Workpiece Recognition and Grasping System Based on Binocular Stereo Vision

Jinghua Liu [a], Wenjuan Li [b], Hui Li [c], Kaige Yang [d]

Shandong University of Science and Technology, Qingdao 266590, China.

[a]1220204543@qq.com, [b]937358742 @qq.com, [c]873519125 @qq.com, [d]2541090225@qq.com

Abstract

In order to meet the requirements of industrial automation, the robot is equipped with "double eyes" to allow it to identify and grab target objects in a certain environment. Using a binocular camera to complete the image acquisition, the SURF_FREAK algorithm is used in the computer to extract features and initially match the workpiece. Under the condition of polar line constraints, the region matching of the workpiece is performed to find a stable matching point pair. In combination with the camera calibration result, the three-dimensional coordinates of the workpiece are obtained, and the pose information of the workpiece is output. This pose information is converted to the robot base coordinate system through Kinematic inverse solution to get the amount of movement of each joint, guide the robot to grab the workpiece in real time. The vision system and the robot grab system were coordinated in the Matlab 2016b and VS 2010 environments, and the algorithm was verified with Vision Library Software Open CV. The experimental results show that the SURF_FREAK algorithm has good real-time performance and robustness in this experiment. It has a great improvement in recognition speed and matching accuracy, and the experimental platform built can successfully let the robot complete the capture of target parts under visual guidance.

Keywords

Binocular Stereo Vision, SURF_FREAK algorithm, Target Recognition, Target Setting.

1. Introduction

There are many methods that can be used for feature matching through the acquired images. The SURF algorithm has been paid attention by researchers because of its good robustness and strong anti-interference. In terms of local feature detection, following the SIFT, SURF, ORB, and BRISK, the FREAK algorithm was proposed by Alexandre Alahi et al. in ICCV on 2012. Its outstanding feature is to introduce the visual mechanism of the retina of the human eye into the random point pair sampling mode, using Saccadic Search to match from coarse to fine. The calculation speed of FREAK algorithm has obvious advantages, but it is not very satisfactory in the matching results. Therefore, SURF_FREAK algorithm is proposed for this situation. The workpiece grabbing experiment platform based on the binocular stereo vision was constructed in this paper, applying SURF_FREAK algorithm to deal with the images collected by the camera, and the whole experimental process is guided by the kinematic inverse solution.

2. Structure of Workpiece Recognition and Robot Grab System Based on Binocular Stereo Vision

As a concrete application in the industrial field, the positioning and recognition technology of the workpiece can accurately measure the coordinates and position of the workpiece. The target recognition and grasping system based on binocular stereo vision in this paper can be divided into binocular stereo vision subsystem and robot grabbing subsystem. In the vision subsystem, in order to accurately identify the target parts, it is necessary to calibrate the camera and adopt an efficient extraction and matching algorithm to complete the identification and positioning of the workpiece. In robot grabbing system, it is necessary to complete the connection among each part of the hardware, establish a D-H model in conjunction with robot kinematics to solve the movement offset of each joint, and then control the robot arm to complete the grabbing process. Based on the six-degree-of-freedom robotic arm, the framework of the binocular stereoscopic vision recognition and grasping system for the workpiece is shown in Figure 1.

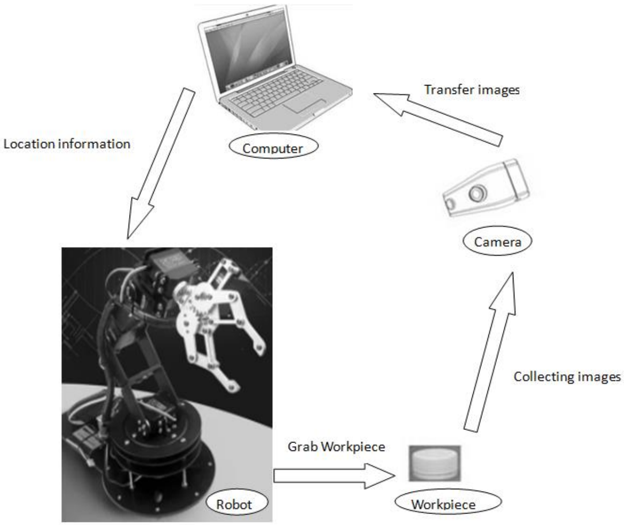

Fig. 1. System structure

3. Vision System Work Process

3.1 Camera Calibration

The role of the camera in the vision system is to capture the target image and the camera needs to be calibrated after installation. The entire process uses the Zhang Zhengyou calibration method combined with the least squares method to obtain the camera's internal and external parameters: Calibrate a single camera to determine the internal parameters of each camera. At the same time, three camera calibrations are performed to obtain the camera's external parameters. Camera calibration can determine the position of the target in the camera image coordinate system and the space coordinate system.

3.2 Target Identification and Matching

It is an important process to select and extract feature points to realize the accurate identification and positioning of target workpiece. In this paper, by combining the SURF and FREAK algorithm, relatively stable feature points can be extracted and preliminary matching can be performed.

SURF is a fast and robust local feature extraction algorithm that can extract a large number of feature points and achieve matching, however, randomness is too high in matching; FREAK algorithm simulates the visual system of the human eye, although the matching speed is improved, the matching rate is extremely low. Therefore, this paper combines the advantages of the two algorithms and uses the SURF_FREAK target recognition and matching algorithm.

First constructing the Hessian matrix, SURF uses Hessian matrix determinant approximation to improve the detection speed. Since feature points need to be scale-independent, Gaussian filtering is required before constructing a Hessian matrix. After filtering, the Hessian matrix of the image point I(x, y) of dimension σ is defined as:

$$H_{ession} = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{yx}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{1}$$

In the process of image feature extraction, SURF uses box filters with different sizes. In order to balance the approximation error, the determinant can be written as:

$$\det(H) = D_{xx}(x)D_{yy}(x) - (0.9D_{xy}(x))^2 \tag{2}$$

The positive and negative values of the Hessian matrix determinant can be used to determine whether the point is an extreme point. The scale space of SURF is composed of O group L layers. The size of the box filter template used between different groups continuously increases, while the size of the image remains unchanged. The response values of the Hessian matrix can be obtained by using box filters of different sizes and integral images, so that feature points of different scales are calculated on the corresponding images. Comparing the size of each pixel processed by the Hessian matrix with the 26 points of its 3-dimensional field, if it is the maximum value or the minimum value, it remains as a preliminary feature point. The 3D linear interpolation method is used to obtain the sub-pixel level feature points, at the same time, the points whose values are smaller than a certain threshold are also removed, and the accurate positioning of the feature points is achieved.

Adding direction information for each feature point ensures that the algorithm has directional rotation invariance. The FREAK algorithm is inspired by the human eye's visual system in which foveola is located in the middle region of human photoreceptor cells and receives high-resolution images. The smaller the distance from foveola, the greater the number of ganglion cells in the receptive field.

The algorithm uses 45 long and symmetrical sampling points to calculate its gradient, as shown in Figure 2.

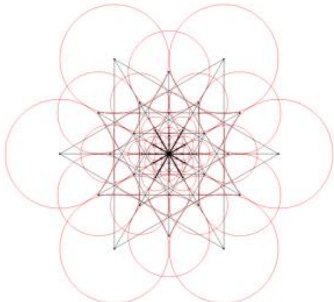

Fig 2. FREAK algorithm sampling mode

The gradient calculation formula is:

$$O=\frac{1}{M}\sum_{P_o}\left(I\left(P_O^{r_1}\right)-I\left(P_O^{r_2}\right)\right)\frac{P_O^{r_1}-P_O^{r_2}}{\left\|P_O^{r_1}-P_O^{r_2}\right\|} \tag{3}$$

Where M is the logarithm of the match, $P_O^{r_1}$ is a two-dimensional vector of the center coordinates in receptive field. Making the feature points extracted by the SURF algorithm a centre, and concentric circles of different radii are constructed. Divide the concentric circles and make a circle around the concentric circles where the points are equally divided. The equal point is the sampling point. The FREAK descriptor uses a binary string to describe the feature point, denoted by F, then

$$F=\sum_{0\le a\le N}2^{a}T(Pa) \tag{4}$$

$$T(P\mathrm{a})=\begin{cases}1 & \left(I\left(P_a^{r_1}\right)-I\left(P_a^{r_2}\right)>0\right)\\ 0 & \text{other}\end{cases} \tag{5}$$

$T(P\mathrm{a})$ is a function, $P\mathrm{a}$ is a sampling pair, N is number of feature vectors, $I\left(P_a^{r_1}\right)$、 $I\left(P_a^{r_2}\right)$ indicates the intensity value of the sample point after Gaussian blur. The descriptor calculated by the SURF_FREAK algorithm is a 512-bit binary bit string, and the similarity measure is performed using the Hamming distance. Assume that H_1、H_2 is the corresponding feature descriptors in the two images. The Hamming distance between them is:

$$D_{\mathrm{hm}}(H_1,H_2)=\sum_{i=1}^{512}(x_i\oplus y_i) \tag{6}$$

$H_1=x_1x_2\cdots x_{512}, H_2=\mathrm{y}_1y_2\cdots y_{512}, \oplus$ indicates the value of x, y is 0 or 1 under the XOR operation. The smaller the Hamming distance value, the higher the matching rate.

3.3 Target Setting

The workpiece identified in this article are objects with relatively regular shapes, so the center position of the object is used as the final anchor point. The image is preprocessed to enhance the contrast. The Canny operator is used to perform edge detection on the processed image. After the first open and close operation, the binary image is obtained, and the average value of the detected edge feature points is taken as the final value of center. The location of the detected center of the target is shown in Figure 3.

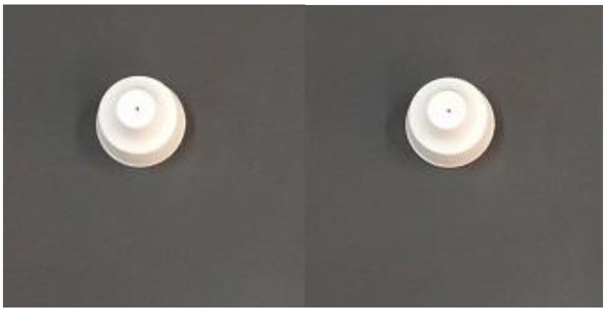

Fig 3. The central position of the workpiece in the left and right image

The spatial coordinates of the center can uniquely determine the position of the object. Assuming that the coordinates of the pair of left and right image matching points is q (u, v) and q '(u, v) the spatial coordinates of the spatial point P is determined by combining the camera calibration. In this paper, a parallel binocular stereo vision system is used. In this system, the parameters such as focal length and center point of the two cameras are the same. The coordinate system of the left camera is defined as the world coordinate system, then the coordinate of the point P in the left camera coordinate system is(x_1,y_1,z_1), and the coordinate in the right camera coordinate system is(x_1-b,y_1,z_1) available from the principle of triangulation:

$$\begin{cases} u_1 - u_0 = f_x \dfrac{x_1}{z_1} \\ u_2 - u_0 = f_x \dfrac{x_1 - b}{z_1} \\ v_1 - v_0 = v_2 - v_0 = f_y \dfrac{y_1}{z_1} \end{cases} \tag{7}$$

From the above formula can be drawn:

$$\begin{cases} x_1 = \dfrac{b(u_1 - u_0)}{u_1 - u_2} \\ y_1 = \dfrac{bf_x(v_1 - v_0)}{f_y(u_1 - u_0)} \\ z_1 = \dfrac{bf_x}{u_1 - u_2} \end{cases} \tag{8}$$

Where u_0、v_0、f_x、f_y has been found by the camera calibration, (u_1,v_1)、(u_2,v_2) are respectively the imaging coordinates of the spatial point in the left and right images. From this, we can get the three-dimensional coordinates of the spatial point (x_1, y_1, z_1).

4. Experimental Results and Analysis

4.1 Camera Calibration Experiment

Use binocular cameras to capture multiple calibration plate images at different locations. In order to reduce the impact of random errors, 12 images were used to calibrate the camera, and tests were performed using checkerboards, as shown in Figure 4. The camera's internal and external parameters and can be obtained by calibration experiment, and the root mean squared error (RMSE) of each calibration image was calculated. The error fluctuation range is shown in Figure 5.

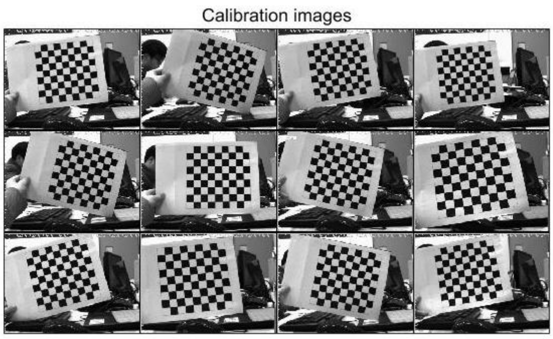

Fig 4. Check_board calibration.

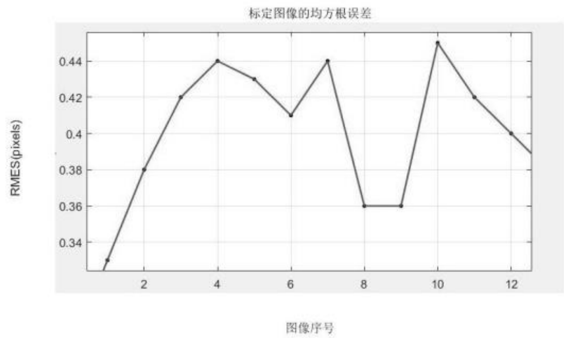

Fig 5. Root mean square error fluctuation range

From the above figure, the calibration method based on the planar template adopted in this paper mainly extracts the corner region information of the calibration plate during the calibration process. The gradient and the direction of the edge of the checkerboard are relatively easy to extract, even if there is a distortion effect of the camera, the error is controlled within 0.5 pixels, which satisfies the total error requirement of the system.

4.2 Stereo Matching Experiment Experiment

The matching points obtained by the SURF_FREAK algorithm after the Hamming distance measurement has many mismatching points. The result is shown in Fig.6 (A).After the region matching based on the polar line constraint, the exact matching pair can be obtained. The result is shown in Fig.6 (B).

(a) Initial match based on SURF_FREAK algorithm. (b) Constraint-based region matching.

Fig 6. Matching results

4.3 Three-Dimensional Positioning Experiment Experiment

Due to exit the external interference such as light and noise, the image acquired in the scene needs to be pre-processed first. Then feature points are extracted, a certain feature point of the left image is selected, this area is used as a matching primitive, search and match are performed in the right image, and then the centre coordinates of the workpiece are obtained. According to the result of the camera calibration, the three-dimensional coordinates of the workpiece can be obtained.

The camera is fixed on the bracket of the grasping platform to complete the connection of the hardware devices, and the workpiece to be grasped is placed within the visual field of the robot. Start the program, in order to verify the robot's grasping accuracy, we carried out many grab experiments, the positioning data as shown in Table 1, the positioning data of the workpiece centre is the first experimental data.

Table 1. Robot positioning experiment results

Numble	Left image	**Right image**	**Workpiece world coordinates**
1	(272,281)	(126,277)	(71,-31.9,-667.5)
2	(206,143)	(51,141)	(79.4,75.6,-628.5)
3	(337,160)	(160,-16.8)	(-16.8,62.5,-630.8)

5. Conclusion

This article uses the binocular stereo vision technology to study the workpiece recognition and grasping system. It elaborates on the construction of the system platform, camera calibration, target recognition and match, target positioning and grabbing. The SURF_FREAK algorithm is adopted for the identification of workpieces. The recognition speed and matching accuracy have been greatly improved, which can meet the real-time and accuracy requirements of field work. The three-dimensional position coordinates obtained are transmitted to the robot control system to guide the robot to complete the grasping work, thereby completing a complete industrial process through the binocular stereo vision subsystem and the robot grabbing subsystem. In actual industrial production, it is of great application value to complete the identification and positioning of the workpiece based on binocular stereo vision.

Acknowledgements

In five months, I completed the writing of the paper and conducted an experimental verification. During this period, I sincerely thank Professor Zhong Pei Si, my respected tutor, for his guidance and encouragement in the preparation of my dissertation. His careful guidance, rigorous scholarship style and tireless teachings have given me endless inspiration and guidance. Thank all the students in the lab for their help and support. At the same time, thank my reviewers for their valuable suggestions.

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. Computer Vision–ECCV 2006, pp. 404– 417

[2] Alexandre Alahi, Raphael Ortiz, Pierre Vandergheynst,FREAK: Fast Retina Keypoint, Ecole Polytechnique Fed´ erale de Lausanne (EPFL), Switzerland, 2012.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: an efficient alternative to sift or surf. 2011.

[4] Wu Yanhai, Zhang Cheng, Wang Jing. Image Registration Method Based on SURF and FREAK. Xi'an University of Science and Technology. ICSPCC2015.

[5] Wang Canjin, Sun Tao, and Chen Juan, "A novel target recognition method based on fast retina key point descriptor for laser active imaging," Chinese Journal of Lasers, vol. 41, no. 3, March, 2014, pp. 1-6.

[6] S. Leutenegger, M. Chli, and R. Siegwart, BRISK: binary robust invariant scalable keypoints, Proceedings of the IEEE International Conference on Computer Vision, pp. 2548-2555,2011.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60,no. 2, pp. 91-110, 2004.

[8] E. Rosten, and T. Drummond, "Machine learning for highspeed corner detection," European Conference on Computer Vision,pp. 430-443, 2006.