{bmatrix}^T \end{aligned} \tag{48}$$

where $\ln(\cdot)$ denotes the natural logarithm. Substituting (47) and (48) into (46) yields

$$e_v = \begin{bmatrix} \frac{X_1}{Z_1} - \frac{X_1^*}{Z_1^*} & \frac{Y_1}{Z_1} - \frac{Y_1^*}{Z_1^*} & \ln\left(\frac{Z_1}{Z_1^*}\right) \end{bmatrix}^T \tag{49}$$

where the ratio $\frac{Z_1}{Z_1^*}$ can be computed from (14) and the decomposition of the estimated Euclidean homography in (17), despite the fact that the individual signals Z_1^* and $Z_1(t)$ are not measurable. To facilitate the subsequent development, an estimate for (49), denoted as $\hat{e}_v(t) \in R^3$, is defined as follows

$$\hat{e}_v = \begin{bmatrix} \hat{m}_{e1} - \hat{m}_{e1}^* & \hat{m}_{e2} - \hat{m}_{e2}^* & \ln\left(\frac{Z_1}{Z_1^*}\right) \end{bmatrix}^T \tag{50}$$

where $\hat{m}_{e1}(t)$, $\hat{m}_{e2}(t)$, $\hat{m}_{e1}^*$, $\hat{m}_{e2}^* \in R$ denote estimates of $m_{e1}(t)$, $m_{e2}(t)$, m_{e1}^*, m_{e2}^*, respectively, that can be calculated from (10) and (11).

By taking the time derivative of (49) and then substituting (36) into the resulting expression for $\omega_r(t)$, the following simplified error dynamics for $e_v(t)$ can be obtained [15]

$$\dot{e}_v = L_v R_r v_r + \lambda_w \left(L_v [t_r]_\times + L_{v\omega}\right) \tilde{R}_r \hat{e}_\omega \tag{51}$$

where (23), (38), and the following fact have been utilized [16]

$$\dot{\bar{m}}_1 = -v_c + [\bar{m}_1]_\times \omega_c. \tag{52}$$

In (51), $L_v(t)$, $L_{v\omega}(t) \in R^{3\times 3}$ denote the following matrices

$$L_v = \frac{1}{Z_1} \begin{bmatrix} -1 & 0 & m_{e1} \\ 0 & -1 & m_{e2} \\ 0 & 0 & -1 \end{bmatrix} \tag{53}$$

$$L_{v\omega} = \begin{bmatrix} m_{e1}m_{e2} & -1 - m_{e1}^2 & m_{e2} \\ 1 + m_{e2}^2 & -m_{e1}m_{e2} & -m_{e1} \\ -m_{e2} & m_{e1} & 0 \end{bmatrix}. \tag{54}$$

[3]The translation error introduced in (46) is described as a hybrid, because it is a composed of the normalized coordinates (obtained from the image-space) along with the reconstructed task-space depth signal.

[4]To develop the translation controller a single feature point can be utilized. Without loss of generality, the subsequent development will be based on the image point O_1, and hence, the subscript 1 will be utilized in lieu of i.

To facilitate the subsequent development, an estimates for (53), denoted by $\hat{L}_v(t) \in R^{3\times 3}$, is defined as follows

$$\hat{L}_v = \frac{1}{\hat{Z}_1} \begin{bmatrix} -1 & 0 & \hat{m}_{e1} \\ 0 & -1 & \hat{m}_{e2} \\ 0 & 0 & -1 \end{bmatrix} \tag{55}$$

where $\hat{m}_{e1}(t)$, $\hat{m}_{e2}(t)$ were introduced in (50), and $\hat{Z}_1(t) \in R$ denotes an estimate of the depth $Z_1(t)$. To determine the estimate $\hat{Z}_1(t)$, we note that

$$Z_1 = \frac{1}{\alpha_1} Z_1^* \tag{56}$$

where α_1 is determined from the decomposition of the homography in (17). Therefore, the estimated depth $\hat{Z}_1(t)$ can be determined as follows

$$\hat{Z}_1 = \frac{1}{\alpha_1} \hat{Z}_1^* \tag{57}$$

where $\hat{Z}_1^* \in R$ denotes a constant estimate of the unknown depth Z_1^*. Given the definition of $\hat{L}_v(t)$ in (55), the following inequality can be developed (details available upon request)

$$x^T \left(\hat{L}_v \hat{L}_v^T\right) x \geq \frac{1}{\hat{Z}_1^2} f(\hat{m}_{e1}, \hat{m}_{e2}) \|x\|^2 \quad \text{for } \forall x \in \mathbb{R}^3, \tag{58}$$

where $f(\hat{m}_{e1}, \hat{m}_{e2}) \in R$ denotes the following positive function

$$\begin{aligned} f(\hat{m}_{e1}, \hat{m}_{e2}) &= \frac{1}{6}\hat{m}_{e1}^2 + \frac{1}{6}\hat{m}_{e2}^2 + \frac{1}{3} \\ &\quad - \frac{1}{3}\sqrt{\left(\frac{1}{2}\hat{m}_{e1}^2 + \frac{1}{2}\hat{m}_{e2}^2 + 1\right)^2 - 1}\,. \end{aligned} \tag{59}$$

From (59), it is clear that if $\hat{m}_{e1}(t), \hat{m}_{e2}(t) \in L_\infty$, then $f(\hat{m}_{e1}, \hat{m}_{e2}) \in L_\infty$. Moreover, it can be proven that $f(\hat{m}_{e1}, \hat{m}_{e2})$ can be lower bounded by a positive constant $c_1 \in R$ as follows (details available upon request)

$$f(\hat{m}_{e1}, \hat{m}_{e2}) > c_1. \tag{60}$$

Based on the structure of the error system developed in (51) and the subsequent stability analysis, the following hybrid translation controller can be developed

$$v_r = -\lambda_v \hat{R}_r^T \hat{L}_v^T \hat{e}_v + \varphi \tag{61}$$

where $\hat{R}_r^T$, $\hat{e}_v(t)$, and $\hat{L}_v(t)$ are introduced in (36), (50), and (55), respectively, and the auxiliary signal $\varphi(t) \in R^3$ is designed as follows

$$\varphi = \left(-k_{n1}\hat{Z}_1^2 - k_{n2}\hat{Z}_1^2 \|\hat{e}_v\|^2\right) \hat{R}_r^T \hat{L}_v^T \hat{e}_v \tag{62}$$

where k_{n1}, $k_{n2} \in R$ represent positive constant control gains, and $\hat{Z}_1(t)$ is defined in (57). In (61), $\lambda_v(t) \in R$ denotes a positive gain function selected as follows

$$\lambda_v = k_{n0} + \frac{\hat{Z}_1^2}{f(\hat{m}_{e1}, \hat{m}_{e2})} \tag{63}$$

where $k_{n0} \in R$ is a positive constant, and $f(\hat{m}_{e1}, \hat{m}_{e2})$ was defined in (59). After substituting (61) into (51) for $v_r(t)$, the closed-loop dynamics for $e_v(t)$ can be obtained as follows

$$\dot{e}_v = -\lambda_v L_v \tilde{R}_r \hat{L}_v^T \hat{e}_v + L_v R_r \varphi + \lambda_w \left(L_v [t_r]_\times + L_{v\omega}\right) \tilde{R}_r \hat{e}_\omega \tag{64}$$

where (35) has been utilized.

Based on (10), (11), (47), (48), and (50), the following property can be determined

$$\hat{e}_v = Be_v \tag{65}$$

where $B \in R^{3\times 3}$ is a constant, invertible matrix defined as follows

$$B = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & 0 \\ 0 & \tilde{A}_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{66}$$

Furthermore, based on the special structure of the matrix $L_v(t)$ and the fact that $\tilde{A}$ defined in (12) is upper-triangular, B can also be expressed as follows

$$B = \eta \hat{L}_v \tilde{A} L_v^{-1} \tag{67}$$

where $\eta \in R$ is defined as

$$\eta = \frac{\hat{Z}_1}{Z_1} \tag{68}$$

and $L_v^{-1}(t)$ is given by the following expression

$$L_v^{-1} = Z_1 \begin{bmatrix} -1 & 0 & -m_{e1} \\ 0 & -1 & -m_{e2} \\ 0 & 0 & -1 \end{bmatrix}. \tag{69}$$

Note that by dividing (57) by (56), it is clear that

$$\frac{\hat{Z}_1}{Z_1} = \frac{\hat{Z}_1^*}{Z_1^*}, \tag{70}$$

and hence, from (68) it is clear that η is a positive constant. After taking the time derivative of (65) and substituting (64) into the resulting expression for $\dot{e}_v(t)$, the closed-loop dynamics for $\dot{\hat{e}}_v(t)$ can be obtained as follows

$$\begin{aligned} \dot{\hat{e}}_v &= -\lambda_v \eta \hat{L}_v \tilde{A} \tilde{R}_r \hat{L}_v^T \hat{e}_v + \eta \hat{L}_v \tilde{A} R_r \varphi \\ &\quad + \lambda_w \hat{L}_v \tilde{A} \left(\eta \left[t_r \right]_\times - \hat{Z}_1 \left[m_1 \right]_\times \right) \tilde{R}_r \hat{e}_\omega \end{aligned} \tag{71}$$

where (67), (68), and the following equality have been utilized

$$L_v^{-1} L_{v\omega} = -Z_1 \left[m_1 \right]_\times . \tag{72}$$

A.2 Stability Analysis

Theorem 2: The kinematic control input given in (61) and (62) ensures that the hybrid translation error signal $e_v(t)$ defined in (49) is exponentially regulated in the sense that

$$\|e_v(t)\| \leq \sqrt{2\zeta_0} \left\| B^{-1} \right\| \exp(-\frac{\zeta_1}{2} t) \tag{73}$$

provided (43) is satisfied, where B is defined in (66), and ζ_0, $\zeta_1 \in R$ denote the following positive constants

$$\begin{aligned} \zeta_0 &= \frac{1}{2} \|\hat{e}_v(0)\|^2 + \frac{\delta \mu^2 \left\| \tilde{A} \right\|^2 \|e_\omega(0)\|^2}{|2\eta\beta_0 - 2\lambda_w \mu \beta_0|} \\ \zeta_1 &= \min\{2\eta\beta_0, 2\lambda_w \mu \beta_0\} \end{aligned} \tag{74}$$

where $\tilde{A}$, μ, β_0, λ_w, and η are defined in (12), (34), (36), (45), and (68), respectively. In (74), $\delta \in R$ denotes a positive constant that can be made arbitrarily small by increasing values of k_{n0}, k_{n1}, k_{n2}. If $\eta = \lambda_w \mu$, a repeated root exists for $e_v(t)$ and (74) is modified as follows

$$\zeta_0 = \frac{1}{2} \|\hat{e}_v(0)\|^2 + \left(\delta \mu^2 \left\| \tilde{A} \right\|^2 \|e_\omega(0)\|^2 \right) t, \qquad \zeta_1 = 2\eta\beta_0.$$

Proof: Details available upon request.

References

[1] F. Chaumette and E. Malis, "2 1/2 D Visual Servoing: A Possible Solution to Improve Image-Based and Position-based Visual Servoings," *Proc. of the IEEE International Conference on Robotics and Automation*, San Francisco, California, pp. 630-635, 2000.

[2] F. Chaumette, E. Malis, and S. Boudet, "2D 1/2 Visual Servoing with Respect to a Planar Object," *Proc. of the Workshop on New Trends in Image-Based Robot Servoing*, pp. 45-52, 1997.

[3] F. Conticelli and B. Allotta, "Nonlinear Controllability and Stability Analysis of Adaptive Image-Based Systems," *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 2, (2001).

[4] F. Conticelli and B. Allota, "Discrete-Time Robot Visual Feedback in 3-D Positioning Tasks with Depth Adaptation," *IEEE/ASME Transactions on Mechatronics*, Vol. 6, No. 3, pp. 356-363, (2001).

[5] P. I. Corke and S. A. Hutchinson, "A New Hybrid Image-Based Visual Servo Control Scheme," *Proc. of the IEEE Conference on Decision and Control*, Sydney, Australia, pp. 2521-2527, 2000.

[6] K. Deguchi, "Optimal Motion Control for Image-Based Visual Servoing by Decoupling Translation and Rotation," *Proc. of the Intl. Conf. on Intelligent Robots and Systems*, B.C. Canada, pp. 705-711, Oct. 1998.

[7] W. E. Dixon, D. M. Dawson, E. Zergeroglu, and A. Behal, "Adaptive Tracking Control of a Wheeled Mobile Robot via an Uncalibrated Camera System," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, Vol. 31, No. 3, pp. 341-352, (2001).

[8] Yongchun Fang, Lyapunov-Based Control for Mechanical and Vision-Based Systems, Ph.D. Dissertation, Clemson University, 2002.

[9] Y. Fang, A. Behal, W. E. Dixon and D. M. Dawson, "Adaptive 2.5D Visual Servoing of Kinematically Redundant Robot Manipulators," *Proc. of the IEEE Conference on Decision and Control*, Las Vegas, Nevada, pp. 2860-2865, Dec. 2002.

[10] Y. Fang, D. M. Dawson, W. E. Dixon, and M. S. de Queiroz, "2.5D Visual Servoing of Wheeled Mobile Robots," *Proc. of the IEEE Conference on Decision and Control*, Las Vegas, Nevada, pp. 2866-2871, Dec. 2002.

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

[12] R. Kelly and A. Marquez, "Fixed-Eye Direct Visual Feedback Control of Planar Robots," *Journal of Systems Engineering*, Vol. 4, No. 5, pp. 239-248, (1995).

[13] R. Kelly, "Robust Asymptotically Stable Visual Servoing of Planar Robots," *IEEE Transactions Robotics and Automation*, Vol. 12, No. 5, pp. 759-766, (1996).

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

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

[16] 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).

[17] C. J. Taylor and J. P. Ostrowski, "Robust Vision-based Pose Control," *Proc. of the IEEE International Conference on Robotics and Automation*, San Francisco, California, pp. 2734-2740, 2000.

[18] E. Zergeroglu, D. M. Dawson, M. de Queiroz, and A. Behal, "Vision-based Nonlinear Tracking Controllers in the Presence of Parametric Uncertainty," *IEEE/ASME Transactions on Mechatronics*, Vol. 6, No. 3, pp. 322-337, (2001).

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