Adaptive Depth Estimation for Eye-In-Hand Robot Control With Homography-Based Visual Servoing

Visual servoing is a crucial control method in robotic systems that utilizes visual information as feedback to measure the environment without the need for physical contact.The absence of depth information in 2D image data poses a challenge in accurately projecting 3D Euclidean space onto image space. Therefore, finding a reliable method to acquire depth information becomes a crucial problem in vision control. In light of the fact that the depth of a feature point can be expressed using its Cartesian feature position, a controller based on dynamic homography-based visual servoing (HBVS) is proposed. Additionally, an enhanced adaptive approach is utilized to estimate the pose. In comparison to the conventional adaptive HBVS, the proposed method is founded on a dynamic design and exhibits notable enhancements in parameter convergence and control effectiveness. This has been validated through experiments conducted on the 7 degrees of freedom(DoF) collaborative robot, Franka Emika Panda, affirming the efficacy of the proposed approach.

Keywords: Collaborative robot,visual servo,dynamic control,adaptive estimation.

1. INTRODUCTION

Visual servoing is a crucial control technique in robot systems, utilizing visual information as feedback for non-contact measurement of the environment. It offers significant improvements in the flexibility and accuracy of robot systems and finds wide application in wheeled robots,manipulator, and Unmanned Aerial Vehicle(UAV) control. Since the early 1990s, visual servoing has emerged as one of the most actively researched topics in robotics. Various methods have been proposed to address the challenges of visual servo control, which can be categorized into three types [(1)]: image-based methods, position-based methods, and hybrid methods. In image-based visual servo control, the feedback signal comprises purely image spatial information. Position-based visual servo control, on the other hand, utilizes reconstructed Euclidean information in the feedback loop. HBVS is advantageous in avoiding singularities in the image Jacobian matrix [(7)]. And it adjusts the translation and rotation error system by reconstructing information from both Euclidean space and image space. The rotation and displacement components are decoupled using the homography relation and polar conditions. This decoupled information is then used to develop two types of visual controllers.

In the context of HBVS, there are various parameter uncertainties, with the time-varying depth of the observed object's feature point being one of the main sources. To compensate for the lack of depth information in two-dimensional image data, researchers have explored alternative sensor options such as laser and sonar technology. Additionally, some have investigated the use of camera-based vision systems combined with other sensors or the utilization of binocular cameras to triangulate corresponding images for depth estimation. However, integrating additional sensors presents practical disadvantages, including increased cost, complexity, reduced reliability, and increased processing requirements. Consequently, recent research has focused on monocular camera-based visual servo strategies that rely on parameter estimation to address the lack of depth information. Among these approaches, adaptive control has proven to be an effective method for handling parameter uncertainties. Numerous adaptive HBVS methods have been developed to tackle uncertain parameters in robots [(8)], including feature position parameters, camera intrinsic and extrinsic parameters, as well as robot kinematics and dynamics parameters. As the depth of a feature point can be represented as a function of a set of unknown parameters, obtaining the desired depth information becomes feasible if the parameter estimates converge to their true values. Indeed, most current adaptive HBVS methods primarily focus on robot kinematics, assuming precise control of robot velocity [(10)]. However, research has shown that incorporating dynamic control can enhance system stability and overall performance [(4)],and Accurately controlling robot speed based solely on kinematics can indeed be challenging to achieve in practical applications [(9)]. Therefore, it is crucial to consider robot dynamics in visual servoing to ensure optimal control performance. The recent work by J Na [(13)] et al. introducing an adaptive estimation method driven by estimation error holds promise in effectively achieving quick convergence of estimated parameters to their exact values. This highlights the potential for further research in developing new parameter estimation learning algorithms in the field of visual servoing. Such advancements could lead to enhanced adaptive control performance, making it a valuable area for further investigation and study.

In this paper, we investigate the problem of 3D visual alignment for an eye-in-hand (EIH) robot using a monocular camera. Prior to executing the control task, it is necessary to directly measure the pose of the feature point, which

*

can be cumbersome and time-consuming. To address this issue, we propose considering unknown Cartesian feature locations, which allows for the arbitrary selection of feature points [14].As depth can be represented by the Cartesian feature position, we design a unified adaptive law that estimates both the feature position and depth parameters. Additionally, we take into account the nonlinear forces that influence robot control and improve the overall stability of the EIH system by integrating robot dynamics. The controller incorporates an adaptive feedforward component that compensates for the unknown feature positions. The key contributions of this study are as follows:

1.we introduce a novel adaptive HBVS control law that incorporates dynamics to achieve 3D pose control of an EIH robot under unconstrained conditions.

2.The proposed control law enables accurate estimation of the time-varying depth when the unknown Cartesian position parameter converges to its true value. This accurate depth estimation is beneficial for online scene reconstruction of planar objects.

3.To validate the effectiveness of the proposed method, we conduct experiments on a collaborative robot with 7-DoF. The experimental results demonstrate the efficacy of the proposed approach in achieving precise 3D pose control for the EIH robot.

In this paper, the structure is organized as follows. Section 2 establishes a geometric relationship by combining the robot's forward kinematics, camera parameters, and feature point coordinates in different coordinate systems. Section 3 outlines the control objective, designs an adaptive law and controller, and provides a stability analysis using the Lyapunov method. Finally, Section 4 presents experimental results conducted on a robot to demonstrate the effectiveness of the proposed method.

2. GEOMETRIC MODEL FORMULATION

2.1 Camera model reconstruction The visual servoing system with an EIH in figure 1 configureuration establishes a direct and fixed connection between the visual sensor and the end-effector of the robot. This system includes the camera frame $\{C\}$, the end-effector frame $\{E\}$, and the world frame $\{B\}$.Initially, the planar target is positioned within the camera's field of view. The reference camera frame $\{C_r\}$ is employed to represent the coordinates of the desired target position as the robot moves.

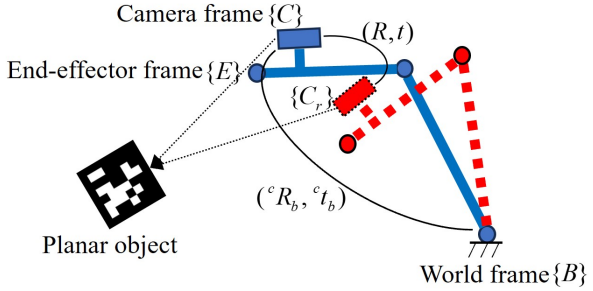

Fig. 1. An articulated robot with an EIH monocular camera for visual servoing

A rotation matrix $R \in \mathbb{R}^{3\times3}$and a translation vector $t \in \mathbb{R}^{3\times1}$ are used to represent the positional relationship between $\{C\}$ and $\{C_r\}$, respectively. Also, denote by ${}^{c}x_i(t) = [x_i(t), y_i(t), z_i(t)]^T \in \mathbb{R}^3$ and ${}^{c_r}x_i(t) = [x_{ri}(t), y_{ri}(t), z_{ri}(t)]^T \in \mathbb{R}^3$ in turn the Euclidean coordinates in $\{C\}$ and $\{C_r\}$ where i denotes the feature points$P_i(i = 1$ to $N)$.Perform the following normalization operation on the Euclidean coordinates

$$x_i(t) = {}^{c}x_i(t)/z_i(t), x_{ri}(t) = {}^{c}x_i(t)/z_{ri}(t). \tag{1}$$

For one feature point, the normalized coordinates can be expressed as$x(t) = [x, y, 1]^T$and $x_r(t) = [x_r, y_r, 1]^T$, the relationships in (1)can be expressed as

$$x_i = \underbrace{\frac{z_{ri}}{z_i}}_{\alpha}\underbrace{\left(R + \frac{t}{d_r}{n_r}^T\right)}_{H}x_{ri}. \tag{2}$$

where $\alpha \in \mathbb{R}$ are scaling terms and $H \in \mathbb{R}^{3\times3}$ denote the Euclidean homographies.

In the pinhole camera model, the relationship between image plane coordinates and pixel coordinates in the camera frame$\{C\}$and$\{C_r\}$ can be expressed as

$$p_i = Kx_i, p_{ri} = Kx_{ri}. \tag{3}$$

where $K \in \mathbb{R}^{3\times3}$ is a constant and invertible intrinsic camera-calibration matrix that is explicitly defined as

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

In(4) $f_x, f_y \in \mathbb{R}$ represent the scale factors in the direction of the camera's u and v axes, and$c_x, c_y \in \mathbb{R}$ are the coordinates of the camera's principal points.

Based on(3)the Euclidean relationship in(2)can be expressed in terms of the image coordinates as

$$p_i = \alpha_i \underbrace{(KHK^{-1})}_{M} p_{ri}. \tag{5}$$

where $M \in \mathbb{R}^{3\times3}$ is a projective homography matrix. The H can be calculated by using four groups of corresponding feature points. By decomposing H, the depth ratio α and the rotation matrix R can be obtained. Alternatively, the virtual parallax method can be employed to fulfill the aforementioned requirements [5]. This method necessitates a minimum of eight feature points, with the additional condition that no four feature points are coplanar, in order to achieve accurate results.

2.2 Linear Parameterization of Camera Models In EIH system, the camera's extrinsic parameters homogeneous matrix ${}^{c}T_e \in \mathbb{R}^{4\times4}$ by ${}^{c}R_e \in \mathbb{R}^{3\times3}$ and ${}^{c}t_e \in \mathbb{R}^3$,respectively. The orientation ${}^{e}R_b(t) \in \mathbb{R}^{3\times3}$ and position ${}^{e}t_b(t) \in \mathbb{R}^3$ between the end-effector frame$\{C\}$ and the world frame $\{B\}$ can be obtained by using robot forward kinematics. With the above description, we can get orientation ${}^{c}R_b(t) = {}^{c}R_e{}^{e}R_b(t)$ and position ${}^{c}t_b(t) = {}^{c}R_e{}^{e}t_b + {}^{c}t_e$ of the world frame $\{B\}$ in the camera frame $\{C\}$ Camera extrinsic parameters ${}^{c}R_e$ and ${}^{c}t_e$ are obtained by offline calibration and are known constants.

Let ${}^{b}x \in \mathbb{R}^{3\times3}$ denote a coordinate of the feature point P_i

in $\{B\}$ The relationship between the projections of the feature points in $\{C\}$ and image plane coordinate systems can be expressed as

$$p_i = \frac{K}{z_i}{}^{c}x_i. \tag{6}$$

where the coordinate ${}^{c}x_i$ is given by

$${}^{c}x_i = {}^{c}R_{b}{}^{b}x_i + {}^{c}t_{b}. \tag{7}$$

Based on(6) and(7), the depth is expressed by

$$z_i = k_3^T({}^{c}R_{b}{}^{b}x_i + {}^{c}t_{b}). \tag{8}$$

where $k_i^T \in {}^{b}3^{\times 3}$ is the i th row of K.Denote the estimate of ${}^{b}x_i$ by ${}^{b}\hat{x}_i$,Then the depth estimate can be expressed as

$$\hat{z}_i = k_3^T({}^{c}R_b{}^{b}\hat{x}_i + {}^{c}t_b). \tag{9}$$

A linearly parameterized formula of (6), which is prominent for adaptive law design and stability analysis, is given as follows.

Property 1: It is assumed that ${}^{b}x_i$, the position of the feature point P_i, is unknown but constant. Multiplying each side of (6) by z_i and substituting (8) into the result, The following results are obtained for the linear parameterization of ${}^{b}x_i$.

$$y_i(p_i) = \Phi_i^T(p_i)^{b}x_i. \tag{10}$$

where $\Phi_i(p_i) = {}^{c}R_b^T K^T(p_i) \in \mathbb{R}^{3\times 3}$ is a regressor vector independent ${}^{b}x_i$ $p_i(p_i) = -K_{c}(p_i){}^{c}t_{b} \in \mathbb{R}^3$ is a certain vector, and $K_{c}(p_i) = [k_1 - u_i k_3 \;\; k_2 - v_i k_3 \;\; 0] \in \mathbb{R}^{3\times 3}$.

3. CONTROL OBJECTIVE

The objective is to design a adaptive controller for torque of the robot such that the camera (or the end-effector) to $\{C_r\}$.In addition, since the monocular camera only provides 2D information, Euclidean reconstruction of the features on image plane is desired. Homography decomposition enables separate control of rotation and translation for robot motion. More specifically, the mismatch rotation R is used for designing a rotation control part, while a combination of the depth ratio α and pixel coordinate p_i of a feature point P_i is utilized to design a translation control part. the control objective is expressed as the desire to achieve: $R \to I, \alpha \to 1, p \to p_r$ as $t \to \infty$ That is, if $R \to I$ and $\alpha \to 1$,one has $R = R_{r}$ and $z = z_{r}$. Then ,with(6),if $p = p_{r}$,one has ${}^{c}x = {}^{c_r}x$, the camera reaches its reference pose.

3.1 Error Definition The matrix R denotes the rotation between $\{C\}$ and $\{C_r\}$. we define a rotation error-like signa l$e_w(t) \in \mathbb{R}^3$ as follows

$$e_\omega(t) = u(t)\,\theta(t). \tag{11}$$

where $u(t) \in \mathbb{R}^3$ denotes a unit rotation axis and $\theta(t) \in \mathbb{R}$ represents the rotation about that is confined to the following region[2].

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

The parameters $u(t)$ and $v(t)$ are related to the rotation matrix R that can be obtained by the following expressing[3]

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

Where R can be obtained by homography decomposition, and$[\xi]_\times \in \mathbb{R}^{3\times 3}$ denotes the skew-symmetry matrix by the following expression

$$[\xi]_\times \triangleq \begin{bmatrix} 0 & -\xi_3 & \xi_2 \\ \xi_3 & 0 & -\xi_1 \\ -\xi_2 & \xi, & 0 \end{bmatrix}, \quad \forall \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}. \tag{14}$$

Differentiating e_w with respect to time t yields[2]

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

where ${}^{b}\omega \in \mathbb{R}^3$ denotes an angular velocity of the camera expressed in the world frame $\{B\}$,and $L_\omega \in \mathbb{R}^{3\times 3}$ is defined by [6]

$$L_\omega \triangleq \mathrm{I}_3 - \frac{\theta}{2}[u]_\times + \left(1 - \frac{\sin c(\theta)}{\sin c^2\left(\frac{\theta}{2}\right)}\right)[u]_\times^2. \tag{16}$$

where I_i denotes an $i \times i$ identity matrix, and the function $\sin c(\cdot)$ is defined as .The matrix can be invertible for all $\theta \in (-\pi, \pi)$.

Let $e_v \in \mathbb{R}^3$ denote the translation error, which expressed as

$$e_v = p_e(t) - p_{er}. \tag{17}$$

where $p_e(t), p_{er} \in \mathbb{R}^3$ are pixel coordinate defined as

$$\begin{cases} p_e = \begin{bmatrix} u(t) & v(t) & -\ln(z(t)) \end{bmatrix}^T, \\ p_{er} = \begin{bmatrix} u_r & v_r & -\ln(z_r) \end{bmatrix}^T. \end{cases} \tag{18}$$

Then, the translation error $e_v = [u(t) - u_r \;\; v(t) - v_r \;\; -\ln(\alpha)]^T$, and the depth ratio α is obtained by homography decomposition, one gets

$$\dot{e}_v = \frac{1}{z}A_{e}(p)\begin{bmatrix} {}^{b}v \\ {}^{b}w \end{bmatrix}. \tag{19}$$

in which ${}^{b}v \in \mathbb{R}^3$denotes a linear velocity of the camera expressed in the world frame $\{B\}$,and $A_{e}(p) \in \mathbb{R}^{3\times 6}$ denotes an auxiliary interaction matrix given by

$$A_{e}(p) := \underbrace{\left(K - p_0 k_3^T\right){}^{c}R_{b}}_{K_{e}}\left[-I_3, \left[{}^{b}x - {}^{b}t_e\right]_\times\right]. \tag{20}$$

with $p_0(t) := [u(t), v(t), 0]^T \in \mathbb{R}^3$,where ${}^{b}t_e \in \mathbb{R}^3$ is the translation of the end-effector frame $\{E\}$ in $\{B\}$,and K_e is an invertible matrix. After defining e_ω in (11) and e_v in (17),the pose control objective (10) becomes $e_\omega \to 0, e_{v} \to 0$ as $t \to \infty$.

3.2 Visual Servo Control Design In this paper, An n-DoF nonlinear robotic system is considered, whose dynamics are described as:

$$M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) = \tau. \tag{21}$$

where $q \in \mathbb{R}^n$ is a joint variable,$M(q) \in \mathbb{R}^{n\times n}$ is an inertial matrix,$C(q,\dot{q}) \in \mathbb{R}^{n\times n}$ is a centripetal-Coriolis matrix,$g(q) \in \mathbb{R}^n$ represents the gravitational torque. The essential properties of the system (21) to be utilizated in this paper are given as

Property 2 [12]:The matrix $\dot{M}(q) - 2C(q,\dot{q})$ is skew symmetric such that $x^{\tau}[\dot{M}(q) - 2C(q,\dot{q})]x = 0$ holds for any

$x, q, \dot{q} \in \mathbb{R}^n$.

Property 3 [12]: The inertia matrix $M(q)$ is bounded,i.e. $\lambda_{\min}\{M\}\mathrm{I} \leq M(q) \leq \lambda_{\max}\{M\}\mathrm{I}$ is true,where $\lambda_{\min}\{M\}$ and $\lambda_{\max}\{M\}$ denote the minimum and maximum eigenvalues of the matrix $M(q)$,respectively. the velocities of the end-effector can be calculated by

$$^{\mathrm{b}}v = J_{\mathrm{v}}(q)\dot{q}, {}^{b}\omega = J_{\omega}(q)\dot{q}. \tag{22}$$

where $J_{\mathrm{v}}(q), J_w(q) \in \mathbb{R}^{3\times n}$ represent the displacement and rotation components of the robot's velocity Jacobi matrix $J(q) \in \mathbb{R}^{6\times n}$,respectively. To achieve the control objective, we propose a dynamics-based control law as follows:

$$\begin{aligned}\tau = g(q) - K_a\dot{q} - \alpha J^T(q)\hat{A}_e^T(p)K_p e_{\mathrm{v}} \\ + J_{\omega}^T(q)^c R_b^T L_{\omega}^T K_{\omega} e_{\omega}.\end{aligned} \tag{23}$$

where $\hat{A}_e(p)$ is an estimate of $A_e(p)$,and $K_d \in \mathbb{R}^{n\times n}$,$K_p \in \mathbb{R}^{3\times 3}$ and $K_{\omega} \in \mathbb{R}^{3\times 3}$ are control gains.

Property 4: For any homogenous vector ρ,the product $\alpha J^T(q)A_e^T(p)\rho$ can be linearly parameterized by

$$\alpha J^T(q)A_e^T(p)\rho = \mathcal{H}^T(p,q,\rho)^{b}x + \beta(p,q). \tag{24}$$

where $\mathcal{H}(p,q,\rho) := -\alpha[K_{\mathrm{e}}^T\rho]_{\times}J_{\omega} \in \mathbb{R}^{3\times 6}$ is a regressor vector, and $\beta(p,q) := -\alpha(J_v^T - J_{\omega}^T[^{b}t_e])K_{\mathrm{e}}^T\rho \in \mathbb{R}^6$ is a certain vector.

Applying (21) and (23),the closed-loop dynamics is expressed as

$$\begin{aligned}M(q)\ddot{q} + (C(q,\dot{q}) + K_d)\dot{q} = J_{\omega}^T(q)^c R_b^T L_{\omega}^T K_{\omega} e_{\omega} \\ -\mathcal{H}^T(p,q,K_p,e_v)^{b}\tilde{x} - \alpha J^T(q)A_c^T(p)K_p e_{\varepsilon}.\end{aligned} \tag{25}$$

where $^{b}\tilde{x}(t) = {}^{b}\hat{x}(t) - {}^{b}x(t) \in \mathbb{R}^3$ is a parameter estimation error.

With (10),Design the following auxiliary filter matrices $P \in \mathbb{R}^{n\times n}$,$Q \in \mathbb{R}^{n\times 1}$ and $W \in \mathbb{R}^{n\times 1}$ [13].

$$\begin{cases}\dot{P} = -lP + \Phi^T\Phi, & P(0) = 0 \\ \dot{Q} = -lQ + \Phi^T y, & Q(0) = 0 \\ W = P^{b}\hat{x} - Q.\end{cases} \tag{26}$$

where $l \in \mathbb{R}$ is a positive design parameter.

The solution of P and Q can be calculated from (26) as

$$\begin{cases}P(t) = \int_0^t e^{-l(t-\lambda)}\Phi^T(\lambda)\Phi(\lambda)d\lambda, \\ Q(t) = \int_0^t e^{-l(t-\lambda)}\Phi^T(\lambda)y(\lambda)d\lambda.\end{cases} \tag{27}$$

From the definition of P,Q and W given in (26) and (27), we have

$$W = P^{b}\hat{x} - Q = P^{b}\hat{x} - P^{b}x + \varepsilon = P^{b}\tilde{x} + \varepsilon. \tag{28}$$

where ε is a bounded residual error.

Applying (27) and (28),we can calculated vector W online. Moreover, the use of this derived variable W in the adaptive law can raise the performance of parameter estimation. Then, one designs a regularized adaptive law

$$^{b}\dot{x} = \Gamma(\mathcal{H}(p,q,K_p,e_v)\dot{q} - \delta W). \tag{29}$$

where $\Gamma \in \mathbb{R}^{3\times 3}$ is a positive-definite diagonal matrix of learning gains. $\delta \in \mathbb{R}^{3\times 3}$ is a weighting factor.

3.3 Closed-Loop System Stability Select a Lyapunov function candidate

$$\begin{aligned}V = \tfrac{1}{2}[\dot{q}^T M(q)\dot{q} + z_r e_v^T K_p e_v \\ + e_{\omega}^T K_{\omega} e_{\omega} + {}^{b}\tilde{x}^T\Gamma^{-1b}\tilde{x}].\end{aligned} \tag{30}$$

The time derivative of V is given by

$$\begin{aligned}\dot{V} = \dot{q}^T M(q)\ddot{q} + \tfrac{1}{2}\dot{q}^T\dot{M}(q)\dot{q} + z_r e_v^T K_p\dot{e}_v \\ + e_{\omega}^T K_{\omega}\dot{e}_{\omega} + {}^{b}\tilde{x}^T\Gamma^{-1b}\dot{\hat{x}}.\end{aligned} \tag{31}$$

Multiplying both sides of (19) by z_r and using (22) yields

$$\begin{cases}z_{\mathrm{r}}\dot{e}_{\mathrm{v}} = \alpha A_e(p)J(q)\dot{q}, \\ \dot{e}_{\omega} = L_{\omega}^c R_{\mathrm{b}} J_{\omega}(q)\dot{q}.\end{cases} \tag{32}$$

Substituting (15), (29) and (32) into (31) and noting Property 3, one immediately obtains

$$\dot{V} = -\dot{q}^T K_d\dot{q} - \delta^{b}\tilde{x}^T W^{b}\tilde{x} \leq 0. \tag{33}$$

4. EXPERIMENTAL STUDIES

To verify the proposed improved adaptive HBVS method, we conducted experiments using the Franka Emika Panda, a 7-DOF collaborative robot. The camera used for the experiments was the Intel RealSense D435i.

Fig. 2. The Franka Emika Panda robot.

The internal and external parameters of the camera were obtained using the camera calibration module in Robot Operating System(ROS). The details about the robot setup can be referred to [11]. The feature points of the April Tag were detected and extracted using the tracking algorithm provided by

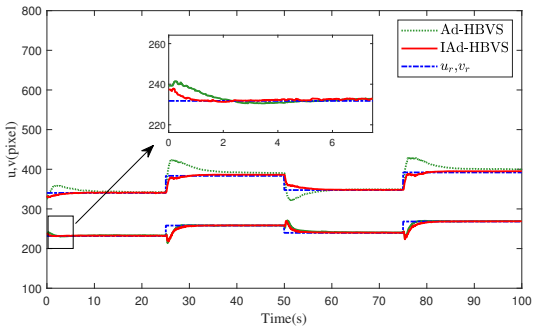

Fig. 3. Pixel tracking trajectories in the $u-$ and $v-$ axes of the image plane, respectively

ViSP, an open-source C++ visual servo platform. The control algorithm was implemented within ViSP as well. To evaluate the results and calculate the estimation error, the pose estimation algorithm in ViSP was utilized to obtain the coordinates ${}^b x$ of the feature points in the world frame $\{B\}$. The

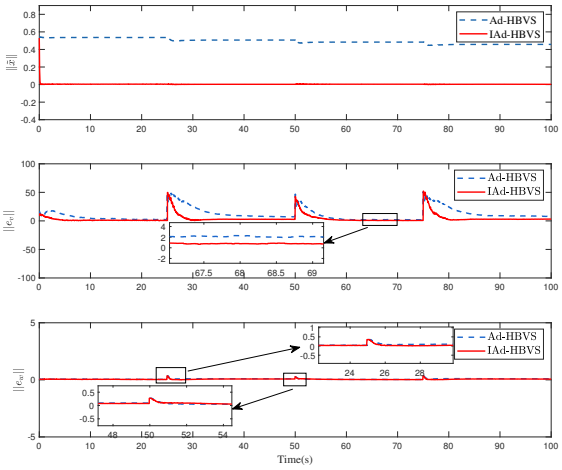

Fig. 4. The estimation error, rotation error and translation error $\|{}^b\tilde{x}\|$, $\|e_v\|$ and $\|e_\omega\|$,respectively.

estimation error ${}^b\tilde{x}$ was then calculated based on these coordinates.the robot initial joint angular position is set as $q(0) = [0.27, -4.32, -0.90, -95.10, -0.52, 90.46, 43.32]^T$.The internal and external parameters of the camera are set as

$$K = \begin{bmatrix} 612 & 0 & 322 \\ 0 & 612 & 234 \\ 0 & 0 & 1 \end{bmatrix}. \tag{34}$$

$${}^e T_c = \begin{bmatrix} 0.72 & -0.69 & 0.05 & 0.016 \\ 0.69 & 0.72 & 0 & -0.068 \\ 0.04 & 0.03 & 1 & -0.025 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{35}$$

The control parameters of the proposed IAd-HBVS with (23) and (29)are set as $K_d = [105, 66, 60, 60, 45, 45, 15]^T$, $K_p = [3 \times 10^{-4}, 3 \times 10^{-4}, 60]^T$, $K_W = [3.5, 3.5.25]^T$, $l = 88$ $\Gamma = 0.005I$,and $\delta = diag[2, 4, 2]$.

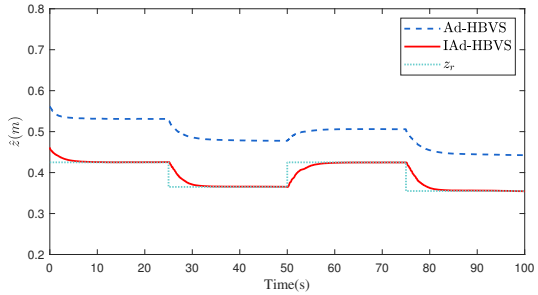

Fig. 5. The estimated and reference depth $\hat{z}$,z_r.

It is important to highlight that the scale of the pixel errors is significantly larger compared to the scale of depth ratio and rotation errors. As a result, the control gain for the pixel errors in the K_d term is much smaller compared to the other terms. In the experiments, the proposed method is compared with traditional adaptive HBVS(Ad-HBVS). The main difference between the two methods lies in the estimation law (equation 23), where the Ad-HBVS method sets $\delta = 0$ and the initial value of ${}^b x = [0.1, 0.1, 0.1]^T$ in the experiment is not specified in the provided information. Furthermore, for the experiment design, the robot is controlled to change its pose every 25 seconds. Four groups of reference points are set as specified in the experiment. This experimental design ensures that the depth represents the stepped time-varying parameters when the target is stationary.

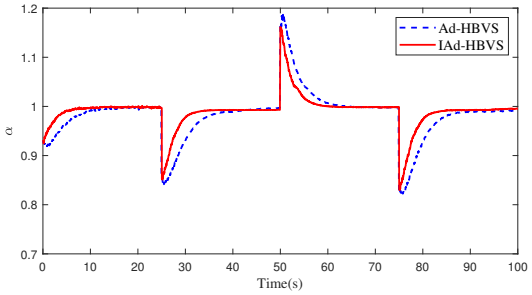

Fig. 6. The depth ratio α.

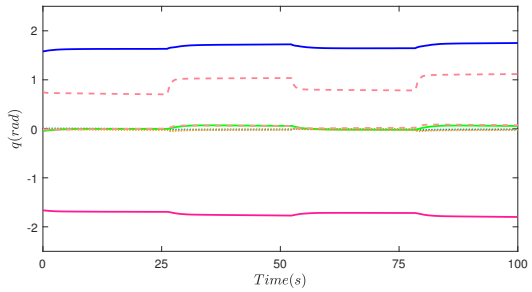

Fig. 7. The joint angular position q by the proposed controller.

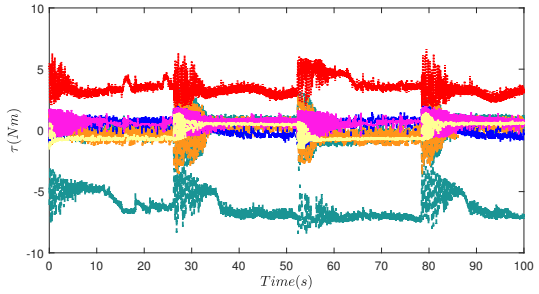

Fig. 8. The control torque τ by the proposed controller.

The experimental comparison results between the two methods are presented in figure 3. In the case of the proposed IAd-HBVS method, the pixel coordinate of feature point P_1 converges to its reference pixel position. The time-varying

depth z also converges to its reference value within the defined four periods, while the depth ratio converges to 1. figure 5 illustrates the estimation error, rotation error, and displacement error as defined in the paper. Comparing IAd-HBVS with Ad-HBVS, it can be observed that the estimation error of IAd-HBVS converges to 0 within 1 second, the rotation error $\|e_\omega\|$ reaches 0.05 rad, and the displacement error $\|e_v\|$ reaches 1 pixel. Furthermore, figure 6 and figure 7 depict the joint variable q and the control torque of the robot, respectively. Both figures show relatively small fluctuations, indicating smooth movement of the robot during the process.

5. CONCLUSIONS

In this paper, a method called IAd-HBVS is proposed to control the pose of the end actuator of a manipulator in cartesian space within the EIH system.Experimental results conducted on a 7-DoF cooperative robot demonstrate that, under the designed simple control task, the target depth is accurately estimated and the estimation error converges quickly. The stability of the system is proven using the Lyapunov stability theorem.The current work focuses on the control of the end actuator's pose in the presence of known dynamics and known internal and external camera parameters. However, future work will consider extending this method to address the tracking problem in scenarios with unknown dynamic interferences and unknown camera parameters.

References

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

(2) C. Jian, D. M. Dawson, W. E. Dixon, and A. Behal, "Adaptive homography-based visual servo tracking for a fixed camera configureuration with a camera-in-hand extension," IEEE Trans. Control Syst. Technol., vol. 13, no. 5, pp. 814–825, 2005,

(3) J. Chen, A. Behal, D. Dawson, and Y. Fang, "2.5D visual servoing with a fixed camera," in Proceedings of the 2003 American Control Conference, 2003., Jun. 2003, pp. 3442-3447 vol.4.

(4) 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.

(5) 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.

(6) 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.

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

(8) 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.

(9) X. Liang, H. Wang, Y.-H. Liu, B. You, Z. Liu, Z. Jing, and W. Chen, "Fully uncalibrated image-based visual servoing of 2DoFs planar manipulators with a fixed camera," IEEE Trans. on Cybern., vol. 52, no. 10, pp. 10 895 10 908, Oct. 2022.

(10) 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.

(11) 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.

(12) D. M. Dawson, C. T. Abdallah, and F. L. Lewis, Robot manipulator control: theory and practice: CRC Press, 2003.

(13) J. Na, Y. Huang, X. Wu, S. -F. Su and G. Li, "Adaptive Finite-Time Fuzzy Control of Nonlinear Active Suspension Systems With Input Delay," in IEEE Transactions on Cybernetics, vol. 50, no. 6, pp. 2639-2650, June 2020, doi: 10.1109/TCYB.2019.2894724.

(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 configureuration," Automatica, vol. 59, pp. 97–105, Sep. 2015.