

Research into Visual Servo Based Haptic Feedback Teleoperation

Tao Ni¹, Lingtao Huang^{1,*}, Huanfei Zheng² and Hongyan Zhang¹

¹School of Mechanical and Aerospace Engineering, Jilin University, Changchun, 130000, China

²Department of Mechanical Engineering, Clemson University, Clemson, 25633, USA

*Corresponding Author: Lingtao Huang. Email: hlt@jlu.edu.cn

Received: 23 June 2020; Accepted: 10 August 2020

Abstract: To overcome the problem that a teleoperation system loses robustness when the target moves outside the robot's visual field or it is far away from the desired position, and to improve the operability and controllability of a master-slave teleoperation system, we present an image servo based haptic feedback (ISBHF) control method for teleoperating. The ISBHF control method involves extracting target feature points and constructing image servo based virtual force. First, the image characteristics of the environment and targets are identified and extracted by a 3D reconstruction method. A composite image Jacobian matrix is used to construct virtual guidance force based on the target's image error. Then, the virtual force is provided for the operator through two joysticks which prevents the target image from drifting out of the visual field of the camera. We conducted experiments to confirm the validity and efficacy of the ISBHF control method. We found that human manipulation may cause the target image to deviate from the camera view, but the image servo based virtual force guides the operator to calibrate the target image back to the center view. The ISBHF control method can easily guide the robot to track the target and prevent image feature points deviating from the camera view. Teleoperation can be improved by combination of the robot visual sensory information and human haptic perception with the ISBHF control method. The ISBHF control method is favorable for the robot to avoid the blind operation of human and the collision with other obstacles.

Keywords: Visual servo; composite image; Jacobian matrix; virtual force; haptic feedback

1 Introduction

Master slave teleoperation robots are well suited to high-temperature, high-pressure, or suffocating environments, such as those in the deep sea, underground, and outer space [1,2]. Operators manipulate such robots with the aid of feedback in the form of images, sound, and haptic information received from the unpredictable and remote working environment without a comprehensive perception of the robot operating area [2]. In vision-based teleoperation, which has been very widely adopted, operators manipulate robot relying on their own senses through the images captured by the end-effector camera. The installation of eye-in-hand camera would provide the operator with a direct visual sense of the robot



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

position [3–6]. However, human perception tends to lack control of security and accuracy [1,2]; for instance, the operator could consider an optimized routine, which is based on the current perspective view, although he/she could find and adopt an alternative that would be the better routine immediately after. In addition, targets tend to fall outside the visual view of the camera and prove difficult to recapture in the continuously changing camera view [1,2,7]. The deviations and collisions that result from this control method adversely affect the operating efficiency and safety of the robot. These are caused by the operator's limited senses, given that the display is the only information available.

We proposed a 3D vision based kinesthetic teaching control strategy [8,9]. It is based on the artificial potential field method [10], whereby a virtual force potential field function is constructed based on the position of obstacles and targets, to make predictions regarding trajectory planning. The method uses position-based visual servo forces to manipulate a slave robot. It converts part of the vision information into force sensations that allows operators to share control while manipulating the robot. However, it needs the prediction of the target position, and the acquisition of target image data will be interrupted with the target moving out of the camera's visual field. When this situation occurs, the operator still has to rely on his/her own senses.

The image based visual servo method establishes an image Jacobian matrix, which reflects the control relationship between the image servo errors and the end-effector position variation, through the application of robot vision control theory [6,10–12]. The method focuses on the target image errors and is able to restrict the target image in the camera view. It does not need an estimate of the target position nor an accurate geometry reconstruction for the working environment [9–11]. In addition, researchers can obtain an online estimate of the image Jacobian matrix, which prevents the singularities and inevitable errors caused by camera calibration or visual depth estimation [11–15]. However, the image based visual servo method can only be applied in relatively small areas. If the target moves outside the robot's field of view or is far away from the desired position, the system will lose robustness.

Consequently, we present the ISBHF control method. It captures an image of a target in its initial position and then establishes an image Jacobian matrix relationship between the image servo errors and the robot position variation. Through the position servo based haptic force (PSBHS) control method, we construct the image servo error based virtual force (virtual force), which includes virtual repulsive force and virtual attractive force. The ISBHF control method can guarantee the target image to be restricted in the center view of camera as well as the end-effector to reach target objects. In addition, we use the RLS algorithm to estimate the Jacobian Matrix without help of camera calibration. On the other hand, the virtual force is fed back to the joysticks and is incorporated with human operation to enhance the reliability of system. As a consequence, teleoperation could be improved by combination of the robot visual sensory information and human haptic perception.

The paper is divided into six sections. Section 2 explains the control method composition and principle. Section 3 explains the method of recognizing and extracting target objects from an image, and how the acquired information is used to estimate the image servo based composite image Jacobian matrix with an RLS algorithm. Section 4 describes how the virtual force is constructed with an estimated composite image Jacobian matrix and then returned to the operator. Section 5 describes experiments that were conducted to verify the validity and efficacy of the ISBHF control method for helping the operator and preventing a target from deviating out of the visual field of the camera. Section 6 concludes this paper.

2 Composition and Principle of Control System

A visual servo based virtual force feedback teleoperation control system transforms the two-dimensional image position deviation information of a target object into force signals and transmits force signals to a

manual device as virtual attractive force and virtual repulsive force. These forces prompt the operator to manipulate a haptic control device in order to complete a teleoperation task, as shown in Fig. 1.

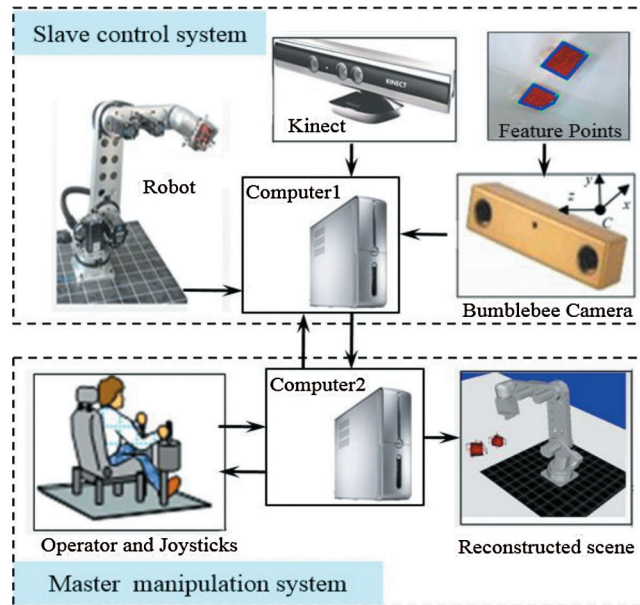


Figure 1: Visual servo based virtual force feedback teleoperation control system

The teleoperation control system consists of the master manipulation system and the slave control system. The slave control system includes a 4-DOF robot, servo motor control modules, a Bumblebee binocular camera (Point Grey Research, Canada) on an end-effector (used to identify and extract the feature points of the target object), a Kinect camera (Microsoft, US; placed over the working environment to acquire 3D data through the application of a point cloud matching technique [16–19]) and a control computer (Computer 1 in Fig. 1). The master manipulation system consists of two joysticks (Logitech WingMan force feedback joysticks) and a control computer (Computer 2 in Fig. 1). Computer 2 receives the information about the robot and image, and then rebuilds the robot and objects in virtual space. Meanwhile, Computer 2 sends the joystick information to Computer 1. The operator manipulates two 2-DOF force feedback joysticks to perform position control for the slave robot. Each joystick contains two motors so that the operator can perceive in two directions. The manipulation is undertaken using the image produced by Computer 2, which reconstructs an image of the 3D working environment.

3 Composite Jacobian Matrix for Image

The creation of virtual force requires the identification and extraction of image characteristic information for the target objects. The Bumblebee binocular camera is used to identify a target object and acquire information on feature points. In addition, the Kinect camera collects video images and depth information of field to enable the regeneration of a 3D virtual working environment, which provides a visual interface to improve the manipulation by the operator. These were specifically demonstrated in our period research in [8,9].

1) Acquisition and Extraction of Targets and Environment Graphic Information

For this study, we assume that there is a distinctive differentiation in terms of color, gradation, and texture between the target objects and working environment background. Given that the work addressed

in this section is mostly a repetition of our period research, we provide only a brief introduction to the identification, extraction and reconstruction processes.

The image information extraction and working environment reconstruction process began with the establishment of a position relationship between the robot arms, cameras and working environment. Binarization segmentation between the targets and the working environment background was acquired according to the local dynamic threshold value for the color and gradation of each pixel. Microscopic pores caused by noise were removed from the binary image by the application of an inflation-corrosion algorithm. Chaotic area less than 100 pixels were removed after each target object had been labeled in the image. Rough contour polygon areas in the targets were extracted through the application of an image segmentation and fusion algorithm. Polygon vertexes were taken as feature points of the targets. The target objects were separated from the working environment background with a geometric and color-based regional growth method. Then, 3D objects for each segment field were constructed with processed point cloud data according to the Power Crust algorithm [20,21]. All constructed 3D objects were placed together to merge the integration of the virtual environment. The extracted key points of the targets are shown in Fig. 2.

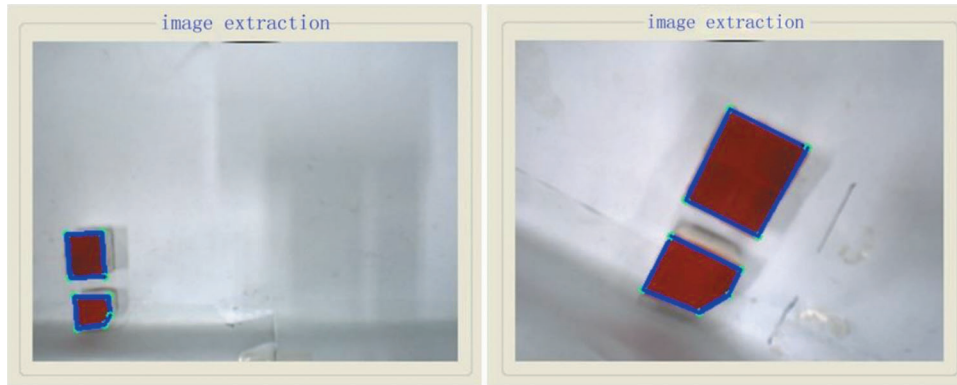


Figure 2: Initial (left) and desired (right) position of targets' image feature points

The virtual force is calculated using a composite image Jacobian matrix. The following parts describe the deduction and value estimation of the composite Jacobian matrix according to the position and orientation of the end-effector.

2) Establishment of Composite Image Jacobian Matrix

A pinhole model of the hand camera is shown in Fig. 3.

C_{xyz} denotes the local coordinate system in which the hand camera is located, while O_{1-uv} denotes the camera's two-dimensional image projecting plane. Further, λ is the focal length of the Bumblebee binocular camera, P is a key point of a target object defined as $p^c = (x_p^c, y_p^c, z_p^c)^T$ in the camera located coordinate system. N , defined as $(u_p, v_p)^T$, is the projection of point P in plane O_{1-uv} . According to the camera pinhole model, the position of a target point in plane O_{1-uv} is related to that in the camera located coordinate system, as follows:

$$\begin{bmatrix} u_p \\ v_p \end{bmatrix} = \frac{\lambda}{z_p^c} \begin{bmatrix} x_p^c \\ y_p^c \end{bmatrix} \quad (1)$$

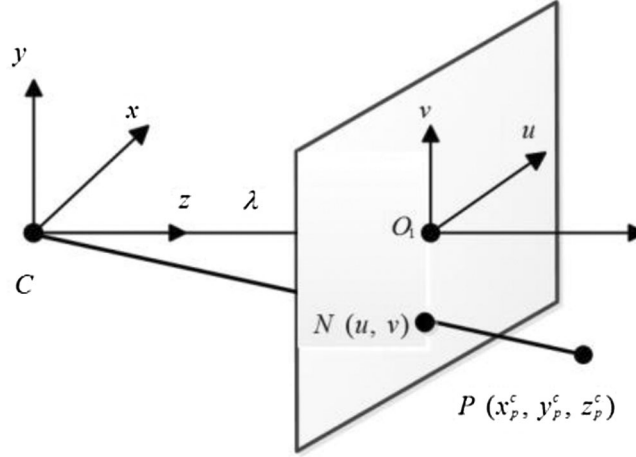


Figure 3: Pinhole model of hand camera

The derivation of the above yields:

$$\begin{bmatrix} \dot{u}_p \\ \dot{v}_p \end{bmatrix} = \frac{\lambda}{z_p^c} \begin{bmatrix} \dot{x}_p^c - \frac{\dot{z}_p^c}{z_p^c} x_p^c \\ \dot{y}_p^c - \frac{\dot{z}_p^c}{z_p^c} y_p^c \end{bmatrix} = \frac{\lambda}{z_p^c} \begin{bmatrix} 1 & 0 & -\frac{u_p}{\lambda} \\ 0 & 1 & -\frac{v_p}{\lambda} \end{bmatrix} \begin{bmatrix} \dot{x}_p^c \\ \dot{y}_p^c \\ \dot{z}_p^c \end{bmatrix} \quad (2)$$

The composite image Jacobian matrix L contributes to the relationship between an feature point's position gradient in the image plane and the hand camera's velocity relative to the robot target object coordinate system O_{xyz} . According to the definition, the relationship is described as $\dot{S} = L \dot{\xi}_c$, which can be explained as:

$$\begin{bmatrix} \dot{u}_p \\ \dot{v}_p \end{bmatrix} = \frac{\partial S}{\partial \xi_c} \left(v_{C_x}^o, v_{C_y}^o, v_{C_z}^o, w_{C_x}^o, w_{C_y}^o, w_{C_z}^o \right)^T \quad (3)$$

S represents the feature point's position in the camera's two-dimensional image projecting plane O_{1-uv} ; $\dot{S} = (\dot{u}_p, \dot{v}_p)^T$ is the velocity of the feature point projected in the two-dimensional image plane, and $\dot{\xi}_c = (v_c^o, w_c^o)^T$, where $v_c^o = (v_{C_x}^o, v_{C_y}^o, v_{C_z}^o)^T$, $w_c^o = (w_{C_x}^o, w_{C_y}^o, w_{C_z}^o)^T$, denotes the position and orientation changes of the hand camera relative to the target object coordinate system, respectively.

According to the definition of the composite image Jacobian matrix, the gradient of each directional image position is described for every dimensional position of the hand camera. Thus, we define $L \in R^{2 \times 6}$ as:

$$L = \frac{\partial S}{\partial \xi_c} = \begin{bmatrix} \frac{\partial u_p}{\partial t_{C_x}^o} & \frac{\partial u_p}{\partial t_{C_y}^o} & \frac{\partial u_p}{\partial t_{C_z}^o} & \frac{\partial u_p}{\partial r_{C_x}^o} & \frac{\partial u_p}{\partial r_{C_y}^o} & \frac{\partial u_p}{\partial r_{C_z}^o} \\ \frac{\partial v_p}{\partial t_{C_x}^o} & \frac{\partial v_p}{\partial t_{C_y}^o} & \frac{\partial v_p}{\partial t_{C_z}^o} & \frac{\partial v_p}{\partial r_{C_x}^o} & \frac{\partial v_p}{\partial r_{C_y}^o} & \frac{\partial v_p}{\partial r_{C_z}^o} \end{bmatrix} \quad (4)$$

u_p and v_p , as has been described, represent the image position of a feature point P . $t_C^o = (t_{C_x}^o, t_{C_y}^o, t_{C_z}^o)^T$, $r_C^o = (r_{C_x}^o, r_{C_y}^o, r_{C_z}^o)^T$ denotes the position and orientation of the hand camera relative to the target object coordinate system O_{xyz} , respectively.

The translational velocity of point P relative to the hand camera can be calculated from the kinematic relationship, as follows:

$$\dot{p}^c = w_o^c \times p^c + v_o^c \quad (5)$$

where v_o^c, w_o^c denotes the position and orientation changes of target object relative to the hand camera located coordinate system, respectively. This can be detailed as:

$$\begin{bmatrix} \dot{x}_p^c \\ \dot{y}_p^c \\ \dot{z}_p^c \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & z_p^c & -y_p^c \\ 0 & -1 & 0 & -z_p^c & 0 & x_p^c \\ 0 & 0 & -1 & y_p^c & -x_p^c & 0 \end{bmatrix} \begin{bmatrix} v_{Cx}^o \\ v_{Cy}^o \\ v_{Cz}^o \\ w_{Cx}^o \\ w_{Cy}^o \\ w_{Cz}^o \end{bmatrix} \quad (6)$$

According to Eqs. (2) and (6), the velocity of the image feature point P in the hand camera image plane can be described as:

$$\begin{bmatrix} \dot{u}_p \\ \dot{v}_p \end{bmatrix} = L \dot{\xi}_c \quad (7)$$

$$L = \begin{bmatrix} -\frac{\lambda}{z_p^c} & 0 & \frac{u_p}{z_p^c} & -\frac{u_p v_p}{\lambda} & \frac{\lambda^2 + u_p^2}{\lambda} & -v_p \\ 0 & -\frac{\lambda}{z_p^c} & \frac{v_p}{z_p^c} & \frac{\lambda^2 + v_p^2}{\lambda} & \frac{u_p v_p}{\lambda} & u_p \end{bmatrix} \quad (8)$$

Considering the offset between the camera and robot end-effector coordinate system, the Jacobian matrix $M \in R^{2 \times 6}$, which contributes to the relationship between the speed of the hand camera's movement and the gradient of the end-effector's posture is given as:

$$\dot{\xi}_c = M \dot{\xi}_r \quad (9)$$

where $\dot{\xi}_r = (v_{rx}^o, v_{ry}^o, v_{rz}^o, w_{rx}^o, w_{ry}^o, w_{rz}^o)^T$ contributes to the velocity vector of the end-effector relative to the target object coordinate system.

The angular velocity $\dot{q} = (\dot{q}_1, \dot{q}_2, \dot{q}_3, \dot{q}_4)$ of each joint has a Jacobian matrix relationship with end-effector posture variation rate $\dot{\xi}_r$, which could be depicted as:

$$J_{rq} = \left[\frac{\partial \xi_r}{\partial q} \right] = \begin{bmatrix} \frac{\partial \xi_{r1}}{\partial q_1} & \dots & \frac{\partial \xi_{r1}}{\partial q_4} \\ \vdots & \ddots & \vdots \\ \frac{\partial \xi_{r6}}{\partial q_1} & \dots & \frac{\partial \xi_{r6}}{\partial q_4} \end{bmatrix} \quad (10)$$

where $J_{rq} \in R^{6 \times 4}$, $q_j (j=1, L, 4)$ are the i th robot joint angles.

The relationship between the velocity vector of feature point \dot{S} and the angular velocity of joints \dot{q} can be defined as:

$$\dot{S} = J_{sq} \dot{q} \quad (11)$$

where $J_{sq} = \frac{\partial S}{\partial q} = \begin{bmatrix} \frac{\partial u_p}{\partial q_1} & \frac{\partial u_p}{\partial q_2} & \frac{\partial u_p}{\partial q_3} & \frac{\partial u_p}{\partial q_4} \\ \frac{\partial v_p}{\partial q_1} & \frac{\partial v_p}{\partial q_2} & \frac{\partial v_p}{\partial q_3} & \frac{\partial v_p}{\partial q_4} \end{bmatrix}$, $J_{sq} \in R^{2 \times 4}$, the composite image Jacobian matrix of

a single feature point is calculated based on Eqs. (8)–(11), as follow:

$$J_{sq} = \frac{\partial S}{\partial \xi_c} \frac{\partial \xi_c}{\partial \xi_r} \frac{\partial \xi_r}{\partial q} = LMJ_{rq} \quad (12)$$

Thus, the composite image Jacobian matrix for multiple feature points can be described as:

$$J_{sq, \Sigma} = \begin{bmatrix} J_{sq,1} \\ \vdots \\ J_{sq,n} \end{bmatrix} = \begin{bmatrix} L_1 MJ_{rq} \\ \vdots \\ L_n MJ_{rq} \end{bmatrix} \quad (13)$$

where $J_{sq,i}$ ($i = 1, 2, 3, 4$) represents the i th feature point's composite image Jacobian matrix, and it is real matrix, L_i is the feature point's image Jacobian matrix.

3) Estimation of Composite Image Jacobian Matrix

To determine the composite image Jacobian matrix, we require the focal length of the hand camera and the depth of image parameters. However, camera calibration and image depth accuracy greatly affect the precision of the composite Jacobian matrix values. Thus, a direct estimation of the composite image Jacobian matrix is neither efficient nor accurate for current ISBHF control strategy.

The recursive least square algorithm (RLS) exploits the exponential weighted average to estimate a composite image Jacobian matrix based on the dynamic quasi-newton method. The iteration of the composite image Jacobian matrix with the RLS algorithm can thus be described as follows:

$$\tilde{J}_k = \tilde{J}_{k-1} + \frac{(\Delta s_k - \tilde{J}_{k-1} \Delta h_k) \Delta h_k^T P_{k-1}}{\rho + \Delta h_k^T P_{k-1} \Delta h_k} \quad (14)$$

$$\tilde{P}_k = \frac{1}{\rho} \left(\tilde{P}_{k-1} - \frac{\tilde{P}_{k-1} \Delta h_k \Delta h_k^T \tilde{P}_{k-1}}{\rho + \Delta h_k^T P_{k-1} \Delta h_k} \right) \quad (15)$$

where $\Delta h_k = [\Delta q_{1,k}, \Delta q_{2,k}, \Delta q_{3,k}, \Delta q_{4,k}]^T$, $\Delta q_{i,k}$ ($i = 1, 2, 3, 4$) represents the displacement differentials of the robot joints in each step; Δs_k is the error matrix between the desired image position and the actual position; ρ is the forgetting factor, and P_k is the covariance matrix. A forgetting factor between 0 and 1 can be set. However, if ρ is equal to 1, the new information is averaged with the previous data, implying that the value stability of \tilde{J}_k would fluctuate greatly with the effect of noise as a result of the infinite divergence of the composite image Jacobian matrix. This would lead to severe safety problems in the control of the position of the robot joints. If ρ approaches 0, previous data is given less weight in the Jacobian estimation, which leads to more iterative steps and more time. Finally, the value of ρ is determined to be 0.5.

To reduce the iteration of the composite image Jacobian matrix, J_k is directly estimated with four steps of linear independent displacement control Δq_1 , Δq_2 , Δq_3 , Δq_4 to initialize the matrix. The initialization of J_k is:

$$J_0 = \begin{bmatrix} \frac{\Delta u_1}{\Delta q_1} & \frac{\Delta v_1}{\Delta q_1} & \dots & \frac{\Delta u_n}{\Delta q_1} & \frac{\Delta v_n}{\Delta q_1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \frac{\Delta u_1}{\Delta q_4} & \frac{\Delta v_1}{\Delta q_4} & \dots & \frac{\Delta u_n}{\Delta q_4} & \frac{\Delta v_n}{\Delta q_4} \end{bmatrix} \quad (16)$$

where Δq_1 , Δq_2 , Δq_3 and Δq_4 are the respective displacements of every robot joint in each of the four steps. Δu_i and $\Delta v_i (i = 1, \dots, n)$ are the finalized respective image error differentials of the target feature points in the image plane. The dynamic estimation of the composite image Jacobian matrix is conducted according to Eqs. (14) and (15) after the initialization.

4 Proposed Method for Visual Servo Based Force Control

We incorporated a visual servo method into the position-based force guidance strategy to construct the virtual force for the teleoperation system based on image servo error. We describe the construction of the virtual attractive and repulsive force model in this section. This method guarantees the convergence of the end-effector's movement trajectory to the desired position. In addition, the visual servo force resists the deviation of the target image from center field of the camera view when the robot end-effector is guided with the virtual force.

1) Image Servo Based Virtual Repulsive Force Model

The image acquisition and multiple feature point extraction for target objects are conducted with the Bumblebee camera as part of the control process. S denotes the current position of the target's multiple feature point defined in Section 3.2. S_D means the desired position in image plane. They are extended as follows:

$$S = [s_1, \dots, s_i, \dots, s_n] = \left[\begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \dots \begin{bmatrix} u_i \\ v_i \end{bmatrix} \dots \begin{bmatrix} u_n \\ v_n \end{bmatrix} \right] \quad (17)$$

$$S_D = [s_{1,D}, \dots, s_{i,D}, \dots, s_{n,D}] = \left[\begin{bmatrix} u_{1,D} \\ v_{1,D} \end{bmatrix} \dots \begin{bmatrix} u_{i,D} \\ v_{i,D} \end{bmatrix} \dots \begin{bmatrix} u_{n,D} \\ v_{n,D} \end{bmatrix} \right] \quad (18)$$

where $s_i = [u_i, v_i]^T$, $u_i \in (u_m, u_M)$, $v_i \in (v_m, v_M)$. u_m and u_M are the minimum and maximum boundaries of pixels u respectively. v_m and v_M are the minimum and maximum boundary values of pixels v respectively. a_u and a_v denote the size of the image margin and both are positive. $C = \{(u, v) | u \in (u_m, u_m + a_u) \cup (u_M - a_u, u_M), v \in (v_m, v_m + a_v) \cup (v_M - a_v, v_M)\}$ is defined as the marginal area of the feature point's image, and C means the shadow area in Fig. 4. The image-based repulsive force equation $\vec{\nabla}_S^T V_R(S)$ for the target is constructed based on the feature points' image boundaries using the artificial potential field method.

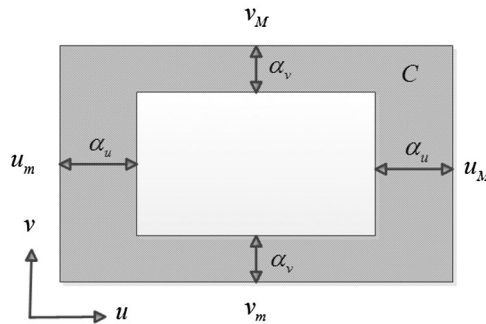


Figure 4: Image boundary diagram

The repulsive potential energy equation $V_R(S)$ describes the current potential energy of the image position [22], and provides an estimation of the stability of the target object image within the visual field of the camera. This equation is defined as:

$$V_R(S) = \begin{cases} -\ln\left(\prod_{i=1}^n \left(1 - \frac{u_i}{u_M}\right)\left(1 - \frac{u_i}{u_m}\right)\left(1 - \frac{v_i}{v_M}\right)\left(1 - \frac{v_i}{v_m}\right)\right) & s_i \in C \\ 0 & s_i \notin C \end{cases} \quad (19)$$

where $i \in \{1 \cdots n\}$ represents the number of target feature points. $s_i = (u_i, v_i)^T$, as has been defined, is the feature point's position. Define $\Delta S = S_D - S$, $\Delta s_i = \Delta s_{i,D} - s_i$, $\Delta u_i = \Delta u_{i,D} - u_i$, $\Delta v_i = \Delta v_{i,D} - v_i$. The transposed first and second gradient vector equation are obtained as Eqs. (20) and (21):

$$\vec{\nabla}_S^T V_R(S) = \left(\frac{\partial V_R}{\partial S}\right)^T = \begin{cases} [g_1 \cdots g_i \cdots g_n]^T & s_i \in C \\ 0 & s_i \notin C \end{cases} \quad (20)$$

$$\text{where } g_i = \left[\frac{u_M + u_m + 2\Delta u_i}{(u_M + \Delta u_i)(u_m + \Delta u_i)}, \frac{v_M + v_m + 2\Delta v_i}{(v_M + \Delta v_i)(v_m + \Delta v_i)} \right].$$

$$\vec{\nabla}_S^{2T} V_R(S) = \begin{cases} \text{diag}\left(\frac{\partial g_1}{\partial s_1}, \dots, \frac{\partial g_i}{\partial s_i}, \dots, \frac{\partial g_n}{\partial s_n}\right) & s_i \in C \\ 0 & s_i \notin C \end{cases} \quad (21)$$

$$\text{where } \frac{\partial g_i}{\partial s_i} = \begin{cases} \text{diag}\left(\frac{1}{(u_M + \Delta u_i)^2} + \frac{1}{(u_m + \Delta u_i)^2}, \frac{1}{(v_M + \Delta v_i)^2} + \frac{1}{(v_m + \Delta v_i)^2}\right) & s_i \in C \\ 0 & s_i \notin C \end{cases}. \quad \vec{\nabla}_S^{2T} V_R(S) \text{ is}$$

positive definite, so the Eq. (19) is convex. $V_R(S)$ increases significantly when feature points deviate from the desired position in area $s_i \in C$, and becomes zero in $s_i \notin C$. As mentioned above, C means the shadow area in Fig. 4. Without considering C , the curve of the Eq. (19) is shown as in Fig. 5.

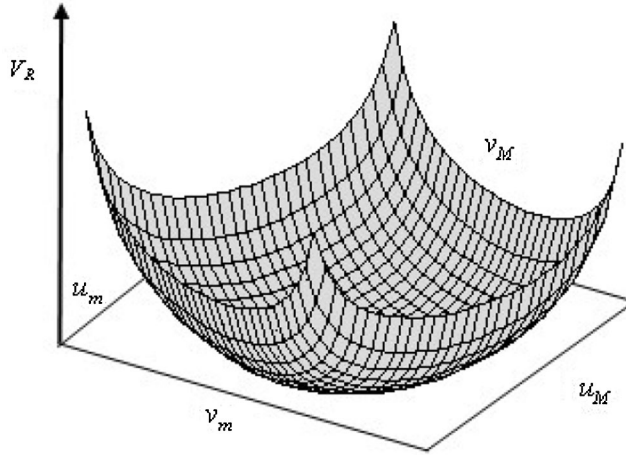


Figure 5: Repulsive potential energy diagram of image feature points

With the negative gradient of the potential equation and composite Jacobian matrix for the image, the image servo based virtual repulsive force F_R is defined as:

$$F_R = -K_R J_{rq} J_{sq}^+ \sum \vec{\nabla}_S^T V_R(S) \quad (22)$$

where K_R is the positive matrix of gain, $J_{sq,\Sigma}^+$ is the pseudo inverse of $J_{sq,\Sigma}$, which is estimated using the RLS algorithm from \tilde{J}_k . The repulsive force Eq. (22) decreases as the points move from marginal limit to the center field of the camera view. It ensures all the feature points located in the inner field of view.

2) Image Servo Based Virtual Attractive Force Model

The image servo based virtual attractive force F_A is constructed with the errors between the desired feature points' position S_D and the current position S on the basis of the artificial potential field [22]. The attractive potential energy equation is defined as:

$$V_A(S) = \begin{cases} \frac{1}{2} \sum_{i=1}^n \|s_i - s_{i,D}\|^2 & s_i \in C \\ 0 & s_i \notin C \end{cases} \quad (23)$$

The transposed gradient vector equation is obtained as:

$$\vec{\nabla}_S^T V_A(S) = \left(\frac{\partial V_A}{\partial S} \right)^T = \begin{cases} [s_1 - s_{1,D}, \dots, s_i - s_{i,D}, \dots, s_n - s_{n,D}]^T & s_i \in C \\ 0 & s_i \notin C \end{cases} \quad (24)$$

$$\vec{\nabla}_S^{2T} V_R(S) = \begin{cases} \text{diag}(I_{2 \times 2,1}, \dots, I_{2 \times 2,i}, \dots, I_{2 \times 2,n}) & s_i \in C \\ 0 & s_i \notin C \end{cases} \quad (25)$$

$\vec{\nabla}_S^{2T} V_R(S)$ is positive definite, so the Eq. (23) is convex. $V_A(S)$ increases when feature points deviate from the desired position in area $s_i \notin C$ and becomes zero in $s_i \in C$. The position of the target feature point's image and robot end-effector approaches the desired position under the guidance of the attractive force:

$$F_A = -K_A J_{rq} J_{sq,\Sigma}^+ \vec{\nabla}_S^T V_A(S) \quad (26)$$

where K_A is the positive matrix of gain.

3) Analysis of Image Servo Based Teleoperation Model

The image servo based virtual attractive force and virtual repulsive force are incorporated through the adaptive coefficients, ζ_1 and ζ_2 . The forces are fed back to the joysticks in the form of haptic signals to guide the operator to control the robot. The feedback guidance force equation can be depicted as:

$$F_{\Sigma} = \zeta_1 F_R + \zeta_2 F_A \quad (27)$$

The operator manipulates the slave robot to reach the desired position with image servo guidance information and control signals from joysticks, as is shown in Fig. 6.

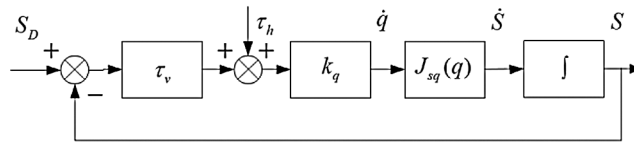


Figure 6: Block diagram of ISBHF control

τ_h and τ_v are the respective torques mapped from the operator manipulating force and the feedback virtual force. The incorporated torques on manipulator joints are used to predict velocity and drive the slave robot. k_q is the impedance matrix. The state equation of control system could be depicted as:

$$\Delta \dot{S} = J_{sq} k_q (\tau_v + \tau_h) \quad (28)$$

$$\tau_v = -\zeta_1 K_R J_{sq}^+ \vec{\nabla}_S^T V_R(S) - \zeta_2 K_A J_{sq}^+ \vec{\nabla}_S^T V_A(S) \quad (29)$$

In the image servo guidance system, the virtual force is generated and is fed back to the joysticks to guide the operator, so while the robot end-effector approaches the target, the target image remains in the center field of the camera view. Namely, the operator does not actively operate the joystick, but the generated virtual force can impel the joystick to guide the operator. $\vec{\nabla}_S^T V_R(S)$ works only in the marginal area $s_i \in C$.

5 Discussion and Experiments

In the experiments, the end-effector of a 4-DOF robot was fitted with a Bumblebee binocular camera that was used to identify and extract the image feature points of a red target object. A Kinect camera was installed near the robot to simultaneously acquire a 3D view of the working environment. The graphic computer (Computer 2 in Fig. 1) reconstructed the 3D scene to provide visual guidance, while the computer (Computer 1 in Fig. 1) determined appropriate virtual force to provide haptic guidance. Two joysticks were used to control the teleoperation robot and to provide the operators with virtual force signals. The operator operated the 4-DO robot by the joysticks according to the reconstructed virtual scene and virtual force. Figs. 7 and 8 shows the platform and the interface working environment, respectively.



Figure 7: Experiment setup in slave side

The ISBHF control method is used to define the desired position and orientation of the robot at the position where the end-effector reached the target object, and the target image feature points move into the center of the camera view. The error threshold $|\Delta s_i^*|$ is restricted to 3.0. The robot tracks the targets as the camera collected and extracted the image feature information, until $|\Delta s_i| < |\Delta s_i^*|$. The image servo based virtual force can guarantee that the target object image is constrained to the center of the camera view.

1) Estimation of Virtual Repulsive Force and Virtual Attractive Force

In order to validate the respective efficacy of the virtual repulsive force and virtual attractive force, we compared the image trajectories caused by virtual repulsive force and virtual attractive force, regardless of the effective area function $V_R(S)$ and $V_A(S)$, conducted experiments under only the virtual repulsive force or attractive forces guidance without human manipulation, as shown in Fig. 9.

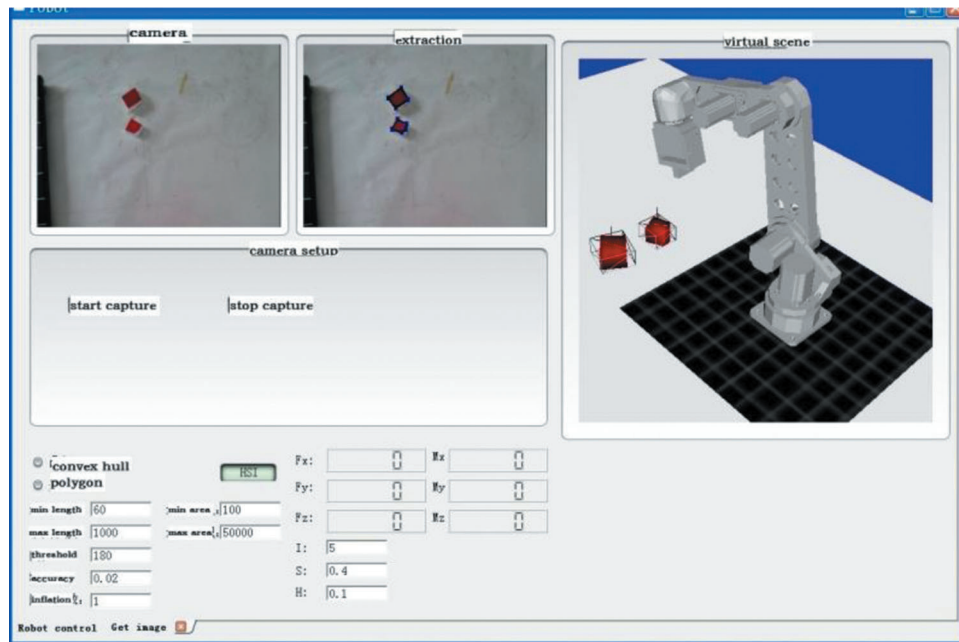


Figure 8: Interface in master side

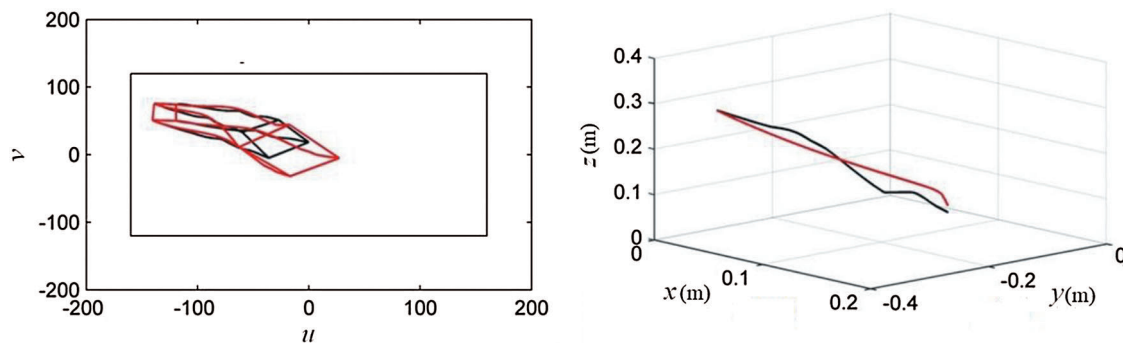


Figure 9: Comparison of feature points (left) and end-effector (right) trajectories between the virtual repulsive force and virtual attractive force

The black curves denote the results guided by the virtual repulsive force, and the red curves mean the results guided by the virtual attractive force in Fig. 9. The coordinate (feature points (left) in Fig. 9) records the trajectories of four feature points in the camera image plane, and the rectangle represents the view boundary lines of the camera. When the feature points' desired position is fully extended in the camera view, the red curves show smooth feature points and end-effector trajectory. It demonstrates well guidance in z-depth direction by the virtual attractive force.

As a contrast, the control guided by the virtual repulsive force shows instable and divergent trajectory results when all the feature points' desired position is extended in a full view. It will cause great and instable guidance forces by every single point in the initial control period according to the definition of repulsive force equation. On the other hand, the virtual attractive force will lose its function if the feature points' desired position is set too close to the center point.

According to teleoperation model in Section 4.3, we needed to incorporate the repulsive and attractive forces. Finally, we made a compromise and set the desired position to be $(-41.72, 11.52)$, $(-1.187, 40.21)$, $(-2.652, -39.82)$ and $(40.27, -3.635)$ through several extensive tests. They can satisfy both the attractive and repulsive force guidance control. In term of the curves in Fig. 9, the feature points and end-effector trajectory can be guided into the desired position with both the attractive and repulsive force guidance control.

2) Comparison between PSBHF and ISBHF

We made a contrast experiment between the PSBHF control method and the ISBHF control method without human manipulation. The obtained trajectories of the target feature points in camera image plane and end-effector trajectory are shown in Fig. 10.

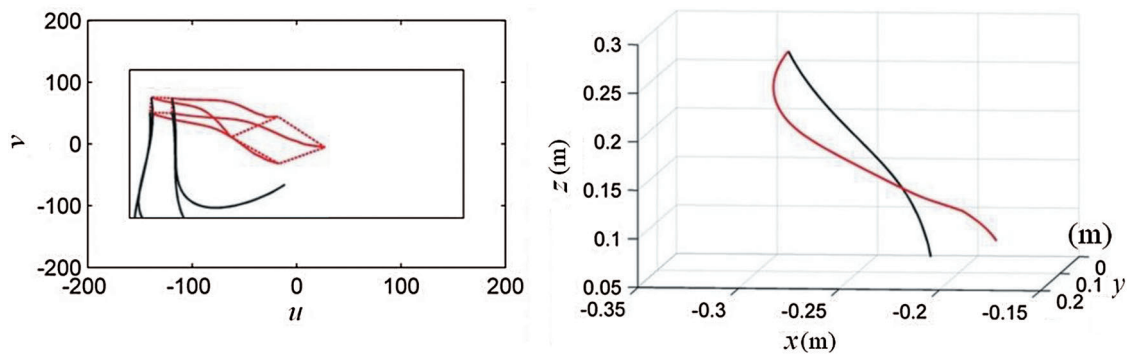


Figure 10: Comparison of feature points (left) and end-effector (right) trajectories between PSBHF and ISBHF

The black curves represent the behavior of feature points and end-effector trajectory under PSBHF control method, while the red ones record the guidance results under ISBHF control method without human manipulation. According to the end-effector trajectories in Fig. 10, both control methods can guide robot to reach the target object, but they show different trajectories of end-effector. In term of the feature points in Fig. 10, the feature points deviate from the camera visual range in the v direction, and they are limited in the -120 pixels by the PSBHF control method. The PSBHF control method cannot constrain the target image during manipulation. On contrary, the feature points' position is calibrated and restricted in the center of the camera field view by the ISBHF method. As a result, the incorporation of repulsive and attractive forces can restrict target image in camera view and guide robot to reach the target without the prediction for target object position.

3) Estimation of Shared Control Method

The image servo based control method may come across with challenge that camera misses the feature points' information. However, we are able to recapture the image of target and manipulate the robot such that it is possible to approach the target based on operator perception.

The shared control method is the ISBHF control method with human manipulation based master-slave teleoperation. In order to reflect the effectiveness of visual servo forces and involvement of human manipulation (the human manipulation forces (top) in Fig. 11), we drove the robot in a way that initially the feature points of image deviated from camera visual range, which is shown in Fig. 11 from 0 to 10 s and then was corrected back into the camera visual range by human manipulation, which is shown in Fig. 11 from 10 to 22 s.

The manipulation by ISBHF is separated into three periods (the visual guidance forces (bottom) in Fig. 11). Firstly, the operator manipulates the robot and makes the feature points deviate from camera

view intentionally, which is shown in Fig. 11 from 0 to 10 s. Then, the operator corrects the feature points' image back into camera view field, which is shown in Fig. 11 from 10 to 20 s. The last period is mainly affected by the incorporated image servo forces and assisted with little human manipulation, which is shown in Fig. 11 from 20 to 30 s.

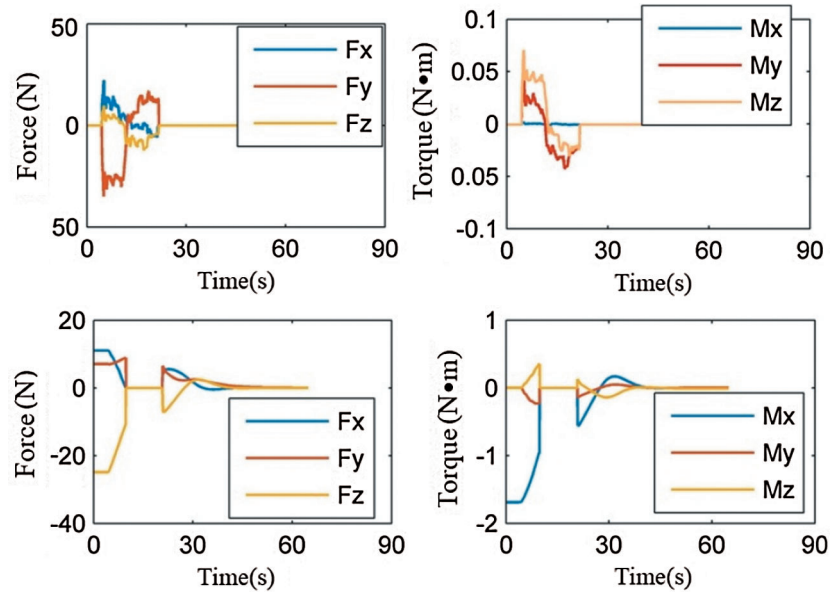


Figure 11: Human manipulation force (top) and visual force (bottom) under ISBHF

As shown in Fig. 12, the multi-colored curves and red curve represent the behavior of feature points and end-effector trajectory only under human manipulation, respectively, and the black ones record the guidance results under human manipulation with ISBHF control method.

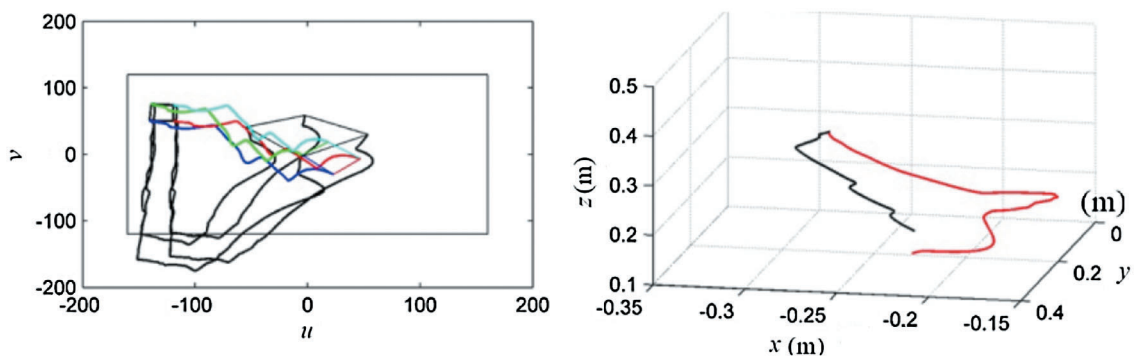


Figure 12: Comparison of feature points (left) and end-effector (right) trajectory between human manipulation and ISBHF

In term of the feature points (left) in Fig. 12, when incorporating human manipulation force with image servo based virtual force, we can use both to affect feature points' position. Even though human manipulation may cause the target image to deviate from camera field view, the image servo based virtual force can guide the operator to calibrate the view back to the center view. When missing the visual sensory information, the operator can rely on the virtual force to correct robot camera view.

Comparing the red and black curves (the end-effector (right) in Fig. 12), it can be seen that the moving distance of robot end-effector is shorter by the ISBHF method, which means the robot can reach the target with less time and less posture adjustment. So the ISBHF control method can easily guide robot to track target and prevent image feature points deviating from camera view. It is favorable for robot to avoid the blind operation of human and the collision with other obstacles.

6 Conclusion

In this study, we proposed the ISBHF control method which is an image-based servo control method with virtual force, and we applied it into teleoperation. The stereo image vision method identified and extracted environment information, and then reconstructed a virtual scene to present visual guidance to the operator. A composite image Jacobian matrix was constructed based on the relationship between the image position change rate in the camera view and the end-effector position. The virtual force was created from the equation and composite image Jacobian matrix. Finally, the virtual force was fed back to the joysticks to restrict or guide the operation.

Experiments and comparisons between the ISBHF control method and PSBHF control method has verified the validity and reliability of the ISBHF control method. Human manipulation under the guidance of the virtual force can decrease the image servo error and end-effector position error. The ISBHF control method proves capable of preventing the image of the target object deviating out of the visual field of the camera. Therefore, the safety and stability of the manipulation can be improved.

Declarations: Ethical Approval and Consent to participate: Approved. Consent for publication: Approved. Availability of supporting data: The data used to support the findings of this study are available from the corresponding author upon request.

Funding Statement: This research was supported by the Key Research and Development Projects in Jilin Province (Grant No. 20200401130GX T. Ni) and National Natural Science Foundation of China (Grant No. 51575219 L.T. Huang). <http://www.nsf.gov.cn>.

Conflicts of Interest: These are no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere. The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] G. Y. Liu, X. D. Geng, L. Z. Liu and Y. Wang, "Haptic based teleoperation with master-slave motion mapping and haptic rendering for space exploration," *Chinese Journal of Aeronautics*, vol. 32, no. 3, pp. 723–736, 2019.
- [2] Y. S. Luo, K. Yang, Q. Tang, J. Zhang and B. Xiong, "A multi-criteria network-aware service composition algorithm in wireless environments," *Computer Communications*, vol. 35, no. 15, pp. 1882–1892, 2012.
- [3] M. Staniak and C. Zielinski, "Structures of visual servos," *Robotics and Autonomous Systems*, vol. 58, no. 8, pp. 940–954, 2010.
- [4] J. M. Zhang, X. K. Jin, J. Sun, J. Wang and A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15095–15115, 2020.
- [5] P. P. Kumar and L. Behera, "Visual servoing of redundant manipulator with Jacobian matrix estimation using self-organizing map," *Robotics and Autonomous Systems*, vol. 58, no. 8, pp. 978–990, 2010.
- [6] L. Y. Xiang, X. B. Shen, J. H. Qin and W. Hao, "Discrete multi-graph hashing for large-scale visual search," *Neural Processing Letters*, vol. 49, no. 3, pp. 1055–1069, 2019.

- [7] F. Liu, S. Z. Mao and H. X. Wu, "On rough singular integrals related to homogeneous mappings," *Collectanea Mathematica*, vol. 67, no. 1, pp. 113–132, 2016.
- [8] T. Ni, H. Y. Zhang, P. Xu and H. Yamada, "Vision-based virtual force guidance for tele-robotic system," *Computers & Electrical Engineering*, vol. 39, no. 7, pp. 2135–2144, 2013.
- [9] T. Ni, X. P. Li, H. Y. Zhang, P. Xu and Z. Ma, "3-D vision-based kinesthesia teaching control strategy for telerobotics," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 44, no. 1, pp. 244–247, 2013.
- [10] Y. T. Chen, J. Wang, R. L. Xia, Q. Zhang, Z. H. Cao *et al.*, "The visual object tracking algorithm research based on adaptive combination kernel," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 12, pp. 4855–4867, 2019.
- [11] X. Y. Li, Q. L. Zhao, Y. X. Li and H. H. Dong, "Binary Bargmann symmetry constraint associated with 3×3 discrete matrix spectral problem," *Journal of Nonlinear Sciences and Applications*, vol. 8, no. 5, pp. 496–506, 2015.
- [12] J. A. Piepmeyer, G. V. McMurray and H. Lipkin, "Uncalibrated dynamic visual servoing," *Robotics and Automation, IEEE Transactions on Robotics & Automation*, vol. 20, no. 1, pp. 143–147, 2014.
- [13] J. Pomares, I. Perea, C. A. Jara, G. J. Garcia and F. Torres, "Dynamic visual servo control of a 4-axis joint tool to track image trajectories during machining complex shapes," *Robotics and Computer-Integrated Manufacturing*, vol. 29, no. 4, pp. 261–270, 2013.
- [14] Z. Cao, L. Yin, Y. Fu and T. Liu, "Predictive control for visual servo stabilization of nonholonomic mobile robots," *Acta Automatica Sinica*, vol. 39, no. 8, pp. 1238–1245, 2013.
- [15] N. R. Gans, G. Hu, J. Shen, Y. Zhang and W. E. Dixon, "Adaptive visual servo control to simultaneously stabilize image and pose error," *Mechatronics*, vol. 22, no. 4, pp. 410–422, 2012.
- [16] X. Li, Y. Jiang, M. Chen and F. Li, "Research on iris image encryption based on deep learning," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 126, 2018.
- [17] K. Khoshelham, "Accuracy analysis of Kinect depth data," *ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3812, no. 5, pp. 133–138, 2011.
- [18] J. Y. Oh, H. S. Choi, S. H. Jung, H. S. Kim and H. Y. Shin, "Development of pallet recognition system using Kinect camera," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 4, pp. 227–232, 2014.
- [19] P. J. Noonan, "The design and initial calibration of an optical tracking system using the Microsoft Kinect," in *2011 IEEE Nuclear Science Sym.Conf. Record*, Valencia, Spain, 24, pp. 441–445, 2011.
- [20] A. Ajouan, C. Fang, N. Tsagarakis and A. Bicchi, "Reduced-complexity representation of the human arm active endpoint stiffness for supervisory control of remote manipulation," *International Journal of Robotics Research*, vol. 37, no. 1, pp. 155–167, 2018.
- [21] L. R. Sanchez, H. A. Marin, H. E. R. Palacios, H. V. Rios-Figueroa and L. F. Marin-Urias, "A real-time 3d pose based visual servoing implementation for an autonomous mobile robot manipulator," *Procedia Technology*, vol. 2013, no. 7, pp. 416–423, 2013.
- [22] Y. Gui and G. Zeng, "Joint learning of visual and spatial features for edit propagation from a single image," *The Visual Computer*, vol. 36, no. 3, pp. 469–482, 2020.