



## **Industrial Robot: An International Journal**

Industrial part localization and grasping using a robotic arm guided by 2D monocular vision  
Zhaohui Zheng, Yong Ma, Hong Zheng, Yu Gu, Mingyu Lin,

### **Article information:**

To cite this document:

Zhaohui Zheng, Yong Ma, Hong Zheng, Yu Gu, Mingyu Lin, (2018) "Industrial part localization and grasping using a robotic arm guided by 2D monocular vision", *Industrial Robot: An International Journal*, <https://doi.org/10.1108/IR-06-2018-0128>

Permanent link to this document:

<https://doi.org/10.1108/IR-06-2018-0128>

Downloaded on: 23 October 2018, At: 05:50 (PT)

References: this document contains references to 25 other documents.

To copy this document: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)

The fulltext of this document has been downloaded 3 times since 2018\*

Access to this document was granted through an Emerald subscription provided by emerald-srm:380143 []

### **For Authors**

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit [www.emeraldinsight.com/authors](http://www.emeraldinsight.com/authors) for more information.

### **About Emerald [www.emeraldinsight.com](http://www.emeraldinsight.com)**

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

\*Related content and download information correct at time of download.

# Industrial part localization and grasping using a robotic arm guided by 2D monocular vision

Zhaohui Zheng, Yong Ma, Hong Zheng and Yu Gu

School of Electronic Information, Wuhan University, Wuhan, China, and Department of Public Courses, Wuhan Railway Vocational College of Technology, Wuhan, China, and

Mingyu Lin

School of Mechanical & Electrical and Information Engineering, Hubei Business College, Wuhan, China

## Abstract

**Purpose** – The welding areas of the workpiece must be consistent with high precision to ensure the welding success during the welding of automobile parts. The purpose of this paper is to design an automatic high-precision locating and grasping system for robotic arm guided by 2D monocular vision to meet the requirements of automatic operation and high-precision welding.

**Design/methodology/approach** – A nonlinear multi-parallel surface calibration method based on adaptive  $k$ -segment master curve algorithm is proposed, which improves the efficiency of the traditional single camera calibration algorithm and accuracy of calibration. At the same time, the multi-dimension feature of target based on  $k$ -mean clustering constraint is proposed to improve the robustness and precision of registration.

**Findings** – A method of automatic locating and grasping based on 2D monocular vision is provided for robot arm, which includes camera calibration method and target locating method.

**Practical implications** – The system has been integrated into the welding robot of an automobile company in China.

**Originality/value** – A method of automatic locating and grasping based on 2D monocular vision is proposed, which makes the robot arm have automatic grasping function, and improves the efficiency and precision of automatic grasp of robot arm.

**Keywords** Multidimensional feature registration (ADK), Pars grasping, Self-adaptive,  $K$ -segment master curve

**Paper type** Research paper

## 1. Introduction

With the development of science and technology, robotic arms are finding increasingly wider application in industrial production, including aerospace, automobile manufacturing, the petrochemical industry, public security and other fields. Yet, traditional robotic arms can only move according to a planned trajectory and cannot obtain outside information, which limits the range of application. To enhance the flexibility of robotic arms, control technologies for robots using visual information obtained from cameras have also been developed in the past few decades.

Usually, this grasping task is called the “bin-picking” problem, and because of its extensive influence on the flexibility and productivity in manufacturing enterprises, it has been widely studied over the past 10 years.

Currently, many bin-picking systems use 3D sensors. Although these systems can completely show the 3D information on the workpiece, they still have some limitations:

- The cost of such 3D industrial sensors is still about three to four times as high as the traditional 2D industrial cameras with high resolution.

- Such systems usually require expensive devices to move the scanning head to be able to scan an entire container.
- For the detection of a very thin plane workpiece, such a piece is usually  $t$  is usually disposed in smooth hills, imposing a high precision in depth estimation which requires very expensive sensors and usually limits the active search area along the  $z$ -axis to a few centimeters.

Our solution provides a reliable, lower cost and less invasive system using a 2D industrial camera for installation in existing robotic cells, as shown in [Figure 1](#). The thickness of the target can be from 0.5 mm to tens of centimeters. Compared with using expensive industrial 3D cameras, or laser triangulation systems, the choice of a single industrial 2D camera solution makes our system much cheaper and more convenient.

Aimed at the requirements of a multi-objective automatic grasping system in robotic arms, we put forward an automatic grasping system for industrial workpiece based on the monocular vision of a machine vision system. Our recognition and localization strategy is successful thanks to an improved and highly engineered calibration and matching algorithm. We adopt the self-adaptive  $k$ -segment master curve algorithm to do nonlinear parallel multi-plane calibration, thereby improving the efficiency of the traditional calibration algorithm of a monocular camera and meanwhile increasing the accuracy of calibration. In image registration, we propose the adaptive multi-dimension

---

The current issue and full text archive of this journal is available on Emerald Insight at: [www.emeraldinsight.com/0143-991X.htm](http://www.emeraldinsight.com/0143-991X.htm)



**Figure 1** The hardware configurations of our system

feature in k-mean clustering (ADK) algorithm to perform registration for target multi-dimensional features using clustering constraints based on  $k$ -means, so that the robustness of the registration and grasping precision is greatly increased.

## 2. Related work

Despite their long history, the use of vision sensors for object detection and pose estimation in industrial robotics applications is still an active research area.

Huang *et al.* (2013) involve a point cloud through color and depth information collected by Kinect and realize the target object recognition through the point cloud segmentation. Papazov *et al.* (2012) use the depth images from Microsoft Kinect sensors to find targets, and then make the robot do bin-picking. They put forward a method comparing the mutual match between the scene point cloud and the object model to perform detection and estimate the posture. Bley *et al.* (2006) put forward another grasping method for point cloud data by studying generic object models. Buchholz *et al.* (2010) present an industrial bin-picking system based on a RANSAC-like (random sample consensus) approach to match a 3D point cloud of the workpieces against the CAD model of the searched objects: pose hypotheses are generated collecting a subset of matches between pairs of oriented points in the model and the 3D data, and the matching quality is given by the amount of contact between the surfaces. Skotheim *et al.* (2012) also propose a vote scheme based on the directional point pairs in a 3D point cloud. They install a laser triangulation sensor on the robot wrist, so it can directly scan the objects from every direction. Botterill *et al.* (2016) present a grasping scheme based on binocular vision, which uses

the double camera or Kinect's depth camera to acquire the target's depth point cloud images, and then the method does the target localization and grasping by point cloud registration.

Collewet and Chaumette (2002) advance a 2.5D visual servo, which overcomes the servo limitation stemming from its positions and images. Weiss *et al.* (1985) raise the image Jacobian matrix, without depending on camera calibration or any geometric models, by means of the deviation between the current image and the specified image and the nonlinear relationships between the robotic movements. Holz *et al.* (2015) propose a system for depalletizing and a complete pipeline for detecting and localizing objects. Their approach is based on multi-resolution surfel models and allows for low cycle times. Nieuwenhuisen *et al.* (2013) extend global navigation techniques by precise local alignment with a transport box. Objects are detected in range images using a shape primitive based approach.

Tian (2009) put forward a grasping scheme based on monocular vision, which finds the target position by image registration, and then guides the grasping directly using a simplex camera. Preto *et al.* (2013) use a monocular vision system in a bin-picking scenario to search the targets with grayscale. The method is based on contour matching of the scene and the target model. They report cycle times of up to 7 s in a statically mounted robot setup. Rahardja and Kosaka (1996) offer a bin-picking system based on stereoscopic vision, where a feature model is chosen by the operator, then the objects being searched for are located roughly, and then other features are found which are commonly small and helpful in eliminating errors and improving localization accuracy. C'ozar *et al.* (2001) present the generalized Hough transform (GHT), which is used for 3D localization of planar objects, and the computational complexity of the GHT is reduced by uncoupling the parameter detection.

## 3. System overview

The automatic grasping system in industrial robotic arms usually consists of the mechanical arm, conveyor belt, light source and visual system. The visual system is composed of the visual sensors and master control computer.

### 3.1 System architecture

Because of the higher accuracy requirements for processing parts in industry, in the system, we selected a single-eye color CCD industrial camera with a Basler Ethernet interface as our visual sensor, the resolution of which is  $2,448 \times 2,050$ , while the actual filming area is  $60 \times 50$  cm.

An LED linear illuminator is used as the light source with a concentrated effect, where the high brightness can shorten the camera exposure time, and at the same time, the light source is stable, thus affording continuous invariance for continuous visual detection.

The conveyor belt is a black rubber belt with high-strength chemical fiber canvas as the basement layer and superior quality rubber as the outer layer and rough surface. This conveyor belt has strength, elasticity, durability and resistance to impact, aging and groove characteristics, so it provides a background with high morphological stability and anti-glare characteristics in the vision system.

The robotic arm is an IRB1600 robot from ABB, and it has excellent repeatable localization ( $\pm 0.05$  mm) precision and

trajectory accuracy. At the same time, it adopts an integrated design and, all cables and flexible tubes are embedded in the robot's upper arm, which is a good choice in arc-welding applications, and the wire and cable contain the dielectric medium for arc welding, including the power supply, solder wire, shielding gas and compressed air.

### 3.2 Workflow

The robotic arm-grasping device, which is designed based on an industrial camera and computer vision system, is shown in Figure 2.

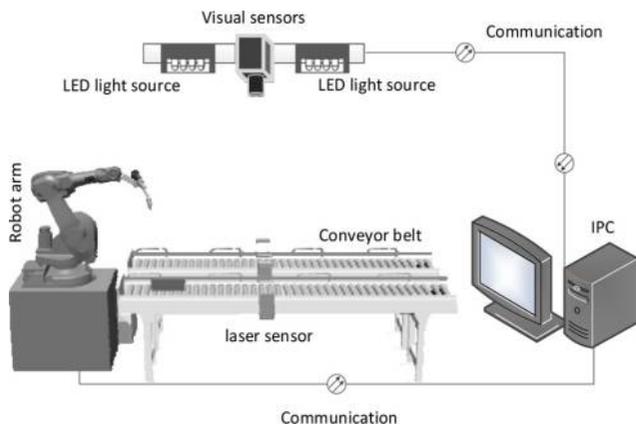
The device consists of the robotic arm, conveyor belt, light source, industrial camera, laser sensor, industrial personal computer and so on, which is placed in an indoor environment to avoid any influence on the test results caused by strong external light changes. The double LED light sources and monocular industrial camera are installed right above the conveyor belt, and are in a position relatively parallel to the conveyor belt. This helps optimize the LED light source localization to minimize changes in light and provides the best lighting for workpiece imaging and ensures that the imaging is directed toward the conveyor belt. The industrial camera links with the IPC are through a Gigabit Ethernet port. When the workpiece move to the laser sensor on the conveyor belt, the conveyor belt stops. Meanwhile, the camera triggers a photo and sends the photo image to the industrial control computer. The IPC calculates the displacement distance and rotation angle using the computer vision algorithm, and then delivers it to the robotic arm. Then, the robotic arm moves and grasps the workpiece on the conveyor belt according to the displacement distances and rotation angles. To improve the workpiece localization accuracy, there is a circular location hole where the grasped workpiece is fastened to the welding groove to confirm the localization accuracy. The workflow diagram of this system is shown in Figure 3.

## 4. Approach

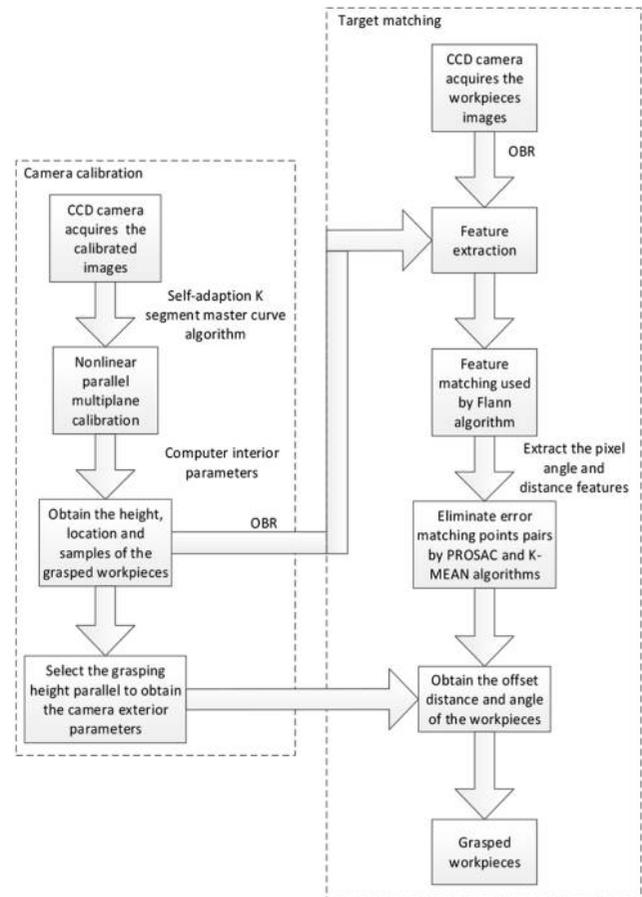
### 4.1 Nonlinear parallel plane calibration algorithm for the camera

The camera is an important sensor in machine vision, as the camera parameters have an effect on the target recognition and positional accuracy. Camera calibration involves finding out

**Figure 2** The structural diagram for the robotic arm grasping device



**Figure 3** The workflow diagram of the robotic arm grasping system



the camera's interior and exterior parameters. The interior parameters mainly refer to the camera's own hardware parameters, such as the camera lens distortion, focal length and the installation position of the imaging chip in the manufacturing process. The exterior parameters mainly refer to the parameters in the World Coordinate System, such as the camera's position and rotational direction, etc.

We chose a single sensor, whereby the complexity of the server information captured by the single sensor is much lower, and the calculation is smaller, but we cannot get the target depth information directly. Considering that the grasping height for each workpiece is the same, we could establish parallel planes of all heights that are parallel to the conveyor belt, and take the points on each parallel plane to carry out the calibration, as shown in Figure 4. Yet, a parallel plane is endless, if all parallel planes are to be used in the calibration, and there are no possibilities to do these calibrations; however, if we select the parallel planes for calibration within a certain interval, there will be considerable error when the height interval is too big. At the same time, a farther distance from the center will lead to a greater error. According to the pinhole imaging principle, the farther the distance from the center, the larger the tilt angle and the greater the drift errors caused by the height difference, as shown in Figure 5. Yet, if the height interval is too small, the amount of calculation is too big and doing this is time-consuming, which causes the operability to

Figure 4 Parallel plane calibration diagram

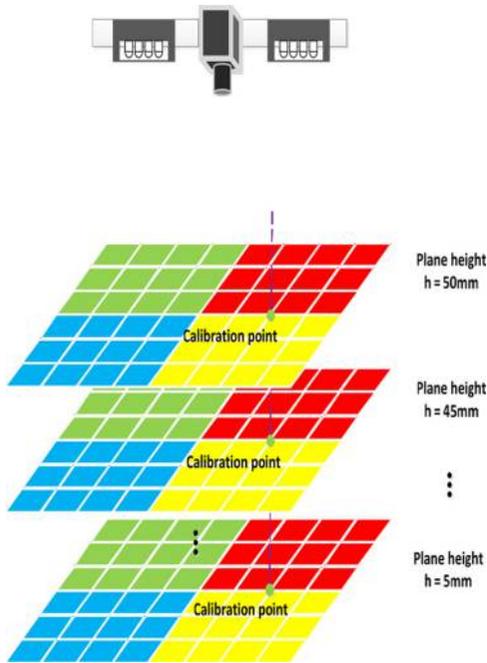
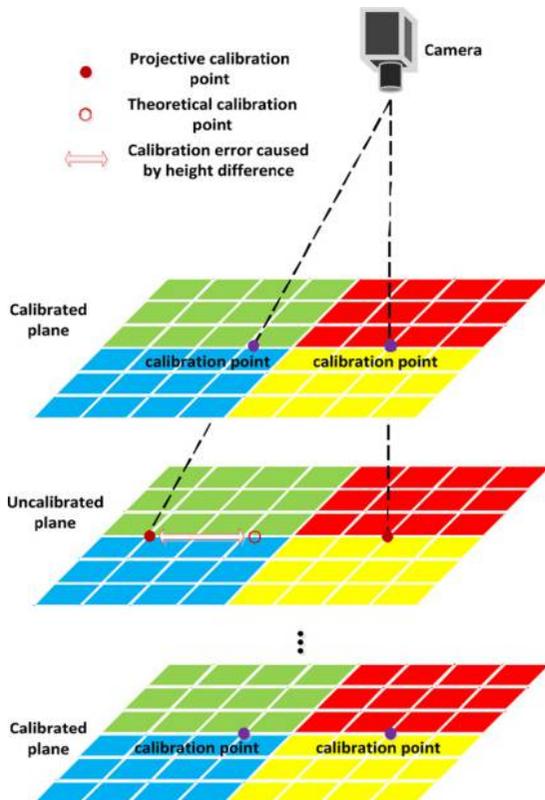


Figure 5 Calibration error caused by using the calibrated parallel plane to derive



diminish. Therefore, we propose a nonlinear parallel plane calibration algorithm.

There are  $2,448 \times 2,050$  pixels in the whole image plane, and we take the middle region of the plane to be divided into 13 rows and 15 columns, 208 points in total, and then move the robotic arm along the location of these 208 points successively, and at the same time record the corresponding World Coordinates, as shown in Figure 6. Here, the conveyor belt is regarded as the initial plane with zero height, because the height at which the workpiece is put on the conveyor belt is commonly 2 to 50 mm. We chose a parallel plane every 5 mm in height, recording the corresponding image coordinates of the 208 World Coordinate points at the given height, so that there are 10 parallel planes in total. Since there is 5 mm height between each parallel plane, we adopt the self-adaptive  $k$ -segment master curve algorithm to fit the corresponding image coordinates of the 208 World Coordinate points of each height plane.

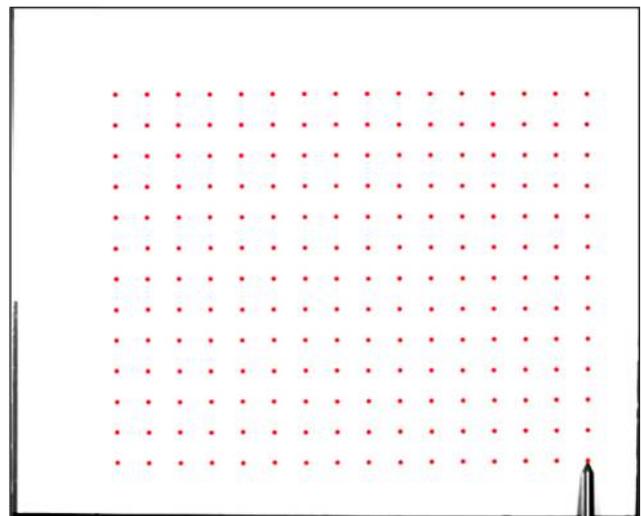
The master curve is the nonlinear generalization of the first principal component. To emphasize finding the smooth curves that are through the “intermediate” of the data distribution, these curves give an overview of the data (Hastie, 1984). Krzyzak and Kegl (1999) thought that certain constraints needed to be attached to some of the curve clusters, and finding the cluster could minimize the distance function. Based on this thought, they gave the definition of  $K$  principal curves.

Regarding the data point set  $X$ , the curve  $f^*$  is known as the master curve of the length  $L$ , and if it is on the curve clusters, where all lengths are less than or equal to  $L$ ,  $f^*$  minimizes the distance function:

$$\begin{aligned} \Delta f &= E[\Delta(X, f)] = E\left[\inf_{\lambda} \|X, f(x)\|^2\right] \\ &= E\left[\|X - f(\lambda_f(X))\|^2\right] \end{aligned} \quad (1)$$

We will take the World Coordinates of the 208 points in the ten parallel planes as the 208 sample sets:

Figure 6 Robotic arm calibration lattice diagram, where the red dots mean 208 points



$$X_i = \{p_{i1}, p_{i2}, \dots, p_{i10}\} \subset R^d (i = 1, 2, \dots, 208) \quad (2)$$

Looking for the function  $f_i$  to make the:

$$\frac{1}{10} \sum_{j=1}^{10} \Delta(p_{ij}, f_i) \quad (3)$$

Value minimum.

First, we work out the shortest first principal component line by using the known World Coordinates of the ten parallel planes, and use the line as the initial curve  $f_i^1$ . Let us suppose that  $f$  is the polygonal line with vertexes like  $u_1, u_2, \dots, u_{k+1}$  and segments such as  $s_1, s_2, \dots, s_k$ , where  $s_k$  is connected  $u_k$  and  $u_{k+1}$ , and sample set  $X_i$  is divided into the  $2k + 1$  different areas such as  $s_1, s_2, \dots, s_k$  and  $u_1, u_2, \dots, u_{k+1}$ . The definition of these vertexes and segments are defined as below:

For any  $P \subset R^d$ , set:

$$\Delta(p, s_j) = \|p - s_j\|^2 \quad (4)$$

$$\Delta(p, u_j) = \|p - u_j\|^2 \quad (5)$$

In the form:

$$U_j = \{P \in X_i : \Delta(p, u_j) = \Delta(p, f), \Delta(p, u_j) < \Delta(p, u_m), m = 1, \dots, j - 1\} \quad (6)$$

$$S_j = \{P \in X_i : \Delta(p, s_j) = \Delta(p, f), \Delta(p, s_j) < \Delta(p, s_m), m = 1, \dots, j - 1\} \quad (7)$$

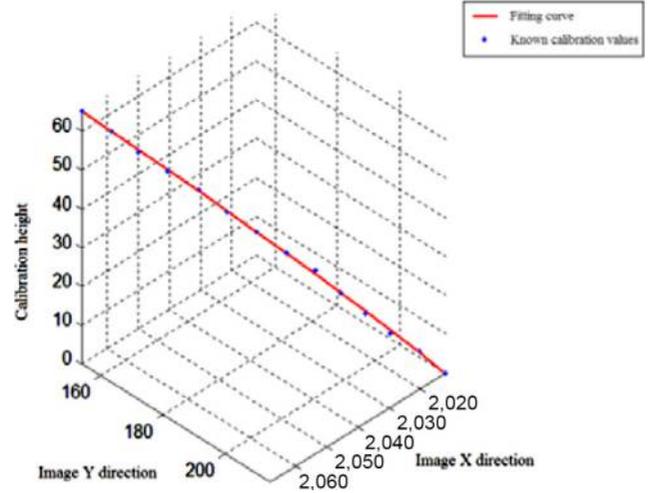
For each vertex, find  $u_i^j$  by fixing the rest of the vertexes to make the distance functions  $G_i(f)$  minimum:

$$G_i(f) = \begin{cases} \sum_{p \in S_j} \Delta(p, s_j) + \sum_{p \in U_j} \Delta(p, u_j), i = 1 \\ \sum_{p \in S_{j-1}} \Delta(p, s_{j-1}) + \sum_{p \in U_j} \Delta(x, u_j) + \sum_{p \in S_j} \Delta(p, s_j) \\ , 1 < i < k + 1 \\ \sum_{p \in S_{j-1}} \Delta(p, s_{j-1}) + \sum_{p \in U_j} \Delta(p, u_j), i = k + 1 \end{cases} \quad (8)$$

Then we select the longest line with the most projection points to make the midpoint of the line be the new vertex. To achieve better robustness,  $k$  must be sensitive to the mean square distance, and it stops when the number of line segments  $k$  is approximately  $10^{1/3}$  and is convergent to the expected square distance at  $o(10^{1/3})$  speed. Thus, we calculate the image coordinates of the 208 points in every parallel plane, as shown in Figure 7.

Through the 208 World Coordinates and the corresponding image coordinates in any of the height planes, we calibrate the camera through the mature and stable calibration algorithm of Zhang (2000) and find the interior parameters and the corresponding height exterior parameters of the camera.

Figure 7 The fitting diagram of a calibration point in different height planes using the master curve algorithm



#### 4.2 Target multidimensional registration algorithm

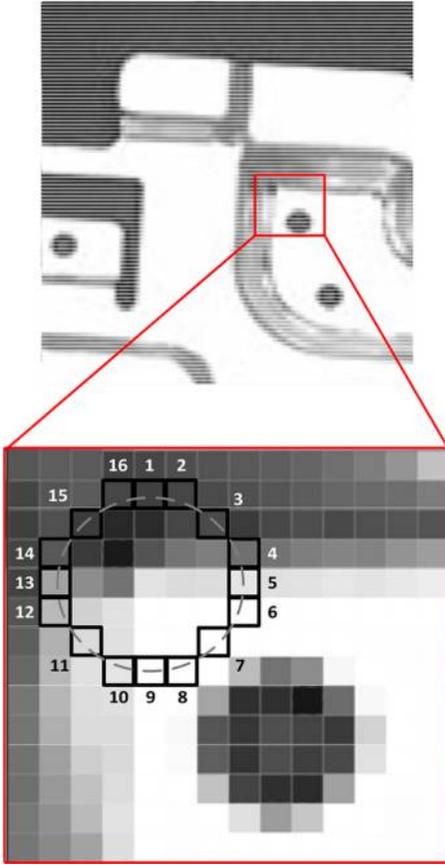
Grasping objects is the most basic ability of a robotic arm, and the level of the ability is an important basis for evaluating a robot's degree of intelligence. The primary task in completely grasping an object is target registration, and feature matching has been widely applied in target registration (Ma et al., 2018, 2017), so we adopt the method of feature point matching for image registration, and the feature points can be sift, surf and so on. We chose the OBR with a faster speed (Rublec et al., 2011). First, we extract the OBR feature points of the two images being registered, then match and filtrate the features in two images using the ADK algorithm and get the perspective transformation matrix (homographic matrix).

OBR detects the feature points with the FAST algorithm. This definition is based on the image's gray values for the surrounding feature points and detecting the pixel values of the candidate feature points in a surrounding circle. If there are enough pixels around the candidate point whose gray values are sufficiently different from the candidate point, it is argued that the candidate point is regarded as a feature point, subject to the following formula:

$$N = \sum_{x \in \text{circle}(p)} |I(x) - I(p)| > \varepsilon_d \quad (9)$$

where  $I(x)$  is the gray of any point in the circumference,  $I(p)$  means the gray of the center and  $\varepsilon_d$  is the threshold of the gray value difference. If  $N$  is greater than the given threshold which generally is three-quarters of the circle points surrounding,  $p$  is regarded as a feature point. For obtaining results faster and improving the efficiency of the comparison, we adopt FAST-9, as shown in Figure 8.

We match the OBR feature points coarsely using the Flann algorithm (Muja and Lowe, 2009), whose feature space is commonly  $n$  real vector space  $R^n$ , and the core of that algorithm is to find the neighbors of the instance points using Euclidean distance. The eigenvectors of the feature points  $p$  and  $q$  are

**Figure 8** Application of the FAST-9 algorithm to an artifact image

represented as  $D_p$  and  $D_q$ , so the Euclidean distance of  $d(p, q)$  can be indicated as in the following formula 10:

$$d(p, q) = \sqrt{(D_p - D_q)(D_p - D_q)^T} \quad (10)$$

We find the initial matching point pair  $(m_{1i}, m_{2i})$  with a minimum distance  $D$  in the image  $I_2$  of the feature points in the image  $I_1$  through the Flann algorithm and then calculate the minimum distance  $\min D$  according to the distances of all the matching point pairs, setting the threshold  $T = \mu \times \min D$  (here, we let  $\mu = 5$ ). If the minimum distance of the matching point pair  $D < T$ ,  $m_2$  will be like the candidate matching point of  $m_1$ , otherwise, do the next feature point matching of image  $I_1$  and eliminate the point  $m_{1i}$ , and finally get the Flann matching point pairs from image  $I_1$  to image  $I_2$ . Do the same for the feature points in image  $I_2$ , and get the Flann matching point pairs from image  $I_2$  to image  $I_1$ . Then judge these two Flann matching point pairs, removing the asymmetry matching point-pairs, so as to get an initial matching set.

Coarse matching in the Flann algorithm will have many false matching points, which has an influence on the matching accuracy to a large extent. Therefore, mismatch removal algorithm is particularly important (Ma et al., 2014, 2013), and we do the validated matching with the ADK algorithm, eliminating error matching point pairs. Before using the ADK algorithm, we preprocess making use of the PROSAC algorithm (Brandt, 2008). The PROSAC algorithm sorts the

individuals according to the similarity level of the figure matching point pairs first and considers that the samples with high similarity are more likely to be the interior point of the correct model, then determines the model parameters based on the data for the interior points.

The PROSAC algorithm can eliminate many false matching points, but will still retain some false matching points with small distance differences from the Figure above. These false matching points will have an effect on the accuracy of the final perspective transformation matrix. If these false matching points map to the corresponding figure spaces, such as the distance feature vector and the direction feature vector between the false matching points, we could find that there is no obvious distribution regularity in the false matching points. However, the correct matching point distributions are relatively concentrated and have local similarity transformation consistency, as shown in Figure 9 and are suitable for removing the false matching points using the  $k$ -means clustering algorithm. Thus, we put forward two new feature vectors, i.e. the distance difference feature vector and the angle difference feature vector, at the same time, and consider using the  $k$ -means clustering method based on the disparity constraint to remove the false matching points through the multidimensional feature vectors.

After filtrating with the PROSAC algorithm, the foremost matching point  $(m_{11}, m_{21})$  has the greatest credibility, so we choose this matching point as a reference point and then calculate the distance difference feature vector  $D$  and the angle difference feature vector  $\theta$  of all matching points set  $H$ , as shown in Figure 10:

The distance difference feature vector  $D$  is below:

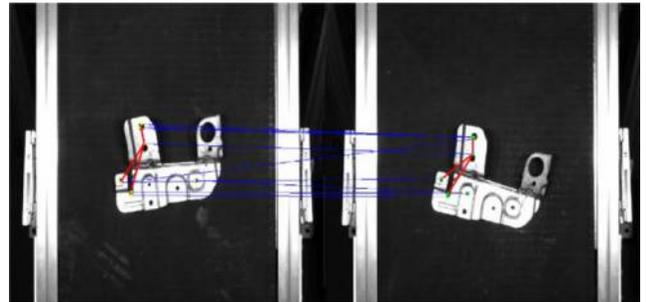
$$D = \{d_{1k} - d_{2k}\} \quad (11)$$

where  $d_{ik}$  is the distance between the matching point and the reference point.

$$d_{ik} = \sqrt{(x_{m_{ik}} - x_{m_{i1}})^2 + (y_{m_{ik}} - y_{m_{i1}})^2}, k \neq 1, m_{ik} \in H, i = 1, 2 \quad (12)$$

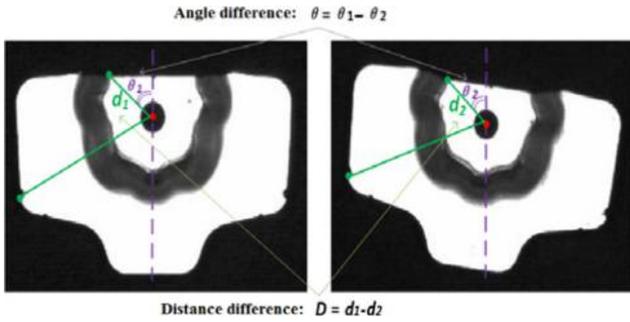
The angle difference feature vector  $\theta$  is below:

$$\theta = \{\theta_{1k} - \theta_{2k}\} \quad (13)$$

**Figure 9** The relative consistency of the correct matching points in the local region

**Notes:** The red line is the geometric structure between the correct matching points

**Figure 10** The distance difference and angle difference of the matching points



where  $\theta_{ik}$  is the angle between the matching point and the reference point.

$$\theta_{ik} = \text{atan} \frac{y_{m_{ik}} - y_{m_{i1}}}{x_{m_{ik}} - x_{m_{i1}}}, \quad (14)$$

$$k \neq 1, m_{ik} \subset H, i = 1, 2$$

The  $k$ -means clustering algorithm is used to cluster the distance difference feature vector  $D$  and the angle difference feature vector  $\theta$  between the sample and the figures being detected because the initial clustering center selection and the number of categories has a strong influence on the  $k$ -means clustering

algorithm (Wang and Zhao, 2010). We observe that the distance feature vectors of the correct matching points are similar according to the feature point matching. If the distances of the matching point pairs are indicated by Euclidean distance, the distance differences are generally not more than five pixels, and the direction feature vectors are also simultaneously very similar. The angle difference generally is within  $3^\circ$ , so we take 5 pixels and  $3^\circ$ , respectively, as the classification boundaries, which make them into a category where the distance difference is less than 5 pixels or the direction difference is no more than  $3^\circ$ , until the aggregation cannot continue, and put the mean values of each clustering category as the initial clustering centers to cluster using the  $k$ -means algorithm.

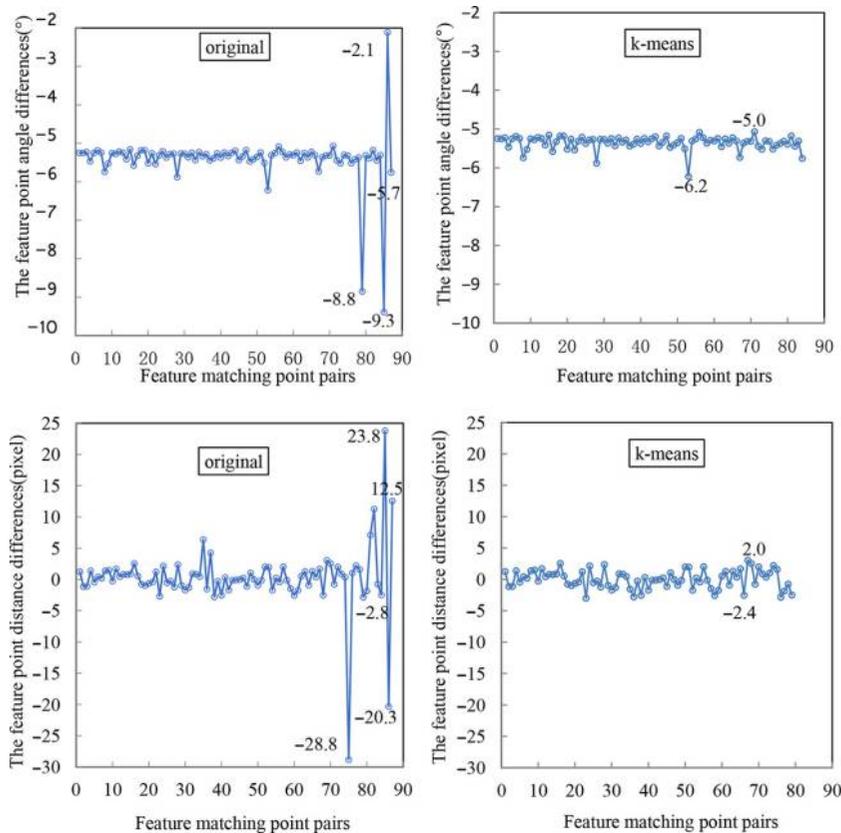
The false matching rate of the matching point pairs processed by the ADK algorithm could fall below 1 per cent, as shown in Figure 11. Finally, we calculate the perspective transformation matrix between image  $I_1$  and image  $I_2$  using these matching point pairs with higher precision.

Finally, the transformation matrix  $w$  is used to calculate the image coordinates of the grabbing point. The formula is as follows:

$$\begin{aligned} x &= x'/w \\ y &= y'/w \end{aligned} \quad (15)$$

where  $(x', y')$  is the grabbing point of the template, and  $(x, y)$  is the grabbing point of the target.

**Figure 11** The first col is the original angle differences and distance difference of the matching point pairs, and the second is the angle differences and distance difference of the matching point pairs after elimination using the  $k$ -means algorithm



After converting the camera coordinate system into the robot coordinate system, the location of the target is compared with the position of the template to find the deviation, which is transmitted to the working robot as the position command.

## 5. Experimental results and analysis

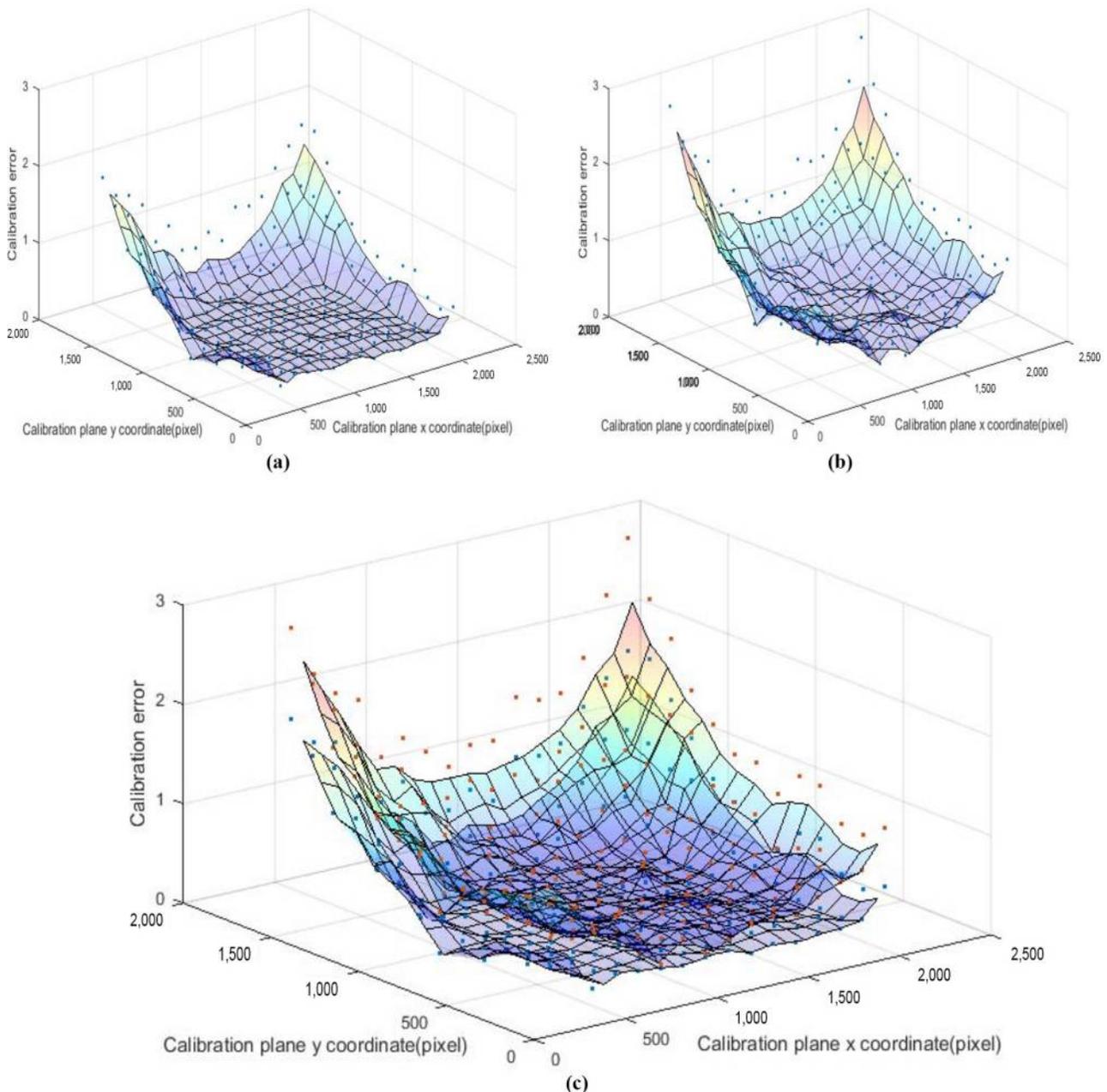
### 5.1 Experimental platform

To verify our design for the robotic arm position and fetching method, we designed a series of experiments to verify the performance of the algorithm. In the process used in the comparative experiments, the parameters of the algorithm are

consistent. These experiments were carried out on a notebook with 2.4 GHz Intel core CPU and an 8 GB memory notebook, and the open source toolbox opencv2.94 was used.

The data sets came from the seven different workpieces, with 1,000 of each workpiece being continuously tested. The image size was  $2,448 \times 2,050$ . To get close to the actual operation of the workpiece, we calculated the real error value through the actual grasping of the robotic arm and defined the corresponding test index: the camera calibration accuracy was the AME of the measurement accuracy for some plane calibration points, where the gripping accuracy refers to the average distance and angle of the offset error for the grasping of

**Figure 12** (a) Projection error with non-linear plane calibration method for  $h = 8\text{mm}$  and (b) projection error with adjacent plane calibration data method for  $h = 8\text{mm}$ ; (c) comparison of the two methods



the robotic arm. At the same time, true positive was defined such that the grasping error distance was less than 3 mm and the grasping error angle was less than  $2^\circ$ , and the false positive was defined as the opposite.

### 5.2 Experiment 1: camera calibration error analysis

We begin by calibrating the 5 mm height plane, and calibrate a plane every 5 mm, so there are 10 planes in all. In the experiments, we selected the 8 mm height plane to test, and tested the camera's interior and exterior parameters of the 8 mm height plane by adopting the method for calibration plane data nearby (for example, the 8 mm height plane adopts the calibration data of the 10 mm height plane) and nonlinear parallel plane calibration, then we obtained the test results of the re-projection errors by comparing the real values, as shown in Figure 12.

### 5.3 Experiment 2: target grasping error analysis

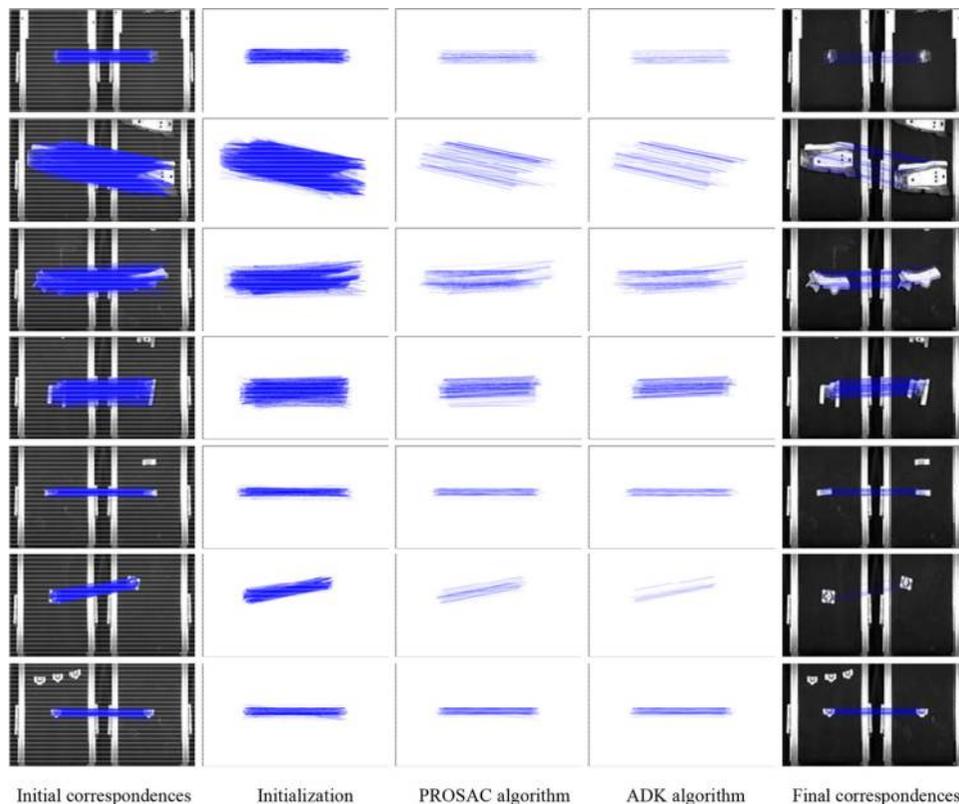
Next, we tested the same workpieces to do image matching according to the different laying angles. Because the rotation angles of the robotic arm are limited, we selected rotation angles within  $0^\circ \sim 60^\circ$  and  $-60^\circ \sim 0^\circ$  to test, as shown in Figure 13. We compared each stage of the matching process, with the columns showing the algorithm process, and at the very start, we established the putative correspondence sets using the Flann algorithm, and all the orb matches were

assumed to be inliers, as shown in the first column. To better display the matching effect, we display the matching points separately, as shown in the second column, where the head and tail of each arrow correspond to the positions of a Flann match in two images. The third column shows the matching results of the PROSAC algorithm, and the mismatching point pairs show a substantial reduction, but there is a small number of mismatching point pairs with small deviation. The fourth column shows that ADK almost converges to a nearly binary decision on the match correctness. The final images matched by our ADK are presented in the last column. From the results, we see that our ADK is able to distinguish inliers from the outliers on all pairs.

At the same time, we present a quantitative comparison of the data sets using the ADK, RANSAC and PROSAC feature matching algorithms. The comparative results are given in Table I, where obviously, our algorithm has higher true positives and relatively lower false positives compared with the other two algorithms. Therefore, it is better suited for our grasp planning approach.

Finally, we performed grasping experiments on the data sets using the three kinds of algorithms and calculated the offset distance and angle of the robot grasping points. Then we obtained the mean errors for each algorithm, and the results are shown in Figure 14. From the results, it is easy to see that PROSAC has better mean accuracy than the RANSAC

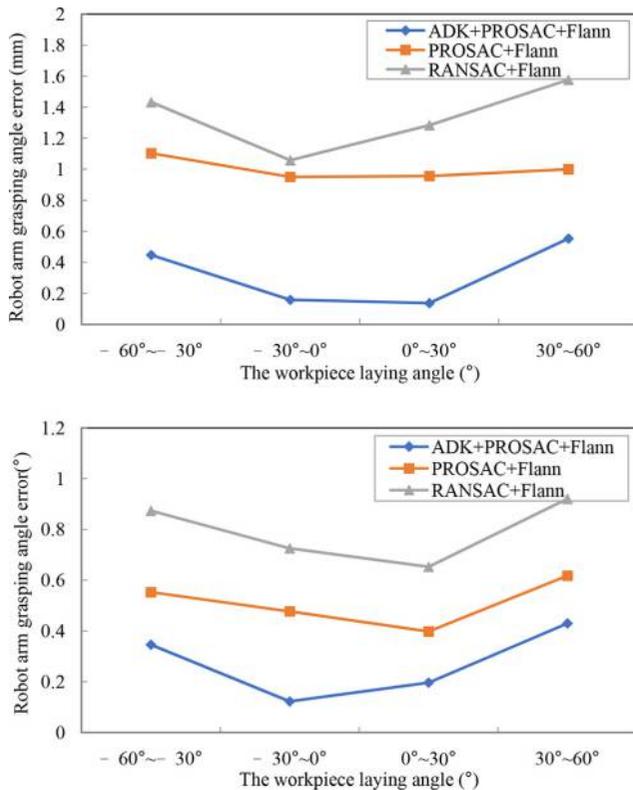
Figure 13 Matching results on several typical workpiece image pairs



**Notes:** The rows show different rigid models. The columns show the matching process during the ADK algorithm, and the level of the blue color indicates to what degree a correspondence belongs to the inlier

**Table 1** Results of the comparison of the three kinds of algorithms

Algorithm type	True positives (%)	False positives (%)
ADK	99.10	0.90
PROSAC	85.30	14.70
RANSAC	83.50	16.50

**Figure 14** The average grasping distance and angle error of the workpieces at different laying angles

algorithm. This is because our application scenario is single, and has less interference, so the probability that a feature with a higher degree of matching is the interior point is much higher than one where there is a lower degree of matching. The matching performance of our ADK algorithm has better accuracy and is more stable than the other two kinds of algorithms.

Considering the test results above, the offset distance range of this grasping method is between 0 and 0.5 mm, and the angle error range is from  $-0.2^\circ$  to  $0.2^\circ$ , which illustrates that from either the angle accuracy or the offset distance accuracy, our grasping method is relatively better compared to a monocular camera and can effectively realize the automatic grasping function of a robotic arm. The stability of the system and positional results can meet the needs of practical production.

## 6. Conclusion

In this paper, we presented a robust, flexible, low-cost and noninvasive smart-picking system that allows detection and

precise location of planar objects randomly placed on a conveyor belt using only a single camera. We improve the accuracy and efficiency of the traditional calibration algorithm of a monocular camera by using a nonlinear parallel plane algorithm and develop an ADK algorithm based on image feature matching, ensuring that its robustness, stability and grasping accuracy is greatly increased without costing much time. The technology has been successfully applied to an automobile part grasping and welding system on the FANUC and ABB robots. In our next work, we will improve the self-adaptive threshold for registration and registration efficiency, and we will apply our algorithm to automatic image matching with speed and high accuracy.

## References

- Bley, F., Schmirgel, V. and Kraiss, K.F. (2006), "Mobile manipulation based on generic object knowledge", *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 411-416.
- Botterill, T., Paulin, S., Green, R., Williams, S. and Lin, J. (2016), "A robot system for pruning grape vines", *Journal of Field Robotics*, Vol. 34 No. 6, pp. 1100-1122.
- Brandt, S. (2008), "Maximum likelihood robust regression with known and unknown residual models", *Statistical Methods in Video Processing Workshop*, pp. 97-102.
- Buchholz, D. Winkelbach, S. and Wahl, F.M. (2010), "Ransac for industrial bin-picking", *ISR 2010 and ROBOTIK 2010*, pp. 1-6.
- Cózar, J.R., Mata, N.G. and Zapata, E.L. (2001), "Detection of arbitrary planar shapes with 3d pose", *Image Vision Computer*, Vol. 19 No. 14, pp. 1057-1070.
- Collewet, C. and Chaumette, F. (2002), "Positioning a camera with respect to planar objects of unknown shape by coupling 2-D visual servoing and 3-D estimations", *IEEE Transactions on Robotics and Automation*, Vol. 18 No. 3, pp. 322-333.
- Hastie, T. (1984), *Principal Curves and Surfaces*, Stanford University, California.
- Holz, D., Topalidou-Kyniazopoulou, A. and Stückler, J. (2015), "Real-Time object detection, localization and verification for fast robotic depalletizing", *IEEE/RSJ International Conference on Intelligent Robots & Systems*, pp. 1459-1466.
- Huang, Z.Y., Huang, J.T. and Hsu, C.M. (2013), "A case study of object identification using a kinect sensor", 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1743-1747.
- Krzyzak, A. and Kegl, B. (1999), *Principal Curves: Learning, Design, and Applications*, Concordia University.
- Ma, J.Y., Jiang, J.J., Liu, C.Y. and Li, Y.S. (2017), "Feature guided Gaussian mixture model with Semi-Supervised EM and local geometric constraint for retinal image registration", *Information Sciences*, Vol. 417, pp. 128-142.
- Ma, J.Y., Jiang, J.J., Zhou, H.B., Zhao, J. and Guo, X.J. (2018), "Guided locality preserving feature matching for remote sensing image registration", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56 No. 8, pp. 4435-4447.
- Ma, J.Y., Zhao, J., Tian, J.W., Bai, X. and Tu, Z.W. (2013), "Regularized vector field learning with sparse approximation for mismatch removal", *Pattern Recognition*, Vol. 46 No. 12, pp. 3519-3532.

- Ma, J.Y., Zhao, J., Tian, J.W., Yuille, A. and Tu, Z.W. (2014), "Robust point matching via vector field consensus", *IEEE Transactions on Image Processing*, Vol. 23 No. 4, pp. 1706-1721.
- Muja, M. and Lowe, D.G. (2009), "Fast approximate nearest neighbors with automatic algorithm configuration", *IEEE Conference on Computer Vision Theory and Applications*, pp. 331-340.
- Nieuwenhuisen, M., Droschel, D., Holz, D. and Stuckler, J. (2013), "Mobile bin picking with an anthropomorphic service robot", *IEEE International Conference on Robotics & Automation*, pp. 2327-2334.
- Papazov, C., Haddadin, S., Parusel, S., Krieger, K. and Burschka, D. (2012), "Rigid 3D geometry matching for grasping of known objects in cluttered scenes", *The International Journal of Robotics Research*, Vol. 31 No. 4, pp. 538-553.
- Pretto, A., Tonello, S. and Menegatti, E. (2013), "Flexible 3d localization of planar objects for industrial bin-picking with monocular vision system", *IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 168-175.
- Rahardja, K. and Kosaka, A. (1996), "Vision-based bin-picking: recognition and localization of multiple complex objects using simple visual cues", *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

- Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011), "ORB: an efficient alternative to SIFT or SURF", *International Conference on Computer Vision*, Vol. 22 No. 11, pp. 2564-2571.
- Skotheim, O., Lind, M., Ystgaard, P. and Fjerdingen, S. (2012), "A flexible 3d object localization system for industrial part handling", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3326-3333.
- Tian, k. (2009), "Based on the visual servo and position of monocular vision", HuaZhong university of science and technology.
- Wang, Y.X. and Zhao, X.J. (2010), "Improving algorithm of K-means based clustering segmentation of colour image", *Computer Applications and Software*, Vol. 30 No. 2, pp. 127-130.
- Weiss, L.E., Sanderson, A.C. and Neuman, C.P. (1985), "Dynamic visual servo control of robots: an adaption image-based approach", *IEEE International Conference on Robotics & Autom.*, pp. 662-668.
- Zhang, Z. (2000), "A flexible new technique for camera calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22 No. 11, pp. 1330-1334.

### Corresponding author

Hong Zheng can be contacted at: [zh@whu.edu.cn](mailto:zh@whu.edu.cn)