Mechatronics 22 (2012) 436–443

Contents lists available at SciVerse ScienceDirect

Mechatronics

journal homepage: www.elsevier.com/locate/mechatronics

Teach by zooming: A unified approach to visual servo control ☆

S.S. Mehta [a,*], V. Jayaraman [b], T.F. Burks [b], W.E. Dixon [c]

[a] Research and Engineering Education Facility, University of Florida, 1350 N. Poquito Rd., Shalimar, FL-32579, United States
[b] Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL-32611, United States
[c] Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL-32611, United States

ARTICLE INFO

Article history:
Available online 23 December 2011

Keywords:
Teach by zooming
Teach by showing
Visual servo control
Nonlinear control

ABSTRACT

Traditionally, a visual servo control problem is formulated in the teach by showing framework with an objective to regulate a camera based on a reference (or desired) image obtained by *a priori* positioning the same camera at the desired task-space location. A new strategy is essential for a variety of applications where it may not be possible to position the camera *a priori* at the desired position/orientation. In this paper, a visual servo control approach, called "teach by zooming", is formulated where the objective is to position/orient a camera based on a reference image obtained by another camera. For example, a fixed camera providing a wide area view of the scene can zoom in on an object and record a desired image for another camera. A non-linear Lyapunov-based controller is designed to regulate the image features acquired by an on-board camera to the corresponding image feature coordinates in the desired image acquired by the fixed camera in the presence of uncertain camera calibration parameters. The proposed control formulation becomes identical to the well-known teach by showing controller when the camera-in-hand can be located *a priori* to the desired position/orientation; thus enabling control in a wide range of applications. Experimental results for regulation control of a 7 degrees-of-freedom robotic manipulator are provided to demonstrate the performance of the proposed visual servo controller.

1. Introduction

Exact knowledge of the camera calibration parameters is required to relate the pixelized image-space information to the task-space. Inevitable discrepancies in the calibration matrix may result in an erroneous relationship between the image-space and task-space. In addition, an acquired image is a function of both the task-space position of a camera and the intrinsic calibration parameters; hence, perfect knowledge of the intrinsic camera parameters is also required to relate the relative position of a camera through respective images as it moves. A typical visual servo control problem is constructed as "teach by showing" (TBS) problem, in which a camera is *a priori* positioned at the desired location to acquire a reference image and then the camera is repositioned at the same desired location by means of visual servo control. TBS control formulation requires that the calibration parameters do not change in order to reposition the camera to the same task-space location given a matching image. See [1–4] for further explanation and an overview of the TBS problem formulation.

A variety of applications may prohibit the use of TBS controller, i.e., it may not be possible to acquire a reference image by *a priori* positioning an on-board camera at the desired location. As stated in [5], TBS problem formulation is "camera-dependent" due to the assumption that intrinsic camera parameters remain unchanged between the teaching stage and servo control. In [5,6], projective invariance is used to construct an error function that is invariant of the intrinsic parameters meeting the control objective despite variations in the intrinsic parameters. A camera can be repositioned with respect to a non-planar target, requiring at least 6 feature points, where local asymptotic stability of the equilibrium point is achieved, i.e., the transformed points in an invariant space must be in the neighborhood of the desired features. However, the goal is to construct an error system in an invariant space, and unfortunately, as stated in [5,6], several control issues and rigorous stability analysis of the invariant space approach have been left unresolved.

The contribution of presented work is in the development of a new visual servo control approach, called "teach by zooming" (TBZ) control [7,8], to position/orient a camera based on a reference image obtained by another camera. The presented controller is unified in the sense that the underlying mathematical framework remains unchanged even when the problem is formulated as TBS control, i.e., when the same camera is used to acquired a reference

☆ This research is supported in part by the Department of Energy, Grant number DE-FG04-86NE37967, as part of the DOE University Research Program in Robotics (URPR).

* Corresponding author. Tel.: +1 850 833 9350; fax: +1 850 833 9366.
E-mail addresses: siddhart@ufl.edu (S.S. Mehta), venkatj@ufl.edu (V. Jayaraman), tburks@ufl.edu (T.F. Burks), wdixon@ufl.edu (W.E. Dixon).

0957-4158/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mechatronics.2011.11.010

image and perform servo control. TBZ problem can be envisioned as a fixed camera providing a wide area view of the scene that can be used to zoom in on an object of interest and record the desired image for another camera. Camera independent TBZ control strategy can be attractive to applications such as navigating ground or air vehicles based on desired images taken by other ground or air vehicles (e.g., a satellite captures a "zoomed in" desired image that is used to navigate a camera on-board an unmanned aerial vehicle or smart-munition, a camera can view the entire tree canopy and zoom in to acquire a desired image of a fruit product for high speed robotic harvesting). The advantage of TBZ control formulation is that the fixed camera can be mounted so that the complete task-space is visible, can selectively zoom in on objects of interest, and can acquire a desired image that corresponds to a desired position and orientation for an on-board camera.

TBZ controller in this paper is designed to regulate image features acquired by an on-board camera to the corresponding image feature coordinates in a reference image acquired by a zooming fixed camera. The challenge lies in developing a meaningful task-space relationship using images acquired from different cameras to achieve not only the image-space regulation but also the desired task-space control objective. As stated in [5], since the reference and current images are obtained using different cameras, irrespective of the visual servo control method, even if the image coordinates match it can not be guaranteed that the task-space objective of positioning a camera at the desired pose with respect to a target is achieved. Also, it is assumed that parametric uncertainty exists in the camera calibration, and hence, the ability to construct a meaningful relationship between the estimated and actual rotation matrix is problematic as the estimated rotation matrix may not lie on $\mathbb{SO}^3$. Generally, image-based visual servo (IBVS) control methods are considered to be computationally inexpensive and robust to camera calibration errors [4,9] than homography-based methods since the control objective is written in terms of image coordinates regulation. For the proposed problem, IBVS control objective can be established in terms of regulation of the current image coordinates to the 'virtual' image coordinates that achieve the desired task-space positioning objective. The virtual image coordinates can be obtained by expressing the desired image coordinates from a reference image captured by a fixed camera in terms of an on-board camera. It can be shown that the resulting virtual image coordinates are functions of the calibration parameters of both the on-board and fixed camera, and therefore IBVS methods may not demonstrate robustness with respect to uncertainties in the intrinsic camera parameters. In addition, it is well known that IBVS control may result in unrealizable and sub-optimal task-space trajectories, the interaction matrix or image Jacobian J may become singular during servoing thus resulting in system instability, a local minima may be reached for certain image-space trajectories, and the solution of J (or J^+) requires time-varying depth measurements [10,11]. To overcome these challenges, the control objective is formulated in terms of normalized Euclidean coordinates that are invariant to changes in the calibration parameters by defining a virtual camera at the desired task space location and by expressing the desired normalized Euclidean coordinates as a function of the mismatch in the camera calibration. This is a physically motivated relationship, since an image is a function of both the Euclidean camera position and the camera calibration. Since the estimates of only the static camera calibration parameters, e.g., corresponding to the minimum zoom setting, are required, it is not necessary to calibrate the fixed camera for different focal lengths. The main contribution of the presented TBZ problem formulation is that it guarantees global exponential stability of the equilibrium point while servoing with cameras having different intrinsic calibration matrices with an uncertainty in the parameters.

This paper builds on our previous efforts that have investigated the advantages of multiple cameras working in a non-stereo pair. Specifically, in [12,13], a new cooperative visual servoing approach was developed and experimentally demonstrated that using information from both an uncalibrated fixed camera and an uncalibrated on-board camera enables the on-board camera to track an object moving in the task-space with an unknown trajectory. A crucial assumption in [12,13] is that the camera and the object motion is constrained to a plane so that the unknown distance from the camera to the target remains constant. However, in contrast to [12,13], an on-board camera motion in this paper is not restricted to a plane. In our previous work [14], exponential regulation of an on-board camera is achieved despite uncertainty in the calibration parameters in contrast to asymptotic stability result in [15] by formulating a model-free rotation and composite translation controller. The proposed controller differs from [14,15] in the sense that the reference image need not be obtained using the same camera, i.e., it does not rely on the TBS paradigm. Further, TBZ control objective is formulated so that we can leverage the control development and stability analysis in [14] to achieve exponential regulation of an on-board camera in contrast to local asymptotic stability result proved in [5]. The developed controller is also invariant to the time-varying camera calibration parameters of a fixed camera since only constant parameter estimates are utilized in the control development, thus allowing the desired trajectory to be encoded using a stationary zooming-camera with e.g., time-varying focal length. Experimental results for regulation control of a 7 degrees-of-freedom (DOF) robotic manipulator are provided to demonstrate the performance of TBZ control.

2. Model development

Consider the orthogonal coordinate systems $\mathcal{F}$, $\mathcal{F}_f$, and $\mathcal{F}^*$ as depicted in Fig. 1. Coordinate system $\mathcal{F}$ is attached to an on-board camera (e.g., a camera held by a robot end-effector, a camera mounted on a vehicle) and $\mathcal{F}_f$ is attached to a fixed camera that has an adjustable focal length to zoom in on an object. A captured image is defined by both the camera calibration parameters and the Euclidean position of the camera; therefore, the feature points of an object as seen in an image acquired by the fixed camera after zooming in on the object can be expressed in terms of $\mathcal{F}_f$ in one of two ways: a different calibration matrix can be used due to the change in the focal length, or the calibration matrix can be held constant and the Euclidean position of the camera is changed to a virtual camera position and orientation. The position and orientation of the virtual camera is described by the coordinate system $\mathcal{F}^*$. A reference plane π is defined by four target points $O_i \; \forall i = 1, 2, 3, 4$ where the three dimensional (3D) coordinates of O_i expressed in terms of $\mathcal{F}$, $\mathcal{F}_f$, and $\mathcal{F}^*$ are defined as elements of $\bar{m}_i(t)$, $\bar{m}_{fi}$ and $\bar{m}_i^* \in \mathbb{R}^3$ as

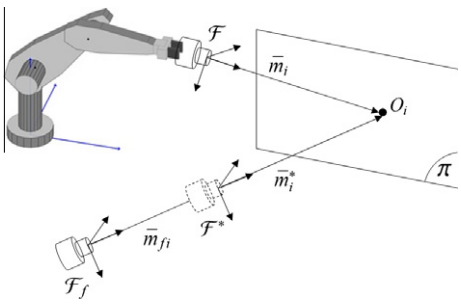

Fig. 1. Camera coordinate frame relationships, where frame $\mathcal{F}$ is attached to an on-board camera, $\mathcal{F}_f$ is attached to a fixed camera, and $\mathcal{F}^*$ represents a virtual camera.

$$\bar{m}_i = [X_i \quad Y_i \quad Z_i]^T$$
$$\bar{m}_{fi} = [X_{fi} \quad Y_{fi} \quad Z_{fi}]^T \tag{1}$$
$$\bar{m}_i^* = [X_{fi} \quad Y_{fi} \quad Z_i^*]^T.$$

The Euclidean-space is projected onto the image-space, so the normalized coordinates of the targets points $\bar{m}_i(t)$, $\bar{m}_{fi}$, and $\bar{m}_i^*$ can be defined as

$$m_i = \frac{\bar{m}_i}{Z_i} = \left[\frac{X_i}{Z_i} \quad \frac{Y_i}{Z_i} \quad 1\right]^T$$
$$m_{fi} = \frac{\bar{m}_{fi}}{Z_{fi}} = \left[\frac{X_{fi}}{Z_{fi}} \quad \frac{Y_{fi}}{Z_{fi}} \quad 1\right]^T \tag{2}$$
$$m_i^* = \frac{\bar{m}_i^*}{Z_i^*} = \left[\frac{X_{fi}}{Z_i^*} \quad \frac{Y_{fi}}{Z_i^*} \quad 1\right]^T$$

Assumption 1. The unknown target depth $Z_i(t)$, Z_i^*, and $Z_{fi} > \varepsilon$, where $\varepsilon \in \mathbb{R}$ denotes a positive definite constant. It represents a standard assumption for vision-based systems that is consistent with the fact that a camera can view objects in front of the image plane.

Based on (2) the normalized Euclidean coordinates of m_{fi} can be related to m_i^* as

$$m_{fi} = \text{diag}\left\{\frac{Z_i^*}{Z_{fi}}, \frac{Z_i^*}{Z_{fi}}, 1\right\} m_i^* \tag{3}$$

where diag $\{\cdot\}$ denotes a diagonal matrix of given arguments.

In addition to having normalized task-space coordinates, each target point will also have pixel coordinates that are acquired from an on-board camera, expressed in terms of $\mathcal{F}$, denoted by $u_i(t), v_i(t) \in \mathbb{R}$, and are defined as elements of $p_i(t) \in \mathbb{R}^3$ as

$$p_i \triangleq [u_i \quad v_i \quad 1]^T. \tag{4}$$

The pixel coordinates $p_i(t)$ and the normalized task-space coordinates $m_i(t)$ are related by the following global invertible transformation (i.e., the pinhole model):

$$p_i = A m_i. \tag{5}$$

Constant pixel coordinates, expressed in terms of $\mathcal{F}_f$ (denoted $u_{fi}, v_{fi} \in \mathbb{R}$) and $\mathcal{F}^*$ (denoted $u_i^*, v_i^* \in \mathbb{R}$) are respectively defined as elements of $p_{fi} \in \mathbb{R}^3$ and $p_i^* \in \mathbb{R}^3$ as

$$p_{fi} \triangleq [u_{fi} \quad v_{fi} \quad 1]^T \quad p_i^* \triangleq [u_i^* \quad v_i^* \quad 1]^T. \tag{6}$$

The pinhole model can also be used to relate the pixel coordinates p_{fi} and $p_i^*(t)$ to the normalized task-space coordinates m_{fi} and $m_i^*(t)$ as

$$p_{fi} = A_f m_{fi} \tag{7}$$
$$p_i^* = A^* m_{fi} \text{ or } p_i^* = A_f m_i^*. \tag{8}$$

In (8), the first expression is where the Euclidean position and orientation of the camera remains constant and the camera calibration matrix changes, and the second expression is where the calibration matrix remains the same and the Euclidean position and orientation is changed. In (5) and (8), the intrinsic calibration matrices A, A_f, and $A^* \in \mathbb{R}^{3\times3}$ denote the following constant invertible intrinsic camera calibration matrices:

$$A \triangleq \begin{bmatrix} \lambda_1 & -\lambda_1 \cot\phi & u_0 \\ 0 & \frac{\lambda_2}{\sin\phi} & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$A_f \triangleq \begin{bmatrix} \lambda_{f1} & -\lambda_{f1} \cot\phi_f & u_{0f} \\ 0 & \frac{\lambda_{f2}}{\sin\phi_f} & v_{0f} \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$
$$A^* \triangleq \begin{bmatrix} \lambda_1^* & -\lambda_1^* \cot\phi_f & u_{0f} \\ 0 & \frac{\lambda_2^*}{\sin\phi_f} & v_{0f} \\ 0 & 0 & 1 \end{bmatrix}.$$

In (9), u_0, $v_0 \in \mathbb{R}$ and u_{0f}, $v_{0f} \in \mathbb{R}$ are the pixel coordinates of the principal point of an on-board camera and fixed camera, respectively. Constants λ_1, λ_{f1}, λ_1^*, λ_2, λ_{f2}, $\lambda_2^* \in \mathbb{R}$ represent the product of camera scaling factors and focal length, and ϕ, $\phi_f \in \mathbb{R}$ are the skew angles between the camera axes for an on-board camera and fixed camera, respectively.

Since the intrinsic calibration matrix of a camera is difficult to accurately obtain, the development in this paper is based on the assumption that the intrinsic calibration matrices are unknown. Since A_f is unknown, the normalized Euclidean coordinates m_{fi} cannot be determined from p_{fi} using Eq. (7). Since m_{fi} cannot be determined, then the intrinsic calibration matrix A^* cannot be computed from (8). For the TBZ problem formulation, p_i^* defines the desired image-space coordinates. Since the normalized Euclidean coordinates m_i^* are unknown, the control objective is defined in terms of servoing an on-board camera so that the images correspond. If the image from an on-board camera and the zoomed image from a fixed camera correspond, then the following expression can be developed from (5) and (8):

$$m_i = m_{di} \triangleq A^{-1} A_f m_i^* \tag{10}$$

where $m_{di} \in \mathbb{R}^3$ denotes the normalized Euclidean coordinates of the object feature points expressed in $\mathcal{F}_d$. $\mathcal{F}_d$ denotes the coordinate system attached to an on-board camera when the image taken from an on-board camera corresponds to the image acquired from a fixed camera after zooming in on an object. Hence, the control objective for uncalibrated TBZ problem can be formulated as the desire to force $m_i(t)$ to m_{di}. Given that $m_i(t)$, m_i^*, and m_{di} are unknown, the estimates $\hat{m}_i(t)$, $\hat{m}_i^*$, and $\hat{m}_{di} \in \mathbb{R}^3$ are defined to facilitate the subsequent control development [15]

$$\hat{m}_i = \widehat{A}^{-1} p_i = \widetilde{A} m_i \tag{11}$$
$$\hat{m}_i^* = \widehat{A}_f^{-1} p_i^* = \widetilde{A}_f m_i^* \tag{12}$$
$$\hat{m}_{di} = \widehat{A}^{-1} p_i^* = \widetilde{A} m_{di} \tag{13}$$

where $\widehat{A}$, $\widehat{A}_f \in \mathbb{R}^{3\times3}$ are constant, best-guess estimates of the intrinsic camera calibration matrices[1] A and A_f, respectively. The calibration error matrices $\widetilde{A}$, $\widetilde{A}_f \in \mathbb{R}^{3\times3}$ are defined as

$$\widetilde{A} \triangleq \widehat{A}^{-1} A = \begin{bmatrix} \widetilde{A}_{11} & \widetilde{A}_{12} & \widetilde{A}_{13} \\ 0 & \widetilde{A}_{22} & \widetilde{A}_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

$$\widetilde{A}_f \triangleq \widehat{A}_f^{-1} A_f = \begin{bmatrix} \widetilde{A}_{f11} & \widetilde{A}_{f12} & \widetilde{A}_{f13} \\ 0 & \widetilde{A}_{f22} & \widetilde{A}_{f23} \\ 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

Remark 1. For the standard TBS visual servo control problem where the camera calibration parameters do not change between the teaching phase and the servo phase, i.e., $A = A_f$, the coordinate frames $\mathcal{F}_d$ and $\mathcal{F}^*$ are identical.

3. Homography development

The following expression can be obtained based on the relationship between coordinate frames $\mathcal{F}$ and $\mathcal{F}^*$ (see Fig. 1):

$$\bar{m}_i = R\bar{m}_i^* + x_f \tag{16}$$

[1] The estimates $\widehat{A}$ and $\widehat{A}_f$ can be obtained by approximate camera calibration or by referring to the manufacturer specifications.

where $R(t) \in \mathbb{R}^{3\times3}$ and $x_f(t) \in \mathbb{R}^3$ denote the rotation and translation, respectively, between $\mathcal{F}$ and $\mathcal{F}^*$. By utilizing (1) and (2), the expression in (16) can be expressed as

$$m_i = \underbrace{\left(\frac{Z_i^*}{Z_i}\right)}_{\alpha_i} \underbrace{\left(R + x_h n^{*T}\right)}_{H} m_i^* \tag{17}$$

where $x_h(t) \triangleq \frac{x_f(t)}{d^*} \in \mathbb{R}^3$ and $d^* \in \mathbb{R}$ denotes an unknown constant distance from $\mathcal{F}^*$ to π along the unit normal n^*. The following relationship can be developed by substituting (17) and (8) into (5) for $m_i(t)$ and m_i^*, respectively:

$$p_i = \alpha_i G p_i^* \tag{18}$$

where $G \in \mathbb{R}^{3\times3}$ is the projective homography matrix defined as $G(t) \triangleq AH(t)A_f^{-1}$. The expressions in (5) and (8) can be used to rewrite (18) as

$$m_i = \alpha_i A^{-1} G A_f m_i^*. \tag{19}$$

The following expression can be obtained by substituting (10) into (19):

$$m_i = \alpha_i H_d m_{di} \tag{20}$$

where $H_d(t) \triangleq A^{-1}G(t)A$ denotes the Euclidean homography matrix that can be expressed as

$$H_d = R_d + x_{hd} n_d^T \quad \text{where} \quad x_{hd} = \frac{x_{fd}}{d_d}. \tag{21}$$

In (21), $R_d(t) \in \mathbb{R}^{3\times3}$ and $x_{fd}(t) \in \mathbb{R}^3$ denote the rotation and translation, respectively, from $\mathcal{F}$ to $\mathcal{F}_d$. The constant $d_d \in \mathbb{R}$ in (21) denotes the distance from $\mathcal{F}_d$ to π along the unit normal $n_d \in \mathbb{R}^3$. Since $m_i(t)$ and m_i^* cannot be determined because the intrinsic camera calibration matrices and A_f are uncertain, the estimates $\hat{m}_i(t)$ and $\hat{m}_{di}$ defined in (11) and (12), respectively, can be utilized to obtain the following:

$$\hat{m}_i = \alpha_i \widehat{H}_d \hat{m}_{di}. \tag{22}$$

In (22), $\widehat{H}_d(t) \in \mathbb{R}^{3\times3}$ denotes the following estimated Euclidean homography [15]:

$$\widehat{H}_d = \widetilde{A} H_d \widetilde{A}^{-1}. \tag{23}$$

Since $\hat{m}_i(t)$ and $\hat{m}_{di}$ can be determined from (11) and (13), a set of linear equations can be developed to solve for $\widehat{H}_d(t)$ (see [16] for additional details regarding the set of linear equations). The expression in (23) can also be expressed as [16]

$$\widehat{H}_d = \widehat{R}_d + \hat{x}_{hd} \hat{n}_d^T. \tag{24}$$

In (24), the estimated rotation matrix, denoted $\widehat{R}_d(t) \in \mathbb{R}^{3\times3}$, is related to $R_d(t)$ as below

$$\widehat{R}_d = \widetilde{A} R_d \widetilde{A}^{-1}, \tag{25}$$

and $\hat{x}_{hd}(t) \in \mathbb{R}^3$, $\hat{n}_d^T \in \mathbb{R}^3$ denote the estimate of $x_{hd}(t)$ and n_d, respectively, and are defined as

$$\hat{x}_{hd} = \gamma \widetilde{A} x_{hd} \tag{26}$$

$$\hat{n}_d = \frac{1}{\gamma} \widetilde{A}^{-T} n_d \tag{27}$$

where $\gamma \in \mathbb{R}$ denotes the following positive constant

$$\gamma = \|\widetilde{A}^{-T} n_d\|. \tag{28}$$

Although $\widehat{H}_d(t)$ can be computed, standard techniques cannot be used to decompose $\widehat{H}_d(t)$ into the rotation and translation components in (24). Specifically, from (25) $\widehat{R}_d(t)$ is not a true rotation matrix, and hence, it is not clear how standard decomposition algorithms (e.g., the Faugeras algorithm [17,18]) can be applied. To address this issue, additional information (e.g., at least four vanishing points) can be used. For example, 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, (24) can be used to conclude that $\widehat{H}_d(t) = \widehat{R}_d(t)$ on the plane at infinity, and the four vanishing point pairs can be used along with (22) to determine $\widehat{R}_d(t)$. Once $\widehat{R}_d(t)$ has been determined, various techniques (e.g., see [17,19]) can be used along with the original four image point pairs to determine $\hat{x}_{hd}(t)$ and $\hat{n}_d(t)$.

4. Control objective

The control objective is to ensure that the position and orientation of the camera coordinate frame $\mathcal{F}$ is regulated to $\mathcal{F}_d$. Based on Section 3, the control objective is achieved if

$$R_d(t) \to I_3 \tag{29}$$

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

$$m_i(t) \to m_{di} \text{ and } Z_i(t) \to Z_{di}. \tag{30}$$

To control the position and orientation of $\mathcal{F}$, a relationship is required to relate the linear and angular camera velocities to the linear and angular velocities of the vehicle/robot (i.e., the actual kinematic control inputs) that enables an on-board camera motion. This relationship is dependent on the extrinsic calibration parameters as [15]

$$\begin{bmatrix} v_c \\ \omega_c \end{bmatrix} = \begin{bmatrix} R_r & [t_r]_\times R_r \\ 0 & R_r \end{bmatrix} \begin{bmatrix} v_r \\ \omega_r \end{bmatrix} \tag{31}$$

where $v_c(t)$, $\omega_c(t) \in \mathbb{R}^3$ denote the linear and angular velocity of the camera, $v_r(t)$, $\omega_r(t) \in \mathbb{R}^3$ denote the linear and angular velocity of the vehicle/robot, $R_r \in \mathbb{R}^{3\times3}$ denotes the unknown constant rotation between an on-board camera and robot end-effector frames, and $[t_r]_\times \in \mathbb{R}^{3\times3}$ is a skew symmetric form of $t_r \in \mathbb{R}^3$, which denotes the unknown constant translation vector between an on-board camera and vehicle/robot frames.

5. Control development

5.1. Rotation controller

To quantify the rotation between $\mathcal{F}$ and $\mathcal{F}_d$ (i.e., $R_d(t)$ given in (21)), a rotation error-like signal, denoted by $e_\omega(t) \in \mathbb{R}^3$, is defined by the angle axis representation as

$$e_\omega = u\theta \tag{32}$$

where $u(t) \in \mathbb{R}^3$ represents a unit rotation axis, and $\theta(t) \in \mathbb{R}$ denotes the rotation angle about $u(t)$ that is assumed to be constrained to the region

$$0 \leqslant \theta(t) \leqslant \pi. \tag{33}$$

The parameterization $u(t)\theta(t)$ is related to the rotation matrix $R_d(t)$ as

$$R_d = I_3 + \sin\theta [u]_\times + 2\sin^2\frac{\theta}{2}[u]_\times^2 \tag{34}$$

where $[u]_\times$ denotes the 3×3 skew-symmetric matrix associated with $u(t)$. The open-loop error dynamics for $e_\omega(t)$ can be expressed as

$$\dot{e}_\omega = -L_\omega R_r \omega_r \tag{35}$$

where $L_\omega(t) \in \mathbb{R}^{3\times3}$ is defined as

$$L_\omega = I_3 - \frac{\theta}{2}[u]_\times + \left(1 - \frac{\operatorname{sinc}(\theta)}{\operatorname{sinc}^2\left(\frac{\theta}{2}\right)}\right)[u]_\times^2. \tag{36}$$

In Eq. (36) the sinc (θ) term is given by (37) as,

$$\operatorname{sinc}(\theta) = \frac{\sin(\theta)}{\theta} \tag{37}$$

Since the rotation matrix $R_d(t)$ and the rotation error $e_\omega(t)$ defined in (32) are unmeasurable, an estimated rotation error $\hat{e}_\omega(t) \in \mathbb{R}^3$ is defined as

$$\hat{e}_\omega = \hat{u}\hat{\theta} \tag{38}$$

where $\hat{u}(t) \in \mathbb{R}^3$, $\hat{\theta}(t) \in \mathbb{R}$ represent estimates of $u(t)$ and $\theta(t)$, respectively. Since $\widehat{R}_d(t)$ is similar to $R_d(t)$ (i.e., $\widehat{R}_d(t)$ has the same trace and eigenvalues as $R_d(t)$), the estimates $\hat{u}(t)$ and $\hat{\theta}(t)$ can be related to $u(t)$ and $\theta(t)$ as [15]

$$\hat{\theta} = \theta \quad \hat{u} = \mu\widetilde{A}u \tag{39}$$

where $\mu(t) \in \mathbb{R}$ denotes the following unknown function

$$\mu = \frac{1}{\|\widetilde{A}u\|}. \tag{40}$$

The relationship in (39) allows $\hat{e}_\omega(t)$ to be expressed in terms of the unmeasurable error $e_\omega(t)$ as

$$\hat{e}_\omega = \mu\widetilde{A}e_\omega. \tag{41}$$

Given the open-loop rotation error dynamics in (35), the control input $\omega_r(t)$ is designed as

$$\omega_r = \lambda_\omega \widehat{R}_r^T \hat{e}_\omega \tag{42}$$

where $\lambda_\omega \in \mathbb{R}$ denotes a positive control gain, and $\widehat{R}_r \in \mathbb{R}^{3\times3}$ denotes a constant best-guess estimate of R_r. Substituting (41) into (42) and substituting the resulting expression into (35) gives the following expression for the closed-loop error dynamics [15]:

$$\dot{e}_\omega = -\lambda_\omega \mu L_\omega \widetilde{R}_r \widetilde{A} e_\omega \tag{43}$$

where the extrinsic rotation estimation error $\widetilde{R}_r \in \mathbb{R}^{3\times3}$ is defined as

$$\widetilde{R}_r = R_r \widehat{R}_r^T. \tag{44}$$

Property 1. *The kinematic control input given in (42) ensures that $e_\omega(t)$ defined in (32) is exponentially regulated in the sense that* [14]

$$\|e_\omega(t)\| \leqslant \|e_\omega(0)\| \exp(-\lambda_\omega \mu \beta_0 t) \tag{45}$$

provided the following inequality is satisfied:

$$x^T(\widetilde{R}_r\widetilde{A})x \geqslant \beta_0\|x\|^2 \text{ for } \forall x \in \mathbb{R}^3 \tag{46}$$

where

$$x^T\left(\widetilde{R}_r\widetilde{A}\right)x = x^T\left(\widetilde{R}_r\widetilde{A}\right)^T x = x^T\left(\frac{\widetilde{R}_r\widetilde{A} + (\widetilde{R}_r\widetilde{A})^T}{2}\right)x \tag{47}$$

for $\forall x \in \mathbb{R}^3$, and $\beta_0 \in \mathbb{R}$ denotes the minimum eigenvalue as below

$$\beta_0 = \lambda_{\min}\left\{\frac{\widetilde{R}_r\widetilde{A} + (\widetilde{R}_r\widetilde{A})^T}{2}\right\}. \tag{48}$$

5.2. Translation controller

The difference between the actual and desired 3*D* Euclidean camera position, denoted by the translation error signal $e_v(t) \in \mathbb{R}^3$, is defined as

$$e_v \triangleq m_e - m_{de} \tag{49}$$

where $m_e(t) \in \mathbb{R}^3$ denotes the extended coordinates of an image point on π expressed in terms of $\mathcal{F}$ and is defined as[2]

$$m_e \triangleq [\, m_{e1}(t) \quad m_{e2}(t) \quad m_{e3}(t)\,]^T = \left[\frac{X_1}{Z_1} \quad \frac{Y_1}{Z_1} \quad \ln(Z_1)\right]^T \tag{50}$$

and $m_{de} \in \mathbb{R}^3$ denotes the extended coordinates of the corresponding desired image point on π in terms of $\mathcal{F}_d$ as

$$m_{de} \triangleq [\, m_{de1} \quad m_{de2} \quad m_{de3}\,]^T = \left[\frac{X_{d1}}{Z_1^*} \quad \frac{Y_{d1}}{Z_1^*} \quad \ln\left(Z_1^*\right)\right]^T \tag{51}$$

where ln $(\cdot)$ denotes the natural logarithm. Substituting (50) and (51) into (49) yields

$$e_v = \left[\frac{X_1}{Z_1} - \frac{X_{d1}}{Z_1^*} \quad \frac{Y_1}{Z_1} - \frac{Y_{d1}}{Z_1^*} \quad \ln\left(\frac{Z_1}{Z_1^*}\right)\right]^T \tag{52}$$

where the ratio $\frac{Z_1}{Z_1^*}$ can be computed from (17) and the decomposition of the estimated Euclidean homography in (22). Since $m_1(t)$ and m_d are unknown (since the intrinsic calibration matrices are unknown), $e_v(t)$ is not measurable. Therefore, the estimate of the translation error system given in (52) is defined as

$$\hat{e}_v \triangleq \left[\hat{m}_{e1} - \hat{m}_{de1} \quad \hat{m}_{e2} - \hat{m}_{de2} \quad \ln\left(\frac{Z_1}{Z_1^*}\right)\right]^T \tag{53}$$

where $\hat{m}_{e1}(t)$, $\hat{m}_{e2}(t)$, $\hat{m}_{de1}$, $\hat{m}_{de2} \in \mathbb{R}$ denote estimates of $m_{e1}(t)$, $m_{e2}(t)$, m_{de1}, m_{de2}, respectively.

To develop the closed-loop error system for $e_v(t)$, we take the time derivative of (52) and then substitute (42) into the resulting expression for $\omega_r(t)$ to obtain

$$\dot{e}_v = L_v R_r v_r + \lambda_w (L_v [t_r]_\times + L_{v\omega}) \widetilde{R}_r \hat{e}_\omega \tag{54}$$

where $L_v(t), L_{v\omega}(t) \in \mathbb{R}^{3x3}$ are defined as

$$L_v \triangleq \frac{1}{Z_1}\begin{bmatrix} -1 & 0 & m_{e1} \\ 0 & -1 & m_{e2} \\ 0 & 0 & -1 \end{bmatrix} \tag{55}$$

$$L_{v\omega} \triangleq \begin{bmatrix} m_{e1}m_{e2} & -1 - m_{e1}^2 & m_{e2} \\ 1 + m_{e2}^2 & -m_{e1}m_{e2} & -m_{e1} \\ -m_{e2} & m_{e1} & 0 \end{bmatrix}. \tag{56}$$

To facilitate the control development, the unknown depth $Z_1(t)$ in (55) can be expressed as

$$Z_1 = \frac{1}{\alpha_1} Z_1^* \tag{57}$$

where α_1 is given by the homography decomposition.

An estimate for $L_v(t)$ can be designed as

$$\widehat{L}_v = \frac{1}{\widehat{Z}_1}\begin{bmatrix} -1 & 0 & \hat{m}_{e1} \\ 0 & -1 & \hat{m}_{e2} \\ 0 & 0 & -1 \end{bmatrix}, \tag{58}$$

where $\hat{m}_{e1}(t), \hat{m}_{e2}(t)$ were introduced in (53), and $\widehat{Z}_1(t) \in \mathbb{R}$ is developed based on (57) as

$$\widehat{Z}_1 = \frac{1}{\alpha_1}\widehat{Z}_1^*. \tag{59}$$

Remark 2. In (59), $\widehat{Z}_1^*$ is an estimate of the constant unknown depth Z_1^* and is assumed to be known with an uncertainty. However, the constant parameter Z_1^* can also be estimated online in a direct adaptive framework [16].

Based on the structure of an error system in (54) and subsequent stability analysis, the following composite translation controller can be developed

[2] 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*.

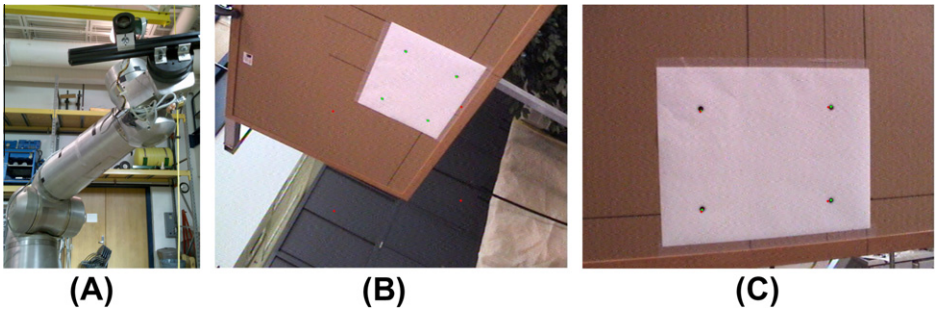

Fig. 2. (A) Robotics Research K-1607 manipulator in a camera-in-hand configuration; (B) initial and (C) final position of the feature points as viewed by the camera-in-hand (green points) and the overlaid desired feature points captured by the fixed camera (red points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

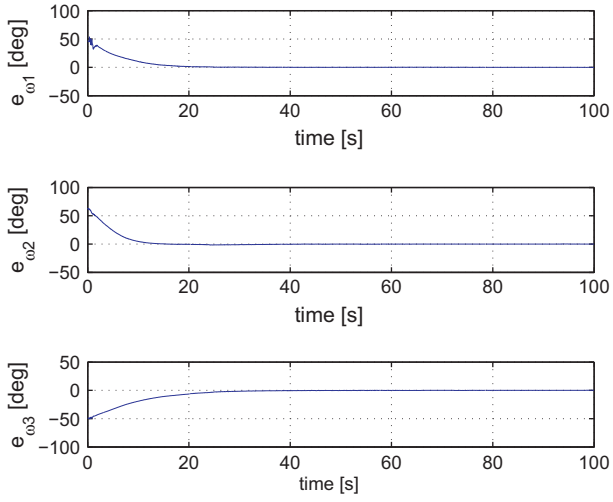

Fig. 3. Rotation error plot indicating angular error about x-axis ($e_{\omega 1}(t)$), y-axis ($e_{\omega 2}(t)$), and z-axis ($e_{\omega 3}(t)$) of the camera coordinate frame.

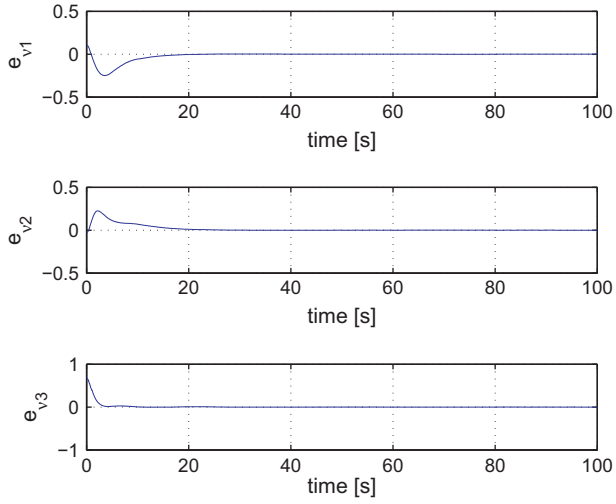

Fig. 4. Linear error plot indicating discrepancy between the extended normalized Euclidean coordinates of the target between the current and desired configurations.

$$v_r(t) = -\lambda_v \widehat{R}_r^T \widehat{L}_v^T \hat{e}_v - \left(k_{n1}\widehat{Z}_1^2 + k_{n2}\widehat{Z}_1^2 \|\hat{e}_v\|^2\right) \widehat{R}_r^T \widehat{L}_v^T \hat{e}_v \tag{60}$$

where $\widehat{R}_r^T$, $\hat{e}_v(t)$, and $\widehat{L}_v(t)$ are introduced in (42), (53), and (58), respectively, k_{n1}, $k_{n2} \in \mathbb{R}$ denote positive constant control gains,

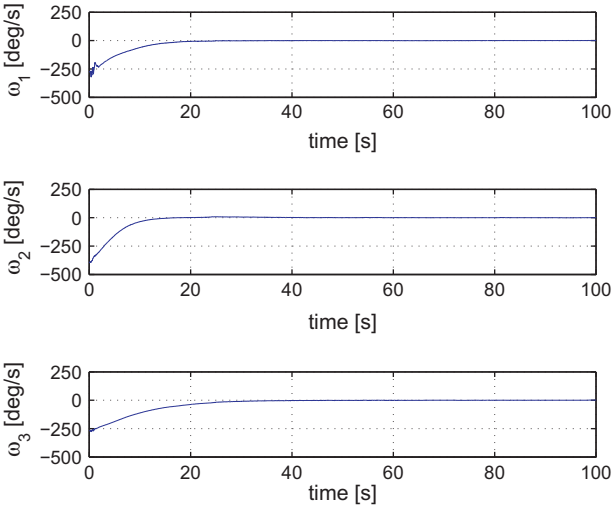

Fig. 5. Rotation velocity control input about x-axis ($\omega_1(t)$), y-axis ($\omega_2(t)$), and z-axis ($\omega_3(t)$) of the camera coordinate frame.

and $\widehat{Z}_1(t)$ is defined in (59). In (60), $\lambda_v(t) \in \mathbb{R}$ denotes a positive gain function defined as

$$\lambda_v = k_{n0} + \frac{\widehat{Z}_1^2}{f(\hat{m}_{e1}, \hat{m}_{e2})} \tag{61}$$

where $k_{n0} \in \mathbb{R}$ is a positive constant, and $f(\hat{m}_{e1}, \hat{m}_{e2})$ is a positive function of $\hat{m}_{e1}$ and $\hat{m}_{e2}$.

Property 2. *The kinematic control input given in (60) ensures that the composite translation error signal $e_v(t)$ defined in (52) is exponentially regulated in the sense that [14]*

$$\|e_v(t)\| \leqslant \sqrt{2\zeta_0}\|B^{-1}\| \exp\left(-\frac{\zeta_1}{2}t\right) \tag{62}$$

provided (46) is satisfied, where $B \in \mathbb{R}^{3\times 3}$ is a constant invertible matrix, and ζ_0, $\zeta_1 \in \mathbb{R}$ denote positive constants.

6. Experimental results

An experiment was performed using a Robotics Research K-1607 7-DOF robotic manipulator, as shown in Fig. 2A, to demonstrate the performance of the TBZ controller given in (42) and (60). The robot end-effector was equipped with a SONY CCD block camera and a mvBlueFox CCD camera fitted with a variable focal length lens served as a fixed camera. Location of the camera-in-

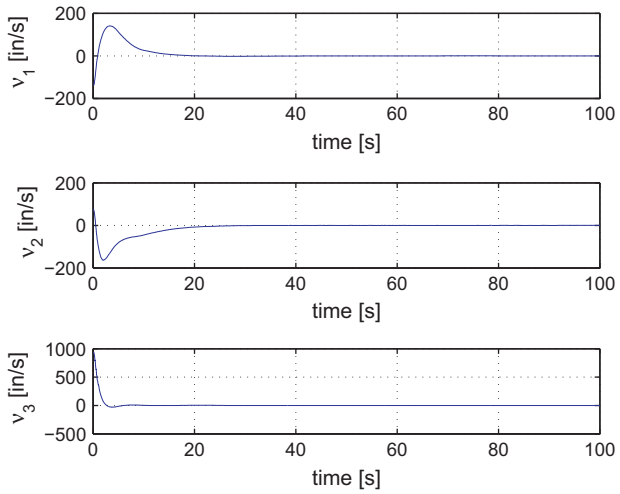

Fig. 6. Linear velocity control input along x-axis ($v_1(t)$), y-axis ($v_2(t)$), and z-axis ($v_3(t)$) of the camera coordinate frame.

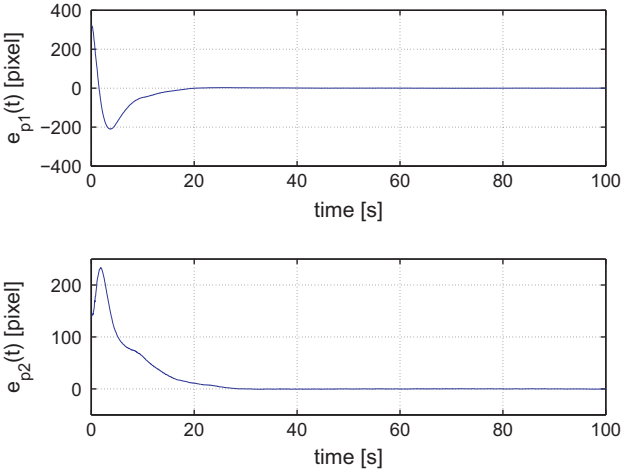

Fig. 7. Image space error $e_p(t)$ showing the difference in pixel coordinates between the current and reference images along the image x- and y-axes.

hand and fixed camera was selected such that a set of four coplanar features can be viewed by both the cameras. The focal length of the fixed camera was varied (zoomed in) to obtain the feature points p_i^* corresponding to the desired position/orientation, while a pyramidal implementation of the Lucas Kanade (KLT) feature tracker provided the time-varying feature points $p_i(t)$ for the camera-in-hand.

In an event where the camera-in-hand can be located *a priori* to the desired position/orientation of a robotic manipulator, the feature points p_i^* as well as $p_i(t)$ are obtained by the camera-in-hand and the TBZ control law presented in (42) and (60) becomes identical to TBS controller.

Caltech camera calibration toolbox for Matlab was used to obtain the estimates of intrinsic calibration parameters for both the cameras. The principal point image coordinates for the camera-in-hand and fixed camera are considered to be $u_0 = 345$, $v_0 = 245$ and $u_{0f} = 302$, $v_{0f} = 272$, respectively; $\lambda_1 = 868$, $\lambda_2 = 878$, $\lambda_{f1} = 571$, $\lambda_{f2} = 571$, $\lambda_1^* = 1596$, $\lambda_2^* = 1597$ denote the product of focal length and scaling factors for an on-board camera, fixed camera, and fixed camera after zooming, respectively; and $\phi = \phi_f = 1.53$ (rad) is the skew angle for each camera. The intrinsic parameters λ_1^* and λ_2^*, corresponding to the zoomed-in position of the lens, were computed to evaluate the correctness of servo control and were not used in the control formulation. The constant rotation and translation between the camera-in-hand and end-effector frame, i.e., the extrinsic camera calibration parameters $\widehat{R}_r$ and $\hat{t}_r$ defined in (31), are measured to be

$$\widehat{R}_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.9974 & 0.0724 \\ 0 & -0.0724 & 0.9974 \end{bmatrix}, \tag{63}$$

$$\hat{t}_r = [\,-100.3 \quad 48.0 \quad -48.0\,]^T. \tag{64}$$

The estimates $\widehat{R}_r$ and $\hat{t}_r$ are obtained using the commercial grade digital level and laser range finder, respectively, within the accuracy of the sensors. Since the z-axis of camera frame $\mathcal{F}$ is parallel to the x-axis of robot end-effector frame there exists a rotation $R_x(\xi)$ about the x-axis of end-effector frame by an angle $\xi \in \mathbb{R}$ that will align camera frame to the end-effector frame. The extrinsic calibration matrix $\widehat{R}_r \in \mathbb{R}^{3\times 3}$ in (63) is then obtained as $R_x(\xi)$ by measuring the constant angle ξ. The translation estimate $\hat{t}_r$ given in (64) is obtained by measuring the components of position vector corresponding to the origin of the camera coordinate frame $\mathcal{F}$ in the end-effector frame using a laser range finder.

The control objective is to regulate the camera-in-hand to the position/orientation of the virtual camera coordinate frame representing the zoomed in position/orientation of the fixed camera. The control gains k_{n0}, k_{n1}, k_{n2}, and λ_w were adjusted to the following values to yield the best performance

$$k_{n0} = 40 \quad k_{n1} = 26.6 \quad k_{n2} = 26.6 \quad \lambda_w = 6.0. \tag{65}$$

The feature points viewed by the camera-in-hand before and at the end of the servo control are shown in Figs. 2B and 2C, respectively, along with the location of desired feature points captured by the fixed camera. The resulting rotation and unitless translation errors are depicted in Figs. 3 and 4, respectively. The angular and linear control input velocities $\omega_r(t)$ and $v_r(t)$ defined in (42) and (60), respectively, are shown in Figs. 5 and 6. It can be observed from Figs. 3 and 4 that the rotation and translation error between camera coordinate frames $\mathcal{F}$ and $\mathcal{F}^*$ vanishes exponentially, thus regulating the camera-in-hand to the position/orientation of the virtual camera coordinate frame representing the zoomed in position/orientation of the fixed camera; Figs. 5 and 6 show that the linear and angular velocity control inputs remain bounded during the closed-loop operation. The image-space error $e_p(t) \in \mathbb{R}^3$ between the desired and current image coordinates of the target point O_1 is defined as

$$e_p(t) = [\,e_{p1} \quad e_{p2} \quad e_{p3}\,]^T = p_1(t) - p_1^*. \tag{66}$$

Fig. 7 shows the plot of the image error $e_p(t)$ defined in (66) to demonstrate the regulation result in an image-space.

7. Conclusion

A unified visual servo control approach – *teach by zooming* – is presented to address the control problem in applications where the camera cannot be *a priori* positioned to the desired position/orientation to acquire a reference image. Specifically, the TBZ control objective is formulated to position/orient an on-board camera based on a reference image obtained by another camera. In addition to formulating the TBZ control problem, another contribution of this paper is to illustrate how to preserve a symmetric transformation from the projective homography to the Euclidean homography for problems when the corresponding images are taken from different cameras with calibration uncertainty. To this end, a desired camera position/orientation is defined where the images correspond, but the Euclidean position differs as a function of the mismatch in the calibration of the cameras. Applications of this

strategy could include navigating ground or air vehicles based on the desired images taken by other ground or air vehicles (e.g., a satellite captures a "zoomed in desired image that is used to navigate an unmanned aerial vehicle (UAV), a camera can view the entire tree canopy and zoom in to acquire a desired image of a fruit product for high speed robotic harvesting). Experimental results are provided to demonstrate the performance of the developed controller.

References

[1] Corke P. Visual control of robot manipulators – a review. World scientific series in robotics and automated systems, vol. 7. World Scientific Press; 1993.
[2] Hashimoto K. A review on vision-based control of robot manipulators. Adv Robot 2003;17(10):969–91.
[3] Malis E, Chaumette F, Bodel S. 2-1/2D visual servoing. IEEE Trans Robot Automat 1999;15(2):238–50.
[4] Hutchinson S, Hager G, Corke P. A tutorial on visual servo control. IEEE Trans Robot Automat 1996;12(5):651–70.
[5] Malis E. Visual servoing invariant to changes in camera-intrinsic parameters. IEEE Trans Robot Automat 2004;20(1):72–81.
[6] Malis M. Vision-based control using different cameras for learning the reference image and for servoing. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, vol. 3; 2001. p. 1428–33.
[7] Dixon W. Teach by zooming: a camera independent alternative to teach by showing visual servo control. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, vol. 1; 2003. p. 749–54.
[8] Mehta S, Dixon W, Burks T, Gupta S. Teach by zooming visual servo control for an uncalibrated camera system. In: Proceedings of the AIAA guidance, navigation and control conference and exhibit; 2005. p. AIAA–2005–6095.
[9] Espiau B. Effect of camera calibration errors on visual servoing in robotics. In: 3rd International symposium on experimental robotics; 1993. p. 187–93.
[10] Chaumette F, Hutchinson S. Visual servo control: i – basic approaches. IEEE Robot Automat Mag 2006;13(4):82–90.
[11] Chaumette F. Potential problems of stability and convergence in image-based and position-based visual servoing. In: Kriegman D, Hager G, Morse A, editors. The confluence of vision and control. Lecture notes in control and information sciences, vol. 237. Berlin/Heidelberg: Springer; 1998. p. 66–78.
[12] Dixon W, Love L. Lyapunov-based visual servo control for robotic deactivation and decommissioning. In: The 9th biennial ANS international spectrum conference; 2002.
[13] Dixon W, Zergeroglu E, Fang Y, Dawson D. Object tracking by a robot manipulator: a robust cooperative visual servoing approach. In: Proceedings of the IEEE international conference on robotics and automation, vol. 1; 2002. p. 211–6.
[14] Fang Y, Dixon W, Dawson D, Chen J. An exponential class of model-free visual servoing controllers in the presence of uncertain camera calibration. Int J Robot Automat 2006;21(4):247–55.
[15] Malis E, Chaumette F. Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods. IEEE Trans Robot Automat 2002;18(2):176–86.
[16] Fang Y, Behal A, Dixon W, Dawson D. Adaptive 2.5D visual servoing of kinematically redundant robot manipulators. In: Proceedings of the 41st IEEE conference on decision and control, vol. 3; 2002. p. 2860–5.
[17] Faugeras O, Lustman F. Motion and structure from motion in a piecewise planar environment. Int J Pattern Recogn Artif Int 1988;2(3):485–508.
[18] Faugeras O. The geometry of multiple images. MIT Press; 2001.
[19] Zhang Z, Hanson A. Scaled euclidean 3D reconstruction based on externally uncalibrated cameras. In: Proceedings of the international symposium on computer vision; 1995. p. 37–42.