

A Novel Camera Calibration Method for Binocular Vision Based on Improved RBF Neural Network

Weike Liu¹, Ju Huo², Xing Zhou², and Ming Yang¹(✉)

¹ Control and Simulation Center, School of Astronautics,
Harbin Institute of Technology, Harbin 150001, China
myang_csc@163.com

² School of Electrical Engineering, Harbin Institute of Technology,
Harbin 150001, China

Abstract. Considering the problems that camera imaging model is complex and operation is complicated, a binocular camera calibration method of RBF neural network based on k-means and gradient method is proposed in this paper. The data center selection method based on the law of clustering error function can obtain hidden nodes and data centers of RBF network accurately. Dynamic learning of data centers, spread constants and weight values based on gradient method can contribute to improving the precision. Experimental results show that the proposed method has high precision and can be well applied in machine vision.

Keywords: RBF neural network · Binocular vision
Camera calibration · K-means · Gradient method

1 Introduction

Machine vision is becoming more and more widely applied in many fields, such as robot vision navigation and positioning, 3-D reconstruction, etc., while camera calibration is the most important part of machine vision. Traditional camera calibration methods are generally based on precision machined 2-D or 3-D targets. Direct linear transformation (DLT) calibration method [1] proposes the relationship between the two-dimensional plane image and the actual object coordinates in the three-dimensional space. This method is based on the ideal pinhole model. The parameters involved are relatively few and the calculation process is simple and fast, but the nonlinear distortion caused by the imaging process is not taken into account, so the precision is poor. Tsai proposed a two-step method based on radial constraint [2], combining the traditional linear method and nonlinear optimization method, but the highprecision three-dimensional calibration block is difficult to process and maintain in practice. In the literature [3], Kruppa equation is established based on the epipolar geometry and curvilinear geometry theory which is used in camera self-calibration technology and the calibration

speed is fast. It is suitable for onsite calibration, but the precision is low and the stability is poor. Zhang proposed a flexible calibration method [4] according to the plane target images shot by camera from different locations to solve the camera parameter model. This method is widely used, but it also needs to establish a complex mathematical model.

Considering the problems that mathematical model constructed is complex, operation is complicated and precision is poor in traditional camera calibration methods, this paper proposes a camera calibration method based on improved RBF neural network, which has high precision and strong real-time, and the method can be applied to more complex environments.

2 Camera Imaging Model

The projection of three-dimension spatial point D on camera imaging plane is usually obtained by ideal pinhole model, which can be written as

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_0 & 0 \\ 0 & f_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = A \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where $X = (X_w^T, 1)^T$ is the homogeneous coordinate of the three-dimensional spatial point D in the world coordinate system. $(u, v, 1)$ is the homogeneous coordinate of the projection point on the image plane. R and T are the rotation matrix and the translation vector that the world coordinate system transforms into the camera coordinate system. (u_0, v_0) is the optical center of the camera. f_u and f_v are the scale factors in u and v axis directions. With the impact of the camera lens optical distortion, as shown in Fig. 1, the actual imaging point is not the intersection of the image plane and the joint line between three-dimensional point and the optical center, but with a certain offset $(\Delta u, \Delta v)$. The distortion of the camera mainly includes radial distortion and tangential distortion.

Binocular stereo vision imaging principle is shown in Fig. 2. The binocular camera imaging equation is expressed as

$$\left\{ s_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = A_l [R_l \ T_l] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, s_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = A_r [R_r \ T_r] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \right. \tag{2}$$

where s is the scale factor. Subscript l and r represent left camera and right camera, respectively. For binocular camera calibration, in addition to calibrating the internal parameters of the cameras, the rotation matrix and the translation matrix between two cameras are also needed.

Above the mentioned, the traditional calibration method based on the camera imaging mathematical model involves a large number of unknown parameters, which takes many factors into account and constructs complexly. While the calibration method based on neural network calibration is able to overcome these problems [5,6].

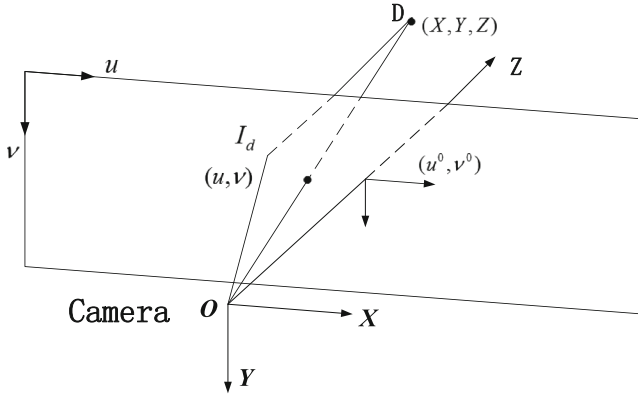


Fig. 1. Camera imaging model of monocular vision

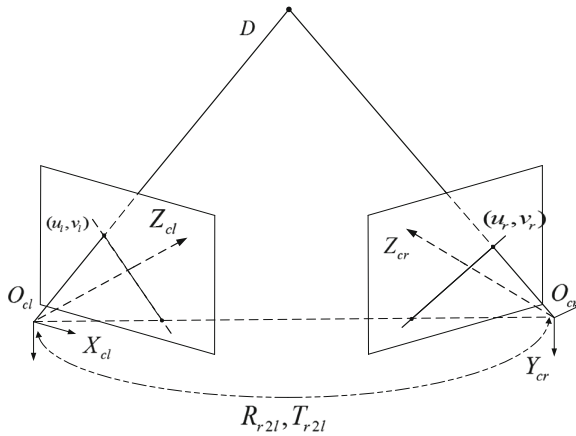


Fig. 2. Camera imaging model of binocular vision

3 RBF Neural Network Based on K-means and Gradient Method

3.1 The Structure of RBF Neural Network

RBF neural network is generally three-tier structure. As shown in Fig. 3, the RBF neural network structure is $n - h - m$, that means, the network has n inputs, h hidden nodes and m outputs. $x = (x_1, x_2, \dots, x_n)^T \in R^n$ is the network input vector. $\omega \in R^{h \times m}$ is the output weight matrix. $y = [y_1, y_2, \dots, y_m]^T$ is the network out-put vector. $\phi(\bullet)$ is the activation function of the i -th hidden node. In this paper, the Euclidean distance function is used as the basis function represented as $\|\bullet\|$ and the Gauss function is used as the activation function, which is as follow

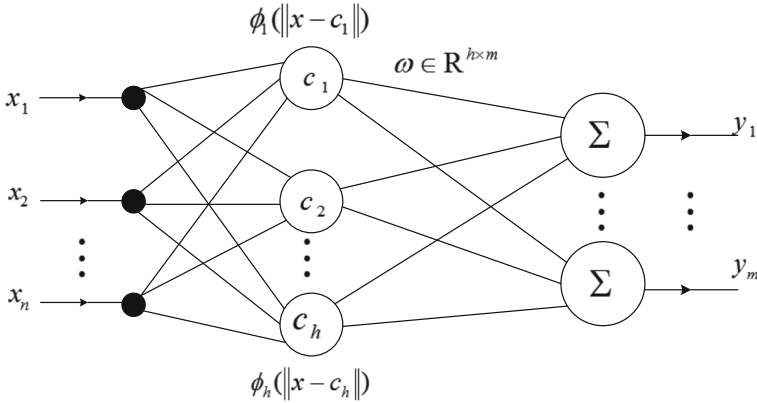


Fig. 3. The structure of RBF neural network

$$\phi(u) = e^{-\frac{u^2}{\delta^2}} \tag{3}$$

where the spread constant of RBF $\delta > 0$. The smaller δ is, the smaller the width of the RBF is and more selective it becomes. In Fig. 3, the j -th output of the RBF neural network can be given as

$$y_j = \sum_{i=1}^h \omega_{ij} \phi_i(\|x - c_i\|), \quad 1 \leq j \leq m \tag{4}$$

3.2 The Learning Algorithm of RBF Neural Network

Clustering method is the most classic RBF neural network learning algorithm [7,8]. The idea is to use the unsupervised learning method to obtain the data centers of the hidden nodes in RBF neural network, and to determine the spread constants according to the distance between the data centers. Then we can use the supervised learning method to train the output weight values of the hidden nodes.

In this paper, we propose a novel method of clustering number selection based on k-means clustering error law to obtain the number of hidden nodes and data centers. In the training, we construct the objective function based on multi-output error and calculate the gradient to update the data centers, the spread constants and the weight values dynamically.

The basis of the k-means algorithm is error sum of squares [9]. If N_i represents the sample size of the i -th class Γ_i , m_i is the mean value of the samples in Γ_i , as follow $m_i = \frac{1}{N_i} \sum_{y \in \Gamma_i} y$. The error sum of squares for each class Γ_i is calculated and the error sum of squares for all classes are added, as follow $J_e = \sum_{i=1}^c \sum_{y \in \Gamma_i} \|y - m_i\|^2$. J_e is called the clustering criteria based on error sum of

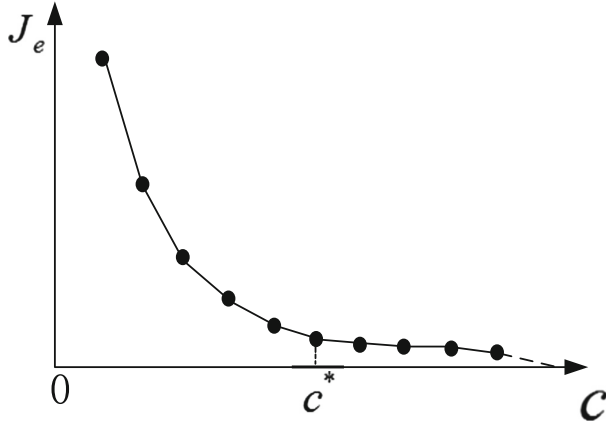


Fig. 4. The law of clustering error

squares. The basic premise of the k-means algorithm is that the clustering number is given in advance, which is not satisfied in unsupervised learning problems such as camera calibration.

For the initial partitioning after a given clustering number, we can get good results by dividing according to the natural order of the samples (the order of the feature points on the calibration board).

Obviously, the clustering error function decreases monotonically with the clustering number increasing. When the clustering number is equal to the size of all samples, we can obtain $J_e(c) = 0$. Thus, each sample itself becomes a class. If there are c^* much clustered classes in the data, will decrease rap-idly when increases from 1 to c^* . But the rate of decreasing will be significantly slower when outnumbers c^* , because the original intensive samples will be separated [10]. Figure 4 shows the curve that represents the change of $J_e(c)$ following c . Based on this law, in order to obtain a better clustering number, we can start from 1 until we select a preferable number. Formula (5) is given below as a criterion of judgment

$$|J_e(c^*) - J_e(c^* - 1)| \leq \alpha \bullet J_e(c^* - 1), \quad c^* = 2, \dots, c \tag{5}$$

where α is the convergence factor, the range is $\alpha = 0.2 - 0.3$. When the formula (5) is satisfied, the constructed neural network can achieve the fitting accuracy, and there will be not too many nodes in the hidden layer that lead to overfitting and affect generalization ability.

By using k-means clustering algorithm which selects the number of clusters automatically, the hidden nodes and data centers of RBF neural networks can be determined. The spread constants of RBF neural network can be determined by $\delta = \frac{d}{\sqrt{2h}}$, where d is the maximum distance of all classes and h is the clustering number of RBF neural network.

When hidden nodes, data centers and the spread constants are initially determined, the gradient method is adopted to train the output weight values of hidden nodes [11]. The gradient method used in this paper not only updates the output weight values, but also updates the data centers and the spread constants dynamically on the basis of the k-means method to ensure the global optimization of the parameters.

We construct an objective function based on the errors of multiple output nodes. The objective function is defined as

$$\varepsilon = \frac{1}{2} \sum_{n=1}^N \|e_n\|_2^2 \tag{6}$$

where N is the number of training samples of the learning process. e_n is the error signal corresponding to the J output nodes and $J \times 1$ is the vector of n dimension.

The j -th component of the error vector (that is the error of the j -th output node) is established as

$${}_j e_n = {}_j d_n - j \left[\sum_{i=1}^I \omega_{ij} \phi_i(\|X_n - c_i\|) \right] \tag{7}$$

where I is the number of hidden nodes. d represents the expected output. The left subscript represents the component of the vector.

We aim at finding the parameters ω_{ij} , c_i and δ_i that minimize ε . For the Gauss activation function used in this paper, the gradient of the objective function to each parameter is given as

$$\begin{cases} \frac{\partial \varepsilon(m)}{\partial \omega_{ij}(m)} = - \sum_{n=1}^N {}_j e_n(m) \phi_i(\|X_n - c_i(m)\|) \\ \frac{\partial \varepsilon(m)}{\partial c_i(m)} = - \sum_{j=1}^J \sum_{n=1}^N \frac{2\omega_{ij}(m)}{\delta_i^3(m)} {}_j e_n(m) \phi_i(\|X_n - c_i(m)\|) \|X_n - c_i(m)\| \\ \frac{\partial \varepsilon(m)}{\partial \delta_i(m)} = - \sum_{j=1}^J \sum_{n=1}^N \frac{2\omega_{ij}(m)}{\delta_i^3(m)} {}_j e_n(m) \phi_i(\|X_n - c_i(m)\|) \|X_n - c_i(m)\|^2 \end{cases} \tag{8}$$

The update law is given as

$$\begin{cases} \omega_{ij}(m+1) = \omega_{ij}(m) + \eta_1 \frac{\partial \varepsilon(m)}{\partial \omega_{ij}(m)} \\ c_i(m+1) = c_i(m) + \eta_2 \frac{\partial \varepsilon(m)}{\partial c_i(m)} \\ \delta_i(m+1) = \delta_i(m) + \eta_3 \frac{\partial \varepsilon(m)}{\partial \delta_i(m)} \end{cases} \tag{9}$$

where m represents the number of iterations. η_1 , η_2 and η_3 are different convergence factors.

In the camera calibration experiment based on improved RBF neural network, we let the image coordinates of the feature points, (u_l, v_l) and

(u_r, v_r) , as inputs and the spatial coordinates in the world coordinate system, (X_W, Y_W, Z_W) , as outputs. A RBF neural network with 4 input nodes and 3 output nodes is constructed. The hidden nodes and data centers are determined automatically and the data centers, spread constants and weight values are updated dynamic until the parameters satisfy the precision requirement.

4 Experiment and Analysis

The calibration experiment utilizes a binocular vision system as shown in Fig. 5. The camera in measuring experiment is from EoSens® 3CL series of Mikrotron company, whose model number is MC3010 and resolution is 1280 1024 pixel. The pixel size is 0.008 mm/pixel and the lens model number is AF Zoom-Nikkor 24-85 mm/1:2.8-4D.



Fig. 5. Binocular vision system

A flat plate with round feature points is used as a calibration board and the points on the calibration board are 10×9 , which arrange in a certain order. We place the calibration board vertically on the linear motion platform and establish the X-axis, Y-axis of the world coordinate system on the plane of the calibration board, with Z-axis perpendicular to the plane of the calibration board. The calibration board moves along Z-axis 30 mm, 60 mm, 90 mm and 120 mm. We use two cameras to capture simultaneously 5 pairs of images at 0, 30 mm, 60 mm, 90 mm and 120 mm. When Z is 0, the images acquired by the left camera and right camera are shown in Fig. 6.

Zernike moments algorithm [12] is used to extract feature points, and 370 sets of data are obtained. These sets of data were divided into two groups, of

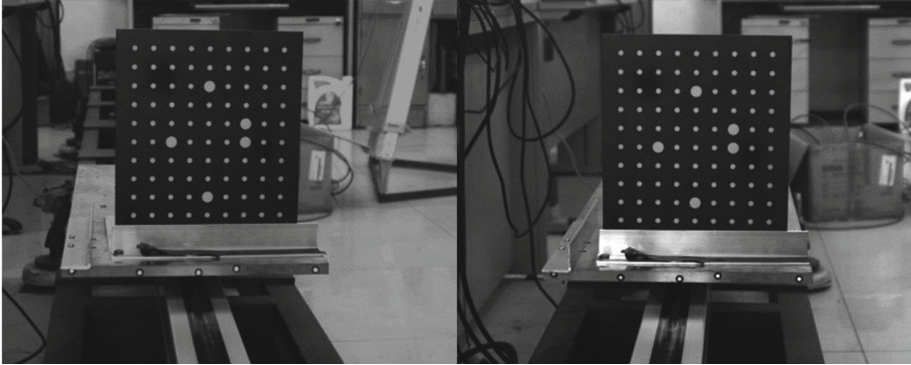


Fig. 6. The images acquired by left camera and right camera

which 280 sets of data were used as training data for the network, and the other 80 groups were used for testing.

RBF network with 4-10-3 structure is trained according to 370 sets of data. The formula for calculating the error of each spatial point is given as

$$e = \sqrt{(\tilde{X}_W - X_W)^2 + (\tilde{Y}_W - Y_W)^2 + (\tilde{Z}_W - Z_W)^2} \tag{10}$$

where $(\tilde{X}_W, \tilde{Y}_W, \tilde{Z}_W)$ expresses the reconstructed coordinates of the spatial points according to the trained RBF neural network. The results of the reconstruction of some points are shown in Table 1.

Table 1. The test result of some points (Unit: mm)

Spatial points			Reconstructed coordinates			Error
X_W	Y_W	Z_W	\tilde{X}_W	\tilde{Y}_W	\tilde{Z}_W	
125	0	0	124.9691	0.0112	0.1913	0.1940
25	50	30	25.0415	50.0821	30.1745	0.2151
200	50	30	200.0829	50.0475	30.0988	0.1375
125	100	60	124.9302	100.0949	60.1935	0.2265
50	150	60	49.9193	148.0450	60.1221	0.1531
200	175	90	200.0505	175.1134	90.1213	0.1735
75	175	90	75.0387	175.0554	90.0869	0.1101
100	50	120	100.0984	50.0955	120.1154	0.1792

The average of the errors of all 80 testing points is 0.1719 mm. It can be seen from Table 1 that the precision of reconstruction in direction X and Y direction

is higher than that in direction Z, because there is an error in the installation of the calibration board on motion platform and the precision of the motion platform is less than that of the calibration board.

The precision of our method is verified by measuring the length of the segment between the two feature points as shown in Fig. 7.

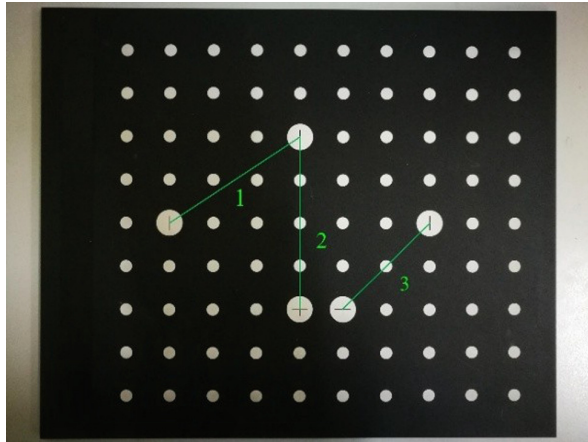


Fig. 7. Three segments that need to be measured

We place the calibration board in any position within the range of the field of view, and take the average of the actual length of the segment between five measurements. Zhang method [4] is used as a comparison. The results are shown in Table 2. It can be seen that the precision of our new method is better than that of Zhang method.

Table 2. Comparison of two methods (Unit: mm)

No.	Actual length	Measurement result		Error	
		Our method	Zhang	Our method	Zhang
1	90.1388	90.3712	90.5723	0.2324	0.4335
2	100	100.2918	100.3652	0.2918	0.3652
3	70.7107	70.8991	71.1634	0.1884	0.4527

5 Conclusion

The camera calibration method based on improved RBF neural network do not need to consider the impact of the lens distortion and environmental factors. Our method can reduce the error caused by the imperfect mathematical model

of the traditional calibration method and contribute to improving measurement precision. The reconstruction error of spatial points is 0.1719 mm in this paper. Measuring experiment shows that the precision of our new method is better than that of Zhang method. The data center selection method and dynamic learning of data centers, spread constants and weight values based on the gradient method further can improve the precision.

Acknowledgements. This work was supported by the National Natural Science Foundation of China [grant number 61473100].

References

1. Abdelaziz, Y.I.: Direct linear transformation from comparator coordinates in close-range photogrammetry. In: ASP Symposium on Close-Range Photogrammetry in Illinois (1971)
2. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. In: Radiometry. Jones and Bartlett Publishers Inc. (1992)
3. Faugeras, O.D., Luong, Q.-T., Maybank, S.J.: Camera self-calibration: theory and experiments. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 321–334. Springer, Heidelberg (1992). https://doi.org/10.1007/3-540-55426-2_37
4. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
5. Qi, Z., Wang, Z., Huang, Z.: Research on system calibration of structured-light measurement based on neural network. *Acta Photonica Sinica* **45**(5), 512002 (2016)
6. Tian, X.C., Zhong-Ke, L.I.: Binocular vision 3D measurement system calibrated by dual plane method. *Electron. Opt. Control* (2015)
7. Zhang, Z.Z., Qiao, J.F.: Design RBF neural network architecture based on online subtractive clustering. *Control Decis.* **27**(7), 997–1002 (2012)
8. Guan, S., Gao, J.W., Zhang, B., Liu, X., Leng, Z.W.: Traffic flow prediction based on k-means clustering algorithm and RBF neural network. *J. Qingdao Univ.* (2014)
9. Wu, S.: Survey on k-means algorithm. *N. Technol. Libr. Inf. Serv.* **29**(3), 433–439 (2011)
10. Ripley, B.D.: *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge (2009)
11. Wei, H., Li, Q., Song, W.: Gradient learning dynamics of radial function networks. *Control Theory Appl.* **24**(3), 355–356 (2017)
12. Cui, J., Tan, J.: Algorithm for edge subpixel location based on Zernike moment. *Opt. Tech.* **31**(5), 775–779 (2005)