Journal of Computers Vol. 35 No. 1, February 2024, pp. 167-175
doi: 10.53106/199115992024023501012

A Method for Industrial Robots to Grasp and Detect Parts of Instrument under 3D Visual Guidance

Qing-Chuan Liu[1,2,3], Xiao-Yang Zhang[1,2], Rui Fan[1,2], Wei-Min Liu[1,2], Jian-Fang Xue[1,3*]

[1] Hebei Institute of Mechanical and Electrical Technology,
Xingtai City 054000, Hebei Province, China

{qingchuan3978, xiaoyang86863, fanrui98792,
weimin78687, jianfang68789}@126.com

[2] Xingtai Intelligent Production Line and Equipment Technology Innovation Center,
Xingtai City 054000, Hebei Province, China

[3] Hebei Province Mechanical and Electrical Equipment Intelligent Sensing and Advanced Control Technology Innovation Center, Xingtai City 054000, Hebei Province, China

Received 15 December 2023; Revised 15 January 2024; Accepted 31 January 2024

Abstract. Guiding industrial robots to complete grasping tasks through machine vision is an important part of achieving autonomous robot operation. This article explores the control method of industrial robots under 3D vision, focusing on the feature that two-dimensional vision can only perform color and pose recognition but lacks depth recognition. Firstly, a high-precision point cloud registration calibration matrix solution method is proposed. Then, an improved recognition model is designed to address the issue of how vision guides robots to grasp and detect. This model integrates feature extraction and object detection modules, and describes the parameters of each module. Finally, the effectiveness of the proposed method is verified in the assembly scene of gas instruments. Finally, experimental results show that, the method proposed in this article can limit the grasping accuracy to within 2 millimeters in guiding robots to grasp detection scenes, achieving the expected effect.

Keywords: 3D vision, hand-eye calibration, intelligent recognition, instrument

1 Introduction

The industrial robot industry is currently in a period of rapid development, with a steady growth in market size. Intelligent manufacturing centered around industrial robots is continuously improving enterprise production efficiency. Robots are widely used in modern chemical plants and hardware processing systems, with a wider range of applications being their ability to grasp and detect workpieces.

The visual system endows robots with recognition ability during the grasping process. Machine vision system is a device that uses optical devices and non-contact sensors to collect feature images of target objects through the camera vision system, extract relevant feature information of target objects through point cloud algorithms, analyze and process them, and then achieve detection and control. Compared with two-dimensional vision, three-dimensional vision can accurately detect the distance between objects and cameras, provide a more comprehensive description of objects, and obtain more visual information [1]. The comprehensive measurement method of 3D camera no longer relies on the overall shape and color of rigid objects, and can achieve recognition and positioning of textureless and occluded objects.

Instruments and meters are often used to indicate important indicators and are frequently used in enterprise production, daily life, teaching demonstrations, and the automotive industry. Due to the fact that the assembly accuracy of instruments determines their usage accuracy, this article designs and implements an industrial robot instrument installation method guided by 3D vision to improve instrument assembly accuracy.

In summary, in response to the current development direction and practical needs of industry, this article focuses on the research of industrial robot instrument assembly problem guided by 3D vision. The work done is as follows:

* Corresponding Author

1) A hand eye calibration method based on point cloud registration was proposed, which incorporates constraints to minimize point cloud registration errors in the process of solving the hand eye calibration matrix and improves calibration accuracy.

2) A recognition model has been designed, which includes feature extraction and grasping detection modules, which can quickly provide robot grasping coordinates and improve the efficiency and accuracy of robot guidance.

The arrangement of each chapter in this article is as follows: Chapter 1 mainly analyzes the current development status of related industries, Chapter 2 discusses some research results of various experts and scholars in the technical field related to the industry, Chapter 3 discusses the hand eye calibration method, Chapter 4 completes the design of the recognition and detection model, Chapter 5 takes gas instrument assembly as an example to verify the feasibility of the method in this article, Chapter 6 is the conclusion section, summarizing the content of this article, At the same time, the shortcomings of this article and further research directions are also presented.

2 Related Work

Hui Xiao, for the detection of instrument thermocouples, two types of automatic motion positioning equipment, multi axis motion robotic arm and three-axis motion positioning device, were used to complete the automatic detection of the instrument, improving the detection accuracy. The positioning method of the robotic arm is worth learning from [2]. Wei Song proposed a 3D part grasping method based on monocular vision guidance, which obtains the robot's grasping information through matrix transformation between coordinate systems and system calibration, thereby achieving 3D part grasping. The position error of the robot's 3D grasping experiment can be limited to within 2 mm, and the angle error can be limited to within 2 ° [3]. Jianchun Liu calculated the three-dimensional coordinates based on the matching point parallax and camera calibration results, obtained the posture of the faucet in the robot base coordinate system, and achieved the positioning and grasping of heterogeneous components. The experimental results show that the errors in the x, y, and z directions are within ± 0.9mm, ± 1mm, and ± 1.4mm, respectively, and the deviation angle error is within ± 3.5 ° [4]. Ruiqi Gao combines two detection methods, namely 3D feature extraction of solder joints and artificial neural network models, and her 3D recognition method for small features is worth learning from [5]. Shengyuan Dai extracted point cloud features for complex edges of small parts, using multiple surface models to fit the neighborhood of sampling points to estimate curvature. Ultimately, he was able to extract complete part feature information, which has reference value for the extraction method of small part features [6].

3 Point Cloud Registration Process Guided by 3D Vision

The traditional method is to use the least squares method [7] to obtain the hand eye matrix, but due to the lack of constraints on the linear equation system, the solution results are unstable and the accuracy is low. Therefore, the article proposes a hand eye calibration method based on point cloud registration, which incorporates constraints to minimize point cloud registration errors in the process of solving the hand eye calibration matrix.

Assuming $P_1 = \begin{bmatrix} X_{C1} & Y_{C1} & Z_{C1} \end{bmatrix}^T$ is the coordinate of the chessboard corner cloud in the first pose of the robotic arm, and P_1 is used as the reference coordinate system, let $P_i^1 = \begin{bmatrix} X_{Ci}^1 & Y_{Ci}^1 & Z_{Ci}^1 \end{bmatrix}$ represent the coordinate of the chessboard corner cloud in the i -th pose of the robotic arm after registration, and let the residual function $f(x) = P_i^1 - P_1$ represent the registration error between the two point clouds. Therefore, the cost function can be obtained as:

$$F(x) = \min \frac{1}{2} \left\| f(x) \right\|^2 = \min \frac{1}{2} \left\| P_i^1 - P_1 \right\|^2 . \quad (1)$$

The registration principle diagram is shown in Fig. 1.

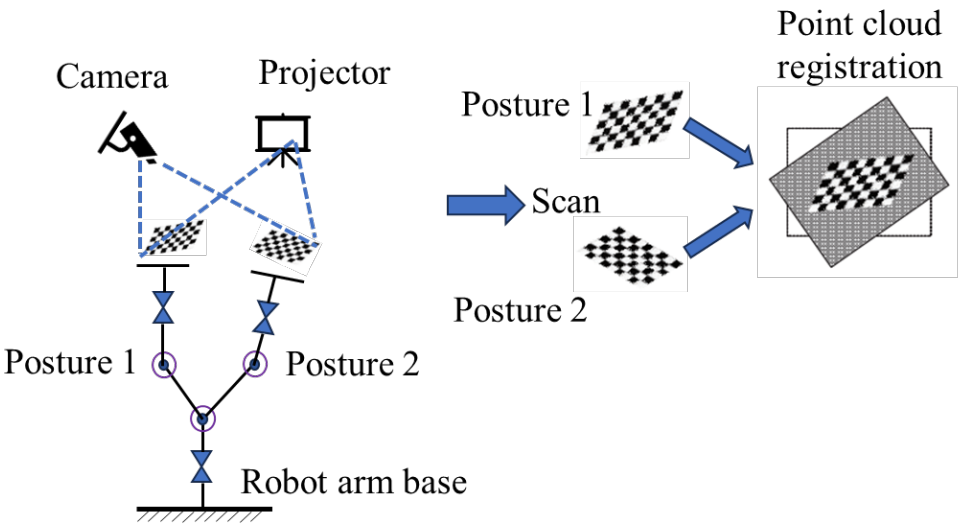

Fig. 1. Schematic diagram of point cloud registration

The formula for transforming the corner point cloud of the i -th chessboard to the baseline is as follows:

$$\begin{bmatrix} X_i^1 \\ Y_i^1 \\ Z_i^1 \\ 1 \end{bmatrix} = X^{-1} A_1 A_i^{-1} X \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}. \tag{2}$$

According to the above formula, it can be concluded that:

$$F(x) = \min \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| X^{-1} A_1 A_i^{-1} X \begin{bmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \\ 1 \end{bmatrix} - \begin{bmatrix} X_{1j} \\ Y_{1j} \\ Z_{1j} \\ 1 \end{bmatrix} \right\|^2. \tag{3}$$

Among them, j is the j -th corner point of the chessboard calibration board. Transform the hand eye matrix from the Lie group space SE (3) to the Lie algebraic space se (3), thereby constructing the cost function in the Lie algebraic space se (3) [8]. The spatial transformation under Lie group space SE (3) is represented as:

$$T = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}. \tag{4}$$

The space transformation under Lie algebraic space se (3) is represented as:

$$\xi^{\wedge} = \begin{bmatrix} \phi^{\wedge} & \beta \\ 0^T & 0 \end{bmatrix}. \tag{5}$$

The relationship between two transformations can be expressed as:

$$T = \exp\left(\xi^{\wedge}\right) = \begin{bmatrix} \exp\left(\phi^{\wedge}\right) & J\rho \\ 0^{T} & 0 \end{bmatrix}. \tag{6}$$

$$\exp\left(\phi^{\wedge}\right) = \exp\left(\theta a^{\wedge}\right) = \cos\theta I + \left(1 - \cos\theta\right) a a^{T}. \tag{7}$$

$$J = \frac{\sin\theta}{\theta} I + \left(1 - \frac{\sin\theta}{\theta}\right) a a^{T}. \tag{8}$$

In the formula, ϕ represents the rotation vector, and θ represents the modulus of the three-dimensional vector ϕ , which is the radius of rotation. a is the unit vector of ϕ, $a^{\wedge}$ is the antisymmetric matrix of the unit vector a, and ρ is the translation vector. Therefore, the cost function is ultimately changed through formulas 7, 8, and 9. The cost function F (x) is iteratively solved using the least squares method to obtain a high-precision hand eye matrix X.

$$F(x) = \min \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| \exp\left(\xi^{\wedge}\right) A_1 A_i^{-1} \exp\left(\xi^{\wedge}\right) \begin{bmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \\ 1 \end{bmatrix} - \begin{bmatrix} X_{1j} \\ Y_{1j} \\ Z_{1j} \\ 1 \end{bmatrix} \right\|^{2}. \tag{9}$$

4 Robot Grab Detection Method

The weak light grabbing detection network proposed in this article mainly consists of a feature extraction module and a grabbing detection module, and the overall model structure is shown in Fig. 2.

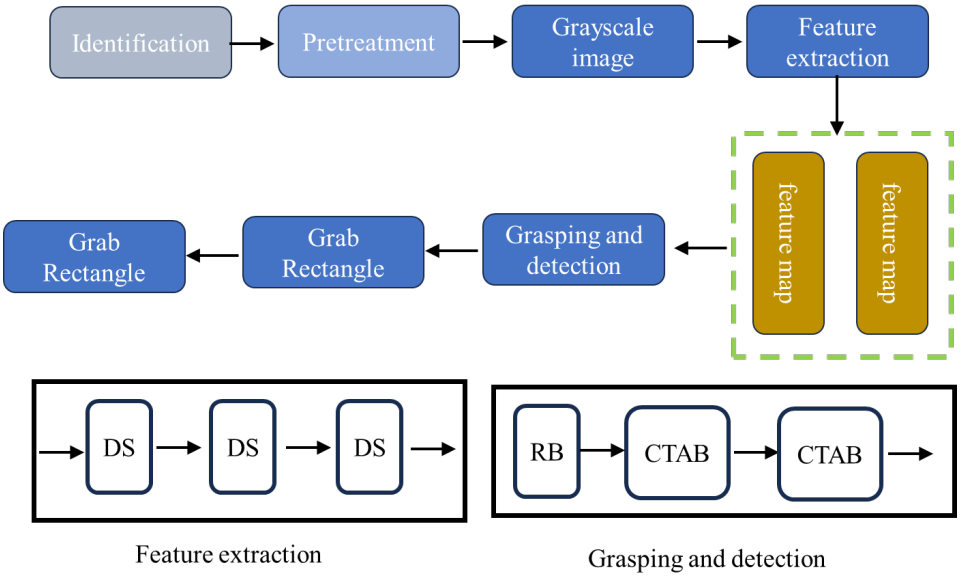

Fig. 2. Overall model structure

4.1 Feature Extraction Module

The feature extraction module consists of four modules, including Conv2D, BN, Corrected ReLU, and Max Pooling, as shown in Fig. 3. Finally, feature maps of 128 and 256 channels are generated, which are then used as inputs for GDM to predict grasping parameters.

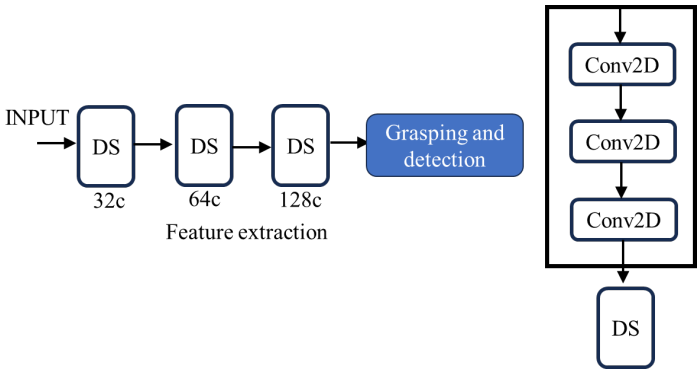

Fig. 3. Structure diagram of feature extraction module

The calculation process of the entire process is represented as:

$$I_{GDM} = DS\left(I_{low}\right)^3 . \tag{10}$$

4.2 Grab Detection Module

The Grab Detection Module (GDM) receives 128 channel feature maps from FEM and uses 5 Residual Blocks (RB) and 3 ConvT2D And BatchNorm convolutional transpose modules (CTAB) to predict the parameters of the grabable region. GDM outputs three images of the same size as the input image, namely: grasping quality, grasping angle, and grasping width. The grabbing detection module is shown in Fig. 4.

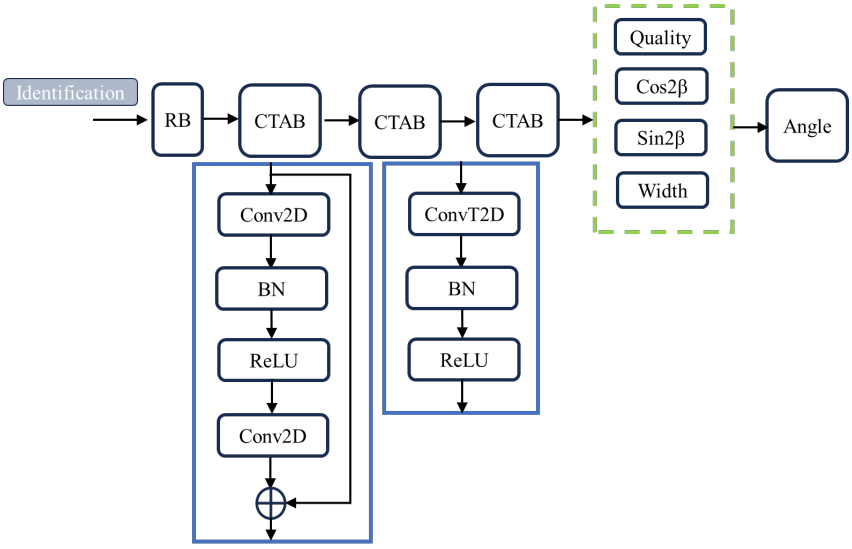

Fig. 4. Grab detection module structure

In order to restore image details and preserve spatial features, convolutional transpose operation is used in the CTAB module to obtain an output feature map with the same size as the input image. The process can be expressed as:

$$I_{output} = CTAB\left(RB\left(I_{GDM}\right)^5\right)^3. \tag{11}$$

RB represents residual module, CTAB module represents convolutional transpose module, and I_{output} represents output feature map. In the grasping stage, this article uses a pixel based grasping representation method, namely:

$$G = \left(p, \theta, w, q\right). \tag{12}$$

Among them, $p = (x, y)$ represents the geometric center position of the grabbing rectangular box, θ represents the angle of rotation of the grabbing rectangular box around the horizontal axis, with a value range of $[-\pi/2, \pi/2]$, ω is the width of the gripper opening, with a value range of $[0, W]$ pixels, and W is the maximum width of the gripper opening; q represents the grasping quality corresponding to the current grasping state, with a value ranging from 0 to 1. Approaching 1 indicates a higher grasping success rate.

4.3 Loss Function

The network loss function in this article is mainly the detection loss s_d of the grabbing detection task, expressed as:

$$s_d = \frac{1}{n}\sum_{k=1}^{n} j_k. \tag{13}$$

$$j_k = 0.5\left(G_{ik} - \hat{G}_{ik}\right)^2. \tag{14}$$

Among them, G_i represents the estimated value of grab detection, and $\hat{G}_i$ represents the corresponding true value.

5 Robot Grab Detection Method

The network model in this article aims to capture the components of household gas instruments. The system configuration is as follows: Ubuntu 18.04.6 LTS system, which runs under the hardware condition of Intel Core i7 9700 CPU (3.00 GHz) × 8) And NVIDIA GeForce RTX 2080Ti (11 GB) graphics card. The software platform is CUDA 11, Python version 3.6, PyTorch environment, PyCharm compilation platform, and result visualization is achieved through OpenCV. The camera uses Intel RealSense D345 to obtain RGB-D images.

The training test set used in the experiment is an extended version of the Cornell Grab dataset, which includes 1035 RGB images with a resolution of 640 pixels-480 pixels, 5110 positive captures and 2909 negative captures. The true value of the annotated grabbing background is represented by several grabbing rectangles, indicating the grabbing possibility of each object. By randomly cropping, scaling, and rotating, an enhanced dataset was created, resulting in effective 40000 grab examples. During the training process, only positive label crawling from the dataset is considered. This article focuses on identifying household gas meters, and the structure of the gas meter is shown in Fig. 5.

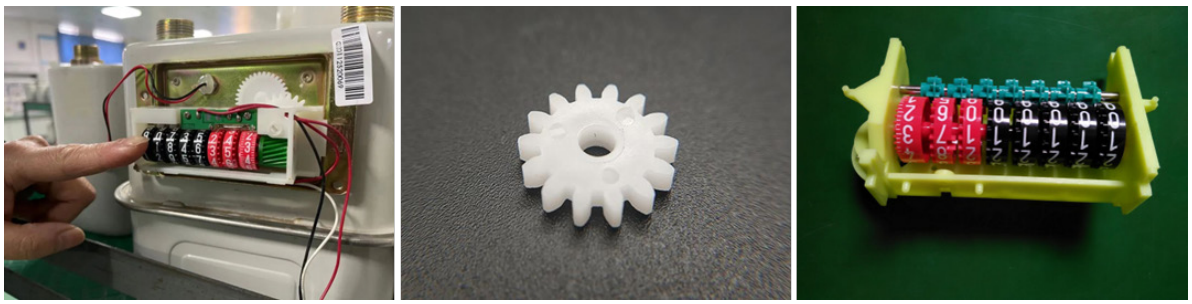

Fig. 5. The composition and structure of the instrument

According to Chapter 3, the transformation matrix obtained at the end of the robotic arm is:

$$T = \begin{bmatrix} 0.027987 & -0.006931 & 0.978924 & 0 \\ -0.982198 & -0.069726 & 0.021902 & 0 \\ 0.061098 & -0.089983 & -0.009212 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (15)$$

Use MoveIt during the grabbing process! Control the robotic arm and gripper. Place the instrument workpiece on the workbench and keep it perpendicular to the tabletop. To ensure assembly quality, the robotic arm always maintains its end perpendicular to the tabletop during downward movement. Obtain the position information of the assembly workpiece through the object detection algorithm, and publish this position information to MoveIt in the form of topic messages under ROS! Controller. MoveIt! The controller controls the robotic arm to start moving from top to bottom and reach the workpiece position, MoveIt! The controller controls the gripper to close, and the robotic arm begins to move upwards to complete the grasping task. The structure of the robot is shown in Fig. 6.

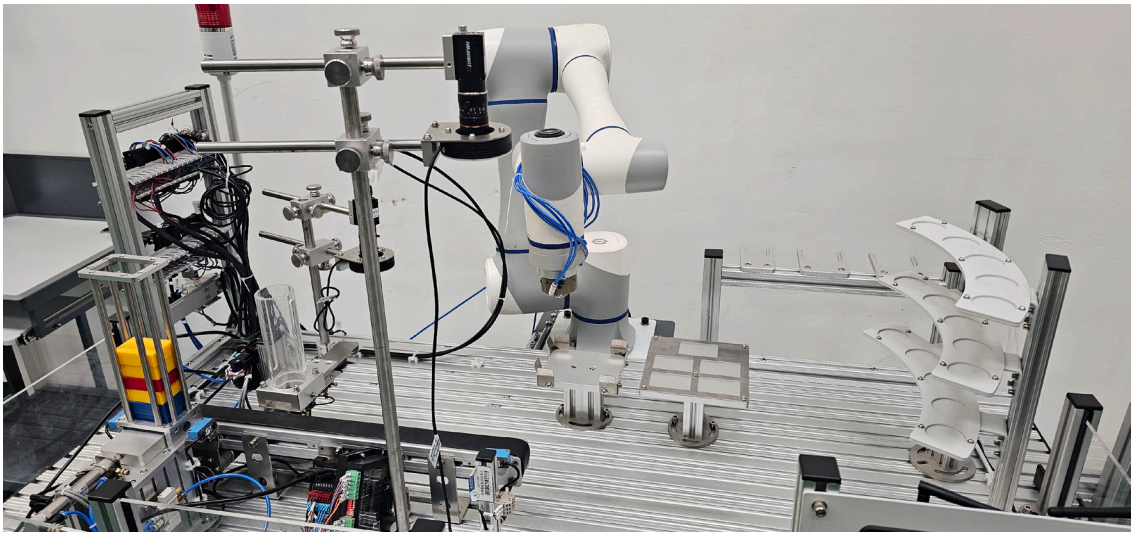

Fig. 6. Robot system

Move_ Group is the core part of Moveit, mainly responsible for organizing feature packages and plugin integration. In the system of this article, the point cloud information collected by the camera, the shutdown status of the robot, and the opening and closing status of the gripper are transmitted to the move in the form of topics_ Group, then ROS plans the trajectory of the robot's running path based on the robot model and motion control plugin. The comparison results are shown in Table 1 and Table 2.

Table 1. Comparison between the matching position and actual position of the target object

	Matching position			Actural position		
	x/mm	y/mm	z/mm	x/mm	y/mm	z/mm
Instrument gear	221.92	-786.23	-198.30	223.36	-787.62	-200.21
Instrument number axis	312.09	-792.31	-202.17	310.72	-790.74	-201.95

Through comparison, it can be seen that the absolute range of recognition errors for the target recognition object in the X, Y, and Z directions is [0.22,1.91] millimeters, which meets the requirements.

Evaluate the assembly and grabbing tasks of instrument components using the success rate of grabbing throughout the entire assembly process. The success rate of grasping is equal to the ratio of successful grasping times to the total number of experiments, with a total of 60 experiments. The grasping results are shown in Table 2.

Table 2. Results of grabbing instrument parts

	Grab tasks		
	Recognition successful	Capture successful	Success rate
Instrument gear	60	57	95%
Instrument number axis	59	53	88.3%

6 Conclusion

This article takes the robot operation skills in the scene of factory instrument grabbing and assembly tasks as the research background, and carries out intelligent recognition and grabbing of instruments based on 3D vision. A hand eye calibration method was proposed to address this issue. In the process of solving the hand eye calibration matrix, constraints were added to minimize point cloud registration errors to improve calibration accuracy. Then, a recognition model was designed that can quickly provide robot grasping coordinates and improve the efficiency and accuracy of robot guidance. Finally, the intelligent grabbing and assembly of instrument components were achieved on the visual perception based grabbing and assembly platform, demonstrating the effectiveness and practicality of the 3D vision guided grabbing detection method. However, there are still some shortcomings in this article, such as the positional errors between the identified and actual positions listed at the end of this article, and the need for improvement in accuracy. Future research directions will focus on the accuracy of calibration, the accuracy of recognition, and the lightweighting of recognition models.

7 Acknowledgement

Research on industrial Robot grasping detection based on deep learning and machine vision (2022ZC026).

References

[1] E.-W. Wen, Y.-W. He, Applications of Robot Disorder Grasping Based on 3D Vision, Development & Innovation of Machinery & Electrical Products 35(6)(2022) 36-39.

[2] H. Xiao, L.-Y. Shi, B. Sun, Y.-B. Shao, Research on Application Scheme of Multi-axis Mechanical Arm and Positioning Device in Thermocouple Automatic Calibration, Metrology & Measurement Technique 50(1)(2023) 13-16.

[3] W. Song, N.-N. Qiu, L.-Y. Shen, Y.-N. Zhang, The Monocular Stereo Matching and Grasping of Robot for Industrial Parts, Robot 40(6)(2018) 950-957.
[4] J.-C. Liu, Y.-K. Gao, Y.-Z. Chen, Research on Binocular Vision Guided Localization and Grasping System of Heterogeneous Parts, Machinery Design & Manufacture (11)(2022) 285-295.
[5] R.-Q. Gao, M.-Q. Tang, F. Lan, Research on Vehicle Radar PCB Solder Joints Defect Detection System Based on 3D Vision, Internal Combustion Engine & Parts (3)(2022) 136-138.
[6] S.-Y. Dai, L.-F. Xu, F.-Y. Liu, B. He, Semantic assistance and edge feature based salient object detection, Journal of Image and Graphics 27(11)(2022) 3243-3256.
[7] X.-D. Ouyang, L. Luo, J. Xu, Z.-Y. Liu, Calibration Method of Joint Robot Positioning Accuracy Based on Least Square Method, Machinery 49(4)(2022) 63-67.
[8] H.-L. Li, H. Jiang, P.-W. Sun, Three-dimensional object space pose detection based on Lie algebras representation, Journal of Shandong University of Science and Technology (Natural Science) 38(6)(2019) 91-97.