

Image moments-based visual servoing control of bagged agricultural materials handling robot

Han Li¹, Zhijiang Zuo^{1*}, Ruijuan Chi², Yuefeng Du², Enrong Mao²

(1. State Key Laboratory of Precision Blasting, Jiangnan University, Wuhan 430056, China;

2. College of Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Manual handling is less efficient and sometimes even hazardous to humans in many areas, for example, agriculture. Using robots in those areas not only avoids human contact with such dangerous agricultural materials but also improves working efficiency. The motion of a robot is controlled using a technique called visual servoing that uses feedback information extracted from a vision sensor. In this study, a visual servoing method was proposed based on learning features and image moments for 3D targets to solve the problem of image moments-based visual servoing. A Gaussian process regression model was used to map the relationship between the image moment invariants and the rotational angles around the X - and Y -axes of the camera frame (denoted as γ and β). To obtain maximal decoupled structure and minimal nonlinearities of the image Jacobian matrix, it was assumed two image moment features, which are linearly proportional to γ and β . Combining the four image moment features of the normalized centroid coordinates, area, and orientation angle, a 6-DOF image moment-based visual servoing controller for the agricultural material handling robot was designed. Using this method, the problem of visual servoing task failure due to the singularity of the Jacobian matrix was solved, and it also had a better convergence effect for the part of the target image beyond the field of view from the initial pose and large displacement visual servoing system. The proposed algorithm was validated by carrying out experiments tracking bagged flour in a six-degree-of-freedom robotic system. The final displacement positioning accuracy reached the millimeter level and the direction angle positioning accuracy reached the level of 0.1° . The method still has a certain convergence effect when the target image is beyond the field of view from the initial pose. The experimental results have been presented to show the adequate behavior of the presented approach in robot handling operations. It provides reference for the application of visual servoing technology in the field of agricultural robots and has important theoretical significance and practical value.

Keywords: image moment, visual servoing, Jacobian matrix, agricultural material handling robot

DOI: 10.25165/j.ijabe.20231601.7050

Citation: Li H, Zuo Z J, Chi R Q, Du Y F, Mao E R. Image moments-based visual servoing control of bagged agricultural materials handling robot. *Int J Agric & Biol Eng*, 2023; 16(1): 212–219.

1 Introduction

With the rapid development of China's agriculture, the agricultural materials are also in high demand. In 2020, China's total food production was 669.49 million t^[1] and its import was 142.62 million t^[2]. Similarly, fertilizer production was 52.5 million t in 2020^[3]. The majority of food, fertilizers, and other agricultural products are transported in bags by rail transport. Most of the bags are loaded and unloaded manually, which requires intensive labor work and large amounts of human resources. Figure 1a shows the manual handling method of handling workers at a railway station in Qingzhou, Shandong Province, China, loading and unloading fertilizer. According to the field survey, the manual handling efficiency is 600 times/h which is less efficient compare to the Kuka Palletizing robot KR 40 PA (load 40

kg) of 3000 bags/h^[4] (Figure 1b) and SIASUN Palletizing robot SRM160A (load 160 kg) of 1400 bags/h^[5] (Figure 1c). In addition, some agricultural materials are chemically corrosive, which is harmful to humans.

With the rapid development of computer technology, the increasing cost of human resources, and the strict pursuit of workers' personal safety, the application of robots has penetrated from the industrial field to many fields such as medical treatment, military, agriculture, and service industry. It puts forward higher requirements for the intelligence level and environmental adaptability of the robot. Vision sensors have become one of the most important sensors because of their rich information, wide application range, and non-contact characteristics. The introduction of vision sensors into the robot control system can increase the adaptability of the robot to the surrounding environment and broaden the application field of the robot. Visual servoing technology uses machine vision information to control the motion of the robot in a closed loop, which can overcome the uncertainty in the model (including the robot, vision system, and environment) and improve the accuracy of visual positioning or tracking. The research of visual servoing has a history of nearly 40 years. However, there are still many problems not well solved in this research. Firstly, The method of image processing is the biggest difficulty of visual servoing in terms of theoretical and practical processing speed. Secondly, for the selection of image features, the performance of visual servoing depends closely on the image features used. In addition, many

Received date: 2021-09-07 **Accepted date:** 2022-10-18

Biographies: Han Li, PhD, Lecturer, research interest: visual servoing control, Email: gracelihan@jhu.edu.cn; Ruijuan Chi, PhD, Associate Professor, research interest: vehicle mechatronics control, Email: chiruijuan@cau.edu.cn; Yuefeng Du, PhD, Associate Professor, research interest: agricultural machinery digital design, Email: dyf@cau.edu.cn; Enrong Mao, PhD, Professor, research interest: vehicle intelligent control, Email: gxy15@cau.edu.cn.

***Corresponding author:** Zhijiang Zuo, PhD, Professor, research interest: machine vision. College of Intelligent Manufacturing, Jiangnan University, P.O. Box 47, Jiangnan University, No. 8, Sanjiaohu Rd., Wuhan Economic & Technological Development Zone, Wuhan 430056, China. Tel: +86-27-84226912, Email: zzjdfcy@163.com.

control methods cannot guarantee global stability, that is, convergence to the desired location from any feasible initial one.



a. Handling worker at a railway station in Qingzhou, Shandong Province loading and unloading fertilizer



b. Kuka Palletizing robot KR 40 PA



c. SIASUN Palletizing robot SRM160A

Note: Figure 1a shows the manual handling method of handling workers at a railway station in Qingzhou, Shandong Province, China, loading and unloading fertilizer. According to the field survey, the manual handling efficiency is 600 times/h which is less efficient compare to the Kuka Palletizing robot KR 40 PA(load 40 kg) of 3000 bags/h^[4] (Figure 1b) and SIASUN Palletizing robot SRM160A(load 160 kg) of 1400 bags/h^[5](Figure 1c).

Figure 1 Efficiency comparison between the handling robot and manual handling method

The vision-guided handling robot using visual servoing technology not only reduces human contact with such dangerous agricultural materials but also improves working efficiency (Figure 1). One of the key problems of visual servoing is the selection of image features, which can be used to describe the environment and to define control tasks. The most commonly used image features are geometric features such as point features^[6], line features^[7], and ellipse features^[8]. However, in practice, it is usually hard or impossible to obtain or even track the geometric features. Moreover, these geometric features cannot describe the general information about the objects. These features are easily influenced by image noise. And when occlusion or beyond the field of view occurs, it is easy to cause servoing task failure.

Some researchers proposed visual servoing methods, such as Fourier descriptors^[9,10], wavelet coefficients^[11,12], histograms^[13,14], luminance signal^[15-17], shape descriptors^[18,19], and so on, based on global image features. These global image features consider all

image data, rather than simple geometric features, and hence have better robustness. Image moments are invariant to 2D translation, 2D rotation, and scale conversion and are not sensitive to the starting point of contour, which is of great interest to researchers. François Chaumette et al.^[20,21] was the first to give the analytical expression of the image Jacobian matrix of image moments-based visual servoing. As a visual servoing feature, the global image feature image moment has better robustness and performs well in visual servoing. However, the feature selection of the rotation control around the X - and Y -axes of the camera (ω_x and ω_y) and the visual servoing of 3D objects is still a difficult problem. Zhao et al.^[22] and Mebarki et al.^[23] improved the image moment features of rotation control around the X - and Y -axes of the camera, however, the analytical expression of the Jacobian matrix using this method is still coupled in ω_x and ω_y direction motion control, which makes the Jacobian matrix still singular in some positions. Some scholars have proposed a feature selection method combining the advantages of neural networks^[24,25], support vector regression^[26], and other intelligent algorithms to achieve complete decoupling of ω_x and ω_y directions, which solves the singularity problem in visual servoing control. Bakhavatchalam et al.^[27] introduced spatial weight into image moments-based visual servoing, which solved the problem of visual servoing task failure when the target image exceeded the field of view. Image moments selected as image features are widely used in visual servoing. Most of the research studies assumed that the image moments are in the same plane, and there exist some technical challenges for image moments to implement 3D visual servoing tasks.

Visual servoing is a very important research area in robotics. Despite many research works in this area, software that allows fast prototyping of 3D visual servoing tasks is not available. The widely used visual servoing simulation platform is based on MATLAB, combined with the Robot and Robot Vision Toolbox developed by Peter Corke^[28]. It is used for visual servoing simulation of points, lines, ellipses, and other simple geometric objects. VISP^[29] is a visual servoing platform developed in C++ on Unix Workstations that has a large library of elementary positioning tasks for various basic control features, such as points, lines, circles, spheres, cylinders, and so on, and an image-processing library for the tracking of visual cues, such as dots, segments, ellipses, splines, and so on.

In this study, Gaussian process regression (GPR)-based image features were adopted to decouple the rotational control around the X - and Y -axes of the camera frame in a six-degree-of-freedom (6-DOF) robotic visual servoing system. Combining the four image moment features of the normalized centroid coordinates, area, and orientation angle, a 6-DOF image moments-based visual servoing controller was also designed for the agricultural material handling robot. A visual servoing simulation platform based on MATLAB and V-REP (virtual robot experiment platform) was also built to achieve the visual servoing simulation of 3D targets. Finally, the proposed algorithm was validated by conducting simulations and experiments on tracking bagged agricultural materials (3D objects) in a 6-DOF robotic system.

2 Image moments-based visual servoing

2.1 Image moments-based visual servoing controller model

The control principle of visual servoing is to regulate the error vector $e(t)=s(t)-s^*$ to 0, where s and s^* are the current and the desired image features, respectively. Using the real-time feedback

of the image error, combined with the appropriate control law, the manipulator is driven to obtain the desired pose. A visual servoing controlled 6-DOF manipulator has 6-DOFs in 3D space: v_x, v_y, v_z (3D translation along the camera coordinate system axis) and $\omega_x, \omega_y, \omega_z$ (3D rotation around the camera coordinate system axis). To avoid redundancy of the visual servoing system, six image features were designed by

$$s=(s_1, s_2, s_3, s_4, s_5, s_6)^T \tag{1}$$

where, s_i ($i=1, 2, 3, 4, 5, 6$) represents a selected image feature.

The control law is classically given by

$$v_c = -\lambda L_s^+ e \tag{2}$$

where, $v_c=(v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)^T$, λ is a positive scalar gain defining the convergence speed of the control law, L_s^+ is the pseudo-inverse of the interaction matrix, and e is the image feature error.

The area of image a , the coordinates of centroid x_g and y_g , and the orientation angle α are selected as the four image features. A better choice can be obtained from these intuitive features, by just adding an adequate normalization. More precisely, it was defined as,

$$\begin{cases} a_n = Z^* \sqrt{\frac{a^*}{a}} \\ x_n = a_n x_g \\ y_n = a_n y_g \\ \alpha = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \end{cases} \tag{3}$$

where, a^* and Z^* are the area and depth of the target image at the desired pose, respectively, $a=m_{00}$; m_{pq} and μ_{pq} are the image moments and central image moments of order $i+j$ defined by

$$\begin{cases} m_{pq} = \iint_{R(t)} x^p y^q f(x, y) dx dy, p, q = 0, 1, 2, \dots \\ \mu_{pq} = \iint_{R(t)} (x - x_g)^p (y - y_g)^q f(x, y) dx dy, p, q = 0, 1, 2, \dots \end{cases} \tag{4}$$

$$\begin{bmatrix} L_{mx} \\ L_{my} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & c_x \cos \alpha \cos \beta \\ 0 & 0 & 0 & c_y (\cos \alpha \sin \beta \sin \gamma - \sin \alpha \cos \gamma) \end{bmatrix}$$

where, α is the rotational angle around the Z -axis with respect to the desired camera frame. When the camera moves to the desired pose, $\alpha=\beta=\gamma=0$, and the interaction matrices of image moment features m_x and m_y are written as

$$\begin{aligned} L_{mx} &= [0 \ 0 \ 0 \ c_x \ 0 \ 0], \\ L_{my} &= [0 \ 0 \ 0 \ 0 \ c_y \ 0]. \end{aligned} \tag{8}$$

If the values of γ and β are known, then we easily calculate the image moment features m_x and m_y in the visual servoing process. From Equation (5), the overall interaction matrix of six image features at the desired pose ($\beta=\gamma=0$) of the camera is written as

$$L^{\parallel} = \begin{bmatrix} -1 & 0 & 0 & a_n r_{11} & -a_n(1+r_{12}) & y_n \\ 0 & -1 & 0 & a_n(1+r_{21}) & -a_n r_{22} & -x_n \\ 0 & 0 & -1 & -a_n r_{31} & a_n r_{32} & 0 \\ 0 & 0 & 0 & c_x & 0 & 0 \\ 0 & 0 & 0 & 0 & c_y & 0 \\ 0 & 0 & 0 & \alpha_{ox} & \alpha_{oy} & -1 \end{bmatrix} \tag{9}$$

2.2 Estimation of γ and β based on GPR

To obtain maximal decoupled structure and minimal nonlinearities of the image Jacobian matrix, a GPR-based regression model was designed to map the relationship between the

where, $R(t)$ is part of images acquired by the camera at time t where the object projects.

The Jacobian matrix of four image moment features (x_n, y_n, a_n, α) is written as

$$\begin{cases} L_{xn} = [-1 \ 0 \ 0 \ a_n r_{11} & -a_n(1+r_{12}) \ y_n] \\ L_{yn} = [0 \ -1 \ 0 \ a_n(1+r_{21}) & -a_n r_{22} \ -x_n] \\ L_{an} = [0 \ 0 \ -1 \ -a_n r_{31} & a_n r_{32} \ 0] \\ L_{\alpha} = [0 \ 0 \ 0 \ \alpha_{ox} & \alpha_{oy} \ -1] \end{cases} \tag{5}$$

where,

$$\begin{cases} r_{11} = r_{22} = 4\mu_{11} / m_{00} - x_g y_g / 2 \\ r_{12} = 4\mu_{20} / m_{00} - x_g^2 / 2 \\ r_{21} = 4\mu_{02} / m_{00} - y_g^2 / 2 \\ r_{31} = 3y_g / 2 \\ r_{32} = 3x_g / 2 \\ \alpha_{ox} = \{5[\mu_{12}(\mu_{20} - \mu_{02}) + \mu_{11}(\mu_{03} - \mu_{21})] + x_g[\mu_{02}(\mu_{20} - \mu_{02}) - 2\mu_{11}^2] + y_g \mu_{11}(\mu_{20} + \mu_{02})\} / \Delta \\ \alpha_{oy} = \{5[\mu_{21}(\mu_{02} - \mu_{20}) + \mu_{11}(\mu_{30} - \mu_{12})] + x_g \mu_{11}(\mu_{20} + \mu_{02}) + y_g[\mu_{20}(\mu_{02} - \mu_{20}) - 2\mu_{11}^2]\} / \Delta \\ \Delta = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \end{cases}$$

To obtain maximal decoupled structure and minimal nonlinearities of the image Jacobian matrix, it was assumed two image moment invariants (to 2D translation, to 2D rotation, and to scale), which are referred to as virtual image moments and denoted as m_x and m_y , respectively, and are written as

$$\begin{cases} m_x = f_x(\gamma) = c_x \gamma \\ m_y = f_y(\beta) = c_y \beta \end{cases} \tag{6}$$

where, c_x and c_y are constants and β and γ are the rotational angles around the X - and Y -axes with respect to the desired camera frame, respectively.

The Jacobian matrix of image moment features m_x and m_y is

$$\begin{bmatrix} c_x \sin \alpha \cos \beta & -c_x \sin \beta \\ c_y (\sin \alpha \sin \beta \sin \gamma + \cos \alpha \cos \gamma) & c_y \cos \beta \sin \gamma \end{bmatrix} \tag{7}$$

image moment invariants and the rotational angles around the X - and Y -axes of the camera frame with respect to the desired pose. Using Hu's seventh-order moment invariants I_1-I_7 and the aspect ratio b/a as learning features, the nonlinear mapping modeling of γ and β based on GPR was established as follows:

$$\gamma = f_{GPRX}(I_1, I_2, I_3, I_4, I_5, I_6, I_7, b/a) \tag{10}$$

$$\beta = f_{GPRY}(I_1, I_2, I_3, I_4, I_5, I_6, I_7, b/a) \tag{11}$$

Table 1 lists the regional features and their invariant attributes.

Table 1 Regional features and their invariance relative to motions

Features	2D translation	2D rotation	Scale
Area	√	√	×
Centroid	×	√	√
Orientation angle	√	×	√
Aspect ratio	√	√	√
Hu's moment invariants	√	√	√

Note: '√' indicates that the feature has motion invariance, '×' indicates that the feature does not have motion invariance

3 Simulations

Figure 2 shows a robot visual servoing simulation platform

based on V-REP and MATLAB. Figure 3 shows the block diagram of visual servoing control based on image moments, where the robot, robot controller, camera image acquisition, and the image preprocessing module are carried out at the V-REP module, which transmits the image of the flour bag to MATLAB, and image moment calculation, the GPR regression model, image moment features, and the image moments-based visual servo controller module are implemented in MATLAB.

3.1 Regression model of γ and β

When the pose of the camera is $\beta=\gamma=0$, the robot was driven continuously to rotate the camera around the X - and Y -axes to collect the sample data of the regression model of γ and β . For Hu's seventh-order moments and aspect ratio b/a , the sampling angle interval is set as $\Delta\theta=1^\circ$, and the number of unidirectional acquisition points corresponding to γ and β is $N_n=20$. The

sampling interval was $\beta \in [-20^\circ, 20^\circ]$, $\gamma \in [-20^\circ, 20^\circ]$, and a total of 1200 sets of data were collected.

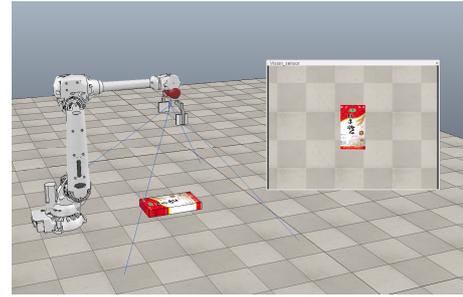


Figure 2 Robot visual servoing joint simulation platform based on R-VEP and MATLAB

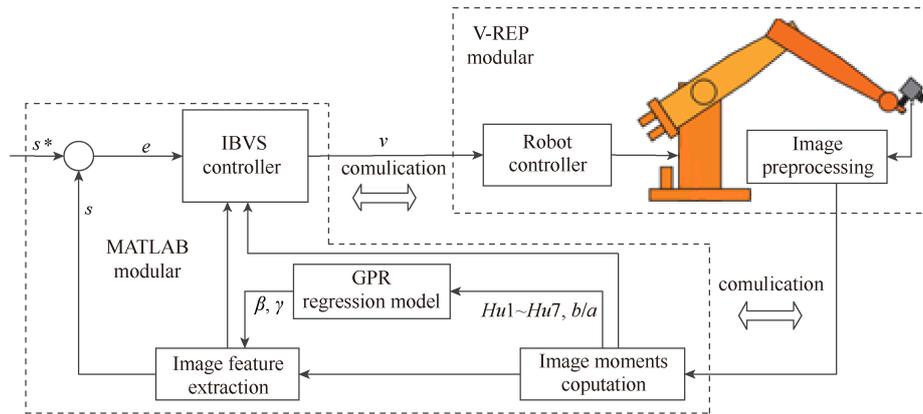


Figure 3 Block diagram of visual servoing control based on image moments

To evaluate the prediction effect, expressions for the evaluation indices such as root mean square error σ_{RMSE} , mean absolute error σ_{MAE} , and determination coefficient R^2 are written as,

$$\sigma_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (\hat{y}(i) - y(i))^2} \quad (12)$$

$$\sigma_{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}(i) - y(i)| \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y(i) - \hat{y}(i))^2}{\sum_{i=1}^N (y(i) - \bar{y})^2} \quad (14)$$

Two groups of comparative experiments were designed. Tables 2 and 3 present the effect of different regression models for Hu's seventh-order moments and aspect ratio b/a (eight features), respectively, and Tables 4 and 5 present those for Hu's seventh-order moments (seven characteristics), respectively. It understands that the addition of feature aspect ratio b/a improves the fitting effect of all regression models to a certain extent.

Table 2 Regression effects of different regression methods on γ (eight features)

Evaluation indices	Estimator γ (aspect ratio b/a and Hu's seventh-order moment invariants)					
	RQ	SE	Matern 5/2	EP	C-SVM	L-SVM
σ_{RMSE}	0.230	0.230	0.218	0.263	0.651	0.660
σ_{MAE}	0.167	0.167	0.158	0.197	0.523	0.542
R^2	1	1	1	1	1	1

Note: σ_{RMSE} : Root mean square error; σ_{MAE} : mean absolute error; RQ: Rational quadratic, SE: Squared exponential; EP: Exponential; C-SVM: Cubic-Support Vector Machine; L-SVM: Linear-Support Vector Machine.

Then GPR and SVM prediction models were compared. The

kernel functions of C-SVM and L-SVM are cubic and linear, respectively. Tables 2-5 list that the GPR model performs better than the SVM prediction model. And among the four different kernel functions (RQ, SE, Matern 5/2, and EP) of the GPR model, Matern 5/2 has the least prediction error than other kernel functions. Using the GPR model of the Matern 5/2 kernel function, σ_{RMSE} is 0.218 (estimator γ) and 0.704 (estimator β), σ_{MAE} is only 0.158 (estimator γ) and 0.439 (estimator β), and R^2 is 1. The results indicate that the prediction performance of the GPR model-adopted kernel function Matern 5/2 was the best.

Table 3 Regression effect of different regression methods on β (eight features)

Evaluation indices	Estimator β (aspect ratio b/a and Hu's seventh-order moment invariants)					
	RQ	SE	Matern 5/2	EP	C-SVM	L-SVM
σ_{RMSE}	1.223	1.224	0.704	1.108	1.113	3.484
σ_{MAE}	0.443	0.443	0.439	0.763	0.824	2.744
R^2	0.99	0.99	1	0.99	0.99	0.91

Table 4 Regression effect of different regression methods on γ (seven features)

Evaluation indices	Estimator γ (Hu's seventh-order moment invariants)					
	RQ	SE	Matern 5/2	EP	C-SVM	L-SVM
σ_{RMSE}	0.234	0.234	0.223	0.286	0.677	0.666
σ_{MAE}	0.170	0.170	0.168	0.216	0.553	0.551
R^2	1	1	1	1	1	1

Table 5 Regression effect of different regression methods on β (seven features)

Evaluation indices	Estimator β (Hu's seventh-order moment invariants)					
	RQ	SE	Matern 5/2	EP	C-SVM	L-SVM
σ_{RMSE}	1.109	1.109	0.734	1.179	1.187	3.499
σ_{MAE}	0.763	0.442	0.441	0.827	0.832	2.745
R^2	0.99	0.99	1	0.99	0.99	0.91

By comparing the regression effects of different regression characteristics and different regression functions, the final regression model for estimation of γ and β was determined, and GPR based on Matern 5/2 kernel function was used to map the relationship between Hu's seventh-order moment invariants I_1-I_7 , the aspect ratio b/a (eight features), and γ and β .

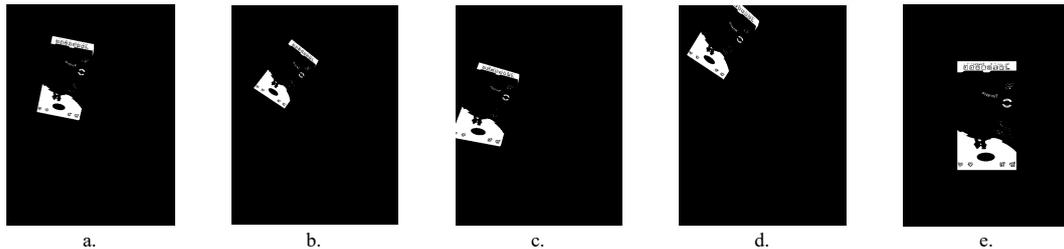
3.2 Simulation results and analysis of image moments-based visual servoing

Simulations were carried out to validate the proposed

algorithm. Four servo positioning tasks were designed, as listed in Table 6 and Figure 4, including four application scenes. The notation $\Delta=(\mathbf{t}, \theta\mathbf{u})=(d_x, d_y, d_z, \delta_x, \delta_y, \delta_z)$ was used, where \mathbf{t} is the translation part of the homogeneous matrix related to the transformation from the current to the desired frame, while its rotation part is expressed under the form $\theta\mathbf{u}$, \mathbf{u} is the unit rotation axis vector, and θ is the rotation angle around this axis. This representation is also considered in the plots to report positioning errors.

Table 6 Four servoing positioning tasks and application scene

Task	Initial pose/(m, m, m, rad, rad, rad)	Desired pose/(m, m, m, rad, rad, rad)	Space displacement/m	Application scene
1	(-0.3, 0.3, 0.3, 0, 0, 0.2)		0.520	4-DOF motion, the image does not exceed the field of view
2	(-0.3, 0.4, 0.3, 0.2, -0.2, 0.6)	(0, 0, 0, 0, 0, 0)	0.705	6-DOF motion, the image does not exceed the field of view
3	(-0.3, 0.3, 0.3, 0.2, -0.2, 0.2)		0.520	6-DOF motion, image beyond the field of view
4	(-0.4, 0.4, 0, -0.2, -0.2, 0.6)		0.825	6-DOF large displacement motion, image beyond the field of view



Note: a-d: The initial image of Task 1-4; e: the desired image of the four tasks.

Figure 4 Initial and desired images of four servoing positioning tasks

The translation errors, rotation errors, and the shape of the 3D path are the indices considered in the performance evaluation of the control law. The translation and rotation errors give the positioning accuracy, whereas the shape of the 3D path gives the spatial motion characteristics of the task and are defined, respectively, as

$$\Delta d = \sqrt{\Delta d_x^2 + \Delta d_y^2 + \Delta d_z^2}, \quad (15)$$

$$\Delta \delta = \sqrt{\Delta \delta_x^2 + \Delta \delta_y^2 + \Delta \delta_z^2}.$$

Table 7 and Figures 5-8 present visual servoing results of Tasks 1-4, respectively. Figure 5 is the simulation results of Task 1 (4-DOF motion; the image does not exceed the field of view). It infers from Figure 5a that the four image moment features are exponentially attenuated, and the trajectory of the camera in 3D space is close to a straight line (the shortest trajectory).

Figure 6 is the simulation results of Task 2 (6-DOF motion; the image does not exceed the field of view). From Figure 6a, it understands that the six image moment features show exponential

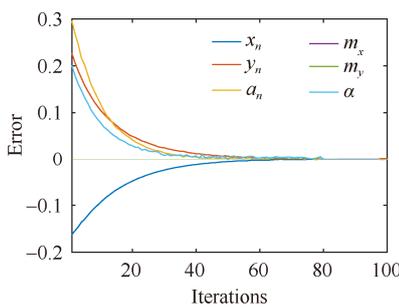
attenuation, and the trajectory of the camera's 3D space is a smooth curve. There is a slight error jitter in m_y , but it basically converges exponentially. Task 2 reflects the visual servoing results of general motion in 3D space, which proves the effectiveness of the proposed method, and has a high positioning accuracy. The final displacement positioning accuracy reached the millimeter level, and the orientation angle positioning accuracy reached the level of 0.1° .

Task 2 adds rotation angles around the X- and Y-axes to the initial pose of Task 1. The 3D space trajectory of the camera in Task 2 is a smooth curve and is longer than in Task 1 because the camera needs to adjust its pose continuously to achieve the final pose after rotation angles were added to the initial pose.

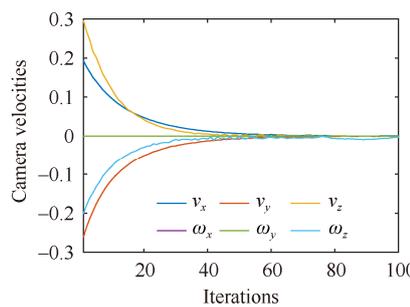
Figure 7 is the simulation results of Task 3 (6-DOF motion; the image is beyond the field of view from the initial pose). As shown in Figure 7a, it understands that features a_n and y_n increase first and then decrease, whereas all other features show exponential attenuation because, near the initial pose, part of the target image exceeds the camera field of view.

Table 7 Simulation results of four servoing positioning tasks

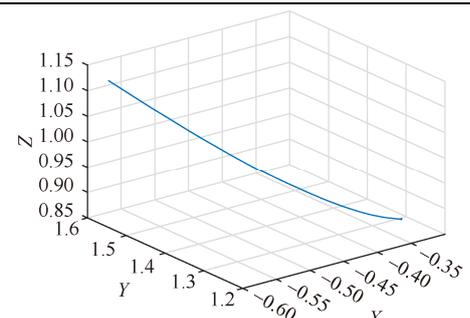
Data	Translation error/m			Rotation error/(°)			Translation error/mm				Rotation error/(°)			
	X	Y	Z	α	β	γ	Δd_x	Δd_y	Δd_z	Δd	$\Delta \delta_x$	$\Delta \delta_y$	$\Delta \delta_z$	$\Delta \delta$
Desired pose	-0.3135	1.2835	0.8600	-180.00	0.00	0.00	--	--	--	--	--	--	--	--
Results of Task 1	-0.3139	1.2832	0.8556	-179.96	0.00	0.19	0.40	0.30	4.40	4.43	0.04	0.00	0.19	0.20
Results of Task 2	-0.3141	1.2851	0.8568	-179.97	-0.11	0.00	0.60	1.60	3.20	3.63	0.03	0.11	0.00	0.12
Results of Task 3	-0.3141	1.2820	0.8571	-179.96	-0.06	0.00	0.60	1.50	2.90	3.32	0.04	0.06	0.00	0.07
Results of Task 4	-0.3139	1.2819	0.8566	-179.93	-0.06	0.00	0.40	1.60	3.40	3.78	0.07	0.06	0.00	0.09



a. Positioning error (in m and rad)



b. Camera velocity (in m/s and rad/s)



c. 3D camera trajectories (in m)

Figure 5 Simulation results of Task 1 in this study

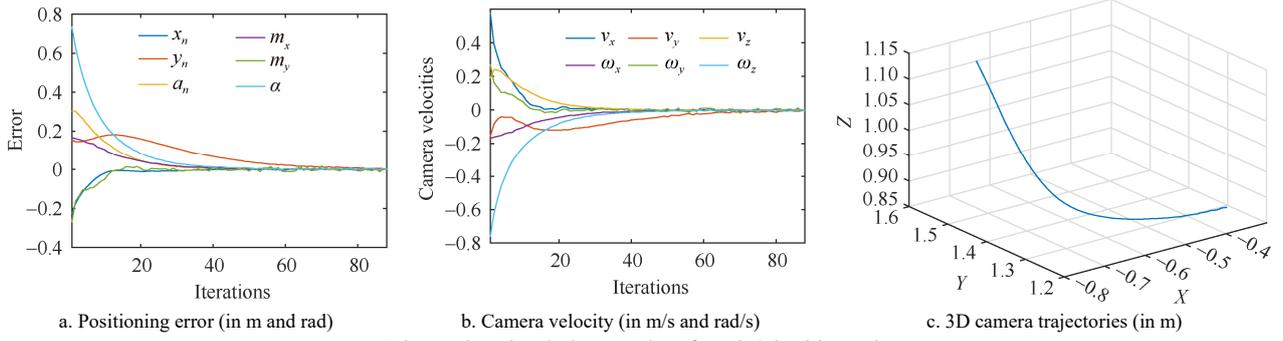


Figure 6 Simulation results of Task 2 in this study

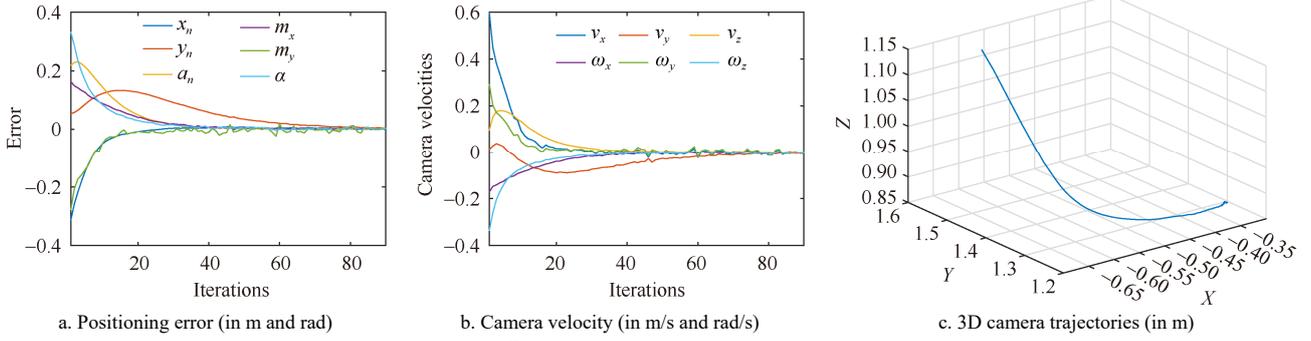


Figure 7 Simulation results of Task 3 in this study

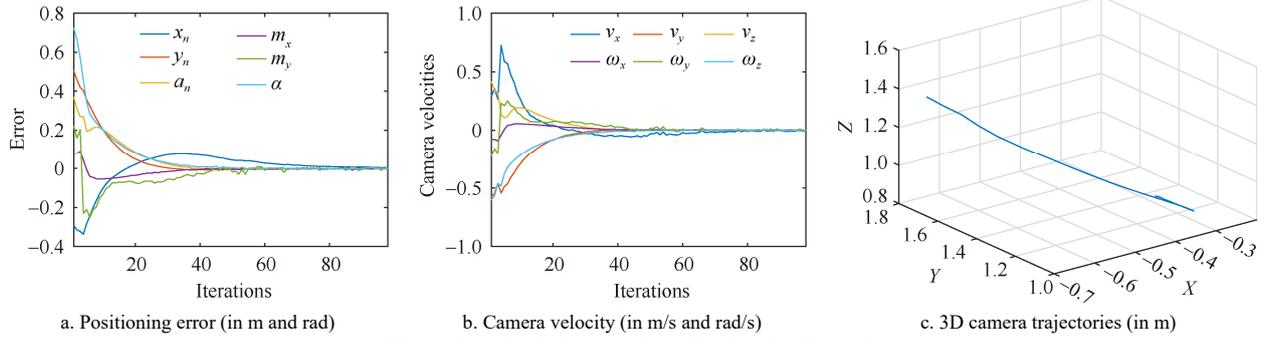


Figure 8 Simulation results of Task 4 in this study

Figure 8 is the simulation results of Task 4 (6-DOF large displacement motion; the image is beyond the field of view from the initial pose). It is a very large initial error for a visual servoing task. It can be seen from Figure 8a that image feature errors have a large oscillation because, at the beginning of the servoing task, part of the target image is beyond the field of view from the initial pose. However, due to the robustness of the servo control system, a good control effect is still achieved, converging near the desired pose.

4 Experiment and results

To verify the performance of the control laws proposed in this study, a visual servoing experimental platform (Figure 9) was built for a bagged flour-handling robot and carried out the visual servoing grabbing experiments on a 6-DOF YASKAWA ES165D robot equipped with a CCD camera mounted on the end effector. The camera frame rate is 30 frames-per-second (fps) for a resolution of 640×480 pixels, and the angle of view is 120°. The camera calibration and the hand-eye calibration were done manually. The image processing and the control law computation were performed on a 2-core 2.3-GHz Intel Core i5 PC. The code was written in C++ and MATLAB. The model of γ and β estimator was trained in advance according to the processes explained in Section 3.1 to calculate m_x and m_y . Figure 10 shows the flow chart of specific visual

servoing.

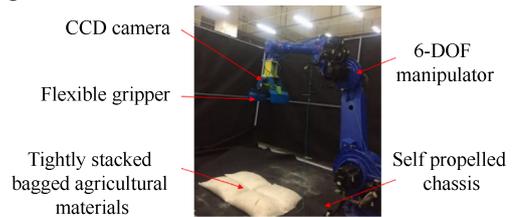


Figure 9 Visual servoing platform for a bagged flour-handling robot

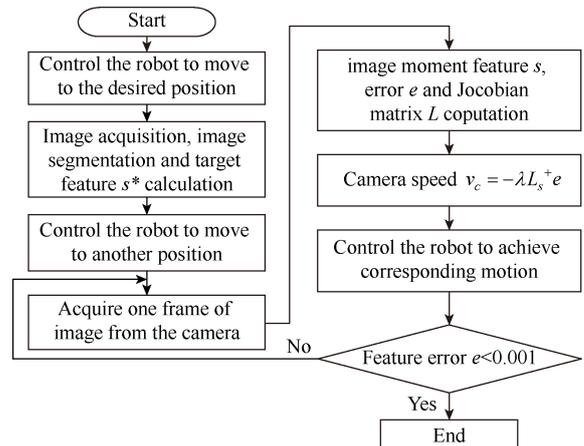


Figure 10 Visual servoing flow chart based on image moments

Two experiments were conducted to verify the performance of the control laws in 4-DOF and 6-DOF visual servoing tasks separately (Table 8 and Figure 11).

The experimental results of 4-DOF and 6-DOF visual servoing tasks, as listed in Table 9. Figures 12 and 13 are in agreement with the simulation results. The image feature errors are all exponentially attenuated, which further verifies the good characteristics of the control laws proposed in this study. Moreover, it has a good convergence effect for 3D flour bags.

Table 8 Initial and desired pose of 4-DOF and 6-DOF visual servoing

Experiment	Translation/mm			Rotation/(°)		
	<i>X</i>	<i>Y</i>	<i>Z</i>	α	β	γ
Desired pose	1509.960	0.0110	134.940	180.00	0.00	0.00
Initial pose of 4-DOF task	1625.190	-242.172	443.766	180.00	0.00	19.34
Initial pose of 6-DOF task	1785.179	-242.198	443.714	175.39	-0.89	19.57

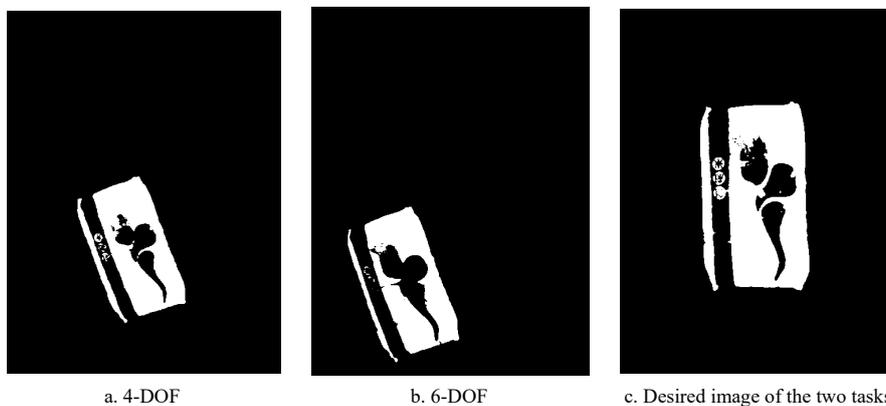


Figure 11 Initial and desired image of 4-DOF and 6-DOF visual servoing

Table 9 Experimental results of 4-DOF and 6-DOF visual servoing tasks

Task	Camera convergence pose/mm			Camera convergence pose/(°)			Translation error/mm				Rotation error/(°)			
	<i>X</i>	<i>Y</i>	<i>Z</i>	α	β	γ	Δd_x	Δd_y	Δd_z	Δd	$\Delta \delta_x$	$\Delta \delta_y$	$\Delta \delta_z$	$\Delta \delta$
4-DOF task	1509.43	0.56	135.76	-179.93	0.00	0.08	0.53	0.55	0.82	1.12	0.07	0.00	0.08	0.11
6-DOF task	1509.31	0.67	130.15	-179.91	0.46	0.24	0.64	0.66	4.79	4.88	0.09	0.46	0.24	0.53

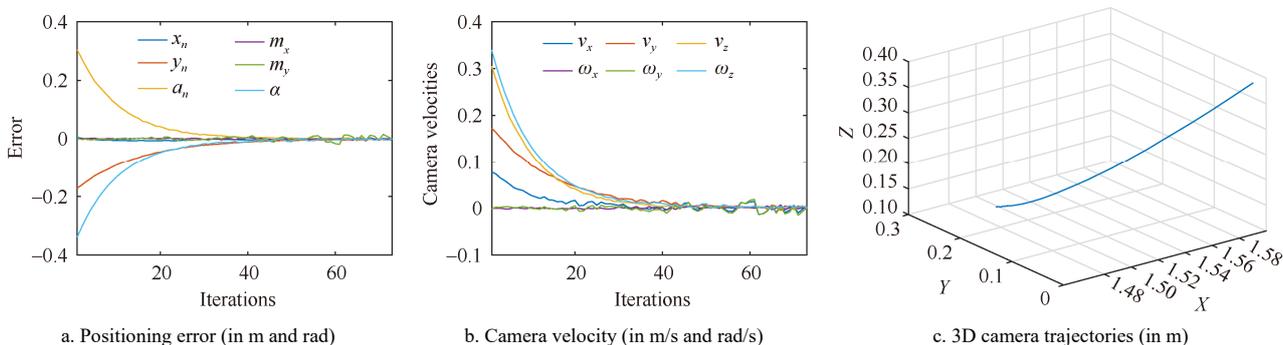


Figure 12 Experimental results of 4-DOF visual servoing using image moments

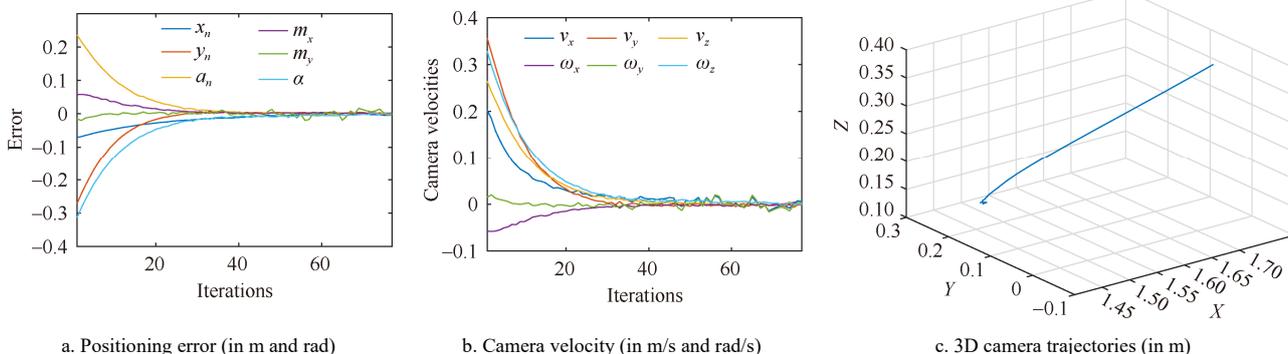


Figure 13 Experimental results of 6-DOF visual servoing using image moments

5 Conclusions

Image moments selected as image features are widely used in visual servoing. Unfortunately, most of the research studies assumed that the image moments are in the same plane, and there exist some technical challenges for image moments to implement 3D visual servoing tasks. Aiming at the problem of image

moments-based visual servoing, a visual servoing method based on learning features and image moments for 3D targets was proposed. To obtain maximal decoupled structure and minimal nonlinearities of the image Jacobian matrix, a GPR-based regression model was designed to map the relationship between the image moment invariants and the rotational angles around the *X*- and *Y*-axes of the camera frame with respect to the desired pose. This method

solved the problem of visual servoing task failure due to the singularity of the Jacobian matrix in the visual servoing system, and it also has a better convergence effect for the part of the target image beyond the field of view from the initial pose and large displacement visual servoing system (about 0.825 m). Experiments were conducted to validate the proposed algorithm tracking bagged flour in a 6-DOF robotic system. The final displacement positioning accuracy reached the millimeter level, and the direction angle positioning accuracy reached the level of 0.1° . The method has a certain convergence effect even if the target image is beyond the field of view from the initial pose. The experimental results have been presented to show the adequate behavior of the presented approach in robot handling operations.

However, the current method only studied the situation in that the desired pose of the camera is parallel to the target plane. It also needs to extract the binary image of the target, and the extraction result directly affects the visual servoing control effect.

In future works, the proposed robot visual servoing scheme will be further extended into two aspects: 1) the situation that the desired pose of the camera is three-dimensional arbitrary can be further studied; 2) visual servoing control without image processing can be studied, such as direct visual servoing control.

Acknowledgments

The authors are grateful to the editors and the anonymous reviewers for providing us with insightful comments and suggestions throughout the revision process. This work was financially supported by the Twelfth Five-Year National Science and Technology Support Program (Grant No. 2015BAD18B03). The authors appreciate the linguistic assistance provided by TopEdit (www.topeditsci.com) during the preparation of this manuscript.

[References]

- [1] Announcement of National Bureau of Statistics of China on 2020 grain output data. Available: http://www.stats.gov.cn/tjsj/zxfb/202012/t20201210_1808377.html. Accessed on [2021-09-01].
- [2] Statistics of China's grain import quantity, import amount and growth rate from 2016 to 2020. Available: <https://www.chyxx.com/shuju/202102/930106.html>. Accessed on [2021-09-01].
- [3] National data from National Bureau of Statistics of China. Available: <https://data.stats.gov.cn/easyquery.htm?cn=C01>. Accessed on [2021-09-01].
- [4] Official website of KUKA. Available: <https://www.kuka.com/zh-cn/%E4%BA%A7%E5%93%81,-a,%E6%88%90%E6%9E%9C/%E6%9C%BA%E5%99%AA%E4%BA%BA%E7%B3%BB%E7%BB%9F/%E5%B7%A5%E4%B8%9A%E6%9C%BA%E5%99%AA%E4%BA%BA/kr-40-pa>. Accessed on [2021-09-01].
- [5] Production introduce of SIASUN SRM160A. Available: <http://www.siasun.com/index.php?m=content&c=index&a=show&catid=24&id=40>. Accessed on [2021-09-02].
- [6] Li T, Zhao H, Chang Y. Visual servoing tracking control of uncalibrated manipulators with a moving feature point. *International Journal of Systems Science*, 2018; 49(9–12): 2410–2426.
- [7] Wang H, Liu Y H, Zhou D. Adaptive visual servoing using point and line features with an uncalibrated eye-in-hand camera. *IEEE Transactions on Robotics*, 2008; 24(4): 843–857.
- [8] Chaumette F, Hutchinson S. Visual servo control, Part II: Advanced approaches. *IEEE Robotics and Automation Magazine*, 2007; 14(1): 109–118.
- [9] Marturi N, Brahim T, Dembélé S, Piat N. Visual servoing schemes for automatic nanopositioning under scanning electron microscope. In: *IEEE International Conference on Robotics & Automation*, Hongkong: IEEE, 2014; pp.981–986. doi: 10.1109/ICRA.2014.6906973.
- [10] Marchand E. Direct visual servoing in the frequency domain. *IEEE Robotics and Automation Letters*, 2020; 5(2): 620–627.
- [11] Dufloy L A, Reichenhofer R, Tamadazte B, Andreff N, Krupa A. Wavelet and shearlet-based image representations for visual servoing. *The International Journal of Robotics Research*, 2019; 38(4): 422–450.
- [12] Ourak M, Tamadazte B, Lehmaan O, Andreff N. Direct visual servoing using wavelet coefficients. *IEEE/ASME Transactions on Mechatronics*, 2019; 24(3): 1129–1140.
- [13] Bateau Q, Marchand E. Histograms-based visual servoing. *IEEE Robotics & Automation Letters*, 2017; 2(1): 80–87.
- [14] Abidi H, Chtourou M, Kaaniche K, Mekki H. Visual servoing based on efficient histogram information. *International Journal of Control Automation & Systems*, 2017; 15(4): 1746–1753.
- [15] Bakthavatchalam M, Tahri O, Chaumette F. A direct dense visual servoing approach using photometric moments. *IEEE Transactions on Robotics*, 2018; 34(5): 1226–1239.
- [16] Dame A, Marchand E. Mutual information-based visual servoing. *IEEE Transactions on Robotics*, 2011; 27(5): 958–969.
- [17] Dame A, Marchand E. Using mutual information for appearance-based visual path following. *Robotics and Autonomous Systems*, 2013; 61(3): 259–270.
- [18] Wang H S, Yang B H, Wang J C, Liang X W, Chen W D, Liu Y H. Adaptive visual servoing of contour features. *IEEE/ASME Transactions on Mechatronics*, 2018; 23(2): 811–822.
- [19] Chang W C, Cheng M Y, Tsai Hong J. Image feature command generation of contour following tasks for SCARA robots employing Image-Based Visual Servoing-A PH-spline approach. *Robotics and Computer Integrated Manufacturing*, 2017; 44: 57–66.
- [20] Chaumette F. Image moments: A general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 2004; 20(4): 713–723.
- [21] Tahri O, Chaumette F. Point-based and region-based image moments for visual servoing of planar objects. *IEEE Transactions on Robotics*, 2005; 21(6): 1116–1127.
- [22] Zhao Y M, Xie W F, Liu S N. Image-based visual servoing using improved image moments in 6-DOF robot systems. *International Journal of Control, Automation, and Systems*, 2013; 11(3): 586–596.
- [23] Mebarki R, Krupa A, Chaumette F. 2-D ultrasound probe complete guidance by visual servoing using image moments. *IEEE Transactions on Robotics*, 2010; 26(2): 296–306.
- [24] Zhao Y M, Xie W F, Liu S N, Wang T T. Neural network-based image moments for robotic visual servoing. *Journal of Intelligent & Robotic Systems*, 2014; 78(2): 239–256.
- [25] Li Y X, Mao Z Y, Tian L F. Visual servoing of 4DOF using image moments and neural network. *Control Theory & Applications*, 2009; 26(10): 1162–1166. (in Chinese)
- [26] Ye G Q, Li W G, Wan H. Image moment-based visual servoing method with learning features. *Journal of South China University of Technology (Natural Science Edition)*, 2017; 45(2): 99–107. (in Chinese)
- [27] Bakthavatchalam M, Chaumette F, Tahri O. An improved modelling scheme for photometric moments with inclusion of spatial weights for visual servoing with partial appearance/disappearance. In: *2015 IEEE International Conference on Robotics & Automation (ICRA)*, Seattle: IEEE, 2015; pp.6037–6043.
- [28] Robotics toolbox in Peter Corke.com. <https://petercorke.com/toolboxes/robotics-toolbox/>. Accessed on [2022-02-20].
- [29] Marchand E, Spindler F, Chaumette F. ViSP for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robotics & Automation Magazine*, 2006; 12(4): 40–52.