Target Positioning Based on Binocular Vision

Ronghua Zhu[1(✉)] and Enyu Hou[2]

[1] School of Physical Science and Information Engineering, Liaocheng University, Liaocheng 252000, China
1045104740@qq.com
[2] SAS Medical Technology (Beijing) Co., LTD, Beijing 100044, China
houey@qq.com

Abstract. In order to improve the accuracy of workpiece positioning in the manufacturing process, this paper presents a binocular vision technology to identify the target object and locate the target area. Firstly, a novel and effective HALCON-based stereo calibration method is proposed to solve the problem of non-standard external polar line geometry of binocular system. Subpixel-accurate-based template matching with scaling and image pyramid algorithm are presented, and target objects are identified accurately and feature points are extracted accurately. Secondly, the stereo matching of feature points can be completed quickly according to the polar line constraint and gray-value-based template matching using the normalized cross-correlation (NCC). Finally, stereo reconstruction of feature points is completed by the combination of the binocular parallax principle and 3D coordinate affine transformation. The experimental results show that the detection radius error is less than 0.4 mm, and the error rate is less than 3%; the positioning depth error is less than 0.2 mm, and the error rate is less than 0.5%.

Keywords: Stereo vision · System calibration · Sub-pixel shape template matching with scaling · Pyramid search strategy · Stereo matching · Stereo reconstruction binocular system calibration

1 Introduction

With the rise of the industrial 4.0 strategy, the combination of vision and robotic systems has become an important means to improve the intelligence of robots [1]. In the current practical industrial applications, 2D vision technology is often used in combination with robots, but the two-dimensional image almost loses all the depth information of the object and it is difficult to obtain the three-dimensional information of the target. Therefore, it is necessary to reconstruct the three-dimensional information of the target from the two-dimensional image, for more comprehensive and true reflection of objective objects, and further improving the intelligence of the robot system [2]. Because binocular vision has the advantages of high efficiency, high precision, non-contact and depth information, it can be widely used in target recognition and positioning, and has important research significance for the precise positioning of mass production workpieces.

Y. Zhao et al. (Eds.): ICIG 2019, LNCS 11903, pp. 157–168, 2019.
https://doi.org/10.1007/978-3-030-34113-8_14

In this paper, we propose an recognition and localization algorithm based on HALCON binocular stereo vision technology. The main contribution of this paper is in three respects. First, the stereo correction of the target image is achieved by binocular system calibration. Second, considering various linear, nonlinear illumination changes and occlusion factors, this paper propose a pyramid search strategy and sub-pixel shape template matching algorithm with scaling to achieve accurate extraction of feature points. This algorithm can effectively cope with various linear and nonlinear illumination changes according to the gradient correlation of the edge of the object and it has strong resistance to occlusion and partial deletion. Lastly, we can quickly complete stereo matching of feature points through a normalized cross-correlation (NCC) based grayscale template matching algorithm, even if there is illumination variation in the image. The experimental results show that the proposed algorithm can realize the high-precision positioning of the target object under the premise of real-time and high efficiency, and the feasibility of the method can be verified by experiments.

2 Binocular Stereo Vision Positioning Principle

Binocular stereo vision can perceive the depth information of the three-dimensional world by simulating human eyes. By using any point in the space at the imaging position of the left and right cameras, the feature point matching relationship and the principle of triangular geometry, the parallax can be calculated to obtain the information of the object's three-dimensional space [3].

In this experiment, we use an axis parallel system structure consisting of two cameras. The physical map is displayed in Fig. 1a, and the binocular system model is shown in Fig. 1b. Assume that the focal lengths of both cameras are f, and the distance between the projection centers is b (also called the baseline). O_luv, O_ruv are two imaging plane coordinate systems whose coordinate directions coincide with the x-axis and y-axis directions, respectively. To simplify calculations, we take the coordinate system of the left camera as the world coordinate system $O - XYZ$. The image coordinates of the space point $P(x, y, z)$ on the left and right camera imaging planes are $P(u_l, v_l)$ and $P(u_r, v_r)$, respectively. The coordinates in the left and right camera coordinate systems are $P(x_l, y_l, z_l)$ and $P(x_r, y_r, z_r)$, respectively. Since the imaging planes of the two cameras are on the same plane, the image coordinates of the spatial point $P(x, y, z)$ have the same v coordinates, that is, $v_l = v_r = v$. According to the triangular geometry, we have:

$$u_l = f\frac{x_l}{z_l} \quad u_r = f\frac{(x_l - b)}{z_l} \quad v_l = v_r = f\frac{y_l}{z_l} \tag{1}$$

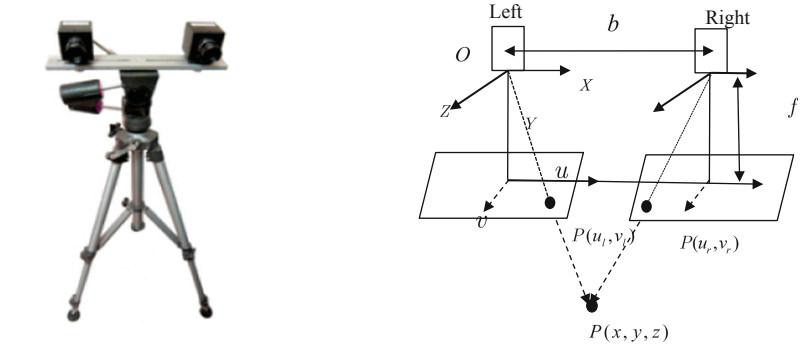
Since the left and right images are on the same plane, the disparity value of the corresponding point is defined as the difference between the coordinates of the corresponding point column in the left and right images. Hence, disparity can be computed:

$$D = u_l - u_r = \frac{b \times f}{z_l} \tag{2}$$

Combining (1) and (2), the formula for calculating the three-dimensional coordinates of the spatial point can be represented by the Eq. (3):

$$x = x_l = \frac{b \times u_l}{d} \quad y = y_l = \frac{b \times v}{d} \quad z = z_l = \frac{b \times f}{d} \tag{3}$$

Where $z = z_l$ is the depth distance of the space point P From the above formula, we can see that before the three-dimensional coordinates of the spatial point are obtained, the task to be solved by stereo vision positioning is to determine camera parameters and conjugate points.



a Binocular stereo system physical map

b Binocular camera model

Fig. 1. Binocular stereo system

3 Binocular System Calibration and Stereo Rectification

3.1 System Calibration

Define camera calibration is a crucial step in stereo imaging [4]. Camera calibration in binocular system is similar to single camera calibration. Firstly, the internal and external parameters of the two cameras are obtained by single camera calibration, and then the positional relationship between two cameras is obtained by using the external parameters of the two cameras.

Camera calibration refers to establish the relationship between the pixel coordinates of the camera image and the three-dimensional coordinates of the scene point. According to the camera model, the internal and external parameters of the camera are solved by the image coordinates and world coordinates of the known feature points [5]. Establishing a camera imaging model, that is, the model parameters is solved by

projection relationship, experiment and calculation method. Throughout the projection process, the conversion relationship between the image pixel coordinate system and the world coordinate system is:

$$
\begin{aligned}
Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= Z_c \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{f}{dx} & 0 & u_0 \\ 0 & \frac{f}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
\end{aligned} \tag{4}
$$

Where: uv is the image pixel coordinate system, d_x, d_y is the pixel unit and f is the distortion coefficient. $O_cX_cY_cZ_c$ is the camera coordinate system and $O_wX_wY_wZ_w$ represents the world coordinate system, M_1, M_2 represents the internal and external parameters of the camera.

This experiment uses Zhengyou Zhang calibration method to calibrate the system. We use two identical cameras and lenses, which is the M-1614MP2 industrial lens and model MV-VS120 CCD color industrial cameras by Vision Digital Image Technology Co., Ltd. Combined with the HALCON software platform, calibration of the camera is achieved via using its algorithmic dynamic library. The calibration plate processing is shown in Fig. 2. The internal and external parameters of the two cameras and the relative positional relationship between the cameras are determined by averaging 20 calibrations. The internal parameter calibration results are shown in Table 1. Table 2 shows the external parameters of the two cameras before and after calibration.

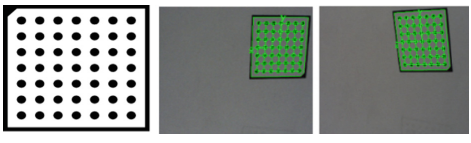


Fig. 2. Calibration plate image processing

Table 1. Calibration results of internal parameter

Internal parameter	f(mm)	k	s_x(m)	s_y(m)	Row (pixel)	Column (pixel)
C1 before correction	14.85	−1080.1	4.6240e−6	4.6500e−6	674.067	685.180
C2 before correction	14.83	−1081.1	4.6256e−6	4.6500e−6	687.468	580.303
C1 after correction	14.15	0.0	4.6500e−6	4.6500e−6	647.319	670.510
C2 after correction	14.15	0.0	4.6500e−6	4.6500e−6	2220.34	670.510

Table 2. Calibration results of external parameter

External parameter	Translation matrix T			Rotation matrix R		
	x	y	z	x	y	z
Before correction	10.37	0.05	0.03	336.4	358.8	0
After correction	10.97	0	0	0	0	0
Error (pixel)	0.29					

According to the above data, the error level is lower than one pixel, and the calibration accuracy is high. After correction, the column coordinates of the pixel points of the two images are equal, and the position of the right image relative to the left image is only translated in the X-axis direction. This shows that the corrected binocular positioning system is a standard external polar line geometry [6], which can greatly save the time of stereo matching.

3.2 Stereo Rectification

After the system calibration, we can calculate the two correction maps by using the internal and external parameters of the two cameras and the relative positional relationship between the two cameras. The two map images combining functions map_image() is used to rectify the acquired stereo image pairs to the polar standard geometry. The rectified images of two cameras are shown in Fig. 3.

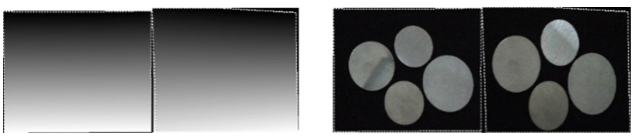


Fig. 3. Stereo rectification

4 Target Recognition and Feature Point Extraction

To improve the accuracy and speed of the matching algorithm in the process of object identification and position detection, subpixel-accurate-based template matching with scaling and image pyramid algorithm are applied in this paper, which is robust to occlusion, chaos, nonlinear illumination changes and contrast global inversion.

4.1 Subpixel Edge Extraction

For subpixel accurate contour extraction, the edge is extracted by a combination of canny operator and subpixel edge detection, which takes the image as input and returns to the XLD contour. Through the edges_sub_pix() operator, the canny filter is used to detect the gradient edge, and the canny operator repeats the gray value at the image boundary to obtain the optimal filter width by the Alpha option. This maintains greater noise invariance and enhances the ability to detect small details. The edge detection effect is shown in Fig. 4.

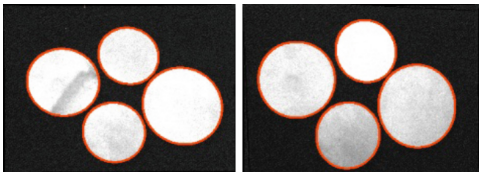


Fig. 4. Edge detection

In this paper, the Tukey weight function is used to fit through three iterations. The Tukey weight function is defined as:

$$\omega(\sigma) = \begin{cases} [1 - (\frac{\sigma}{\tau})^2], & |\sigma| \leq \tau \\ \frac{\tau}{|\sigma|}, & |\sigma| > \tau \end{cases} \tag{5}$$

Where: parameter τ represents the distance threshold. When the distance from the point to the circle is greater than the threshold, the weight function is equal to the reciprocal of the distance multiplied by the threshold. When the distance from the point to the circle is less than or equal to the threshold, the weight is the square of the difference between the square of the distance divided by the threshold and 1.

4.2 Shape Template Matching and Feature Point Extraction

To improve the speed and accuracy of the matching algorithm, support X/Y direction scaling and nonlinear illumination changes, subpixel-accurate-based template matching with scaling algorithm to detect position are applied in this paper. The algorithm is based on the direction vector of the edge point obtained by Sobel filtering, and defines

the similarity measure. Furthermore, combined with the image pyramid hierarchical search strategy, the shape information is used for template matching.

The similarity measure of the shape-based template matching algorithm is the sum of the gradient vector of the point in the template and the gradient vector of the point in the image, the similarity measure s:

$$S = \frac{1}{n}\sum_{i=1}^{n} d_i^T e_{q+p} = \frac{1}{n}\sum_{i=1}^{n} t_i v_{r+r_i,c+c_i} + u_i w_{r+r_i,c+c_i} \tag{6}$$

Where: d is the gradient vector of the point in the template, and e is the gradient vector of the point in the image.

Normalized similarity measure s:

$$S = \frac{1}{n}\sum_{i=1}^{n} \frac{d_i^T e_{q+p}}{\|d_i^T\|\|e_{q+p}\|} = \frac{1}{n}\sum_{i=1}^{n} \frac{t_i v_{r+ri,c+ci} + u_i w_{r+ri,c+ci}}{\sqrt{t_i^2 + u_i^2}\sqrt{v^2_{r+ri,c+ci} + w^2_{r+ri,c+ci}}} \tag{7}$$

Since the gradient vector is normalized, the similarity measure s will return a value less than or equal to 1. When $s = 1$, the template corresponds to the image one-to-one.

In the image matching process, each potential point of the search image is subjected to a normalization calculation of one traversal, the calculation amount is very large. Therefore, to improve the speed of the algorithm, it is necessary to try to reduce the number of poses examined and the points in the template, and the pyramid hierarchical search strategy can simultaneously reduce the two parts to improve the operation speed [7]. The image pyramid is shown in Fig. 5.

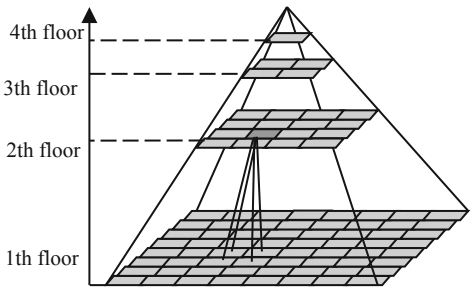
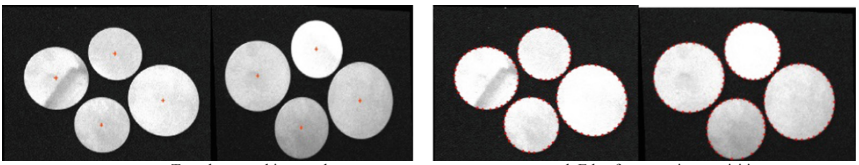


Fig. 5. Pyramid image

Performing the same edge detection and filtering on each layer of image when creating the template, and then searching from top to bottom layer by layer until the similarity measure is greater than the threshold, and finally we can get the row and column coordinates of the matching template. The results of workpiece recognition and center extraction are shown in Fig. 6a. The cross mark is the central feature point The fitted XLD edge is segmented by a basic geometric element such as a straight line and

an arc by a Ramer [8] algorithm for a large number of subpixel edge coordinate data. The edge feature points obtained are shown in Fig. 6b.



a Template matching results b Edge feature point acquisition

Fig. 6. Feature point extraction

5 Target Three-Dimensional Position

5.1 Stereo Matching

Stereo matching is the most critical step in the binocular vision algorithm. Its main task is to find the corresponding relationship of the same point in space in different images under different observation angles [9]. To complete the stereo matching of feature points quickly, we use the polar line constraint and region matching algorithm based on normalized cross-correlation (NCC). Since the two cameras have different viewing angles, the illumination will also have a certain difference, and a method that does not change with the change of illumination is needed, that is, the normalized cross-correlation (NCC) [7]. The NCC is defined as:

$$NCC(x,y,d) = \frac{1}{(2n+1)(2m+1)} \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m}([I_L(x+i,y+j) - \overline{I_L(x,y)}][I_R(x+i,y+j+d) - \overline{I_R(x,y+d)}])}{\sqrt{\delta^2(I_L) \times \delta^2(I_R)}} \tag{8}$$

Here, $\overline{I(x,y)} = \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m} I(x+i,y+j)}{(2n+1)(2m+1)}$ is the average gray value of all pixels in the neighborhood of the current location search point, $\delta(I) = \sqrt{\frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m} I^2(x,y)}{(2n+1)(2m+1)} - I(y,y)}$ represents the variance. The value of NCC is $[-1, 1]$. When the absolute value of NCC is larger, it indicates that the sub-window is more closely matched with the neighborhood of the search point. When $NCC = 1$ it indicates that the two polarities are the same; when $NCC = -1$ the two polarities are opposite, this means that the results of the normalized correlation coefficients are not affected by linear illumination changes [10].

The stereo matching procedure can be summarized as the following steps: Centering the pixel point P_L to be matched in the image, and interceping a rectangular sub-window B_L. In the image to be matched searching for window B_R, which is most similar to the gray value of B_L according to the principle from left to right and from top

to bottom. Calculating the center P_R of the window B_R, and getting the matching pixel points P_L and P_R in the left and right image (see Fig. 7).

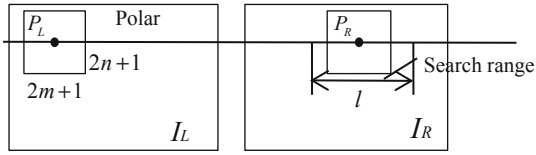


Fig. 7. Template matching schematic

According to the above steps, the point to be matched of the first feature point of the left image is searched; the point to be matched of the second feature point is searched… until the feature points of the left image are traversed, that is, the stereo matching task is completed. The stereo matching result is shown in Fig. 8.

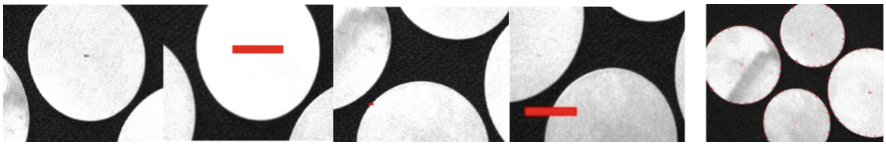


Fig. 8. Stereo matching

5.2 Stereo Reconstruction and 3D Affine Transformation

After the system calibration, stereo correction and stereo matching are completed, our next task is to calculate the depth information of the target. The three-dimensional reconstruction of the binocular system is to acquire two images of the scene simultaneously by two cameras and find the matching point pairs of the same point in the two images in space. The three-dimensional coordinates of the point can be obtained by combining the principle of binocular vision imaging [11]. The three-dimensional point cloud map is shown in Fig. 9.

Fig. 9. The three-dimensional point cloud map

To improve the accuracy of the evaluation positioning system, we need to perform 3D affine transformation on the 3D point cloud coordinates of 3D reconstruction and convert it into the world coordinates under the left camera as the reference coordinate system. The principle expression is given by:

$$\begin{pmatrix} Q_z \\ Q_y \\ Q_x \\ 1 \end{pmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot \begin{pmatrix} P_z \\ P_y \\ P_x \\ 1 \end{pmatrix} = (R \cdot \begin{pmatrix} P_z \\ P_y \\ P_x \\ 1 \end{pmatrix} + T) \tag{9}$$

Where (P_x, P_y, P_z) is the input point and returns the resulting point to (Q_x, Q_y, Q_z).

6 Experimental Results and Analysis

To detect the size of the target object, each edge point and the corresponding center can be calculated by the space curve fitting formula $l = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$ to complete the detection of the workpiece radius. The experimental results are shown in Fig. 10.

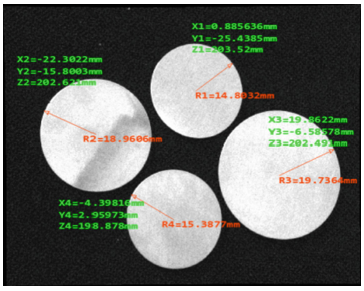


Fig. 10. 3D spatial information and detect results

To analyze the reconstruction accuracy of the algorithm, we compute the error between the four target detection data and the actual data. Table 3 shows comparison results. The measured actual radius error of the workpiece is less than 0.4 mm, and the error rate is less than 3%; the actual depth of the workpiece is less than 0.2 mm, and the error rate is less than 0.5%. It shows that the target positioning method can accurately locate the target object with high precision, which verifies the feasibility and accuracy of the method under certain conditions.

Table 3. Comparison of test results with actual results

Number	Measuring Z(mm)	Actual Z (mm)	Error (mm)	Error rate (%)	Measuring R(mm)	Actual R (mm)	Error (mm)	Error rate (%)
1	203.520	204.5	1.020	0.499	14.8032	15.00	0.1968	1.312
2	202.621	202.0	0.621	0.307	18.9606	18.75	0.2106	1.123
3	202.491	203.5	1.009	0.496	19.7364	20.00	0.2636	1.318
4	198.878	199.3	0.422	0.212	15.3877	15.00	0.3877	2.585

7 Conclusion

In this paper, we study the whole process of target recognition and localization based on binocular vision. In the target recognition and feature point extraction stage, the canny sub-pixel edge extraction, edge fitting and sub-pixel shape template matching algorithm with scaling are used to target the target area, which has a great improvement in feature point extraction speed. In the target positioning stage, we use the polar line constraint and region matching algorithm based on NCC gray correlation to complete the stereo matching of the feature positioning points, which solves the problem that the target feature points in the left and right images do not match. Finally, the three-dimensional positioning of the target object is realized by the principle of three-dimensional reconstruction. The experimental results show that the whole process method improves the speed of target recognition and the accuracy of feature point extraction, reduces the possibility of matching errors or repeated matching, and effectively achieves accurate positioning of the target object. The accuracy can be better satisfied in the working space of the robot, which is more conducive to the three-dimensional reconstruction and positioning of the robot vision system.

According to the current experimental research situation, and for the complex and varied manufacturing environment of the workpiece, we need further research and experimentation to realize a more general and effective image recognition and localization algorithm. It is necessary to improve the existing experimental positioning results and obtain more accurate position information of the target object in three-dimensional space.

References

1. Huang, N., Liu, G., Zhang, Y., et al.: Unmanned aerial vehicle vision navigation algorithm. Infrared Laser Eng. **45**(7), 269–277 (2016)
2. Men, Y., Ma, Y., Zhang, G., et al.: A stereo matching algorithm based on Census transform and improved dynamic programming. J. Harbin Inst. Technol. **47**(3), 60–65 (2015)
3. Shen, T., Liu, W., Jing, W.: Target ranging system based on binocular stereo vision. Electron. Measur. Technol. **38**(4), 52–54 (2015)
4. Chen, X., Wei, Y.: Target positioning based on binocular stereo vision. Autom. Technol. Appl. **56**(2), 224–229 (2017)
5. Yu, D., Wang, Y., Mao, J., et al.: Vision-based object tracking method of mobile robot. J. Optoelectron.·Laser **40**(1), 227–235 (2019)

6. Zhang, X., Wu, B.: Research on three-dimensional information acquisition technology of small field using integral imaging. J. Optoelectron. Laser **28**(11), 1240–1245 (2017)
7. Ruru, Z., Guangying, G., Zhe, S., et al.: 3D reconstruction based on binocular stereo vision. J. Yangzhou Univ. (Natural Sci. Ed.) **21**(3), 5–10 (2018)
8. Lin, H., Pan, W.: Research and application of image edge feature extraction algorithm based on variable weighted least squares method. Combined Mach. Tool Autom. Process. Technol. **6**(2), 66–71 (2015)
9. Zhang, J., Wu, S., Chen, B., et al.: Binocular vision based multi-dimensions on-line measuring system for workpieces. Instrum. Tech. Sensor **32**(10), 75–80 (2018)
10. Guo, A., Xiao, D., Zou, X.: Computation model on image segmentation threshold of litchi cluster based on exploratory analysis. J. Fiber Bioeng. Inform. **7**(3), 441–452 (2014)
11. Lu, B., Liu, Y., Sun, L.: Error analysis of binocular stereo vision system based on small scale measurement. Acta Photonica Sinica **8**(2), 232–237 (2015)