Homography-Based Visual Servoing of Eye-in-Hand Robots With Exact Depth Estimation

Beixian Lai, Zhiwen Li, Weibing Li, *Member, IEEE*, Chenguang Yang, *Senior Member, IEEE*, and Yongping Pan, *Senior Member, IEEE*

***Abstract*—Visual servoing can effectively control robots using visual feedback to improve their intelligence and reliability. For a feature point detected by a monocular camera, the time-varying depth appearing nonlinearly in the Jacobian matrix is difficult to be measured without the prior geometry knowledge of the observed object. Therefore, the depth of the feature point is one of the major uncertain parameters in visual servoing. Considering unknown Cartesian feature positions, this article presents a robot dynamics-based homography-based visual servoing (HBVS) controller for the 3-D pose regulation of eye-in-hand robot arms with monocular cameras. The uncertain depth is represented into a linear form of its Cartesian feature position, and a composite learning law is applied to estimate position parameters accurately, resulting in exact depth estimation. Compared to existing adaptive HBVS methods, the distinctive feature of the proposed method is that it is a dynamics-based design and guarantees exact depth estimation under a much weaker condition termed interval excitation compared to persistent excitation. Simulations and experiments on a collaborative robot with seven degrees of freedom named Franka Emika Panda have verified the effectiveness of the proposed method.**

***Index Terms*—Collaborative robot, composite learning, parameter convergence, unknown feature position.**

I. Introduction

VISUAL servoing is an effective methodology for controlling robots using visual feedback to improve their intelligence and reliability. There exist two basic categories of visual servoing approaches [1]: position-based visual servoing (PBVS) and image-based visual servoing (IBVS). PBVS reconstructs the Cartesian pose (i.e., position and orientation) of the observed object based on the object's visual projection information and accurate geometry information to achieve Cartesian pose control and can specify control tasks directly in the Cartesian space [2]. However, it excludes image information from the feedback loop, and thus, the object is hard to be kept within the field of view (FOV) of the camera, possibly leading to servoing failures [3]. IBVS directly controls the robot in the image space but has some drawbacks, such as the singularity of image Jacobian, local minima, and local stability [4]. Besides, it is more difficult to describe control tasks using image features in IBVS compared to PBVS [5], i.e., Cartesian trajectories may be unsatisfied or physically impossible despite image features converging to their desired locations [6]. Despite these drawbacks, IBVS is more likely to maintain the object within the FOV and is less sensitive to image noise and camera calibration errors than PBVS [2]. A hybrid approach known as homography-based visual servoing (HBVS) can inherit the advantages of both PBVS and IBVS while avoiding their disadvantages [7]. The control objective of HBVS is divided into two parts: one in the 3-D Cartesian space and the other in the 2-D image space so that it is convenient to achieve robot pose control.

Various parameter uncertainties exist in robot visual servo systems [8], where the time-varying depth of the feature point on the observed object is one of the major uncertain parameters which appears nonlinearly in the image Jacobian matrix and is difficult to be measured without the prior geometry knowledge of the object for a monocular camera. Adaptive control is effective in handling parameter uncertainties. Numerous adaptive IBVS [9], [10], [11], [12], [13], [14], [15], [16], [17] and HBVS [18], [19], [20], [21], [22], [23] methods have been developed to handle uncertain parameters in robots, e.g., feature position parameters, camera intrinsic and extrinsic parameters, and robot kinematic and dynamic parameters. In the above adaptive IBVS methods, the depth of the feature point can be represented as a function of a set of unknown parameters such that it is obtainable if parameter estimates converge to their true values. A major drawback of these methods is that the strict condition termed persistent excitation (PE) for parameter convergence is difficult to be fulfilled in practice [23]. Some of the above adaptive HBVS methods, such as [19] and [21], estimate Cartesian position

Manuscript received 10 October 2022; revised 27 February 2023; accepted 3 May 2023. Date of publication 22 May 2023; date of current version 27 October 2023. This work was supported in part by the Fundamental Research Funds for the Central Universities, Sun Yat-sen University, China, under Grant 23lgzy004, and in part by the Guangdong Provincial Pearl River Talents Program of China under Grant 2019QN01X154. *(Corresponding author: Yongping Pan.)*

Yongping Pan is with the School of Advanced Manufacturing, Sun Yat-sen University, Shenzhen 518100, China (e-mail: panyongp@mail.sysu.edu.cn).

Beixian Lai, Zhiwen Li, and Weibing Li are with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China (e-mail: laibx@mail2.sysu.edu.cn; lizhw63@mail2.sysu.edu.cn; liwb53@mail.sysu.edu.cn).

Chenguang Yang is with the College of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: cyang@ieee.org).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TIE.2023.3277072.

Digital Object Identifier 10.1109/TIE.2023.3277072

and depth parameters jointly such that PE is still needed for parameter convergence, and the rest of them estimate the depth directly so that it is easier to achieve exact depth estimation as PE holds easily for estimation with only one parameter.

In the above adaptive IBVS methods, pixel error convergence can be achieved only when the number of degrees of freedom (DoFs) for robots is not less than $2N$ [9], and N is not more than three to satisfy a full-rank condition [16], in which N is the number of feature points. Therefore, under the convergence conditions mentioned above, these adaptive IBVS methods can only theoretically achieve convergence of pixel errors with no more than three feature points. Yet, only three feature points cannot define a unique 3-D robot pose in visual servoing [24], whereas four feature points on the plane are sufficient to define its unique pose [2], [6]. Simulations and experiments of the aforesaid adaptive IBVS methods were carried out on robots with no more than three effective DoFs. Adaptive HBVS provides a promising decoupling approach to achieve robot pose control under parameter uncertainties without resorting to the above impractical condition. However, existing adaptive HBVS methods [18], [19], [20], [21], [22], [23] are only based on the robot kinematics with an assumption that the robot velocity is controlled precisely while the robot dynamics is negligible. It has been shown that the robot dynamics is critical for control accuracy and system stability in visual servoing [13], [25] such that it should be considered in visual servoing to guarantee the control performance [14].

This article focuses on the 3-D visual regulation problem of eye-in-hand (EIH) robots with monocular cameras. If Cartesian feature positions must be known in advance (as in some adaptive IBVS methods with eye-to-hand (ETH) cameras [9], [12], [13]), directly measuring these positions is needed before executing control tasks, which is tedious and time-consuming. Considering unknown Cartesian feature positions is beneficial to the arbitrary selection of feature points [14]. Since the depth can be expressed by Cartesian feature positions, a unified adaptive law can be designed to estimate feature position and depth parameters simultaneously. The problem mentioned above is suitable for the scenario where the robot end-effector manipulates (under various desired poses) a static or slow-moving planar object with an unknown Cartesian pose. We aim to address the following major problems. 1) Some dynamics-based IBVS methods cannot be theoretically proven their effectiveness for 3-D robot pose control. 2) The assumption of accurate robot velocity control in kinematics-based HBVS does not generally hold in real-world environments [26]. 3) The PE condition is difficult to be fulfilled in practice to achieve exact depth estimation.

In this study, a dynamics-based composite learning HBVS (CL-HBVS) method is proposed to address the aforementioned problems. The design procedure of the proposed method is as follows. First, a homography decomposition based on Euclidean reconstruction is presented to extract the depth ratio and rotation matrix of image features that are helpful in forming the pose of the end-effector; second, the time-varying depth of the feature point is represented as a linear form of its unknown position parameters in the Cartesian space to facilitate depth estimation; third, a dynamics-based HBVS controller is designed to regulate the end-effector to its desired one; fourth, a composite learning law that combines joint velocity and Cartesian prediction errors is proposed to guarantee position parameter convergence under a condition of interval excitation (IE) that relaxes the stringent PE condition. Our main contributions lie in the following.

1) The novel dynamics-based CL-HBVS control law is proposed to achieve 3-D pose control of EIH robots without the restricted conditions in [9] and [16].
2) An exact estimation of the time-varying depth is achieved when the unknown Cartesian position parameters converge to their true values, which is helpful in reconstructing the scene of the planar object online.
3) Experiments on a collaborative robot with 7 DoFs are given to show the effectiveness of the proposed method.

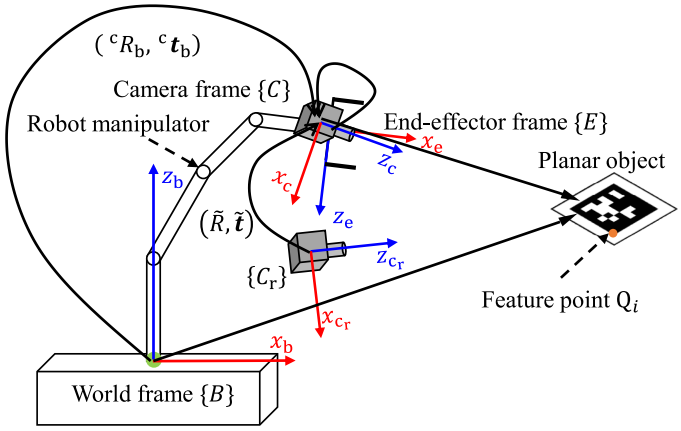


Fig. 1. Articulated robot with an EIH monocular camera for visual servoing, where a planar object is fixed nearby the world frame $\{B\}$.

Throughout this article, $\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}^n$, and $\mathbb{R}^{m\times n}$ denote the spaces of real numbers, positive real numbers, real n-vectors, and real $m \times n$-matrices, respectively, $\mathbb{N}$ is the set of natural numbers, $\max(\cdot)$ is the maximum operator, $\|\boldsymbol{x}\|$ is the Euclidean norm of $\boldsymbol{x} \in \mathbb{R}^n$, $\mathrm{diag}(x_1, x_2, \ldots, x_n)$ is a diagonal matrix with elements $x_i \in \mathbb{R}$ and $i = 1$ to n, $\arccos(x_i)$ is the arc-cosine function of x_i, $\ln(x_i)$ is the natural logarithm of x_i, $\mathrm{rank}(A)$ is the rank of $A \in \mathbb{R}^{m\times n}$, $\det(B)$ is the determinant of $B \in \mathbb{R}^{n\times n}$, $\mathrm{tr}(B)$ is the trace of B, where $i, m, n \in \mathbb{N}$.

II. Geometric Model With EIH Camera

A. Camera Projection and Euclidean Reconstruction

An articulated robotic system with an EIH camera attached to the end-effector is shown in Fig. 1 to illustrate the Euclidean relationships among a world frame $\{B\}$, a camera frame $\{C\}$, and an end-effector frame $\{E\}$. It is assumed that: 1) a planar object fixed in the environment is always captured by the camera; 2) there are $N \geq 4$ ($\in \mathbb{N}$) noncollinear feature points Q_i ($i =$ 1 to N) lying on the object. A stationary camera frame $\{C_r\}$ in Fig. 1 is used to represent a constant reference pose that provides a reference image of the planar object for $\{C\}$.

Let $\tilde{R}(t) \in \mathbb{R}^{3\times3}$ be a mismatch rotation matrix and $\tilde{\boldsymbol{t}}(t) \in \mathbb{R}^3$ be a translation vector between $\{C\}$ and $\{C_r\}$, and ${}^{c}\boldsymbol{x}_i(t) = [x_i(t), y_i(t), z_i(t)]^T \in \mathbb{R}^3$ and ${}^{c}\boldsymbol{x}_{ri} = [x_{ri}, y_{ri}, z_{ri}]^T \in \mathbb{R}^3$

be Euclidean coordinates of Q_i in $\{C\}$ and $\{C_\mathrm{r}\}$, respectively, where $z_i(t)$, $z_{\mathrm{r}i} \in \mathbb{R}^+$ are the depths of Q_i in $\{C\}$ and $\{C_\mathrm{r}\}$, respectively. Then, the normalized Euclidean coordinates of Q_i in $\{C\}$ and $\{C_\mathrm{r}\}$ can be defined by

$$\boldsymbol{x}_i(t) := {}^{\mathrm{c}}\boldsymbol{x}_i/z_i,\ \boldsymbol{x}_{\mathrm{r}i} := {}^{\mathrm{c}}\boldsymbol{x}_{\mathrm{r}i}/z_{\mathrm{r}i} \tag{1}$$

respectively. From Fig. 1, the Euclidean relationship between $\boldsymbol{x}_i$ and $\boldsymbol{x}_{\mathrm{r}i}$ can be obtained as follows [20]:

$$\boldsymbol{x}_i = \underbrace{(z_{\mathrm{r}i}/z_i)}_{\alpha_i}\overbrace{(\tilde{R} + \underbrace{(\tilde{\boldsymbol{t}}/d_\mathrm{r})}_{\tilde{\boldsymbol{t}}_\mathrm{d}}\boldsymbol{n}_\mathrm{r}^T)}^{H}\boldsymbol{x}_{\mathrm{r}i} \tag{2}$$

where $\alpha_i(t) \in \mathbb{R}^+$ is a depth ratio, $H(t) \in \mathbb{R}^{3\times3}$ is an Euclidean homography matrix, $\boldsymbol{n}_\mathrm{r} \in \mathbb{R}^3$ is a constant unit normal of the object in $\{C_\mathrm{r}\}$, $d_\mathrm{r} \in \mathbb{R}^+$ is a constant distance from $\{C_\mathrm{r}\}$ to the object along $\boldsymbol{n}_\mathrm{r}$, and $\tilde{\boldsymbol{t}}_\mathrm{d}(t) \in \mathbb{R}^3$ is a scaled translation vector. Let $\boldsymbol{p}_i(t) = [u_i(t), v_i(t), 1]^T \in \mathbb{R}^3$ and $\boldsymbol{p}_{\mathrm{r}i} = [u_{\mathrm{r}i}, v_{\mathrm{r}i}, 1]^T \in \mathbb{R}^3$ be homogeneous pixel coordinates of Q_i projected to the image plane in $\{C\}$ and $\{C_\mathrm{r}\}$, respectively. The following relationships can be obtained by a pinhole camera model [27]:

$$\boldsymbol{p}_i = K\boldsymbol{x}_i,\ \boldsymbol{p}_{\mathrm{r}i} = K\boldsymbol{x}_{\mathrm{r}i} \tag{3}$$

where $K \in \mathbb{R}^{3\times3}$ denotes an intrinsic parameter of the camera. Using (2) and (3), one obtains [18]

$$\boldsymbol{p}_i = \alpha_i \underbrace{KH(t)K^{-1}}_{G}\boldsymbol{p}_{\mathrm{r}i} \tag{4}$$

where $G(t) \in \mathbb{R}^{3\times3}$ is a projective homography matrix.

Given four pairs of noncollinear correspondence points ($\boldsymbol{p}_i$, $\boldsymbol{p}_{\mathrm{r}i}$) ($i = 1$ to 4), the mismatch rotation matrix $\tilde{R}$, scaled translation vector $\tilde{\boldsymbol{t}}_\mathrm{d}$, unit normal $\boldsymbol{n}_\mathrm{r}$ and depth ratio α_i can be extracted via homography decomposition [28]. If four coplanar feature points are unavailable, the above development can exploit a virtual parallax method [23] that requires at least eight feature points in any position where any four of them are not coplanar.

B. Linear Parameterization of Camera Models

Let ${}^{\mathrm{c}}R_\mathrm{b}(t) \in \mathbb{R}^{3\times3}$ and ${}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b}(t) \in \mathbb{R}^3$ be the orientation and position of the world frame $\{B\}$ expressed in the camera frame $\{C\}$, respectively, and ${}^{\mathrm{e}}R_\mathrm{b}(t) \in \mathbb{R}^{3\times3}$ and ${}^{\mathrm{e}}\boldsymbol{t}_\mathrm{b}(t) \in \mathbb{R}^3$ be the orientation and position of $\{B\}$ in the end-effector frame $\{E\}$, respectively (i.e., robot forward kinematics). It is assumed that the camera's extrinsic parameters ${}^{\mathrm{c}}R_\mathrm{e} \in \mathbb{R}^{3\times3}$ and ${}^{\mathrm{c}}\boldsymbol{t}_\mathrm{e} \in \mathbb{R}^3$, which represent the orientation and position of $\{E\}$ in $\{C\}$, respectively, are available from offline calibration. Then, the kinematic relationships ${}^{\mathrm{c}}R_\mathrm{b}$ and ${}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b}$ are obtainable by ${}^{\mathrm{c}}R_\mathrm{b}(t) = {}^{\mathrm{c}}R_\mathrm{e}\,{}^{\mathrm{e}}R_\mathrm{b}(t)$ and ${}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b} = {}^{\mathrm{c}}R_\mathrm{e}\,{}^{\mathrm{e}}\boldsymbol{t}_\mathrm{b} + {}^{\mathrm{c}}\boldsymbol{t}_\mathrm{e}$. Since ${}^{\mathrm{c}}R_\mathrm{e}$ and ${}^{\mathrm{c}}\boldsymbol{t}_\mathrm{e}$ are constant for EIH robots, the above assumption is reasonable and simplifies the adaptive HBVS design.

Let ${}^{\mathrm{b}}\boldsymbol{x}_i \in \mathbb{R}^3$ denote a constant Euclidean coordinate of the feature point Q_i in $\{B\}$. Afterward, the camera projection model (3) can be modified into [11]

$$\boldsymbol{p}_i = K\,{}^{\mathrm{c}}\boldsymbol{x}_i/z_i \tag{5}$$

where the Euclidean coordinate ${}^{\mathrm{c}}\boldsymbol{x}_i$ is given by

$${}^{\mathrm{c}}\boldsymbol{x}_i = {}^{\mathrm{c}}R_\mathrm{b}\,{}^{\mathrm{b}}\boldsymbol{x}_i + {}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b}. \tag{6}$$

Then, the depth z_i is expressed by

$$z_i = \boldsymbol{k}_3^T({}^{\mathrm{c}}R_\mathrm{b}\,{}^{\mathrm{b}}\boldsymbol{x}_i + {}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b}) \tag{7}$$

where $\boldsymbol{k}_i^T \in \mathbb{R}^3$ is the ith row of K. Let ${}^{\mathrm{b}}\hat{\boldsymbol{x}}_i \in \mathbb{R}^3$ be an estimate of ${}^{\mathrm{b}}\boldsymbol{x}_i$. From (7), a depth estimate $\hat{z}_i \in \mathbb{R}$ can be obtained by

$$\hat{z}_i = \boldsymbol{k}_3^T({}^{\mathrm{c}}R_\mathrm{b}\,{}^{\mathrm{b}}\hat{\boldsymbol{x}}_i + {}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b}). \tag{8}$$

A linearly parameterized formula of (5), which is crucial for adaptive design and stability analysis, is given as follows.

Property 1: It is assumed that ${}^{\mathrm{b}}\boldsymbol{x}_i$, the position of the point Q_i, is unknown but constant. Multiplying each side of (5) by z_i and substituting (7) into the result, one gets the camera projection model (5) in a linear form of ${}^{\mathrm{b}}\boldsymbol{x}_i$ as follows:

$$\boldsymbol{y}_i(\boldsymbol{p}_i) = \Phi_i^T(\boldsymbol{p}_i)\,{}^{\mathrm{b}}\boldsymbol{x}_i \tag{9}$$

where $\Phi_i(\boldsymbol{p}_i) := {}^{\mathrm{c}}R_\mathrm{b}^T K_\mathrm{c}^T(\boldsymbol{p}_i) \in \mathbb{R}^{3\times3}$ is a regressor independent of ${}^{\mathrm{b}}\boldsymbol{x}_i$, $\boldsymbol{y}_i(\boldsymbol{p}_i) := -K_\mathrm{c}(\boldsymbol{p}_i)\,{}^{\mathrm{c}}\boldsymbol{t}_\mathrm{b} \in \mathbb{R}^3$ is a certain vector, and $K_\mathrm{c}(\boldsymbol{p}_i) := [\boldsymbol{k}_1 - u_i\boldsymbol{k}_3\ \boldsymbol{k}_2 - v_i\boldsymbol{k}_3\ \boldsymbol{0}] \in \mathbb{R}^{3\times3}$.

III. Composite Learning Dynamic HBVS

This section aims to develop an adaptive HBVS controller to regulate the end-effector (or the camera) to its reference pose. In HBVS, the translation and rotation components of the end-effector pose can be controlled separately. More specifically, the mismatch rotation $\tilde{R}$ is utilized to design a rotation control part, while a combination of the pixel coordinate $\boldsymbol{p}_i$ and depth ratio α_i of a feature point Q_i is utilized to design a translation control part. Any Q_i can be used to form the translation component of the end-effector pose [19], and thus, Q_1 is chosen for the sake of simplicity, and the subscript 1 is omitted here. The pose control objective can be given by

$$\tilde{R} \to I, \alpha \to 1, \boldsymbol{p} \to \boldsymbol{p}_\mathrm{r} \text{ as } t \to \infty. \tag{10}$$

That is, if $\tilde{R} = I$ and $\alpha = 1$, one has $R = R_\mathrm{r}$ and $z = z_\mathrm{r}$. Then, with $z\boldsymbol{p} = K\,{}^{\mathrm{c}}\boldsymbol{x}$, if $\boldsymbol{p} = \boldsymbol{p}_\mathrm{r}$, one has ${}^{\mathrm{c}}\boldsymbol{x} = {}^{\mathrm{c}}\boldsymbol{x}_\mathrm{r}$, where ${}^{\mathrm{c}}\boldsymbol{x}$ and ${}^{\mathrm{c}}\boldsymbol{x}_\mathrm{r}$ are current and reference Euclidean coordinates of the feature point Q, respectively. When $R = R_\mathrm{r}$ and ${}^{\mathrm{c}}\boldsymbol{x} = {}^{\mathrm{c}}\boldsymbol{x}_\mathrm{r}$, the end-effector reaches its reference pose.

A. Pose Error Definition

To quantify the mismatch rotation $\tilde{R}$ in (2) between the frames $\{C\}$ and $\{C_\mathrm{r}\}$, a rotation error $\boldsymbol{e}_\omega \in \mathbb{R}^3$, i.e., an axis-angle representation of $\tilde{R}$, is defined by

$$\boldsymbol{e}_\omega(t) := \boldsymbol{\mu}(t)\phi(t) \tag{11}$$

where $\boldsymbol{\mu}(t) \in \mathbb{R}^3$ denotes the unit axis of rotation, and $\phi(t) \in \mathbb{R}$ denotes the rotation angle error about the axis $\boldsymbol{\mu}(t)$ that can be confined to the following region [19]:

$$-\pi < \phi(t) < \pi. \tag{12}$$

Both ϕ and $\boldsymbol{\mu}$ can be calculated by $\tilde{R}$ as follows [29]:

$$\begin{cases} \phi(t) = \arccos\left(\frac{1}{2}(\mathrm{tr}(\tilde{R}) - 1)\right) \\ [\boldsymbol{\mu}(t)]_\times = \frac{\tilde{R} - \tilde{R}^T}{2\sin(\phi)} \end{cases} \tag{13}$$

where $\tilde{R}$ is from homography decomposition as in Section II-A, and $[\cdot]_\times \in \mathbb{R}^{3\times 3}$ is a skew-symmetric operator given by

$$[\boldsymbol{\mu}]_\times := \begin{bmatrix} 0 & -\mu_3 & \mu_2 \\ \mu_3 & 0 & -\mu_1 \\ -\mu_2 & \mu_1 & 0 \end{bmatrix} \quad \forall \boldsymbol{\mu} := \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \in \mathbb{R}^3. \tag{14}$$

Differentiating $\boldsymbol{e}_\omega$ with respect to time t yields [19]

$$\dot{\boldsymbol{e}}_\omega = -L_\omega {}^{\mathrm{c}}R_{\mathrm{b}} {}^{\mathrm{b}}\boldsymbol{\omega} \tag{15}$$

in which ${}^{\mathrm{b}}\boldsymbol{\omega}(t) \in \mathbb{R}^3$ denotes an angular velocity of the end-effector expressed in the world frame $\{B\}$, and $L_\omega(t) \in \mathbb{R}^{3\times 3}$ denotes a Jacobian-like matrix defined by [7]

$$L_\omega(t) := I - \frac{\phi}{2}[\boldsymbol{\mu}]_\times + \left(1 - \frac{\mathrm{sinc}(\phi)}{\mathrm{sinc}^2(\frac{\phi}{2})}\right)[\boldsymbol{\mu}]_\times^2 \tag{16}$$

with $\mathrm{sinc}(\phi) := \sin(\phi)/\phi$ and $\mathrm{sinc}(0) = 0$. Then, one has

$$\det(L_\omega) = 1/\mathrm{sinc}^2(\phi/2) \tag{17}$$

which is singular only for $\phi = 2k\pi \quad \forall k \in \mathbb{N}, k \neq 0$.

To control the robot translation, define current and reference extended translation vectors $\boldsymbol{p}_{\mathrm{e}}(t), \boldsymbol{p}_{\mathrm{er}} \in \mathbb{R}^3$ by [7]

$$\begin{cases} \boldsymbol{p}_{\mathrm{e}}(t) := [u(t), v(t), \ln(z(t))]^T \\ \boldsymbol{p}_{\mathrm{er}} := [u_{\mathrm{r}}, v_{\mathrm{r}}, \ln(z_{\mathrm{r}})]^T \end{cases} \tag{18}$$

and an extended translation error $\boldsymbol{e}_{\mathrm{v}}(t) \in \mathbb{R}^3$ by

$$\boldsymbol{e}_{\mathrm{v}} := \boldsymbol{p}_{\mathrm{e}}(t) - \boldsymbol{p}_{\mathrm{er}} = [\tilde{u}, \tilde{v}, -\ln(\alpha)]^T \tag{19}$$

where $\tilde{u}(t) := u(t) - u_{\mathrm{r}}$ and $\tilde{v}(t) := v(t) - v_{\mathrm{r}}$ are pixel errors in the u- and v-axes of the image plane, respectively, and α is the depth ratio obtained by homography decomposition as in Section II-A. Using (5) and (7), one gets

$$\dot{\boldsymbol{e}}_{\mathrm{v}} = \frac{1}{z} A_{\mathrm{e}}(\boldsymbol{p}) \begin{bmatrix} {}^{\mathrm{b}}\boldsymbol{v} \\ {}^{\mathrm{b}}\boldsymbol{\omega} \end{bmatrix} \tag{20}$$

where ${}^{\mathrm{b}}\boldsymbol{v}(t) \in \mathbb{R}^3$ denotes a linear velocity of the end-effector expressed in the world frame $\{B\}$, and $A_{\mathrm{e}}(\boldsymbol{p}) \in \mathbb{R}^{3\times 6}$ denotes an extended interaction matrix given by

$$A_{\mathrm{e}}(\boldsymbol{p}) := \underbrace{(K - \boldsymbol{p}_0 \boldsymbol{k}_3^T) {}^{\mathrm{c}}R_{\mathrm{b}}}_{K_{\mathrm{e}}} \left[-I, [{}^{\mathrm{b}}\boldsymbol{x} - {}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{e}}]_\times\right] \tag{21}$$

with $\boldsymbol{p}_0(t) := [u(t), v(t), 0]^T \in \mathbb{R}^3$, in which ${}^{\mathrm{e}}\boldsymbol{t}_{\mathrm{b}} = {}^{\mathrm{e}}R_{\mathrm{b}} {}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{e}}$ is used to get the above result, ${}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{e}} \in \mathbb{R}^3$ is the origin position of the end-effector frame $\{E\}$ in $\{B\}$, $[\cdot]_\times$ is defined in (14), and $K_{\mathrm{e}}(t) \in \mathbb{R}^{3\times 3}$ is an invertible matrix (as K and ${}^{\mathrm{c}}R_{\mathrm{b}}$ are invertible). With the definitions of $\boldsymbol{e}_\omega$ in (11) and $\boldsymbol{e}_{\mathrm{v}}$ in (19), the pose control objective (10) becomes

$$\boldsymbol{e}_\omega \to \boldsymbol{0}, \boldsymbol{e}_{\mathrm{v}} \to \boldsymbol{0} \text{ as } t \to \infty. \tag{22}$$

B. Visual Servo Control Design

The dynamics of n-DoF robots is described by

$$M(\boldsymbol{q})\ddot{\boldsymbol{q}} + C(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{g}(\boldsymbol{q}) = \boldsymbol{\tau} \tag{23}$$

in which $\boldsymbol{q}(t) \in \mathbb{R}^n$ is a joint angular position, $M(\boldsymbol{q}) \in \mathbb{R}^{n\times n}$ is an inertial matrix, $C(\boldsymbol{q}, \dot{\boldsymbol{q}}) \in \mathbb{R}^{n\times n}$ is a centripetal-Coriolis matrix, $\boldsymbol{g}(\boldsymbol{q}) \in \mathbb{R}^n$ is a gravitational torque, and $\boldsymbol{\tau}(t) \in \mathbb{R}^n$ is a control torque. The following two significant properties of the robot dynamics are available for stability analysis [30].

Property 2: $M(\boldsymbol{q})$ is symmetric positive-definite and satisfies $m_0 I \leq M(\boldsymbol{q}) \leq m_1 I$ with some constants $m_0, m_1 \in \mathbb{R}^+$.

Property 3: $\dot{M}(\boldsymbol{q}) - 2C(\boldsymbol{q}, \dot{\boldsymbol{q}})$ is skew-symmetric such that $\boldsymbol{x}^T(\dot{M}(\boldsymbol{q}) - 2C(\boldsymbol{q}, \dot{\boldsymbol{q}}))\boldsymbol{x} = 0, \forall \boldsymbol{q}, \dot{\boldsymbol{q}}, \boldsymbol{x} \in \mathbb{R}^n$.

Let $J(\boldsymbol{q}) := \left[J_{\mathrm{v}}^T(\boldsymbol{q})\ J_\omega^T(\boldsymbol{q})\right]^T \in \mathbb{R}^{6\times n}$ be the robot Jacobian matrix, where $J_{\mathrm{v}}(\boldsymbol{q}) \in \mathbb{R}^{3\times n}$ is the translational Jacobian, and $J_\omega(\boldsymbol{q}) \in \mathbb{R}^{3\times n}$ is the rotational Jacobian. Then, the velocities of the end-effector can be calculated by

$${}^{\mathrm{b}}\boldsymbol{v} = J_{\mathrm{v}}(\boldsymbol{q})\dot{\boldsymbol{q}}, \ {}^{\mathrm{b}}\boldsymbol{\omega} = J_\omega(\boldsymbol{q})\dot{\boldsymbol{q}}. \tag{24}$$

To achieve the pose control objective, we propose a dynamics-based CL-HBVS control law as follows:

$$\begin{aligned} \boldsymbol{\tau} = \boldsymbol{g}(\boldsymbol{q}) - K_{\mathrm{d}}\dot{\boldsymbol{q}} - \alpha J^T(\boldsymbol{q})\hat{A}_{\mathrm{e}}^T(\boldsymbol{p})K_{\mathrm{p}}\boldsymbol{e}_{\mathrm{v}} \\ + J_\omega^T(\boldsymbol{q}) {}^{\mathrm{c}}R_{\mathrm{b}}^T L_\omega^T K_\omega \boldsymbol{e}_\omega \end{aligned} \tag{25}$$

where $\hat{A}_{\mathrm{e}}(\boldsymbol{p}) := A_{\mathrm{e}}(\boldsymbol{p})|_{{}^{\mathrm{b}}\boldsymbol{x} = {}^{\mathrm{b}}\hat{\boldsymbol{x}}} \in \mathbb{R}^{3\times 6}$ denotes an estimate of $A_{\mathrm{e}}(\boldsymbol{p})$ in (21), and $K_{\mathrm{d}} \in \mathbb{R}^{n\times n}$, $K_{\mathrm{p}} \in \mathbb{R}^{3\times 3}$ and $K_\omega \in \mathbb{R}^{3\times 3}$ are positive-definite diagonal matrices of control gains. Note that $\mathrm{rank}(\hat{A}_{\mathrm{e}}(\boldsymbol{p})) = 3$ is always true since the estimate ${}^{\mathrm{b}}\hat{\boldsymbol{x}}$ does not affect the rank of $\hat{A}_{\mathrm{e}}(\boldsymbol{p})$.

Property 4: With ${}^{\mathrm{b}}\boldsymbol{x}$ being unknown, for any given $\boldsymbol{\eta} \in \mathbb{R}^3$, $\alpha J^T(\boldsymbol{q})A_{\mathrm{e}}^T(\boldsymbol{p})\boldsymbol{\eta}$ can be linearly parameterized by

$$\alpha J^T(\boldsymbol{q})A_{\mathrm{e}}^T(\boldsymbol{p})\boldsymbol{\eta} = Y^T(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\eta}) {}^{\mathrm{b}}\boldsymbol{x} + \boldsymbol{\upsilon}(\boldsymbol{p}, \boldsymbol{q}) \tag{26}$$

where $\boldsymbol{\upsilon}(\boldsymbol{p}, \boldsymbol{q}) := -\alpha(J_{\mathrm{v}}^T - J_\omega^T[{}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{e}}]_\times)K_{\mathrm{e}}^T\boldsymbol{\eta} \in \mathbb{R}^6$ is a certain vector, $Y(\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\eta}) := -\alpha[K_{\mathrm{e}}^T\boldsymbol{\eta}]_\times J_\omega \in \mathbb{R}^{3\times 6}$ is a regressor matrix independent of ${}^{\mathrm{b}}\boldsymbol{x}$, and $[\cdot]_\times$ is defined in (14).

Applying (25) and *Property 4* to (23), one obtains the closed-loop dynamics as follows:

$$\begin{aligned} M(\boldsymbol{q})\ddot{\boldsymbol{q}} + (C(\boldsymbol{q}, \dot{\boldsymbol{q}}) + K_{\mathrm{d}})\dot{\boldsymbol{q}} = J_\omega^T(\boldsymbol{q}) {}^{\mathrm{c}}R_{\mathrm{b}}^T L_\omega^T K_\omega \boldsymbol{e}_\omega \\ - Y^T(\boldsymbol{p}, \boldsymbol{q}, K_{\mathrm{p}}\boldsymbol{e}_{\mathrm{v}}) {}^{\mathrm{b}}\tilde{\boldsymbol{x}} - \alpha J^T(\boldsymbol{q})A_{\mathrm{e}}^T(\boldsymbol{p})K_{\mathrm{p}}\boldsymbol{e}_{\mathrm{v}} \end{aligned} \tag{27}$$

where ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}(t) := {}^{\mathrm{b}}\hat{\boldsymbol{x}}(t) - {}^{\mathrm{b}}\boldsymbol{x} \in \mathbb{R}^3$ is a parameter estimation error. The definitions of excitation are given below [30].

Definition 1: A bounded signal $\Phi(t) \in \mathbb{R}^{3\times 3}$ is of IE if $\exists$ $T_{\mathrm{e}}, \zeta_{\mathrm{d}}, \sigma \in \mathbb{R}^+$ such that $\int_{T_{\mathrm{e}}-\zeta_{\mathrm{d}}}^{T_{\mathrm{e}}} \Phi(\zeta)\Phi^T(\zeta)\mathrm{d}\zeta \geq \sigma I$.

Definition 2: A bounded signal $\Phi(t) \in \mathbb{R}^{3\times 3}$ is of PE if $\exists$ $\tau_{\mathrm{d}}, \sigma \in \mathbb{R}^+$ such that $\int_{t-\zeta_{\mathrm{d}}}^{t} \Phi(\zeta)\Phi^T(\tau)\mathrm{d}\zeta \geq \sigma I \quad \forall t \geq 0$.

Define a generalized prediction error

$$\boldsymbol{\xi} := \Psi_{\mathrm{e}}(t) {}^{\mathrm{b}}\hat{\boldsymbol{x}} - \Psi_{\mathrm{e}}(t) {}^{\mathrm{b}}\boldsymbol{x} \in \mathbb{R}^3 \tag{28}$$

where $\Psi_{\mathrm{e}} \in \mathbb{R}^{3\times 3}$ is given by

$$\Psi_{\mathrm{e}}(t) := \begin{cases} \Psi_{\mathrm{o}}(t), & t < T_{\mathrm{e}} \\ \Psi_{\mathrm{o}}(T_{\mathrm{e}}), & t \geq T_{\mathrm{e}} \end{cases} \tag{29}$$

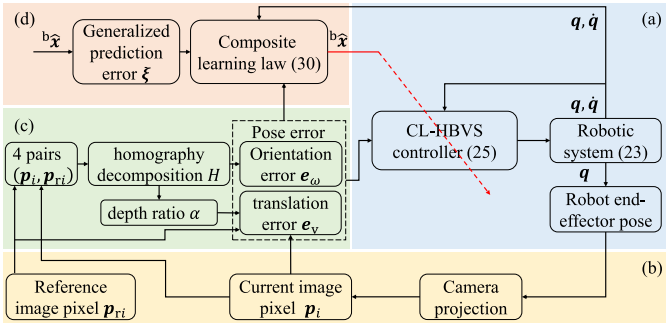


Fig. 2. Block diagram of the proposed CL-HBVS robot control scheme. (a) HBVS robot control loop. (b) Camera projection and image acquisition. (c) Homography decomposition. (d) Composite learning law.

where $\Psi_o(t) := \int_{t-\zeta_d}^{t} \Phi(\zeta)\Phi^T(\zeta)d\zeta$ and $\Phi(t) := \Phi(\boldsymbol{p}(t))$ for brevity, and $\Psi_e^b \boldsymbol{x} = \int_{t-\zeta_d}^{t} \Phi(\zeta)\boldsymbol{y}(\zeta)d\zeta$ is calculable via Φ and $\boldsymbol{y}$ in (9). Because Ψ_e is symmetric positive-semidefinite, one obtains $\Psi_e = U\Sigma U^T$ with $\Sigma := \text{diag}\,(\lambda_1\ \lambda_2, \lambda_3) \in \mathbb{R}^{3\times3}$, where $\lambda_i \geq 0$ ($i = 1$ to 3) is the ith eigenvalue of Ψ_e, and $U \in \mathbb{R}^{3\times3}$ is orthonormal, i.e., $UU^T = I$. Then, one designs a regularized composite learning law

$$ {}^b\dot{\hat{\boldsymbol{x}}} = \Gamma(Y(\boldsymbol{p}, \boldsymbol{q}, K_p\boldsymbol{e}_v)\dot{\boldsymbol{q}} - \kappa\Psi_\varrho^+\boldsymbol{\xi}) \tag{30}$$

in which $\Gamma \in \mathbb{R}^{3\times3}$ is a positive-definite diagonal matrix of learning rates, $\kappa \in \mathbb{R}^+$ is a weighting factor, and $\Psi_\varrho^+ \in \mathbb{R}^{3\times3}$ is a regularized inversion of Ψ_e defined by

$$ \Psi_\varrho^+ := U\Sigma_\varrho U^T \tag{31}$$

with $\Sigma_\varrho := \text{diag}(\lambda_{\varrho1}, \lambda_{\varrho2}, \lambda_{\varrho3})$, where $\lambda_{\varrho i} := 1/\max(\lambda_i, \varrho) \in \mathbb{R}^+$ ($i = 1$ to 3) denotes the ith eigenvalue of Ψ_ϱ^+, and $\varrho \in \mathbb{R}^+$ denotes a small threshold.

Lemma 1: If Ψ_e in (29) is positive-semidefinite, Ψ_ϱ^+ in (31) is positive-definite. Besides, $\boldsymbol{0} \leq \Psi_\varrho^+\Psi_e \leq I$, where $\Psi_\varrho^+\Psi_e = I$ iff $\Psi_e \geq \varrho I$, and $\Psi_\varrho^+\Psi_e = \boldsymbol{0}$ iff $\Psi_e = \boldsymbol{0}$.

Proof: According to Ψ_e in (29), one immediately obtains Ψ_e being positive-semidefinite. Similarly, based on Σ_ϱ under (31), Ψ_ϱ^+ is positive-definite. Then, one obtains

$$ \Psi_\varrho^+\Psi_e = U\,\text{diag}\,(\lambda_{r1}, \lambda_{r2}, \lambda_{r3})\,U^T $$

where $\lambda_{ri} := \lambda_i\lambda_{\varrho i} \in \mathbb{R}$ ($i = 1$ to 3) denotes the ith eigenvalue of $\Psi_\varrho^+\Psi_e$. Noting $\lambda_i \geq 0$ and $\varrho > 0$, it yields $0 \leq \lambda_{ri} \leq 1$ ($i =$ 1 to 3), where the left equality holds iff $\lambda_i = 0$, and the right equality holds iff $\lambda_i \geq \varrho$. Then, based on the fact of $UU^T = I$, one gets $\boldsymbol{0} \leq \Psi_\varrho^+\Psi_e \leq I$. As Ψ_ϱ^+ is positive-definite, one gets that $\Psi_\varrho^+\Psi_e = \boldsymbol{0}$ iff $\Psi_e = \boldsymbol{0}$, and $\Psi_\varrho^+\Psi_e = I$ iff $\lambda_i \geq \varrho$. ■

C. Stability and Convergence Analysis

A block diagram of the proposed CL-HBVS robot control scheme is provided in Fig. 2 to demonstrate how signals flow among the robot visual servoing system shown in Fig. 1. The following theorem states the stability and convergence results of the proposed visual servoing method.

Theorem 1: Consider the robot visual servoing system with an EIH configuration in Fig. 1, where a planar object with at least 4 coplanar feature points is fixed in the environment. If the system (23) is driven by the control law (25) with (30), closed-loop stability can be achieved in the sense that:

1) $\dot{\boldsymbol{q}}(t) \to \boldsymbol{0}$ asymptotically on $t \in [0, \infty)$;
2) $\boldsymbol{e}_v(t) \to \boldsymbol{0}$ and $\boldsymbol{e}_\omega(t) \to \boldsymbol{0}$ asymptotically, which implies the end-effector reaches its reference pose;
3) if there exist $T_e, \tau_d, \sigma \in \mathbb{R}^+$ such that the IE condition $\Psi_o(T_e) \geq \sigma I$ in *Definition 1* holds, one has ${}^b\tilde{\boldsymbol{x}}(t) \to \boldsymbol{0}$, $\boldsymbol{e}_v(t) \to \boldsymbol{0}$ and $\boldsymbol{e}_\omega(t) \to \boldsymbol{0}$ asymptotically, which implies the exact depth estimation as $\hat{z}(t) \to z(t) \to z_r$.

Proof: 1) Select a Lyapunov function candidate

$$ V = [\dot{\boldsymbol{q}}^T M(\boldsymbol{q})\dot{\boldsymbol{q}} + z_r\boldsymbol{e}_v^T K_p\boldsymbol{e}_v]/2 + [\boldsymbol{e}_\omega^T K_\omega\boldsymbol{e}_\omega + {}^b\tilde{\boldsymbol{x}}^T\Gamma^{-1}{}^b\tilde{\boldsymbol{x}}]/2. \tag{32}$$

The time derivative of V is given by

$$ \dot{V} = \dot{\boldsymbol{q}}^T M(\boldsymbol{q})\ddot{\boldsymbol{q}} + \frac{1}{2}\dot{\boldsymbol{q}}^T\dot{M}(\boldsymbol{q})\dot{\boldsymbol{q}} + z_r\boldsymbol{e}_v^T K_p\dot{\boldsymbol{e}}_v + \boldsymbol{e}_\omega^T K_\omega\dot{\boldsymbol{e}}_\omega + {}^b\tilde{\boldsymbol{x}}^T\Gamma^{-1}{}^b\dot{\tilde{\boldsymbol{x}}}. \tag{33}$$

Multiplying both sides of (20) by z_r and using (24) yields

$$ \begin{cases} z_r\dot{\boldsymbol{e}}_v = \alpha A_e(\boldsymbol{p})J(\boldsymbol{q})\dot{\boldsymbol{q}}, \\ \dot{\boldsymbol{e}}_\omega = L_\omega\,{}^cR_b J_\omega(\boldsymbol{q})\dot{\boldsymbol{q}}. \end{cases} \tag{34}$$

Substituting (27), (30), and (34) into (33) and noting *Property 3*, one immediately obtains

$$ \dot{V} = -\dot{\boldsymbol{q}}^T K_d\dot{\boldsymbol{q}} - \kappa{}^b\tilde{\boldsymbol{x}}^T\Psi_\varrho^+\Psi_e{}^b\tilde{\boldsymbol{x}} \leq 0. \tag{35}$$

Thus, the Lyapunov function (32) never increases, which implies that the closed-loop system (27) is stable in the sense of $\dot{\boldsymbol{q}}$, $\boldsymbol{e}_v$, $\boldsymbol{e}_\omega$ and ${}^b\tilde{\boldsymbol{x}}$ being bounded. With the Barbalat's Lemma, one gets $\lim_{t\to\infty}\dot{\boldsymbol{q}}(t) = \boldsymbol{0}$ asymptotically as in [16].

2) According to the invariant set theorem [31, Th. 2.3], the trajectories $(\dot{\boldsymbol{q}}, {}^b\tilde{\boldsymbol{x}})$ eventually reach an invariant set $\Omega := \{(\dot{\boldsymbol{q}}, {}^b\tilde{\boldsymbol{x}}) | \dot{V} \equiv 0\}$ with $\dot{\boldsymbol{q}} \equiv \boldsymbol{0}$ as the right-hand side of (27) is bounded, which in turn implies that in Ω one has

$$ -\alpha J^T(\boldsymbol{q})\hat{A}_e^T(\boldsymbol{p})K_p\boldsymbol{e}_v - J_\omega^T(\boldsymbol{q})\,{}^cR_b^T L_\omega^T K_\omega\boldsymbol{e}_\omega \equiv \boldsymbol{0} $$

from (25) and (27). Rearranging the above equation in a matrix form, one obtains the following result:

$$ \underbrace{J^T(\boldsymbol{q})\overbrace{\begin{bmatrix} \alpha K_e & \boldsymbol{0} \\ [{}^b\hat{\boldsymbol{x}} - {}^b\boldsymbol{t}_e]_\times K_e & -{}^cR_b^T \end{bmatrix}}^{T} K_g}_{S}\,\boldsymbol{e} = \boldsymbol{0} \tag{36}$$

where $\boldsymbol{e} := [\boldsymbol{e}_v^T, \boldsymbol{e}_\omega^T]^T \in \mathbb{R}^6$ is a pose-like error, $[\cdot]_\times$ is defined in (14), and $K_g \in \mathbb{R}^{6\times6}$ is an invertible matrix (i.e., $\text{rank}(K_g) =$ 6 as K_p, K_ω, and L_ω are all invertible) defined by

$$ K_g := \begin{bmatrix} K_p & \boldsymbol{0} \\ \boldsymbol{0} & L_\omega^T K_\omega \end{bmatrix}. $$

For an articulated robot with no less than 6 DoFs, one obtains $\text{rank}(J(\boldsymbol{q})) = 6$ except for a few singular joint positions. One also has $\text{rank}(T) = 6$ (as K_e and ${}^cR_b$ are invertible, and $\alpha > 0$). Then, one gets $\text{rank}(S) = 6$. At the invariant set Ω, $\text{rank}(S) = 6$

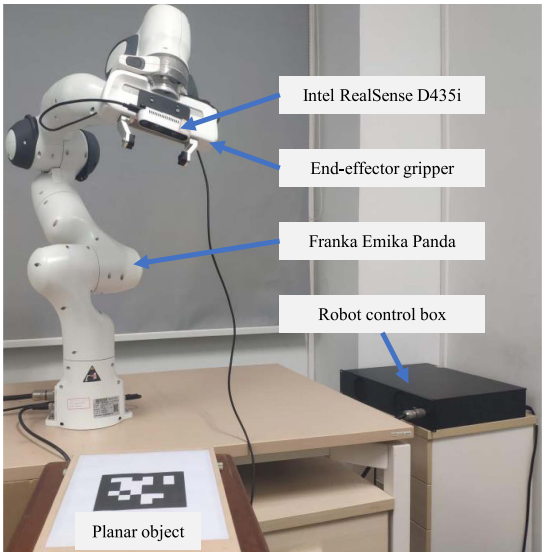


Fig. 3. Experimental environment of EIH robot visual servoing comprised of a 7-DoF collaborative robot named Franka Emika Panda, an Intel RealSense camera D435i, a planar object, and a control box.

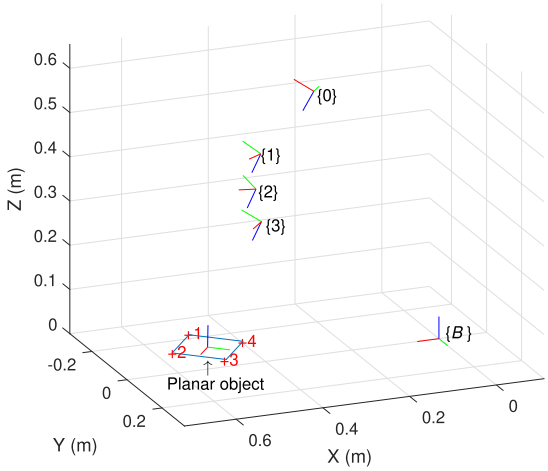


Fig. 4. Reference poses of the robot end-effector (or camera) in the world frame $\{B\}$ for capturing the reference pixel positions $\boldsymbol{p}_{\mathrm{r}}$ of a planar object, where the planar object has 4 feature points laying on the 4 corners. Note that the pose with "{0}" represents the initial end-effector pose $({}^{\mathrm{b}}R_{\mathrm{e}}(0), {}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{e}}(0))$, and the poses with "{1}" to "{3}" represent 3 reference end-effector poses $({}^{\mathrm{b}}R_{\mathrm{er}}, {}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{er}})$, where ${}^{\mathrm{b}}R_{\mathrm{e}}$ and ${}^{\mathrm{b}}R_{\mathrm{er}}$ denote the current and reference end-effector orientations in $\{B\}$, respectively, and ${}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{e}}$ and ${}^{\mathrm{b}}\boldsymbol{t}_{\mathrm{er}}$ denote the current and reference end-effector positions in $\{B\}$, respectively. The end-effector pose trajectory is with the order: {0} → {1} → {2} → {3} → {2}.

results in the unique solution $\boldsymbol{e} = \boldsymbol{0}$, which implies that the control objective (22), as well as (10), are accomplished, i.e., the end-effector reaches its reference pose.

3) To show parameter convergence, one rewrites (30) as

$$ {}^{\mathrm{b}}\dot{\tilde{\boldsymbol{x}}} + \kappa\Gamma\Psi_{\varrho}^{+}\Psi_{\mathrm{e}}\,{}^{\mathrm{b}}\tilde{\boldsymbol{x}} = \Gamma Y(\boldsymbol{p},\boldsymbol{q},K_{\mathrm{p}}\boldsymbol{e}_{\mathrm{v}})J_{\mathrm{v}}(\boldsymbol{q})\dot{\boldsymbol{q}}. \tag{37} $$

The IE condition implies $\lambda_{\min}(\Psi_{\mathrm{e}}) \geq \sigma$ on $t \in [T_{\mathrm{e}}, \infty)$ with some constants $T_{\mathrm{e}}, \sigma, \zeta_{\mathrm{d}} \in \mathbb{R}^{+}$. Based on *Lemma 1*, one gets $\Psi_{\varrho}^{+}(t)\Psi_{\mathrm{e}}(t) > 0$. Let (37) be a stable filter with a filter parameter $\kappa\Gamma\Psi_{\varrho}^{+}(t)\Psi_{\mathrm{e}}(t) > 0$ and an input $\Gamma Y(\boldsymbol{p},\boldsymbol{q})J_{\mathrm{v}}(\boldsymbol{q})\dot{\boldsymbol{q}}$. Then, it is derived that $\lim_{t\to\infty} {}^{\mathrm{b}}\tilde{\boldsymbol{x}}(t) = \boldsymbol{0}$ asymptotically on $t \in [T_{\mathrm{e}}, \infty)$ as $\lim_{t\to\infty} \dot{\boldsymbol{q}}(t) = \boldsymbol{0}$. Based on (7), (8), and the convergence of ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$, one gets $\hat{z}(t) \to z(t) \to z_{\mathrm{r}}$. ■

Remark 1: For the adaptive IBVS methods under multiple feature points [9], [10], [11], [12], [13], [14], [15], [16], [17], the convergence condition of pixel errors is $n \geq 2N$. The essential need to achieve 3-D pose control is that there exist at least four coplanar feature points to determine the unique pose, which makes these methods impractical. On the contrary, the proposed HBVS can achieve 3-D pose control by separating rotation control from translation control and only needs to control the pixel position of one feature point.

Remark 2: The motion of an articulated robot with less than 6 DoFs is restricted, i.e., the 3-D pose of the end-effector cannot be controlled completely in theory. Consider a 3-DoF robot as an example. If the 3-D position of the 3-DoF robot is controlled, the end-effector is regulated to its reference position such that the control objective (22) becomes $\boldsymbol{e}_{\mathrm{v}} \to \boldsymbol{0}$, and (36) is rewritten as $-\alpha J^{T}(\boldsymbol{q})\hat{A}_{\mathrm{e}}^{T}(\boldsymbol{p})K_{\mathrm{p}}\boldsymbol{e}_{\mathrm{v}} = \boldsymbol{0}$. Then, $\boldsymbol{e}_{\mathrm{v}} \to \boldsymbol{0}$ is obtained if $\mathrm{rank}(J(\boldsymbol{q})) = 3$. Although only a simple case about a 3-DoF robot is discussed above, similar conclusions can be drawn for robots with different numbers of DoFs.

Remark 3: Composite learning takes advantage of dynamic regressor extension with online data memory to achieve parameter convergence in adaptive control without the stringent PE condition [30], where unknown parameters are estimated by the integration of position tracking and generalized prediction errors such that only the much weaker IE condition is required to guarantee parameter convergence. Composite learning has been applied to several real-world robot control problems [32], [33], [34], [35], [36], [37], [38]. However, parameter convergence is seldom considered in existing adaptive visual servoing methods, and applying composite learning to adaptive robot visual servoing has not been reported before. This study fills this gap and provides a promising CL-HBVS method that can achieve 3-D pose control with exact depth estimation for EIH robots.

Remark 4: It follows from the proof of *3)* in *Theorem 1* that the parameter error ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$ converges to $\boldsymbol{0}$ under the composite learning law (30), which brings several benefits. 1) The Euclidean coordinate ${}^{\mathrm{b}}\boldsymbol{x}$ of the feature point Q is determined. 2) The depth z expressed in the camera frame $\{C\}$ is exactly estimated by (8). 3) Based on the above two benefits, the structure of the planar object can also be determined, which is helpful in reconstructing the scene online [23]. Noting that the Euclidean coordinate ${}^{\mathrm{c}}\boldsymbol{x}$ can be determined by ${}^{\mathrm{c}}\boldsymbol{x} = zK^{-1}\boldsymbol{p}$, thanks to the exact estimation of z after the convergence of ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$. Meanwhile, the unit normal $\boldsymbol{n}_{\mathrm{r}}$ of the planar object in $\{C_{\mathrm{r}}\}$ can be obtained by homography decomposition as in Section II-A. Then, as presented in [7], the unit normal of the planar object in $\{C\}$, denoted by $\boldsymbol{n} \in \mathbb{R}^{3}$, can be calculated as $\boldsymbol{n} = \tilde{R}\boldsymbol{n}_{\mathrm{r}}$, which implies that the planar object is defined by $\boldsymbol{n}^{T\,\mathrm{c}}\boldsymbol{x} = d$ with $d \in \mathbb{R}^{+}$ being the distance from the original of $\{C\}$ to the planar object along $\boldsymbol{n}$. Therefore, the scene can be reconstructed online.

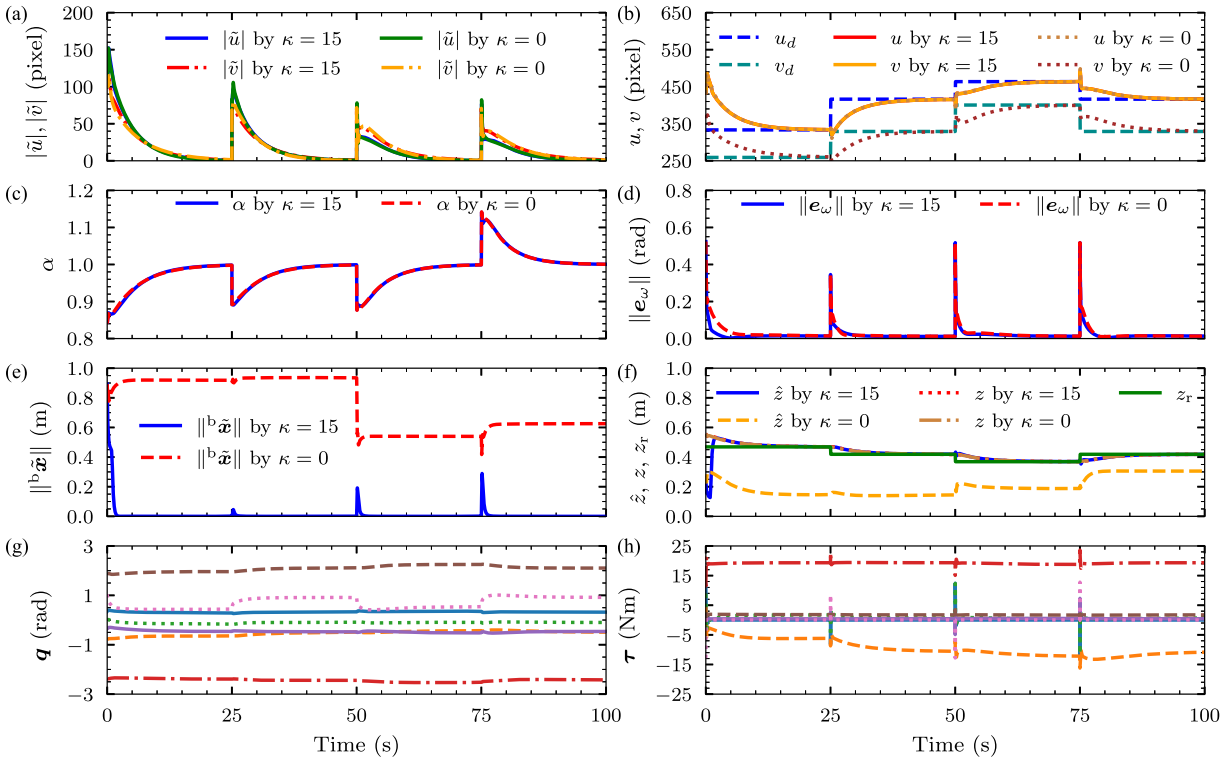


Fig. 5. Control results by the proposed CL-HBVS controller ($\kappa = 15$) and the Ad-HBVS controller ($\kappa = 0$) in simulations. (a) Pixel errors $\tilde{u}$ and $\tilde{u}$ in the u- and v-axes of the image plane, respectively. (b) Pixel tracking trajectories in the u- and v-axes of the image plane, respectively. (c) Depth ratio α. (d) Norm of the rotation error $\boldsymbol{e}_\omega$ (equal to ϕ). (e) Norm of the parameter estimation error ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$. (f) Current, estimated, and reference depth, z, $\hat{z}$, and z_{r}, respectively. (g) Joint angular position $\boldsymbol{q}$ by the proposed controller. (h) Control torque $\boldsymbol{\tau}$ by the proposed controller.

IV. Simulation Studies

An EIH robot visual servoing system exhibited in Fig. 3 is simulated via MATLAB toolboxes in [39], where a monocular camera named Intel RealSense D435i is mounted at the end-effector to acquire visual signals at a frame rate of 30 Hz, and an AprilTag with four visual feature points at four corners is located on a desk to serve as the planar object. For simulations, the dynamic parameters of the 7-DoF Panda robot are set as those in [40], the initial joint angular position is set as $\boldsymbol{q}(0) = [0.349, -0.785, 0, -2.356, -0.349, 1.920, 1.047]^T$ rad, the initial position ${}^{\mathrm{c}}\boldsymbol{t}_{\mathrm{b}}(0)$, and orientation ${}^{\mathrm{c}}R_{\mathrm{b}}(0)$ between the camera frame $\{C\}$ and the world frame $\{B\}$ are set as

$$
{}^{\mathrm{c}}\boldsymbol{t}_{\mathrm{b}}(0) = [-0.0818, 0.5898, 0.3400]^T \mathrm{m}
$$

$$
{}^{\mathrm{c}}R_{\mathrm{b}}(0) = \begin{bmatrix} 0.1425 & -0.9586 & 0.2467 \\ -0.8849 & -0.2350 & -0.4022 \\ 0.4435 & -0.1610 & -0.8817 \end{bmatrix}
$$

respectively, and the camera's intrinsic parameter K is retrieved from the software named Intel RealSense SDK 2.0 as

$$
K = \begin{bmatrix} 612 & 0 & 322 \\ 0 & 612 & 234 \\ 0 & 0 & 1 \end{bmatrix}.
$$

The control parameters of the proposed CL-HBVS (25) with (30) are set as $K_{\mathrm{d}} = \mathrm{diag}(22, 20, 22, 20, 5, 5, 4)$, $K_{\mathrm{p}} = \mathrm{diag}(2.6 \times 10^{-4}, 2.6 \times 10^{-4}, 60)$, $K_\omega = 25I$, $\Gamma = 0.3I$, $\kappa = 15$, $\varrho = 1$, $T_{\mathrm{e}} = 5$, and $\tau_{\mathrm{d}} = 5$. The proposed approach is compared with an adaptive HBVS (Ad-HBVS) controller, which is obtained by simply setting $\kappa = 0$ and keeping other control parameters the same as those of the proposed controller. Note that the control gains for the pixel errors $\tilde{u}$ and $\tilde{v}$ are set much smaller as their scales are much larger than those of the depth ratio α and rotation error $\boldsymbol{e}_\omega$. The initial position estimate is set as ${}^{\mathrm{b}}\hat{\boldsymbol{x}}(0) = [1.0, 0.5, 0.5]^T$ m, and its true value is ${}^{\mathrm{b}}\boldsymbol{x} = [0.5328, -0.1074, 0.0433]^T$ m. We design a regulation task, where the reference pose of the end-effector is changed at an interval of 25 s, and there are four reference poses as illustrated in Fig. 4. The nonhomogeneous pixel positions $\boldsymbol{p}(0)$ and $\boldsymbol{p}_{\mathrm{r}}$ of all feature points Q_i ($i = 1$–4) with respect to each robot pose are given in Table I, where Q_1 (working as Q) is used for translation control.

Simulation results by the two controllers are depicted in Fig. 5. For the proposed CL-HBVS, the pixel position $\boldsymbol{p} = [u, v, 1]^T$ of the end-effector reaches the reference pixel points $\boldsymbol{p}_{\mathrm{r}}$ [see Fig. 5(b)] with high accuracy [see Fig. 5(a)], the depth ratio α converges to 1 rapidly [see Fig. 5(c)], and the rotation error $\boldsymbol{e}_\omega$ converges to $\boldsymbol{0}$ rapidly [see Fig. 5(d)], which implies that the robot pose can be well controlled. Despite both controllers achieve almost the same accuracy in $|\tilde{u}|$ and $|\tilde{v}|$, α and $\|\tilde{\boldsymbol{e}}_\omega\|$ [see Fig. 5(a), (c), and (d)], the estimation error ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$ by the proposed CL-HBVS converges to $\boldsymbol{0}$ at around 2.5 s compared to the Ad-HBVS [see Fig. 5(e)], which leads to the exact approximation of the ture depth z by the estimated one $\hat{z}$ [see Fig. 5(f)]. The

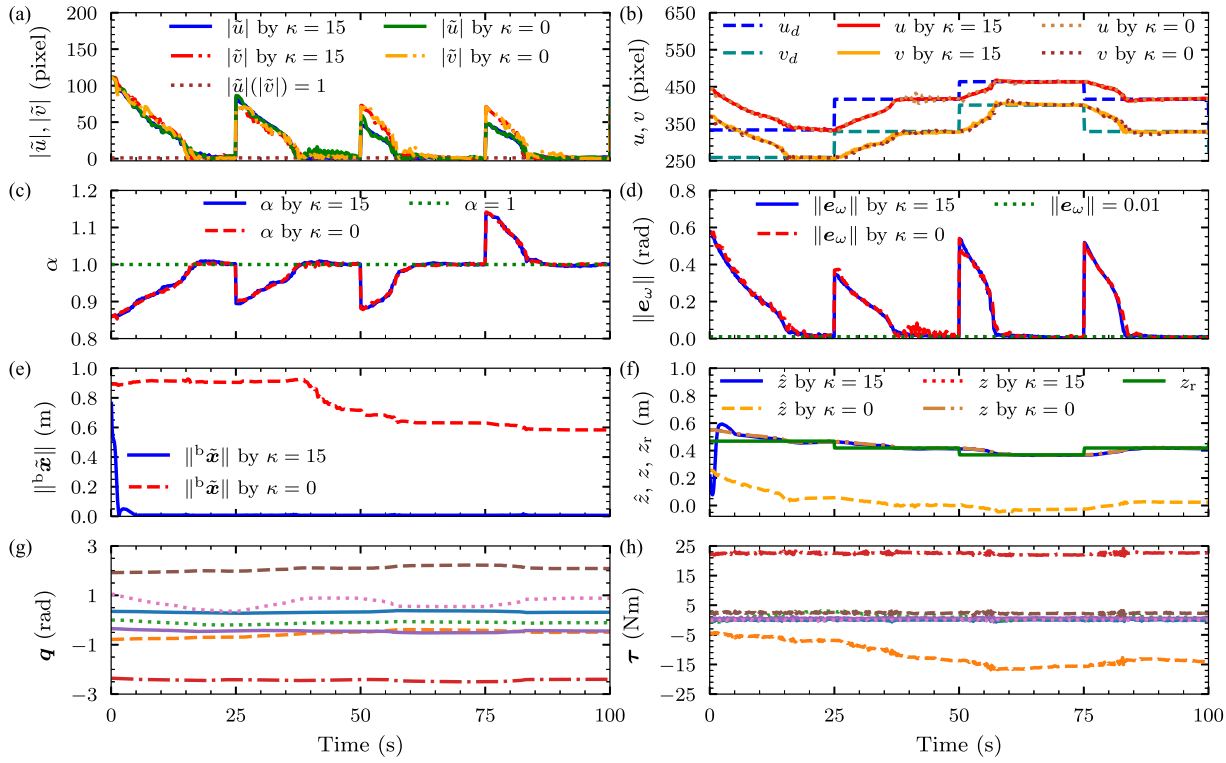


Fig. 6. Control results by the proposed CL-HBVS controller ($\kappa = 15$) and the Ad-HBVS controller ($\kappa = 0$) in experiments. (a) Pixel errors $\tilde{u}$ and $\tilde{u}$ in the u- and v-axes of the image plane, respectively. (b) Pixel tracking trajectories in the u- and v-axes of the image plane, respectively. (c) Depth ratio α. (d) Norm of the rotation error $\boldsymbol{e}_\omega$ (equal to ϕ). (e) Norm of the parameter estimation error ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$. (f) Current, estimated, and reference depth, z, $\hat{z}$, and z_{r}, respectively. (g) Joint angular position $\boldsymbol{q}$ by the proposed controller. (h) Control torque $\boldsymbol{\tau}$ by the proposed controller.

TABLE I
PIXEL COORDINATES $\boldsymbol{p}(0)$ AND $\boldsymbol{p}_{\mathrm{r}}$ OF ALL FEATURE POINTS Q_i ($i = 1$ TO 4) CORRESPONDING TO FOUR ROBOT POSES

Robot pose	{0}	{1}	{2}	{3}
Feature point Q_1	(444,372)	(333,259)	(416,329)	(464,400)
Feature point Q_2	(356,270)	(304,103)	(326,177)	(462,199)
Feature point Q_3	(257,351)	(143,120)	(168,264)	(269,197)
Feature point Q_4	(351,444)	(175,292)	(270,435)	(272,427)

Note that the nonhomogeneous pixel coordinates $\boldsymbol{p}(0) = (u(0), v(0))$ and $\boldsymbol{p}_{\mathrm{r}} = (u_{\mathrm{r}}, v_{\mathrm{r}})$ are used for simplicity, and each value of $\boldsymbol{p}_{\mathrm{r}}$, corresponding to a reference robot (or camera) pose depicted in Fig. 4, is obtained from a "teach by showing" approach [2] with the physical robot and camera.

TABLE II
CURRENT DEPTH z AND ITS ESTIMATE $\hat{z}$ AT SEVERAL INSTANTS

Time t (s)	24	49	74	99
Current true depth z (m)	0.4650	0.4128	0.3669	0.4131
Estimated depth $\hat{z}$ (m)	0.4678	0.4166	0.3689	0.4170

joint angular position $\boldsymbol{q}$ and the control torque $\boldsymbol{\tau}$ by the proposed CL-HBVS are shown in Fig. 5(g) and (h), respectively. As the update of ${}^{\mathrm{b}}\hat{\boldsymbol{x}}$ in (30) is a function of the joint velocity $\dot{\boldsymbol{q}}$, the abrupt change of $\dot{\boldsymbol{q}}$ at each time of switching reference pose, caused by the abrupt change of $\boldsymbol{\tau}$ [see Fig. 5(h)], leads to a sudden boost in ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$ [see Fig. 5(e)]. Note that the above phenomenon is observed in both of the applied controllers.

V. EXPERIMENTAL STUDIES

Experiments based on the physical Panda robot are carried out to verify the proposed controller, where the details about the robot setup can be referred to [41]. The initial settings (i.e., $\boldsymbol{q}(0)$, ${}^{\mathrm{c}}R_{\mathrm{b}}(0)$, ${}^{\mathrm{c}}\boldsymbol{t}_{\mathrm{b}}(0)$, ${}^{\mathrm{b}}\hat{\boldsymbol{x}}(0)$) and camera intrinsic parameter K remain the same as those in simulations. A visual tracker algorithm integrated into ViSP (an open source C++ library for visual servoing [42]) is utilized to extract the visual feature, and the control algorithm is implemented in the ViSP. The Euclidean coordinate ${}^{\mathrm{b}}\boldsymbol{x}$ of any feature point in the world frame $\{B\}$ is unknown for the control algorithm in experiments such that the estimation error ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$ can not be exactly calculated. For an evaluation purpose, we use a pose estimation algorithm in the ViSP [43] with the prior geometry knowledge of AprilTag to get ${}^{\mathrm{b}}\boldsymbol{x} = [0.5328, -0.1074, 0.0433]^T$ m for calculating ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$. The control task and control parameters (i.e., K_1, K_{p}, K_ω, Γ, κ, ϱ, T_{e}, τ_{d}) are set the same as those in simulations.

Experimental results by the two controllers are illustrated in Fig. 6. For the proposed CL-HBVS, the pixel position $\boldsymbol{p}$ of the feature point Q reaches its reference pixel position $\boldsymbol{p}_{\mathrm{r}}$ [see Fig. 6(b)] with about a steady-state error of 1 pixel for both $\tilde{u}$ and $\tilde{v}$ [see Fig. 6(a)], the depth ratio α converges to around 1 [see Fig. 6(c)], and the rotation error norm $\|\boldsymbol{e}_\omega\|$ converges to about 0.01 rad [see Fig. 6(d)]. Although both controllers achieve almost the same accuracy in the tracking errors $\boldsymbol{e}_{\mathrm{v}}$ and $\boldsymbol{e}_\omega$ [see Fig. 6(a)–(d)], the parameter error ${}^{\mathrm{b}}\tilde{\boldsymbol{x}}$ by the proposed CL-HBVS converges to $\boldsymbol{0}$ at around 5 s compared to the Ad-HBVS, which

results in the convergence of the estimated depth $\hat{z}$ to its true value z [see Fig. 6(e)–(f)]. The errors between z and $\hat{z}$ are less than 5 mm at instants $t = 24$, 49, 74, and 99 s, which illustrates the accurate depth estimation by the CL-HBVS [see Table II]. The joint angular position $\boldsymbol{q}$ and the control torque $\boldsymbol{\tau}$ by the CL-HBVS are depicted in Fig. 6(g) and (h), respectively. As shown in Figs. 5(a)–(d) and 6(a)–(d), the pose-like error $\boldsymbol{e}$ in (36) in simulations converges faster than the one in experiments (especially $\boldsymbol{e}_\omega$), which means that $\dot{\boldsymbol{q}}$ in simulations is faster than that in experiments. Thus, $\boldsymbol{\tau}$ by the CL-HBVS [see Fig. 6(h)] in experiments is smoother than that in simulations, which results in a smoother parameter estimation [see Fig. 6(e)].

The following content analyzes the 3-D pose control accuracy of the proposed CL-HBVS from perceptual intuition. As $\alpha = 1$ at the steady state, one has a depth error $\tilde{z} := |z - z_\mathrm{r}| = 0$. Based on (1) and $\boldsymbol{n}_\mathrm{r}^T {}^\mathrm{c}\boldsymbol{x}_\mathrm{r} = d_\mathrm{r}$ [20], (2) can be rewritten as ${}^\mathrm{c}\boldsymbol{x} = \tilde{R}{}^\mathrm{c}\boldsymbol{x}_\mathrm{r} + \tilde{\boldsymbol{t}}$. Then, one has $\tilde{\boldsymbol{t}} = {}^\mathrm{c}\tilde{\boldsymbol{x}} - (\tilde{R} - I){}^\mathrm{c}\boldsymbol{x}_\mathrm{r}$ with ${}^\mathrm{c}\tilde{\boldsymbol{x}} := {}^\mathrm{c}\boldsymbol{x} - {}^\mathrm{c}\boldsymbol{x}_\mathrm{r} \in \mathbb{R}^3$. As $\|\boldsymbol{e}_\omega\| \approx 0.01$ rad, $(\tilde{R} - I){}^\mathrm{c}\boldsymbol{x}_\mathrm{r}$ is small enough such that $\tilde{\boldsymbol{t}} \approx {}^\mathrm{c}\tilde{\boldsymbol{x}}$. If there is a pixel error $\tilde{\boldsymbol{p}} = [1, 1, 0]^T$ pixel and a depth $z = 1$ m, one has the translational position error $\tilde{\boldsymbol{t}} = [0.0016, 0.0016, 0]^T$ m by (5). Thus, the proposed CL-HBVS achieves accurate 3-D pose control.

VI. Conclusion

In this article, a CL-HBVS method had been developed for 3-D visual regulation of EIH robots with monocular cameras under unknown Cartesian feature positions. Asymptotic convergence of the pose error was rigorously proven without any excitation condition, and position parameter convergence was obtained under the weakened IE condition, resulting in an exact estimation of the time-varying depth. Simulations and experiments on the 7-DoF Panda robot had demonstrated that the proposed method successfully achieved 3-D robot pose control using four coplanar feature points and exact depth estimation under simple regulation tasks. Due to the benefits of homography decomposition, the proposed method had the potential to be extended to the case with extra uncertainties from the camera's extrinsic parameters for both regulation and tracking problems, but this can result in the overparameterization problem [16] that would be taken into consideration in further studies.

References

[1] F. Chaumette, S. Hutchinson, and P. Corke, "Visual servoing," in *Springer Handbook of Robotics*, 2nd ed., B. Siciliano and O. Khatib, Eds. Berlin, Germany: Springer, 2016, pp. 841–866.

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

[3] 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. Vis.*, vol. 37, no. 1, pp. 79–97, Jun. 2000.

[4] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *Proc. Conf. Vis. Control*, London, U.K., 1998, pp. 66–78.

[5] Y. Shen, D. Sun, Y.-H. Liu, and K. Li, "Asymptotic trajectory tracking of manipulators using uncalibrated visual feedback," *IEEE/ASME Trans. Mechatronics*, vol. 8, no. 1, pp. 87–98, Mar. 2003.

[6] F. Chaumette and S. Hutchinson, "Visual servo control Part I. Basic approaches," *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

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

[8] C. C. Cheah, C. Liu, and J. J. E. Slotine, "Adaptive Jacobian vision based control for robots with uncertain depth information," *Automatica*, vol. 46, no. 7, pp. 1228–1233, Jun. 2010.

[9] Y.-H. Liu, H. Wang, C. Wang, and K. K. Lam, "Uncalibrated visual servoing of robots using a depth-independent interaction matrix," *IEEE Trans. Robot.*, vol. 22, no. 4, pp. 804–817, Aug. 2006.

[10] H. Wang, Y.-H. Liu, and D. Zhou, "Adaptive visual servoing using point and line features with an uncalibrated eye-in-hand camera," *IEEE Trans. Robot.*, vol. 24, no. 4, pp. 843–857, Aug. 2008.

[11] Y.-H. Liu and H. Wang, "An adaptive controller for image-based visual servoing of robot manipulators," in *Proc. World Congr. Intell. Control Autom.*, Jinan, China, 2010, pp. 988–993.

[12] H. Wang, Y.-H. Liu, and W. Chen, "Uncalibrated visual tracking control without visual velocity," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 6, pp. 1359–1370, Nov. 2010.

[13] X. Liang, X. Huang, M. Wang, and X. Zeng, "Improved stability results for visual tracking of robotic manipulators based on the depth-independent interaction matrix," *IEEE Trans. Robot.*, vol. 27, no. 2, pp. 371–379, Feb. 2011.

[14] X. Liang, H. Wang, Y.-H. Liu, W. Chen, and J. Zhao, "A unified design method for adaptive visual tracking control of robots with eye-in-hand/fixed camera configuration," *Automatica*, vol. 59, pp. 97–105, Sep. 2015.

[15] H. Wang, C. C. Cheah, W. Ren, and Y. Xie, "Passive separation approach to adaptive visual tracking for robotic systems," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 6, pp. 2232–2241, Apr. 2018.

[16] Y. Li, H. Wang, Y. Xie, C. C. Cheah, and W. Ren, "Adaptive image-space regulation for robotic systems," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 2, pp. 850–857, Dec. 2021.

[17] Y. Zhang and C. Hua, "A new adaptive visual tracking scheme for robotic system without image-space velocity information," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 52, no. 8, pp. 5249–5258, Aug. 2022.

[18] Y. Fang, W. E. Dixon, D. M. Dawson, and J. Chen, "Robust 2.5D visual servoing for robot manipulators," in *Proc. Amer. Control Conf.*, Denver, CO, USA, 2003, pp. 3311–3316.

[19] J. Chen, D. M. Dawson, W. E. Dixon, and A. Behal, "Adaptive homography-based visual servo tracking for a fixed camera configuration with a camera-in-hand extension," *IEEE Trans. Control Syst. Technol.*, vol. 13, no. 5, pp. 814–825, Sep. 2005.

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

[21] J. Chen, V. K. Chitrakaran, and D. M. Dawson, "Range identification of features on an object using a single camera," *Automatica*, vol. 47, no. 1, pp. 201–206, Jan. 2011.

[22] S. Mehta, V. Jayaraman, T. Burks, and W. Dixon, "Teach by zooming: A unified approach to visual servo control," *Mechatronics*, vol. 22, no. 4, pp. 436–443, Jun. 2012.

[23] A. Parikh, R. Kamalapurkar, H.-Y. Chen, and W. E. Dixon, "Homography based visual servo control with scene reconstruction," in *Proc. IEEE Conf. Decis. Control*, Osaka, Japan, 2015, pp. 6972–6977.

[24] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*. London, U.K.: Springer, 2010.

[25] J. A. Gangloff, M. De Mathelin, and G. Abba, "6 DoF high speed dynamic visual servoing using GPC controllers," in *Proc. IEEE Int. Conf. Robot. Autom.*, Leuven, Belgium, 1998, pp. 2008–2013.

[26] X. Liang et al., "Fully uncalibrated image-based visual servoing of 2 DoFs planar manipulators with a fixed camera," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10895–10908, Oct. 2022.

[27] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2004.

[28] E. Malis and M. Vargas, "Deeper understanding of the homography decomposition for vision-based control," *INRIA, Sophia Antipolis Cedex, France, Res. Rep. RR-6303*, Sep. 2007.

[29] J. Chen, A. Behal, D. Dawson, and Y. Fang, "2.5D visual servoing with a fixed camera," in *Proc. Amer. Control Conf.*, Denver, CO, USA, 2003, pp. 3442–3447.

[30] Y. Pan and H. Yu, "Composite learning robot control with guaranteed parameter convergence," *Automatica*, vol. 89, pp. 398–406, Mar. 2018.

[31] I. Barkana, "Defending the beauty of the invariance principle," *Int. J. Control*, vol. 87, no. 1, pp. 186–206, 2014.

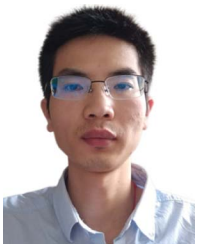
[32] C. Yang, Y. Jiang, W. He, J. Na, Z. Li, and B. Xu, "Adaptive parameter estimation and control design for robot manipulators with finite-time convergence," *IEEE Trans. Ind. Electron.*, vol. 65, no. 10, pp. 8112–8123, Oct. 2018.

[33] K. Guo, Y. Pan, and H. Yu, "Composite learning robot control with friction compensation: A neural network-based approach," *IEEE Trans. Ind. Electron.*, vol. 66, no. 10, pp. 7841–7851, Dec. 2019.

[34] K. Guo, Y. Pan, D. Zheng, and H. Yu, "Composite learning control of robotic systems: A least squares modulated approach," *Automatica*, vol. 111, Jan. 2020, Art. no. 108612.
[35] D. Huang, C. Yang, Y. Pan, and L. Cheng, "Composite learning enhanced neural control for robot manipulator with output error constraints," *IEEE Trans. Ind. Inform.*, vol. 17, no. 1, pp. 209–218, Dec. 2021.
[36] K. Guo, M. Li, W. Shi, and Y. Pan, "Adaptive tracking control of hydraulic systems with improved parameter convergence," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7140–7150, Jul. 2022.
[37] K. Guo, Y. Liu, B. Xu, Y. Xu, and Y. Pan, "Locally weighted learning robot control with improved parameter convergence," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13236–13244, Dec. 2022.
[38] K. Guo and Y. Pan, "Composite adaptation and learning for robot control: A survey," *Annu. Rev. Control*, to be published, doi: 10.1016/j.arcontrol.2022.12.001.
[39] P. Corke, *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*, 2nd ed.Cham, Switzerland: Springer, 2017.
[40] C. Gaz, M. Cognetti, A. Oliva, P. R. Giordano, and A. De Luca, "Dynamic identification of the Franka Emika Panda robot with retrieval of feasible parameters using penalty-based optimization," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 4147–4154, Jul. 2019.
[41] X. Liu, Z. Li, and Y. Pan, "Preliminary evaluation of composite learning tracking control on 7-DoF collaborative robots," *IFAC-PapersOnLine*, vol. 54, no. 14, pp. 470–475, 2021.
[42] E. Marchand, F. Spindler, and F. Chaumette, "ViSP for visual servoing: A generic software platform with a wide class of robot control skills," *IEEE Robot. Autom. Mag.*, vol. 12, no. 4, pp. 40–52, Dec. 2005.
[43] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Daejeon, South Korea, 2016, pp. 4193–4198.

Beixian Lai received the B.Eng. degree in optic information science and technology from the Hefei University of Technology, Hefei, China, in 2014.

He is currently working toward the M.Eng. degree in computer science and technology with the Sun Yat-sen University, Guangzhou, China. His research interests include visual servoing and adaptive control.



Zhiwen Li received the B.Eng. degree in measurement and control technology and instrumentation from the Southeast University, Nanjing, China, in 2013, and the M.Eng. degree in control science and engineering from the Northwestern Polytechnical University, Xi'an, China, in 2016.

He is currently working toward the Ph.D. degree in computer science and technology with the Sun Yat-sen University, Guangzhou, China. His research interests include collaborative robots and adaptive control.

Weibing Li (Member, IEEE) received the Ph.D. degree in mechanical engineering from the University of Leeds, Leeds, U.K., in 2018.

From 2018 to 2020, he was a Postdoctoral Fellow with the Chow Yuk Ho Technology Centre for Innovative Medicine, The Chinese University of Hong Kong, Hong Kong. He is currently an Associate Professor with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. His research interests include medical robotics, industrial robotics, modular robotics, neural networks, and cybernetics.

Chenguang Yang (Senior Member, IEEE) received the B.Eng. degree in measurement and control from Northwestern Polytechnical University, Xi'an, China, in 2005, the Ph.D. degree in control engineering from the National University of Singapore, Singapore, in 2010.

He received the postdoctoral training in human robotics from the Imperial College London, London, UK. His research interest lies in human-robot interaction and intelligent system design.

Dr. Yang was a Recipient of the IEEE Transactions on Robotics Best Paper Award (2012) and IEEE Transactions on Neural Networks and Learning Systems Outstanding Paper Award (2022) as the lead author. He is a Fellow of the Institution of Mechanical Engineers (IMechE), a Fellow of the Institute of Engineering and Technology (IET), and a Fellow of the British Computer Society (BCS).

Yongping Pan (Senior Member, IEEE) received the Ph.D. degree in control theory and control engineering from the South China University of Technology, Guangzhou, China, in 2011.

He spent one year in the industry as a Control Systems Engineer in Shenzhen and Guangzhou, China during 2007–2008, and has long-term research experience with the Nanyang Technological University, Singapore, National University of Singapore, Singapore, and University of Tokyo, Tokyo, Japan. He is currently a Professor who leads the Robot Control and Learning Group with the Sun Yat-sen University, Shenzhen, China. He has authored or coauthored over 150 peer-reviewed academic papers, including more than 100 refereed journal papers. His research interests include automatic control and machine learning for robotic applications.

Dr. Pan is the Founding Chair of the IEEE Robotics and Automation Society Guangzhou Chapter. He has been invited as an Associate Editor of several top-tier journals, such as IEEE Transactions, has organized three special issues as the Lead Guest Editor in reputable journals, and has served as an Organizing Committee Member of several international conferences.