Recognition and Localization Method for Automotive Axle Holes in Assembly Robots

Junsheng Zhang

Tianjin University of Science and Technology, 1038 Dagu South Road, Hexi District, Tianjin, 300222, China

Dongpo Ma

Tianjin University of Science and Technology, 1038 Dagu South Road, Hexi District, Tianjin, 300222, China

Yizhun Peng*

Tianjin University of Science and Technology, 1038 Dagu South Road, Hexi District, Tianjin, 300222, China

Email: pengyizhun@126.com

Abstract

During the production of automotive axle holes, the roundness error at the pipe opening leads to low detection efficiency due to manual measurements, rendering real-time inspection unfeasible. This paper proposes a method for detecting and sorting the roundness of automotive axle holes based on visual inspection. Ensuring accuracy in grasping, it establishes the kinematic model of the robot Aubo_I5. Targeting automotive axle holes for grasping, it employs adaptive threshold segmentation to highlight the section features of the axle hole. The Canny algorithm is then used to extract edge information, and finally, the least squares method is utilized to detect roundness errors for sorting the axle holes based on this error.

Keywords: Axle Hole, Assembly, Vision Deep, Learning

1. Introduction

The assembly of mechanical components constitutes the subsequent phase in the entire manufacturing process. Manual assembly is characterized by low efficiency, and the assembly speed is affected by factors such as worker labor intensity. Traditional assembly robots have mostly been controlled through teaching or offline programming [1], [2]. Consequently, these robots strictly follow predefined assembly procedures, lacking the ability to promptly respond to changes in the surrounding environment or objectives, thus exhibiting poor robustness. With the advancements in computer and visual technologies, automation machinery on production lines gradually possesses visual perception capabilities. This enables real-time dynamic recognition and localization of targets, facilitating visual-guided adjustments of robot poses and grasping [3].

The emergence of deep learning has improved issues regarding detection accuracy in visual recognition. However, due to the complexity of the assembly process, uncertainties such as lighting conditions, backgrounds, workpiece sizes, shapes, and other factors in the working environment are critical issues affecting target recognition and localization.

2. Kinematic Modeling

Establish the joint coordinate system of the Aubo_I5 robot as shown in Fig. 1, and the structural model and dimensions of each joint as illustrated in Fig. 2. To more accurately define the sorting model coordinate system, the gripper coordinate system G_{xy} is established at the end joint.

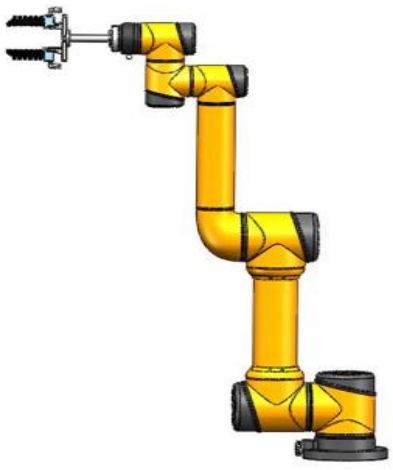

Fig. 1 Structure of the Aubo_I5 Robot

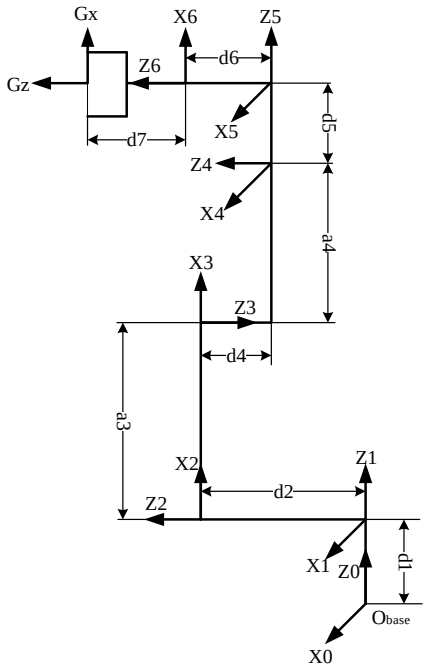

Fig. 2 Robot Joint Coordinate System

Fig 2: Establishing the Joint Coordinate System of Aubo_I5

Based on the joint coordinate system, the corresponding D-H parameter table is shown in Table 1 [4], where each link parameter is defined as follows:

α i-1 represents the distance moved along Xi-1 axis from Zi-1 to Z_i;

ai-1 represents the angle rotated around Xi-1 axis from Zi-1 to Z_i;

d_i represents the distance moved along Z_i axis from Xi-1 to X_i;

θ _i represents the angle rotated around Z_i axis from Xi-1 to X_i.

Table 1 D-H Parameter Table for AUBO_I5 Robot

i	α_{i-1}/deg	a_{i-1}/mm	d_i/mm	θ_i/deg
1	0	0	$d_1 = 98.5$	θ_1
2	90°	0	$d_2 = 140.5$	θ_2
3	180°	$a_3 = 408$	0	θ_3
4	180°	$a_4 = 376$	$d_4 = -19$	θ_4
5	−90°	0	$d_5 = 102.5$	θ_5
6	90°	0	$d_6 = 94$	θ_6

2. 1 Robot Forward Kinematics Analysis

Forward kinematics analysis involves determining the pose relationship between the tool coordinate system of a robot relative to the polar coordinate system, given the known motion angles of each joint. This is accomplished using the homogeneous transformation matrix ${}^{i-1}_{i}T$ between adjacent link pairs, as shown in Equation (1):

$$ {}^{i-1}_{i}T = \begin{bmatrix} C\theta_i & -S\theta_i & 0 & a_{i-1} \\ S\theta_i C\alpha_{i-1} & C\theta_i c\alpha_{i-1} & -S\alpha_{i-1} & -S\alpha_{i-1}d_i \\ S\theta_i S\alpha_{i-1} & C\theta_i S\alpha_{i-1} & C\alpha_{i-1} & C\alpha_{i-1}d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1) $$

In the equation: $C\theta_i = cos\,\theta_i$, $S\theta_i = sin\,\theta_i$

According to the D-H parameter table of Aubo_I5, calculate the homogeneous transformation matrices for each of the seven joints ${}^{0}_{1}T$、${}^{1}_{2}T$、${}^{2}_{3}T$、${}^{3}_{4}T$、${}^{4}_{5}T$、${}^{5}_{6}T$、${}^{6}_{7}T$ As shown below:

$$ {}^{0}_{1}T = \begin{bmatrix} C1 & -S1 & 0 & 0 \\ S1 & C1 & 0 & 0 \\ 0 & 0 & 1 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2) $$

$$ {}^{1}_{2}T = \begin{bmatrix} C2 & -S2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ S2 & C2 & 0 & d_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3) $$

$$ {}^{2}_{3}T = \begin{bmatrix} C3 & -S3 & 0 & a_3 \\ -S3 & -C3 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4) $$

$$ {}^{3}_{4}T = \begin{bmatrix} C4 & -S4 & 0 & a_4 \\ -S4 & -C4 & 0 & 0 \\ 0 & 0 & -1 & d_4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5) $$

$$ {}^{4}_{5}T = \begin{bmatrix} C5 & -S5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -S5 & -C5 & 0 & d_5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6) $$

$$ {}^{5}_{6}T = \begin{bmatrix} C6 & -S6 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ S6 & C6 & 0 & d_6 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7) $$

Combining the above formulas, the robot's forward kinematics equation can be obtained by simultaneous.

$$ {}^{0}_{6}T = {}^{0}_{1}T\,{}^{1}_{2}T\,{}^{2}_{3}T\,{}^{3}_{4}T\,{}^{4}_{5}T\,{}^{5}_{6}T = \begin{bmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8) $$

2.2 Robot Inverse Kinematics Analysis

Inverse kinematics involves calculating the joint angles θ_i of a robot given the pose of the tool coordinate system. Considering the robot's link coordinate systems, the axes of joints 4, 5, and 6 intersect at one point, satisfying the Pieper criterion, indicating the existence of a closed-form

solution. Assuming the end pose ${}^{0}_{6}T$ in Equation (8) is known, solving is performed using an analytical method, and the inverse solutions for each joint are presented in Table 2.

Table 2: Inverse solution formula of joint

$$\theta_1 = arctan\frac{u_1}{\pm\sqrt{1-{u_1}^2}} - arctan\frac{p_y}{p_x}$$

$$\theta_5 = arctan\frac{\pm\sqrt{1-{u_2}^2}}{u_2}$$

$$\theta_6 = arctan\frac{o_x S_1 + o_y C_1}{n_y C_1 + n_x S_1}$$

$$\theta_3 = arctan\frac{\pm\sqrt{1-g^2}}{g}$$

$$\theta_2 = arctan\frac{k}{\pm\sqrt{1-k^2}} - arctan\frac{u_4}{v_4}$$

$$\theta_4 = arctan\frac{v_5}{u_5} - \theta_3 + \theta_2$$

Among which:

$$u_1 = \frac{d_4 - d_5}{\sqrt{{p_x}^2 + {p_y}^2}}$$

$$u_2 = -a_x S_1 - a_y C_1$$

$$u_3 = C_1(p_x + d_6 o_x - d_5 C_6 a_x - S_6 n_x) + S_1(d_5 C_6 a_y - p_y - d_6 o_y - S_6 n_y)$$

$$g = \frac{{u_3}^2 + {v_3}^2 - {a_3}^2 - {a_4}^2}{2a_3 a_4}$$

$$u_4 = a_3 + a_4 C_3 - a_4 S_3$$

$$v_4 = a_3 + a_4 S_3 + a_4 C_3$$

$$k = \frac{u_4 + v_4}{\sqrt{{u_4}^2 + {v_4}^2}}$$

$$u_5 = C_1(o_x S_5 + n_x C_5 C_6 + a_x S_6) - S_1(o_y S_5 + n_y C_5 C_6 + a_y S_6)$$

$$v_5 = o_z S_5 + C_5(n_z C_6 + a_z S_6)$$

2.3 Mage Preprocessing

Before conducting roundness measurement, in order to eliminate the effects of metal debris, noise from the surrounding environment, and surface wear on the workpiece, the captured images need preprocessing. This involves removing noise and the influence of lighting while retaining the edge features of the workpiece [5]. Image preprocessing mainly includes:

1） Image Denoising: Using specific filtering algorithms to reduce noise in the image while preserving edge features.

2） Threshold Segmentation: Highlighting the target contour features more effectively.

(1) Image Filtering

Image filtering refers to the suppression of noise in an image while preserving detailed features, and its effectiveness directly impacts subsequent image processing. Common filtering algorithms include mean filtering, median filtering, Gaussian filtering, and others.

1） Mean filtering is a typical linear filtering algorithm that replaces the grayscale value of a particular pixel with the average pixel value in its surrounding area. Its mathematical expression is as follows:

$$g(x,y) = \frac{1}{mn}\sum_{(i,j)\in S_{xy}} f(i,j) \qquad (9)$$

Where represents the center point at, and is the size of the filtering window. By controlling the window size, the filtering effect can be adjusted. A larger value for results in a more pronounced filtering effect.

2） The median filtering algorithm is a common non-linear filter that calculates the median value of points within a neighborhood range in an image for a particular pixel, replacing the original pixel value. The mathematical expression is as follows:

$$g(x,y) = median_{(i,j)\in S_{xy}}\{f(i,j)\} \qquad (10)$$

Median filtering controls the size of to adjust the filtering effect. Compared to mean filtering, median filtering not only removes salt-and-pepper noise but also better preserves edge information.

3) Gaussian filtering is also a type of linear smoothing filter. It involves the weighted averaging process of the entire image and is primarily used to remove uneven lighting or sensor noise.

To compare the effects of the three filtering methods, experiments were conducted on the images, and the results are shown in Fig. 3.

From the results, it's apparent that after applying median filtering, not only has the original noise in the image and the interference from oil stains on the workpiece surface reduced, but it has also effectively preserved the edge features. Hence, median filtering is employed to filter the images of the pipe end face.

(2) Threshold Segmentation

Image threshold segmentation is a crucial step in segmenting the target region. Determining the threshold is a key aspect of image segmentation. If the threshold is set too high, target points may be considered as background

and ignored, while setting it too low might lead to background points being identified as target features. Common threshold segmentation methods include adaptive thresholding algorithms, Otsu's method, iterative threshold segmentation, and others. Due to varying surface qualities of the hole, there's no fixed threshold selection. Therefore, this paper conducts comparative experiments among several commonly used segmentation algorithms to choose the appropriate threshold segmentation algorithm suitable for the cooling pipe imaging environment [6].

(a)Mean Filtering

(b)Median Filtering

(c)Gaussian Filtering

Fig. 3 Comparison of filtering algs

After applying median filtering to the captured borehole regions, comparative experiments of three segmentation algorithms were conducted, and the results are shown in Fig. 4.

From the segmentation results above, it's evident that the adaptive algorithm has better segmentation results, preserving the edge features of the pipe end more completely. Therefore, this paper adopts the adaptive segmentation algorithm.

From the segmentation results above, it's evident that the adaptive algorithm has better segmentation results, preserving the edge features of the pipe end more completely. Therefore, this paper adopts the adaptive segmentation algorithm.

(a)Adaptive Algorithm

(b)OTSU Algorithm

(c)Iterative Algorithm

Fig. 4 Comparison of threshold segmentation

2.4 Edge Detection

After threshold segmentation, the next step is edge detection for the borehole. To ensure better contour detection, reducing noise interference is essential to maintain the integrity and clarity of edge information. Commonly used edge detection algorithms include the Canny algorithm, Sobel algorithm, Laplacian algorithm, etc. Compared to other edge detection algorithms, the Canny algorithm is less susceptible to noise interference [7].

The Canny algorithm first applies a Gaussian filter to smooth the image, reducing its sensitivity to noise. Then, it calculates the gradients in the X and Y directions using the Sobel operator, performs non-maximum suppression, and finally uses dual thresholds to determine edges. The Canny edge detection effect for the port block is shown in Fig. 5.

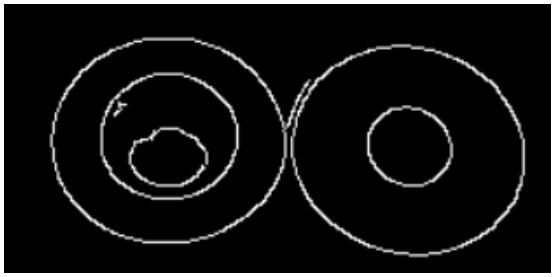

Fig. 5 Canny detection algs

2.5 Borehole Circular Detection

After edge extraction, detecting circles is a crucial part of roundness measurement. The most commonly used methods for circular detection are the Least Squares Circle Fitting and Hough Transform [8]. To measure circles accurately, it's essential to filter out the circular contours for measurement. As shown in the image, after pre-processing the image, the contour of the borehole is extracted using Canny edge detection. The contour points of the borehole are then sorted for subsequent roundness measurement.

Using the detected circle center as the reference [9], [10], the roundness error is represented by the difference between the maximum contour circle radius Rmax and the minimum contour circle radius Rmin, expressed as Rerr= Rmax − Rmin.

$$erro = \sqrt{(x_{\max} - x_c)^2 + (y_{\max} - y_c)^2} - \sqrt{(x_{\min} - x_c)^2 + (y_{\min} - y_c)^2} \quad (11)$$

among which : $(x_{\max}, y_{\max}), (x_{\min}, y_{\min})$ The coordinates of the farthest and nearest points from the fitted circle center, respectively, correspond to the contour.

The principle of the least squares fitting circle involves several steps: 1) detecting the actual contour, 2) applying the least squares fitting to find a circle, 3) creating a circle with the least squares circle center and the maximum distance to the contour, and 4) creating a circle with the least squares circle center and the minimum distance to the contour. The difference between the maximum and minimum radii of these circles is used as the basis for evaluating roundness.

The circle fitted using the least squares circle algorithm for the axle hole contour is depicted in Fig. 6. The coordinates of the fitted circle's center are (232.87733, 85.63195), and the fitted circle's radius is R = 63.36658.3

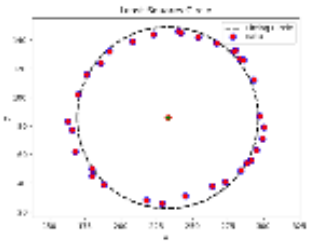

Fig. 6 Least squares circle fitting

3. Target recognition network based on YoloV4

The YoloV4 algorithm can be divided into four parts: input end, backbone feature extraction network, enhanced feature extraction network and output end. The input end includes image preprocessing, converting the size of the input image to the specified input size, and normalizing the image information. Secondly, it is the backbone feature extraction network. CSPDarknet53 formed after YoloV4 is added to CSPNet serves as the backbone feature extraction network. YoloV4 's enhanced feature extraction network consists of a PANet structure, which is used to improve the diversity of feature expressions and strengthen the fusion of feature information; the last is the output end, which is used to perform classification and regression tasks and output the final prediction results.

YoloV4 has three outputs after passing through the backbone feature extraction network, namely L3, L4 and L5. Among them, L3 and L4 will be output to the enhanced feature extraction network for corresponding feature fusion after one 1 × 1 convolution; L5 will be input to the spatial pyramid pooling layer SPP after three convolutions, and then enter the enhanced feature Extract the network for feature fusion. Finally, the final detection result can be obtained by sending it to the output terminal.

YoloV4''s backbone extraction network CSPDarknet53 is an improvement based on Darknet53. The original residual module is changed to the CSP structure as shown in Fig. 7. The CSP structure changes the original residual module into two parts, one part goes through the residual network and the other part Directly merging with the output of the residual network, this method can reduce the among of parameters and calculations while maintaining a high accuracy.

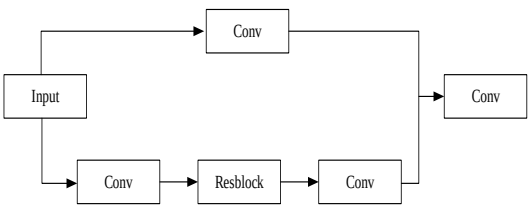

Fig. 7 CSP structure block

The PANet network is used to build a feature pyramid, analyze semantic features from top to bottom, and use the feature layer results obtained by down sampling and up sampling to reduce the loss of underlying feature information during the propagation process of the FPN

algorithm by increasing the bottom-up feature fusion path. problem [69], which improves the richness of the feature map and the detection accuracy of the algorithm. The feature map after feature fusion is predicted in three different sizes by Yolo-head. The prediction results include offset x, y, detection frame size w, h, and confidence level c of target information. Finally, redundant ones are eliminated through IOU and NMS. Detect the and NMS. Detect the frame and complete the target prediction.

3.1 Shaft hole type identification

As shown in Fig. 8, after the roundness detection of shaft holes of different sizes is completed, the YoloV4 algorithm needs to be processed to predict the category and position frame of the shaft hole. Before attitude detection, in order to remove the shaft hole target in the image, The effects of other complex backgrounds, noise, and light damage on the target surface existing in the sorting environment can be used for subsequent shaft hole type identification. We have preprocessed the image above, and then use YoloV4 to predict the image.

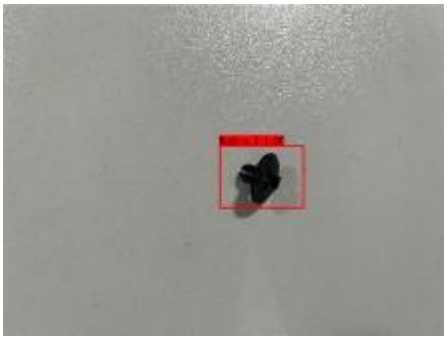

Fig. 8 Shaft hole image after inspection by YoloV4

4 Experiments and Analysis

4.1 Dataset and environment configuration

The data set used in the experiment of this article contains a total of 6 categories, with 600 data for each category. In order to prevent overfitting due to too little data, this article uses Gaussian noise and random color transformation for data enhancement. The final data the quantity is 3600 sheets.

The training and testing environment of this article is running on a Windows 10 system, NVIDIA GeForce RTX 4060TI GPU, and 8G video memory. The 3600 data sets are divided, 90% of which are used for training and 10% for verification. Train for 100 epochs, and freeze the next 50 epochs. The initial value of the learning rate for the first 50 epochs is 0.01, and the batch size is 32. The initial value of the learning rate for the last 100 epochs is 0.001, and the batch size is 16. The Adam optimizer is used during training, and the cosine annealing learning rate decay strategy is adopted.

4.2 Experimental results and analysis

From the global loss function curve in Fig. 9, we can see that the loss curve drops rapidly in the early stages of iteration, indicating that the model is fitting quickly and the model's learning efficiency is high. When epoch is 40, the network model gradually becomes stable. After 100 iterations of training, the loss function of the final model converged to 1.50. In addition, by observing the consistency of the loss curves of the training set and the verification set, it can be seen that the generalization ability of the model has reached the best state.

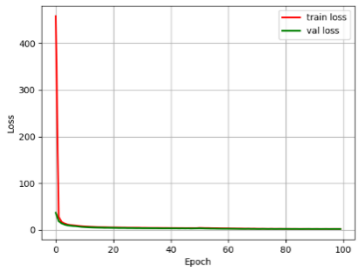

Fig. 9 Loss value curve

Fig. 10 shows the detection effect of the Yolov4 model. It can be seen that the network model accurately identified all buckles without missing detection. It can be seen from the above results that this method has a high accuracy in detecting targets and is practical.

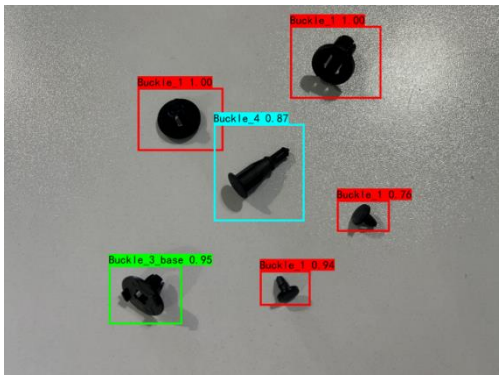

Fig. 10 Yolov4 detection results

In this experiment, a plastic bottle was placed next to the shaft hole as an interference object through the above steps, and 50 sorting tests were conducted, in which the positions of the plastic bottle and the shaft hole were randomly placed. The robot grabbing effect is shown in Fig. 11. The robot successfully avoided other interference and grasped the target shaft hole.

Fig. 11 Robotic arm grabs shaft hole

5 Conclusion

The vision-based automobile shaft hole roundness detection system proposed in this article improves the applicability of detection and can be widely used in roundness error detection of various workpiece nozzles. In the image pre-processing stage, by comparing various threshold segmentation algorithms and filtering algorithms, the adaptive threshold segmentation algorithm combined with the median filter algorithm was selected, which can not only remove the noise in the image, but also retain the edge features of the nozzle. The Canny detection algorithm is used to identify the contour of the nozzle, and the least squares method is used to detect the roundness error of the nozzle, and then combined with the YoloV4 algorithm to complete type recognition, and finally the robot completes the shaft hole identification and grabbing work. The results show that the designed detection system can not only improve the real-time performance and efficiency of detection, but also complete the subsequent grabbing work.

References

1 CHURSIN Y A, REDKO L A, FEDOROV E M. Enlargement of measuring zone in laser gauges without sacrificing measurement accuracy［J］. Measurement, 2019, 131: 647－653.

2 LIAN M . Methods and key technologies of automatic ultrasonic thickness measurement of large thin－wall parts［D］. Dalian: Dalian University of Technology, 2019 .

3 URBAN S, WURSTHORN S, LEITLOFF J, et al . MultiCol bundle adjustment: a generic method for pose estimation, simultaneous self－calibration and reconstruction for arbitrary multi － camera systems［J］. International Journal of Computer Vision, 2017, 121(2) : 234－252 .

4 CRAIG J J. Introduction to Robotics［M］. Yun Chao, waiting for translation. Beijing: Machinery Industry Press, 2006.

5 CHILLARÓN M , VIDAL V , VERDÚ G . Evaluation of image filters for their integration with LSQR computerized tomography reconstruction method［J］. PLOS One, 2020, ne, 2020, 15(3) : e0229113.

6 FU S J, ZHONG W B, LIANG L Z. Optical fiber diameter non－contact measurement with computer image processing［J］. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2000, 39(2) : 24－27.

7 LI Y M , TU J K , XIANG H Z , et al . Measurement of optical fiber geometric parameters with Canny operator and binarization filtering［J］. Optical Technique, 2018, 44(5) : 513－518.

8 SHEN X J, DUAN X Y, YUAN W L, et al. Research on circle detection algorithm based on connected region labeling algorithm［J］. Computer Engineering and Applications, 2018, 54(21) : 95－98.

9 LEI X Q, ZHANG Y D, MA W S, et al. Least square fitting and error evaluation of the convex contour of bearing roller［J］. Optics and Precision Engineering, 2018, 26(8) : 2039－2047.

10 CAI Z, WANG J L, LYU M F, et al. Roundness error assessment based on improved cuckoo search algorithm ［J］. Modular Machine Tool & Automatic Manufacturing Technique, 2020(7) : 40－44.

Authors Introduction

Junsheng Zhang

In 2021, he will be admitted to the School of Electronic Information and Automation, Tianjin University of Science and Technology, China, majoring in electronic information. He is currently a master's student at Tianjin University of Science and Technology, China.

Mr. Dongpo Ma

H He received his bachelor's degree in physics and optoelectronic engineering from Taiyuan University of Technology in 2019 and is now a graduate student in Tianjin University of Science and Technology

Dr. Yizhun Peng

He is an Associate Professor in Tianjin University of Science &Technology. He received a doctor’ s degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Science, in 2006. His research field is intelligent robot and intelligent control.