

Received April 13, 2020, accepted May 2, 2020, date of publication May 6, 2020, date of current version June 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992652

# Research on the Calibration of Binocular Camera Based on BP Neural Network Optimized by Improved Genetic Simulated Annealing Algorithm

LONG CHEN<sup>1</sup>, FENGFENG ZHANG<sup>1,2</sup>, AND LINING SUN<sup>1,2</sup>

<sup>1</sup>School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215006, China

<sup>2</sup>Collaborative Innovation Center of Suzhou Nano Science and Technology, Soochow University, Suzhou 215123, China

Corresponding author: Fengfeng Zhang (zhangfengfeng@suda.edu.cn)

This work was supported in part by the National Key Research and Development Project under Grant 2018YFB1307700.

**ABSTRACT** The Back Propagation (BP) neural network has the problems of low accuracy and poor convergence in the process of binocular camera calibration. A method based on BP neural network optimized by improved genetic simulated annealing algorithm (IGSAA-BP) is proposed to solve these problems to complete the binocular camera calibration. The method of combining Gaussian scale space and Harris corner detection operator is used for corner detection. A matched algorithm of homonymous corner is proposed by combining point-to-point spatial mapping and grid motion statistics. The pixel values of the homonymous corner and three-dimensional coordinate values are taken as the input and output of BP neural network respectively. The crossover and mutation probability of genetic simulated annealing algorithm and the annealing criterion are improved, the IGSAA-BP neural network is used to calibrate the binocular camera. The average calibration accuracy of BP neural network and IGSAA-BP neural network is 0.71mm and 0.03mm, respectively. The average calibration accuracy of binocular camera is improved by 96%. The iteration speed is increased by 20 times and global optimization ability is improved. It can be seen that the IGSAA-BP neural network can improve the calibration accuracy of binocular camera and accelerate convergence speed.

**INDEX TERMS** Camera calibration, genetic simulated annealing algorithm, Gaussian scale space, corner detection, homonymous corner match, BP neural network.

## I. INTRODUCTION

The three-dimensional (3D) reconstruction technology based on binocular vision can restore the 3D contour information of objects in a simple and convenient way [1]–[3]. In recent years, due to the rapid development of this technology, robot navigation, visual detection, artificial intelligence and other industries have also changed with significant breakthroughs. Binocular vision is different from monocular vision in that it can obtain the depth information of the object according to different positions in the object space [4]–[6]. In the process of actual 3D measurement, it has the advantages of high efficiency and accuracy. The parallax principle is used to calculate the position deviation between the homonymous points in two images obtained from different positions so as to obtain the 3D information of the target object [7]–[9]. Binoc-

ular vision technology is the foundation of many industries, among which one of the key research fields is the calibration of binocular camera. In the process of practical application, the calibration accuracy and the required time are the primary research objectives [10], [11]. Therefore, it is necessary to improve the accuracy and reduce the time of the binocular camera calibration.

The calibration accuracy of the binocular camera directly affects the accuracy of the final imaging system. The traditional calibration methods of binocular camera mostly rely on the accurate imaging mathematical model, which takes little account of various nonlinear distortion factors in the imaging process, and the calibration process is relatively complicated [12]–[14]. The calibration of binocular camera determines the internal and external parameters of the imaging system for binocular stereo vision. Compared with monocular stereo camera, its calibration process is relatively tedious, requiring higher requirements for algorithm optimization and

The associate editor coordinating the review of this manuscript and approving it for publication was Liang-Bi Chen.

target change. In terms of the essence of binocular camera calibration, it is actually to obtain the mapping relationship between camera image coordinates and 3D real world coordinates. The binocular camera calibration does not necessarily need to solve the internal and external parameters of the imaging system for binocular stereo vision [15]–[18]. In recent years, many scholars have proposed binocular calibration methods based on neural network. The projection point coordinates of two cameras are taken as input signals to match the expected values of 3D coordinates of the network [19], [20]. Therefore, how to quickly obtain the accurate mapping relationship between pixel coordinates and 3D physical coordinates will become one of the hot topics in binocular camera calibration.

As a new technology in artificial intelligence field, artificial neural network has been proved to be able to approximate any continuous function with any accuracy [21], [22]. Based on this feature, the mapping relationship between pixel coordinates and the object's 3D real world coordinates can be directly established without determining the internal and external parameters of the binocular camera or knowing the relevant camera imaging model. All the nonlinear factors are included in the neural network, which is currently used as an implicit calibration [23]–[25]. In the study of the calibration of binocular camera based on neural network, the BP neural network is the most common method. Its feasibility in binocular camera calibration has been proved by An and Zhi [26], Yao *et al.* [27]. Yuan *et al.* proposed a simple and flexible camera calibration method based on BP neural network as early as 2009 [28]. The projection matrix of the camera is fitted by the weights between the network input layer and the hidden layer. Jin *et al.* studied the influence of different number of hidden layers of BP network on camera calibration accuracy, compared and analyzed the final experimental results for many times, and concluded that hidden layers have a direct influence on the calibration accuracy [29]. It can be seen from many experimental results that the binocular camera based on BP neural network has long iteration time, low accuracy and poor convergence. Holland *et al.* first proposed genetic algorithm, which originated from the imitation of Da Vinci's theory of species evolution [30]. To solve this problem, Yao *et al.* used the fast convergence rate of genetic algorithm to optimize BP neural network and obtained good convergence [31]. Genetic algorithm is easy to fall into the local minimum, thus reducing the ability of global optimization [32]. Liu *et al.* made a similar study and used Particle Swarm Optimization (PSO) algorithm to optimize BP neural network, but the final result was not ideal because PSO algorithm was prone to premature convergence [33], [34]. In most of the current studies, the main problem in the application of BP neural network to the calibration of binocular camera is that it cannot meet the requirements of short calibration time, high accuracy and good convergence.

In this paper, a method based on BP neural network optimized by improved genetic simulated annealing algorithm

(IGSAA-BP) is proposed for binocular camera calibration, which overcomes the disadvantages of BP neural network in the process of binocular camera calibration, such as long iteration time, low accuracy and poor convergence. The improved corner detection and the algorithm of homonymous corner match are proposed to solve the problem of low corner detection rate and high rate of homonymous corner mismatch in camera calibration process. The contributions of our work are pointed as follows:

1. A new corner detection method based on the combination of Gaussian scale space and Harris corner detection operator is proposed to improve the success rate of corner detection in checkerboard images.
2. An automatic matching algorithm based on the rough matching of point-to-point spatial mapping and the grid motion statistics is proposed to reduce the rate of the homonymous corner mismatch in the checkerboard images.
3. The improved genetic simulated annealing algorithm is to optimize BP neural network and overcome the disadvantages of BP neural network in binocular camera calibration.
4. The experiments are carried out to verify the proposed improved methods of corner detection, homonymous corner match and IGSAA-BP, so as to prove their effectiveness in the process of binocular camera calibration.

The paper is structured as follows: Section II describes the hardware platform and calibration process of binocular camera as well as the algorithm design and principle, homonymous corner match and so on in the calibration process. In section III, the improved algorithms proposed in this paper are verified by experiments, and the experimental results are simply counted and analyzed. Section IV discusses the binocular camera calibration method of IGSAA-BP neural network. Section V summarizes the specific improvements of the proposed methods in this paper on the calibration result of binocular camera based on BP neural network.

## II. MATERIALS AND METHODS

### A. A HARDWARE PLATFORM AND COMPONENTS OF BINOCULAR CAMERA CALIBRATION

The experimental hardware platform for binocular camera calibration is composed of the monitor, workstation, binocular camera and calibration target as shown in Figure 1. The calibration target is a cube calibration block with a side length of 9cm, and the six sides are evenly covered with a black and white checkerboard with a side length of 1cm. As shown in Figure 1, the X, Y and Z axis is respectively established by the three edges of the calibration block, and the intersection of the three axes is at the origin O, so as to construct the real 3D world coordinate system of the object. According to this world coordinate system, the true 3D spatial coordinates of each checkerboard corner on the cube can be calculated.

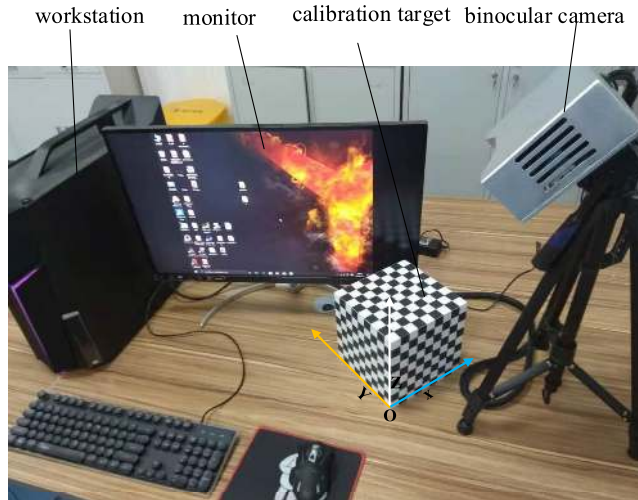


FIGURE 1. Calibration platform for binocular camera.

## B. ALGORITHM DESIGN AND PRINCIPLE OF CORNER DETECTION

Corner detection is an important part of binocular camera calibration. Corner detection is to obtain pixel coordinates of corners, so the precision of corner detection directly affects the final calibration accuracy of binocular camera. Harris corner detection algorithm is one of the common corner detection algorithms [35]. The main principle of common Harris corner detection algorithm for checkerboard corner detection is as follows: The corner of the checkerboard images can be simply considered as the intersection of two sides. Corner recognition is usually done in a small local area or window. If the small window is moved in all directions, the gray gradient of the area in the window changes greatly, then it is considered that corners are encountered in the window. If the gray gradient of the images in the window does not change as the window moves in all directions of the checkerboard images, there are no corners in the window. If the gray gradient of the images in the window changes greatly when the window moves in one direction and does not change in other directions, then the images in the window may be a line segment. The change in motion here is measured by the self-similarity of the panned window. Assuming that the gray value of pixels at the point  $(x,y)$  of the image is  $I(x,y)$ , when the distance is shifted  $(u,v)$  at the point  $(x,y)$ , the self-similarity  $(E(u,v))$  is:

$$E(u, v) = \sum_{(u,v) \in W(x,y)} w(x, y) [I(x+u, y+v) - I(x, y)]^2 \quad (1)$$

where  $W(x, y)$  is a window centered on the point  $(x, y)$ .  $I(x, y)$  is the pixel value at point  $(x, y)$ .  $w(x, y)$  is the weighted function.  $I(x+u, y+v)$  represents the pixel gray value when the distance is shifted  $(u, v)$  at the point  $(x, y)$ . When  $E(u, v)$  value changes in all directions in the window, it is more likely to be the corner. For small local movement  $(u, v)$ , equation (1)

can be simplified as:

$$E(u, v) \cong [uv]M \begin{bmatrix} u \\ v \end{bmatrix} \quad (2)$$

$M$  is the second-order differential matrix, and its calculation includes two processes: calculation of the differential, and calculation of the differential value in the accumulation window:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (3)$$

$I_x$  and  $I_y$  are the partial derivatives of  $I(x, y)$ . The matrix  $M$  is diagonalized, where  $\lambda_1$  and  $\lambda_2$  represent the rate of change of gray value in  $x$  and  $y$  directions:

$$M = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad (4)$$

The common Harris corner detection algorithm has gray invariance and rotation invariance, but has no scale invariance [36], [37]. In small scale, the corner of the checkerboard images may be considered as the edge of the checkerboard images after being enlarged, resulting in the failure of detection of some corners. Therefore, scale invariance is crucial for the local features of the images. To solve this problem, an improved Harris corner detection algorithm is adopted in this paper, which combines Gaussian scale space with Harris corner detection operator to carry out corner detection of the checkerboard images. Based on the common Harris corner detection algorithm, Gaussian scale space is introduced [38]. In the original feature point space, the feature points of other scale space are added. These additional feature points correspond to the images of different scale spaces, which increase the robustness of the target scale variation and make it have a certain degree of scale invariance. Let  $w(x,y)$  take Gaussian weight  $g(\delta_m)$ , and the new second moment  $M$  can be expressed as:

$$M = \delta_n^2 g(\delta_m) \otimes \begin{bmatrix} L_x^2(x, y, \delta_n) & L_x L_y(x, y, \delta_n) \\ L_x L_y(x, y, \delta_n) & L_y^2(x, y, \delta_n) \end{bmatrix} \quad (5)$$

$g(\delta_m)$  is the Gaussian convolution kernel of scale  $\delta_m$ .  $L_x(x, y, \delta_n)$  and  $L_y(x, y, \delta_n)$  represent the result of smoothing the images with Gaussian  $g(\delta_n)$  function. The new matrix  $M$  adds two Gaussian scale parameters compared to the original ( $\delta_m$ ) is called integral scale, and it is the variable that determines the current scale of Harris corner.  $\delta_n$  is the differential scale or local scale. It is the variable that determines the variation of the differential value near the corner. In fact, it can be considered as a Gaussian smoothing parameter).

It is not convenient to judge corner by judging the values of two variables  $\lambda_1$  and  $\lambda_2$ . The corner detection algorithm of checkerboard images firstly defines a set of integral and differential scales to  $\delta_m$  and  $\delta_n$  in advance. Then, the response function  $R$  of the corner is defined to roughly judge the corner of the checkerboard images from the position space:

$$R = \det M - k(\text{trace}M)^2 \quad (6)$$

$$\begin{cases} \det(M) = \lambda_1 \lambda_2 \\ \text{trace}(M) = \lambda_1 + \lambda_2 \end{cases} \quad (7)$$

Trace (M) is the direct trace of matrix M and (detM) is the determinant of matrix M. k is an empirical constant, which is generally set at 0.04-0.06. Search the candidate corner of the checkerboard images in the position space. At this time, when R>0, it is temporarily considered to be the corner of the checkerboard images. When R<0, it is temporarily considered as an edge of the checkerboard images. When R value is very small, it is temporarily regarded as a smooth transition region.

Since the candidate corner of the position space is not necessarily the candidate corner of the scale space, we also need to search the scale space to find the so-called characteristic scale value of the corner. On the basis of the results of the position space search, the search for the characteristic scale value is mainly divided into two steps:

Given a threshold value  $T_t$ , Laplace's response value  $F(x, \eta)$  is calculated for the candidate points of each checkerboard images that have been searched in the position space, and its absolute value is greater than the given threshold condition:

$$F(x, \eta_n) = \eta_n^2 |L_{xx}(x, \eta_n) + L_{yy}(x, \eta_n)| \geq T_t \quad (8)$$

$$\delta_n = \gamma \eta_n \quad (9)$$

$\gamma$  is the local Gaussian smoothing parameter (generally 0.7~1), and  $\eta$  is the initial scale parameter.

Compare the value of  $F(x, \eta)$  with the Laplace response value of the adjacent two scale Spaces to make it satisfy:

$$\begin{aligned} F(x, \eta_n) &> F(x, \eta_{n-1}) \\ F(x, \eta_n) &> F(x, \eta_{n+1}) \end{aligned} \quad (10)$$

The scale value satisfying condition 1) and 2) is the characteristic scale value of the corner. Combined with the search results of the position space, the Harris corner that satisfies the conditions of both the position space and the scale space is found, that is, the corner of checkerboard images that is finally determined. This method improves the accuracy of checkerboard corner detection and overcomes the low detection accuracy of common Harrias corner detection algorithm. The steps of corner detection algorithm for checkerboard images based on the combination of Gaussian scale space and Harris corner detection operator are shown in the Table 1:

**C. ALGORITHM DESIGN AND PRINCIPLE OF HOMONYMOUS CORNER MATCH**

The homonymous corner is the one-to-one corresponding corner in the left and right checkerboard images. Homonymous corner match is also an essential step for binocular camera calibration using BP neural network. It is necessary to match the corresponding homonymous corners on the left and right checkerboard images and obtain their corresponding pixel coordinates in the left and right checkerboard images respectively. The traditional method of homonymous corner match mainly uses the feature operators such as Scale Invariant Feature Transform (Sift) and Speed Up Robust Feature

**TABLE 1. Corner detection of checkerboard images based on improved Harrias algorithm.**

Steps	Algorithm steps of corner detection of checkerboard images
1	Search for the candidate point of corner of the checkerboard images from the position space
2	Compute the gray gradient values $I_x$ and $I_y$ for all pixels on the checkerboard images
3	Compute the matrix M for each pixel in the neighborhood of (x,y)
4	Compute $R = \det M - k(\text{trace}M)^2$
5	If $R>0$ , It is considered as the candidate point of corner of the checkerboard images.
6	From the scale space, we can further judge whether the candidate point is corner of the checkerboard images
7	For the candidate point of corner of each checkerboard images searched through the position space, the $F(x, \eta)$ is calculated
8	If $F(x, \eta_n) = \eta_n^2  L_{xx}(x, \eta_n) + L_{yy}(x, \eta_n)  \geq T_t$ and $F(x, \eta_n) > F(x, \eta_{n-1})$ and $F(x, \eta_n) > F(x, \eta_{n+1})$ , it is considered that the candidate point meet the requirements of scale space.
9	The candidate point satisfying both the position space and the scale space can be determined to be the corner of the checkerboard images

(Surf) in Open Source Computer Vision Library (Opencv) to detect the corner of the checkerboard images, then the algorithm of BruteForceMatcher or FlannBasedMatcher is used for homonymous corner match [39], [40]. Because the pattern of checkerboard is too similar, the rate of homonymous mismatch is high and the matched effect is poor when the traditional methods are used to match the homonymous corner.

In this paper, a method combining point-to-point spatial mapping algorithm and grid motion statistics is proposed to match the homonymous corner of the checkerboard images, so as to overcome the disadvantage of high rate of homonymous corner mismatch of the checkerboard images. The algorithm of corner detection proposed in this paper is used to detect the corner of left and right checkerboard images. On this basis, the algorithm of homonymous corner match is divided into two parts. In principle, it will mainly include:

Part I: The point-to-point spatial mapping algorithm is used to roughly match the homonymous corner of checkerboard images.

Part II: On the basis of the results of rough match in the Part I, the grid motion statistics algorithm proposed by Bian et al. is used to optimize the results of rough match to reduce the rate of homonymous corner mismatch and obtain high-quality results.

In the Part I, a correlation coefficient matrix  $M_c$  is assumed based on the needs of point-to-point spatial mapping algorithm, which is used to describe the correlation matrix of all corners in the left and right images (only comparison between different images, i.e. the correlation of corners of the same image need not be calculated). Assuming that the number of corner in the first image is  $m$  and the number of corner in the second image is  $n$ , then the size of the correlation coefficient matrix is  $m \times n$ :

$$M_c = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & \ddots & \ddots & x_{2n} \\ \vdots & \ddots & \ddots & \ddots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (11)$$

The matched window of  $3 \times 3$  is built by centering on the position  $p(p_x, p_y)$  of the pixels to be matched in the left and right checkerboard images. In the same way, the target function is established to measure the correlation of the matched window by constructing the matched window of neighborhood at the position of the target pixel  $p'(p_x + d, p_y)$ . A correlation coefficient  $\xi(p, d)$  is defined to describe the correlation between the corners of the two checkerboard images. The interval range is  $[-1, 1]$  (The greater the value of the correlation coefficient, the better its correlation will be): (12), as shown at the bottom of the next page.

$I_1(x, y)$  represents the pixel value at point  $(x, y)$  in the sub-graph of the left checkerboard image.  $\bar{I}_1(p_x, p_y)$  and  $\bar{I}_1(p_x + d, p_y)$  represent the average of the position of the pixel to be matched and the target pixel, respectively.  $d$  represents the distance between the position of the queried pixel in the right image and the position of the  $p_x$  in the horizontal direction.  $W_p$  represents a matching window with coordinates of matched pixels as the center (Generally, it is a  $3 \times 3$  matched window). If  $\xi(p, d) = -1$ , it means that the two matched windows are completely unrelated. On the contrary, if  $\xi(p, d) = 1$ , it means that the two matched windows are highly correlated. On this basis, several important parameters such as correlation coefficient threshold ( $t_s$ ), distance ratio threshold ( $r$ ) and corner distance threshold ( $q$ ) in the algorithm are set. The main body of the algorithm for homonymous corner match of the checkerboard images based on point-to-point spatial mapping is shown in the following Table 2:

After rough matching of point-to-point spatial mapping, the rough matched results are relatively poor, and there are some mismatched corners. At this time, the Part II of homonymous corner match: the grid motion statistical algorithm is used to optimize the rough matched results, and the wrong matching is quickly eliminated, and the high-quality results are obtained. According to the grid motion statistics algorithm proposed by Bian et al. [41], the results of mismatching after rough matching are eliminated. The main steps of the algorithm for optimizing the results of homonymous corner match based on the grid motion statistics are as follows:

- 1) Find the nearest corner in the right image corresponding to each corner in the left image after the rough matched

**TABLE 2. Rough matching algorithm for corner based on point-to-point spatial mapping.**

Steps	The body of the rough matched algorithm for corner based on point-to-point spatial mapping
1	Compute $\xi$ and relevance of matching windows of two checkerboard images
2	Find the position $(m, n)$ of the maximum value in $M_c$ of the left and right images. (Row $m$ , column $n$ )
3	Save the maximum value $I_{best}$ in the $M_c$ . Meanwhile, the value $(m, n)$ in $M_c$ is set as $-1$ .
4	Find the maximum value $I_{(ax_m)}$ in row $m$ of the current $M_c$ and the maximum value $I_{(ax_n)}$ in column $n$ .
5	Set the values of $t_s$ , $r$ and $q$ .
6	If two corners in the left and right images satisfy $1 - I_{best} < (1 - I_{(ax_m)}) \times r$ , $1 - I_{best} < (1 - I_{(ax_n)}) \times r$ and $Dis_{ab} < q$ , the rough matching between the two corners is successful.
7	Record this pair of corners, and set all elements in the row $m$ and column $n$ to $-1$ in the $M_c$ .
8	Repeat the process until the maximum value in the $M_c$ does not exceed $t_s$ .

step of point-to-point spatial mapping algorithm in the Part I.

- 2) Divide the left and right images into multiple grids. After the grid is divided, each grid is regarded as a small neighborhood.
- 3) If the two corners after rough matching can be matched correctly, a small area near the two corners can be regarded as corresponding to the same 3D position. All the matched corners in each grid need to be counted only once, instead of counting each corner separately.
- 4) Set the threshold value  $z$  for the number of corner in the divided grid. According to the smoothness of the motion, the correctly matched pair of corners near the correctly matched corner should be larger than the correctly matched pair near the incorrectly matched corner.
- 5) The number of correctly matched corners near corner  $C$  after rough matching is calculated as  $N_u \cdot N_u$  and threshold  $z$  are used to determine whether the corner is correctly matched.
- 6) If  $N_u > z$ , it is considered to be a correct match; otherwise, it is considered to be a wrong match, and the wrong match is deleted and rematched.

#### D. CALIBRATION OF BINOCULAR CAMERA BASED ON NEURAL NETWORK

##### 1) PRINCIPLE AND STRUCTURE DESIGN OF BINOCULAR CAMERA CALIBRATION BASED ON TRADITIONAL BP NEURAL NETWORK

BP neural network adopts the training method of supervised feedforward neural network. The main principle of using BP neural network for binocular camera calibration is as follows:

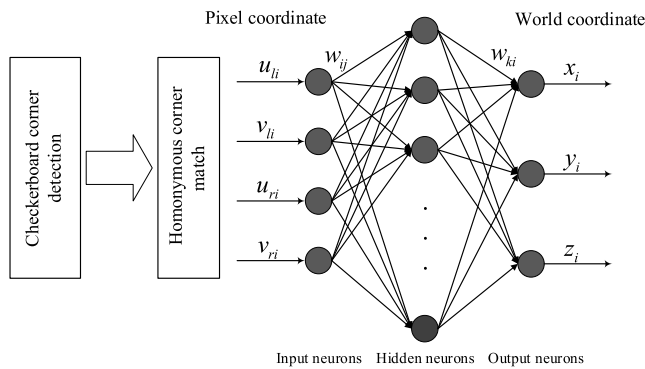


FIGURE 2. Design of binocular camera calibration structure based on BP neural network.

Based on the pixel coordinates of the corner of the checkerboard images of the binocular camera, the 3D coordinates of the network output of the corresponding corner of the checkerboard on the calibration block are calculated in the forward. The weight of the network is adjusted and optimized by the difference between the output of the network and the actual output of the three-dimensional coordinates.

Suppose that the input and output dimensions of BP neural network are  $m_o$  and  $n_o$  respectively. The number of hidden neurons and training samples are  $N_q$  and  $N$ , respectively. The weight between the input layer and the hidden layer of the neural network is  $W_{ij}$ , and the weight between the output layer and the hidden layer is  $W_{ki}$ . In this study, BP neural network is used to calibrate binocular camera. The specific calibration design of binocular camera based on BP neural network is shown in Figure 2. The BP neural network used in this study has three layers, including the input layer, the hidden layer and the output layer. The input layer contains four neurons, which are the coordinates of the horizontal and vertical pixels of the corner in the left and right checkerboard images. The output layer contains three neurons, representing the true 3D world coordinates of the corner corresponding to the 2D pixel coordinates of the input. The number of hidden layer neurons in this study is 9 (In general, if the number of neurons in the input layer is  $n$ , the number of neurons in the hidden layer is  $2n+1$ ). Set the maximum number of iterations ( $M_N = 1000$ ), the number of training samples ( $N_x = 1000$ ) and the accuracy of the given training target ( $\epsilon = 0.0001$ ).

2) PRINCIPLE OF BINOCULAR CAMERA CALIBRATION BASED ON BP NEURAL NETWORK OPTIMIZED BY TRADITIONAL GENETIC SIMULATED ANNEALING ALGORITHM  
The binocular camera calibration based on BP neural network can establish the mapping relationship between the coordi-

nates of 2D pixels and the coordinates of 3D real objects without complicated imaging model. However, BP neural network tends to fall into the minimum value and cannot reach the global optimal value. As the BP neural network randomly selects the threshold and weight, the convergence speed is slow, which leads to the low accuracy and slow speed of binocular camera calibration. To solve the above problems, BP neural network is optimized by traditional genetic simulated annealing algorithm (TGSAA-BP) [42], [43]. When TGSAA is used to optimize BP neural network, genetic algorithm is mainly used as the core frame. Annealing operation is introduced in the process of genetic variation operation. TGSAA can find the optimal threshold and weight of BP network in the calibration process of binocular camera through the survival of the fittest through population and genetic operators. Firstly, genetic algorithm is used to generate the initial population. Then selection, crossover, mutation and other operations are carried out to generate new individuals. Finally, each new individual is annealed to obtain a new population. The whole process is cycled until the pre-set algorithm end condition is reached. The optimal threshold and weight are assigned to BP neural network as the initial threshold and weight of BP network.

3) PRINCIPLE AND PROCESS OF BINOCULAR CANERA CALIBRATION BASED ON BP NEURAL NETWORK OPTIMIZED BY IMPROVED GENETIC SIMULATED ANNEALING ALGORITHM

In the process of binocular camera calibration, genetic algorithm improves the optimization ability of initial threshold and weight of BP neural network to some extent. At the same time, the risk of the calibration results falling into local extremum is increased. The simulated annealing algorithm is not easy to fall into the local optimal solution, but its convergence speed is relatively slow. In this study, the genetic algorithm is combined with simulated annealing algorithm, which improves the performance of BP network in the process of camera calibration, but the effect is not ideal. In this paper, the disadvantages of genetic algorithm and simulated annealing algorithm in the process of camera calibration are analyzed, and some improvements are proposed. In this way, the convergence and global optimization of the genetic simulated annealing algorithm can be improved. Thus, it further provides the optimal weight and threshold for BP neural network, improves the performance of BP neural network, and finally improves the accuracy and speed of camera calibration. In this study, aiming at the disadvantages of the traditional genetic simulated annealing algorithm in optimizing the BP neural network for binocular

$$\xi(p, d) = \frac{\sum_{x,y \in W_p} (I_1(x, y) - \bar{I}_1(p_x, p_y)) \cdot (I_2(x + d, y) - \bar{I}_2(p_x + d, p_y))}{\sqrt{\sum_{x,y \in W_p} (I_1(x, y) - \bar{I}_1(p_x, p_y))^2 \cdot \sum_{x,y \in W_p} (I_2(x + d, y) - \bar{I}_2(p_x + d, p_y))^2}} \tag{12}$$

camera calibration, the following improvement measures are taken from the following three aspects: fitness, crossover and mutation probability and annealing algorithm.

#### a: FITNESS STRETCH OF GENETIC ALGORITHM

During the binocular camera calibration test, the inverse of the sum of the squares of the errors between the actual 3D coordinates and the predicted 3D coordinates finally output by the BP neural network is used as the fitness value. In the later stage of genetic algorithm, the individual difference is small and is easily limited to local minimum. In this study, the fitness stretching method is used to improve the competitiveness between individuals to overcome this problem. At this point, the fitness  $F$  of the  $i$ th individual is shown in (13):

$$F = \frac{\exp(\frac{f_i}{T})}{\sum_{i=1}^m \exp(\frac{f_i}{T})} \quad (13)$$

$f_i$  represents the fitness of the  $i$ th individual before improvement.  $T$  represents the current temperature of the simulated annealing algorithm.

#### b: ADAPTIVE IMPROVEMENT OF CROSSOVER AND MUTATION PROBABILITY

The crossover probability of genes determines the diversity of the population studied, while the mutation probability determines whether the genetic algorithm can find the global optimum [44]. In the traditional genetic algorithm, the crossover and mutation probability of the gene in the chromosome is a constant value set in advance, which can only be manually adjusted. This process is very complicated, it is difficult to reconcile their best value. At present, many researches have proposed the completely adaptive genetic algorithm, which has the advantages of adaptive adjustment to the mutation and crossover probability of genes and does not depend on human operation [45]. However, this algorithm increases the risk of falling into local extremum and cannot find the global optimal solution accurately. In this study, the relationship between maximum fitness  $f_m$  and average fitness  $f_a$  of the population is defined as shown in (14). The traditional adaptive crossover probability  $P_c$  and mutation probability  $P_m$  in (15) and (16) are improved by using different degrees of population fitness concentration and dispersion.

$$\theta = \arcsin(\frac{f_a}{f_m}) \quad (14)$$

$$P_c = \begin{cases} 2x_1 - \theta < \frac{\pi}{6} \\ x_1 - 2x_1 - \theta \geq \frac{\pi}{6} \end{cases} \quad (15)$$

$$P_m = \begin{cases} 2x_2 - \theta \geq \frac{\pi}{6} \\ x_2 - 2x_2 - \theta < \frac{\pi}{6} \end{cases} \quad (16)$$

where  $x_1$  and  $x_2$  are the adaptive adjustment parameters of the population. Compared with the linear adaptive change

of the traditional genetic algorithm, the crossover and mutation probability are improved in this study so that it can carry out nonlinear adaptive change. It can be seen from equations (15) and (16) that, when  $\theta < \frac{\pi}{6}$ , as  $\theta$  decreases, the diversity of the camera data population increases. At this time, according to the adaptive increase of the crossover probability value, the crossover between individuals can be completed more fully, and high-quality individuals can be generated continuously. The adaptive value of mutation probability is reduced, the probability of good individuals being destroyed is reduced, and the convergence speed of the whole algorithm is accelerated. On the contrary, when  $\theta \geq \frac{\pi}{6}$ , with the increase of  $\theta$ , the crossover between genes is reduced and the convergence of the algorithm is accelerated. The adaptive probability of variation is increased to reduce the risk of the algorithm falling into local minima and improve the ability of global optimization. This improvement directly reduces the time of binocular camera calibration and improves the accuracy of binocular camera calibration.

#### c: IMPROVEMENT OF SIMULATED ANNEALING ALGORITHM CRITERION

In the process of binocular camera calibration, the convergence speed of the traditional simulated annealing algorithm is slow, and the diversity of camera calibration data population cannot be guaranteed under the traditional Metropolis criterion [46]. In view of these problems, this study is based on the evolution degree of the corresponding individuals in the new and old populations, and the Metropolis criterion is improved. The individuals in the new population are modified according to different conditions to increase the diversity of individuals and speed up the algorithm convergence to reduce the camera calibration time and improve the calibration accuracy. The simulated annealing algorithm is improved from the following two aspects:

(i) Definition of mutation probability ( $P_t$ ) of old and new individuals

$$P_t = \exp(\frac{f_{new}(i) - f_{old}(i)}{KT}) \quad (17)$$

where  $K$  is the fixed parameter,  $f_{new}(i)$  is the fitness of the  $i$ th individual of the new population, and  $f_{old}(i)$  is the fitness of the  $i$ th individual of the old population.  $T$  is the current temperature.

(ii) Establish rules for the replacement of old and new individuals

The fitness values of the corresponding individuals in the new and old populations are compared. When the fitness values of the new individuals are greater than the fitness values of the old individuals, the new individuals are stored as the next generation of the new population. On the contrary, all the genes in the new individual are swapped with random probability to produce the new individual, and the new individual produced at this time is stored as the next generation of the new population with  $P_t$ .

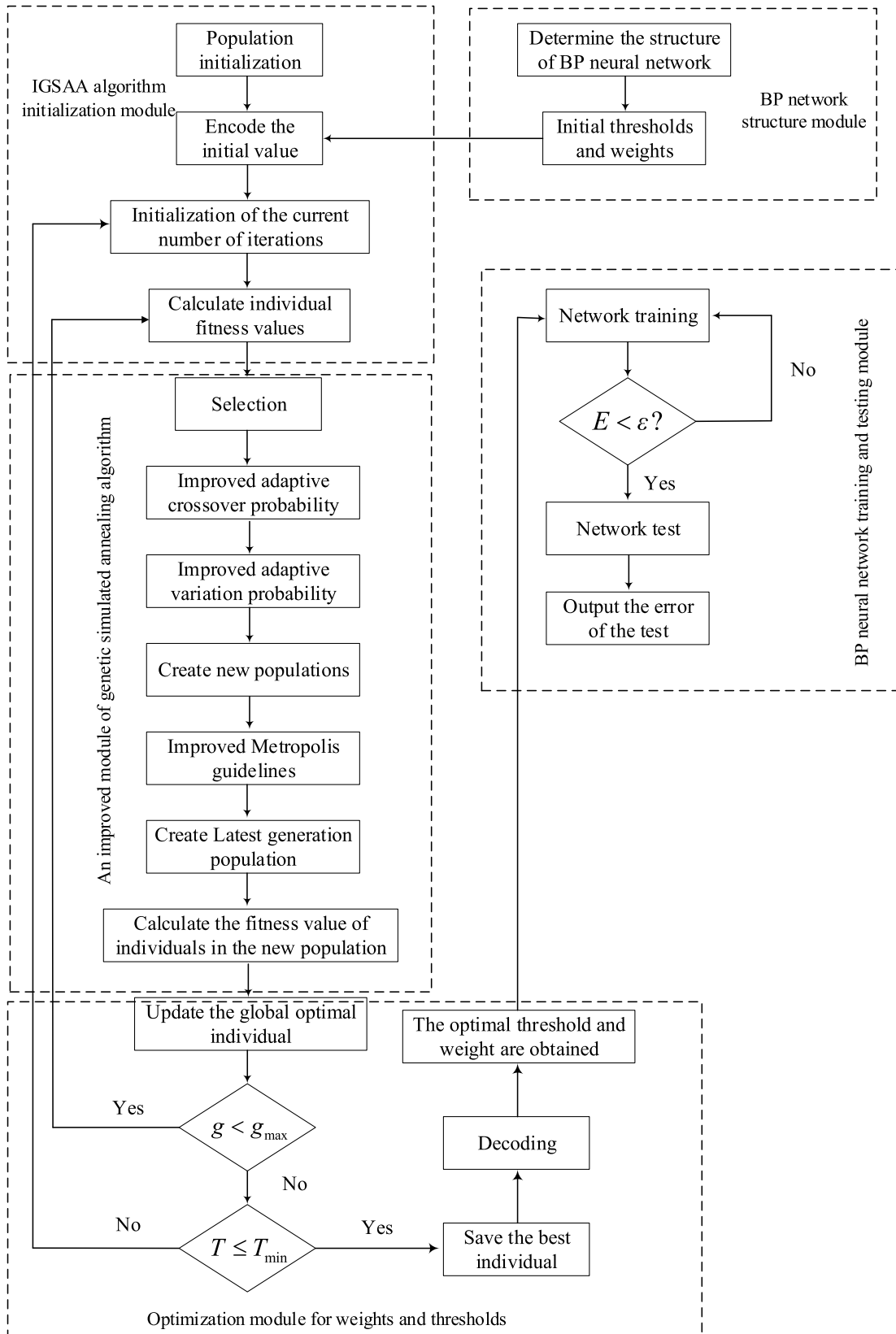


FIGURE 3. The calibration process of binocular camera based on IGSA-BP neural network.





FIGURE 4. Checkerboard images for binocular camera calibration.

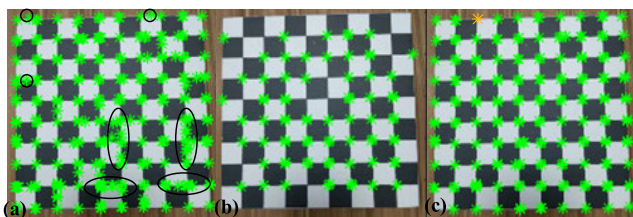


FIGURE 5. Corner detection based on checkerboard images: (a) Results of corner detection by using Surf corner detection algorithm; (b) Results of corner detection by the common Harris corner algorithm; (c) Results of corner detection based on improved Harris corner detection algorithm.

#### d: THE CALIBRATION PROCESS OF BINOCULAR CAMERA BASED ON BP NEURAL NETWORK OPTIMIZED BY IMPROVED GENETIC SIMULATED ANNEALING ALGORITHM

The calibration of binocular camera based on BP neural network optimized by IGSA can be divided into five parts: Parameters initialization module of genetic simulated annealing algorithm, structure module of BP neural network, improvement module of genetic simulated annealing algorithm, weight and threshold optimization module of BP neural network based on IGSA, training and testing module of BP neural network based on camera images. The flow chart of the BP neural network model optimized based on the improved genetic simulated annealing algorithm (IGSAA-BP) is shown in Figure 3. It is specified that the initial temperature of the simulated annealing algorithm is  $T_0 = 100^\circ$ , the current temperature is  $T$ , and the end temperature is  $T_{\min}$ .

### III. EXPERIMENTS AND RESULTS

The hardware platform of binocular camera calibration shown in Figure 1 is used to take images of calibration blocks. The left and right camera images are saved correspondingly, and the checkerboard images obtained through the binocular camera are shown in Figure 4. These images prepare for subsequent checkerboard corner detection and match.

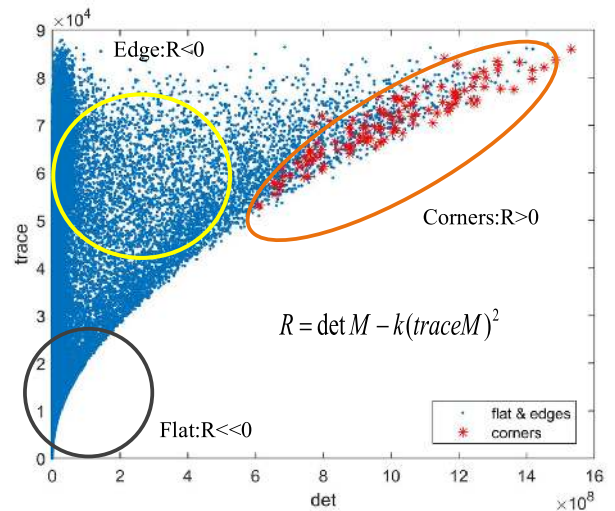


FIGURE 6. The distribution diagram of  $\det(M)$  and  $\text{trace}(M)$ .

#### A. EXPERIMENTS OF CORNER DETECTION BASED ON CHECKERBOARD IMAGES

At present, there are many methods for corner detection, but most of them are not satisfactory when applied to corner detection of checkerboard images. As shown in Figure 5 (a), traditional Surf feature point detection algorithm is used to detect the corner of the checkerboard images. It can be seen from the effect of corner detection circled by black circle in Figure 5 (a) that part of the edges of the checkerboard images are detected as corners, and there are problems of repeated detection and low detection accuracy at the real corners of the checkerboard images. As shown in Table 3, the success rate of corner detection of the checkerboard images by this method is about 60%, which cannot meet the requirements of the homonymous corner match. The common Harris corner detection algorithm is also often used for corner detection of calibration blocks such as checkerboard, but as shown in Figure 5 (b), its actual detection effect is not satisfactory. As can be seen from Table 3 and Figure 5 (b), many corners of the checkerboard images cannot be detected. The algorithm can only detect some corners of the checkerboard images, and the success rate of corner detection is only about 50%.

After the introduction of Gaussian scale space to improve common Harris algorithm, the relationship between  $\det(M)$  and  $\text{trace}(M)$  as shown in Figure 6. It can be seen more clearly that the number of corners and edges can be roughly judged according to the response function  $R$  of corners, and the function of response function of corners can be understood more clearly. The corner effect of the checkerboard images detected by the improved Harris corner detection algorithm is shown in Figure 5(c). It can be seen from Figure 5(c) that compared with Surf corner detection and common Harris corner detection, the accuracy of corner detection has been greatly improved. From the data of corner detection shown in Table 3, it can be seen that the accuracy of corner detection using the improved Harris checkerboard corner detection

**TABLE 3. The results of corner detection of three different algorithms.**

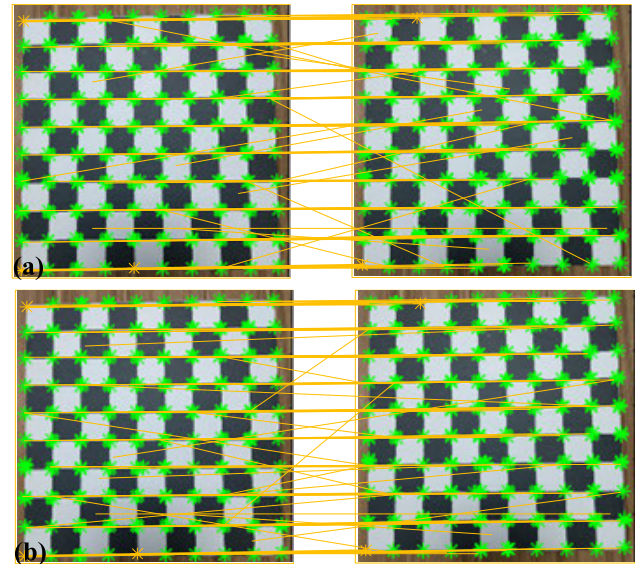
Name of algorithm	Number of corners in the checkerboard images	Number of corners detected	Success rate of corner detection
Surf	100	62	62%
Common Harrias	100	54	54%
Improved Harrias	100	100	100%

algorithm can reach 100%, which greatly reduces the error rate of corner detection.

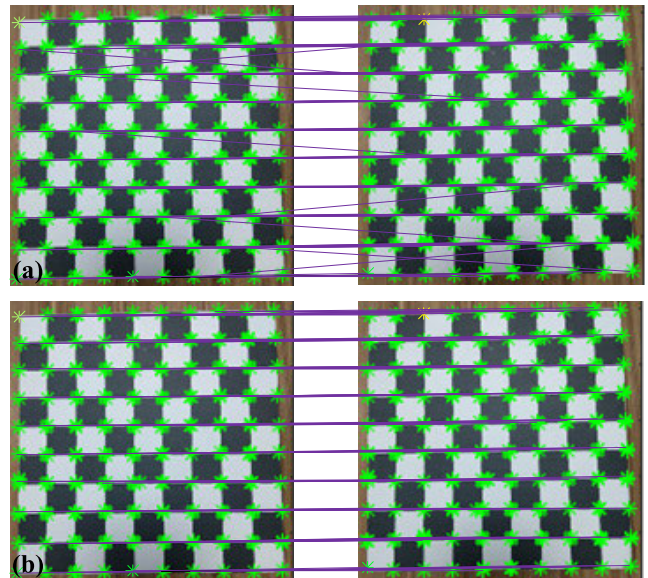
### B. EXPERIMENTS OF HOMONYMOUS CORNER MATCH BASED ON CHECKERBOARD IMAGES

As a calibration object for binocular camera, the checkerboard is helpful to improve the calibration accuracy. The pattern of checkerboard images has a high similarity, so as a camera calibration object, it also has a high mismatching rate of homonymous corner. In this study, two different traditional methods of homonymous corner match are adopted to realize homonymous corner match of left and right checkerboard images. The specific data of homonymous corner match is shown in Table 4. The first method mainly uses the traditional BruteForceMatcher algorithm in Opencv to match the homonymous corner in the left and right checkerboard images, and the final matched effect is shown in Figure 7 (a). It can be seen from Figure 7(a) that the matched algorithm of BruteForceMatcher can be used to complete homonymous corner match of the checkerboard images, but it still produces a large number of mismatches of the homonymous corner. According to the data in Table 4, the success rate of the homonymous corner match is only about 60%, and the process of the homonymous corner match takes about 3.6s. The second method uses the FlannBasedMatcher matched algorithm in Opencv to realize the homonymous corner match, and the final matched effect is shown in Figure 7 (b). It can still be seen from the data in Table 4 that the success rate of homonymous corner match by using FlannBasedMatcher matched algorithm is about 65%, which is 5% higher than the first method. The whole matched process takes 3.8s, which increased by about 6%. From the results of experiments, it can be seen that the two traditional matched methods used in this paper can not meet the actual requirements.

Aiming at the problem of high mismatched rate of homonymous corner, this paper proposes a matched algorithm combining point-to-point spatial mapping algorithm and grid motion statistics algorithm to solve the problem. In the first step, the spatial mapping algorithm of points and points is used to carry out rough matching to the homonymous corner. The final rough matched effect is shown in Figure 8 (a). It can be clearly seen from Figure 8 (a) that mismatches still occur in some homonymous corners. According to the statistical data in the Table 4, the success rate of the homonymous corner match is about 85%, and the whole matched process takes about 1.6s. In order to further improve the success rate of the



**FIGURE 7. Traditional method of the homonymous corner match: (a) Results of homonymous corner match based on BruteForceMatcher algorithm; (b) Results of homonymous corner match based on FlannBasedMatcher algorithm.**



**FIGURE 8. The results of the homonymous corner match algorithm are presented in this paper: (a) Rough matching of the homonymous corner based on point-to-point spatial mapping; (b) Improved matching of the homonymous corner points based on point-to-point spatial mapping and motion grid statistics algorithm.**

homonymous corner match, the second step is carried out: the grid motion statistics algorithm is adopted to eliminate the mismatched corners in the first step and make a second match. The final matched result is shown in Figure 8 (b). By combining Table 4 with Figure 8 (b), it can be seen that the success rate of homonymous corner match is up to 100%. At this time, a good effect is achieved. The whole process requires 1.8s. In terms of the time of homonymous corner match, the matched time of the method proposed in this paper is about 3.4s, which is about 5% less than that

**TABLE 4. Results of the homonymous corner MATCH of different methods.**

Name of algorithm	Pairs of homonymous corners in the checkerboard images	Successful pairs of homonymous corner	Matched success rate	Time (s)
BruteForceMatcher	100	61	61%	3.6
FlannBasedMatcher	100	67	67%	3.8
Point-to-point spatial mapping	100	84	84%	1.6
Our research	100	100	100	3.4

**TABLE 5. 2D pixel coordinates and 3D coordinates of the homonymous corner5(Part of the data).**

Group	Left camera image (pixel)		Right camera image (pixel)		Real 3D coordinates of the calibration object (cm)		
	u	v	u	v	x	y	z
1	525.02	41.12	505.23	26.54	3	3	4
2	536.18	194.65	533.37	181.05	2	0	3
3	555.26	65.74	538.57	49.65	3	2	4
4	584.43	88.63	575.29	72.38	3	1	4
5	530.41	268.32	527.54	255.31	2	0	2
6	457.55	541.27	418.34	530.69	0	3	0
7	687.22	109.54	678.06	92.45	4	0	4
8	657.34	395.36	649.78	376.45	4	0	0
9	602.36	262.10	598.33	246.04	3	0	2
10	654.29	84.58	640.87	66.12	4	1	4

of the traditional matched algorithm. However, based on the success rate of the matched algorithm, the success rate of the algorithm proposed in this study is about 40% higher than that of the traditional matched algorithm.

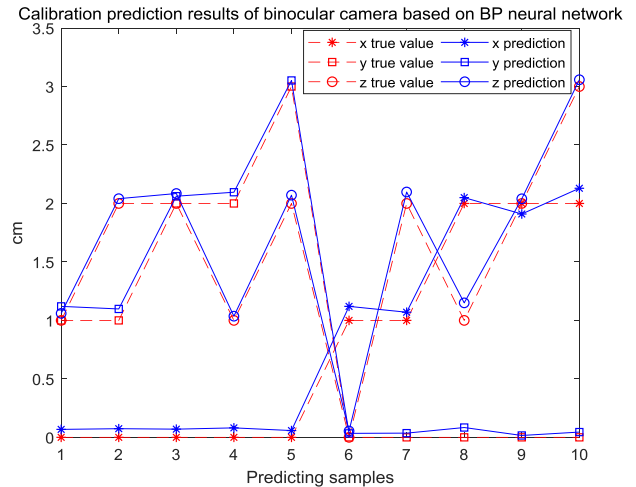
### C. CALIBRATION EXPERIMENTS OF BINOCULAR CAMERA BASED ON NEURAL NETWORK

Based on the left and right checkerboard images, the checkerboard corner detection is carried out and the corresponding homonymous corner match is carried out. The 2D pixel coordinate values corresponding to the homonymous corner is extracted and the actual 3D coordinate values corresponding to the corner is calculated. The pixel coordinates and 3D coordinates of the final extracted part of the homonymous corners are shown in Table 5.

#### 1) EXPERIMENTS OF CALIBRATION ACCURACY OF BINOCULAR CAMERA

##### a: EXPERIMENTS ON CALIBRATION ACCURACY OF BINOCULAR CAMERA BASED ON TRADITIONAL BP NEURAL NETWORK

The pixel coordinates and actual 3D coordinates of 1000 groups of homonymous corners in the checkerboard

**FIGURE 9. Test of calibration accuracy of binocular camera based on traditional BP neural network.**

images are selected as the training set of BP neural network. 1000 sets of experimental data are trained and 10 sets of data are imported into the test set to predict the 3D coordinate of binocular camera calibration using BP neural network (i.e., the prediction of calibration accuracy). The predicted and actual values of the 3D coordinates of the obtained corner are shown in Figure 9. The red curve in the Figure 9 represents the actual 3D coordinate values of the corner, while the blue curve represents the predicted 3D coordinate values of the corner. As can be seen from Figure 9, the trend of the two curves is basically the same, but there are obvious differences between them. It can be observed that there is a certain error between the predicted values and the actual values of the 3D coordinates of the corner, that is, the calibration accuracy is low when the traditional BP neural network is used for binocular camera calibration.

##### b: EXPERIMENTS ON CALIBRATION ACCURACY OF BINOCULAR CAMERA BASED ON TGSA-BP NEURAL NETWORK

In part a), the traditional BP neural network can be used to calibrate binocular camera, and its calibration accuracy is relatively low, and there is a large gap between the actual values and the predicted values of 3D coordinates of the checkerboard corner. Therefore, on this basis, the TGSA is introduced to optimize the BP neural network, so as to improve its calibration accuracy. As shown in Figure 10, compared with Figure 9, the comparison between curves can be seen more clearly. The BP neural network optimized by the TGSA improves the accuracy of camera calibration obviously. The gap between the true 3D coordinates and the predicted values of the corner of the checkerboard images is gradually reduced, and it has been greatly improved compared with that before the optimization within the visible scope of the naked eye. However, it can still be seen from Figure 10 that there is still a certain gap between the real 3D coordinate values and the predicted values of corner, and the accuracy of its calibration still needs to be further improved.

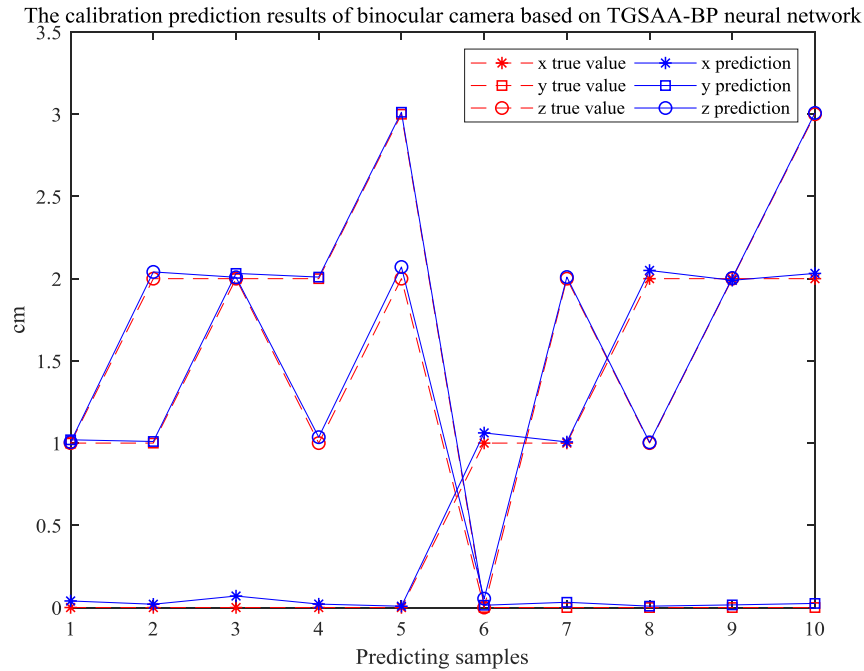


FIGURE 10. Test of calibration accuracy of binocular camera based on TGSAA-BP neural network.

## 2) EXPERIMENTS ON CALIBRATION ACCURACY OF BINOCULAR CAMERA BASED ON IGSA-BP NEURAL NETWORK

In order to further improve the calibration accuracy of binocular camera to meet the final requirements of the experiment, the algorithm is improved on the basis of the TGSAA. BP neural network can be better optimized, so as to achieve higher calibration accuracy of binocular camera. After the improvement of the TGSAA, the predicted values of the obtained 3D coordinates is shown in Figure 11. It can be seen from Figure 11 that the curves of the actual values and the calibrated predicted values of 3D coordinates fit perfectly. The IGSA has achieved a good effect, so as to optimize the BP neural network and further improve the calibration accuracy of binocular camera.

## 3) EXPERIMENTS OF CALIBRATION ITERATION SPEED OF BINOCULAR CAMERA

The time spent in the calibration experiment of binocular camera is analyzed from the number of convergent iteration of BP neural network. The fewer iterations of BP neural network, the better its convergence and the less time needed for binocular camera calibration. The iteration target value in the calibration process of binocular camera is set to 0.0001, and the training test is conducted with 1000 sets of data from the training set. As shown in Figure 12 (a), the number of iterations of the binocular camera calibration experiment based on BP neural network is 1860. As shown in Figure 12(b), the number of iterations for binocular camera calibration based on the TGSAA-BP neural network is 475. Compared with the traditional BP neural network, the convergence rate of the TGSAA-BP neural network is improved by 3 times. On this basis, in order to further accelerate the convergence

speed and reduce the calibration time of binocular camera, the TGSAA is improved. On the basis of the TGSAA-BP neural network, the quadratic optimization is carried out. As shown in Figure 12(c), the final iteration number of the binocular camera calibration based on IGSA-BP neural network is 85. Compared with Figure 12(b), the number of iterations is reduced by 82%. Compared with the binocular camera calibration based on the traditional BP neural network, the iteration speed is increased by about 20 times. In the process of calibration experiment, the BP neural network is optimized and a good result is obtained.

## 4) ANALYSIS OF CALIBRATION RESULTS OF BINOCULAR CAMERA

The calibration errors before and after BP neural network optimization are tested by 1000 training sets. In this experiment, the average error between the 3D actual coordinate values of the training set and the predicted coordinate values is adopted, that is, the mean value of the 3D coordinate errors of X, Y and Z. the mean error of calibration test for 60 training sets randomly selected is shown in Figure 13. It can be seen from the Figure 13 that the calibration accuracy of the binocular camera based on IGSA-BP neural network is the highest and good results have been achieved. Comparing the results before and after optimization, the accuracy of the camera calibration is the lowest when the traditional BP neural network is used before optimization. Its average error range is between 0.3mm and 0.7mm, which is large and relatively unstable. The calibration accuracy of TGSAA-BP neural network is around 0.1-0.4mm, and its stability has been significantly improved. The calibration accuracy of the binocular camera based on IGSA-BP neural network is about 0.02-0.06mm, and the error range is relatively concentrated and stable. It can

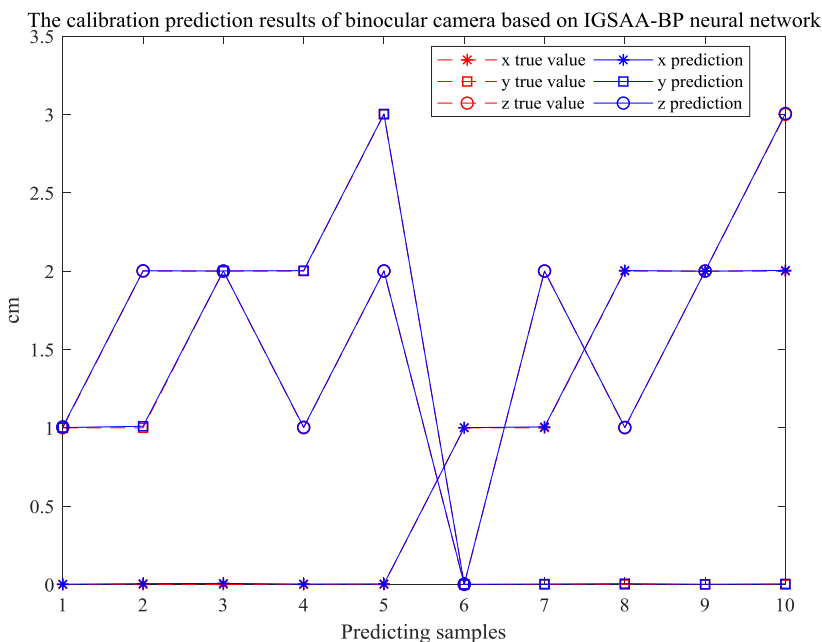


FIGURE 11. Test of calibration accuracy of binocular camera based on IGSA-BP neural network.

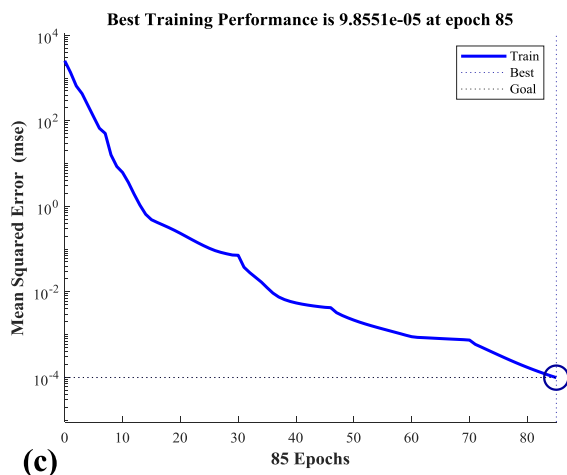
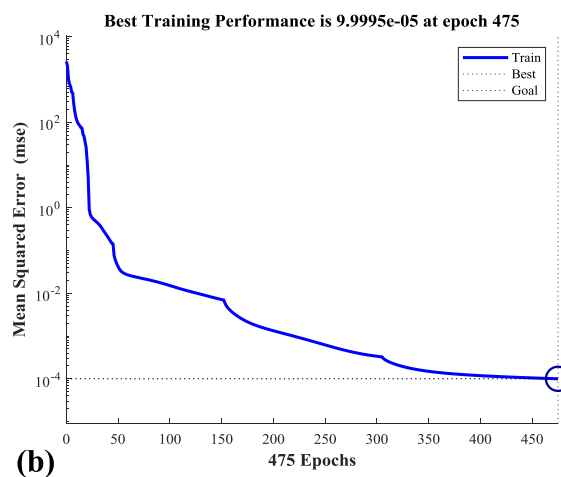
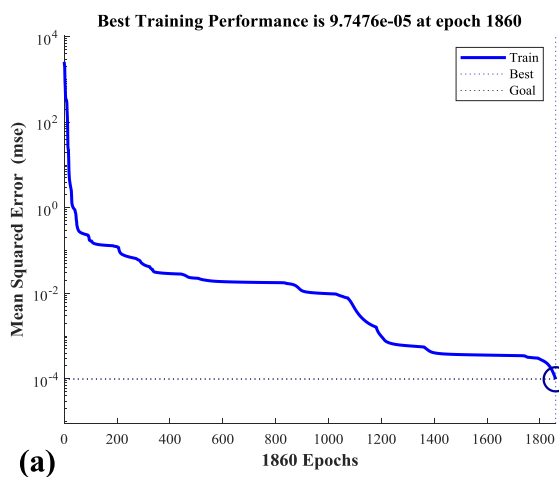


FIGURE 12. The number of iterations in the binocular camera calibration process: (a) The number of iterations of camera calibration based on traditional BP neural network; (b) The number of iterations of camera calibration based on TGSA-BP neural network; (c) The number of iterations of camera calibration based on IGSA-BP neural network.

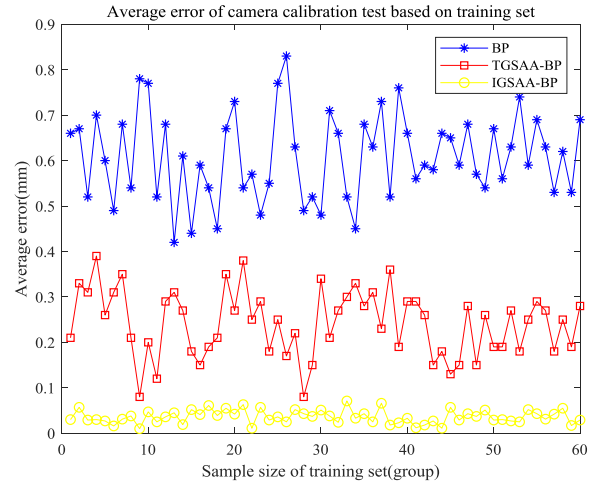
**TABLE 6. Prediction error of calibration of 3D coordinates for binocular camera.**

Group	Prediction error based on BP neural network (mm)			Prediction error based on IGSSA-BP neural network (mm)		
	x	y	z	x	y	z
1	0.60	1.12	0.63	0.01	0.01	0.06
2	0.71	0.93	0.41	0.05	0.09	0.02
3	0.73	0.62	0.85	0.07	0.02	0.05
4	0.82	0.96	0.36	0.02	0.03	0.04
5	0.58	0.52	0.71	0.03	0.03	0.01
6	1.12	0.34	0.54	0.01	0.02	0.03
7	0.75	0.36	0.98	0.06	0.03	0.01
8	0.52	0.84	1.05	0.04	0.06	0.03
9	0.91	0.16	0.41	0.01	0.02	0.01
10	1.28	0.45	0.58	0.05	0.04	0.06

be seen that the IGSA has achieved good results in the optimization of traditional BP network and greatly improved the calibration accuracy of binocular camera.

Based on the analysis of the X, Y and Z coordinate errors of the 10 test sets, the 3D coordinate errors of the binocular camera calibration using traditional BP neural network and IGSA-BP are shown in Table 6. As can be seen from Table 6, the error of each coordinate by using IGSA-BP neural network is obviously smaller than that using traditional BP neural network. In the process of camera calibration, the total mean error of the predicted values of 3D coordinate using traditional BP neural network is 0.71mm. The total mean error of the predicted 3D coordinates using the IGSA-BP neural network is 0.03mm. Compared with traditional BP neural network, the calibration accuracy of IGSA-BP neural network improved by 96%.

The calibration errors of X, Y and Z coordinates are fitted on the surface. According to Figure 14, it can be seen that the error of calibration prediction accuracy using IGSA-BP neural network is more concentrated. In the coordinates of X, Y and Z, the average accuracy of calibration prediction using IGSA-BP neural network is 0.036mm, 0.034mm and 0.025mm respectively. The average accuracy of calibration prediction of traditional BP neural network in X, Y and Z coordinate is 0.79mm, 0.63mm and 0.64mm respectively. Compared with the traditional BP neural network, the calibration accuracy of X, Y and Z coordinate with IGSA-BP neural network is improved by 95%, 94% and 96% respectively. As can be seen from Figure 14 (a), about 60% of the errors in the camera calibration prediction results based on traditional BP neural network are concentrated in the blue region, and the error range is about 0.3mm-0.7mm. Compared with



**FIGURE 13. Average accuracy of binocular camera calibration based on training set.**

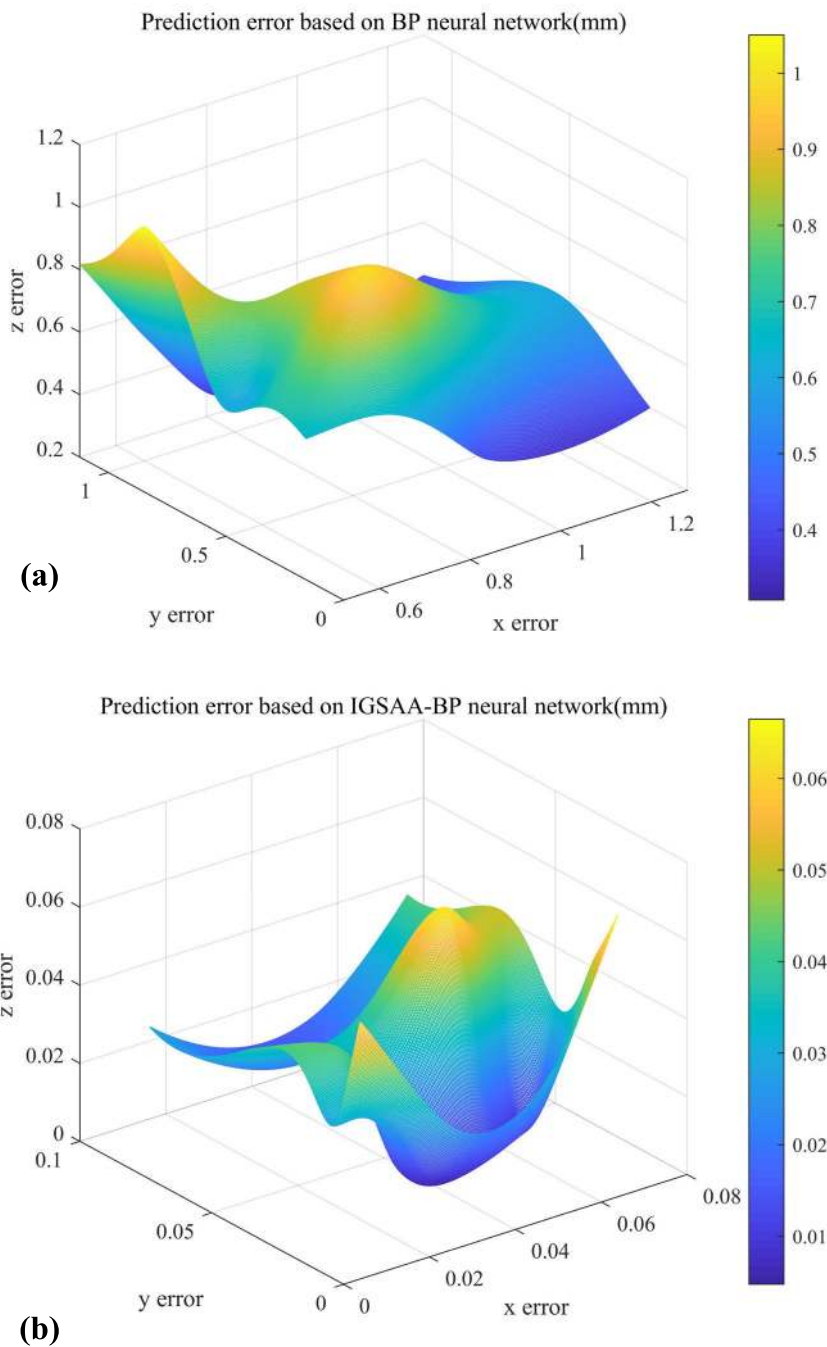
traditional BP neural network, the error range of IGSA-BP neural network in Figure 14 (b) is more concentrated and the error is smaller, ranging from 0.02mm to 0.06mm.

As shown in Figure 15, with the increase of evolutionary algebra, the fitness value of the optimal individual increases in a stepped manner. IGSA-BP network is superior to TGSAA-BP neural network in global optimization of binocular camera calibration. The fitness value of 125 is found when the TGSAA-BP neural network converged for 4 times to 15 generations. The fitness value 185 is found when the IGSA-BP neural network converged four times to 25 generations. It can be seen that the IGSA has significantly improved the optimization ability in camera calibration test. The IGSA has strong climbing ability and is not easy to fall into local optimization. It can be seen from the Figure 15 that the IGSA rises in a stepped manner, resulting in multiple conversions. Finally, the IGSA-BP network gets rid of the local convergence and shows strong adaptability.

**IV. DISCUSSION**

Black and white checkerboard is a common calibration object for camera calibration. How to quickly and accurately detect the corner of the checkerboard, extract the accurate 2D pixel coordinates and complete the homonymous corner match in the left and right checkerboard images is the problem that the current research is concerned about. Traditional corner detection and matched methods have poor robustness. Interference (such as slight occlusion) can cause the checkerboard corner detection to fail. The checkerboard images with a large tilt angle will also lead to the failure of corner detection. The checkerboard images are similar to each other, so the probability of a false match is much higher than that of normal images.

As early as 2012, Andreas Geiger et al. used the energy growth algorithm to detect and match the corner of the checkerboard images to solve these problems [47]. This



**FIGURE 14.** 3D coordinate accuracy prediction of binocular camera calibration based on test set: (a) The prediction error of 3D coordinates based on traditional BP neural network; (b) The prediction error of 3D coordinates based on IGSA-BP neural network.

method is robust and does not need to specify the size of checkerboard in advance. However, this method requires a large amount of calculation. At the same time, when the checkerboard images is polluted, the checkerboard lattice cannot be grown. The corner detection and matched algorithm proposed in this study overcomes the problem of low accuracy and large computation. However, this method requires the positions and shapes of the left and right

camera images to be similar when matching the checkerboard images.

BP neural network is one of the important methods for camera calibration. The accuracy of camera calibration based on BP neural network is discussed. In 2010, Jing *et al.* used BP neural network to complete the calibration of CCD camera [48]. The final calibration accuracy is around 1.2mm, which is not up to the actual calibration requirements. On the

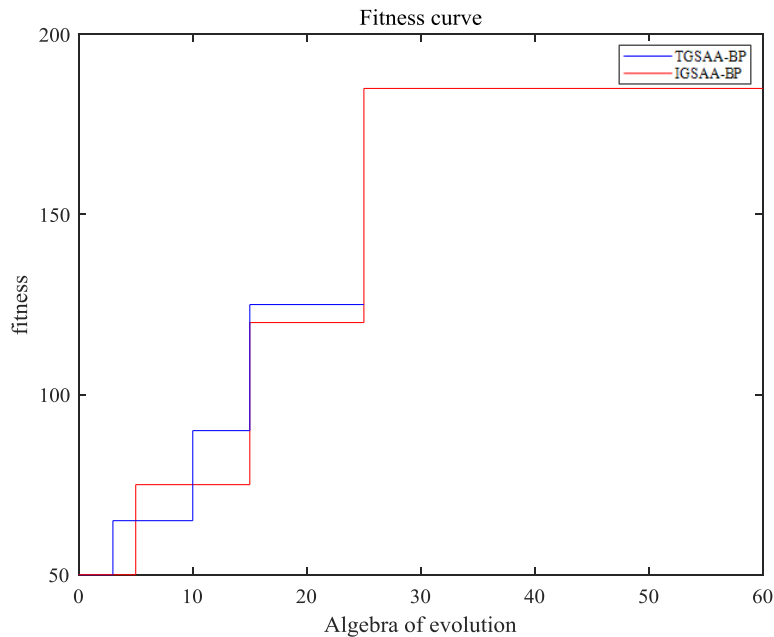


FIGURE 15. The curve of fitness and evolutionary algebra.

basis of BP neural network, Jiang *et al.* proposed using ant colony algorithm to optimize BP neural network [49]. This method improves the calibration speed of the camera, but does not improve the calibration accuracy. Munoz Rodriguez *et al.* introduced the genetic algorithm to optimize the BP neural network, and finally obtained the accuracy around 0.1mm [50]. Compared with the traditional BP neural network, the calibration accuracy has been greatly improved. At the same time, the stability of genetic algorithm is not satisfactory because it is easy to fall into the problem of local optimization. The calibration method of IGSAA-BP neural network adopted in this study can reach about 0.02mm in calibration accuracy. Compared with the traditional BP neural network has a great improvement, at the same time to improve the global optimization ability. The convergence speed is accelerated and the calibration time is reduced.

The calibration of binocular camera based on neural network depends on a large amount of data. Relatively speaking, under the same circumstances, when the amount of data in the training set is larger, the final calibration accuracy will be higher. Due to the limitation of experimental conditions, some experimental data are selected in this paper, so the actual calibration accuracy can be further improved.

## V. CONCLUSIONS

In this paper, an improved genetic simulated annealing algorithm is proposed to optimize BP neural network in the application of binocular camera calibration. Firstly, the corners of the left and right checkerboard images are detected by combining Gaussian scale space and Harris corner detection operator. On this basis, the method of combining point-to-point spatial mapping algorithm and grid motion statistics

is used to match the homonymous corner and obtain the 2D pixel coordinates of the homonymous corner. Finally, the success rate of the homonymous corner match reaches 100%. Then, according to the improved adaptability of crossover and mutation probability and annealing criterion, the genetic simulated annealing algorithm is improved. The improved genetic simulated annealing algorithm is used to optimize the BP neural network to calibrate the binocular camera. The average calibration accuracy of traditional BP neural network is 0.71mm. The average calibration accuracy of BP neural network optimized by improved genetic simulated annealing algorithm is 0.03mm. Compared with the traditional BP neural network, the calibration accuracy is improved by 96%. The iterative speed of binocular camera based on BP neural network optimized by improved genetic simulated annealing algorithm is 20 times higher than that of the traditional BP neural network. The global optimization capability of the whole calibration process is improved.

## ABBREVIATION

Back Propagation (BP)

Particle Swarm Optimization (PSO)

Scale Invariant Feature Transform (Sift)

Speed Up Robust Feature (Surf)

Improved genetic simulated annealing algorithm (IGSA)

Traditional genetic simulated annealing algorithm (TGSA)

BP neural network is optimized by improved genetic simulated annealing algorithm (IGSAA-BP)

BP neural network is optimized by traditional genetic simulated annealing algorithm (TGSAA-BP)



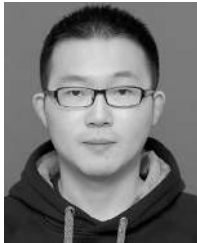
## ACKNOWLEDGMENT

The authors would like to thank Soochow University Robotics and the Microsystems Control Center.

## REFERENCES

- [1] P. Xi-Xi, L. Yong, and D. Xin-Ming, "3-D reconstruction of textureless and high-reflective target by polarization and binocular stereo vision," *J. Infr. Millim. Waves*, vol. 36, no. 4, pp. 432–438, 2017.
- [2] G. Xu, J. Chen, and X. Li, "3-D reconstruction of binocular vision using distance objective generated from two pairs of skew projection lines," *IEEE Access*, vol. 5, pp. 27272–27280, 2017.
- [3] S.-E. Shih and W.-H. Tsai, "Optimal design and placement of omni-cameras in binocular vision systems for accurate 3-D data measurement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1911–1926, Nov. 2013.
- [4] C.-L. Chen, C.-L. Tai, and Y. F. Lio, "Virtual binocular vision systems to solid model reconstruction," *Int. J. Adv. Manuf. Technol.*, vol. 35, nos. 3–4, pp. 379–384, Dec. 2007.
- [5] R. Xia, R. Su, J. Zhao, Y. Chen, S. Fu, L. Tao, and Z. Xia, "An accurate and robust method for the measurement of circular holes based on binocular vision," *Meas. Sci. Technol.*, vol. 31, no. 2, Feb. 2020, Art. no. 025006.
- [6] J. Sun, Y. Zhang, and X. Cheng, "A high precision 3D reconstruction method for bend tube axis based on binocular stereo vision," *Opt. Express*, vol. 27, no. 3, p. 2292, Feb. 2019.
- [7] G. Westheimer, "Three-dimensional displays and stereo vision," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 278, no. 1716, pp. 2241–2248, Aug. 2011.
- [8] Y. Tong, H. Li, J. Chen, M. Zhao, and L. Liu, "Dual-band stereo vision based on heterogeneous sensor networks," *Signal Process.*, vol. 126, pp. 87–95, Sep. 2016.
- [9] M. Wang, W. Wu, N. El-Sheimy, and Z. Xian, "Minimum sigma set SR-UKF for quadrifocal tensor-based binocular stereo vision-IMU tightly-coupled system," *J. Navigat.*, vol. 71, no. 6, pp. 1329–1343, Nov. 2018.
- [10] Y. Wang, X. Wang, Z. Wan, and J. Zhang, "A method for extrinsic parameter calibration of rotating binocular stereo vision using a single feature point," *Sensors*, vol. 18, no. 11, p. 3666, 2018.
- [11] W. Li, S. Shan, and H. Liu, "High-precision method of binocular camera calibration with a distortion model," *Appl. Opt.*, vol. 56, no. 8, pp. 2368–2377, Mar. 2017.
- [12] J. Zhang, X. Wang, M. Ma, F. Li, H. Liu, and H. Cui, "Camera calibration for anamorphic lenses with three-dimensional targets," *Appl. Opt.*, vol. 59, no. 2, pp. 324–332, Jan. 2020.
- [13] E. Shen and R. Hornsey, "Multi-camera network calibration with a non-planar target," *IEEE Sensors J.*, vol. 11, no. 10, pp. 2356–2364, 2011.
- [14] M. Yang, X. Chen, and C. Yu, "Camera calibration using a planar target with pure translation," *Appl. Opt.*, vol. 58, no. 31, pp. 8362–8370, Nov. 2019.
- [15] M. E. Deetjen and D. Lentink, "Automated calibration of multi-camera-projector structured light systems for volumetric high-speed 3D surface reconstructions," *Opt. Express*, vol. 26, no. 25, pp. 33278–33304, Dec. 2018.
- [16] S. Yang, Y. Gao, Z. Liu, and G. Zhang, "A calibration method for binocular stereo vision sensor with short-baseline based on 3D flexible control field," *Opt. Lasers Eng.*, vol. 124, pp. 30–45, Jan. 2020.
- [17] C. Cai, X. Weng, B. Fan, and Q. Zhu, "Calibration and rectification of vertically aligned binocular omnistereo vision systems," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, pp. 46–54, Dec. 2017.
- [18] D.-Y. Ge, X.-F. Yao, and Z.-T. Lian, "Binocular vision calibration and 3D reconstruction with an orthogonal learning neural network," *Multimedia Tools Appl.*, vol. 75, no. 23, pp. 15635–15650, Dec. 2016.
- [19] T. Zhen, Z. Qi, and X. Jiulong, "Large-scale camera calibration with neural network," *Acta Optica Sinica*, vol. 31, no. 4, 2011, Art. no. 0415001.
- [20] B.-M. Chung, "Neural-network model for compensation of lens distortion in camera calibration," *Int. J. Precis. Eng. Manuf.*, vol. 19, no. 7, pp. 959–966, Jul. 2018.
- [21] M. H. Rafiei, W. H. Khushefati, R. Demirboga, and H. Adeli, "Neural network, machine learning, and evolutionary approaches for concrete material characterization," *ACI Mater. J.*, vol. 113, no. 6, pp. 781–789, Dec. 2016.
- [22] R. Luo, T. Shao, H. Wang, W. Xu, X. Chen, K. Zhou, and Y. Yang, "NNWarp: Neural network-based nonlinear deformation," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 4, pp. 1745–1759, Nov. 2020.
- [23] L. D. Suits, T. C. Sheahan, H. Zhao, and L. Ge, "Camera calibration using neural network for image-based soil deformation measurement systems," *Geotechnical Test. J.*, vol. 31, no. 2, pp. 192–197, 2008.
- [24] D. O. Yongtae, "Camera calibration using neural network with a small amount of data," *J. Sensor Sci. Technol.*, vol. 28, no. 3, pp. 182–186, 2019.
- [25] B.-M. Chung, "Neural-network model for compensation of lens distortion in camera calibration," *Int. J. Precis. Eng. Manuf.*, vol. 19, no. 7, pp. 959–966, Jul. 2018.
- [26] C. An and Y. Zhi, "Self-adaptive neural network for binocular camera calibration," *Comput. Eng. Appl.*, vol. 45, no. 21, pp. 55–56, 2009.
- [27] D.-Y. Ge, X.-F. Yao, C. Hu, and Z.-T. Lian, "Nonlinear camera model calibrated by neural network and adaptive genetic-annealing algorithm," *J. Intell. Fuzzy Syst.*, vol. 27, no. 5, pp. 2243–2255, 2014.
- [28] D.-Y. Ge and X.-F. Yao, "Camera calibration and precision analysis based on BP neural network," in *Proc. Int. Congr. Image Signal Process.*, 2009, pp. 1–5.
- [29] W. Jin and Y. Zhou, "Study on calibration of binocular stereovision based on BP neural network with different layers," *Opt. Techn.*, vol. 41, no. 1, pp. 72–75, 2015.
- [30] J. H. Holland, *Adaptation in Natural and Artificial Systems*, vol. 6, no. 2. Ann Arbor, MI, USA: Optics Communications, 1975, pp. 126–137.
- [31] H. Yao and Z. Zhang, "Research of camera calibration based on genetic algorithm BP neural network," in *Proc. IEEE Int. Conf. Inf. Automat. (ICIA)*, Aug. 2016, pp. 350–355.
- [32] M. A. El-Shorbagy, A. Y. Ayoub, and I. M. El-Desoky, "Anovel geneticalgorithm based k-means algorithm for cluster analysis," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl. (AMLTA)*, 2018, pp. 92–101.
- [33] J. Liu, S. Yuan, and K. Jiang, "Neural network on camera calibration based on Zernike moment and PSO algorithm," *Journal of Optoelectronics Laser*, vol. 21, no. 9, pp. 1311–1314, 2010.
- [34] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, 1995, pp. 1942–1948.
- [35] C. G. Harris and M. Stephens, "A combined corner and edge detector," *Opt. Commun.*, vol. 3, no. 1, pp. 23–26, 1988.
- [36] Y. Deng, H. Wang, and S. Liu, "Analysis of the ship target detection in high-resolution SAR images based on information theory and Harris corner detection," *EURASIP J. Wireless Commun. Netw.*, vol. 8, no. 15, pp. 234–243, 2018.
- [37] H. Sheng, "Medical image registration method based on tensor voting and Harris corner point detection," *J. Med. Imag. Health Informat.*, vol. 8, no. 3, pp. 583–589, Mar. 2018.
- [38] J. Sporring, M. Nielsen, and L. Florack, *Gaussian Scale-Space Theory*, vol. 8. Amsterdam, The Netherlands: Springer, 1997.
- [39] G. D. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 3, no. 2, pp. 1150–1157.
- [40] B. Herbert, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2006.
- [41] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4181–4190.
- [42] Z. Kang and Z. Qu, "Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in lanzhou," in *Proc. 2nd IEEE Int. Conf. Comput. Intell. Appl. (ICCIA)*, Sep. 2017, pp. 155–160.
- [43] E. Kurtulu, A. R. Yildiz, S. M. Sait, and S. Bureerat, "A novel hybrid Harris hawks-simulated annealing algorithm and RBF-based metamodel for design optimization of highway guardrails," *Mater. Test.*, vol. 62, no. 3, pp. 251–260, Mar. 2020.
- [44] J. Zhang, S. H. Chung, and W. L. Lo, "Clustering-based adaptive crossover and mutation probabilities for genetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 11, no. 3, pp. 326–335, Jun. 2007.
- [45] Z. Li-Min and Z. Li, "Mobile WSN coverage method based on improvedadaptive genetic algorithm," *Comput. Appl. Res.*, vol. 36, no. 5, pp. 1510–1514, 2019.
- [46] J. Chen, L. He, Y. Quan, and W. Jiang, "Application of BP neural networks based on genetic simulated annealing algorithm for shortterm electricity price forecasting," in *Proc. Int. Conf. Adv. Electr. Eng. (ICAEE)*, Jan. 2014, pp. 1–6.
- [47] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 3936–3943.

- [48] J. Li, F. Yuan, and Z. Ding, "Linear CCD camera calibration of exterior attitude measurement system based on BP neural network," *Chin. J. Sci. Instrum.*, vol. 31, no. 5, pp. 1138–1141, 2010.
- [49] J. Xiang-Kui and W. You-Ming, "BP neural network camera calibration based on ant colony genetic algorithm," *Mach. Electron.*, 2013.
- [50] J. A. M. Rodríguez and F. C. Mejía Alanís, "Binocular self-calibration performed via adaptive genetic algorithm based on laser line imaging," *J. Mod. Opt.*, vol. 63, no. 1, pp. 1–14, Jan. 2016.



**LONG CHEN** is currently a Graduate Student with the School of Mechanical and Electrical Engineering, Soochow University. His research interests include surgical robot, augmented reality, and 3-D reconstruction.



**FENGFENG ZHANG** received the B.S., M.S., and Ph.D. degrees from the School of Mechanical and Electrical Engineering, Harbin Institute of Technology, in 2001, 2003, and 2009, respectively. He is currently an Associate Professor with the School of Mechanical and Electrical Engineering, Soochow University. He has authored over 30 articles and over eight inventions. His research interests include surgical robot, virtual reality, and biomechanics.



**LINING SUN** is currently the Director of the Robotics and Microsystems Center, and the President of the College of Mechatronic Engineering, Soochow University. He received the China National Funds for Distinguished Young Scientists. He has more than 300 academic articles being published and more than 20 patents of invention being authorized. His current research interests include micronano operational robot and equipment, advanced robot and control, and electromechanical integration equipment. He has received two National Science and Technology Awards Grade II and three Provincial Science and Technology Prizes Grade I.

...